

**Developing Models for the Data-Based
Mechanistic Approach to Systems Analysis:
Increasing Objectivity and Reducing Assumptions**

By David Mindham

Lancaster Environment Centre

Table of Contents:

Abstract - *1*

Authors Publications – *2*

Abbreviation of Terms – *3*

Mathematical Terms – *4*

Chapter 1: Introduction – *6*

1.9 Aim and Objectives – *14*

Chapter 2: Development of Arbitrary Sampling Technique for State Space Models – OBJECTIVES 1 - *16*

Chapter 3: Unification and DBM Alignment of (Multiple) State Dependent Parameter Methodologies – OBJECTIVE 2 & 3 – *27*

Chapter 4: Incorporating Arbitrary Sampling Technique to Dynamic Harmonic Regression – OBJECTIVE 4 – *51*

Chapter 5: Applications of Arbitrary Sampling Dynamic Harmonic Regression – OBJECTIVE 4 – *67*

Chapter 6: Application of the Data-Based Mechanistic Approach to Catchment Hydrology – OBJECTIVE 5 – *69*

Chapter 7: Conclusion and recommendations – *97*

Acknowledgements – *98*

References – *99*

Appendix – *104*

Abstract

Stochastic State-Space Time-Varying Random Walk models have been developed, allowing the existing Stochastic State Space models to operate directly on irregularly sampled time-series. These TVRW models have been successfully applied to two different classes of models benefiting each class in different ways.

The first class of models - State Dependent Parameter (SDP) models and used to investigate the dominant dynamic modes of nonlinear dynamic systems and the non-linearities in these models affected by arbitrary State Variables. In SDP locally linearised models it is assumed that the parameters that describe system's behaviour changes are dependent upon some aspect of the system (it's 'state'). Each parameter can be dependent on one or more states. To estimate the parameters that are changing at a rate related to that of it's states, the estimation procedure is conducted in the state-space along the potentially multivariate trajectory of the states which drive the parameters. The introduction of the newly developed TVRW models significantly improves parameter estimation, particularly in data rich neighbourhoods of the state-space when the parameter is dependent on more than one state, and the ends of the data-series when the parameter is dependent on one state with few data points.

The second class of models are known as Dynamic Harmonic Regression (DHR) models and are used to identify the dominant cycles and trends of time-series. DHR models the assumption is that a signal (such as a time-series) can be broken down into four (unobserved) components occupying different parts of the spectrum: trend, seasonal cycle, other cycles, and a high frequency irregular component. DHR is confined to uniformly sampled time-series. The introduction of the TVRW models allows DHR to operate on irregularly sampled time-series, with the added benefit of forecasting origin no longer being confined to starting at the end of the time-series but can now begin at any point in the future. Additionally, the forecasting sampling rate is no longer limited to the sampling rate of the time-series.

Importantly, both classes of model were designed to follow the Data-Based Mechanistic (DBM) approach to modelling environmental systems, where the model structure and parameters are to be determined by the data (Data-Based) and then the subsequent models are to be validated based on their physical interpretation (Mechanistic). The aim is to remove the researcher's preconceptions from model development in order to eliminate any bias, and then use the researcher's knowledge to validate the models presented to them. Both classes of model lacked model structure identification procedures and so model structure was determined by the researcher, against the DBM approach. Two different model structure identification procedures, one for SDP and the other for DHR, were developed to bring both classes of models back within the DBM framework.

These developments have been presented and tested here on both simulated data and real environmental data, demonstrating their importance, benefits and role in environmental modelling and exploratory data analysis.

Authors Publications

The author has written or contributed to the following pieces of work and they form part of the thesis, with the first two appearing in the main text and the rest supplied in the appendices:

Mindham, D., Tych, W., Chappell, N. 2018. Extended State Dependent Parameter modelling with a Data-Based Mechanistic approach to nonlinear model structure identification. *Environmental Modelling and Software*, 104, 81-93.

Mindham, D., Tych, W. 2019. Dynamic harmonic regression and irregular sampling; avoiding pre-processing and minimising modelling assumptions. *Environmental Modelling and Software*, 121, 104503.

Yu, Y., Katsoyiannis, A., Bohlin Nizzetto, P., Brotstrom-Lunden, E., Ma, J., Zhao, Y., Wu, Z., Tych, W., Mindham, D., Sverko, E., Barresi, E., Dryfhout-Clark, H., Fellin, P., Hung, H. 2019. Polycyclic Aromatic Hydrocarbons not declining in Arctic air despite global emission reduction. *Environmental Science and Technology*, 53 (5), 2375-2382.

Snell, M., Barker, P. A., Surridge, B. W. J., Benskin, C. M. H., Barber, N., Reaney, S., Tych, W., Mindham, D., Large, A., Burke, S., Haygarth, P. M. 2019. Strong and recurring seasonality revealed within stream diatom assemblages. *Scientific Reports*, 9 (1), 3319.

Magliano, P. N., Mindham, D., Tych, W., Murray, F., Nosetto, M. D., Jobbagy, E. G., Niborski, M. J., Rufino, M. C., Chappell, N. A. 2019. Hydrological functioning of cattle ranching impoundments in the Dry Chaco rangelands of Argentina. *Hydrology Research*, 50 (6), 1596-1608.

Abbreviation of Terms

AIC – Akaike Information Criterion
AR – Auto Regression
ARMA – Auto Regressive Moving Average
ARX – Auto Regressive eXogenous
ASDHR – Arbitrary Sampled Dynamic Harmonic Regression
DAR – Dynamic Auto Regression
DARX – Dynamic Auto Regressive eXogenous
DBM – Data Based Mechanistic
DHR – Dynamic Harmonic Regression
DLR – Dynamic Linear Regression
ESDP – Extended State Dependent Parameter
FFT – Fast Fourier Transform
FIS – Fixed-Interval Smoother
GRW – Generalised Random Walk
IC – Information Criterion
IRW – Integrated Random Walk
KF – Kalman Filter
LF – Likelihood Function
LTI – Linear Time Invariant
MC – Monte Carlo
ML – Maximum Likelihood
MSDP – Multiple State Dependent Parameter
MSI – Model Structure Identification
NARX – Nonlinear Auto Regressive eXogenous
NVR – Noise Variance Ratio
RIV – Refined Instrumental Variable
RW – Random Walk
SDP – State Dependent Parameter
TF – Transfer Function
TV – Time Varying
TVP – Time Varying Parameters
YIC – Young's Information Criterion

Mathematical Terms

General:

- y, Y – output
- u, U – input
- δ – time delay
- R^2 – Regression efficiency (Nash criterion)
- R_t^2 – Simulation efficiency measure
- t – time-step
- $\| \|$ - norm of

Transfer Functions:

- α, β, a, b – constant parameters
- del – pure time-delay
- p – operator
- g – system gain
- s – lap place transform operator
- T – time constant
- z^{-1} – backward transform operator
- Δt – sampling rate
- f – fast processes
- s – slow processes

State Space:

- a_t, b_t – time-varying parameters
- s – states (a variable that a parameter is dependent on)
- n, N – total number of samples
- x_t – state variable
- \mathbf{x}_k – state vector
- k – knot
- q – random walk type
- $\boldsymbol{\eta}$ – system disturbance
- \mathbf{F}_k – transition matrix
- \mathbf{G}_k – noise matrix
- Δ_k – temporal distance between knots
- \mathbf{H}_k^T – observation matrix
- e_k – observation disturbance
- \mathbf{z}_t^T – observation vector
- $\mathbf{p}_t, \mathbf{p}_k$ – parameter vector
- T – transform
- i – parameter number
- m – monkeys?
- \mathbf{Q} – state covariance matrix
- σ^2 – variance of observation disturbance
- ζ_k – approximation knot

ESDP validation:

- \tilde{a} – parameter map for the a parameter
- a_e – time varying parameter for estimation period
- a_v – time varying parameter for validation period
- T_e – temporal samples for estimation
- T_v – temporal samples for validation
- \hat{y}_t – estimated output

Mode Structure Identification (ESDP):

- R – regressors
- S – maximum number of states
- M_{max} – maximum model order
- Δ_{max} – maximum time-delay
- N_{CM} – number of candidate models
- N_{BM} – number of basic models
- n_{par} – effective number of parameters
- dt – subsampling ratio
- N_s – total number of states used in candidate model
- Q_t – streamflow
- R_t – rainfall

Dynamic Harmonic Regression:

- T_t – trend component
- C_t – cyclic component
- S_t – seasonal component
- ω_i – fundamental and harmonic frequencies associated with the seasonality
- f_i – frequencies associated with the longer cyclical component
- $\alpha_{i,t}, \beta_{i,t}$ – stochastic time varying parameters
- R_s – rabbits?
- R_c – more rabbits?
- e_k – model residuals
- Sk – estimated seasonal component
- R_s^2 – proportion of the data explained by seasonal or cyclic component

Chapter 1: Introduction

1.0 Introduction

There is a vast range of environmental systems, all of which are different in terms of phase, medium, mechanisms and timescales. However, many are essentially the same overall system involving the transport of 'something' from 'somewhere' to 'somewhere else'. That 'something' can be interpreted as an input into a system, which is transported to 'somewhere else' and then usually treated as the system's output. The 'something' is either a pollutant, a nutrient, energy or a resource, the sources of which can be natural, man-made, or a combination of both, while the 'somewhere' and 'somewhere else' lie within natural environments or man-made environments. Typically, both the 'something', the 'somewhere else' and, sometimes even the transport medium, are related to other environmental systems, often crossing the traditional disciplines such as hydrology, chemistry or social sciences. Coupled with the fact that environmental systems are constantly changing with time, this makes it very difficult to fully understand individual environmental systems as they are interlocking, overlapping systems which never repeat any specific behaviour.

While the underlying transport mechanisms of a system do not change, their rates and dominant transport processes do. This is due to the variables in the systems constantly changing with time, e.g. temperature, rate of input from source(s), conditions of the transport medium. This constantly changing state of the system means the exact same state will only ever happen once; there will be many occasions where/when the state is similar, but never the same.

The overlapping and interlocking nature of environmental systems means the overall relationship between the input to the system and its output is never linear; even if the 'something' from the source is at a constant rate, all the constantly changing variables in the system that lead to transportation to 'somewhere else' means the output rate will always vary and have natural constraints. Therefore, environmental systems are non-linear systems. However, depending on the timescales used to investigate these systems it is often possible to approximate them well using linear difference or differential equations.

Another issue is our inability to view all the variables, where incomplete information makes the system appear less "well behaved" or even chaotic. Because it is currently impossible to observe and measure (at least without damaging or destroying part of the system) all aspects of any environmental system, we cannot gain a full understanding of how it works. Thus, mathematical modelling, understood as building numerical approximations of the processes involved, has been adopted to attempt to fill in any gaps in knowledge and improve our understanding. Mathematical modelling is about deriving equations that describe a system's typical behaviour; the level of complexity in the equations reflects the number of assumptions about the structure, the variables and the parameters incorporated in the model. Two general approaches are used for deriving equations that describe a system.

The first is to generate theories, build models based on them and then compare to reality (model output versus observed). The main issue with this approach is the potential for 'researcher bias', the theories and subsequent model could reflect the researcher's views more than the actual system and any discrepancies between the model and the observed could be wrongly dismissed, or worse, the model could be modified with an indescribable or unattributable variable or mode of operation that accounts for the discrepancies but has no physical meaning. This class of models, essentially hypotheses on how the system works fitted to data by tuning the parameters, is often called physics-

based modelling (for example: VanderKwaak and Loague, 2001. Anderson and Segall, 2011. Zhang et al., 2018).

The second approach is to build models for a system based solely on data from that system. Instead of building specific models based on theory, general model classes are adopted and then identified using observed data. These general models, usually known as Black Box models (for example: Rajurkar et al., 2004. Ando et al., 2000) describe the transition of the selected input into the output and are in this work linear dynamic approximations including scaled terms, which in turn include delayed variables or their rates of change. The non-linearity in these linear dynamic models is incorporated or approximated through the parameters, which can be time-varying or dependent upon other system variables and is the key aspect of this approach to modelling. Simply put, an environmental system is first assumed to be a simple linear dynamic input/output relationship with an associated linear equation describing it, and then the equation is modified to incorporate the non-linearities observed in the data. To avoid ‘researcher bias’, this modification step is done using statistical methods that provide the researcher with the statistically most likely equation to describe a system based on the available data. The work present here is based on this second approach.

1.1 Data-Based Mechanistic Philosophy

This latter approach can be expanded to become Data Based Mechanistic (DBM) modelling (Young and Lees, 1993 and Young, 1999) which uses a two-stage approach. The first stage involves the statistical identification and estimation of a model from the available data, ‘data-based’, that could represent the stochastic dynamic system. The underlying equations of this model are chosen from a general and widely applicable class of models where the structure and parameters are identified and estimated from the available data. The statistical identification and estimation processes are made as objective as possible to avoid researcher bias, with the idea being to remove the researcher as much as possible from the model creation process and allow the observed data to create the model that represents it. This first stage typically results in a simple low order, parametrically efficient (lowest number of parameters required to capture the key dynamics), parsimonious model that represents the dominant dynamic mode(s) of the system observed in the data. The second stage of this modelling procedure is to ascertain if this model is physically sensible. Models derived in this manner will always be mathematically ‘sensible’ and so standard ways of determining model validity will not all apply. The model’s validity must be considered in ‘mechanistic’ terms so that the estimated parameters and thus the overall model, are physically interpretable within the confines of the system under study.

1.2 Linear Transfer Function Models

A simple and well-established linear dynamic model is described by the equation (1.1):

$$\dot{y}_t = -\frac{1}{\alpha}y_t + \beta u_{t-\delta} \quad (1.1),$$

where, y_t is the output at time t , u_t is the input, δ is the pure time-delay (between the input and the output), α and β are constant real parameters and \dot{y}_t is the time derivative of the output y_t .

This ordinary differential equation form is often used in time series and control engineering context transformed using one of the commonly used operators such as Laplace differentiation operator s or backward shift operator z^{-1} (which is applied to the difference equation approximating (1.1) in discrete time).

For the continuous time model using Laplace operator s and ignoring the initial conditions (common practice in long time series practice) (1.2) becomes:

$$sY(s) = -\frac{1}{\alpha}Y(s) + \beta U(s)e^{-s\delta} \quad (1.2),$$

where $Y(s)$, usually denoted Y is the Laplace transform of the series y_t , same with $U(s)$ and $e^{-s\delta}$ is the Laplace transform of the pure time delay by δ time units. A simple algebraic manipulation of (1.2) produces form (1.3).

$$Y = \frac{ge^{-s\delta}}{Ts+1}U \quad (1.3),$$

where, Y is the output, U is the input, s is the Laplace operator, g is the system gain (1.3a) determining the scale of magnitude of the output in relation to the input, and T is the time-constant (1.3b), which is the time it takes for the system to respond to a step change in the input.

$$g = \frac{\beta}{\alpha} \quad (1.3a), \quad T = \frac{1}{\alpha} \quad (1.3b).$$

Manipulation of (1.1) into an approximate difference equation and applying the backward shift operator z^{-1} defined as in $z^{-1}y_{k\Delta t} = y_{(k-1)\Delta t}$, k being the sample number in the uniformly sampled series – every Δt – produces a similar transfer function form.

$$Y = \frac{bz^{-d}}{1+az^{-1}}U \quad (1.4),$$

where, z^{-1} is the backward transform operator, d is the pure time-delay measured in samples, and a and b are fixed parameters, where the gain and time-constant are defined as (1.4a) and (1.4b) respectively.

$$g = \frac{b}{1+a} \quad (3a), \quad T = -\frac{\Delta t}{\ln(a)} \quad (3b),$$

where, Δt is the sampling rate in this difference equation.

For example, in hydrology rainfall-runoff models are often described by TF models (Beven, 2012) and modelling the response to Storm Desmond at Temple Sowerby in the Eden catchment, UK, as a first order linear dynamic system resulted in a TF model (discrete time) with a pure time-delay of 17 samples, a steady state gain of 16.6, and a time constant of 37.9 samples (Figure 1.1). As the data was observed in 15 minute-intervals, the pure time-delay translates into 4.25 hours and the time constant translates into 8.4 hours, while the gain's scaling is due to the catchment's contributing area and the units used.

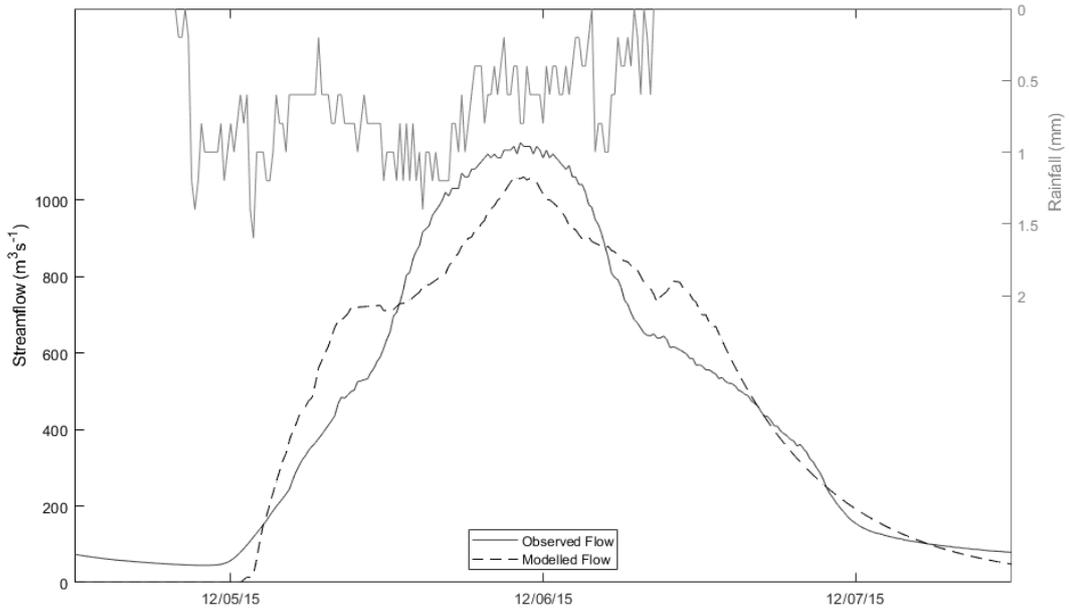


Figure 1.1. Comparing modelled response to storm Desmond at Temple Sowerby in the Eden, UK.

The TF model can be expressed with an additional complexity by increasing the order (number of parameters), where the first order model in (1.3) becomes (for example) a second order discrete time model (1.4).

$$Y = \frac{b_0 z^{-1} + b_1 z^{-2}}{1 + a_1 z^{-1} + a_2 z^{-2}} U \quad (1.4)$$

One physical interpretation of this transfer function is to view it as a parallel connection of two first order models (1.5), where one describes the combination of faster transport processes and the other describes the group of slower processes (Young, 1992).

$$y_t = \left(\frac{b_f z^{-1}}{1 + a_f z^{-1}} + \frac{b_s z^{-1}}{1 + a_s z^{-1}} \right) u_{t-\delta} \quad (1.5),$$

where subscript f refers to the fast processes and s to the slower processes.

Using the above hydrological data, a second order TF model was identified (Figure 1.2) with a pure time-delay of 15 samples, steady state gains of 99.2 and -67.4, and time constants of 1.6 and 27.4. This means a pure time-delay of 3.75 hours with time constants of 0.4 hours for fast processes, and 6.9 hours for slow processes. While adding the complexity leads to data better explained by the model (better fit), the negative gain means that the physical interpretation as a second order parallel process is not viable in the DBM sense and the 2nd order model is rejected on this basis.

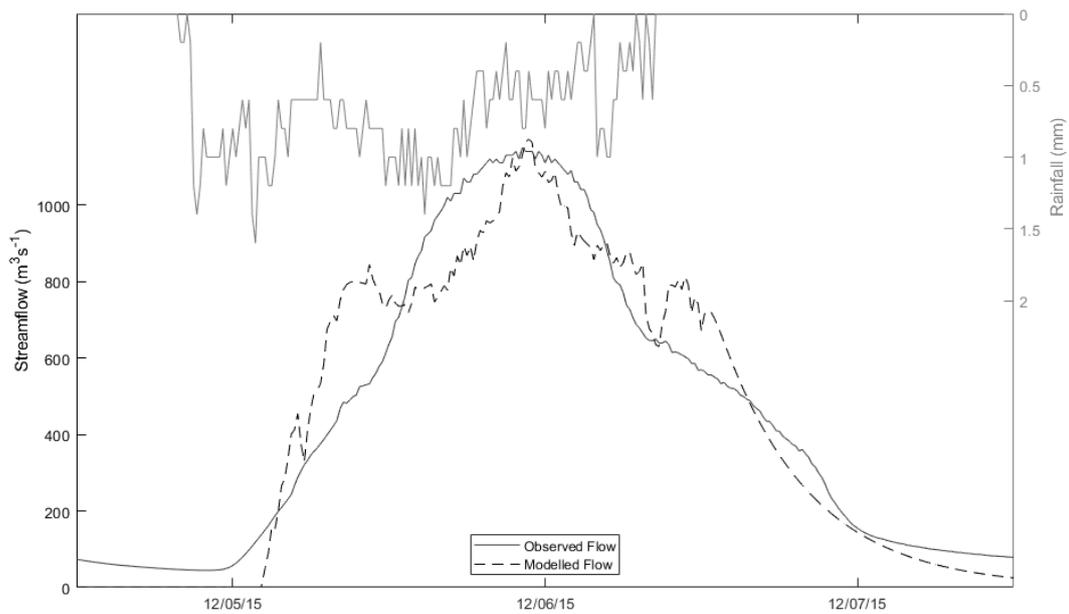


Figure 1.2. Comparing second order modelled response to storm Desmond at Temple Sowerby in the Eden, UK.

Increasing the order of the TF models does increase the amount of detail captured by the model: first order model's output (Figure 1.1) is smoother (simpler) than the second order model (Figure 1.2), as there are more parameters to account for the variation of the data. There is increased likelihood of over-parameterisation with the parameters becoming non-physically interpretable. This leads to a 'balancing act' between having the model fitting data better and preserving physical interpretability. In addition, increased model complexity will usually mean that it will not apply (or fit well) to a different data set, often from the same location.

Environmental systems are generally nonlinear, and, so while using linear models to describe them may seem incorrect, or at least risky, the aim of modelling using the DBM approach is to consider, detect or quantify the dominant processes occurring within the system (Young, 1999a). Linear models, particularly the TF models described here, when linked with the statistical identification and estimation tools described in the next section, can be seen as quantifications of input/output relationships. The TFs are operators mapping the input data series onto the output data series, and the associated statistical tools quantify the quality of this mapping. When using multiple input single output models they can indicate which input is most dominant, and which input is least important in terms of output generation. These tools also support model reduction, by identifying the dominant dynamic modes of the system.

A visualisation of model output and the observed output variable, as in Figures 1.1 and 1.2, can show exactly when there is no relationship between input and output, providing a useful and difficult to quantify diagnostic. This helps to formulate the questions of why this happens, where the answers could lead to interesting insights into the system under study, or point to problems with the measurement techniques, such as poor sensor locations.

Some linear constant parameters TF models (such as rainfall-runoff) do tend to get worse in terms of model fit and interpretation as time span increases while the sampling rate is kept the same. This is because the changes and nonlinearities in the system become more apparent as the slower

processes become more visible as time span increases. In addition, the estimated parameters of these models describe the ‘average behaviour’ of the system for the full time-series; including favouring the part of the time-series with the largest changes or highest amplitudes in the input/output relationship because of the estimation based on the variance or quadratic norm of residuals. For example, in hydrology, a TF model of a single rainfall event in the example catchment above, will be very good and will account for >90% of the observed variance as the timescale is very small, 2-3 days, so only looking at catchment dynamics for that small snapshot in time. But when looking at a years’ worth of rainfall events at the same temporal resolution, then all the slower catchment dynamics such as annual cycles in weather are present in the data and the TF model may only account for ~35% of the observed variance, with the high flows being underestimated and the low flows being overestimated.

1.3 Statistical Identification and Estimation Tools for TF Models

In this work, the Refined Instrumental Variable (RIV) estimation procedure is used to identify the TF model structure (Young et al., 1980) and estimate its parameters (Young and Jakeman, 1979), and is used in its MATLAB implementation using the CAPTAIN Toolbox (Young et al., 2007).

Model structure identification of TF models uses several statistical criteria to evaluate the statistical likelihood a specific model structure is the best for explaining the observed data. The identification criteria, typically used with RIV are Young’s Information Criterion (YIC, Young, 2011) and R_t^2 (Beven, 2012, chapter 4), the latter - a variant of the Nash criterion (R^2 , Nash and Sutcliffe, 1970). Both R^2 and R_t^2 use variance of model residuals to evaluate models and in this thesis R_t^2 is used to compare output data with a full dynamic TF simulation output, and R^2 is used to compare the one step-ahead predictions with the output data. These same criteria are also used to evaluate parameter estimation.

The model structure identification procedure uses a ‘brute force’ approach, where all possible combinations of model structure (from first order to a specified maximum order) are normally evaluated and the most likely model structure is the one with the highest R_t^2 or lowest YIC. R_t^2 is more a measure of data fitting, while YIC balances data fitting with keeping the model order (and hence parametric uncertainty) low and so both will sometimes identify different model structures for the same data set. Ultimately, the choice of criteria depends on the aims of the modelling exercise.

1.4 Nonlinear Models

For a linear dynamic model to incorporate the nonlinearity of a system, its parameters need to vary with time, where the extent of the nonlinearity determines the rate at which the parameters need to change. A simple time-varying model, and thus nonlinear model, is equation (1.6), where the output at a particular time-step is the sum of the output at the previous time-step plus the input at a time-step lagged from the current. This type of model that uses past outputs to estimate the current output is known as a recursive model.

$$y_t = a_t y_{t-1} + b_t u_{t-\delta} \quad (1.6),$$

where, a_t and b_t are the time-varying parameters, which can either simply vary in time (which can be estimated using the constraint of smoothness) or be functions of other measured or known variables.

The physical interpretation of this model is still the same as for the linear TF model described above, except every time-step will have its own estimates of the steady state gain and time constant. The

recursive time-varying nature of this model means it will typically explain >98% of the observed variance, no matter the timescale. While this has obvious advantages over the linear models with their fit often declining with increasing timescale, it means model fit (percentage of data variance captured by the model) cannot be used in validation. All the parameters across the time-series must be physically interpretable (the DBM principles apply strongly here) for the model to be accepted as a good approximation of the system. The parameters of these time-varying models can be estimated using a Kalman Filter (KF, Kalman, 1960) and Fixed-Interval Smoother (FIS, Ait-El-Fquih and Desbouvries, 2008) combination (detailed in Chapters 2, 3 and 4).

The use of KF/FIS estimation introduces a hyper-parameter (this is a parameter used to aid estimation of the model parameters) called Noise Variance Ratio (NVR, Young, 2011, chapter 17), which essentially controls the speed at which the parameters can change (or the relative time-scale of these changes). This control allows the researcher to adjust the process(es) speed the parameters are responding to. For example, a low NVR means the parameters change slowly over time and can be related to the dominant slow process(es) within the system, while a high NVR means the parameters can change quickly and thus reflect the dominant fast process(es) of the system.

This type of nonlinear model applies for systems where the system dynamics are changing at a rate slower than the system variables (input/output). However, many environmental system dynamics change at a rate commensurate with the system variables and capturing such fast changes are difficult using the previously explained method.

This next type of nonlinear model uses the same equation (Eq. 6 for the first order example) as before, but the parameters are said to be dependent on some variable within or outside the system; i.e. the input or output at previous time-steps or another variable altogether. This is known as State Dependent Parameter (SDP, Young, 2000) modelling (1.7).

$$y_t = a(s_{t,1})_t y_{t-1} + b(s_{t,2})_t u_{t-\delta} \quad (1.7),$$

where, $s_{t,x}$ is the 'state' the 'parameters' a and b are 'dependent' on. The state for each parameter can be the same variable or different.

Just like with the TF model, higher orders can be considered (such as 1.8), but this can lead more quickly to over-parameterisation and resulting in non-physical parameter values as more parameters are being estimated.

$$y_t = a_1(s_{t,1})_t y_{t-1} + b_1(s_{t,2})_t u_{t-\delta} + a_2(s_{t,3})_t y_{t-2} + b_2(s_{t,4})_t u_{t-\delta-1} \quad (1.8).$$

It should also be noted that each parameter can be dependent on more than one state and this led to the development of Multiple State Dependent Parameter (MSDP, Tych et al., 2012) modelling, which would assume each parameter can be dependent on two states. In theory it could handle more, but the subsequent analysis of the model gets tricky, particularly in visualisation; one state gives a 2-D picture of state vs parameter, using two states gives a 3-D picture of state1 vs parameter vs state2, and thus m states requires a $(m+1)$ -D picture.

In order to estimate parameters that need to change at a rate commensurate with the system variables, the estimation of each parameter needs to be done in the state space domain of $s_1 \dots s_4$ respectively for $a_1 \dots b_2$ and not the time domain. The time domain is where the modelling is done in the normal temporal sequence ($t=1,2,3,4,\dots,n$). In the state space domain, the modelling is done along a trajectory in the state space of $s_1 \dots s_4$ (for the case of 1.8) using an arbitrary sequence not based on time, and in this case, the sequence is determined by the ascending order of a metric of the state. This is done for each of the parameters separately. In the time domain, the parameters

would need to change rapidly to match the rapid dynamics in the time-series, but in the state space domain the parameters would change smoothly to match the smoothed dynamics along their trajectory in the state space. After the parameters have been estimated they are put back into temporal order as a functional expression (for more detail see Chapter 2).

Unfortunately, the original method of smoothing in the state space domain was based on fixed differences between the states' samples, while the arbitrary sequence of sampling was based on the sequence of states sorted in the order of each of $s_1 \dots s_4$, thus limiting the number of states driving each parameter to one. For each parameter, the distance between each state sample was fixed, when, in reality the distance needs to be based on the state value of each sample. This generalisation of the original algorithm forms objective one of this thesis.

1.5 Statistical Estimation Tools for Nonlinear SDP Models

As this subheading suggests, when using these nonlinear models, the researcher only has the statistical methods for estimating model parameters. The model selection criteria are largely absent, mainly because of the use of time varying parameters. For example when using a simple Random Walk model within the KF/FIS framework applied in this work, the number of degrees of freedom for the model will vary between $N-1$ for $NVR=0$ (N – the number of samples) right down to nearly zero for very high NVR values. So while estimation fell under the DBM remit: using maximum likelihood (Hannan, 1960, chapter 7) for NVR estimation and Least Squares (effectively R^2) for parameter estimation, the model structure was still down to the researcher's discretion and assumptions, and as mentioned earlier the aim of DBM is to reduce the chance of researcher bias by only letting the data determine the model structure, and thus SDP and MSDP did not strictly follow the DBM philosophy they were designed under. This leads to objective two: the development of a statistical identification tool for SDP models with a purpose-built objective function. Importantly, model structure identification for SDP models is more than identifying the number of parameters to estimate, but also identifying which system variables those parameters should depend upon, a complex and novel problem.

1.6 Non-uniform and Irregular Sampling Problem

For many environmental systems, it is either extremely difficult, too expensive or just not possible to collect environmental data in a uniform way. This leads to data-series with either large gaps or an irregular time-base. This causes problems for modelling environmental systems as many identification and estimation tools require uniformly sampled data to function properly. This has led to various data reshaping techniques that take the irregular sampled data and extrapolate it on to a regular time-base. While this allows environmental data to be analysed and modelled, this analysis falls outside of the DBM philosophy as the data undergo some form of modification and it is no longer just the data defining the models but is also influenced by the modification technique with a strong potential for injecting artefacts into the data.

Regularising the data sampling intervals can range from the interpolation (Davis, 1975) of a few missing values, to the complete temporal rescaling of a time-series. In both cases missing data values are estimated from the existing ones and the underlying models used for this estimation can inject assumptions into the data. While the interpolation and sampling regularisation techniques may be very sophisticated, they still are filling in the blanks with the best guess, bluntly put, they are making up the data. The DBM philosophy states the data should define the model, so logically, if there is no data, there should be no analysis because there is nothing to analyse.

This leads to objective three, development of a technique based on existing algorithms for handling data on an irregular time-base that can be incorporated into any univariate model that relies on KF and FIS algorithms as they have a wide range of applications.

1.7 Dynamic Harmonic Regression

An incredibly useful modelling approach that is plagued by the frequently encountered irregularly sampled time-series but only works for uniform time-bases is Dynamic Harmonic Regression (DHR, Young et al., 1999). DHR involves the deconstruction of a signal (time-series in the case of environmental data) into four component parts;

- Trend, the underlying trajectory of the signal.
- Annual cycle, most environmental systems are dominated by this.
- Other cycles.
- Noise, this should hopefully be uncorrelated white noise with Gaussian distribution.

DHR is crucial for studying patterns in environmental systems, understanding the dominant cycles that control the system, whether it is the annual cycle or some other, and for forecasting; such as predicting pollutant levels.

DHR utilises a KF/FIS algorithm, which means the data has to be brought into regular spacing before use, but also means the arbitrary sampling technique developed for objective three could be applied.

Objective four is the modification to the DHR methodology to incorporate the irregular sampling technique and to demonstrate the benefits over the current methods.

1.8 Role of DBM in Data Analysis

While there are a wide range of tools available for the DBM modelling of environmental systems (such as those found in the CAPTAIN Toolbox, Young et al., 2007), DBM is not about feeding data into a chosen modelling algorithm and being given 'the answer'. The tool chosen must not only be appropriate for the modelling exercise and its aim, but also suitable for the available data. This requires experience with the various tools to know their benefits and their limits, and the type of data (typically in terms of timescale and sampling rate) they work best for. The trap here is not to let experience and assumptions drive the data analysis, when it should be the data and the pure mechanistic interpretation that drives the analysis.

The overall role of the DBM philosophy in data analysis is to let the data drive the systems analysis, while the researcher searches for physical meaning in all model outputs and dismisses outputs that have no physical basis. This leads to objective five, the application of the DBM philosophy and available (and newly developed) tools to environmental systems to investigate their nonlinear relationships and drivers.

1.9 Aim and Objectives

The aim of this thesis is to develop modelling tools for use within the DBM framework and enhance existing tools where appropriate and demonstrate their value. The thesis now tackles the presented objectives, reiterated here:

1. Develop a methodology for handling irregularly sampled data that can be applied to all the univariate KF/FIS based models (such as DLR, DHR, DAR).

2. Generalise the (M)SDP methodologies: improving their estimation and unifying them under one term, Extended State Dependent Parameters (ESDP).
3. Develop a statistical based method for identifying ESDP model structures including which states the parameters should depend upon – further aligning ESDP with the DBM philosophy.
4. Apply the newly developed irregular sampling algorithm (from 3) to the DHR method for signal decomposition and demonstrate the benefits by applying to environmental research questions.
5. Application of the DBM approach, including the newly developed tools, to catchment hydrology.

Chapter 2: Development of Arbitrary Sampling Technique for State Space Models – OBJECTIVE 1

As previously mentioned, state space modelling approaches for signal analysis and parameter estimation are powerful tools but they require uniformly sampled data. Here a technical note (**paper 1**) was drafted (not published) that develops a robust method for handling irregular sampled data (termed arbitrary sampling) and demonstrates its effectiveness.

2.1 Introduction

Irregular sampled time-series data can occur for a number of reasons, the chief amongst them being by design. In environmental science when measurements are difficult, expensive or time consuming the methodology is usually adapted so measurements are only taken when the system under study reaches certain conditions and this leads to irregular sampling (for examples see Madrid and Zayas, 2007, Raghavan et al., 2006 and Tandeo, et al., 2011). In process control, navigation and tracking irregular time sampling intervals are used to monitor a system (for examples see Penarrocha et al., 2012, Xue-bo et al., 2012 and Yan et al., 2011), whereas in networked control systems packet losses and network-induced time delays cause randomly missing packets of information and thus produces irregular sampled time-series (Shi and Fang, 2010). In remote exploration, adaptive sensing is used to increase the number of measurements taken over areas of anomalous terrain and results in a generally uniform time-series with sections of non-uniformity (Thompson et al., 2013). Additional reasons for irregular sampled time-series data is from equipment malfunction, losses from data transfer from remote devices and the removal of anomalous measurements during quality checks. It has to be highlighted at this point that the aim here is not to interpolate the missing samples or create a uniformly sampled time series, but to create a recursive approximation algorithm with the regularising properties of the Kalman Filter and Fixed Interval Smoothers. The recursive approach provides an attractive alternative to sampling regularisation and spline based techniques (Gu, 2013). Especially with long data sets, the traditional approximation techniques usually require operations on large matrices of dimensionality growing with the square of the length of the series (Bellman, 1957). This dimensionality curse and the associated numerical problems make the development of suitable recursive and robust algorithms a necessity in some applications.

2.1.1 Recursive Signal Processing

The Kalman Filter (KF) is a powerful and popular tool for estimating the underlying signal (state) of a linear dynamic system by using measurements linearly related to the state but influenced by white noise (Grewal and Andrews, 2015). The original KF was designed for uniformly sampled time-series, non-uniformly sampled time series KF versions are described in Penarrocha et al., 2012, Xue-bo et al., 2012, Shi and Fang, 2010, and Li, et al., 2008 and involves inferring the missing data by performing only the prediction step of KF/FIS. The irregular sampling arises also in the context of continuous time KF related methods (Fang and Callafon, 2012) - this is closely related to the Generalised Random Walk process used as the main example here. The Fixed-Interval Smoother (FIS) uses the output from KF to further refine the estimate of the state (Ait-El-Fquih and Desbouvries, 2008) based on the entire data-set. Less work has been done on non-uniform time-series smoothing and so currently there is only an Expectation-Maximisation based algorithm which infers missing data points of irregular sampled time-series (Raghavan et al., 2006, Tandeo, et al., 2011, Yuz et al., 2011, and Isaksson, 1990), and so does not address the aim of the presented algorithm.

2.1.2 Scope

This note describes, with examples, a generalised recursive state-space method for using irregularly sampled data in a KF/FIS combination algorithm to estimate the state of dynamic systems on the irregular time-base.

2.2 Method

2.2.1 State Space Equations for irregularly sampled series

A simple stochastic dynamic Generalised Random Walk (GRW) model (Ng and Young, 1990) will be used to demonstrate the proposed technique. GRW models are commonly used in stochastic State Space modelling (Young, 2000, Harvey, 1989 and Jonsen et al., 2005) where smoothing or forecasting is required, and also to estimate derivatives effectively. Assuming that the state variable x_t has a $(q + 1)^{th}$ derivative at point t , then by Taylor's expansion for x_t in the local neighbourhood of t_k we have:

$$x(t) = x(t_k) + x'(t_k)(x_t - x(t_k)) + \frac{x''(t_k)}{2!}(x_t - x(t_k))^2 + \dots + \frac{x^{(q)}(t_k)}{q!}(x_t - x(t_k))^q \quad (2.1).$$

Smoothing splines (Yandell, 1989) use approximation knots at every point in a time-series and so if the value of state variable x and its derivatives are known at the k^{th} knot as $x_k = [x(t_k)x'(t_k)x''(t_k) \dots x^q(t_k)]^T$ and the highest derivative of x with respect to t with $x^{(q+1)}(t_k) = \eta_k$, where $\eta_k \sim N(0, \sigma_\eta^2)$ and t_k is the approximation knot at sample k . This leads to the GRW model with the system equations (2.4, (Sadeghi, 2006)):

$$\mathbf{x}_k = \mathbf{F}_k \mathbf{x}_{k-1} + \mathbf{G}_k \eta_{k-1} \quad (2.2),$$

where, \mathbf{x}_k is an n dimensional stochastic state vector and $\eta_{(k-1)}$ is (in general) a n dimensional vector of system disturbance.

$$\Delta_k = t_k - t_{k-1} \quad (2.3),$$

$$\mathbf{F}_k = \begin{bmatrix} 1 & \Delta_k & \frac{\Delta_k^2}{2!} & \dots & \frac{\Delta_k^q}{q!} \\ 0 & 1 & \Delta_k & \dots & \frac{\Delta_k^{q-1}}{(q-1)!} \\ 0 & 0 & 1 & \dots & \frac{\Delta_k^{q-2}}{(q-2)!} \\ & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}, \quad \mathbf{G}_k = \left[\frac{\Delta_k^{q+1}}{(q+1)!} \quad \frac{\Delta_k^q}{q!} \quad \frac{\Delta_k^{q-1}}{(q-1)!} \dots \Delta_k \right]^T \quad (2.4),$$

where, $q = 0,1,2,3$, referring to respectively Random Walk, Integrated Random Walk (IRW), Double IRW, Triple IRW etc. These equations then give us the observation equation,

$$y_k = \mathbf{H}_k^T \mathbf{x}_k + e_k \quad (2.5),$$

where, y_k is a scalar stochastic observed variable, e_k is the observation disturbance, \mathbf{H}_k^T is a $1 \times n$ observation matrix, with the following assumptions about the dynamics of the system described by equation 2.2 and 2.4:

- η_k and e_k are serially independent random variables with distributions $N(0, Q)$ and $N(0, \sigma^2)$ respectively and the Noise Variance Ratio (Q_r) is defined as $Q_r = Q/\sigma^2$ for the scalar

observation case we consider here. The Normality assumption is relaxed to symmetry of the distribution in the original Kalman's formulation (Kalman, 1960).

- The initial stochastic state x_0 is independent of η_{k-1} and e_k for each k .
- The system matrix H_k^T is defined by the model structure whereas the system noise covariance (Q) is usually estimated.
- The initial conditions for the states and their covariance matrix (x_0 and P_0 , respectively) are assumed to be zero mean and high covariance (diffused priors approach).
- Variance, σ^2 forms the assumed or estimated Q_r .

2.2.2 Arbitrary Sampling

The irregularly sampled time series cannot be analysed without the accompanying time scale, so naturally for each sample x_k the value of t_k the time when the sample was taken is required. Vector Δ of the same size as that of the observation vector y contains the times ($t_k - t_{k-1}, k = 1, N$) between each sample in the data series (for an example see Figure 2.1).

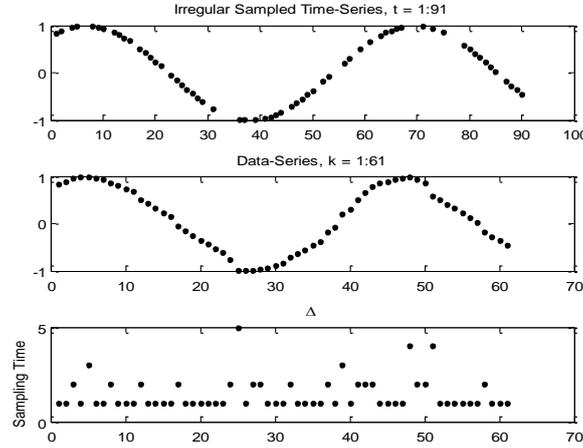


Figure 2.1. Top: Irregular sampled time-series. Middle: Data-series. Bottom: Δ vector of sample time increment.

C. Estimation Algorithm for a first order IRW/SRW model

1. Prepare the data and create the Δ vector.
2. Forward pass through KF, for $1:N$
 - i. Initial conditions: $\hat{P}_0 = \begin{bmatrix} 100 & 0 \\ 0 & 100 \end{bmatrix}$, $\hat{x}_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$
 - ii. Update F_k and G_k matrices (Eq. (4-5))
 - iii. Prediction:

$$\hat{x}_{k|k-1} = F \hat{x}_{k-1}$$

$$\hat{P}_{k|k-1} = F \hat{P}_{k-1} F^T + G Q_r G^T$$

- iv. Correction:

$$\hat{x}_k = \hat{x}_{k|k-1} + \hat{P}_{k|k-1} H_k^T [1 + H_k \hat{P}_{k|k-1} H_k^T]^{-1} \{y_k - H_k \hat{x}_{k|k-1}\}$$

$$\hat{P}_k = \hat{P}_{k|k-1} - \hat{P}_{k|k-1} H_k^T [1 + H_k \hat{P}_{k|k-1} H_k^T]^{-1} H_k \hat{P}_{k|k-1}$$

3. Backward pass through FIS, for N-1:1

- i. Update F_k matrix
- ii. $A_k = \hat{P}_{k|k} F_k^T [\hat{P}_{k+1|k}]^{-1}$
 $x_{k|N} = \hat{x}_{k|k} + A_k [x_{k+1|N} - \hat{x}_{k+1|k}]$
 $P_{k|N} = \hat{P}_{k|k} + A_k [\hat{P}_{k+1|N} - \hat{P}_{k+1|k}] A_k^T$

2.3 Example

In order to compare the proposed techniques performance with standard KF/FIS algorithm handling missing samples with just the prediction step we used fixed sampling rate and removed a number of samples from the data. A simple diminishing amplitude sine wave with periodicity of 2115 samples and with 1% noise variance was simulated with 25% of the signal being randomly removed to represent an irregular sampling case (Figure 2.2). The above algorithm was used to estimate the underlying sine wave from the noisy, irregularly sampled data with a Q of 10^{-8} . The algorithm took 0.8s to process the 7500 long data-series and estimated the sine wave with the goodness of fit R^2 of 0.99 (Figure 2.3).

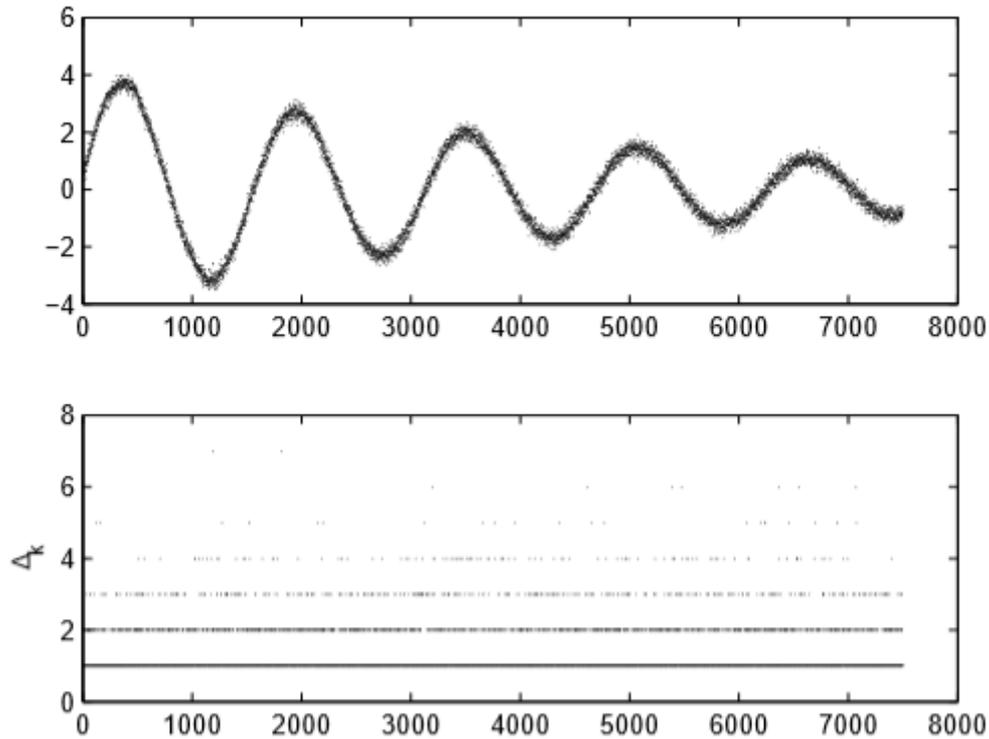


Figure 2.2. Top: Noisy Data-Series. Bottom: Δ - sample time increments.

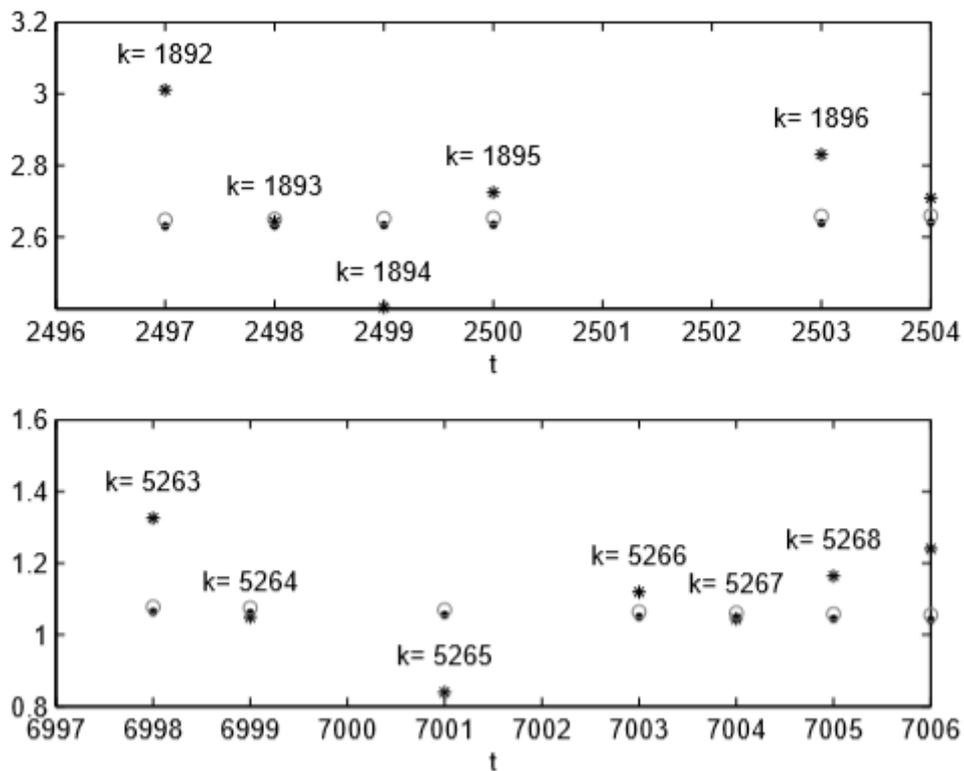


Figure 2.3. Sections of the irregularly sampled time-series with noisy simulated data (*), simulated underlying data (circles) and estimated points (dots).

2.4 Key Aspects of Method

2.4.1. Effects of Noise

The effect of the noise on the ability of the algorithm to estimate the underlying signal is minimal, this was tested using Monte Carlo (MC) simulation by varying the noise variance between 0% and 100% for the same example (Figure 2.4). Difference between the median residual errors of the 100% variance case and the 0% variance case was 0.01 and the R^2 never dropped below 0.99.

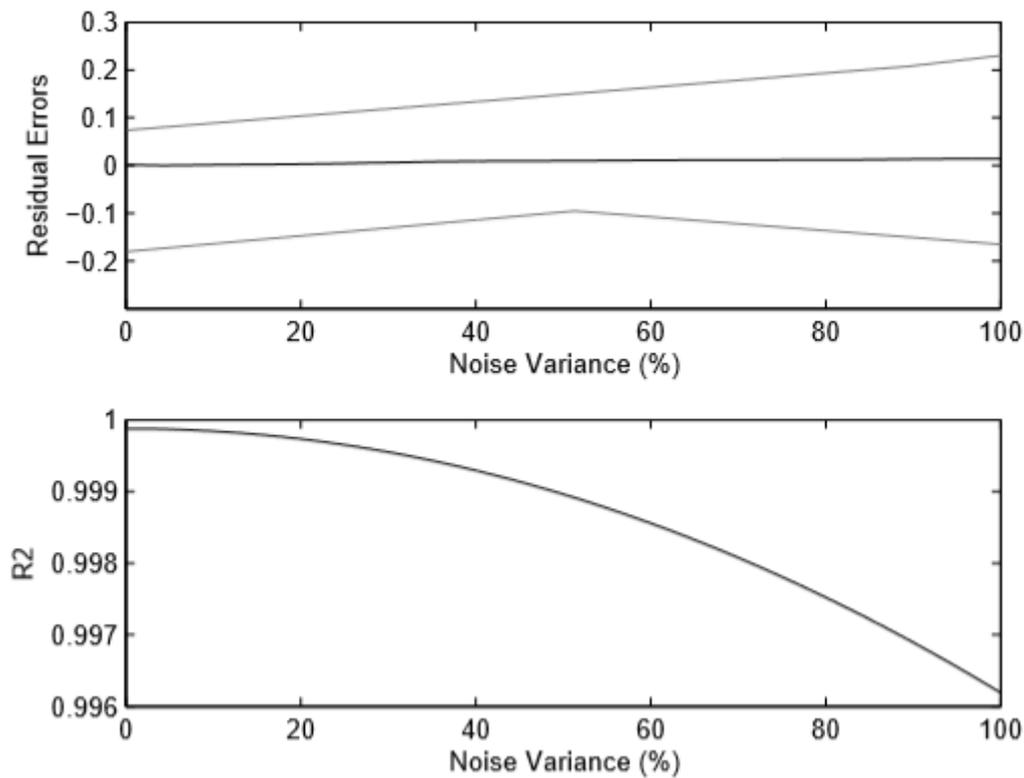


Figure 2.4. Monte Carlo Simulation. Top, residual errors versus observation noise variance, maximum residual error (upper grey), median residual error (black) and minimum residual error (lower grey). Bottom, R^2 of each MC simulation.

2.4.2. Sparse and Irregularly Sampled Time-Series

An important aspect of this method is its ability to perform for severely sparse time-series. This is demonstrated using MC simulation where the percentage of data removed in this numerical experiment was varied from 1% to 99% (100 missing samples and 9900 missing samples respectively) for the same example (10000 long time-series) (Figure 2.5).

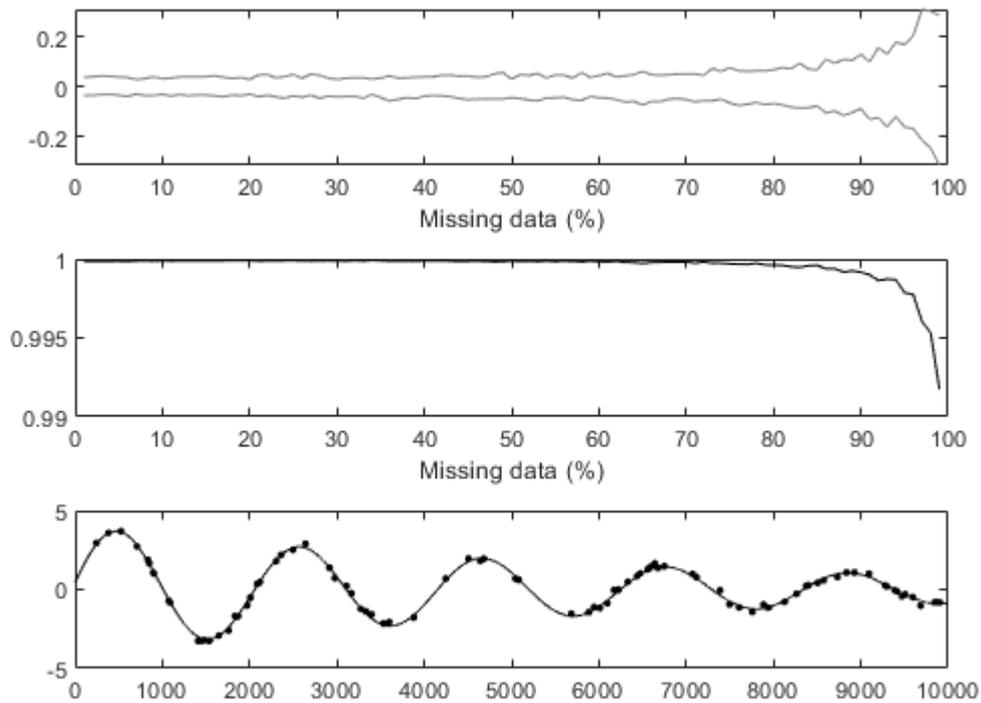


Figure 2.5. Monte Carlo Simulation of under-sampling. Top, percentiles of residual errors; 99% (upper dashes), 1% (lower dashes). Middle, R^2 for each percentage of missing data. Bottom, 99% missing data case, simulated time-series (solid line), estimated values of available data (dots).

2.4.3. Size and Position of Missing Data in a Single Gap

A third important aspect of this method is that the size of the gap, in terms of number of sequential missing samples, is not limiting. The method will perform for any size of missing data gap and is demonstrated using MC simulation where the size of a missing data gap is increased from 100 to 9500 samples, with the starting position fixed (Figure 2.6).

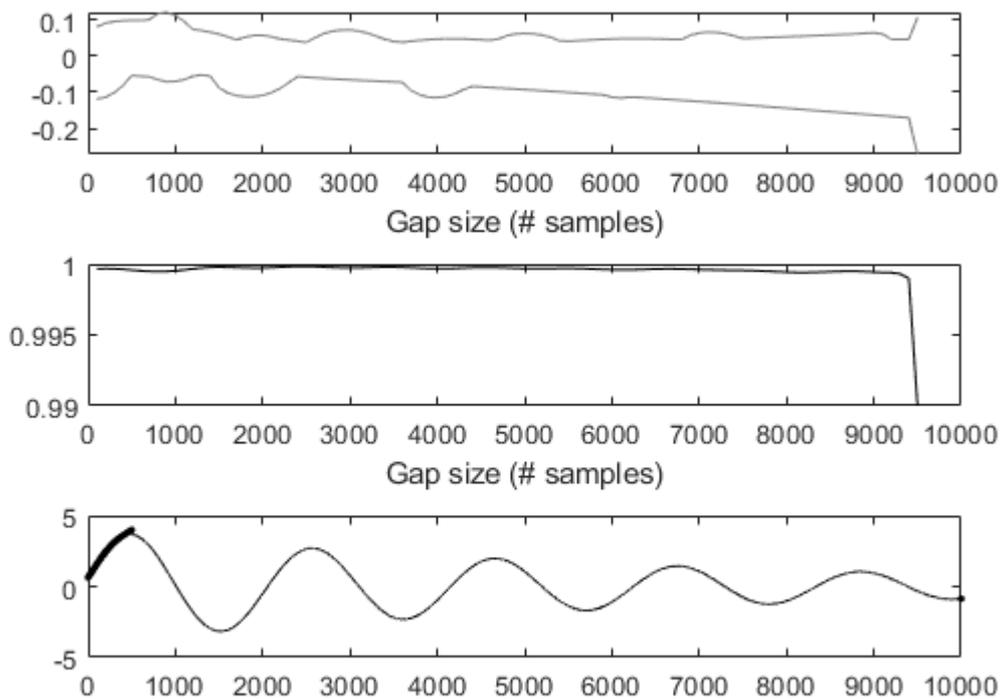


Figure 2.6. Monte Carlo Simulation of missing data gap size. Top, percentiles of residual errors; 99% (upper dashes), 1% (lower dashes). Middle, R^2 for each size of gap. Bottom, case of gap size 9500 samples, simulated time-series (solid line), estimated values of available data (dots).

The points prior to and after a sizeable gap can be poorly estimated, depending where those start, and end points lay within the signal. If there is a significant change in trajectory during the missing data gap, then the estimation either side of the gap will be poor (Figure 2.8), but estimation of the rest of the underlying signal will be unaffected (2.7).

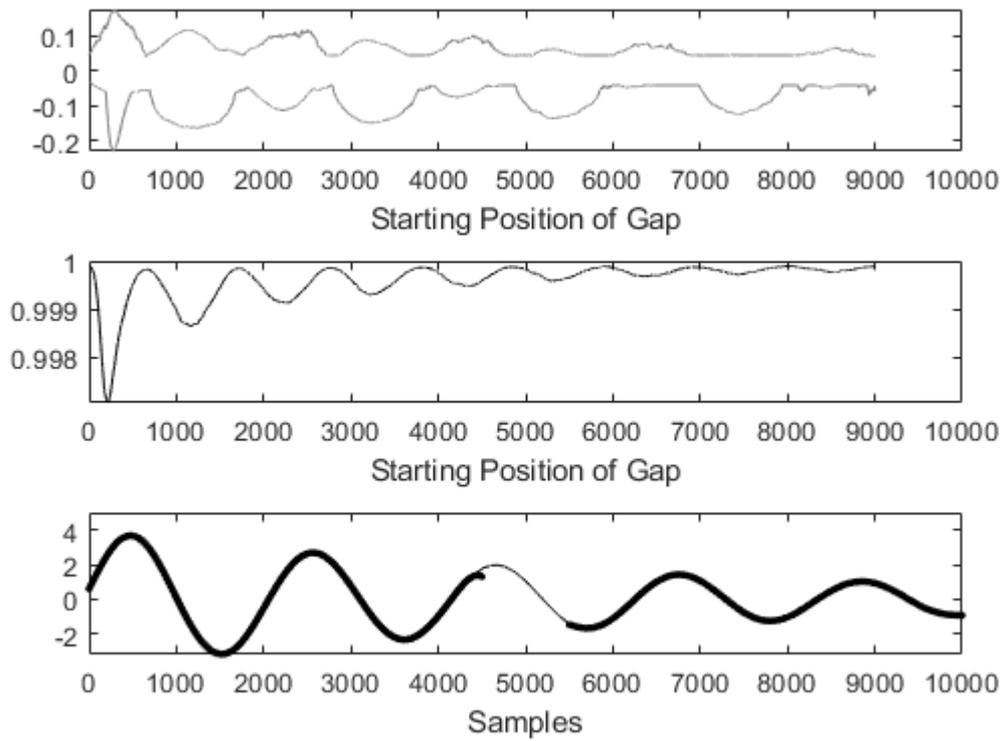


Figure 2.7. Monte Carlo Simulation of position of gap (1000 samples). Top, percentiles of residual errors; 99% (upper dashes), 1% (lower dashes). Middle, R^2 for each position. Bottom, case where gap begins middle of the signal, simulated time-series (solid line), estimated values of available data (dots).

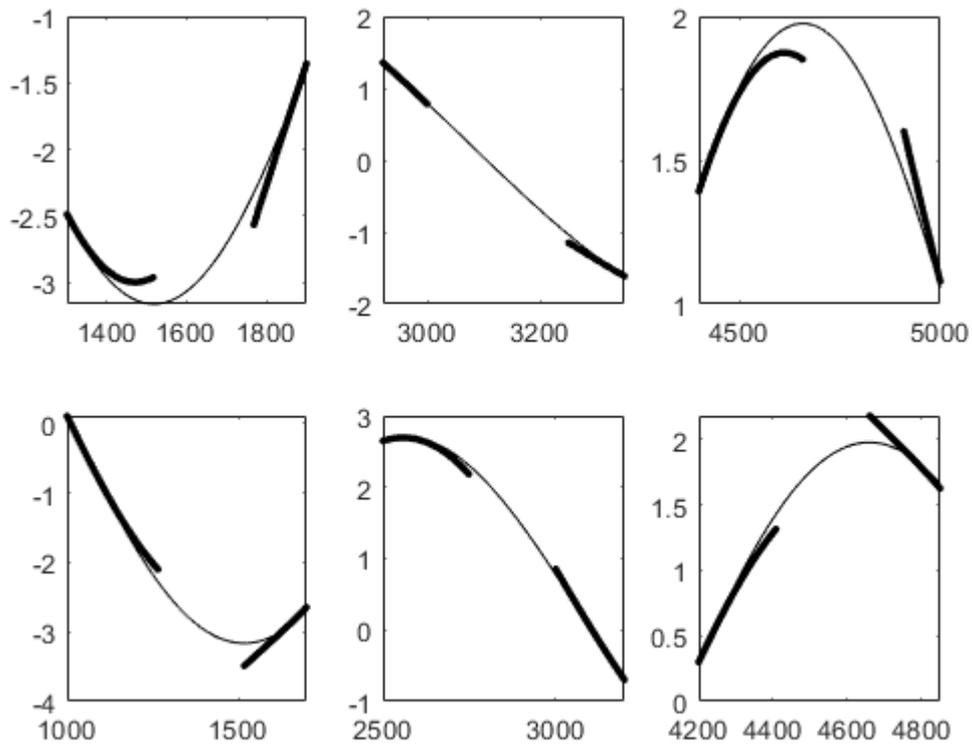


Figure 2.8. Examples of poor estimation either side of a missing data gap (250 samples), simulated time-series (solid line), estimated values of available data (dots).

2.5. Example – Adaptive Sampling with Equipment Failure

Another way to demonstrate the robustness of the algorithm is to simulate a more complicated example (Figure 2.9). Still using the same underlying signal as above, an adaptive sampling with equipment failure situation is simulated. The slopes are sampled every 5th sample, while the peaks and troughs are sampled every sample, and the equipment failure occurs about a third way through. A noise variance of 5% was added to the underlying signal, to further simulate a real case.

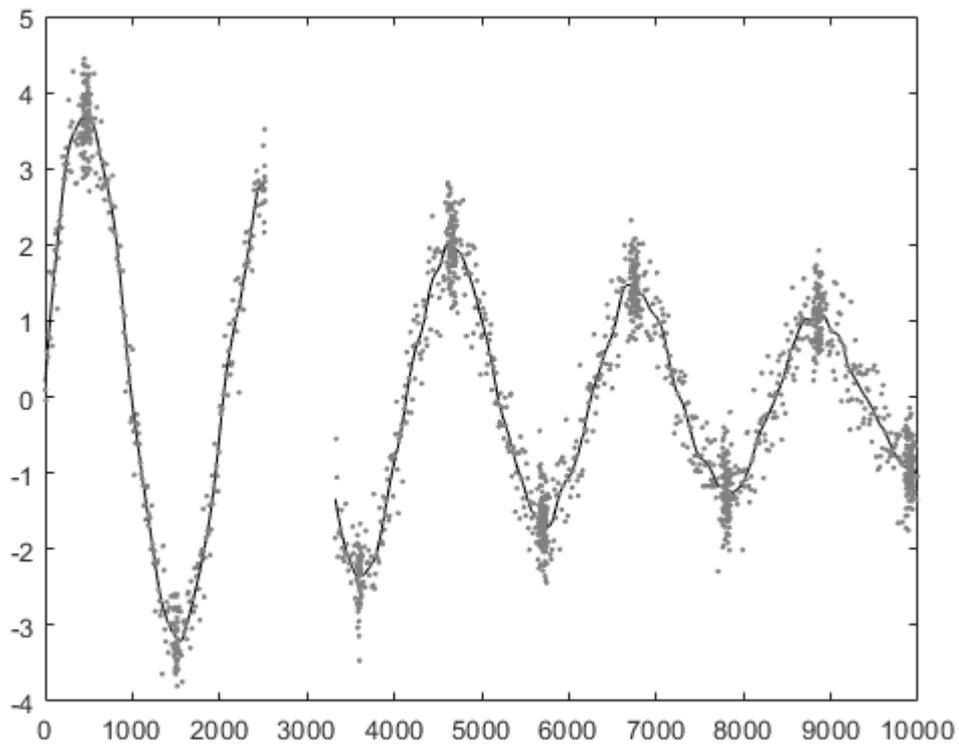


Figure 2.9. Estimation of the underlying signal for a simulated adaptive sampling with equipment failure situation, simulated noisy signal (grey dots), estimated underlying signal (black line).

2.6.0 Conclusion

A generalised recursive state-space method for using irregularly sampled data was presented and its effectiveness demonstrated. The method is robust to noise and data sparsity, and while estimation either side of a large missing data gap can be poor, this can be avoided if the data is broken into two separate signals, which is typically how the data would be handled without the presented method. The method can be incorporated into any KF, FIS or combined data analysis methods and while the smoother used here was the Rauch-Tung-Striebel smoother, other smoothers can be used as the method modifies the Random Walk matrices all smoothers utilise.

Chapter 3: Unification and DBM Alignment of (Multiple) State Dependent Parameter Methodologies – OBJECTIVES 2 & 3

The arbitrary sampling technique developed has a range of applications, one of which is to solve the problem of shifting from the time-domain to the time and state-space domain that the (M)SDP methodologies require for generalisation. This technique was applied to the (M)SDP methodologies, resulting in better estimation and the subsequent work lead to a single unifying methodology including model structure identification and parameter estimation known as ESDP. During the development it became clear an objective identification step was required and this led to the development of a statistical identification method for model structure, including which states the parameter are most likely to depend upon. The acceptance of **paper 2** by the Journal of Environmental Modelling and Software (**Mindham et al., 2018**) demonstrates the importance of the work.

3.1 Introduction

The Data-Based Mechanistic methodology (DBM, Young, 1999a) is built on the premise that the model structure and parameters are to be determined through statistical analysis of observed data ('data-based') which then, along with model metrics leads to a physical interpretation of the model ('mechanistic').

The presented approach completes the nonlinear DBM modelling process by adding an objective identification stage to the nonlinear model selection. Multiple and single State Dependent Parameter (MSDP and SDP) modelling follows the DBM methodology by not parameterising the individual nonlinearities, however the selection of the model structure, including that of the nonlinear drivers, is the key element missing from the current method.

While MSDP employs a very different numerical engine to SDP, conceptually and in terms of the outcome, it is a multi-variable extension of the original SDP and thus, a generalisation of SDP that is not confined to one state dependency. However, both in SDP and then MSDP, the states' values were assumed equidistant, having the same distance in the state-space between each sample, which is a simplifying assumption. The solution introduced in this paper removes this assumption and makes the method fully general.

The new model structure identification procedure allows for the first-time identification of nonlinear structural relationships in an objective manner using a robust and tested model form. This is demonstrated in the paper using high frequency hydrological observations, where the output variable is thought to be affected by more than one nonlinear process.

The terminology, explanation and clarification for the above are laid out below in a logical and methodical manner designed to lead the reader first through existing SDP and MSDP methodologies, then through the process of updating and extending these methodologies into one methodology described in this paper (ESDP – Extended SDP) with useful output tools. Finally, through the process of producing a generalised Model Structure Identification (MSI) procedure to identify the structure from a given data set for the application of ESDP. The MSI procedure is generalised in that it considers - no state dependency for each parameter (linear model) and one or more state dependencies for one or more parameters (nonlinear model).

3.1 Objectives and Structure of this Paper

This paper presents three key updates and additions to the SDP and MSDP methodologies leading to their unification and generalisation (items A,...,C), and one major development (item D) for applying the DBM approach to model structure identification in this new setting:

- A. Transition to a true state-space for parameter estimation by moving from equidistant states to arbitrary-distant states, based on the state values. The terms and context of ‘state-space’ and ‘states’ are clarified and discussed in section 1.4 and onwards. (Section 2.0)
- B. Formation of multivariable parameter maps from M-dimensional state dependent parameters for the purposes of basic model validation and more importantly, for forecasting, scenario investigation and on-line simulation of live events. (Section 2.1)
- C. Use of model validation techniques to not only quantify the ability of the presented algorithms in parameter estimation, but also to validate any models identified from the model structure identification development step below. (Section 2.2)
- D. Development of a DBM approach for model structure identification (MSI) from a group of data sets for a given model type so that the data informs us which measured variables are more influential to the observed model output – importantly this methodology also considers a linear model, allowing for a ‘pure’ DBM approach. (Section 3.0)

The whole approach is then applied to a hydrological example using a dynamic model of streamflow generation, thus forming objective E. (Section 4.0)

3.2 Applications

The methodologies presented are general and can be applied to any system as long as time-series data for all the required variables (including inputs, outputs and additional states) are available. In terms of specific environmental applications, we have evaluated the approach for data in the two applications below, and present the former in this paper:

- Flood scenarios – how the flood response of a stream may be strongly affected by more than multiple nonlinear process, not solely the nonlinear effects of varying catchment wetness (Chappell et al., 2017).
- Water quality dynamics – how the dynamics of one output water quality variable (e.g., Dissolved Organic Carbon concentration) may be affected by more than one nonlinear process, related to separate effects of e.g., rainfall, soil temperature and solar radiation (Jones et al., 2014).

For clarity and in order to introduce the notation, this paper also briefly covers the progression from Transfer Function (TF) to SDP TF (for a more detailed account see Young, 2000) and MSDP TF, with the novel generalisation elements introduced. Significantly, the development of the structural identification methodology for this wide class of nonlinear models is then covered.

3.3 Background to SDP

There is extensive work on modelling input-output dynamic time-series data using Transfer Functions (TF or equivalent Auto-Regressive with eXogenous inputs, or ARX models) where linear or approximated linear relationships are used (Ockenden and Chappell, 2011; Tych et al, 2014; Ampadu et al, 2015; Chappell et al., 2017) as well as extensions into Time Varying Parameter (TVP) TF (Gou, 1990) and further extensions into State Dependent Parameter (SDP) TF (Young et al, 2001) with

latter approaches using nonlinear functional relationships between states of the system and the ARX or TF parameters.

SDP modelling assumes that the system is truly nonlinear in that the TF parameters are time varying; importantly, the rate of change of the parameters is at a rate related to the rate of change within the state variables. This is unlike the more commonly seen time varying parameter TF models, where the parameters change smoothly. Here, the parameters are functions of the input or other states of the system under study. SDP, as originally published by Young (2000) bears the assumption that each parameter is a function of one variable only, and has been successfully applied to many nonlinear systems (e.g. Young et al, 2007a; Taylor et al., 2009; McIntyre et al., 2011). However, many systems, particularly in the natural environment, are complex dynamic systems with many variables that have interlinking relationships, e.g. water quality (Jones et al., 2014), atmospheric chemistry (Seinfeld and Pandis, 2016), and climate change (Ashkenazy et al., 2003; Young and Garnier, 2006). The parameters of models describing these environmental systems, or even just their specific aspects, could be functions of more than one variable and hence the need to generalise SDP modelling into the of Multiple State Dependent Parameter (MSDP) modelling.

3.4 ARX - Transfer Function (TF) – Time Invariant and Time Varying Parameter (TVP) Issues

We begin with a simple linear discrete time dynamic model with time varying parameters (ARX or equivalent TF structure (Young, 1999b)) that relates a single input variable (u_t) to an output variable (y_t) and can be written as a difference equation (3.1). Due to the time-varying character of parameters the standard backward shift TF models are not applicable.

$$y_t = -a_{1,t}y_{t-1} - \dots - a_{n,t}y_{t-n} + b_{0,t}u_{t-\delta} + b_{1,t}u_{t-\delta-1} + \dots + b_{m,t}u_{t-\delta-m} + e_t \quad (3.1),$$

where δ is a pure time delay (measured at this stage in sampling intervals), e_t is a zero mean, serially uncorrelated input with variance σ^2 and Gaussian amplitude distribution. The latter (Normality) assumption is usual, but not required for the Kalman Filter to function (Kalman, 1960 – the distribution needs to be finite and symmetric).

Expressing (3.1) as a vector equation we obtain the TVP observation equation:

$$y_t = \mathbf{z}_t^T \mathbf{p}_t + e_t \quad (3.2),$$

where,

$$\mathbf{z}_t^T = [-y_{t-1} \quad -y_{t-2} \quad \dots \quad -y_{t-n} \quad u_{t-\delta} \quad \dots \quad u_{t-\delta-m}]$$

$$\mathbf{p}_t = [a_{1,t} \quad a_{2,t} \quad \dots \quad a_{n,t} \quad b_{0,t} \quad \dots \quad b_{m,t}]^T$$

When \mathbf{p}_t changes slowly/smoothly there are methods to estimate these changes taking advantage of the smoothness assumption (see for example Dynamic Transfer Function or DARX models – Young, 2011).

However, many environmental systems can be described as complex and nonlinear, where the rates of change of the parameters vary at a rate commensurate to that of other, exogenous variables. This means the changes in \mathbf{p}_t are too rapid to apply the smoothness assumption (slow parametric changes) and so other estimation methods are required. If the parameters are varying at a rate similar to that of the rate of another system variable, then that system variable must be incorporated into the model in some form. This leads to SDP modelling where it is assumed that each parameter is dependent on a system variable. For complex systems the number of variables could be large, for example in a streamflow generation (rainfall-runoff) system there is more to

consider than just rainfall and streamflow - solar radiation (controlling evaporation), soil saturation or water table depth within a deep underlying rock aquifer, may also affect the processes in nonlinear ways (e.g., Ockenden and Chappell, 2011; Deutscher et al., 2016).

3.5 State Dependent Parameter (SDP) Input-Output models

In (2) each of the parameters \mathbf{p}_t can be further defined as a function of identified, but in general arbitrarily chosen, independent variables treated as states \mathbf{s}_t of the same system, hence the name State Dependent Parameters. This is done in the SDP approach (Young, 2000) for a single independent state variable.

Following Young (2000) and then Tych et al (2012), we describe the time varying model parameters (\mathbf{p}_i) using stochastic dynamic system definition, where the i th parameter ($-a_i$ for $i \leq n$, $b_{i-(n+1)}$ if $i > n$) can be defined as a Generalised Random Walk process. Then, in the specific case of Integrated Random Walk every parameter is defined by a two-dimensional stochastic state vector $\mathbf{x}_{i,t} = [l_{i,t} \ d_{i,t}]^T$, where $l_{i,t}$ and $d_{i,t}$ are, respectively, the changing level and slope of the associated time varying parameter (Young, 2011) model. Note that the state vector \mathbf{x} in this context is used to formulate the estimation procedure and is not the same as the SDP vector \mathbf{s} of states driving the parameters. These are states within different spaces – \mathbf{x} is the state used to model and estimate the parametric change as a stochastic process, and \mathbf{s} is a part of the SDP model.

In the time series model the time-varying parameter estimation is conducted in temporal order ($t=1:N$). With State Dependent Parameters we estimate these parameters within a different space, a state space domain defined by the state variables driving the parameters. There the parameters are estimated in an arbitrary order ($k=1:N$) along a trajectory \mathcal{K} within that state space (not the time sequence t). For a parameter driven by a single independent state variable (as in Young, 2000), typically the order is based upon the ascending order of the independent state.

These parameters can be then estimated recursively using a Kalman Filter (KF, Kalman 1960) and Fixed-Interval Smoothing (FIS) combination along the trajectory \mathcal{K} resulting in the stochastic state space equation for the i th parameter (3), $i = 1, \dots, n+m$.

Variation along the trajectory in the state-space (SS) of each parameter i of the $n+m$ model parameters is modelled as a stochastic state space model:

$$\mathbf{x}_k^i = \mathbf{F}\mathbf{x}_{k-1}^i + \mathbf{G}\eta_{k-1}^i \quad (3.3),$$

with:

$$\mathbf{F} = \begin{bmatrix} 1 & 1 & \frac{1}{2} & \dots & \frac{1}{q_i!} \\ 0 & 1 & 1 & \dots & \frac{1}{(q_i-1)!} \\ 0 & 0 & 1 & \dots & \frac{1}{(q_i-2)!} \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}, \quad \mathbf{G} = \left[\frac{1}{(q_i+1)!} \quad \frac{1}{q_i!} \quad \frac{1}{(q_i-1)!} \dots 1 \right]^T,$$

where, $\mathbf{H}_{i,k}^T$ is a $1 \times n$ observation matrix (input variable corresponding to the i th parameter), η_{k-1} is a (in general) $q+1$ dimensional vector of system disturbance, with $q = 0, 1, 2 \dots$ referring to respectively; random walk, integrated random walk, double integrated random walk, etc.

For the $n+m$ parameter model this becomes a block-diagonal SS model of higher dimension. The observation equation for this SS is Equation 2 with parameters vector \mathbf{p} being a subset of the above state vector \mathbf{x} as in the original SDP approach (see e.g. Young, 2000).

The individual variance terms of η^i form the state covariance matrix \mathbf{Q} (diagonal in this case). With the univariate observation series the variance of the observation disturbance ratio of σ_i^2 is a scalar used to standardise the KF variance parameters (as in Young, 2011). The resulting Noise Variance Ratio (NVR) matrix – the meta-parameter of the filtering and smoothing process becomes:

$$NVR = \frac{Q}{\sigma^2}$$

An example of a simple SDP is illustrated with a Nonlinear ARX (NARX) structure and is a forced logistic growth equation (3.4) similar to that found in Young (2000). In (3.4) and further on, \mathbf{s} is defined as a vector variable that the parameter is dependent on, known as ‘state variable’ or ‘state’ vector.

$$y_t = a(s_t)_t \cdot y_{t-1} + b(1)_t \cdot u_t + e_t \quad (3.4),$$

where, $a(s_t)_t = 2 - 2y_{t-1}$, $B_t = 1$, so that a is a SDP that is a function of y_{t-1} (the value of which is the state s which drives the parameter, in this case a past system output – an available observation) and b is a constant coefficient. In this simulation example, u_t is a sequence of random numbers between -0.2 and 0.2 and e_t is a serially uncorrelated Gaussian observation disturbance with standard deviations of approx. 5% of standard deviation of y_t . The SDP function from the Captain Toolbox (Taylor et al., 2007) was used to estimate the parameters for this model (Figure 3.1).

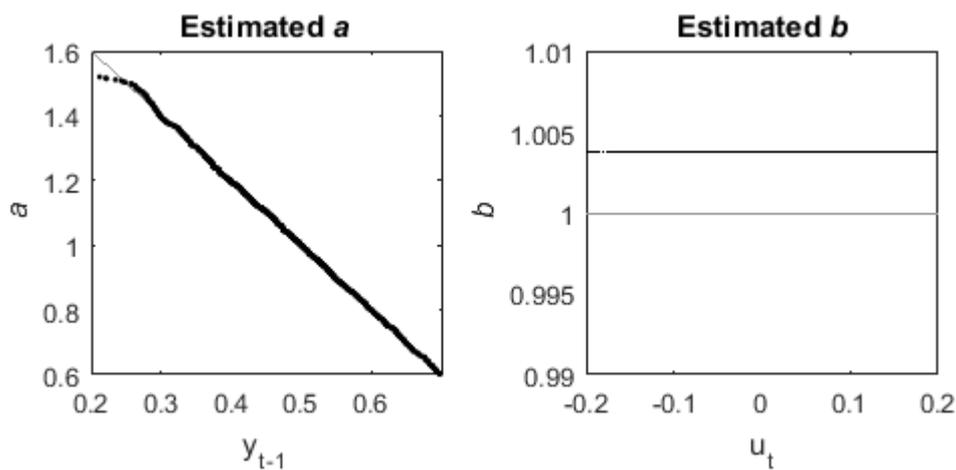


Figure 3.1. Illustration of a simple SDP. Left hand plot, estimation of a , which is a function of y_{t-1} . Right hand plot, estimation of b , which is a constant of 1. Light grey shows the actual values, black - the estimated values.

To avoid any confusion in the sequel, which would be due to the estimation technique using state-space approach of the state-dependent parameters, we use s_t reordered from the temporal sequence into s_k along the trajectory \mathcal{K} to denote the exogenous, independent variables affecting the input-output model parameters, and \mathbf{x}_k - the estimates of parameters sequence \mathbf{p}_k tracked by the variable step KF/FIS algorithm based on (3.3).

As already mentioned the SDP approach only considers one state per parameter and in more complex nonlinear systems it is likely that more than one state will influence each parameter. This leads to the extension into Multiple SDP (MSDP).

3.6 Multiple State Dependent Parameter Models

The approach taken in SDP applies to MSDP with two exceptions; ordering along trajectory \mathcal{K} based on multiple states \mathbf{s}_k , and for robustness, \mathbf{x}_k is averaged over the closest points along its trajectory in the multi-dimensional state space (Tych et al., 2012).

The arbitrary order ($k=1, \dots, N$) of parameter estimation is based in our study on the ordering of multiple states, as an ascending sequence of Euclidean norms of the original n-dimensional states driving \mathbf{p}_i , normalised, so that they lie within an n-dimensional unit hyper-cube, and ordered according to distance from an arbitrarily chosen starting point. In this study the bottom left of the hyper-cube (Figure 3.2) is set as the origin for the sequence.

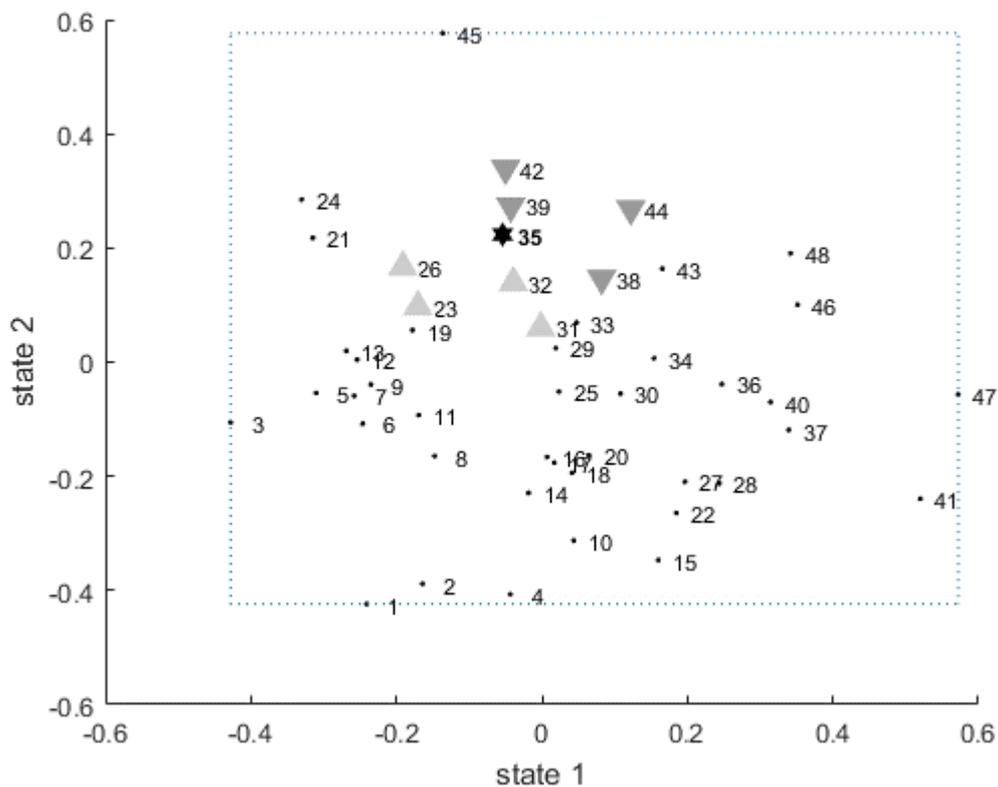


Figure 3.2. Illustration of a simulated sequence within the hyper-cube for 2-dimensions also showing the predecessors and successors sequence. When $k=35$, upwards-pointing triangles are the closest four predecessors (23,26,31,32) and downwards-pointing triangles are the closest four successors (38,39,42,44).

Parameter estimates ($\mathbf{p}_{i,k}$) are no longer based on $\mathbf{x}_{1:(k-1)}$ in filtering or $\mathbf{x}_{1:N-k}$ in smoothing but on the average of estimates from the nearest preceding states or succeeding states within the multi-

dimensional space (Figure 3.2), $(\sum_{1:j}(x_{(1:k-1)_j})) / j$ or $(\sum_{1:j}(x_{(1:N-k)_j})) / j$, where j = number of closest states to x_k .

A simple example, similar to that found in Tych et al., 2010, is illustrated in Figure 3.3 with similar structure and variables to (4) but with the A parameter being a function of two states (3.5)

$$y_t = a(s_t)_t \cdot y_{t-1} + b(1)_t \cdot u_t + e_t \quad (3.5),$$

where, $a(s_t)_t = 0.5 \times \tan^{-1}(y_{t-2} \times u_{t-1})$, with y_{t-2} and u_{t-1} being the states

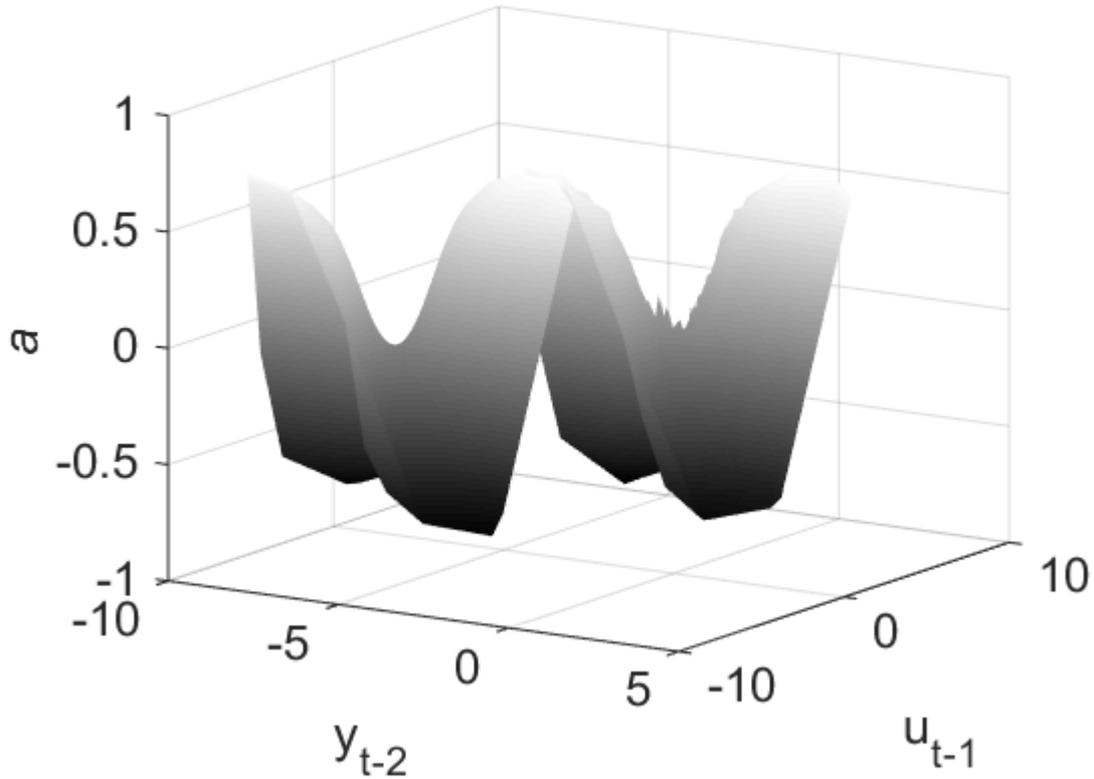


Figure 3. Illustration of a MSDP. Left hand side, simulated a parameter (offset for clarity). Right hand side, estimated a parameter.

3.2 Arbitrary Sampling within the State Space (OBJECTIVE A)

For TVP the distance between each parameter value Δ is based on the temporal distance between each sample and is typically uniform where $\Delta_t = 1$. For both MSDP and SDP, despite the shift from temporal ordering to arbitrary ordering in the state-space, this is still the case and $\mathbf{p}_{i,k}$ is still estimated with $\Delta_k = 1$. It should be apparent that this is incorrect and Δ_k should in fact vary depending on the changing distances between the ordered state values.

This leads to two key changes in the (M)SDP methodology, the first is the calculation of $\Delta_{i,k}$ for $\mathbf{p}_{i,k}$ and the second a modification to the random walk models (3.3) that incorporate a changing Δ .

Going back to the equations in (3.3) focusing on single parameter estimation and defining a as $\mathbf{p}_{1,k}$: if $a^v(s_k)$ is the v^{th} derivative of $a(s_k)$, and the form of the function $a(\cdot)$ is not specified, a data point distant from s_k provides very little information about $a(s_k)$ (Sadeghi, 2006). Using the local

polynomial modelling reasoning (e.g. Fan and Gijbels, 1996) only the local data points in the vicinity of \mathbf{s}_k are used. Assuming $a(\mathbf{s}_k)$ has derivative of order $(q+1)$ at the point \mathbf{s}_k , then following Taylor's expansion for \mathbf{s} in the local neighbourhood of \mathbf{s}_k we have:

$$a(\mathbf{s}) = a(\mathbf{s}_k) + a'(\mathbf{s}_k)(\mathbf{s} - \mathbf{s}_k) + \frac{a''(\mathbf{s}_k)}{2!}(\mathbf{s} - \mathbf{s}_k)^2 + \dots + \frac{a^{(q)}(\mathbf{s}_k)}{q!}(\mathbf{s} - \mathbf{s}_k)^q \quad (3.6).$$

If the value of parameter a and its derivatives with respect to its driving state \mathbf{s} are known at the k^{th} point as $\mathbf{x}_k = [a(\zeta_k) \ a'(\zeta_k) \ a''(\zeta_k) \ \dots \ a^{(q)}(\zeta_k)]^T$ and the highest derivative of $a(s)$ with respect to s with $s_k = \zeta_k : a^{(q+1)}(\zeta_k) = \eta_k$, where $\eta_k \sim \mathcal{N}(0, \sigma_\eta^2)$ and ζ_k is the approximation point (knot) at sample k , then Taylor's expansion (6) can be applied in the local neighbourhood of ζ_k for all derivatives of a resulting in the GRW model with equations:

$$\mathbf{x}_k^i = F_k^i \mathbf{x}_{k-1}^i + G_k^i \eta_{k-1}^i \quad (3.7)$$

$$F_k = \begin{bmatrix} 1 & \Delta_k & \frac{\Delta_k^2}{2!} & \dots & \frac{\Delta_k^q}{q!} \\ 0 & 1 & \Delta_k & \dots & \frac{\Delta_k^{q-1}}{(q-1)!} \\ 0 & 0 & 1 & \dots & \frac{\Delta_k^{q-2}}{(q-2)!} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}, \quad G_k = \begin{bmatrix} \frac{\Delta_k^{q+1}}{(q+1)!} & \frac{\Delta_k^q}{q!} & \frac{\Delta_k^{q-1}}{(q-1)!} & \dots & \Delta_k \end{bmatrix}^T,$$

where, $\Delta_k = \|\zeta_k - \zeta_{k-1}\|$ is the Euclidean distance between the points ζ_k and ζ_{k-1} .

With this generalisation, we now have extended the State Dependent Parameter methodology and to demonstrate the improvement of the addition of arbitrary distance RW algorithm the previous two examples (Figure 3.1 and 3.3) are used as comparisons (Figure 3.4 and 3.5). We refer to, R^2 as the coefficient of determination for a linear regression, or proportion of variance explained by one-step ahead prediction.

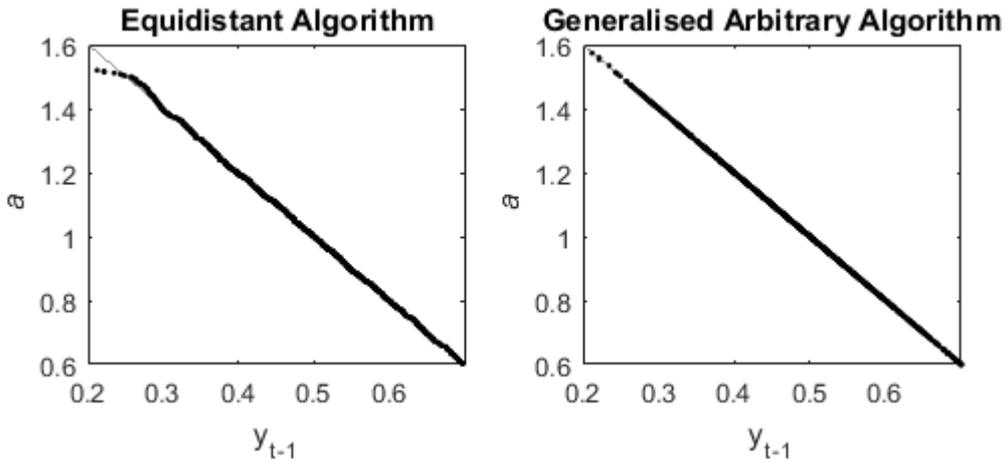


Figure 3.4. Comparing equidistant SDP with the arbitrary distance ESDP. Left hand side, equidistant SDP (R^2 of 0.9985) from Figure 3.1. Right hand side, arbitrary distance ESDP (R^2 of 1.000). Light grey shows the actual values, black - the estimated ones.

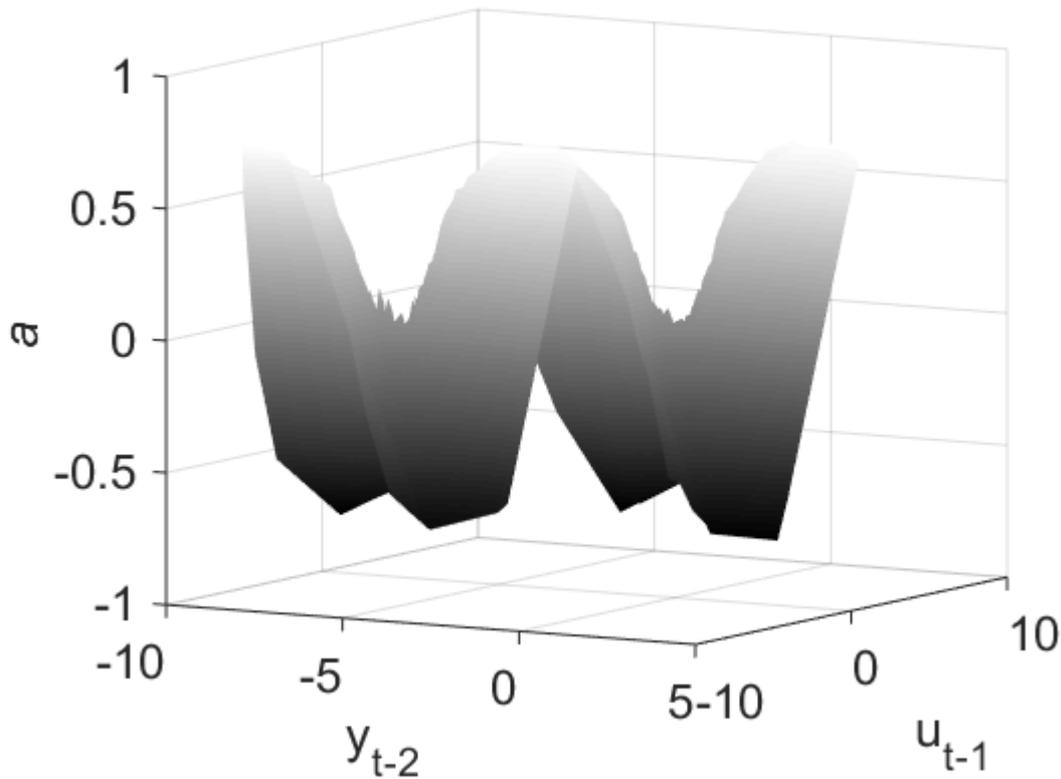


Figure 3.5. Comparing A parameter estimates of equidistant MSDP with arbitrary distance ESDP. Left hand side, equidistant MSDP (R^2 of 0.9775) from Figure 3.3 (offset for clarity). Right hand side, arbitrary distance ESDP (R^2 of 0.9810).

Generally, the introduction of arbitrary distance algorithm improves the parameter estimates but with a sizable improvement in the periphery estimates from single states (noticeable from Figure 3.4) and a reasonable improvement in densely populated estimates from multiple states (reduction of poor estimation in the middle of the saddle shapes from Figure 3.5).

3.2.1 Parameter Mapping (OBJECTIVE B)

Once an ESDP model has been estimated, the surface formed with its driving states can be interpolated to form a parameter map where using a different data-set of the driving states as the coordinates of the parameters on the map can be used to find a suitably interpolated value of the state-dependent parameter. For SDP this had been done before (Ratto et al., 2007), however it had not been done for MSDP. For example, one of our validation procedures consisted of estimating the MSDP $a_e(\mathbf{s}_1, \mathbf{s}_2)$ for the training (estimation) subset of temporal samples $t_e=1, \dots, 500$ and finding the interpolant to form the parameter map $\tilde{a}(\mathbf{s}_1, \mathbf{s}_2)$ followed by using the validation subset samples $t_v=501, \dots, 1000$ to find the interpolated values of $a_v(\mathbf{s}_{1,t_v}, \mathbf{s}_{2,t_v})$ from coordinates $\mathbf{s}_{1,t_v}, \mathbf{s}_{2,t_v}$.

This not only provides a method of validation; in this case answering the question whether the estimated model parameter maps work for different data sets. These maps also provide a tool for studying different scenarios as well as on-line simulation of live events. They place this extension of the MSDP technique within the paradigm of the more general DBM methodology.

To demonstrate this model validation and data visualisation tool, the model from (3.5) was used to simulate data for $n=2000$ where the first half ($t_e=1, \dots, 1000$) was used to estimate $a_e(\mathbf{s}_{1,t_e}, \mathbf{s}_{2,t_e})$ and

then interpolated to find map $\tilde{A}(s_1, s_2)$. The second half of the data ($t_v=1001, \dots, 2000$) was then used to interpolate $a_v(s_{1,t_v}, s_{2,t_v})$ (Figure 3.6) and combined with the estimated B parameter from the first half of the data; these two methods of validation can be used to fully validate the ESDP procedure.

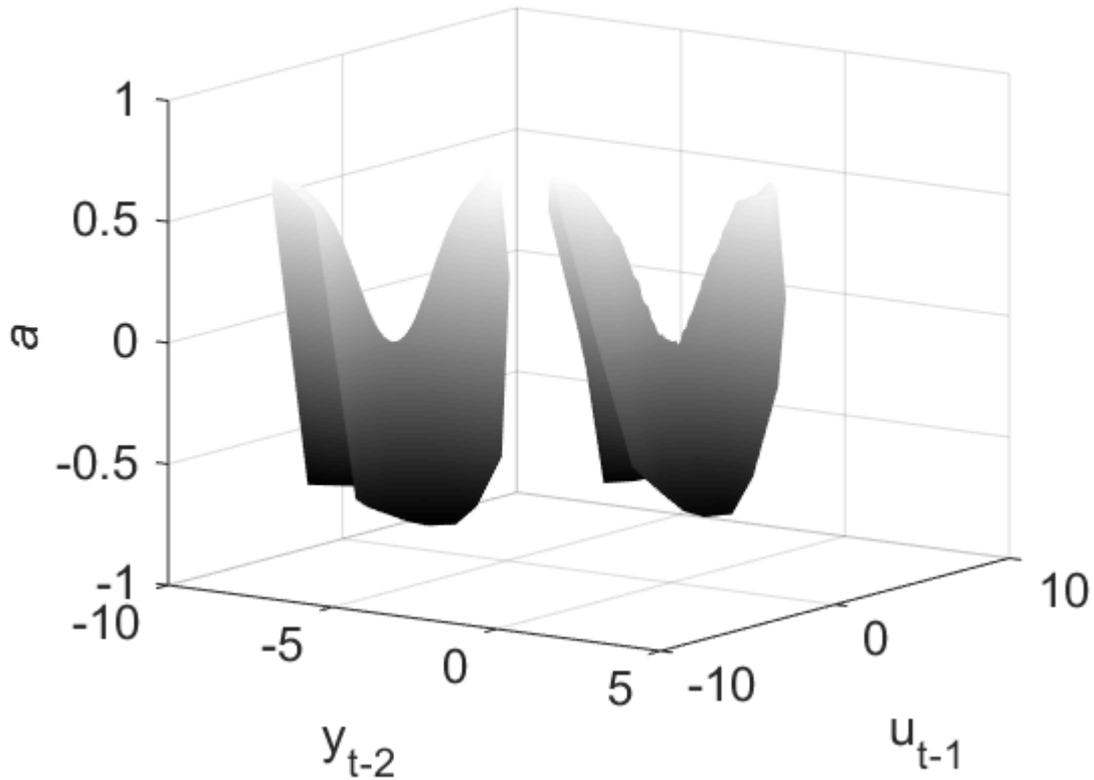


Figure 3.6. Interpolated a parameter for the second half of the data. Left hand side, simulated A (offset for clarity). Right hand side, interpolated a (R^2 of 0.9787).

3.2.2 Model Validation (OBJECTIVE C)

There are four levels of testing that can be applied to validate the performance of these types of models at the stage of their development. The first two support the DBM model development phase, where the suitability of the model structure becomes apparent, while the third and fourth levels constitutes the final validation steps.

1. Within-sample model fit – R^2 of: model output, estimated A parameter and estimated B parameter (the latter two – for simulated models).
2. One-step ahead prediction model fit R^2 – where the simulation output (y_t) is recalculated using the estimated parameters instead of the function of states to form the estimated output (\hat{y}_t) with the regressor values (y_{t-1}, u_{t-1}) still used as above.
3. Full input-output model simulation fit (R_t^2) – again where (y_t) is calculated but not only using the estimated parameters but importantly, the estimated past output \hat{y}_{t-1} such that

$$\hat{y}_t = a_t \cdot \hat{y}_{t-1} + b_t \cdot u_t$$

4. Monte Carlo simulations of the full-simulated model using the estimated parameter standard errors with normally distributed parametric variation – generating error bands for the full input-output model simulation to check model sensitivity and uncertainty.

The first of these tests has been done with the R^2 of the estimated parameter for each model supplied within the relevant figure caption and the overall model fit R^2 was 0.9943 for (Figure 4) and 0.9959 for (Figure 5). The middle two tests were implemented first for the first half of the data and then secondly for the second half of the data from the previous section; first half, one-step ahead and full model simulation produced R^2 of 0.9973 and R_t^2 of 0.9950 respectively. Second half, one-step ahead and the full model simulation produced R^2 of 0.9908 and R_t^2 of 0.9869 respectively.

The MC simulations were also implemented on both halves of the data and shows the model stability (Figure 3.7).

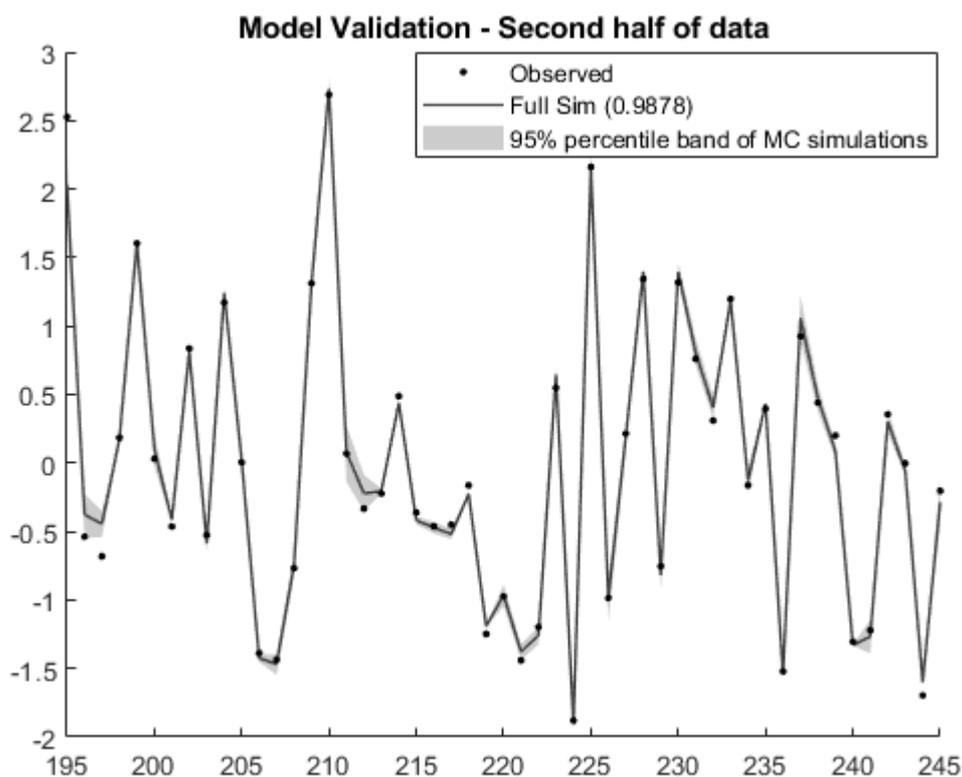


Figure 3.7. Uncertainty of the full model simulation, validation on the second half of data: 95% percentile band of MC simulations using the estimated parameter standard error with normally distributed parametric variation (zoomed-in for visibility). R_t^2 for the first half of the data (training set) was 0.995.

It should be noted that, in the realm of true nonlinear and non-stationary systems, such as those found in the natural environment, that performing full model simulation on different time periods (estimation and validation) may not be practical. The inherent nonlinearity can result in two time periods never behaving similar enough for this type of validation across different time periods to be a test of model validity.

The identification process for estimation period takes into account the prevailing conditions (such as soil saturation), and if they do not change, they are made constant (not an SDP) in the model. These

same conditions may be different during the validation period. Enlarging the model would be needed for covering both periods, but this is prohibited by the efficient identification process selecting the parsimonious model for the estimation period.

3.2.3 Extended State Dependent Parameter (ESDP) Methodology

With updating the variable state-space distancing, addition of multiple dimensional parameter maps and introduction of the latter 2 steps of model validation, the MSDP and SDP methodologies have been updated to perform better (in terms of parameter estimation and model fit) and extended to include more tools for system analysis. This, with the observation that MSDP is a generalisation of SDP to include more than one state, leads to term what is presented here as Extended State Dependent Parameters (ESDP) methodology.

With the inclusion of the following model structure identification procedure, ESDP will have all the tools necessary for analysis of nonlinear dynamic systems in a complete DBM setting.

3.3 Model Structure Identification (MSI) for ESDP (OBJECTIVE D)

It has to be stated at this point, that the work presented in this section is, by its very nature, a step towards the general structure identification for a very general class of models. Therefore, it should be treated as a pragmatic study, providing constructive directions and effective solutions for a specific class of models, but not the 'final' and general answer to the question.

Since the general MSI algorithms for linear TF models are such a useful tool in investigating system dynamics, it is logical to assume MSI for nonlinear SDP models would be just as useful. Historically, SDP has been employed to study specific interactions where the main model structure is pre-selected by the researcher, and the nonlinearity was the main subject of identification.

However, within the realm of nonlinear systems, identification of model structure is not as clearly defined as within the class of LTI (Linear Time-Invariant) models and no well-defined general approach has been developed so far. While model selection within the Nonlinear ARX class (1) is slightly better defined due to the constraints on the TF models, in general model structure identification largely remains an open topic for nonlinear systems.

Most existing approaches to tackling nonlinear structure identification involve assuming some aspect of linearity (Haber and Unbehauen, 1990):

- Breaking a nonlinear system into parallel subsystems with different degrees of nonlinearity (usually static) with at least one dynamic linear subsystem.
- Box models containing static nonlinear and dynamical linear terms such as simple Hammerstein or Wiener models.
- Cascade models with single-valued static (no memory) nonlinear terms, such as the Wiener-Hammerstein cascade model.
- A nonlinear process that can be described by a semi-linear system with signal-dependent parameters where a static polynomial function is assumed for the parameters.
- Nonlinear dynamic models, which are difference equations linear in the parameters.

The simplifying assumptions for these five types of nonlinear systems lead to easier model estimation and identification, as the nonlinearities are typically stationary and single-valued (with no memory) whereas for general nonlinear systems these assumptions cannot be made.

For more general nonlinear systems, there are a few structural identification methods:

- Fuzzy Model approach where a linear model is first identified and then fuzzified (Sugeno and Kang, 1988)
- Neural Networks approach, which involves adjusting the weights of layers of 'neurons' (that form a network) to identify model structures (Narendra and Parthasarathy, 1990), effectively creating a large number of parametric nonlinear regressions.
- Genetic Algorithms (GA) and Genetic Programming (GP) involves the creation and evolution of model structure from a dataset and; in the former, a set of adaptive algorithms (Whitley, 1994) and in the latter, an existing function library that is chosen by the researcher (Gray, et al., 1998), with Evolutionary Algorithms effectively selecting the 'fittest' model structure. This family of techniques could apply to MSDP and will be evaluated in this context in the forthcoming studies.

The issue with Neural Networks and Fuzzy Models is that due to their complexity, even for simple models, they can be unreliable for making model-based predictions because of their high parameterisation leading to potentially large parametric uncertainty, not to mention the resulting long execution times. By contrast, the aim of this paper is to study a robust and generally usable method with a wide range of applications.

In general, when identifying the structure of a model an important assumption is made prior to choosing the MSI method: is the system linear or nonlinear? If the system is assumed to be linear, then any possible nonlinear aspect of that system will be ignored, and likewise if the system is assumed to be nonlinear, then the dominance of its linear modes may be overlooked, and an overly complex structure may be assumed. The methodology presented here considers both linearity and nonlinearity when identifying the structure of a system. This is particularly important when investigating state dependencies as the proposed MSI has the flexibility to consider the case of no-state dependency, i.e. a linear system, for a particular dataset. In addition, the methodology presented here considers true nonlinearity, in that the system is nonstationary / time variable.

In general, the TF model MSI algorithms find the statistically optimal model structure from the available time-series data by evaluating each possible model through the following process (Young, 2011):

- Estimate the parameters and output of a model candidate.
- Calculate an identification index value that statistically represents the likelihood that this model is the correct one for this data set, or a different model quality index.
- Repeat the above steps for all possible structure combinations.
- The highest or lowest, depending on type of identification index used, is considered the statistical optimum model structure for this data set.

For linear TF models the number of possible models depends on the maximum model polynomial orders (PO_{max}), the number of inputs and the maximum time-delay (δ_{max}) with MSI only needing to find the statistically optimum model polynomial orders and time-delay.

For nonlinear ESDP models the number of possible candidate models depends, in addition to the above, on the maximum number of potential states which could drive each of the parameters (S), and on the maximum number of dimensions (M_{max}), with MSI required to find the statistically optimum model polynomial orders and time delay, as well as how many, and which, states drive each parameter.

This additional complexity of model selection creates a non-trivial identification problem, potentially generating numbers of models which would be unrealistic to evaluate in finite time. This is why the polynomial orders are kept pragmatically low as not to introduce headaches and over-complicated models.

3.3.1 Number of Candidate Model Structures (N_{CM}) to Evaluate

Before an algorithm can be developed for MSI, an understanding of how many models are possible (in the sequel termed number of candidate models (N_{CM}) for more clarity) from a given configuration of hyper-parameters is required; number of regressors (R), S , M_{max} and δ_{max} . Note, to avoid over-complication the number of $n+m$ parameters from differing polynomial orders is restrained to $n=m$. In full implementation this need not be the case. This means R gives the number of parameters to estimate (1 parameter per regressor). If polynomial orders were not restrained to $n=m$, PO_{max} would also be a hyper-parameter to determine the $n+m$ parameters per model.

The total number of candidate models for a given set of hyper-parameters is given by (8) and Table 3.1 illustrates just how quickly the complexity of MSI becomes, as more state variables and higher polynomial orders are considered.

$$N_{CM} = \delta_{max} \times \left(S + 1 + \sum_{i=1}^{M_{max}-1} \left((S + 1 - i) \times \frac{(S-i)}{2} \right) \right)^R \quad (3.8).$$

Table 3.1. Showing how the number of candidate models (N_{CM}) varies depending on the maximum dimensional space and the number of possible state variables. Polynomial orders are restrained to $n=m$ and $\delta_{max}=1$

M_{max}	S	N_{CM}		
		1 st Order	2 nd Order	3 rd Order
1	3	16	256	4,096
1	5	36	1,296	46,656
2	3	49	2,401	11,7649
2	5	256	65,536	16,777,216
3	3	64	4,096	262,144
3	5	484	234,256	113,379,904

However, to create an algorithm to explore each candidate model, other equations are required; their number and form dependent upon R and M_{max} :

$$\text{Number of basic models } (N_{BM}) = (M_{max} + 1)^R.$$

For the class of models with NARX structure (4), where $A(\cdot)$ and $B(\cdot)$ are either functions of a number of states (SDP/MSDP) or are constants (coefficients).

$$y_t = a(\cdot)_{t-\delta} \cdot y_{t-\delta} + b(\cdot)_{t-\delta} \cdot u_{t-\delta} \quad (3.9).$$

When $M_{max}=1$ (as in SDP), $N_{BM}=4$: (numbers in brackets in the lists below are the number of states influencing that parameter, for example $A(0)$ would mean just a constant parameter, while $A(1)$ would mean A dependent on one state):

1. $y_t = a(0)_{t-\delta} y_{t-\delta} + b(0)_{t-\delta} u_{t-\delta}$

2. $y_t = a(0)_{t-\delta}y_{t-\delta} + b(1)_{t-\delta}u_{t-\delta}$
3. $y_t = a(1)_{t-\delta}y_{t-\delta} + b(0)_{t-\delta}u_{t-\delta}$
4. $y_t = a(1)_{t-\delta}y_{t-\delta} + b(1)_{t-\delta}u_{t-\delta}$

When $M_{max}=2$, $n_{BM}=9$, the 4 from above and the following:

5. $y_t = a(0)_{t-\delta}y_{t-\delta} + b(2)_{t-\delta}u_{t-\delta}$
6. $y_t = a(2)_{t-\delta}y_{t-\delta} + b(0)_{t-\delta}u_{t-\delta}$
7. $y_t = a(1)_{t-\delta}y_{t-\delta} + b(2)_{t-\delta}u_{t-\delta}$
8. $y_t = a(2)_{t-\delta}y_{t-\delta} + b(1)_{t-\delta}u_{t-\delta}$
9. $y_t = a(2)_{t-\delta}y_{t-\delta} + b(2)_{t-\delta}u_{t-\delta}$

When $M_{max}=3$, $n_{BM}=16$, the 9 from above and the following:

10. $y_t = a(0)_{t-\delta}y_{t-\delta} + b(3)_{t-\delta}u_{t-\delta}$
11. $y_t = a(3)_{t-\delta}y_{t-\delta} + b(0)_{t-\delta}u_{t-\delta}$
12. $y_t = a(1)_{t-\delta}y_{t-\delta} + b(3)_{t-\delta}u_{t-\delta}$
13. $y_t = a(3)_{t-\delta}y_{t-\delta} + b(1)_{t-\delta}u_{t-\delta}$
14. $y_t = a(2)_{t-\delta}y_{t-\delta} + b(3)_{t-\delta}u_{t-\delta}$
15. $y_t = a(3)_{t-\delta}y_{t-\delta} + b(2)_{t-\delta}u_{t-\delta}$
16. $y_t = a(3)_{t-\delta}y_{t-\delta} + b(3)_{t-\delta}u_{t-\delta}$

The basic models above may be repeated for different combinations of dependent states, depending on how many are available, for example when $M_{max}=2$ and $S=2$, basic models number 2,3,7,8 are repeated twice and the basic model 4 is repeated four times. The formula for the number of possible combinations depends on the basic model (see Appendix A, Table A.1 for each formula per basic model).

One aspect that is not taken into consideration here is the possibility of different time-delays for each regressor and each state as that would dramatically increase N_{CM} , which is already considerable (Table 3.1). Additionally, time delay might well be state-dependent, as may happen with catchment routing models (Li et al, 2013) and that too would significantly increase N_{CM} and the overall model complexity.

One thing that needs to be pointed out is the number of calculations involved in a single model evaluation. For linear MSI the parameters are constant and so only the order of $n+m$ calculations are made. However, for nonlinear MSI the parameters are not fixed and so the order of $N \times n+m$ calculations are made. With long series of high frequency data, where the application of nonlinear models would be useful, hundreds of thousands of samples are often used. This requires either long computational times or the utilisation of parallel computation.

However, limiting the model orders, including prior identification of the system's dominant modes using linear or simply time varying analysis techniques (Young, 1999a) should allow for a pragmatic two-stage procedure. Additionally, the search process is highly 'parallelizable', which is not difficult to implement using contemporary multi-processor computers and tools such as Matlab's Parallel Toolbox.

3.3.2 Evaluation of each Candidate Model: *Noise Variance Ratio (NVR)* vs. effective parameterisation level (number of degrees of freedom)

An information value needs to be calculated for each model evaluated in order to determine the optimum model structure; typically, this value is called Information Criterion (IC) and can be formulated in a number of ways (Akaike 1972; Sakamoto, et al., 1986; Young, 2011).

For the MSI methodology presented here the likelihood function (LF) was initially used and compared with Akaike IC (AIC):

$$AIC = 2 \times n_{par} - 2 \log LF \quad (3.10),$$

where LF is the likelihood function (calculated as part of the ESDP procedure) and the effective number of parameters (n_{par}), similar in concept to the number of degrees of freedom, is equivalent to a single time varying model coefficient is found from:

$$dt = 0.5 \times \left(\frac{1605}{NVR}\right)^{0.25} \quad n_{par} = \frac{N}{dt-2} \quad (3.11).$$

dt can be interpreted as a subsampling ratio (when $dt=2$ every second sample is taken with the accompanying loss of temporal resolution – filtering/smoothing) (Ng and Young, 1990), and it reflects the time-scale of variability of the parameter. Naturally, n_{par} is calculated for each of the TVP (3.10) separately, so the AIC value would be calculated using the total of these values for all the TV parameters.

Normally ESDP optimises the NVR during parameter estimation convergence, however for the proposed MSI methodology the parameters for each candidate model are estimated once with a fixed NVR . This greatly reduces computational time and would render (3.11) irrelevant as the NVR never changes, thus AIC and LF performed identically in terms of model identification.

However, through extensive simulation testing using a broad range of both NVR values and nonlinearities within the NARX class of models, it was found that the total number of dependent states had a stronger influence over which model was selected than the effective number of parameters n_{par} , *i.e.* models with higher numbers of states were being chosen over models with fewer states, resulting in misidentification. This required the introduction of a heuristic weighting value (3.12) to offset this bias which was obtained empirically from a review of a broad range of NARX model structures using simulated data to warrant reproducibility.

$$w = 0.5 \times \log_{10}(1 + N_S) \quad (3.12),$$

where, N_S is the total number of states used in the candidate model and then the overall IC equation becomes:

$$IC = LF + 0.5 \times \log_{10}(1 + N_S) \quad (3.13).$$

The IC of all evaluated models is sorted and the highest value corresponds to the model structure considered to be statistically optimal.

3.3.3 Sensitivity of NVR vs. Smoothness of Data vs. Complexity of the Model Sought

Through a search of a broad range of NVR values for several scenarios (such as the simulated examples above, and a complex nonlinear industrial benchmark), it was found that the proposed IC was robust enough to identify the correct model structure for a wide range of $NVRs$ (10^{-10} : 10^{-2}). However, for hydrological data (such as that found in Section 4), it was found that higher $NVRs$ resulted in a different model structure being identified to lower $NVRs$, e.g. for the example in Section 4, $NVRs$ between 10^{-10} : 10^{-4} identified one model structure and $NVRs >10^{-4}$ identified a different model structure.

This is logical as simulated examples are typically simple and built to have one timescale whereas hydrological data is composed from a number of cycles and system dynamics meaning applying different timescales results in different models. Selecting an *NVR* is similar to selecting a timescale as low *NVRs* only enable identification of the broader picture and high *NVRs* provide all the temporal detail seen in frequently sampled data.

3.3.4 The Need for Parsimonious Models

As is clearly shown in Table 1, the number of models becomes prohibitively large if too complex models are specified. This is for fixed, pre-selected *NVR* values. The problem would be compounded if optimising *NVR* values. This is due to the fact that for each model *NVR* is optimised (Maximum Likelihood), and the ML objective function response surfaces become very flat for over-parameterised (mis-specified) model structures, which makes the optimisation much slower. This reinforces the need for parsimonious modelling approach (Young, 2011), not unique to the realm of nonlinear models, but particularly critical in this area.

3.3.5 Simulated Example

A model of NARX structure similar to (3.4) was simulated:

$$y_t = a(\mathbf{s}_{3,t}, \mathbf{s}_{2,t})_t \cdot y_{t-1} + b(1)_t \cdot u_{t-1} \quad t = 1, \dots, N \quad (3.14),$$

where, $a(\mathbf{s}_{3,t}, \mathbf{s}_{2,t})_t = e^{-s_{3,t-1}^2 - s_{2,t-1}^2}$ and $b(1)_t = 1$, with a 5% noise level by standard deviation.

All the variables were supplied to the MSI algorithm; y_t , u_{t-1} , \mathbf{s}_1 , \mathbf{s}_2 , \mathbf{s}_3 and \mathbf{s}_4 along with hyper-parameters $\delta_{\max}=4$ and $NVR=1e^{-2}$. States \mathbf{s}_1 , \mathbf{s}_2 , \mathbf{s}_3 and \mathbf{s}_4 are simulated as uncorrelated Gaussian sequences with zero mean and unit variance. 484 models were evaluated and formed the IC values (Figure 3.8) that correctly identified model structure number 35 with a time-delay of 1.

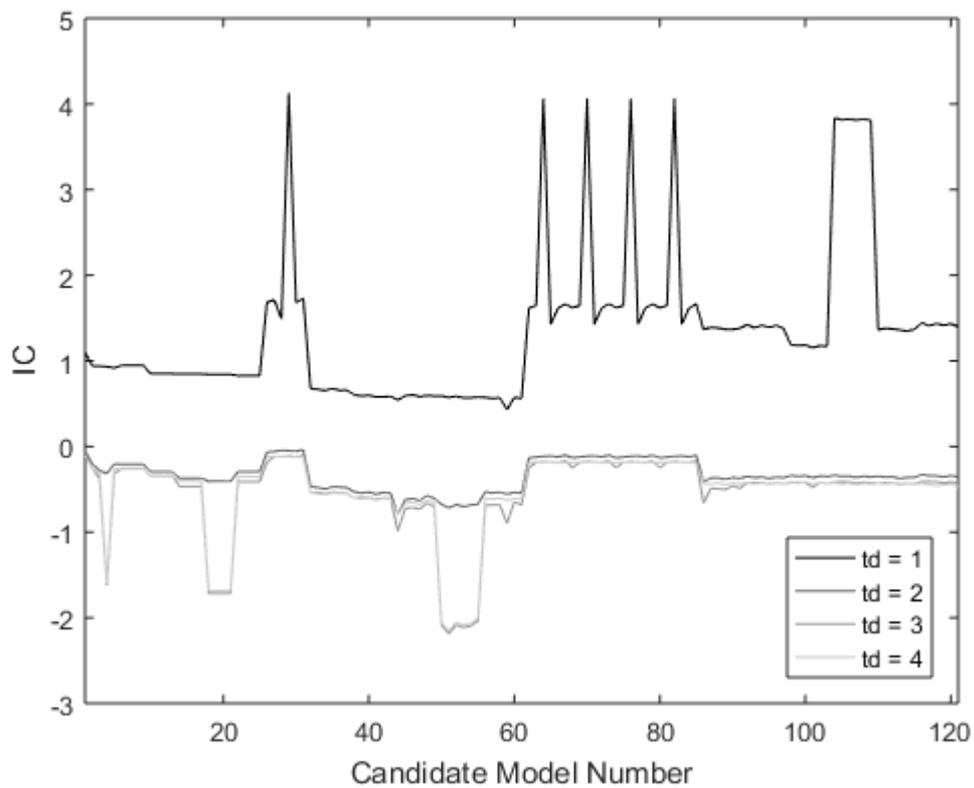


Figure 3.8. Information Criterion values from the proposed MSI methodology. Highest peak is for model structure number 29 with a time-delay of 1, which corresponds to the model structure in (3.14).

The key observation (Figure 8) is the sensitivity to time-delay, making identifying the time-delay the easiest part of MSI. The four other peaks correspond to candidate models with the same state configuration for a but with b being a single SDP driving by s_1 , s_2 , s_3 and s_4 respectively.

After identifying the structure, the model was then estimated using ESDP producing a fit a R^2 of 0.999 and estimated the a parameter (Figure 3.9) and constant b of 1.

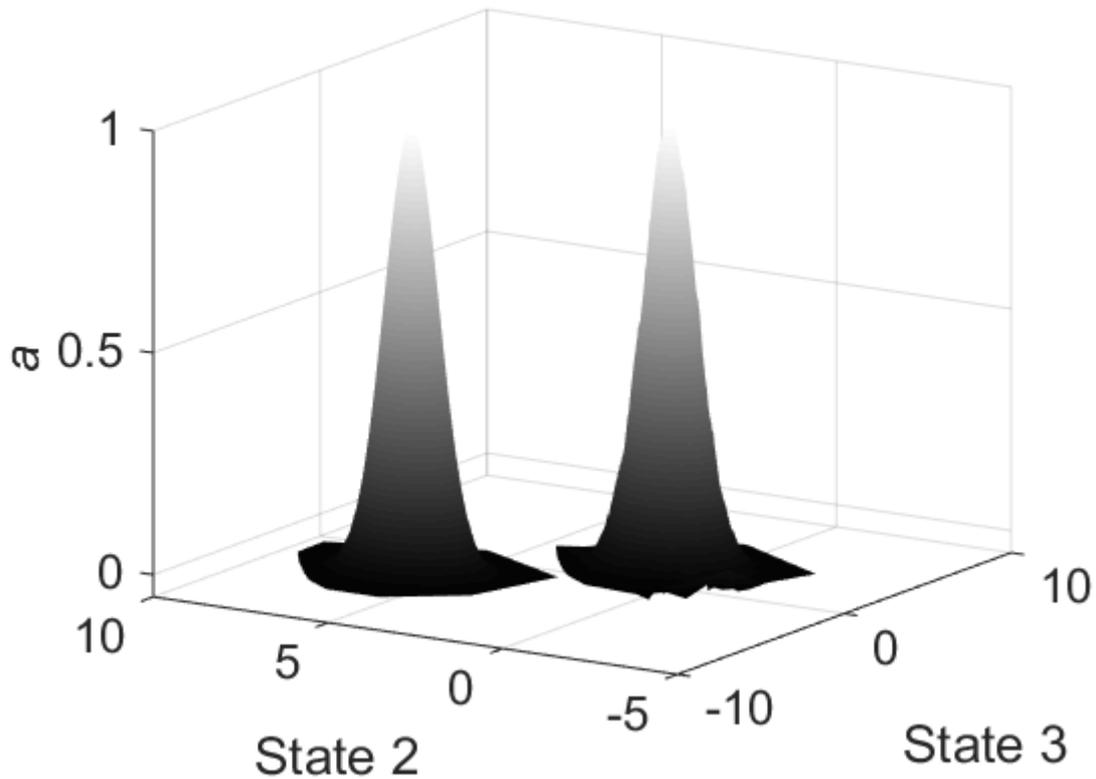


Figure 3.9. Illustration of the a parameter. Left hand side, simulated a parameter (offset for clarity). Right hand side, estimated a parameter.

3.4 Streamflow Generation Example (Objective E)

A commonly used hydrological model in flood simulation is $Q = \mathcal{F}(R)$, where R is rainfall within a catchment, Q is the streamflow per unit basin area (or ‘channel runoff’) and f is a functional representing the dynamic processes involved in translating rainfall into runoff. This model may be described by the equations (1 or 2) in atypical catchment systems where a linear relationship is observed (Ockenden and Chappell, 2011). Rainfall-runoff is however, often a nonlinear process (McIntyre and Al-Qurashi 2009; Beven, 2012), so an SDP methodology may be applied to capture this nonlinearity. Since it is not always clear whether nonlinearity in rainfall-runoff response is caused solely by temporal variations in the moisture state of the whole active hydrological system including rock aquifers, rather than a combination of nonlinear hydrological processes also related to differing rainstorm characteristics (Rodríguez-Iturbe et al., 1982; Chappell et al., 2017), changes in wet-canopy evaporation during storms (Love et al., 2010), diurnal solar radiation effects on soil moisture via transpiration dynamics (Deutscher et al., 2016) or temperature effects on snow melt (Loheide et al., 2009; Mutzner et al., 2015). There is therefore an intrinsic value in applying an ESDP methodology to discover the dominant mechanism active in the current data set - strictly following the DBM philosophy.

Similar to equation (3.9) we have the following nonlinear rainfall-runoff model:

$$Q_t = a(\cdot)_{t-\delta_1} Q_{t-1} + b(\cdot)_{t-\delta_2} R_{t-\delta_3} \quad (3.15),$$

where $a(\cdot)$ and $b(\cdot)$ are defined as in (3.9), with SDP coefficients. The base-line of the flow series is estimated at the same time, in this case as a constant, within the same estimation process. The aim

here is to use MSI to identify the most statistically likely state variables, if any, driving parameters a and b and then to apply ESDP to optimise these parameters and finally investigate the parameter-state surfaces for any meaningful physical interpretation of the rainfall-runoff processes.

Even though δ_1 , δ_2 and δ_3 could theoretically be different values, for these examples they will all have the same value, determined by MSI and so similar to the commonly used Nash rainfall-runoff model (Nash, 1957). This is done to reduce the number of models MSI needs to evaluate to in order to minimise model complexity and to reduce runtime.

The examples that follow are using 15-minute interval data collected from the 0.76 km² Nant-y-Craflwyn catchment in upland Wales, UK (Jones and Chappell, 2014; Jones et al., 2014). The available variables are: rainfall, streamflow per unit basin area ('channel runoff'), air temperature, stream temperature and solar radiation (Figure 3.10). Rainfall, air temperature and solar radiation were measured at an automatic weather station while stream temperature and streamflow were measured at water quality station and flume, respectively (Fig 1 in Jones and Chappell, 2014).

This means the number of candidate models to evaluate is 768 ($n_{CM}=256$ and assuming one time-delay with $\delta_{max}=3$). By contrast if each time-delay was to be different then $n_{CM}=6912$.

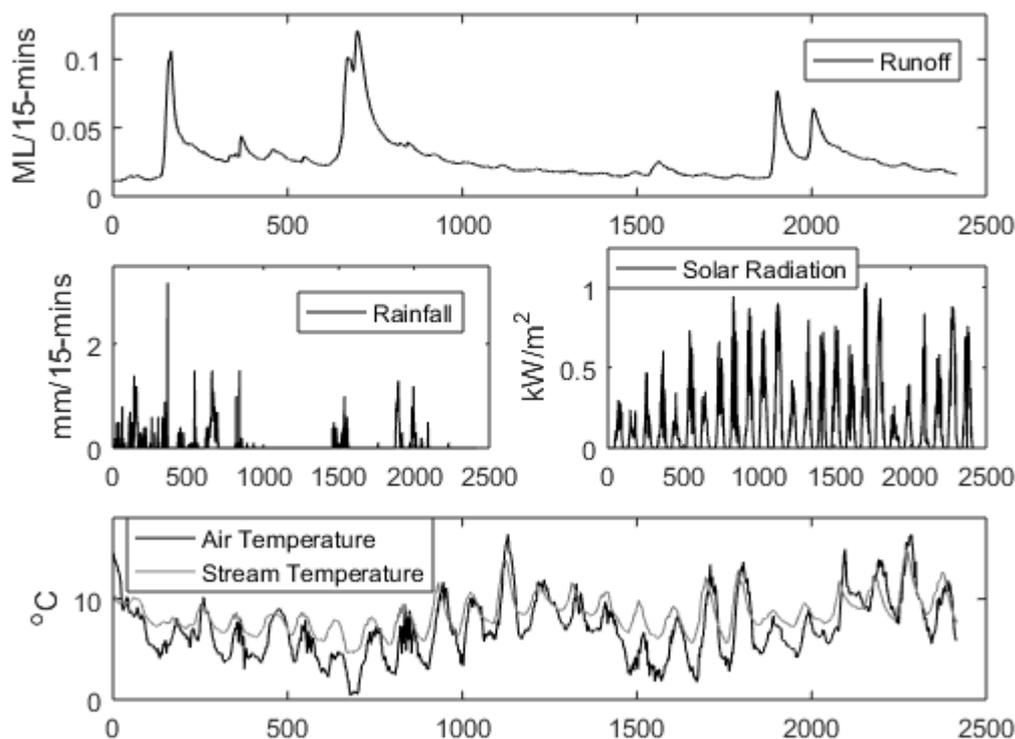


Figure 3.10. Available variables over a 25-day period starting 28/03/2012 in 15-minute interval steps.

One possible feature in streamflow time-series are diurnal (or diel) cycles, even though these may be visible only between storms in the absence of rainfall. These cycles may be caused by diurnal cycles in snow melt or ice melt (Mutzner et al., 2015) or diurnal cycles in transpiration (Deutscher et al., 2016), driven primarily by the diurnal cycle in solar radiation inputs. Within the 15-minute streamflow data for the Nant-y-Craflwyn catchment there is a visible diurnal cycle that corresponds to the solar radiation data (Figure 3.11). The MSI approach is evaluated to see if solar radiation can

be combined with rainfall data to derive a more complete DBM model of the drivers of streamflow response at this location.

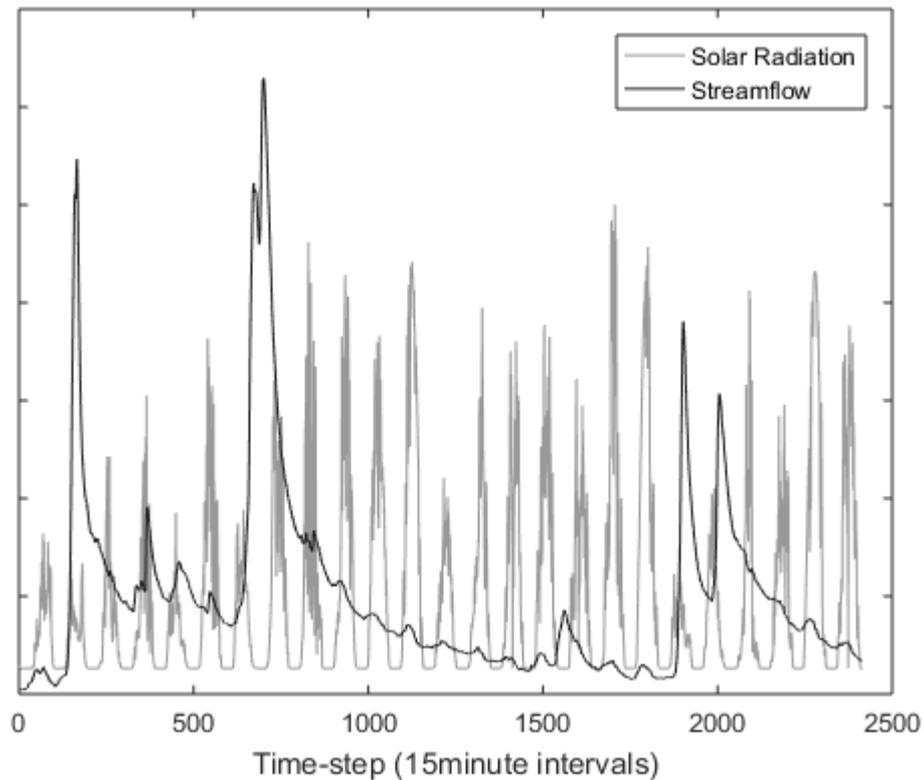


Figure 3.11. Showing the diurnal-modulated nature of the streamflow records for the Nant-y-Craflwyn stream as compared to that found in the solar radiation (both have been standardised for easy of comparison). This is picked-up by the state dependency estimate of the a parameter.

Hourly sampling rate has been chosen for this model to avoid oversampling. The 15-minute data was converted into hourly data, where the rainfall was totalled over the hour and the other variables were pre-processed using an Integrated Random Walk (IRW) smoothing and decimation algorithm from the Captain Toolbox to limit their spectrum to avoid bias (Young et al., 2007).

MSI identified a structure with the decay coefficient dependent on solar radiation and rainfall with time-delays of 1 hour. ESDP produced an optimised model with an R^2 of 0.998, estimated A parameter with range 0.95-1.06 (dimensionless) and constant b (referred to as 'effective rainfall coefficient' in Young 2003) of 2.3×10^{-3} with units consistent with observations of discharge.

The outliers for the 2D parameter estimates made visualisation of the parameter surface difficult, probably due to observation disturbance in measurements of all the variables in the model. Smoothing the 2-D parameter estimate using Smoothing Splines ANOVA (SS ANOVA; Gu, 2013) allows a useful representation of the surface (Figure 3.12) and its subsequent parameterisation for simulation purposes, conveniently including a measure of uncertainty of the smoothed surface.

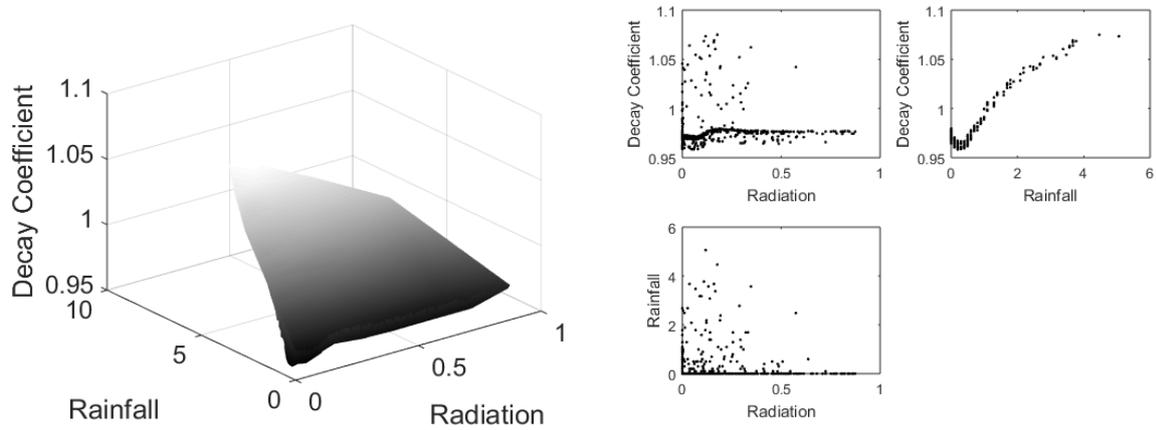


Figure 3.12. Decay coefficient and its two dependent states. Left: 3D visualisation of the decay coefficient and the states. Right: projection plots of each state/decay coefficient vs the other state/decay coefficient.

Another way to visualise the 2-D parameters is to compare their changing value with the values of their two potentially explanatory states (Figure 3.13) as time series, this not only shows how the diurnal cycle in the solar radiation has been captured by the a parameter on the shorter time scale, but also how it is superseded by the effect of rainfall on the longer time scale. The a parameter is generally lower than one and can be interpreted as a momentary decay coefficient, with values higher than one meaning momentary 'growth' or 'production'.

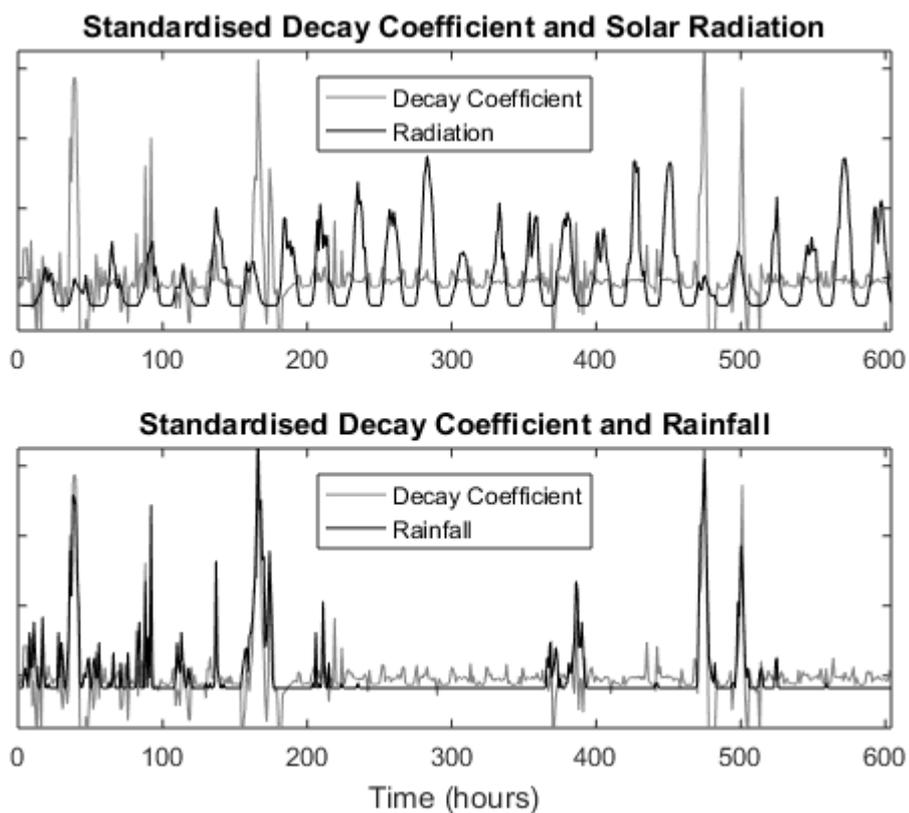


Figure 3.13. Comparing the decay coefficient estimates to its driving states. Top, comparing the standardised values of the decay coefficient and radiation. Bottom, comparing the standardised values of the decay coefficient and rainfall.

It is worth noting that this interpretation of the a coefficient is not the same as for Linear Time Invariant systems (LTI) where the value larger than one means the system is unstable, and the values lower than one relate to the recession process of the autonomous (no input) solution. This does not hold for nonlinear (time varying) systems, which are analysed here. Such values of similar SDP parameters have been seen in previous publications (Young, 2000; Young, 2003), while Ratto et al (2007) have limited the a coefficient to not go above one, based on their specific application. In this case leaving a to track unconstrained is a part of the DBM identification process; any physical constraints can be included at the final application stage.

As noted earlier, the apparent diurnal cycle in streamflow may be caused by diurnal cycles in transpiration losses from the subsurface pathways generating streamflow (Graham et al., 2013; Deutscher et al., 2016) or diurnal cycles in air temperature affecting the production of snowmelt (Loheide et al., 2009), but may also be caused by residual thermal artefacts in the pressure transmitter output (Liu and Higgins, 2015; Moore et al., 2016). For the diurnal cycles in the Nant-y-Crawlyn time-series, *in-situ* field tests show that all of the cyclical behaviour may be explained by residual thermal artefacts. The modelling approach would, however, have been able to quantify the diurnal component of response if it had been caused by transpiration or snowmelt effects, as seen within other catchments.

3.4.1 Validation

The validation processes described in section 3.2.2 were applied to the above hydrological data and showed good model validity (Figure 3.14) with the one-step ahead and full model simulation having R^2 of 0.9900 and R_t^2 of 0.9883 respectively.

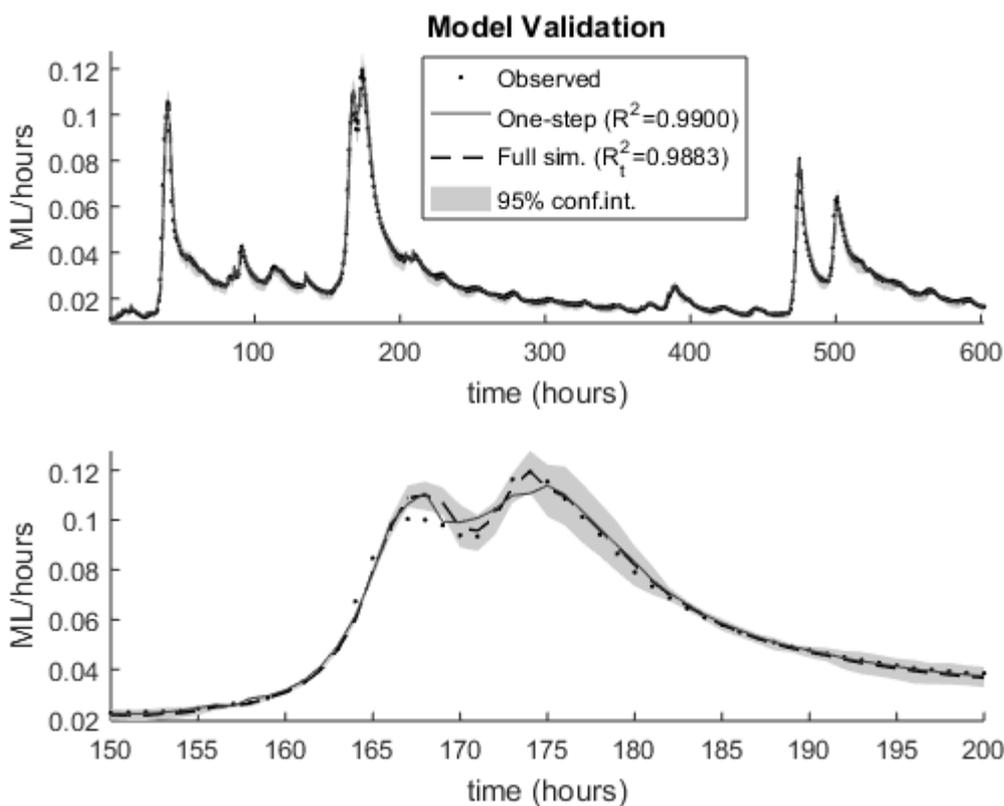


Figure 3.14. One-step ahead and full model simulation with uncertainty: 95% percentile band of MC simulations using the estimated parameter standard error with normally distributed parametric variation.

3.5 Conclusions

The paper unified and improved upon the SDP and MSDP methodologies to form a Generalised State Dependent Parameter methodology. In addition, this paper introduced a generalised Model Structure Identification methodology that allows for a closer following of the DBM approach to investigating nonlinear system dynamics. The introduction of the arbitrary RW not only generalised the methodology but also improved the parameter estimates.

The MSI approach to identification of this class of nonlinear systems allows researchers to statistically explore interactions between system variables for a specific process and to give insights into which variables are more important to that process. It also enhances the DBM nature of ESDP by more rigorously determining the model structure, not relying on the researcher's preconceptions. The 'brute force' nature of the algorithm could lead to computational runtime limitations if dealing with many variables and very long data sets, but this is a minor limitation given the ever-increasing power of computers and the option of parallel computation.

The parameter map allows the outputs of ESDP to be quickly utilised for scenario analysis and on-line simulation of live events.

The approach was evaluated and validated in several ways, including the use of separate data sets, regressive, one step-ahead and full simulation model fit.

3.5.1 Further Developments

Two groups of developments aimed at improving the usability and applicability of the MSI/MSDP methods may be identified at this stage.

1. Generalisation of the MSI methodology:
 - Removal of the constraint of $n=m$ on polynomial orders – greatly increasing the number of models to evaluate but allowing closer following of the DBM approach.
 - To other models (multiple inputs, ARMA structures etc.), with the possibility of leading to the future development of an algorithm that chooses the model type as well as the structure.
2. Numerical effort is a serious consideration in this case. Applicability of the method, which naturally involves a structure search and optimisation based estimation, is limited in the current implementation by the time taken by the algorithm, mainly the structural identification search. Thus:
 - Only single-input-single-output models were demonstrated but the extension into multiple input and even multiple output is possible – it would drastically increase the number of models to evaluate as seen from Table 1.
 - Currently time-delays are fixed, in that all time-delays have the same value. Allowing MSI to evaluated models with several different time-delays would enhance the exploration of nonlinear dynamics – it would dramatically increase the number of models to evaluate.
 - For some systems time-delay can be state dependent and so introducing state dependent time-delay would benefit the modelling of those systems – it would also heavily increase the number of models to evaluate.

Chapter 4: Incorporating Arbitrary Sampling Technique into Dynamic Harmonic Regression – OBJECTIVE 4

The arbitrary sampling technique could also be applied to DHR, to remove the requirement of uniformly sampled data, and resulted in a new methodology known as ASDHR. This work (**paper 3**) was published in the Journal of Environmental Modelling and Software (**Mindham and Tych, 2019**), highlighting its importance.

4.1 Introduction

In analysing environmental data and modelling environmental processes there is a significant need to identify and estimate trends, cycles, and seasonal components. Dynamic Harmonic Regression (DHR) provides a cogent analytical tool to generate such results. It is, however, firmly based in the classical time series analysis domain and relies on the data being sampled at specific intervals. However, in many disciplines dealing with the natural environment, data-sets are not sampled at regular intervals. Presented here is a significant update to DHR allowing direct use of irregularly sampled time series data in estimation of trends, cycles and seasonal components. In addition, there is a distinct need for such a method to be accompanied by software that is easy to use and with results directly interpretable in the terms of the specific discipline, be it climatology, hydrology or environmental chemistry, to cite a few of the disciplines where these methods have been successfully applied by the authors.

Dynamic Harmonic Regression is a nonstationary time-series analysis approach used to identify trends, seasonal, cyclical and irregular components within a state space framework (Young, 1989). The DHR method is implemented in the CAPTAIN Toolbox for Matlab and has been used extensively by many researchers (Young et al., 1999, Taylor et al., 2007). The DHR methodology has a wide range of applications, and is particularly useful in analysing environmental data; such as atmospheric pollutants (Becker et al., 2006, Venier et al., 2012) where, importantly, it is cited as a recommended method in the 2011 UNEP Air Report in the Persistent Organic Pollutants section (UNEP, 2011). Other significant applications include paleoclimatology data based on isotope dating (Smith et al., 2016), impacts on catchment water balance (Chappell and Tych, 2012), groundwater-surface water fluxes (Keery et al., 2007), geomorphology (Carling et al., 2005), water quality cycles (Halliday et al., 2013), or solar irradiation forecasting (Trapero et al., 2015), but also in forecasting of phone-call numbers within the call-centre context (Tych et al., 2006), medicine (Sofianopoulou et al., 2017) and finance (Bhar, 2010). However, when data are irregularly sampled, the existing DHR and related methods cannot be applied directly. For instance, paleoclimatic data-series from core samples and speleothems are interpolated onto a fixed time-base prior to analysis (e.g. Smith et al., 2016). Historic water quality data, geomorphological data and atmospheric chemistry data (e.g. Becker et al., 2006, Carling et al., 2005) are treated in the same way using prior processing.

The problem is that the original state-space filtering-based DHR cannot handle irregular sampling without applying resampling techniques making the time sampling uniform prior to the analysis. As useful as resampling is, it is still manipulation of the observed data and leads to increased uncertainties in model outputs and to potential artefacts resulting from the interpolation techniques applied, such as aliasing (Chappell et al., 2017) or spectral features of the approximation functional base applied in the interpolation process. Importantly, where the actual samples become sparser it can lead to 'false certainty' - introducing interpolated samples where there are no data available. The uncertainty estimates then become tainted, usually unduly lower. Conversely, where samples are denser it can lead to a removal of information, also leading to increasing uncertainty (fewer samples - less averaging) and potentially losing information in the upper part of the signal spectrum.

Common anti-aliasing methods for down-sampling (such as low-pass filtering) will have the latter problem.

With interpolation, resampling is a step away from the Data Based Mechanistic approach (DBM, Young, 1999) to modelling and data analysis which DHR is designed to be consistent; allowing the observed data to tell us about the systems prior to process interpretation. This is because interpolation always has an underlying model or assumptions, which may form an introject affecting the data.

It has to be pointed out that the developed algorithm is not aimed purely at dealing with irregular sampling, but at augmenting the existing DHR model which has proven to be highly effective and widely used in Environmental Science due to the natural interpretation of its object and of the model components, as well as the inherent stochastic information provided by it.

Analysing irregularly sampled time series data has a large body of literature addressing it. Irregularly sampled Auto-Regression is one of them. Broersen et al. (2004) derive a method for handling AR models with missing samples (a very specific and limited form of irregular sampling – with missing samples the sampling is regular). In general, AR and other methods relying on solving stochastic equations (such as Brockwell's (2001) Levy process driven approach) are not directly comparable with the proposed technique because they are usually much more general, and so rely on additional mathematical analysis and assumptions in every specific application. Other approaches to irregularly sampled time series address whole spectrum estimation (such as irregularly spaced approaches to Fourier estimation, such as O'Toole et al. (2007) or wavelet-based approaches of e.g. Mathias et al., 2004), which is exactly what is avoided here in order to reduce the uncertainty of results. The reduction of uncertainty in the presented approach is achieved through minimising the number of spectral components that are estimated, to only the dominant periodicities. Various machine learning approaches tend to require high data volumes and suffer from difficulty with obtaining justified uncertainty estimates. We work with often expensive to obtain environmental data sets of necessarily limited length, and normally address univariate time series, so direct comparison with most published machine learning approaches is not easily achieved.

The term "arbitrary sampling" is introduced in the specific DHR context and used to describe how the irregular sampled time-series is used within the irregular sampled DHR technique. The temporal distance between each sample in the irregularly sampled time-series is stored as a $1 \times (n-1)$ vector complementing the irregularly sampled time-series itself. The term 'temporal distance' is used here deliberately to highlight the possibility of using this technique in analysis of spatial series, as for example in Carling et al. (2005) where DHR was used on regularised spatial data to analyse the pool-riffle sequence in river geomorphology. The arbitrary sampling processing uses both series – the measured values and their sample times. This approach can be used also for sparse regularly sampled series with many missing values, where the time index for the missing values are removed, creating an irregular sampled time-series. The creation of the temporal distance vector is the only pre-processing required for the updated DHR and for the purposes of differentiating between the current DHR methodology and the proposed updated methodology, the latter will be referred to as 'Arbitrary Sampled Dynamic Harmonic Regression' (ASDHR).

The update implements an arbitrary sampling technique in the Kalman Filter (KF) and Fixed Interval Smoother (FIS) algorithms. While the irregular sampling has been previously used with Kalman Filtering (e.g. Li et al., 2008), the FIS algorithm implementation here is a novel element, not used elsewhere (except for Mindham et al., 2018) and necessary for the use of ASDHR. Overall the Arbitrary sampling technique eliminates the need for any pre-analysis or resampling and puts DHR

back in line with the DBM approach by not inserting any assumptions or artefacts into the observed data.

The aim of this paper is to introduce the arbitrary sampling technique and to demonstrate the benefits of ASDHR for analysis of environmental data sets, which so often are irregularly sampled or contain numerous missing values. This is achieved by:

- Providing a brief background to DHR and then introduce the arbitrary sampling methodology (Section 4.2.0)
- Demonstrating the capability, benefits and necessity of ASDHR when using environmental data:
 - Paleo-climatology (Smith et al., 2016) – comparing ASDHR and DHR outputs to demonstrate the arbitrary sampling capability (Section 4.3.1).
 - Persistent organic pollutants (Becker et al., 2006) – demonstrating the necessity for extending DHR to accommodate irregular sampled time-series (Section 4.3.2, 4.3.3) especially for noisy data series.
 - Forecasting Atmospheric CO₂ – introducing and demonstrating ‘arbitrary forecasting’, the ability to forecast at arbitrary points into the future with different sampling times to the observed data (Section 4.3.4).
- Evaluation of ASDHR robustness to data sparseness and observational noise (Section 4.4.0)

4.2 Dynamic Harmonic Regression

The DHR model assumes that the observable variable of a system is composed of four components (1); trend (T), sustained cyclical (C) with period different to the seasonality, seasonal (S) and white noise (e) (Young et al., 1993).

$$y_t = T_t + C_t + S_t + e_t \quad (4.1).$$

The measured values of y are the output (observations) series of a system of stochastic state space equations, which can then be broken down to allow for estimation of the four components.

T_t is the trend component, which can be considered a stochastic time-varying ‘intercept’ parameter and is interpreted spectrally as a zero frequency term ($i=0$, where ω_0 or $f_0 = 0$), in practice - occupying the lowest part of the spectrum, and modelled as Integrated (or Smoothed) Random Walk (see Young et al., 1999) with states termed level and slope of the trend.

The seasonal component S_t is defined as:

$$S_t = \sum_{i=1}^{R_s} \{a_{i,t} \cos(\omega_i t) + b_{i,t} \sin(\omega_i t)\} \quad (4.2),$$

where $a_{i,t}$ and $b_{i,t}$ are stochastic Time-Varying Parameters (TVP) and ω_i are the fundamental and harmonic frequencies associated with the seasonality in the series ($i=1,2,\dots,R_s$).

$$C_t = \sum_{i=1}^{R_c} \{\alpha_{i,t} \cos(f_i t) + \beta_{i,t} \sin(f_i t)\} \quad (4.3),$$

where $\alpha_{i,t}$ and $\beta_{i,t}$ are stochastic TVP and f_i are the frequencies associated with the longer cyclical component ($i=1,2,\dots,R_c$). The cyclic component C_t has an identical definition to the seasonal and is isolated here to allow for a different physical interpretation.

Whereas the white noise component e_t is the remaining information after the other 3 components have been removed from y (i.e. model residuals). Note that the full Unobserved Components Model (Young et al., 1999) also incorporates the Irregular component, here omitted for simplicity.

Typically, the TVP ($a_{i,t}$, $b_{i,t}$, $\alpha_{i,t}$, $\beta_{i,t}$ and both T_t states) are defined by a two dimensional state vector $x_{i,t} = [l_{i,t}, d_{i,t}]^T$, where $l_{i,t}$ and $d_{i,t}$ are, respectively; the changing level and the changing slope, of the associated TVP. The stochastic evolution of each $x_{i,t}$ is assumed to be described by a generalised random walk process (4.4).

$$x_{i,t} = F_i x_{i,t-1} + G_i \eta_{i,t-1} \quad i = 1, 2, \dots, R \quad (4.4),$$

where, $R = 1 + R_c + R_s$, and F and G defined in their time-varying form in (4.7a) and (4.7b) respectively (see also Young et al., 1999 for fixed form).

4.2.1 State Space and Observation Equations

The state space model is constructed by aggregation of the subsystem matrices defined in (4.2) and is defined in Young et al. (1999). However, both the state transition and noise-input matrices are fixed and thus can only work for uniformly sampled data, hence the need to apply regularisation techniques on irregularly sampled data. The method proposed here replaces these fixed matrices with time-step dependent ones, where the values depend on the time between each sample; thus, allowing them to work for irregularly sampled data.

For the rest of the paper, the temporal positioning of samples (t) at regular intervals Δt is replaced with the arbitrary positioning of samples (Δt_k), where k – the sample number - keeps the temporal order.

If $y^{(q)}(\mathbf{y}_k)$ is the q^{th} derivative of $y(\mathbf{y}_k)$, and the form of the function $y(\cdot)$ is not specified, a data point distant from \mathbf{y}_k provides very little information about $y(\mathbf{y}_k)$. Using the local polynomial modelling reasoning (e.g. Fan and Gijbels, 1996) only the local data points in the vicinity of \mathbf{y}_k are used. Assuming $y(\mathbf{y}_k)$ has the $(q+1)^{\text{th}}$ derivative at the point \mathbf{y}_k , then following Taylor's expansion for y in the local neighbourhood of \mathbf{y}_k we have:

$$y(\mathbf{y}) = y(\mathbf{y}_k) + y'(\mathbf{y}_k)(\mathbf{y} - \mathbf{y}_k) + \frac{y''(\mathbf{y}_k)}{2!}(\mathbf{y} - \mathbf{y}_k)^2 + \dots + \frac{y^{(q)}(\mathbf{y}_k)}{q!}(\mathbf{y} - \mathbf{y}_k)^q \quad (4.5).$$

If the value of y and its derivatives are known at the t^{th} point as $\mathbf{x}_k = [y(\zeta_k) \ y'(\zeta_k) \ y''(\zeta_k) \ \dots \ y^{(q)}(\zeta_k)]^T$ and the highest derivative of $y(\mathbf{y})$ with respect to \mathbf{y} with $\mathbf{y}_k = y(\zeta_k)$ and $y^{(q+1)}(\zeta_k) = \eta_k$, where $\eta_k \sim \mathcal{N}(0, \sigma_\eta^2)$ and ζ_k is the approximation point (knot) at sample k , then Taylor's expansion (4.3) can be applied in the local neighbourhood of ζ_k for all derivatives of y resulting in the GRW model with state space (4.6a) and observation (4.6b) equations with the now time-varying state transition (in 7a) and system noise-input (in 4.7b) matrices:

$$\mathbf{x}_k = \mathbf{F}_k \mathbf{x}_{k-1} + \mathbf{G}_k \eta_{k-1} \quad (4.6a),$$

$$\mathbf{y}_k = \mathbf{H}_k^T \mathbf{x}_k + e_k \quad (4.6b),$$

where H_k is the observation matrix (dimension of $n \times R$ for RW trend and RW harmonics' amplitudes). Observation equation (4.6b) implements the regressive structure of DHR, with x_k being the estimated amplitudes of harmonics (or trend levels) for each k and their corresponding elements of H_k contain the i -th's harmonic values ($\cos(\omega_i t_k)$ or $\sin(\omega_i t_k)$) or ones for the trend level.

$$F_k = \begin{bmatrix} \alpha & \Delta_k & \frac{\Delta_k^2}{2!} & \dots & \frac{\Delta_k^q}{q!} \\ 0 & 1 & \Delta_k & \dots & \frac{\Delta_k^{q-1}}{(q-1)!} \\ 0 & 0 & 1 & \dots & \frac{\Delta_k^{q-2}}{(q-2)!} \\ & \vdots & & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix} \quad (4.7a), \quad G_k = \left[\frac{\Delta_k^{q+1}}{(q+1)!} \quad \frac{\Delta_k^q}{q!} \quad \frac{\Delta_k^{q-1}}{(q-1)!} \dots \Delta_k \right]^T \quad (4.7b),$$

where, $\Delta_k = \zeta_k - \zeta_{k-1}$ is the temporal distance between the temporal samples number k and $k-1$ (time difference between the knots ζ_k and ζ_{k-1}). When $\alpha = 1$ and $q = 0, 1, 2, \dots$ this equation describes, respectively, random walk, integrated random walk, double integrated random walk, etc. When α ranges between 0 and 1 referring to smoothed random walk with orders q as above.

The choice of Random Walk order depends on the application; where variation is expected, q of 1 or higher is used naturally depending on the shape of the trend, where no assumptions are made and suspicion of stationarity needs to be evaluated, the simple RW model ($q=0$) is usually more appropriate. In practice the analysis starts with q of 0 for the harmonic components and 1 for the trend.

4.2.2 Algorithmic Considerations

To reiterate, the term ‘arbitrary sampling’ is used because the new state space equations (section 4.2.1) do not constrain the time between each sample to be fixed, as long as they are in temporal order; the regularity or amount of irregularity in the sampling process does not impact the modelling process so long as the temporal positions of each sample are known and the finite Taylor expansion (4.5) is a satisfactory approximant of y .

To keep the algorithm numerically well defined (keeping the spectrum of the transition matrix “sensible” as its inverse is used in the smoothing algorithm), the temporal distances between each sample (Δ_k) may need scaling to keep the ‘majority’ or average sampling rate close to one. This is equivalent to choosing the time unit suitable for the analysis, e.g. if the sampling process is typically once a week, then Δ_k provided in days needs to be divided by seven. In formal terms, as Δ_k affects the spectrum of the transition matrix F_k and its invertibility (condition), the time must be expressed in the units that will not cause poorly conditioned F_k for any (or at least for very few) steps k . In the state space matrices (section 2.1) a Δ_k of 1 for all k implies regular sampling as a special case.

4.2.3 Variance Intervention

Amongst the numerous algorithmic advantages of the Stochastic State-Space techniques, such as the ease of forecasting, smoothing and interpolation, including arbitrary times between the existing samples, the variance intervention technique seems particularly well suited to environmental data analysis and modelling. Very often the researcher is looking for confirmation or detection of a discrete change in the system. Variance intervention technique has been introduced by Box and Tiao (1975) and in the context of stochastic state space models it was developed and evaluated by Young and Ng (1989). With regular DHR, intervention points are used to account for abrupt changes in the data, such as a sharp calibration change or a shift in environmental system behaviour (e.g., Chappell and Tych, 2012). It can be used to model and evaluate the potential for specific breaks in the time series, whether background level, slope changes or sudden amplitude changes of the harmonic components in the dynamic harmonic regression context. Without intervention points any abrupt changes are smoothed in the model estimate and give poorer models that are not true to the system, not reflecting the mechanisms governing the observed processes and so not consistent with

the DBM approach. In Bayesian terms interventions amount to introducing diffused priors at the intervention *a-priori* step, causing the recursive estimator to “doubt” the current estimate by increasing the covariance matrix significantly.

Introducing intervention points requires either assumptions or knowledge of the time of the change. Alternatively, a search for a significant parametric change may be made using a sliding intervention technique, as applied in Chappell and Tych (2012) to detect discrete changes in streamflow and evaporation records due to forest cover change, or to other such interventions.

This advantage is not lost with the introduction of the arbitrary sampling technique, as one of the examples below demonstrates.

4.2.4 Period Identification

In many environmental data series there is a need for identification of dominant periodicities, as these are likely to indicate the phenomena active or dominant in the processes generating the time series. Spectral estimators such as FFT and the various families of parametric spectral estimators (from Burg’s to wavelets), while commonly used, are (a) sensitive to noise, especially coloured noise, and (b) their uncertainty is very high: as Fisher (1929) has shown, the uncertainty of spectral estimators is of the same order as their magnitude. These issues are aggravated for noisy processes within time varying spectra, for the simple reason of the number of estimated values of the spectral characteristics being of a similar magnitude to that of the number of data points. So, while using the standard methods, we are getting a picture for a range of frequencies, this picture is highly uncertain for spectral estimators. With DHR and ASDHR, one periodicity (a handful of harmonic frequencies) is analysed in each step of the periodicity sweep, as we show below. The powerful handling of uncertainty by the Kalman Filter and Fixed Interval Smoother allows for a significant improvement of detection and estimation process, and in addition, importantly and nearly uniquely, provides an uncertainty estimate of the identified periodicities.

The periodicity identification method used for ASDHR involves scanning through a predetermined selection of periods and selecting the most statistically likely period(s). This is relatively time consuming for wider sweeps, but the search could be optimised using variations of common search algorithms. Computation time is arguably not such a critical issue, as (a) this is an off-line process, and (b) with high speed modern computers it is not significant.

The question of identification criterion for spectral peaks is quite critical. The standard R^2 , being the proportion of variance of observed data explained by the model (4.8),

$$R^2 = 1 - \frac{\|e_k\|^2}{\|y_k - \bar{y}\|^2} \quad (4.8),$$

with e_k being the model residuals, could be used as the statistical measure, but this was found to be less reliable, as many periods were found to have similar R^2 values leading to poor sensitivity, especially in trend-dominated series, and so the statistically best period was hard to distinguish from other significant periods. The same approach will apply to multi-periodic signals thanks to the orthogonality of the harmonic components.

Introduced here is an analogue of the standard R^2 , a new measure easily described as the proportion of data explained by the seasonal or cyclic component (4.9)

$$R_S^2 = \frac{\|S_k\|^2}{\|y_k - T_k\|^2} \quad (4.9),$$

where, S_k is the estimated seasonal component (section 4.2.0). Its quadratic norm (variance) is compared with that of the detrended data term in the denominator of (4.9). Environmental data often have a significant slow (or trend) component, which dominates the standard R^2 , while R_s^2 focuses on the detrended data and seasonal component. Note that in the process of identification the trend is estimated together with the seasonal component for each case, so there is no danger of introducing artefacts due to this procedure, effectively a spectral decomposition. In addition, (4.9) provides a standardised measure of a relative strength of periodicity, comparable across different time series.

4.2.5 Noise Variance Ratio (NVR)

The introduction of Taylor's expansion to the GRW models indirectly influences the selection of hyperparameters by introducing the sampling rate directly into the state transition equation. For the original DHR spectral model fit was used, as it was relatively easy to formulaically express the DHR spectrum and compare it to the AR spectrum of the data. This becomes more complex and ambiguous for irregularly sampled data. However, NVRs also carry the interpretation of time scaling of the model (spectral boundary of the low pass filter interpretation of the process, as explained in Young, 1999). In this case, with the challenging application examples we found that choosing the NVRs needs to reflect the time scale of the modelled process, rather than the best fit in any particular sense. The latter would have been arbitrary and would introduce additional assumptions into the modelling process.

4.3 Demonstrating ASDHR on environmental time-series

The examples compare, where possible, the proposed ASDHR methodology with the original DHR methodology. Direct comparison is often difficult as both methods operate on different data sets, DHR works with regularised and sometimes pre-filtered data whereas ASDHR works with the raw or 'unedited', observed data. In terms of the DBM approach or philosophy it should already be apparent that ASDHR is a more appropriate tool for analysing environmental systems.

The first example (4.3.1) has more comparable data sets (interpolated and raw data) and is a good demonstration of a working ASDHR methodology that is comparable to DHR. The second example (4.3.2) demonstrates the need for ASDHR to achieve data analysis that is in line with the DBM philosophy, i.e. the data tells the story and not the assumptions used to manipulate the data for DHR analysis. The third example (4.3.3) introduces a new type of forecasting, termed 'arbitrary forecasting', that allows predictions to occur at chosen points in the future, not just from the end of the observed time-series, and at different sampling rates and points to that of the observed time-series.

4.3.1 Paleo-climatology example

Analysing patterns such as trends and cycles in paleoclimatic data is a common theme in this discipline. The example we used was a typical one and previously published in Nature: Scientific Reports indicating the importance of the problem for the community. A carbonate oxygen isotope ($\delta^{18}\text{O}$) record derived from speleothems contained within Cueva de Asiul situated in the Matienzo depression (Cantabria), North Spain, was used to reconstruct the precipitation delivery to northern Spain during the last 12100 years using DHR analysis to find the trends and periods (for full details and the analysis see Smith et al., 2016).

Here the DHR analysis, in terms of trend and harmonic amplitudes, is compared with the proposed ASDHR. Similar results are expected as both methods are operating on the same data, but the result

specifics should be different due to the differences in the data pre-processing and analysis algorithms.

4.3.1.2 Data Pre-processing

With current DHR the data needs pre-processing to make uniform the sampling rate and this involves linear interpolation followed by removing the interpolated samples where there are gaps in the observed data to avoid introduction of artefacts. In this case (analysed in Smith et al., 2016 by the corresponding author) a highly irregularly sampled data-series of 1919 values is reduced to a uniformly sampled data-series of 815 (62 of which are missing values, mainly due to a single large gap in the data-series), which is a significant data loss, even if the key characteristics of the data are preserved (Figure 4.1).

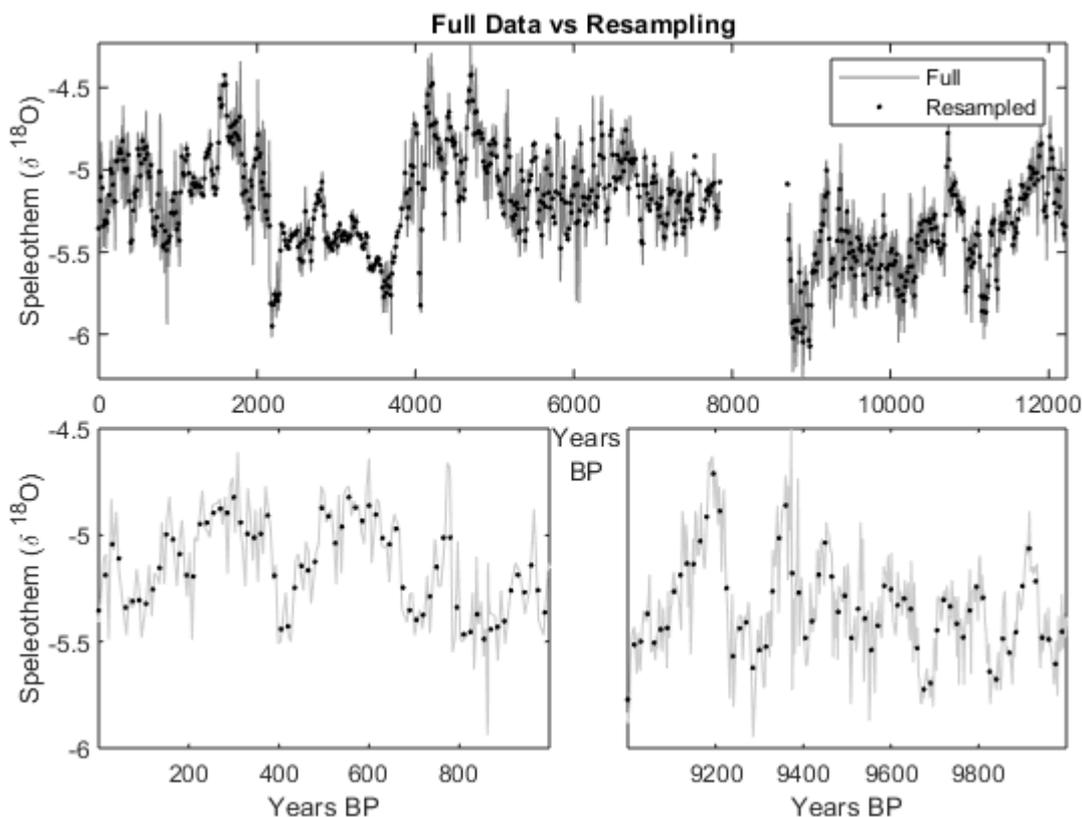


Figure 4.1. Comparing the full collected data and the resampled data. The two lower plots are zoomed-in to clearly demonstrate the differences between the full data range and the resampled range.

With current DHR, the large gap (872 years) in the data-series needs to be interpolated and due to its size, any interpolation across it is meaningless (in terms of physical interpretation) and could affect the immediate estimates either side of it.

With the proposed ASDHR procedure all that is required is to provide the temporal distance between each pair of samples. This means ASDHR has the full range of data (1919 values) to utilise and ignores the effect of the large gap (there is no interpolation).

4.3.1.3 Period Identification

The periodicities of the data were identified by scanning across a range of periods to find the two that fit the data best (as in Smith et al., 2016), although as noted above, the current DHR method

only uses 815 resampled values while ASDHR has the original 1919 values to use; current DHR method (1290 and 1490 years), ASDHR (1320 and 1540 years), well within the accuracy of the age estimate based on ^{18}O isotope levels. This new result confirms the findings of Smith et al. (2016), only providing a small adjustment when compared to the samples' timing error.

4.3.1.4 Comparing Current DHR with Proposed ASDHR

To compare the model fit of the two methods, the fit from DHR was rescaled back to the original time base and this showed that the arbitrary sampling procedure yielded slightly better results (Figure 4.2). The estimated trends and amplitudes were similar between the two methods (Figures 4.3 and 4.4 respectively) with the main difference with the behaviour over the gaps. The grey shaded area highlights the large gap in the data and a period of suspect data immediately before it.

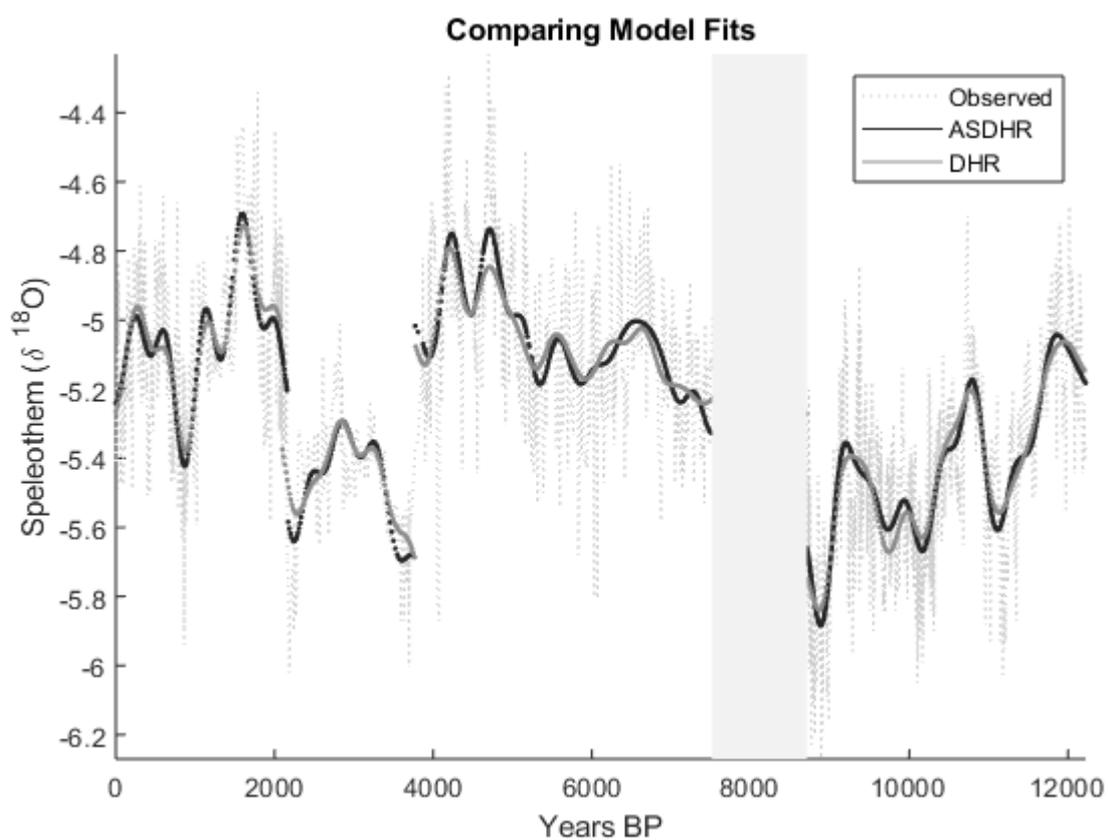


Figure 4.2. Model fit - comparing ASDHR (R^2 of 0.6741) with DHR (R^2 of 0.6452) on the original time base.

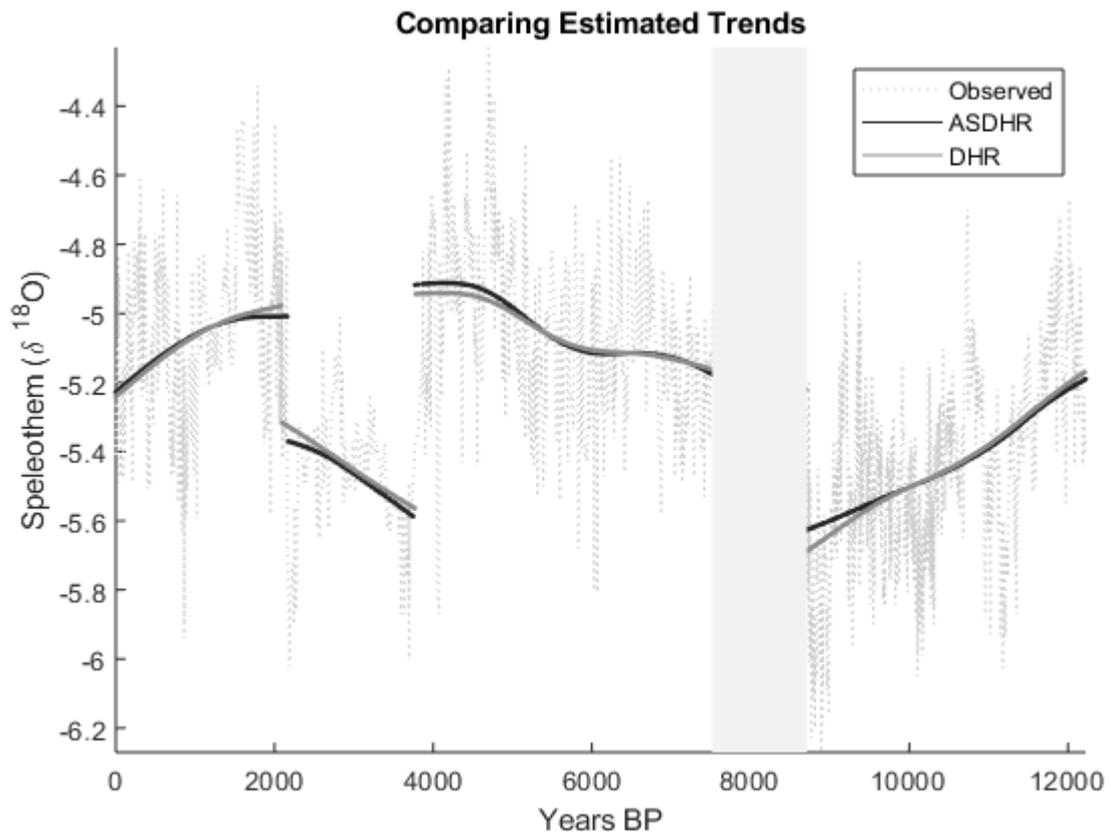


Figure 4.3. Trend - comparing ASDHR with DHR on the original time base.

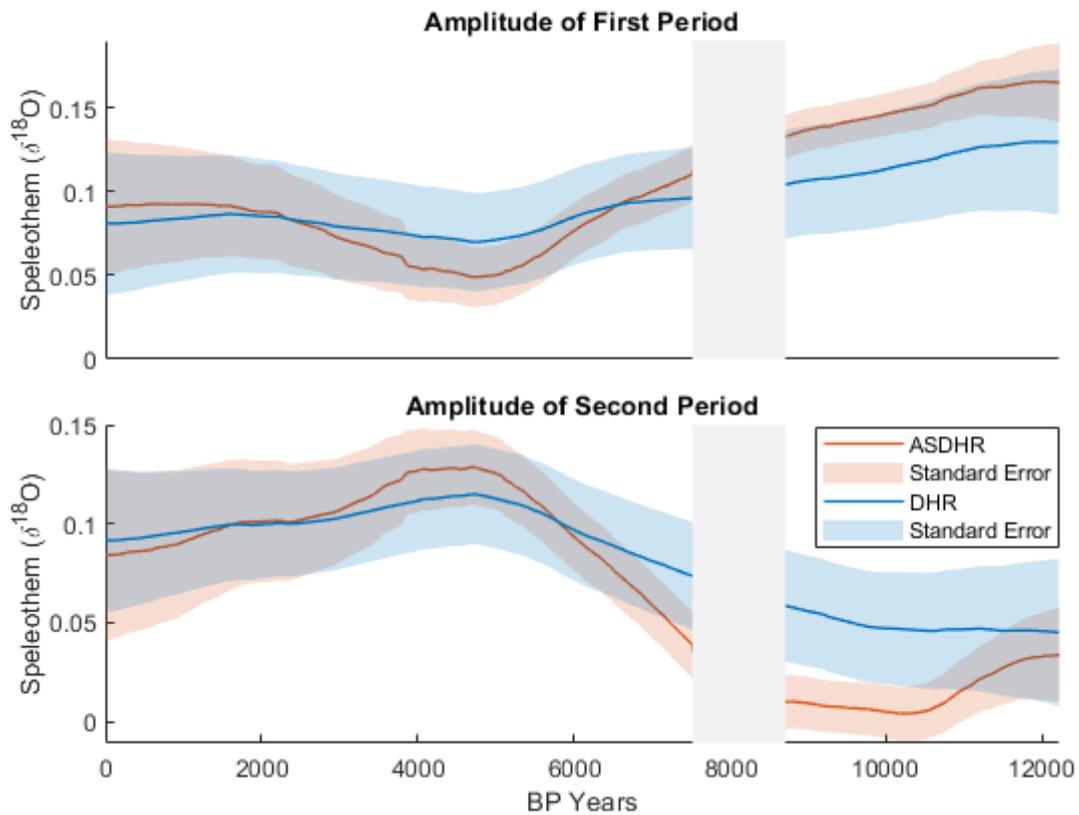


Figure 4.4. Amplitudes of seasonal components – comparing ASDHR with DHR on the original time base.

The distributions of residual errors were also similar between the two methods, with both being approximately Gaussian and very close to symmetric (Figure 4.5). This demonstrates how the introduction of the arbitrary sampling technique does not affect the fundamentals of the DHR method.

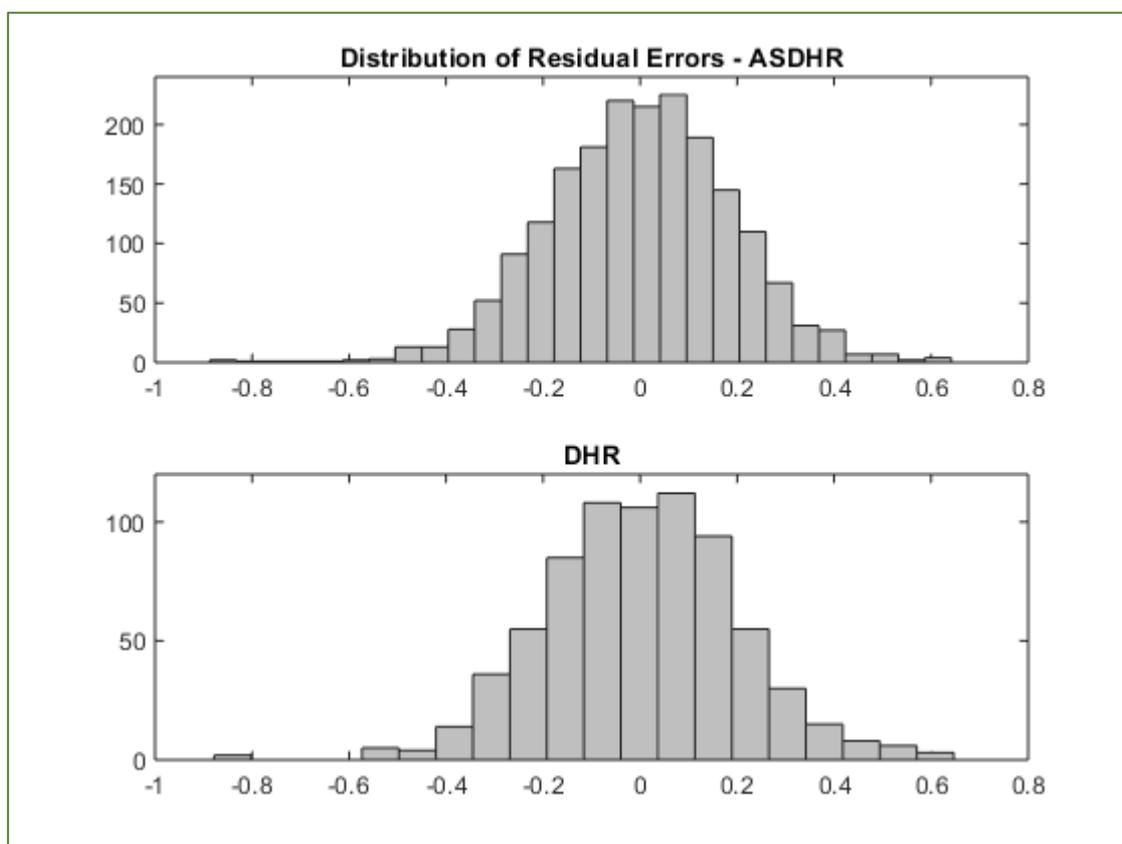


Figure 4.5. Residual error distributions of both methods.

Using ASDHR for paleo-climatic series not only simplifies the analysis procedure (no prior data manipulation) and preserves the data (i.e. no data loss/artefacts introduced), but also returns slightly better models (in this case a more pronounced seasonal component).

4.3.2 Persistent Organic Pollutant example

Identifying patterns and trends in atmospheric concentrations of Persistent Organic Pollutants (POPs) is an important part of monitoring and understanding how anthropogenic activities affect them. Weekly air samples have been collected since January 1992 at the High Arctic station of Alert in Canada and are filtered for various POPs and have a high signal to noise ratio (3:1). For further background and DHR analysis see Becker et al. (2006).

Two examples are taken from the data collected at Alert, Benzo(a)pyrene (reported in Becker et al., 2006) and α -Hexachlorocyclohexane (α -HCH, not reported) and both are members of the polycyclic aromatic hydrocarbons subset of POPs. Prior to DHR analysis the data were pre-processed; missing values due to data below the instrumental detection limits were replaced by values 2/3 of the detection limits, and data points situated outside 3x the standard deviation of any fitted trend were considered outliers and removed from the data set. The data were then resampled to fortnightly, due to the weekly data having significant irregularity, and finally ran through a low pass filter to reduce aliasing.

In the first example the observed data, with missing values and outliers, were used with ASDHR and compared to pre-processed data with DHR. In the second example both use the pre-processed data but without the resampling or pre-filtering for ASDHR. Both model fit and trend were compared between the two techniques as that was the aim of DHR in the original paper.

4.3.2.1 Benzo(a)pyrene

In Becker et al. (2006) an annual cycle was identified, and using the highly irregularly sampled raw data, this same annual cycle was identified using the ASDHR identification procedure.

The subsequent estimated fit (Figure 4.6) and trend (Figure 4.7) show that data pre-processing (unless required for a specific analysis question) is no longer necessary for DHR analysis if the arbitrary sampling technique is used, where a good model fit and trend estimate were obtained from the raw, unfiltered data.

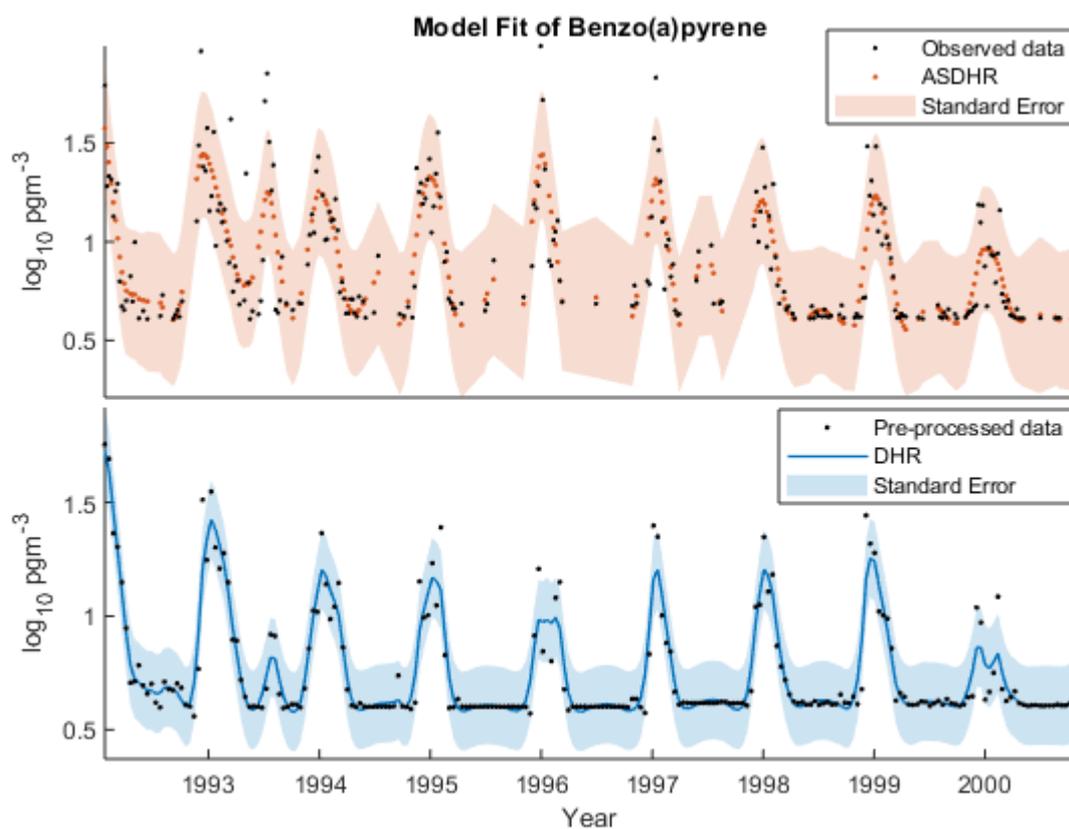


Figure 4.6. Comparing model fit and data used.

The uncertainty estimate is higher for ASDHR, which can be attributed mainly to the prefiltering used necessarily prior to DHR application, where data were brought onto fortnightly time base. This pre-filtering reduced the variance of the irregular component, which is clearly visible in Figure 4.6.

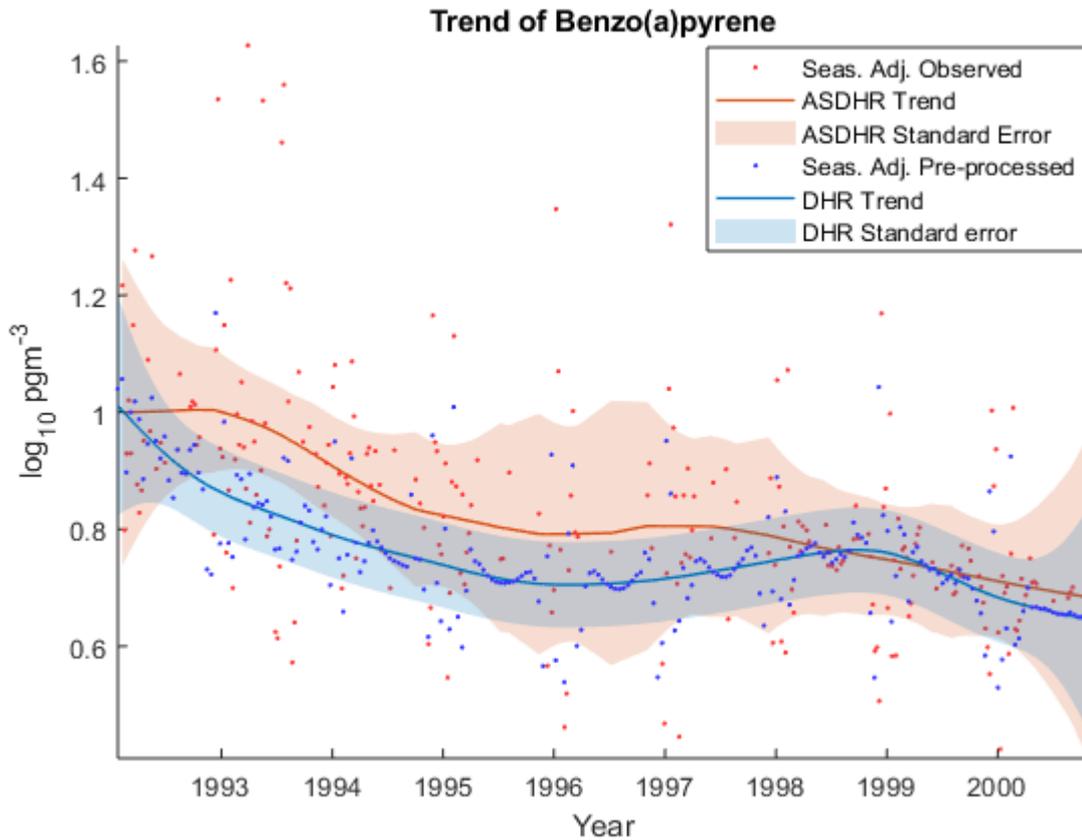


Figure 4.7. Comparing trend and seasonal adjusted data.

In addition, in ASDHR, as expected, uncertainty grows when data are absent, so the less frequent sampling between 1995 and 1999, as visible in Figure 4.7, leads to the increase of estimated uncertainty.

4.3.2.2 α -Hexachlorocyclohexane

Here both DHR and ASDHR used the pre-processed data, where values under detection limits were set to $2/3$ of the instrumental detection limits, but for DHR the data were then resampled and pre-filtered.

An annual cycle was identified again, and the subsequent estimation of trend shows how pre-filtering the data can lead to bias in its analysis. In this case (Figure 4.8), the trend estimated by ASDHR is 'pulled' down by the observed data, data that are missing in the resampled and pre-filtered time-series

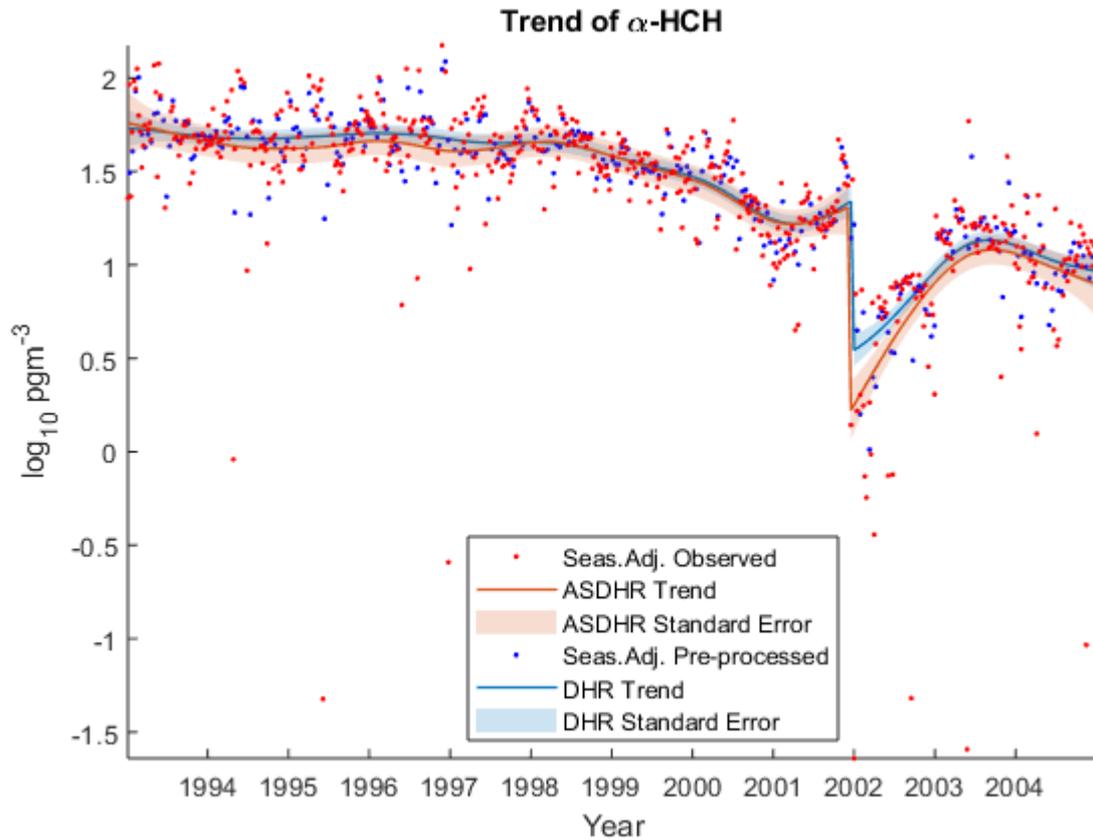


Figure 4.8. Showing the extent of the data loss from resampling and pre-filtering the data.

4.3.3 Arbitrary Forecasting – Atmospheric CO₂ example

Below we present a simple example of forecasting of the well-known Keeling Mauna Loa CO₂ series (see Acknowledgments). In this example we do not aim at perfect forecasts, but rather at showing the flexibility and versatility of even the basic version ASDHR in this application. Better forecasts could be achieved with more assumptions being included in the model, such as a model of the business cycle or the industrial growth projections.

The Mauna Loa CO₂ monthly mean data set was used to demonstrate arbitrary forecasting, that is forecasting from any point in the future and not just from the end of the data-set. Additionally, the time-base of the forecasting period is not limited to that of the observed data, something that the original DHR cannot do.

This is easily implemented by extending vector Δ_k by the values corresponding to the arbitrary points chosen. In this example (Figure 4.9) the monthly Mauna Loa data are used to forecast the year 2018 on a weekly basis, for five different periods of data: 1960 to 1970, to 1980, to 1990, to 2000, and to 2010 as shown with colour code in Figure 8. In this example the forecast shown as purple is based solely on the data up to 1970.

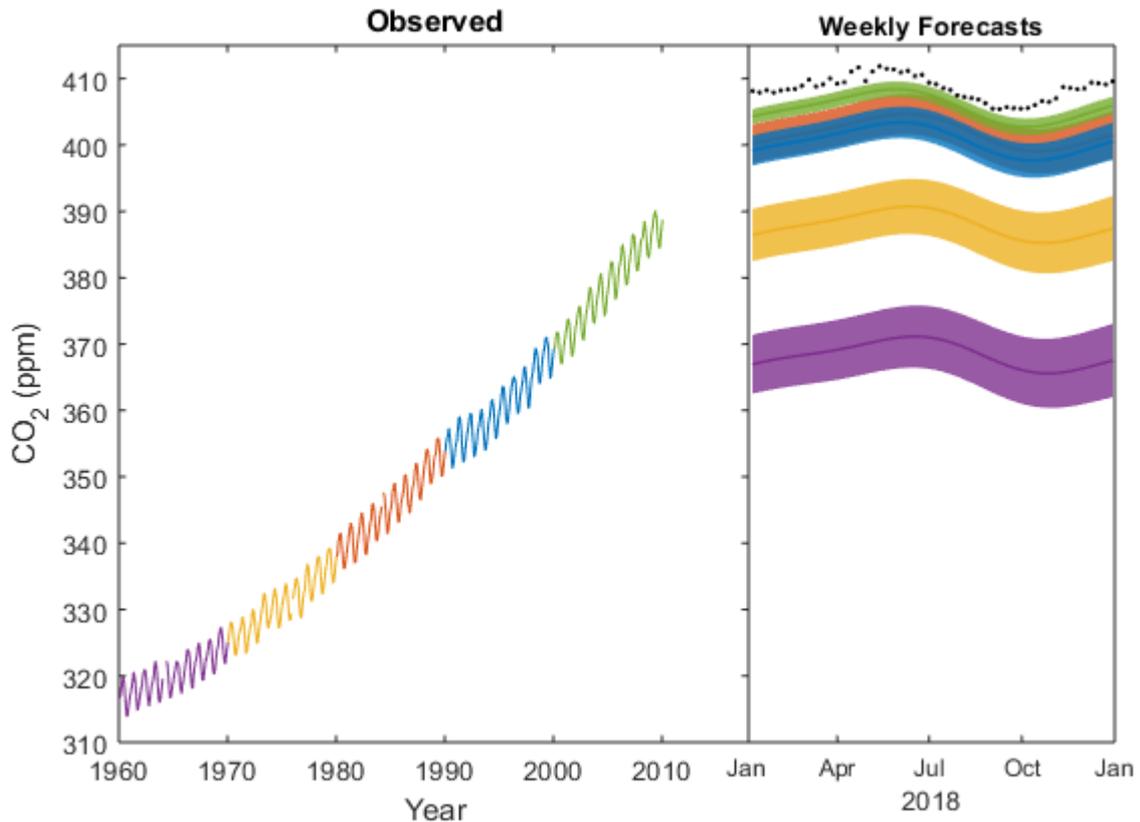


Figure 4.9. Arbitrary forecasting of Mauna Loa CO₂ data. Observed data is monthly averages and forecasts are weekly averages. Purple, 1960:1970. Yellow, 1960:1980. Orange, 1960:1990. Blue, 1960:2000. Green, 1960:2010. Shading, uncertainty of each forecast. Black dots, observed weekly average data for 2018.

This (Figure 4.9) demonstrates how the rate of change of CO₂ grows with time – as the knowledge is increasing (e.g. the 1960:1970 based predictions for 2018 have levels observed in 2000), except during the 1990s where the 1960:2000 based predictions are lower than that of the 1960:1990, clearly relating to the visible dip during the 1990s.

The size of the uncertainty, shown here as a single standard error estimate band, increases as the forecasting horizon increases; purple for 48 years, green for 8 years. Testing on this data set and on simulated data sets indicates that the forecasts are largely unaffected by data sparseness or the forecasting horizon magnitude. Only the observation noise level impacts the predictions.

In the past the DHR technique was successfully used for forecasting various processes, in both broadly environmental and industrial applications: from demand for electricity (Young and Pedregal, 1998) to numbers of calls to a banking call centre (Tych et al, 2002) to, anecdotally, demand for beer (unpublished Lancaster University thesis). ASDHR naturally broadens the application area of this robust method.

4.4 Robustness and reliability evaluation

Before the methodology was applied to real data, it was tested on a variety of challenging simulated data-sets (with known “unobserved” components) to test the robustness and sensitivity to data sparseness, size of temporal distancing, and observational noise.

The methodology was found to be very robust with very sparse data-sets. It also produces meaningful estimates for a very wide range of temporal distances between the samples, limited by the Taylor expansion (4.3) and condition of the recursive estimate of the covariance matrix within the Fixed Interval Smoother algorithm applied. Subject to these constraints, the technique can be used to estimate the observed values and components no matter how irregularly sampled the time-series is.

Sensitivity to observation noise is similar between the time-tested and commonly used DHR (with about 300 citations) and ASDHR.

In terms of physical interpretation, the extent of data sparseness needs to be considered when interpreting the estimated components as very sparse data may effectively under-sample higher frequency components.

4.5 Conclusions

Presented here is a technique to improve the DHR analysis of irregularly sampled time-series that removes the need for data pre-processing: regularisation, decimation, interpolation etc. The technique also does not involve estimation of missing values and so any subsequent analysis is free from assumptions and bias and only uses the available observed data. This brings DHR closer to the DBM philosophy of allowing the data to inform us of the processes and mechanisms that result in the observed time-series. It also makes it uniquely suitable for analysis of environmental irregularly sampled observational data.

Data pre-processing was a necessary step to allow DHR to work on irregularly sampled data-sets but it comes with artefacts, bias and increased uncertainties in the model estimates. However, with the arbitrary sampling technique, this step can be avoided and provides model estimates with lower uncertainties and no bias.

The technique has been tested on challenging simulated data and is robust enough to work on extremely sparse data, however, in terms of physical interpretation there is a limit to how sparse the data can be due to e.g., under-sampling (Chappell et al., 2017). Additionally, the technique was found to have similar observation noise sensitivity to that found in standard DHR method.

The technique has been demonstrated here on three different types of observed environmental time-series data and has yielded slightly better model outputs than the standard DHR method. Without data pre-processing, there will be no introduction of any assumptions, artefacts or bias into the data prior to analysis and thus these results should be closer to observed reality.

The technique also allows for forecasting at arbitrary points and at different sampling rates than in the observed data. This means the frequency of the forecast is not limited to the frequency of the observations, and with a non-stationary forecast horizon may allow forecasting to yield more insights into environmental processes.

Finally, while it may not be apparent from the equations, ASDHR is easily and inherently generalised so that all aspects of estimation are either time-varying or state-dependent: from periodicity to dynamics of random walk model, to NVRs. For each sample k , the periodicity, random walk model and NVR can be set. So, for example, one section of observed data could be analysed for one set of periodicities and another section analysed for another set, or the random walk model could be changed to match a significant change in the data, or the NVRs can be varied to suit the smoothness of the data.

Chapter 5: Applications of Arbitrary Sampling Dynamic Harmonic Regression – OBJECTIVE 4

ASDHR was applied to two different studies (**paper 4 and 5**), one looking for dominant cycles in stream diatom assemblages (**Snell et al., 2019**) and the other looking at the trends of polycyclic aromatic hydrocarbons found in the Arctic (**Yu et al., 2019**). In both studies the ASDHR algorithms designed above were used to drive the trend and cycle analysis provided by the present author. In both cases the analysis is central to each paper's objectives. In terms of contribution, the present author undertook the ASDHR modelling for both works and supplied model outputs, cycle comparisons for the former and trend analysis for the latter. Figure 2 of both publications are direct outputs of the present authors work.

These papers are not included in the body of the Thesis as they are not wholly the present Author's text. While the present author made a substantial contribution to both, they are included as appendices to show that the developed methods and their software implementation work and can be directly used across various disciplines.

5.1 Snell et al., 2018

The seasonality of the diatom Ecological Quality Ratio, a highly irregularly sampled data set, was identified and compared to the seasonality of other climate factors (rainfall, streamflow, total reactive phosphorus, net solar radiation) using ASDHR, and revealed a strong link with net solar radiation (Figure 5.1).

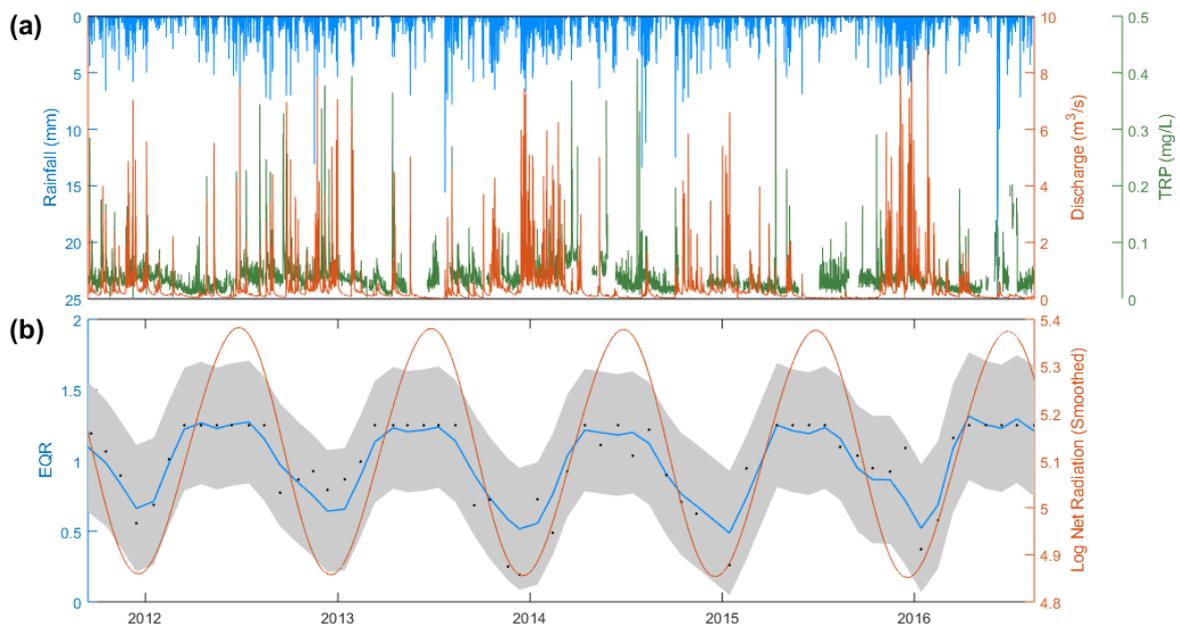


Figure 5.1. Flow and water chemistry data for the period from 14 September 2011 to 16 August 2016 for Newby Beck: (a) Hourly rainfall, discharge and Total Reactive Phosphorus (TRP); (b) Monthly diatom EQR and Log Net Radiation. ASDHR model fit with 95% confidence interval explaining over 80% of data variance, and showing the strength of seasonal character of the process. Log Net Radiation Smoothed is the natural log of Net Radiation prior to using DHR-demonstrating Net Radiation seasonality + smoothing

5.2 Yu et al., 2019

The underlying trend of 3 Polycyclic Aromatic Hydrocarbons (PAH), irregularly sampled, from 3 sites was identified using ASDHR (Figure 5.2) and formed part of the analysis that concluded that PAHs were not declining in the Arctic air.

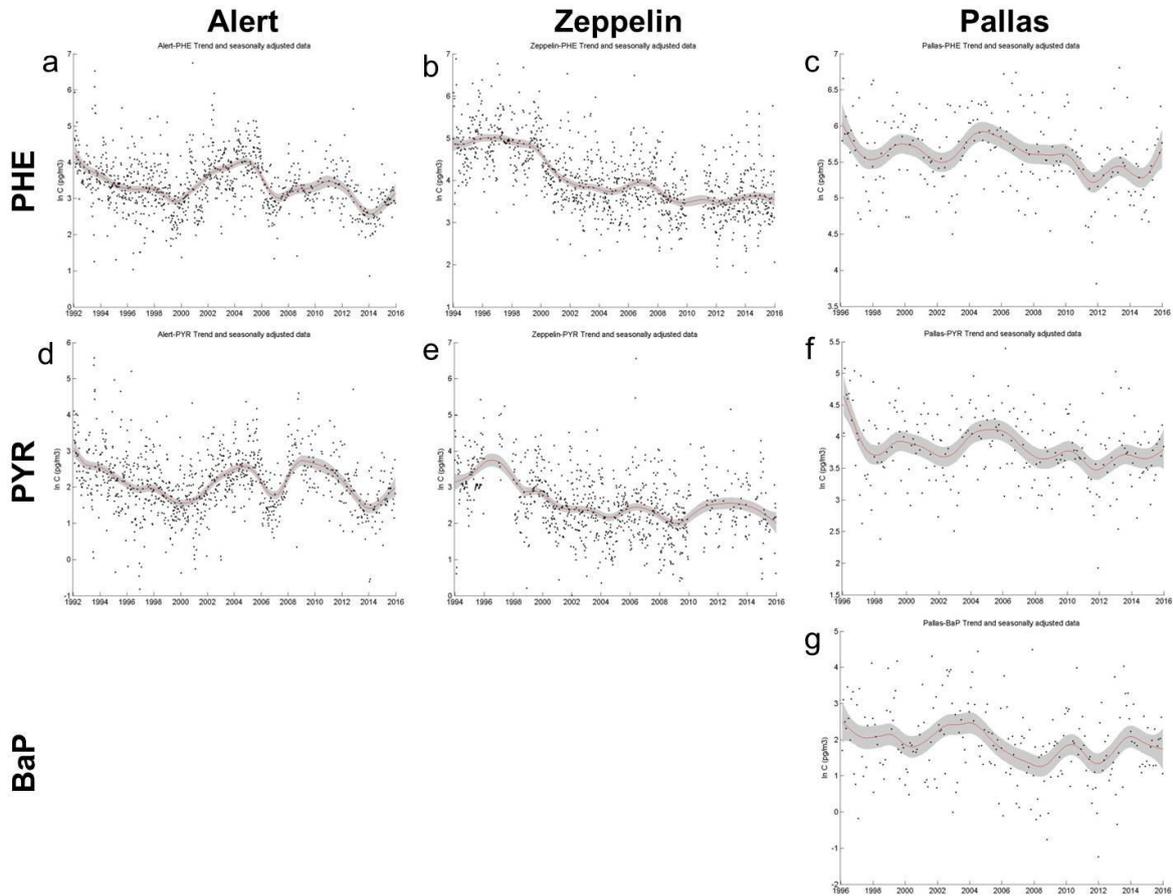


Figure 5.2. Long-term trends of (a, b, c) PHE and (d, e, f) PYR at 3 sites and (g) BaP at Pallas.

Chapter 6: Applications of the Data-Based Mechanistic approach to Catchment Hydrology – OBJECTIVE 5

The DBM philosophy is perfect for exploratory data analysis of environmental systems as the emphasis is on the observed data, and any observations and conclusions obtained will be based solely on the available data. In **Paper 6**, DBM methodology was applied to the study of ranch ponds in the dry Chaco rangelands of Argentina, to understand the mechanisms behind the recharge and draining of ponds in this region. Standard hydrological understanding does not apply to these regions due to their dry nature having very limited groundwater storage, so no subsurface contributions to surface waters, and thus any drinking water for animals comes from rainwater collected in ponds. Various ESDP and linear TF models were evaluated and from the knowledge gained in developing ESDP it became apparent that the limits of the available data meant that simpler TF models were more appropriate for explaining these systems (**Magliano et al., 2019**). In terms of contribution, the present author undertook the lead on the modelling aspects of the research, and produced the figures. Along with the other application papers, this work is included as an appendix, not in the main body of the Thesis.

The second application of the DBM approach is the temporal evolution of rainfall-runoff response with respect to borehole, soil moisture and extreme rainfall data. ASDHR was applied to investigate annual cycles, some of the data sets had gaps in the data, which was easily handled by the new tool. The analysis also involves investigating the concept of ‘catchment wetness’, a proxy for describing antecedent conditions and ESDP was used to compare several variables as potentials for this proxy. This is presented in the **final paper**, which is in draft form, where the present author led both on the modelling and methodological sides of the work.

6.1 Introduction

Catchment hydrology is no simple matter, catchments are highly heterogeneous systems with varying land use, soil cover (depth/type/preferential pathways), subsurface composition (depth/rock type/fractures/aquifer).

As catchment size increases the more heterogeneous it is. The spatial heterogeneity of catchments leads to the temporal heterogeneity of streamflow response to rainfall, as all the aspects of the catchment that facilitate the transfer of rainfall to runoff are not only diverse but constantly changing with time. Catchments, in the context of runoff generation, are nonlinear dynamic systems.

This leads to significant issues when modelling the streamflow response to rainfall. While the underlying system is simple: it rains and ‘sometime later’ the flow responds. The mechanisms and processes involved in determining when that ‘sometime later’ is, are complex. The general theory is that how a stream responds to an input of rainfall depends on the antecedent conditions of the catchment; the ‘wetter’ the catchment is, the greater and quicker the stream responds to a new input of rainfall.

This leads further to the idea of ‘catchment wetness’, a measure of the conditions prior to the current rainfall event. ‘Catchment wetness’ cannot be directly measured, but is assumed that it can be inferred by using some other catchment variable as a proxy.

The identification of such a proxy may seem simple, given the few measurable possibilities, but they may not be good indicators of 'catchment wetness'. Soil moisture and groundwater are two such possibilities.

Soil moisture, measured as the volumetric water content (%) of soil, controls the portion of rainfall that can infiltrate the soil, with the remaining portion forced into overland flow. The higher the volumetric water content, the more rainfall that should become runoff. While selection of this variable seems quite logical, the heterogeneity of soil across the catchment means many soil moisture measurement sites would be required to get a detailed picture of soil moisture as the 'catchment wetness' variable. Then there would be the issue of incorporating all these soil moisture variables into a rainfall-runoff model.

Groundwater, measured as a level (m), represents the water table level and controls the amount of water that can infiltrate into the subsurface: the higher the level, the less amount of water that can access the subsurface and thus more water has to travel as runoff. This issue with using groundwater level is the same as when using soil moisture; the heterogeneity of the subsurface composition and thus the associated issues occur as described above.

The third possibility for a 'catchment wetness' proxy, is rainfall, measured as a level (mm). Rainfall is the driver for runoff response and as rainfall duration and intensity increase, the runoff response is quicker and larger. Filtering rainfall gives a measure of rainfall intensity, which has potential as a 'catchment wetness' proxy.

6.1.1 Study Site

The River Eden is approximately 130km long and rises South of Kirkby Stephen on the Cumbria/Yorkshire border and generally flows North West until it passes through Carlisle and West into the Solway Firth. It is fed by tributaries that come from the Pennines in the East and the Lake District in the West. The catchment (2400km²) is largely rural with three urban areas along the Eden; Kirkby Stephen, Appleby-in-Westmorland and Carlisle. Land use is dominated by agriculture, with a transition from more pasture to more crops as the river flows downstream.

The catchment is well gauged (Figure 6.1) with many streamflow (m³s⁻¹), rain (mm) and groundwater (m) gauges scattered across the area, additionally, at Moor House, on the Pennines side of the catchment is a soil moisture (%) site. While there are two decades of streamflow and rainfall data (15-minute intervals), there are only around 4 years of groundwater (hourly intervals) and soil moisture (30-minute intervals). The bulk of the analysis is done using hourly intervals since that is the lowest sampling frequency and is coincidentally largely appropriate for most catchment dynamics.

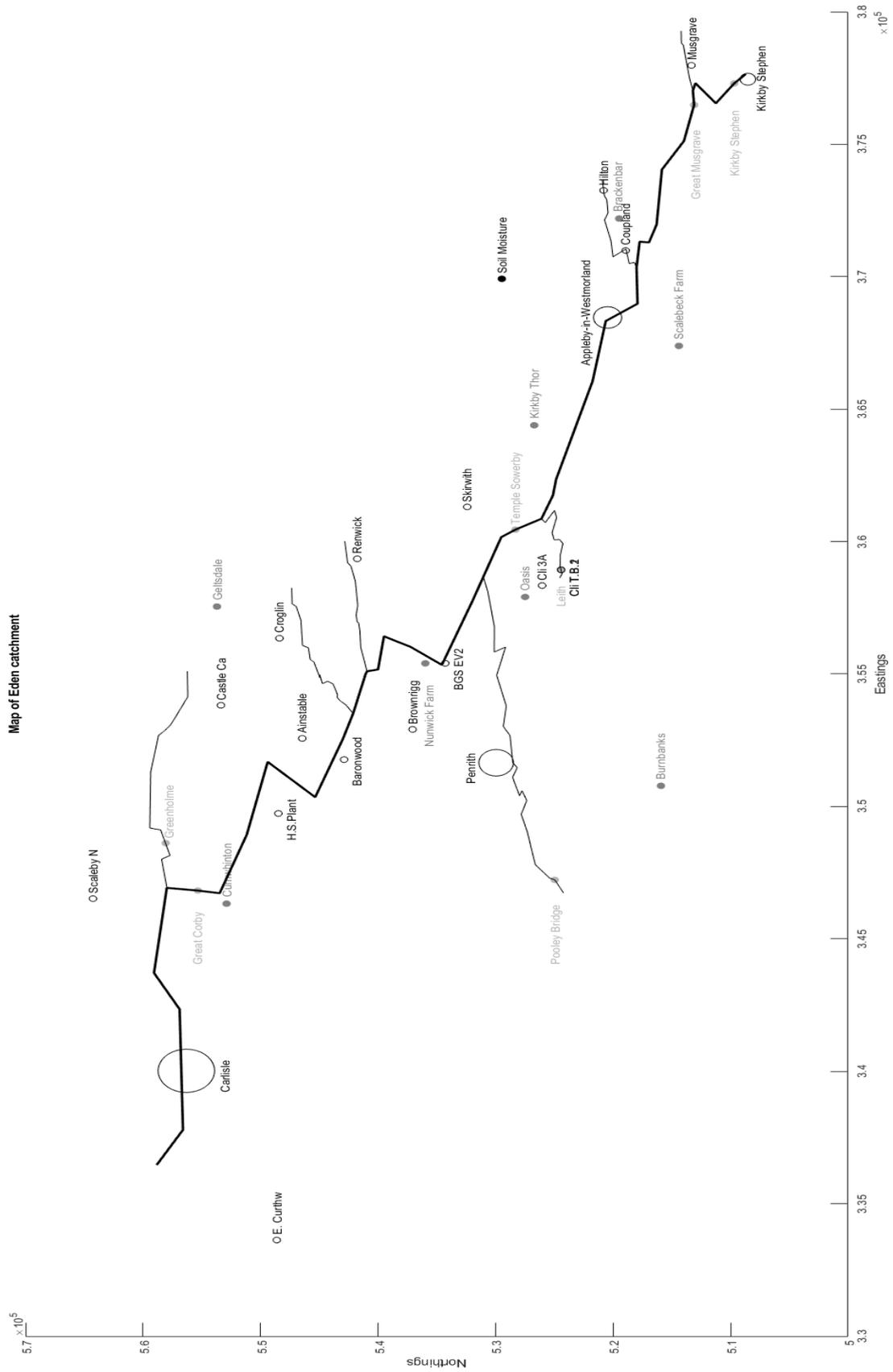


Figure 6.1. Map of the Eden Catchment showing the position of gauging stations: light grey – streamflow, dark grey – rainfall, black circles – groundwater, black dot - soil moisture.

6.1.2 Aim and Objectives

The aim of this paper is to explore the available data for a well gauged catchment that includes groundwater and soil moisture, and then delve into rainfall-runoff modelling. The paper can be broken down into two objectives:

1. Visualisation and comparison of the various gauged observations in a UK catchment, and explain, where possible, their relationships.
2. Using the DBM approach to objectively investigate the rainfall-runoff nonlinearity, captured by nonlinear models.
3. Compare the variables that appear to have potential as a 'catchment wetness' proxy using a class of models known as state dependent parameter models and determine which variable, if any, is best suited for the 'catchment wetness' proxy.

6.2 Exploratory Data Analysis

Here data visualisations and variable comparisons, on various timescales, are used to look for any relationships between the variables. Arbitrary Sampled Dynamic Harmonic Regression (ASDHR, Mindham and Tych, 2019) is used to estimate the annual cycles, and linear Transfer Function models (TF, Taylor et al., 2007) are used to check input/output relationships between variables that visually appear to be closely related.

6.2.1 Rainfall Driver

The Eden catchment discharge dynamics is dominated by frontal rainfall that runs broadly West to East with an annual cycle that peaks around November-December (Figure 6.2, top) with each gauge peaking at slightly different times due to the nature of frontal rainfall systems. Typically, when looking at individual events, the rain gauges on both sides of the catchment receive rainfall at the same time with an occasional one-hour difference, but not necessary between West to East, but also East to West. Additionally, the upland part of the catchment (South from Temple Sowerby) generally receives more rainfall than the lowlands, and logically the West side of the catchment tends to receive more rainfall than the East (Table 6.1).

Table 6.1. Rainfall totals for six rain gauges for a four-year period from 24/02/2014 to 10/01/2018. For actual gauging locations refer to Figure 6.1.

	Rain Gauge					
	Burnbanks	Brackenbar	Oasis	Nunwick Farm	Cumwhinton	Geltsdale
Relative Location	West Uplands	East Uplands	Central Uplands	Central Lowlands	West Lowlands	East Lowlands
Rainfall Total (mm)	9458.0	3880.4	4534.8	4001.2	4131.4	4868.0
Maximum Rainfall over one hour (mm)	47.2	26.4	13.2	18.6	52.0	17.8

6.2.2 Streamflow Response

Overall, the catchment responds uniformly to an impulse of rainfall, the peaks and low flows all occur around the same time (Figure 6.2, bottom), with the lag times (between rainfall event and stream response) and amplitudes (size of response) being dependent on position within the catchment. I.e. up in the headwaters has a lower lag time than down in the lowlands, due to the water having a shorter distance to travel, and amplitude increases as drainage area increases (Table 6.2). While the lag time between a rainfall event and the flow response is several hours, the overall lag time between peak rainfall and peak flow is one to two months (Figure 6.2), with the individual rain gauges experiencing peaks at slightly different times; due to frontal rainfall.

Table 6.2. Overview of how the River Eden changes as it flows down the catchment. For actual gauging locations refer to Figure 6.1.

Flow Gauge	Relative Location	Time-delay (R>Q) (Hours)	Min Flow (m ³ s ⁻¹)	Max Flow (m ³ s ⁻¹)
Kirkby Stephen	Uplands	2	0.14	140
Great Musgrave	Uplands	2	0.63	372
Temple Sowerby	Central	4	1.93	1140
Great Corby	Lowlands	5	7.53	1490

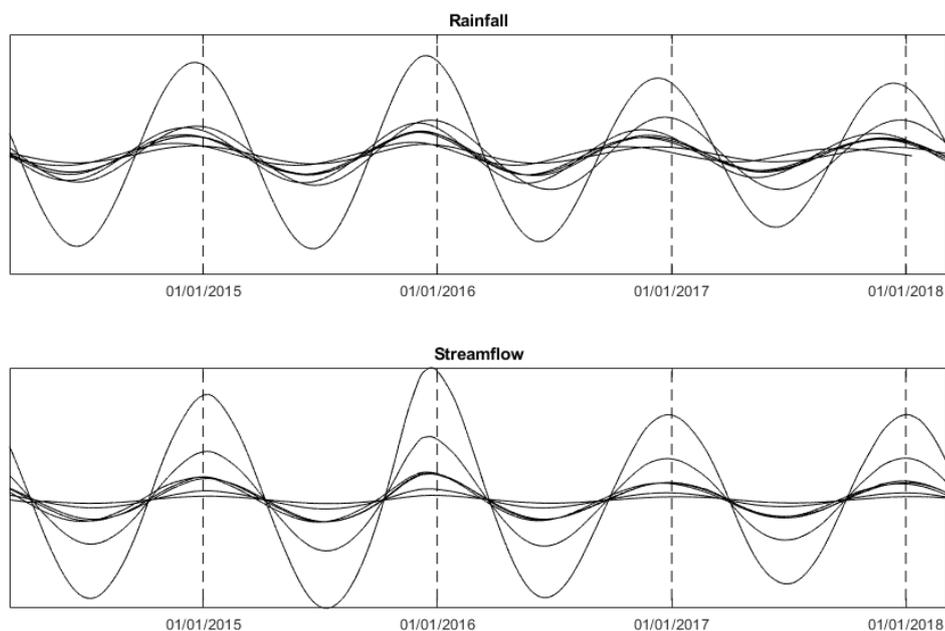


Figure 6.2. Showing the cyclic nature of the rainfall and streamflow at annual timescale in the Eden, extracted using ASDHR as the seasonal component. Top, annual cycles of five rainfalls. Bottom, annual cycles of seven streamflows. Vertical dashed lines at the start of each year showing the approximate times the flows peak.

6.2.3 Groundwater Response

With the catchment responding uniformly to rainfall, it would be logical to expect the groundwater response to be uniform as well. Unfortunately, this is not the case, all 17 boreholes respond differently (Figure 6.3), but with some similar underlying characteristics. These boreholes could be categorised into 3 types:

1. 'Flow like' – have a flow-like response.
2. 'Drain/Refill' – slow response to rainfall; period of draining followed by a period of refilling, annual cycle.
3. 'Random' – no discernible pattern, noisy response to rainfall.

These overall trend patterns are also characterized by significant uncertainty levels; as well as having one of the three characteristics described above there are daily/sub-daily variations, the frequency and amplitude being different for each borehole.

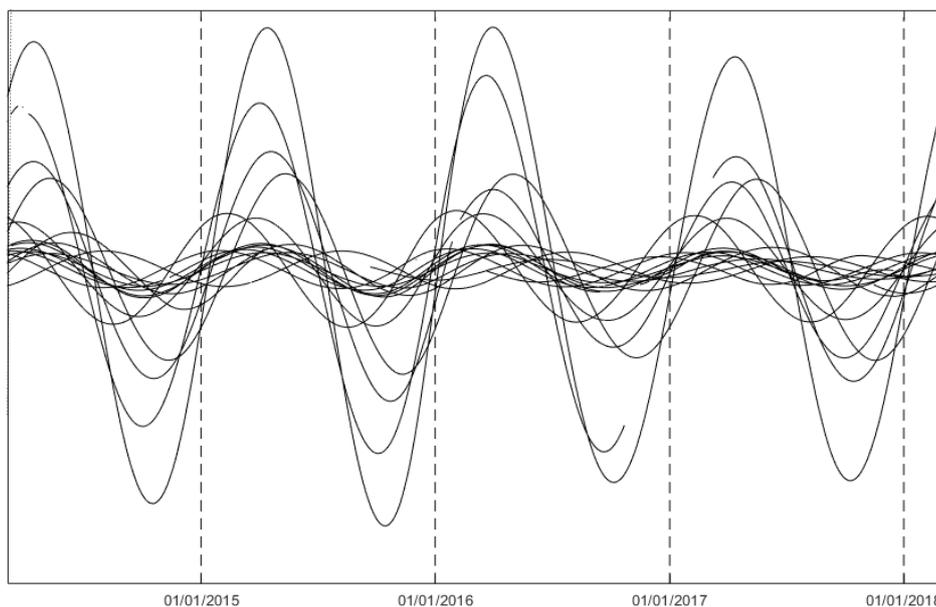


Figure 6.3. Annual cycles of all boreholes extracted using ASDHR. Vertical dashed lines at the start of each year showing the approximate times the flow peaks.

6.2.3.1 'Flow Like'

Three boreholes display 'flow like' behaviour and clearly demonstrate the connectivity of those boreholes to the stream. In the case of two of them, Cliburn Tower Bridge 1 (deep, 47.68m deep) and 2 (shallow, 10.87m deep), they are situated a few meters from the River Leith, a tributary of the River Eden, and are 31m and 25m, respectively, from the flow gauge. Both mimic the Leith flow pattern with the deep borehole showing a smoother response (Figure 6.4).

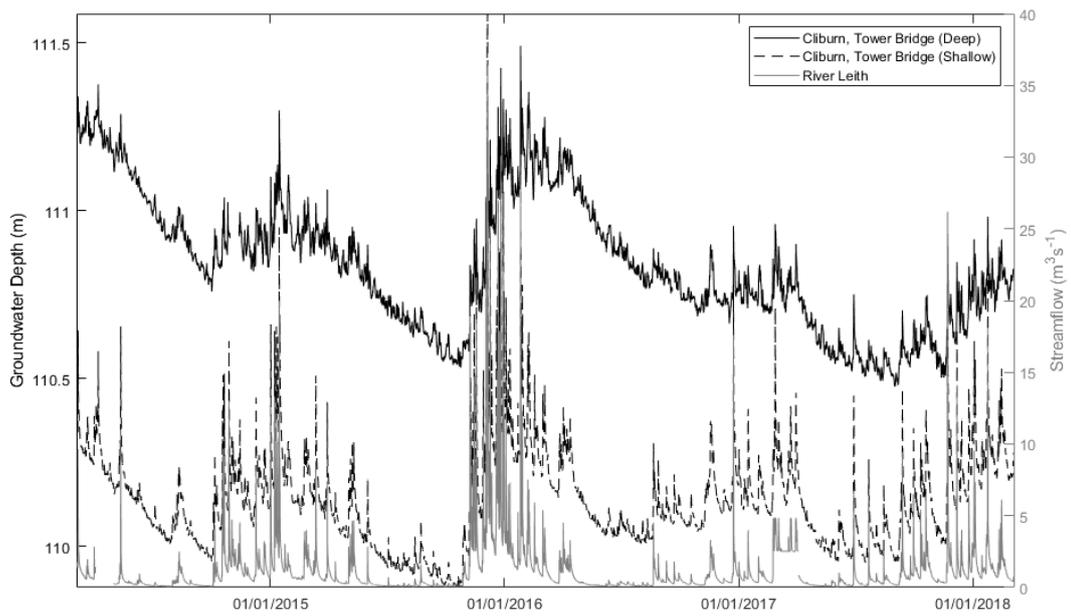


Figure 6.4. Comparing Cliburn Tower Bridge boreholes to the River Leith over the available data period.

The Great Musgrave (GM) borehole is situated 1.51km east of the nearest flow gauge on the Eden (but 408m from the closest stream) and behaves similarly to the flow but with a dampened response (Figure 6.5).

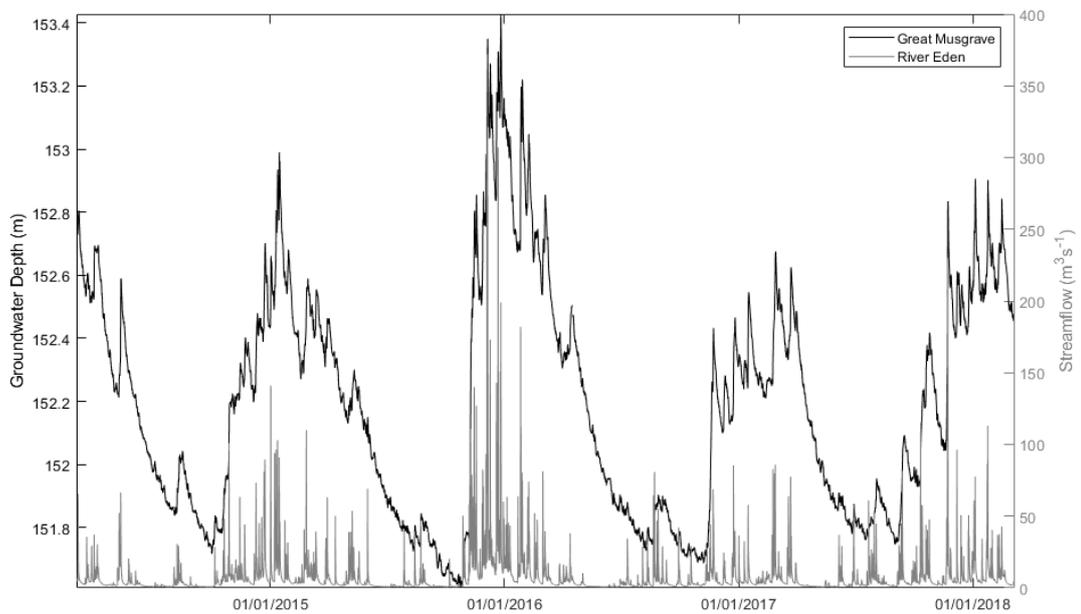


Figure 6.5. Comparing the Great Musgrave borehole with the Great Musgrave flow (Eden) over the available data period.

A closer visualisation of these boreholes with flow show how well connected the Cliburn boreholes are to the River Leith (Figure 6.6, top). The shape of the response of the shallow borehole is almost identical to that of the flow, while the deep borehole shows a smoothed similarity. This suggests there is very strong connectivity between the River Leith and the local groundwater. However, comparing the TF models of rainfall to flow, rainfall to shallow borehole and shallow borehole to flow, do not reflect this. The rainfall to flow model was identified to have a 3-hour pure time-delay, whereas the models of rainfall to borehole and borehole to flow, failed to identify a physically interpretable model.

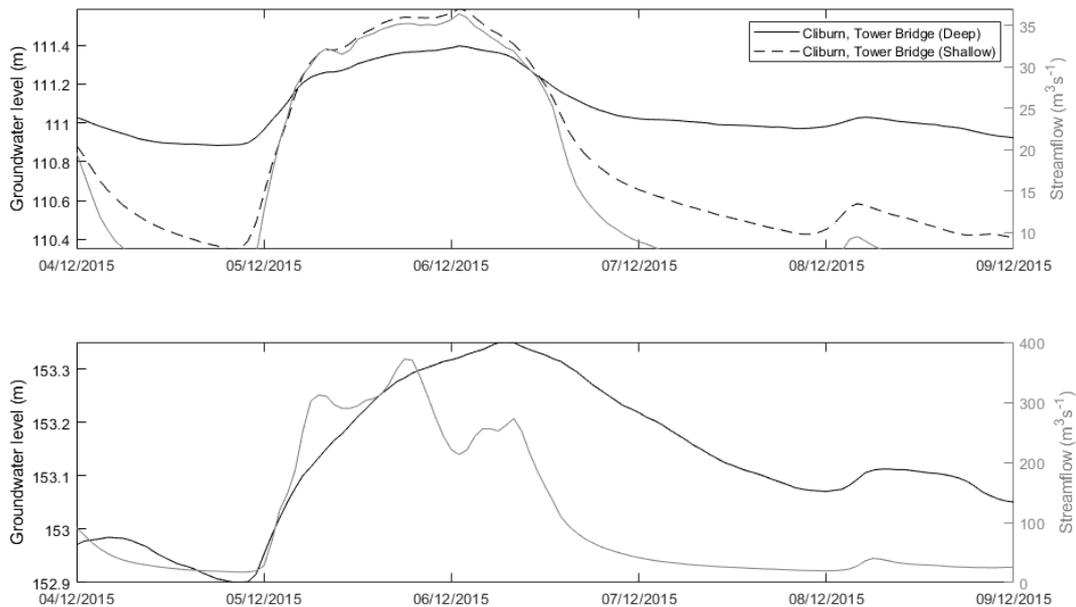


Figure 6.6. Comparing borehole response to the flow response during Storm Desmond, 2015. Top, Cliburn boreholes and the River Leith. Bottom, Great Musgrave borehole and the River Eden.

The visualisation of the GM borehole (Figure 6.6, bottom) demonstrates that its flow like behaviour is likely not due to connectivity with a stream; the response is nothing like that of the closest flow gauge. Unlike the rest of the Eden, the Great Musgrave borehole is situated in an area of Brockram; rock composites of carbonate breccias and conglomerates containing carbonate clasts and calcite cements. It is likely these carbonates have been locally dissolved creating large fractures allowing for easy drainage and rapid recharge (BGS, 2014), giving the flow like behaviour. While the GM borehole starts to respond at a similar time as the streamflow, the borehole takes longer to reach a peak with this delay time (between flow peak and borehole peak) varying slightly across the time-series.

6.2.3.2 'Drain/Refill'

Eleven of the boreholes exhibit a much slower response with long periods of draining followed by a shorter refill period before draining again (Figure 7 and Figure 8). While these boreholes have this pattern, none of them are similar. They all peak at different times of the year with different recession periods, and three appear to have a two-year cycle (Figure 9).

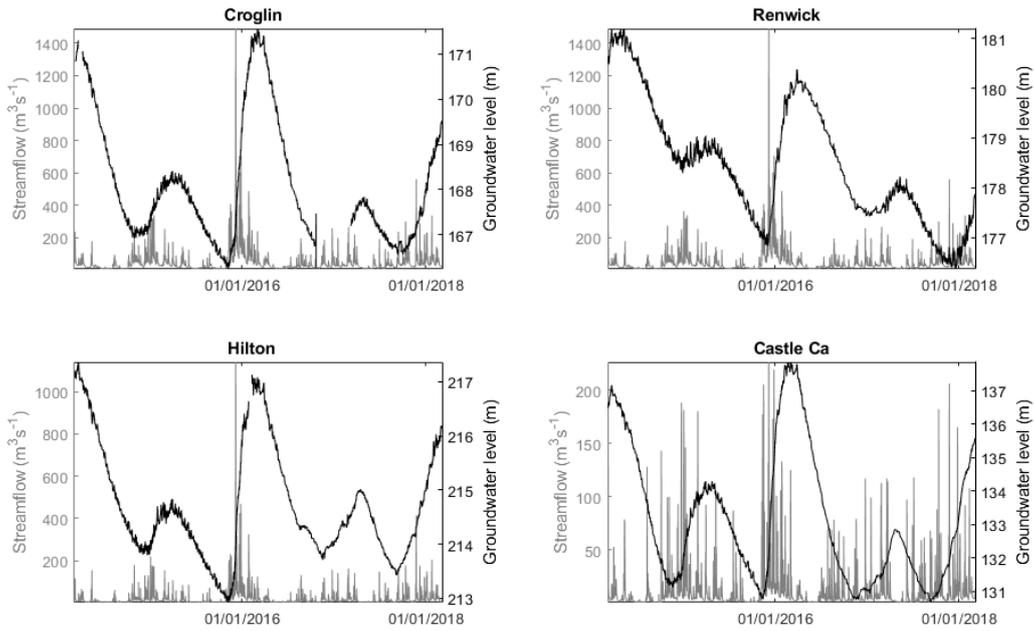


Figure 6.7. Comparing boreholes with their closest flow gauge, over the available data period.

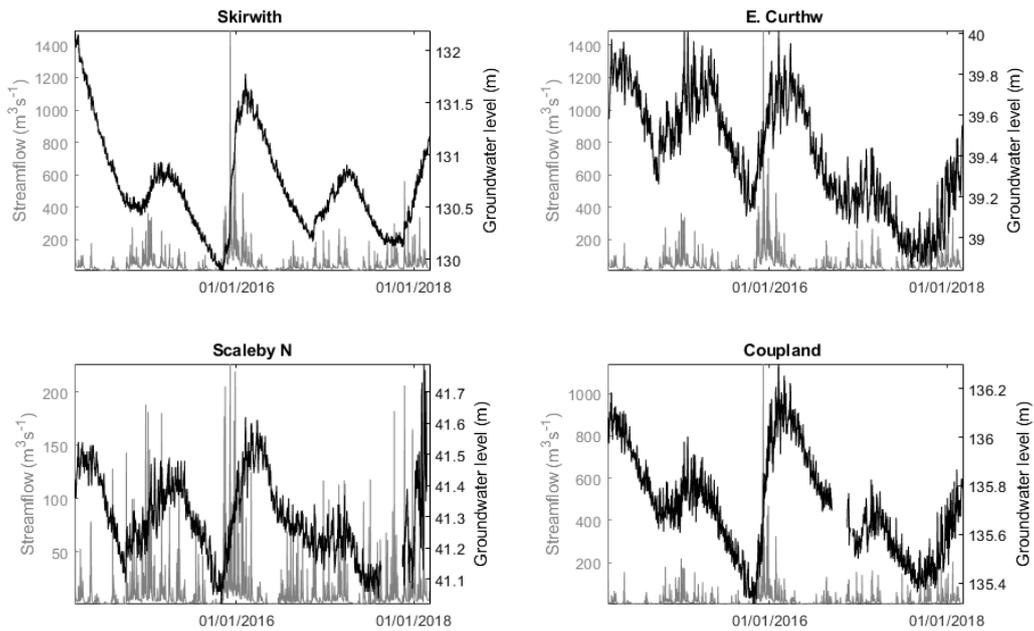


Figure 6.8. Comparing boreholes with their closest flow gauge, over the available data period.

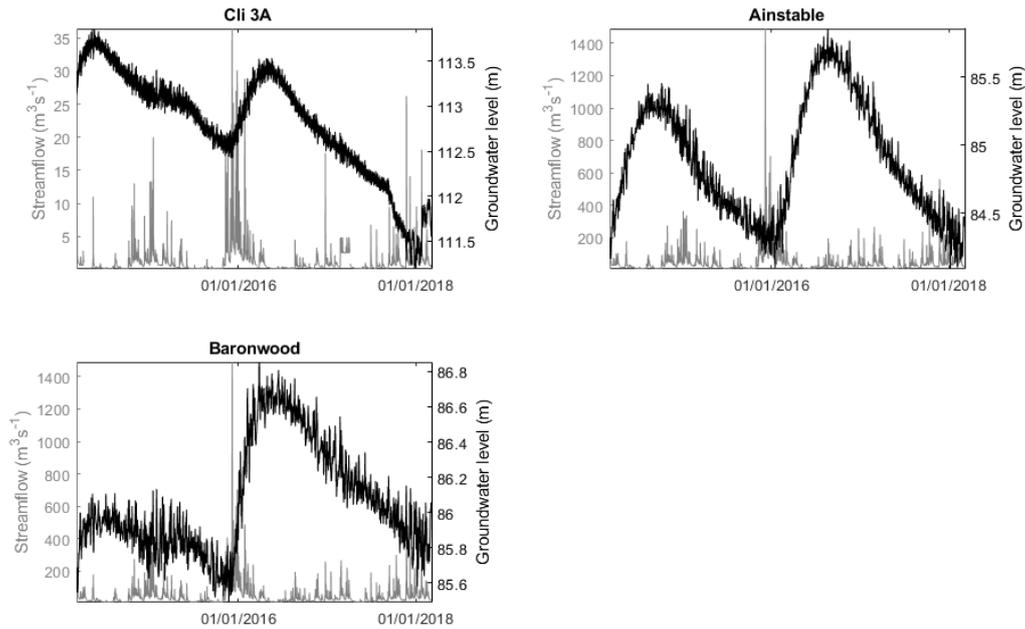


Figure 6.9. Comparing boreholes with apparent 2 year cycle with their closest flow gauge, over the available data period.

The large peaks occurring around the beginning and middle of most of these time-series are linked to the stormy winter of 2013 and Storm Desmond at the end of 2015 and most likely account for the apparent 2-year cycle of three of the boreholes (Figure 6.9). Two of the boreholes, East Curthwaite and North Scaleby, appear to be unaffected by these storms and both are in the lowlands away from waterways. Baronwood is unaffected by the storms of 2013 but is heavily impacted by Storm Desmond.

These boreholes display no connectivity to surface waterways, even the two that are only ~30m away from the Hilton Beck (Coupland and Hilton). It is likely they play no part in the streamflow response to rainfall and simply 'fill up' as rainfall intensity increases and drain once the intensity levels or drops off (Figures 6.10 to 6.12). Rainfall intensity in this case was estimated using a direct form II transposed implementation of a standard difference equation, with parameters $b = 1$ and $a = [1-0.9996]$;

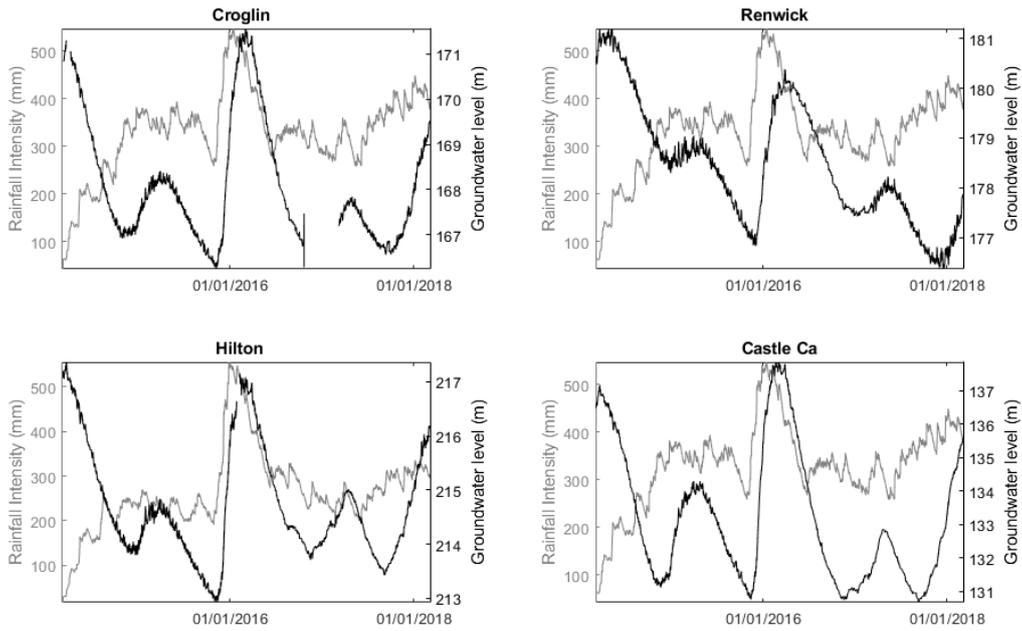


Figure 6.10. Comparing rainfall intensity from the nearest rain gauge to each borehole to groundwater, over the available data period.

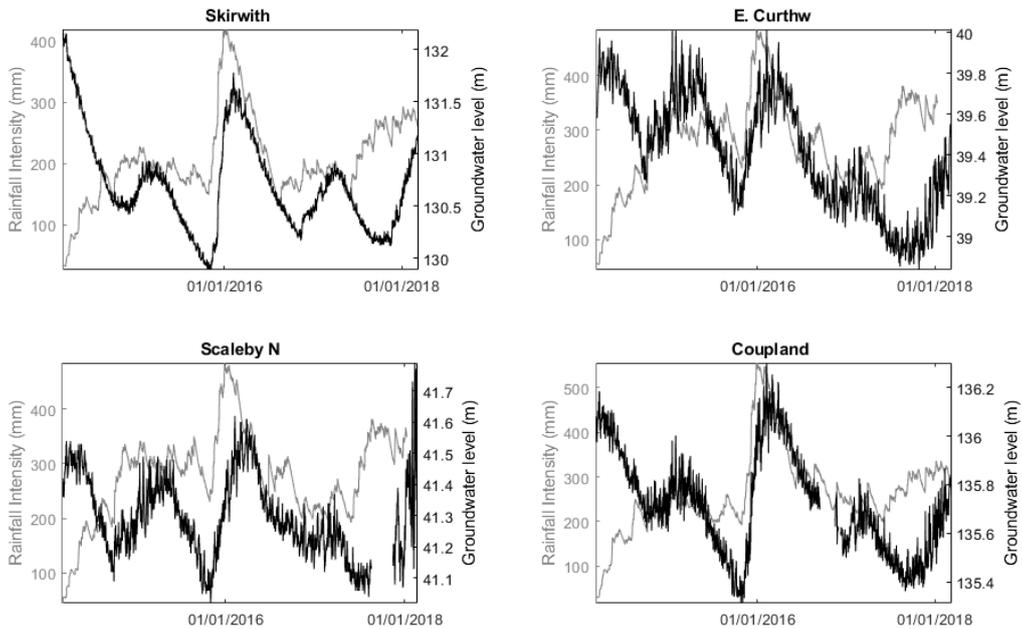


Figure 6.11. Comparing rainfall intensity from the nearest rain gauge to each borehole to groundwater, over the available data period.

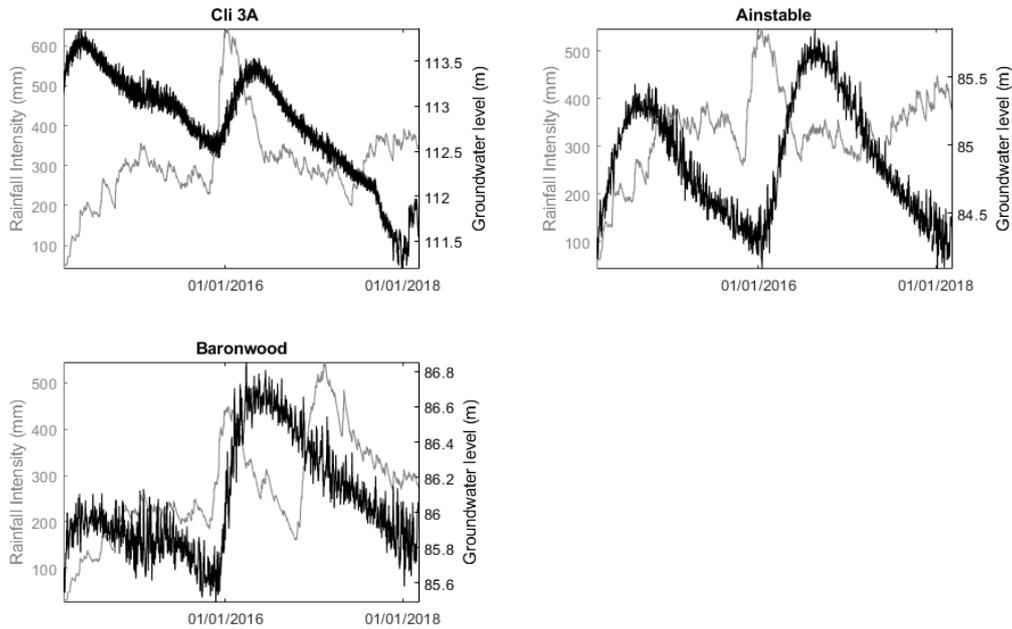


Figure 6.12. Comparing rainfall intensity from the nearest rain gauge to each borehole to groundwater, over the available data period.

For the most part, after rainfall intensity increases (the length of time for ‘after’ varies on borehole location), there is a corresponding increase in the groundwater level and as rainfall intensity levels off or decreases, the groundwater level decreases. The Cliburn 3A, Ainstable and Baronwood boreholes (Figure 6.12) are still responding to the winter storms of 2013 and 2015 and seem to largely ignore the rainfalls of 2014, 2016 and 2017 (some hints of response to winter of 2014). This suggests these three boreholes have very slow drainages and are very poorly connected to any water system.

6.2.3.3 ‘Random’

Three boreholes can be described as random as there is no discernible pattern, or that pattern is hidden by high levels of noise (Figure 6.13).

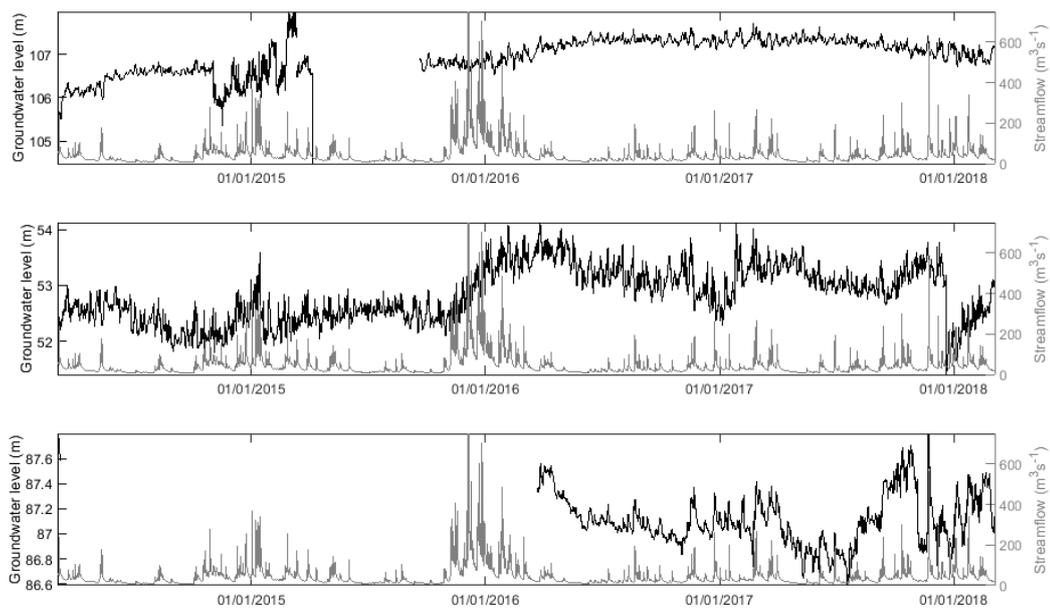


Figure 6.13. Comparing boreholes with Great Corby flow (Eden) over the available data period (note, Storm Desmond’s peak flows cut off for visualisation reasons). Top, Brownrigg. Middle, H.S. Plant. Bottom, BGS EV2.

All three boreholes are located in the lowlands, surrounded by fields (H.S. Plant is in a section of woodland that is surrounded by fields), and are far from waterways; BGS EV2, 240m, Brownrigg, 1km, and H.S. Plant 3km away.

H.S. Plant demonstrates a step increase in level after Storm Desmond (before, 52-53m and after 53-54m) and Brownrigg suggests something similar. However, the reasons for this behaviour are not apparent.

6.2.3.4 Sub-Daily Variations

All the boreholes have sub-daily variations and they are likely the responses to individual rainfall events and while this is clear in the three ‘flow like’ boreholes, it is harder to see in the others. The amplitudes and frequencies of the sub-daily variations vary between borehole and are likely influenced by the localized conditions of each borehole. For instance, the five boreholes (Croglin, Renwick, Hilton, Castle Carrock and Skirwith) that have the lowest variation are all situated in the St. Bees sandstone aquifer, which has been shown to be more homogenous and have lower vertical hydraulic conductivities than the Penrith sandstone aquifer (Younger and Milne, 1997), which contains the other twelve boreholes.

6.2.4 Soil Moisture Response

Soil moisture responds as expected to rainfall (Figure 6.14), increases with rainfall and decreases rapidly without. However, the soil does seem to start saturating around a volumetric water content of 77%, and this is very clear during Storm Desmond where soil moisture fluctuates between 77% and 78% for around 3 months.

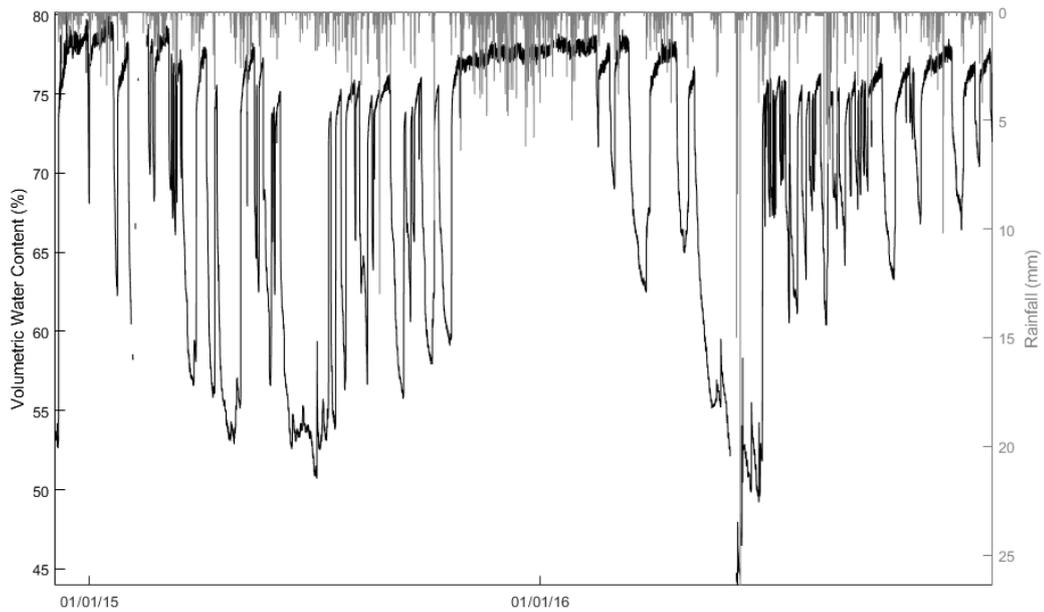


Figure 6.14. Comparing Pennines soil moisture with the Brackenbar rain gauge.

Given the logical response to rainfall, then it is expected that the shape of the soil moisture response should be similar to that of the streamflow. This shows a clear nonlinear effect; as volumetric water content increases, flow increases and when the flow enters recession, so does the water content (Figure 6.15). Additionally, when the soil moisture has reached its upper limit, the flow is elevated until soil moisture drops below this limit.

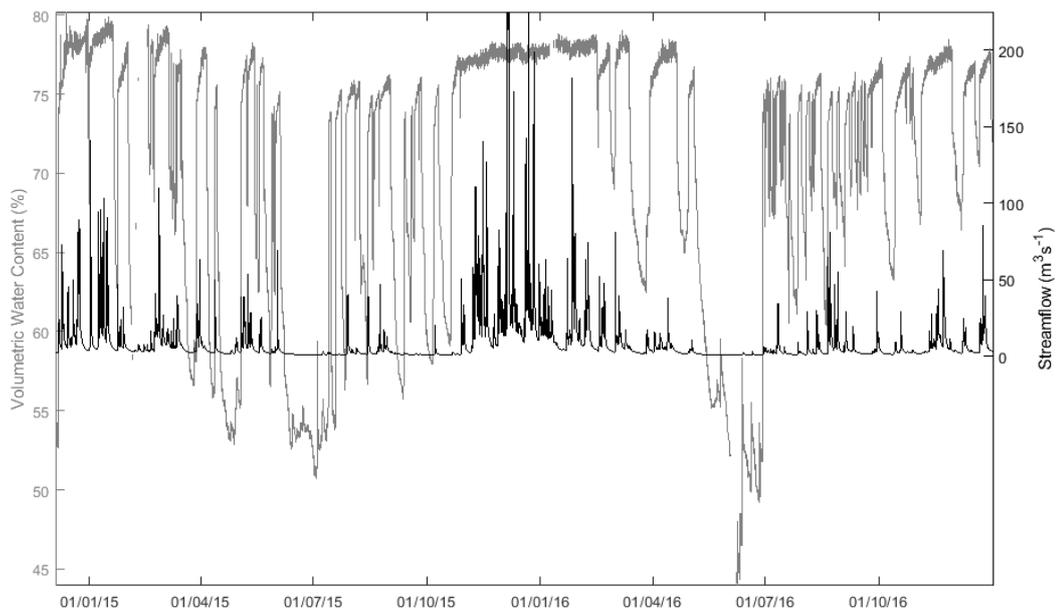


Figure 6.15. Comparing Pennines soil moisture to the upland flow gauge at Great Musgrave.

6.2.5 Exploratory Data Analysis Conclusions

The rainfall drivers are typical for frontal systems, with less rain as the front passes over the catchment and more rain in the uplands.

In terms of response to rainfall, it appears the surface responds largely uniformly, while the subsurface responds highly non-uniformly. This means it is difficult to link subsurface processes to streamflow generation and highlights just how spatially and temporally dynamic catchments are under the surface (Table 6.3).

Table 6.3. Displaying the variety of the subsurface in the vicinity of the boreholes. For location refer to Figure 6.1.

Borehole	Aquifer	Superficials	Level Range (m)	Category
Croglin	St. Bees	None - aquifer only	5.32	Drain/fill
Renwick	St. Bees	Unconsolidated sand/gravel (0-3m deep)	4.84	Drain/fill
Hilton	St. Bees	Boulder Clay (0.5-6m deep) Sand/gravel (6-10m deep)	4.49	Drain/fill
Castle Carrock	St. Bees	Sand/gravel (glacio-fluvial, 0-3m deep)	7.19	Drain/fill
Skirwith	St. Bees	Boulder Clay (including igneous material) (0-3m deep)	2.50	Drain/fill
East Curthwaite	Penrith	Unknown	1.18	Drain/fill
Scaleby North	Penrith	Glacial till	7.17	Drain/fill
Cliburn Hill 3A	Penrith	Unconsolidated sand (0.5-2m deep)	2.68	Drain/fill
Coupland	Penrith	Sandy Clay (0-5m deep)	0.99	Drain/fill
Ainstable	Penrith	Boulder Clay (0-15m deep) Sand and Pebbles (15-18m deep)	1.77	Drain/fill
Baronwood	Penrith	Sand/gravel (glacio-fluvial, 0-3m deep)	1.49	Drain/fill
Great Musgrave	Penrith	Boulder Clay (0-24m deep)	1.82	Flow like
Cliburn Tower Bridge 2 (Deep)	Penrith	Alluvial deposits (0-1.5m deep)	1.02	Flow like
Cliburn Tower Bridge 1 (Shallow)	Penrith	Alluvial deposits (0-1.5m deep)	1.71	Flow like
East Brownrigg	Penrith	Sand/boulders (0.4-4.5m deep) Boulder clay (4.5-10m deep)	3.50	Random
H.S. Plant	Penrith	Unknown	2.73	Random
BGS EV2	Penrith	Unknown	1.60	Random

Soil moisture response in the Pennines does not resemble anything like the groundwater response for boreholes in similar settings, making it difficult to study the soil-subsurface interface. However, there is a good link between soil moisture response and streamflow response and indicates the potential of soil moisture data being useful for rainfall-runoff analysis.

6.3 Rainfall-Runoff Models

A simple time-varying model, and thus nonlinear model (6.1) is used to model the runoff response to rainfall.

$$Q_t = \alpha_t Q_{t-1} + \beta_t R_{t-\delta} \quad (6.1),$$

where, Q is streamflow, R is rainfall, t is time, δ is the time-delay for the input to be observed in the output, and α and β are parameters.

The α parameter represents the time-constant (T) of the system (6.1), a measure of how quickly the system responds to an input. The β parameter is related to the steady state gain (ssg) of the system (6.2), the amount the input influences the output.

$$ssg_t = \frac{\beta_t}{1+\alpha_t} \quad (2), \quad T_t = -\frac{\Delta t}{\ln(\alpha_t)} \quad (6.2).$$

These time-varying parameters (α and β) are estimated, in turn using a Kalman Filter (Kalman, 1960) and Fixed-Interval Smoother back-fitting algorithm (Young, 2000). One important hyper-parameter of these types of models is the Noise Variance Ratio (NVR, Young, 2011, chapter 17), this essentially controls the speed at which the parameter estimates change. The higher the NVR, the faster the parameters can change, meaning the greater variance in the parameter estimate reflecting increasing numbers of dynamics captured. This means models estimated using high NVRs will reflect the faster dominant pathways and models estimated using lower NVRs will reflect the slower dominant pathways.

The estimated gain (β) is considered to reflect the dominant modes of runoff generation and the aim here is to see which variables of the catchment are closely linked to the estimated gain, at both high and low NVRs. To do this, the α parameter will not be allowed to vary and so will be estimated as a constant parameter. However, to ensure that it is predominately the gain that is the important parameter, the same procedure used to compare the gain with the catchment variables, will be used to compare the time constant with the variables; α parameter will be time-varying and the β will be estimated as a constant.

6.3.1 Fast Processes

The fast processes are ones that translate rainfall into streamflow within a few hours (event by event), such as saturated soil causing more overland flow to be generated. Catchment variables with high frequencies will be compared to the estimated gain using a high NVR, those variables being soil moisture, 'flow-like' boreholes and rainfall.

6.3.1.2 Leith – A Small Upland System

Estimating Time-varying gain with a constant α

The pure time-delay for rainfall to flow was estimated at 1 hour. Using an NVR of $1e^{-2}$, the α parameter was estimated to be 0.966, which gives a time-constant of 29.5 hours, with the estimate of the gain fluctuating around zero (Figure 6.16), where positive values scale up the input and negative values mean the input does not lead to an increase in the output and means the output is decreasing (scaling with a negative value turns the additive of the input into a subtractive).

Physically, the input of rainfall is always additive (rainfall cannot go in reverse), the reason why the gain can go negative has several possible reasons.

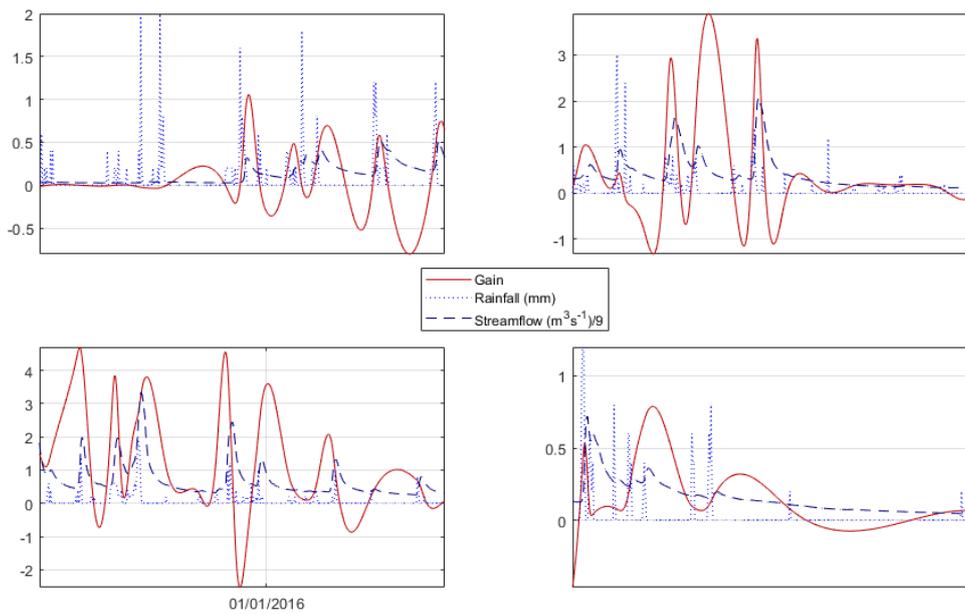


Figure 6.16. Comparing flow, rainfall and estimate gain for the River Leith system.

The first, is that rainfall at $t - \delta$ is not responsible for the increase in flow at t , suggesting an error in time-delay; how quickly a catchment responds to an input of rainfall depends on the current conditions of the catchment (how wet it is) and so time-delay is actually time-varying, but the model has a fixed time-delay. Secondly, if the flow is in recession at t and the amount of rainfall at $t - \delta$ is small, then it is possible the flow does not respond and since it is in recession, the gain has to be estimated as negative to account for the decrease in flow. Thirdly, a temporary change in the runoff pathways or mechanisms could prevent an input of rainfall from affecting the flow.

It should be noted that mathematically when the input is zero, the estimated gain will have no effect on the output (as it is multiplied by the input). So, a more accurate view of the gain would be to replace the estimates of gain at t with zero when $R_{t-\delta} = 0$, this then gives a better picture (Figure 6.17) of which rainfall events have an impact on the flow and to what extent, and which have no impact. For example, in Figure 6.17 (bottom left) a short series of rainfall events have a relatively high negative gain as there is no response in flow during its recession, whereas some of the rainfall events in Figure 6.17 (top right) have high positive gains due to large response by the flow.

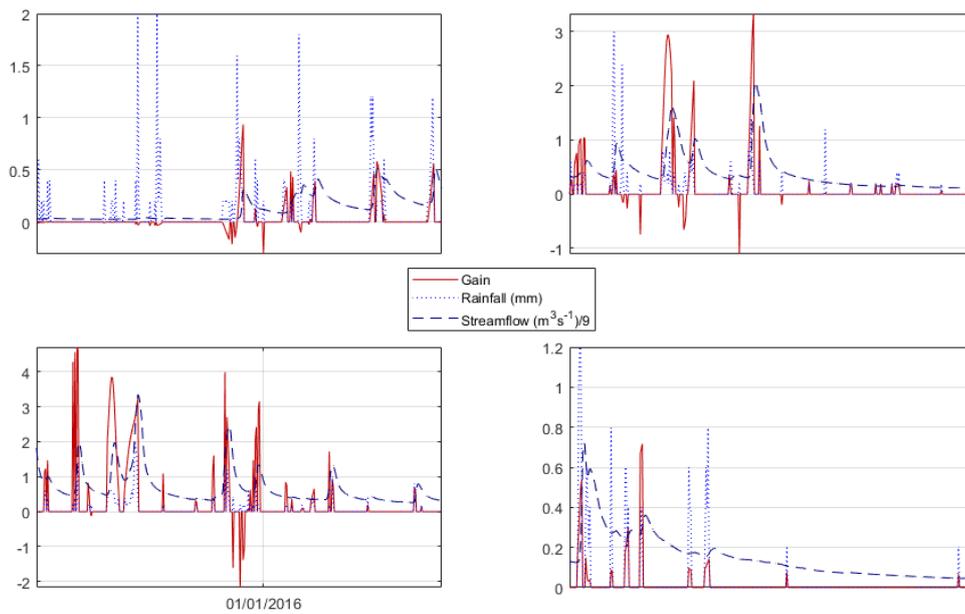


Figure 6.17. Comparing flow and rainfall with the modified gain ($b_t=0$ when $R_{t-\delta}=0$).

While there is a direct link between rainfall and gain, there is not so much between soil moisture and gain, despite there being a good link between rainfall and soil moisture (Figure 6.18). When gain > 1 then soil moisture is typically increasing or at saturation point, but outside this, there is no relationship.

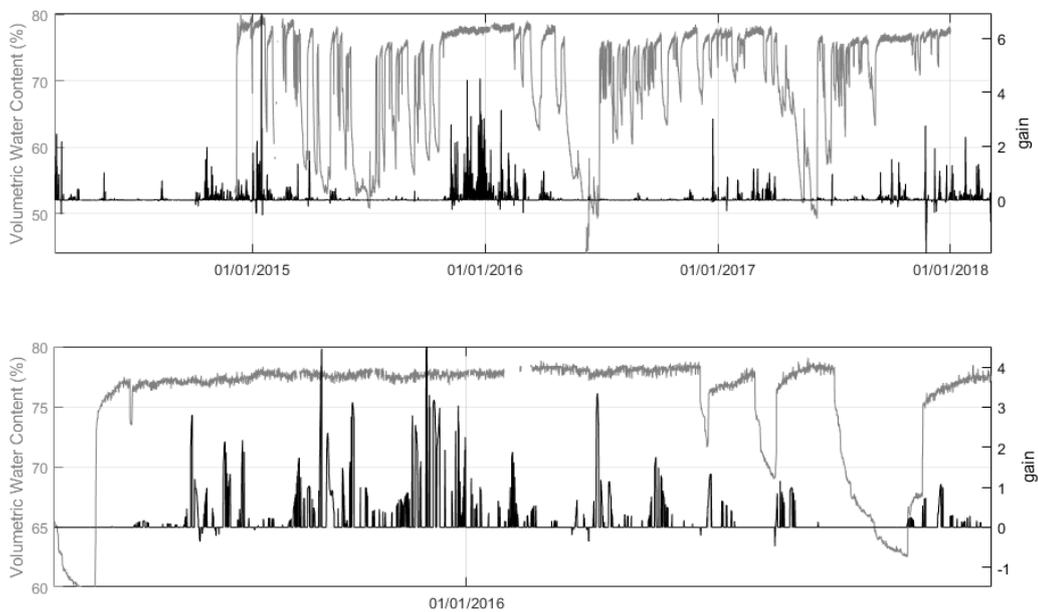


Figure 6.18. Comparing soil moisture and modified gain ($b_t=0$ when $R_t=0$).

Since the gain was estimated using flow, it is logical that there is a good link between the ‘flow like’ borehole at Cliburn (shallow), typically when the gain is zero the borehole is draining and there is lagged response between gain rising and then borehole refilling (Figure 6.19).

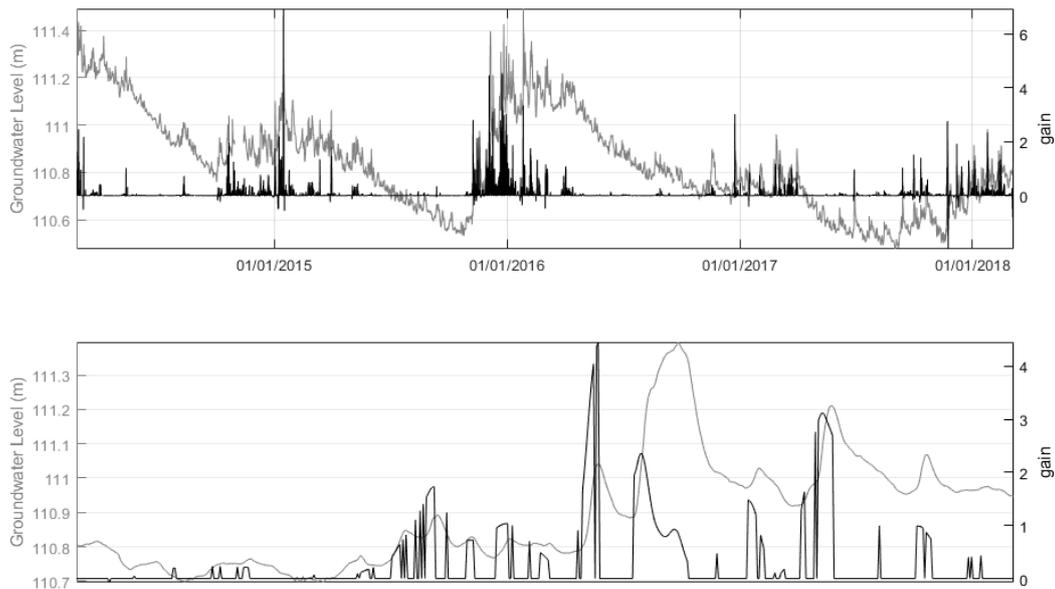


Figure 6.19. Comparing ‘flow-like’ borehole Cliburn (shallow) and modified gain ($b_t=0$ when $R_{t-\delta}=0$).

Estimating Time-varying α with a constant gain

With a high NVR of $1e^{-2}$, the estimation of α fluctuates around 1, where $\alpha > 1$ implies increased gain of the streamflow (more than the input would lead to) and a $\alpha < 1$ implies a loss in streamflow. Typically, $\alpha > 1$ occurs when flow increases and $\alpha < 1$ occurs when the flow is in recession (Figure 20).

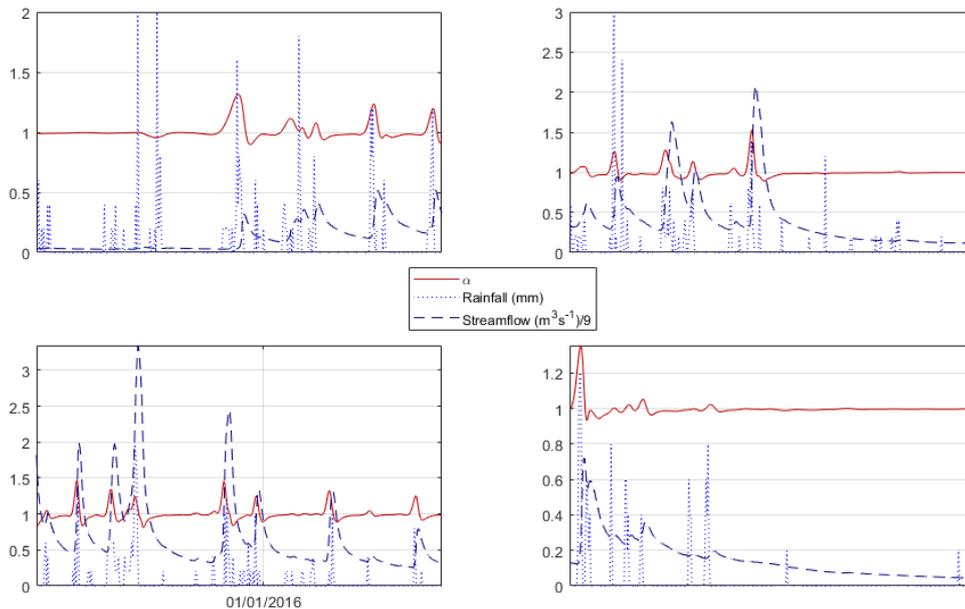


Figure 6.20. Comparing flow and rainfall with α .

The usual physical interpretation of α as a time constant is not valid for these types of time-varying hydrological models, as $\alpha > 1$ implies a negative time constant, which indicates an unstable system

and for constant parameters is physically nonsense. For the rainfall-runoff models here, the α parameter can be interpreted as a water recession or loss parameter, additional storage growth parameter in absence of rainfall input. $\alpha < 1$ implies the flow is in recession, $\alpha = 1$ implies accumulation of flow, and $\alpha > 1$ implies the addition of a source of water (such as uncounted tributary or aquifer involvement) not related to the input ($R_{t-\delta}$). The additional source(s) of water in this case is more likely due to either missed rainfall events or the rainfall intensity at the gauge was less than elsewhere in the catchment, both being symptoms of the nature of frontal weather systems. It could be possible that new tributaries open during long or intense rainfall events, which allow more rainfall to reach the flow gauge quicker, but it would be practically impossible to know for sure.

The α parameter has a worse relationship with soil moisture than the gain does (Figure 6.21).

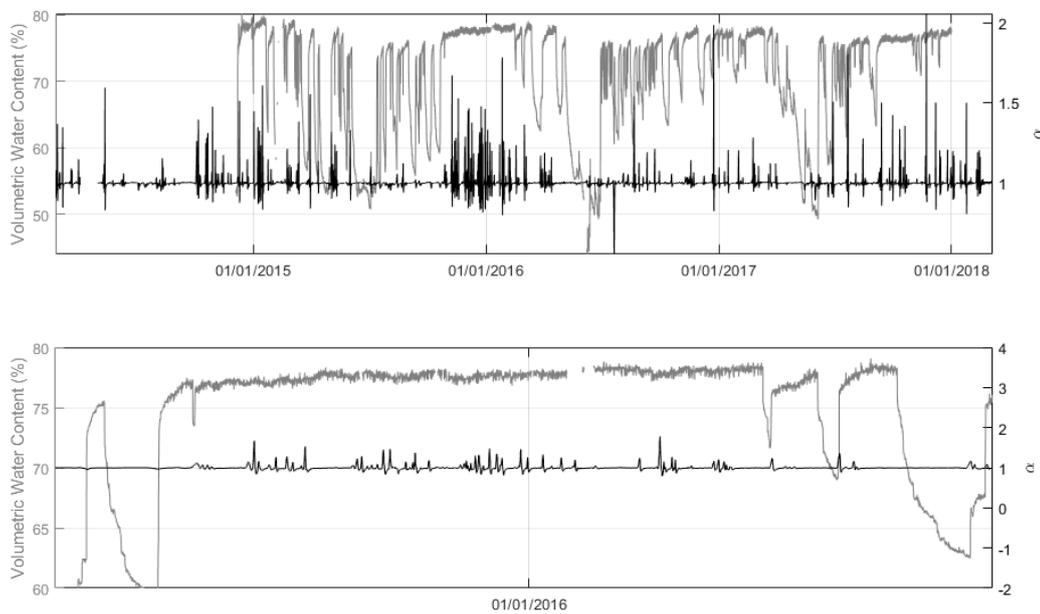


Figure 6.22. Comparing 'flow-like' borehole Cliburn (shallow) and α .

Similarly, to gain, the α parameter has a lagged relationship with the Cliburn shallow borehole in that α responds first (by becoming >1) and then the groundwater level increases, also when $\alpha < 1$ the groundwater level is typically in decline (Figure 6.22).

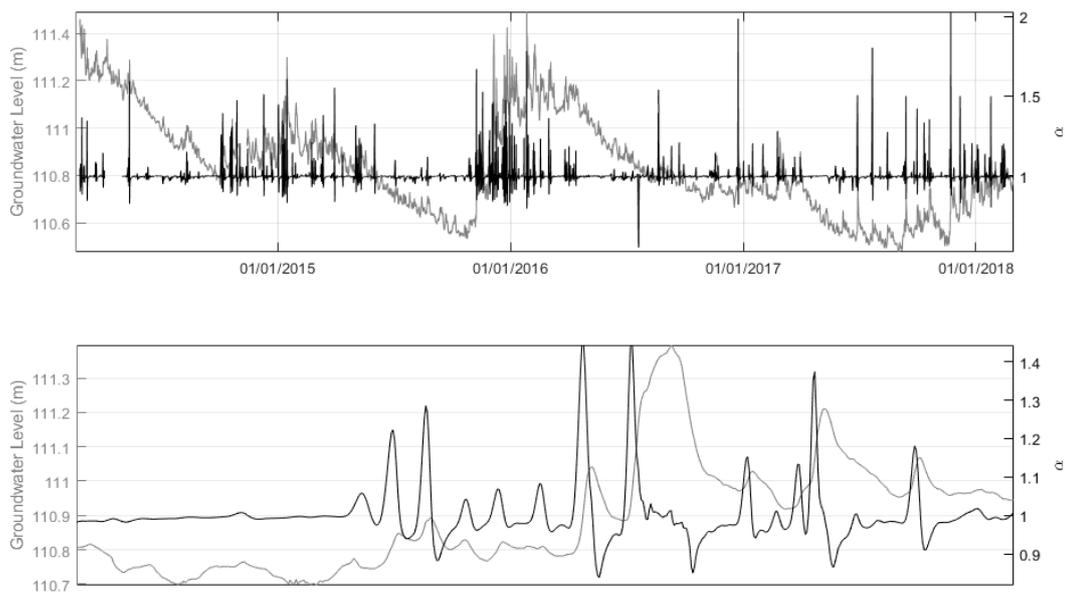


Figure 6.22. Comparing groundwater

6.3.2 Slow Processes

Estimating Time-varying gain with a constant α

The slow processes can be interpreted as the subsurface pathways rainfall takes, keeping the stream constantly flowing. Catchment variables with low frequencies will be compared with the estimated gain using a low NVR, those variables being the slow boreholes and rainfall intensity.

6.3.2.1 Temple Sowerby – Large Upland System

The time-delay for rainfall to flow was estimated at 4 hours. Using an NVR of $1e-10$, the α parameter was estimated to be 0.9705, which gives a time-constant of 33.4 hours, while the estimate of the gain had a similar profile to the annual cycle of streamflow and rarely went below zero (Figure 6.23). While for the fast processes setting the gain to zero when rainfall was zero was a logical step as we were interested in the highest frequency relationships. Here we are interested in the low frequency relationships and while applying $b_t=0$ when $R_{t-4}=0$ doesn't hide this, it makes the visual comparisons tricky (Figure 6.23).

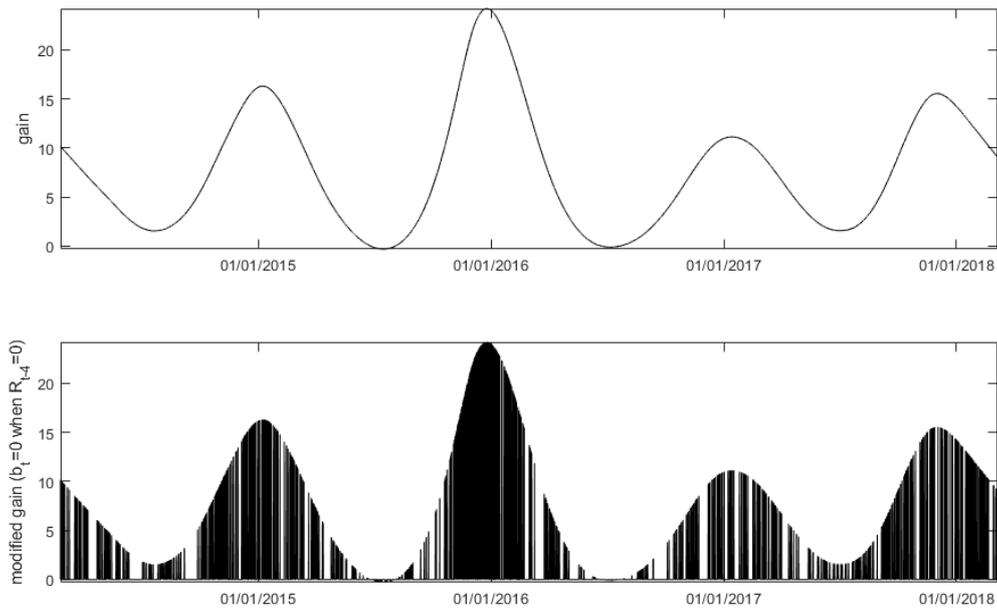


Figure 6.23. Visualisation of the gain and how the modified gain makes visual comparisons difficult. Top, estimated gain. Bottom, modified gain.

Because the gain resembles the annual cycle of flow, it is logical to see some similarity with the 'drain/refill' boreholes in the upland (Figure 6.24) as they also display annual cycles. However, because these cycles are not in time with each other, this suggests there is no link between the slow processes associated with groundwater and the slow evolution of the gain. The gain is much more similar to the streamflow cycle (Figure 6.25).

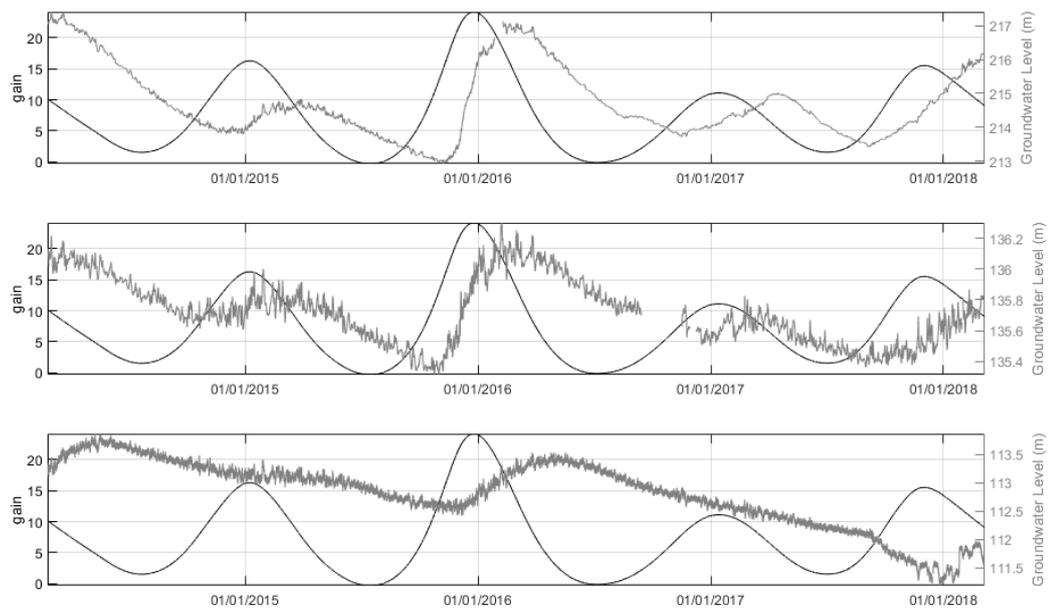


Figure 6.24. Comparing gain with the 'drain/refill' boreholes. Top, Hilton. Middle, Coupland. Bottom, Cliburn Hill 3A

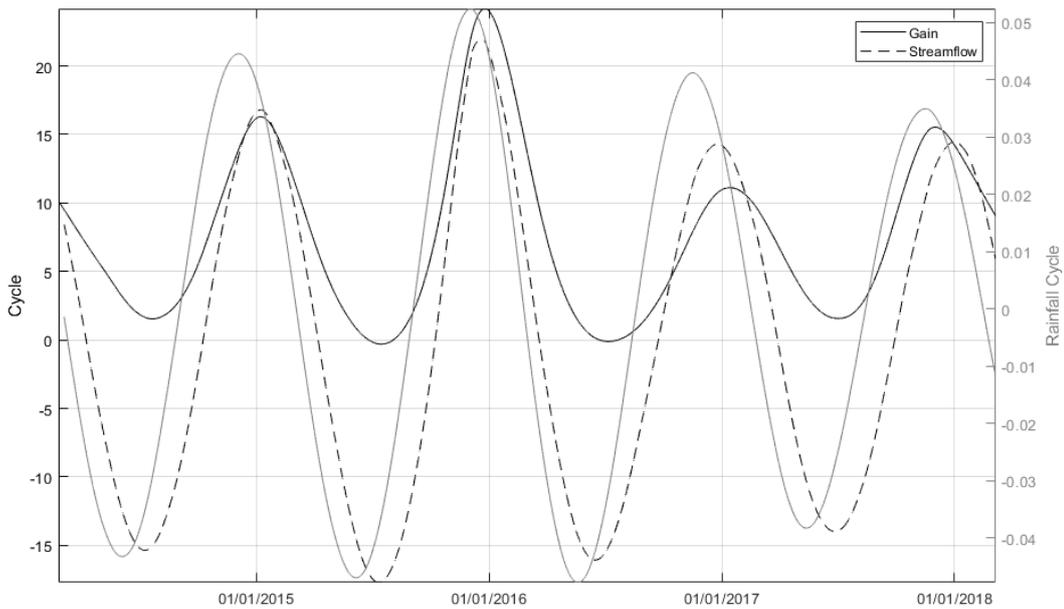


Figure 6.25. Comparing gain with the annual cycles of rainfall and streamflow.

Rainfall intensity, that has something in common with the slower boreholes, does not have much in common with the gain, except for the response to Storm Desmond (Figure 6.26) and could suggest other mechanisms of refill are more important.

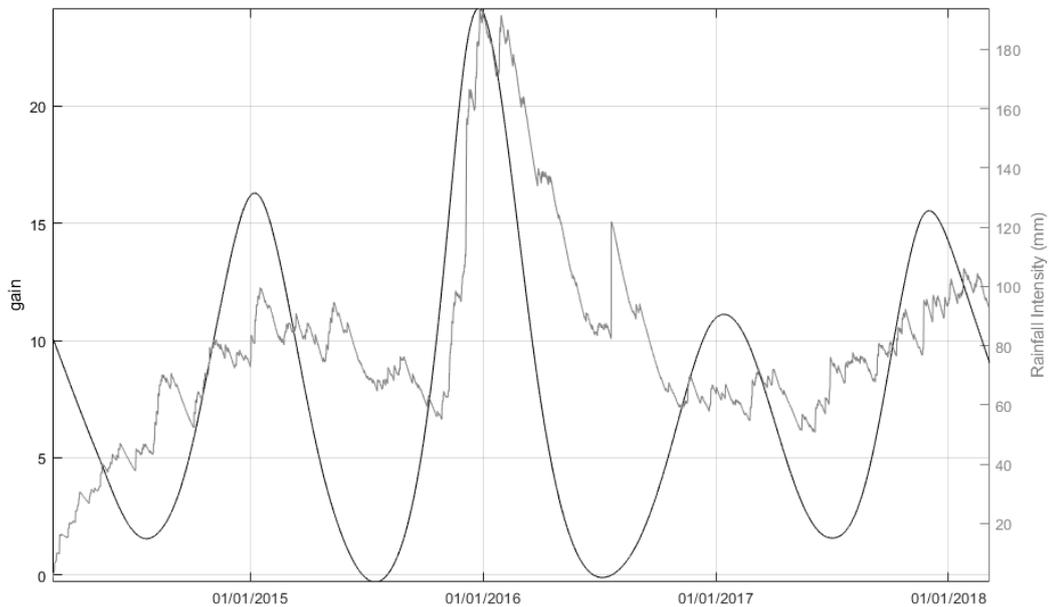


Figure 26. Comparing gain with rainfall intensity.

Estimating Time-varying α with a constant gain

With a NVR of $1e^{-12}$, the estimated α parameter fluctuates between 0.970 and 0.995, but the shape is not like anything observable in the slower variables (Figure 6.27 and 6.28), suggesting the time-

varying α parameter is not physically linked to anything observable and is likely just a product of fitting the data.

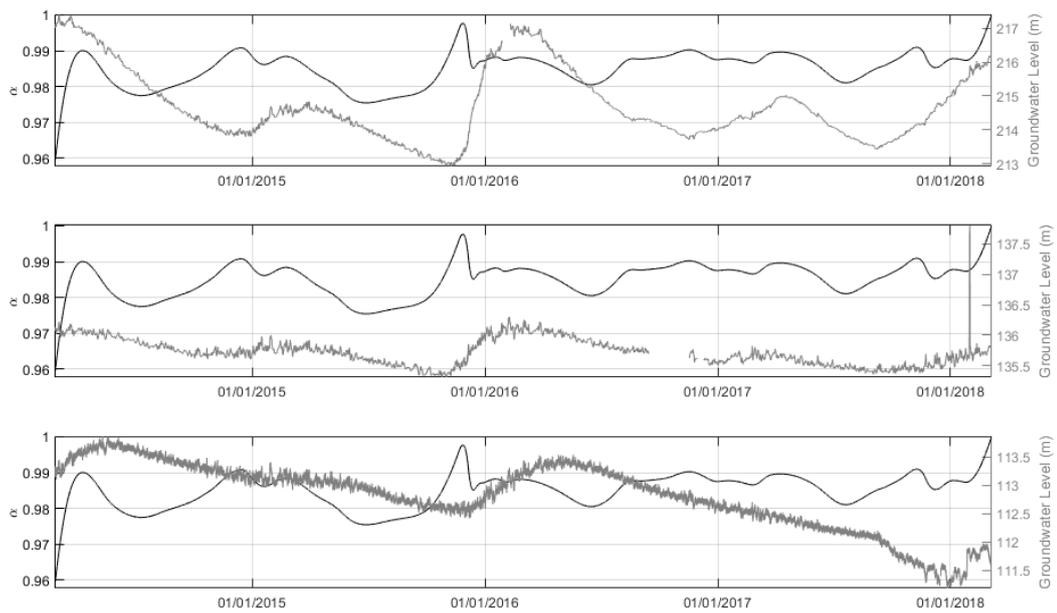


Figure 6.27. Comparing α with the 'drain/refill' boreholes. Top, Hilton. Middle, Coupland. Bottom, Cliburn Hill 3A

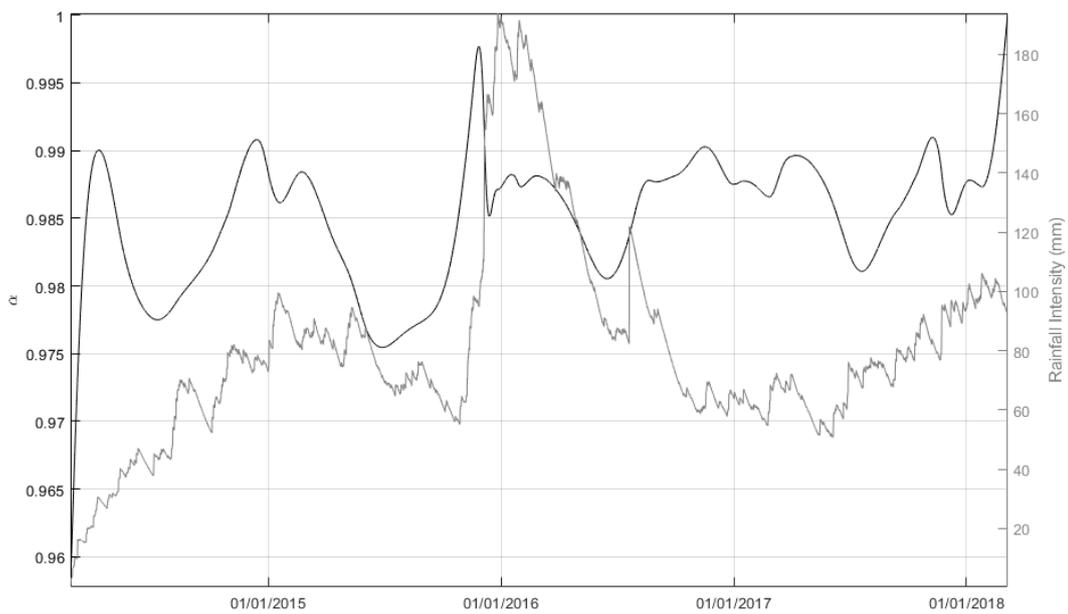


Figure 6.28. Comparing α with rainfall intensity.

6.3.3 Rainfall-Runoff Complications

Firstly, the α parameter is not linked to any of the slower processes, while the gains only link to the slower processes is through the annual weather cycle, particularly the streamflow cycle. There is a clear link between the faster processes and both parameters, which is tied directly to rainfall.

This may indicate that the rainfall-runoff process is all about the fast processes and rainfall is not only the driver but the dominant aspect of the system. However, the sheer dynamic nature of the catchment means there is no obvious link between gain/ α , rainfall/flow; a given value of either parameter could have different values of rainfall and/or flow associated with it (Figure 6.29 and 6.30).

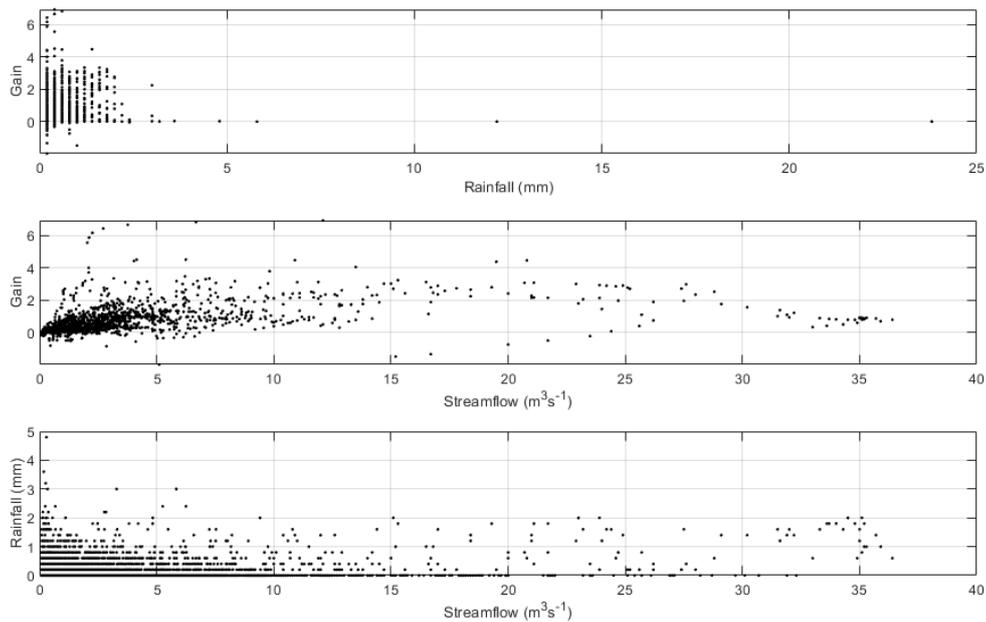


Figure 6.29. Showing how the same rainfall and flow values can lead to the same gain. Values of gain when $R_{t-6}=0$ are not displayed for ease of visualisation. Note, bottom plot does not show rainfall values greater than 5 for ease of visualisation (2 values not shown).

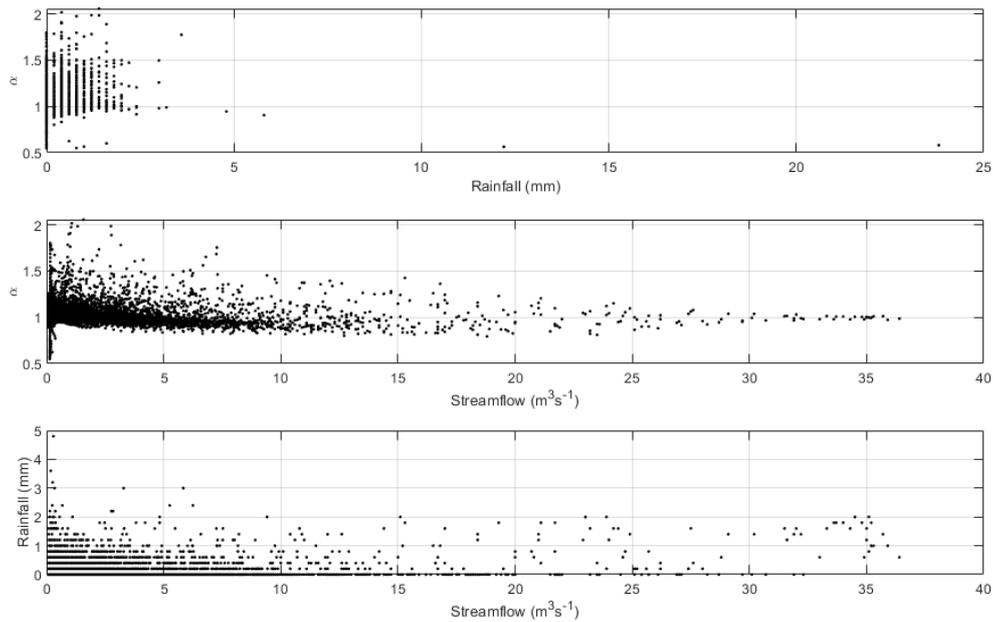


Figure 6.29. Showing how the same rainfall and flow values can lead to the same α . Note, bottom plot does not show rainfall values greater than 5 for ease of visualisation (2 values not shown).

It can be seen (Figure 6.29, bottom) that a streamflow value can have many corresponding rainfall values, and this leads to the issue of the estimated parameters having different values for the same corresponding value of rainfall or streamflow. This clearly demonstrates the nonlinear nature of the rainfall-runoff system and not only complicates the modelling procedure but also any forecasting procedures; if rainfall events of the same magnitude can lead to a variety of streamflow responses, then accurate forecasting becomes extremely difficult.

6.3.4 Extended State Dependent Parameter models

The rainfall-runoff complexity is assumed to be controlled by some other aspect of the catchment, such as soil moisture or groundwater, that represents the antecedent conditions that control how the stream responds to a rainfall event.

The trick is to find and then incorporate the variable representing antecedent conditions into our models. One method of incorporating this variable is Extended State Dependent Parameter (ESDP, Mindham et al., 2018) modelling, which is an extension of the time varying rainfall-runoff model described by (6.1). ESDP assumes that each parameter is dependent on zero to two variables from the catchment, including lagged inputs and outputs (6.3), and uses those variables in estimating the parameters.

$$Q_t = \alpha(\cdot)_t Q_{t-1} + \beta(\cdot)_t R_{t-\delta} \quad (6.3),$$

where, (\cdot) represents any variables ('states') the parameter is dependent on; $()$ for no state dependency, $(s1)$ for one state, $(s1, s2)$ for two states.

The variables to consider as states that could dominate control of the antecedent conditions and thus be the 'catchment wetness' proxy are:

- Volumetric water content (soil moisture) – as this controls how much of the rain water will infiltrate the subsurface and how much will flow over land.
- Groundwater level adjacent to the stream (Cliburn shallow borehole) – as this is an indicator of the water table level and an increase in table level will lead to an increase in streamflow.
- Groundwater level of a fast responding, non-stream linked (Great Musgrave borehole) – as this borehole responds quickly to rainfall events and recedes in similar manner to streamflow. This could provide a useful proxy measurement for the catchment antecedent conditions.
- Rainfall intensity (filtered rainfall) – as rainfall intensity is a measure of the cumulative effects of past rainfall events and as the duration and intensity of events increase, the likelihood the streamflow will increase.

Eight models were compared, each model had one constant parameter and the other dependent on one of the above variables, and then a full model simulation was conducted and R_t^2 for model fit was calculated. This is the estimation of the output using the input and the estimated parameters; instead of Q_{t-1} from (6.3) being the observed value, it is the estimated value \hat{Q}_{t-1} , with $\hat{Q}_1 = 0$. These models were for the small River Leith sub-catchment.

From the available data it is difficult to see which of the four variables would be a better ‘catchment wetness’ proxy, nor which parameter should depend on them (Table 6.4). They all result in reasonable rainfall-runoff models, but underestimate the high flows and overestimate the low flows.

Table 6.4. Comparing rainfall-runoff models where one parameter is constant and the other is dependent on another catchment variable. Hourly data from 16/02/2015 1100 to 02/06/2016 1200. Soil moisture was smoothed to remove the rapid variation when the saturation limit was reached and to interpolate a few gaps in the data*.

Variable	Time-varying α			Time-varying gain		
	α_t range	β	Full Sim R_t^2	α	β_t range	Full Sim R_t^2
Soil Moisture*	0.786 to 1.051	0.964	0.70	0.912	-0.510 to 1.652	0.77
Great Musgrave borehole	0.961 to 1.022	0.959	0.42	0.959	-0.170 to 1.945	0.74
Cliburn (shallow) borehole	0.966 to 0.992	0.961	0.63	0.967	0.009 to 1.997	0.69
Rainfall Intensity	0.879 to 1.021	0.965	0.66	0.951	-0.513 to 2.087	0.73

6.4 Conclusions

The Exploratory data analysis demonstrated how varied the subsurface response to rainfall is, while in stark contrast the surface response is relatively uniform. This suggests that the involvement of the subsurface in runoff response is minimal and rainfall-runoff modelling showed that it is the fast processes, dominated by rainfall, that are important.

In terms of identifying a 'catchment wetness' proxy to incorporate the antecedent conditions that control the runoff response, ESDP modelling suggests that soil moisture, rainfall intensity and 'flow like' boreholes all have potential and further exploratory data analysis and rainfall-runoff modelling is required, with more data than is currently available (soil moisture in particular).

Chapter 7: Conclusions and Recommendations

7.0 Conclusions

The modelling tools developed here allow for a more objective analysis of environmental data by reducing the researcher's impact on the modelling procedure. First, for SDP based models, there is now a statistical tool for identifying the most likely model structure that not only includes number of parameters, but the states that these parameters depend upon. This tool largely removes the researcher from the model structure selection step, a must for following the DBM approach, and allows the researcher to see which variables, in an objective way, are the dominant drivers for the observed processes in the system under study. The development of ESDP has led to better estimates of the model parameters, particularly at the ends of the time-series (due to less information) for single states, and in the dense clusters of state-space data points for multiple states (in general – in situations of highly non-uniform distribution of the driving states).

Secondly, for state space models, the complete removal of the data resampling step, means that only the observed data is analysed and the artefacts and assumptions associated with sampling regularisation and data interpolation are avoided, as these methods are no longer required, simplifying data handling and analysis of irregularly sampled environmental systems. Additionally, the forecasting potential of DHR has been increased with the development of ASDHR. Firstly, the forecasting horizon no longer must begin when the data ends, but at any point in the future, within reason. And secondly, the forecasting sampling interval is no longer determined by the sampling rate of the data and can be specified by the researcher. However, discretion is advised here, while forecasting at a lower sampling frequency than the observed data is fine, forecasting at a sampling frequency higher than the observed data becomes dubious due to the observed frequency range of the data.

The DBM approach is about analysing the data in an objective manner as possible, with as few assumptions as possible. The developments here add to the objectivity of DBM modelling to a broader range of systems and reduce the amount of assumptions used, while the knowledge gained during their development help the author guide the application of DBM philosophy to data analysis.

7.1 Recommendations

The drive for application of the DBM philosophy to environmental systems is ongoing and while this work increased the number of tools available, there are always improvements to be made and other analyses to be done. Of these the following three major extensions need addressing in the first place.

- The arbitrary sampling technique developed as part of objective three deliverables is currently implemented only for univariate signal analysis and as such needs to be generalised to work for all forms of state space modelling, such as the time-varying parameter models that use equation (Introduction: 7), including vector observations and Input-Output models, such as currently used at fixed sampling rates by Beven (2012) and others in rainfall driven flood forecasting methods.
- The parameter estimation of ESDP models could be further improved by applying the RIV method of model identification and parameter estimation used for linear TF models. This is no trivial matter as RIV would have to be expanded to work for time-varying parameters.

- While ASDHR can identify long term cycles in paleo-climatic series, there is currently no uncertainty estimate for the timing of the cycles. A starting point would involve identifying cycles of simulated time-series with known cycles and the use of Monte Carlo analysis to estimate the error variance for varying noise levels, time-series length, and cyclic periods to gain an understanding of period identification sensitivity.

Acknowledgements

The author would like to thank his primary supervisor Wlodek Tych for his patience and guidance, particularly during the technical and mathematical aspects of the thesis, as well as his general kindness and pleasantness to the author. The author would also like to thank his secondary supervisor Nick Chappell for his guidance on context and perspective during various application aspects of the thesis. The author also needs to thank his parents, Emma and Tony Mindham, for their continuing moral and financial support before and during thesis development.

Finally, the author is obliged (read 'forced') to thank his cats, Freyja and Hnoss, for their benevolence in allowing their servant to work on this thesis. They wrote their acknowledgement section for the author to use, but it was far too long. Part of their benevolence involved sacrificing many hours of attention for the thesis to be completed and aided in funding by providing many bodies.

References

- Ait-El-Fquih, B., Desbourvries, F. 2008. On bayesian fixed-interval smoothing algorithms, *Automatic Control, IEEE Transactions*, 53(10), 2437-2442.
- Akaike, H. 1972. Information theory and an extension of the maximum likelihood principle. *Proc. 2nd Int. Symp. Information Theory, Supp. to Problems of Control and Information Theory*, 267-281.
- Ampadu, B., Chappell, N. A., Tych, W. 2015. Data Based Mechanistic modelling optimal utilisation of raingauge data for rainfall-riverflow modelling of sparsely gauged tropical basin in Ghana. *Mathematical Theory and Modelling*, 5(8), 29-49.
- Anderson, K., Segall, P. 2011. Physics-based models of ground deformation and extrusion rate at effusively erupting volcanoes. *Journal of Geophysical Research*, 116(B7).
- Ando, B., Baglio, S., Graziani, S., Pitrone, N. 2000. Models for air quality management and assessment. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 30(3), 358-363.
- Ashkenazy, Y., Baker, D. R., Gildor, H., Havlin, S. 2003. Nonlinearity and multifractality of climate change in the past 420,000 years. *Geophysical Research Letters*, 30(22).
- Becker, S., Halsall, C. J., Tych, W., Hung, H., Attewell, S., Blanchard, P., Li, H., Fellin, P., Stern, G., Billeck, B., Friesen, S., 2006. Resolving the long-term trends of polycyclic aromatic hydrocarbons in the Canadian arctic atmosphere. *Environ. Sci. Technol.* 40, 3217–3222.
- Bellman, R. 1957. *Dynamic Programming*, New Jersey: Princeton University Press.
- Beven, K. J., 2012. *Rainfall-Runoff Modelling: The Primer*. 2nd Ed. Wiley-Blackwell.
- Bhar, R. 2010, *Stochastic Filtering with Applications in Finance*, World Scientific Publishing
- Billeck, B., Friesen, S. 2006. Resolving the Long-Term Trends of Polycyclic Aromatic Hydrocarbons in the Canadian Arctic Atmosphere. *Environmental Science & Technology*, 40, 3217-3222.
- Box, G. E. P., Tiao, G. C. 1975. Intervention Analysis with Applications to Economic and Environmental Problems. *Journal of the American Statistical Association*, 70 (349), 70-79.
- Brockwell, P. J. (2001). Levy-Driven Carma Processes. *Ann. Inst. Statist. Math*, 53 (1), 113-124.
- Broersen, P. M. T., de Waele, S., Bos, R. (2004). Autoregressive spectral analysis when observations are missing. *Automatica*, 40 (9), 1495-1504.
- Carling, P.A., Tych, W., Richardson, K. 2005, The hydraulic scaling of step-pool systems, *River, coastal and estuarine morphodynamics*, in Parker, G. and Garcia, M.H. (eds.), 144, 55-63, Taylor and Francis
- Chappell, N.A., Jones, T.D., Tych, W., Krishnaswamy, J. 2017. Role of rainstorm intensity underestimated by flood models: emerging global evidence from subsurface-dominated watersheds. *Environmental Modelling and Software*, 88, 1-9.
- Chappell, N. A., Tych, W. 2012. Identifying step changes in single streamflow and evaporation records due to forest cover change. *Hydrological Processes*, 26 (1), 100-116.
- Chappell, N.A., Jones, T.D., Tych, W. 2017. Sampling frequency for water quality variables in streams: systems analysis to quantify minimum monitoring rates. *Water Research*, 123, 49-57.

- Davis, P. J. 1975. Interpolation and Approximation. New York: Dover Publications Inc.
- Deutscher, J., Kupec, P., Dundek, P., Holík, L., Machala, M., Urban, J. 2016. Diurnal dynamics of streamflow in an upland forested micro-watershed during short precipitation-free periods is altered by tree sap flow. *Hydrol. Process.* 30, 2042.
- Fan, J., Gijbels, I. 1996. Local polynomial modelling and its applications: monographs on statistics and applied probability 66, (Vol. 66). CRC Press.
- Fang, H., Callafon, R. 2008. Stability analysis and application of kalman filtering with irregularly sampled measurements, in American Control Conference (ACC), 2012, June 2012, 4795–4800.
- Fisher, R. A. 1929. Tests of Significance in Harmonic Analysis. *Proceedings of the Royal Society of London, A*, 125, 54-59.
- Gou, L. 1990. Estimating time-varying parameters by the Kalman filter based algorithm: stability and convergences. *IEEE Transactions on Automatic Control*, 35(2), 141-147.
- Graham, C. B., Barnard, H.R., Kavanagh, K.L., Mc Namara, J.P. 2013. Catchment scale controls the temporal connection of transpiration and diel fluctuations in streamflow. *Hydrol. Process.* 27, 2541–2556.
- Gray, G. J., Murray-Smith, D. J., Li, Y., Sharman, K. C., Weinbrenner, T. 1998. Nonlinear model structure identification using genetic programming. *Control Engineering Practice*, 6, 1341-1352.
- Grewal, M., Andrews, A. P. 2015. Kalman Filtering: Theory and Practice Using Matlab, 4th ed. John Wiley & Sons.
- Gu, C. 2013. Smoothing Spline ANOVA Models. New York: Springer-Verlag New York Inc.
- Haber, R., Unbehauen, H. 1990. Structure Identification of Nonlinear Dynamic Systems – A Survey of Input/Output Approaches. *Automatica*, 26(4), 651-677.
- Halliday, S. J., Skeffington, R. A., Wade, A. J., Neal, C., Reynolds, B., Norris, D., Kirchner, J. W. 2013. Upland streamwater nitrate dynamics across decadal to sub-daily timescales: a case study of Plynlimon, Wales. *Biogeosciences*, 10, 8013-8038.
- Harvey, A. 1989. Structural Time Series Models and the Kalman Filter. Cambridge University Press, Cambridge.
- Isaksson, A. 1993. Identification of ARX-models subject to missing data, *Automatic Control, IEEE Transactions*, 38(5), 813-819.
- Jones, T., Chappell, N. A. 2014. Streamflow and hydrogen ion interrelationships identified using data-based mechanistic modelling of high frequency observations through contiguous storms. *Hydrology Research*, 45(6), 868-892
- Jones, T.D., Chappell, N. A., Tych, W. 2014. First dynamic model of dissolved organic carbon derived directly from high frequency observations through contiguous storms. *Environmental Science and Technology*, 48(22), 13289-13297.
- Jonsen, I. D., Flemming, J. M., Myers, R. A. 2005. Robust state space modelling of animal movement data, *Ecology*, 86(11), 2874-2880.

- Kalman, R.E. 1960. A new approach to linear filtering and prediction problems, *Journal of Basic Engineering*, 82(1), 35–45.
- Keery, J., Binley, A., Crook, N., Smith, J. W.N. 2007. Temporal and spatial variability of groundwater-surface water fluxes: Development and application of an analytical method using temperature time series. *Journal of Hydrology*, 336 (1-2), 1-16.
- Li, W., Shah, S.L., 2008, Kalman Filters in non-uniformly sampled multirate systems For FDI and beyond, *Automatica*, 44, 199-208.
- Li, Y., Ryu, D., Western, A. W., Wang, G. J. 2013. Assimilation of stream discharge for flood forecasting: The benefits of accounting for routing time lags. *Water Resources Research*, 49(4), 1887-1990.
- Liu, Z., Higgins, C.W. 2015. Does temperature affect the accuracy of vented pressure transducer in fine-scale water level measurement? *Geosci. Instrum. Methods Data Syst.* 4, 65–73.
- Loheide, S.P., Lundquist, J.D. 2009. Snowmelt-induced diel fluxes through the hyporheic zone. *Water Resources Research*, 45, W07404.
- Love, D., Uhlenbrook, D., Corzo-Perez, G., Twomlow, S., van der Zaag, P. 2010. Rainfall–interception–evaporation–runoff relationships in a semi-arid catchment, northern Limpopo basin, Zimbabwe, *Hydrological Sciences Journal*, 55(5), 687-703.
- Madrid, Y., Zayas, Z. P. 2007. Water sampling: Traditional methods and new approaches in water sampling strategy, *TrAC Trends in Analytical Chemistry*, 26(4), 293-299.
- Mathias, A., Grond, F., Guardans, R., Seese, D., Canela, M., Diebner, H.H. 2004, Algorithms for spectral analysis of irregularly sampled time series, *Journal of Statistical Software*, 11 (2).
- McIntyre, N., Al-Qurashi, A. 2009. Performance of ten rainfall-runoff models applied to an arid catchment in Oman. *Environmental Modelling & Software*, 24(6), 726-738.
- McIntyre, N., Young, P. C., Orellana, B., Marshall, M., Reynolds, B., Wheeler, H. 2011. Identification of nonlinearity in rainfall-flow response using data-based mechanistic modelling. *Water Resources Research*, 47(3), doi:10.1029/2010WR009851.
- Mindham, D. A., Tych, W., Chappell, N. A. 2018. Extended State Dependent Parameter modelling with a Data-Based Mechanistic approach to nonlinear model structure identification. *Environmental Modelling & Software*, 104 (2018), 81-93.
- Moore, M.F., Vasconcelos, J.G., Zech, W.C., Soares, E.P. 2016. A procedure for resolving thermal artifacts in pressure transducers. *Flow Meas. Instrum.* 52, 219-226.
- Mutzner, R., Weijs, S. V., Tarolli, P., Calaf, M., Oldroyd, H. J. 2015. Controls on the diurnal streamflow cycles in two subbasins of an alpine headwater catchment. *Water Resources Research*, 51(5), 3403-3418.
- Hannan, E. J. 1960. Time Series Analysis. Methuen.
- Narendra, K. S., Parthasarathy, K. 1990. Identification and Control of Dynamical Systems Using Neural Networks. *IEEE Transactions on Neural Networks*, 1(1), 4-27.
- Nash, J.E. 1957. The form of the instantaneous unit hydrograph. *Int. Assoc. Hydrol. Sci., General Assembly of Toronto., Publ*, 45, 114–119.

- Ng, C. N., Young, P. C. 1990. Recursive estimation and forecasting of non-stationary time series. *Journal of Forecasting*, 9(2), 173-204.
- Ockenden, M. C., Chappell, N. A. 2011. Identification of the dominant runoff pathways from data-based mechanistic modelling of nested catchments in temperate UK. *Journal of Hydrology*, 402(1-2), 71-79.
- O'Toole, S. J., Butler, R. P., Tinney, C. G., Jones, H. R. A., Marcy, G. W., Carter, B., McCarthy, C., Bailey, J., Penny, A. J., Apps, K., Fischer, D. 2007. New Planets around Three G Dwarfs. *Astrophysics Journal*, 660 (2), 1636-1641.
- Penarrocha, I., Sanchis, R., Romero, J. 2012. State estimator for multisensor systems with irregular sampling and time-varying delays, *International Journal of Systems Science*, 43(8), 1441-1453.
- Raghavan, H., Tangirala, A. K., Gopaluni, R. B., Shah, S. L. 2006. Identification of chemical processes with irregular output sampling. *Control Engineering Practice*, 14(5), 467-480.
- Rajurkar, M. P., Kothiyari, U. C., Chaube, U. C. 2004. Modeling of the daily rainfall-runoff relationship with artificial neural network. *Journal of Hydrology*, 285(1:4), 96-113.
- Ratto, M., Young, P.C., Romanowicz, R., Pappenberger, F., Saltelli, A., Pagano, A. 2007. Uncertainty, sensitivity analysis and the role of data based mechanistic modeling in hydrology, *Hydrol. Earth Syst. Sci*, 11, 1249-1266.
- Rodríguez-Iturbe, I., Gonzales-Sanabria, M. and Camano, G. 1982. On the climatic dependence of the IUH: a rainfall-runoff theory of the Nash model and the geomorphoclimatic theory. *Water Resour. Res.*, 18, 887-903.
- Sadeghi, J. 2006. Modelling and control of non-linear systems using state-dependent parameter (sdp) models and proportional -integral-plus (PIP) control method. Ph.D Thesis, Lancaster University.
- Sakamoto, Y., Ishiguro, M., Kitagawa, G. 1986. Akaike information criterion statistics. Tokyo: KTK Scientific Publishers.
- Seinfeld, J. H., Pandis, S. N. 2016. Atmospheric Chemistry and Physics: From Air Pollution to Climate Change. 3rd Ed. New Jersey: John Wiley & Sons, Inc.
- Shi, Y., Fang, H. 2010. Kalman filter-based identification for systems with randomly missing measurements in a network environment, *International Journal of Control*, 83(3), 538-551.
- Smith, A. C., Wynn, P. M., Barker, P. A, Leng, M. J., Noble, S. R., Tych, W. 2016. North Atlantic forcing of moisture delivery to Europe throughout the Holocene. *Scientific Reports*, 6:24745.
- Sofianopoulou, E., Pless-Mulloli, T., Rushton, S, Diggle, P. J. 2017, Modelling seasonal and spatiotemporal variation: the example of respiratory prescribing, *American Journal of Epidemiology*, 186 (1), 101-108.
- Sugeno, M., Kang, G. T. 1988. Structure Identification of Fuzzy Model. *Fuzzy Sets and Systems*, 28, 15-33.
- Tandeo, P., Ailliot, P., Autret, E. 2011. Linear gaussian state-space model with irregular sampling: application to sea surface temperature, *Stochastic Environmental Research and Risk Assessment*, 25(6), 793-804.

- Taylor, C. J., Pedregal, D. J., Young, P. C., Tych, W. 2007. Environmental time series analysis and forecasting with the CAPTAIN toolbox. *Environmental Modelling & Software*, 22, 797–814.
- Taylor, C.J., Chotai, A., Young, P. C. 2009. Non-linear control by input-output state variable feedback pole assignment. *International Journal of Control*, 82(6) 1029-1044.
- Thompson, D., Cabrol, N., Furlong, M., Hardgrove, C., Low, B., Moersch, J., Wettergreen, D. 2013. Adaptive sensing of time series with application to remote exploration, Robotics and Automation (ICRA), 2013 IEEE International Conference on, May 2013, 3463-3468.
- Trapero, J. R., Kourentzes, N., Martin, A. 2015. Short-term Solar Irradiation forecasting based on Dynamic Harmonic Regression. *Energy*, 84, 289-295.
- Tych, K., Wood, C., Tych, W. 2014. A simple transfer-function-based approach for estimating material parameters from terahertz time-domain data. *IEEE Photonics Journal*, 6(1).
- Tych, W., Pedregal, D.J., Young, P.C., Davies, J. 2002, An unobserved component model for multi-rate forecasting of telephone call demand: the design of a forecasting support system, *International Journal of Forecasting*, 18, 673-695.
- Tych, W., Sadeghi, J., Smith, P. J., Chotai, A., Taylor, C. J. 2012. Multi-state Dependent Parameter Model Identification and Estimation. In: Wang, L., Garnier, H. ed. *System Identification, Environmental Modelling, and Control System Design*. Springer London, 191-210.
- UNEP, 2011. Climate Change and POPs: Predicting the Impact. Report of the UNEP/AMAP Expert group, Geneva, Switzerland.
- VanderKwaak, J. E., Loague, K. 2001. Hydrologic-Response simulations for the R-5 catchment with a comprehensive physics-based model. *Water Resources Research*, 37(4), 999-1013.
- Venier, M., Hung, H., Tych, W., Hites, R. A. 2012. Temporal Trends of Persistent Organic Pollutants: A Comparison of Different Time Series Models. *Environmental Science & Technology*, 46 (7), 3928-3934.
- Whitley, D. 1994. A genetic algorithm tutorial. *Statistics and Computing*, 4(2), 65-85.
- Xue-bo, J., Jing-jing, D., Jia, B. 2012. Target tracking of a linear time invariant system under irregular sampling, *International Journal of Advanced Robotic Systems*, 9, 219-231.
- Yan, B., Lev-Ari, H., Stankovic, A. 2011. Multi-sensor networked estimation in electric power grids, in Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), 2011 4th IEEE International Workshop, 117-120.
- Yandell, B. S. 1989. Spline smoothing and nonparametric regression, *Technometrics*, 31(3), 379-380.
- Young, P. C., Jakeman, A. 1979. Refined instrumental variable methods of recursive time-series analysis Part I. Single input, single output systems. *International Journal of Control*, 29(1), 1-30.
- Young, P. C., Jakeman, A., McMurtrie, R. 1980. An instrumental variable method for model order identification. *Automatica*, 16(3), 281-294.
- Young, P. C., Ng, C. 1989. Variance intervention. *Journal of Forecasting*, 8 (4), 399-416.

- Young, P. C. 1992. Parallel Processes in Hydrology and Water Quality: A Unified Time-Series Approach. *Water and Environmental Journal*, 6(6), 598-612.
- Young, P.C., Pedregal, D.J. 1998. Adaptive Electricity Demand Forecasting Using a Novel Non-Linear, Unobserved Components Model Estimated in the Frequency Domain, *Colloquium on Electricity Demand Forecasting*, London Business School, London, July 6th 1998.
- Young, P. C., Pedregal, D. J., Tych, W. 1999. Dynamic Harmonic Regression. *Journal of Forecasting*, 18, 369-394.
- Young, P. C. 1999a. Data-based mechanistic modelling, generalised sensitivity and dominant model analysis. *Computer Physics Communications*, 117, 113-129.
- Young, P. C. 1999b. Nonstationary time series analysis and forecasting. *Progress in Environmental Science*, 1, 3-48.
- Young, P. C. 2000. Stochastic, dynamic modelling and signal processing: time variable and state dependent parameter estimation. *Nonlinear and nonstationary signal processing*, 74-114.
- Young, P.C., McKenna, P., and Bruun, J. 2001. Identification of nonlinear stochastic systems by state dependent parameter estimation, *Int. J. Control*, 74, (18), 1837–1857.
- Young, P. C. 2003. Top-down and data-based mechanistic modelling of rainfall-flow dynamics at catchment scale. *Hydrol. Process.* 17, 2195-2217.
- Young, P. C., Garnier, H. 2006. Identification and estimation of continuous-time, data-based mechanistic (DBM) models for environmental systems. *Environmental modelling & software*, 21(8), 1055-1072.
- Young, P. C., Castelletti, A., Pianosi, F. 2007a. The data-based mechanistic approach in hydrological modelling. In: Castelletti, A and Sessa R. S. eds. *Topics on System Analysis and Integrated Water Resource Management*. Amsterdam: Elsevier. 27-48.
- Young, P. C. 2011. Recursive Estimation and Time-Series Analysis: An introduction for the student and practitioner. 2nd Ed. Springer.
- Yuz, J., Alfaro, J., Agüero, J., Goodwin, G. 2011. Identification of continuous-time state-space models from non-uniform fast-sampled data, *Control Theory Applications, IET*, 5(7), 842-855.
- Zhang, S., Zhao, L., Delgado-Tellez, R., Bao, H. 2018. A physics-based probabilistic forecasting model for rainfall-induced shallow landslides at regional scale. *Natural Hazards and Earth Sciences*, 18, 969-982.

Appendix

As follows are the three papers the present author contributed too:

- Snell et al., 2019
- Yu et al., 2019
- Magliano et al., 2019

SCIENTIFIC REPORTS



OPEN

Strong and recurring seasonality revealed within stream diatom assemblages

M. A. Snell^{1,2}, P. A. Barker¹, B. W. J. SurrIDGE¹ , C. McW. H. Benskin¹, N. Barber³, S. M. Reaney³ , W. Tych¹ , D. Mindham¹, A. R. G. Large⁴, S. Burke⁵ & P. M. Haygarth¹ 

Improving stream water quality in agricultural landscapes is an ecological priority and a legislative duty for many governments. Ecosystem health can be effectively characterised by organisms sensitive to water quality changes such as diatoms, single-celled algae that are a ubiquitous component of stream benthos. Diatoms respond within daily timescales to variables including light, temperature, nutrient availability and flow conditions that result from weather and land use characteristics. However, little consideration has been given to the ecological dynamics of diatoms through repeated seasonal cycles when assessing trajectories of stream function, even in catchments actively managed to reduce human pressures. Here, six years of monthly diatom samples from three independent streams, each receiving differing levels of diffuse agricultural pollution, reveal robust and repeated seasonal variation. Predicted seasonal changes in climate-related variables and anticipated ecological impacts must be fully captured in future ecological and water quality assessments, if the apparent resistance of stream ecosystems to pollution mitigation measures is to be better understood.

In the context of a changing climate and agricultural intensification, it is recognised that freshwater ecosystems are vulnerable to multiple anthropogenic stressors, including excess sediment and nutrient delivery, whilst also being modulated by climate-dependent flow, water temperature and event-driven transfers from catchments^{1–3}. Low-order streams comprise the headwaters of river networks and drain a significant proportion of the UK's agricultural land. Such streams are central to the functioning of all river networks, forming important corridors linking terrestrial and aquatic ecosystems. These reaches are especially vulnerable to stress as their small channel size relative to catchment area makes them poorly buffered to changes in external climatic, physicochemical and energetic factors^{4–6}. However, low-order streams play an important role in regulating downstream processes and, ultimately, the quality of coastal or other receiving waters^{7,8}.

Ecosystem health is assessed using ecological surveys of stream benthos to identify anthropogenic impacts and trajectories of change^{9–11}. Within Europe, biological monitoring has been formalised through the Water Framework Directive (WFD)¹². This legislation requires an assessment of ecological health, reported as an Ecological Quality Ratio (EQR) which represents the value of an observed biological parameter to that expected under minimally-impaired conditions characteristic of the waterbody type under consideration¹³. The EQR can vary between values indicative of 'high' to 'bad' status, with intermediate values indicative of 'good', 'moderate' or 'poor' status. The EQR is based on specific groups of organisms, including macrophytes, fish, macroinvertebrates and diatoms, and informs subsequent development of restorative measures to reduce significant anthropogenic pressures. Full consideration of organism colonisation, resilience to high-energy events and seasonal controls is needed to produce robust biological datasets that capture ecosystem function and variability. Failure to consider this temporal dynamic could bias inferences drawn from ecological assessment and contribute to poorly-targeted mitigation within catchments, potentially incurring significant economic cost.

The incorporation of weather-related factors within monitoring programmes, such as solar radiation, temperature and rainfall that are descriptive of multi-annual seasonal trends, is critical because future climate scenarios are anticipated to have both direct impacts on stream ecosystems and indirect effects mediated via wider

¹Lancaster Environment Centre, Lancaster University, Lancaster, LA1 4YQ, UK. ²Agri-Food and Biosciences Institute, Newforge Lane, Belfast, BT9 5PQ, UK. ³Department of Geography, Durham University, Durham, DH1 3LE, UK. ⁴School of Geography, Politics and Sociology, Newcastle University, Newcastle upon Tyne, NE1 7RU, UK. ⁵British Geological Survey, Environmental Science Centre, Nicker Hill, Keyworth, Nottingham, NG12 5GG, UK. Correspondence and requests for materials should be addressed to M.A.S. (email: m.snell@lancaster.ac.uk)

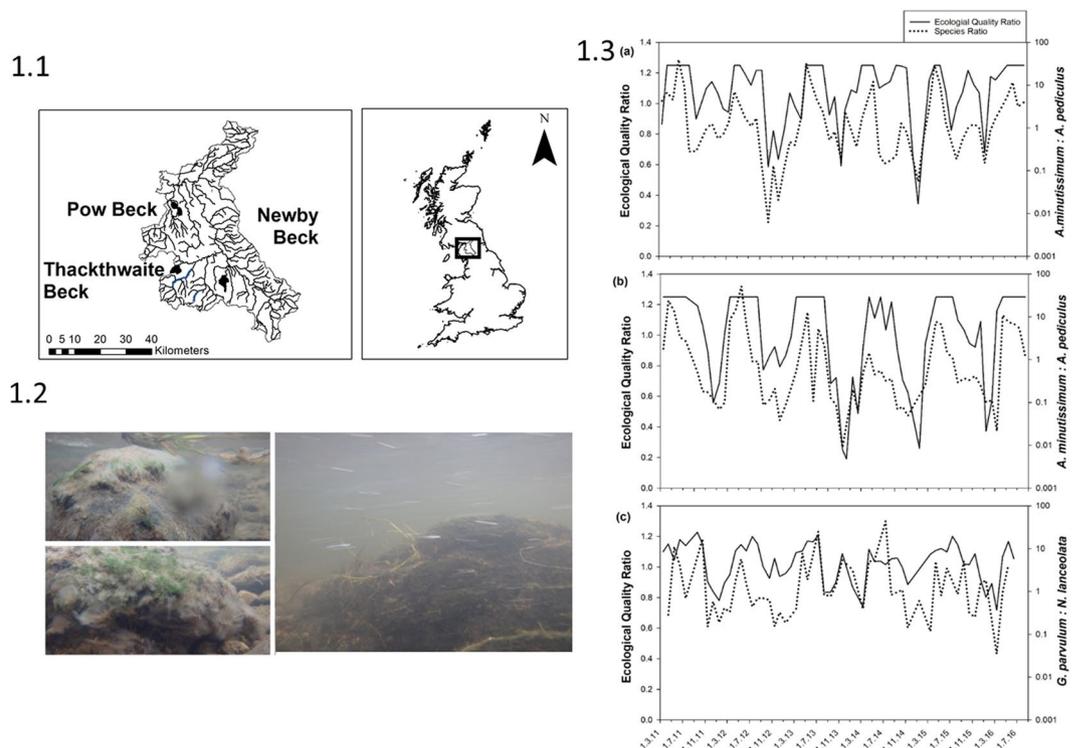


Figure 1. 1.1) Newby, Thackthwaite and Pow Beck catchments of the River Eden, NW England. Crown Copyright/database right 2014. An Ordnance Survey/EDINA supplied service; 1.2) Images of the diatom biofilm assemblages *in situ* at Newby Beck November 2015; 1.3) (a) Monthly Ecological Quality Ratio (EQR: a diatom-based WFD community metric which measures ecological status) for Pow Beck (solid line) and ratio of *Achnantheidium minutissimum* to *Amphora pediculus* (dashed line) (b) Monthly EQR for Newby Beck (solid line) and ratio of *Achnantheidium minutissimum* to *Amphora pediculus* (dashed line) and (c) Monthly EQR for Thackthwaite Beck (solid line) and ratio of *Gomphonema parvulum* and *Navicula lanceolata* (dashed line) from March 2011 to August 2016. EQR status boundaries are: High/good = 0.8; Good/moderate = 0.6; Moderate/poor = 0.4; Poor/bad = 0.2 (DARLEQ 2).

catchment processes. For example, the export of phosphorus (P) from agricultural land to waterbodies is predicted to increase in the future, independent of land use change¹⁴. Moreover, climate scenarios emphasise an acute increase in the extremes of seasonal cycles with warmer, wetter, winters and hotter, drier, summers in parts of NW Europe¹⁵. However, the relative contribution of weather, catchment land use and reach-scale processes on benthic stream ecology has not been rigorously assessed to date via multi-annual, high-resolution *in situ* environmental monitoring programmes. This is particularly true within highly dynamic low-order streams. We hypothesise that robust and recurrent seasonal controls on stream ecosystems could overwhelm, or at least partially mask, efforts to reduce the effects of anthropogenic stress on streams. Here we present six full years of monthly diatom data from three separate intensively instrumented tributaries of the River Eden, NW England. For the first time, these data enable us to evaluate the resilience and recurrence of seasonal patterns in benthic diatom assemblages. These diatom data are evaluated alongside high-resolution environmental monitoring data to determine the relative discrete and combined contributions of land use, the seasonality of precipitation, temperature and light, and in-stream environmental variables that combine to define the niche experienced by benthic diatoms.

Results

Distinct and recurring summer and winter diatom assemblages and EQR values are revealed in the monthly data from all three sub-catchments (Fig. 1.3). The greatest range in ecological status was observed within Newby Beck where EQR values varied between those indicative of 'high' to 'bad' water quality. In contrast, within Thackthwaite Beck which is subject to the lowest intensity of agricultural production, EQR values cycled within the 'high' to 'good' status classes through all seasons. In all three sites, the EQR indicates much better water quality in spring/summer compared to autumn/winter (Fig. 1.3a–c). Strong stability of diatom assemblages within Pow Beck (Fig. 1.3a) and Newby Beck (Fig. 1.3b) is evidenced through the dominance, in terms of relative abundance determined via valve counts, of two key species; *Achnantheidium minutissimum* and *Amphora pediculus*. Periods of improved water quality, as determined by the EQR, are associated with an increase in the ratio of *A. minutissimum* to *A. pediculus*, a pattern that is repeated through all six annual cycles. *A. minutissimum* is a small non-colonial pioneer species (length 15 µm, width 3.5 µm)¹⁶, which colonises and reproduces rapidly¹⁷ and is often abundant under low phosphorus availability¹⁸. *A. pediculus*, while also a small pioneer species adapted to dynamic discharge conditions (length 11.5 µm, width 3 µm)¹⁶, favours environments in which nutrient availability is enhanced¹⁹. At Thackthwaite Beck, the least impacted of the three sites in terms of nutrient enrichment, *A. minutissimum*

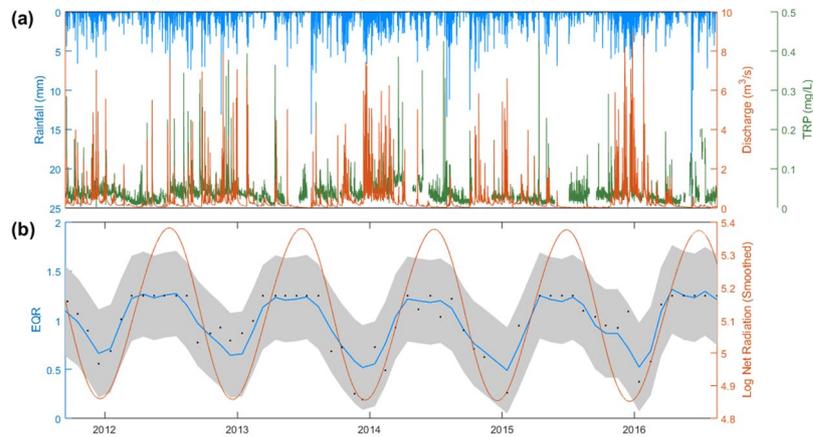


Figure 2. Flow and water chemistry data for the period from 14 September 2011 to 16 August 2016 for Newby Beck: **(a)** Hourly rainfall, discharge and Total Reactive Phosphorus (TRP); **(b)** Monthly diatom EQR and Log Net Radiation. ASDHR model fit with 95% confidence interval explaining over 80% of data variance, and showing the strength of seasonal character of the process. Log Net Radiation Smoothed is the natural log of Net Radiation prior to using DHR—demonstrating Net Radiation seasonality + smoothing.

dominated throughout each year and increased in relative abundance during spring and summer. At this site *A. minutissimum* was almost seven times more abundant than the next two most abundant species; *Gomphonema parvulum* and *Navicula lanceolata*. Nevertheless, these latter two species also display repeated seasonal patterns throughout the time series, with the motile guild species *N. lanceolata* increasing in dominance under autumn/winter conditions (Fig. 1c) suggesting that higher current velocities and nutrient concentrations could be behind this temporal pattern. The dominance of pioneer species across our dataset suggests that these diatom assemblages are kept in a dynamic equilibrium state throughout the annual cycle. Further functional responses are found in Thackthwaite Beck, where repeated switching in dominance between *G. parvulum* and *N. lanceolata* represents the temporal interaction between a high-profile, larger species characteristic of nutrient enrichment (*G. parvulum*), and species indicative of the motile guild that are better adapted to high flow-low nutrient availability conditions (*N. lanceolata*). Although different species composition was found between the three independent streams, the same co-aligned seasonal pattern of species change was observed.

Environmental controls on species composition were investigated using a 5-year subset of data from Newby Beck, comprising the most complete flow and water chemistry data (Fig. 2a). These high frequency *in situ* measurements reveal a strong relationship between weather-related variables (including rainfall, radiation and temperature), discharge conditions, and nutrient and sediment concentrations^{14,20}. The cyclical nature of the monthly diatom data is strongly supported by ASDHR results (Arbitrary Sampled Dynamic Harmonic Regression; Fig. 2b, showing the data against the model with its 95% confidence band). This method allows estimation of periodic or seasonal signal components in irregularly-sampled time-series data, providing both an estimate of the periodic components themselves as well as their uncertainty estimates (see Methods). The model explained 82% of data variance ($R^2 = 0.82$) in the EQR metric and the ANOVA F-test was significant at $p = 5.2 \cdot 10^{-10}$.

Diatom assemblages typically develop over 1–3 weeks²⁰, so the mean value of environmental variables (water temperature, net radiation, nitrate-N, dissolved oxygen, pH, rainfall, turbidity, discharge, total reactive phosphorus and conductivity) for the 21 days prior to diatom sampling were used to explore species-environment interactions. Spring and summer diatom samples scored highly against Principal Component Analysis (PCA) axis 1 with an eigenvalue of 0.2, demonstrating the primary importance of light. These samples were characterised by the presence of *Gomphonema olivaceum* and *A. minutissimum*. *G. olivaceum* is a high-profile guild species indicating resource-unlimited but disturbance-stress conditions, while *A. minutissimum*, the dominant species in terms of relative abundance as determined by cell counts, is a low-profile guild species that is comparatively resistant to physical disturbance. Under spring/summer conditions, water temperature and net radiation, dissolved oxygen and nitrogen (N) availability are key drivers of the diatom assemblage composition. Variability in spring/summer conditions was indicated through low-profile (*Meridion circulare*, *Rhoicosphenia abbreviata*), high-profile (*Encyonema minutum*) and motile (*Surirella crumena*) guild species. In contrast, the PCA demonstrates that winter and autumn diatom samples are associated with in-stream conditions, primarily turbidity and discharge, precipitation-related climate variations and phosphorus resource availability, which all plot in the lower left quadrant of Fig. 3. The percentage of *A. minutissimum* is inversely related to *A. pediculus*, the dominant pioneer species under rainfall-driven discharge-phosphorus winter conditions¹⁴. The presence of low-profile (*Cocconeis euglypta*) and motile (*Navicula cryptotenella* and *Caloneis bacillum*) guild species along with *A. pediculus* indicates the importance of variables occurring under winter conditions. Seasonality and nutrient resource availability are important in determining contrasting summer and winter ecological states, with in-stream variables of conductivity, water temperature and pH switching in relative importance as drivers of diatom assemblage composition between seasons.

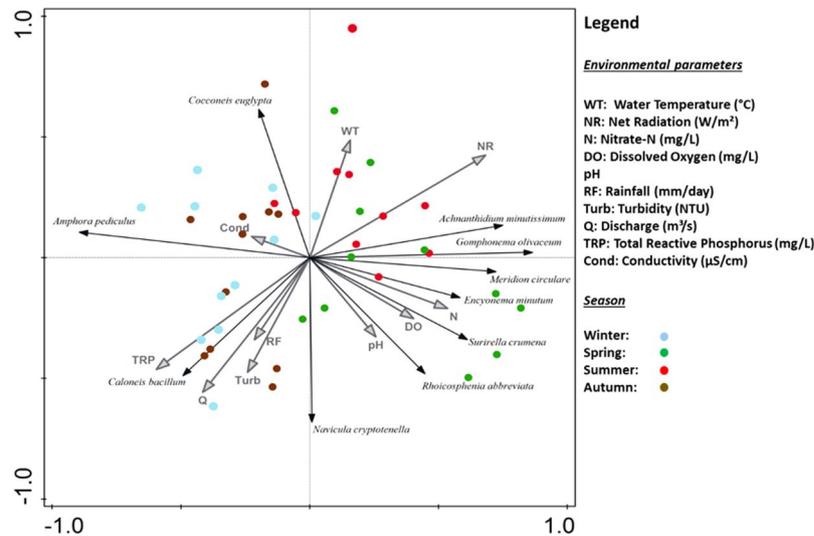


Figure 3. Principal Component Analysis with monthly diatom species scores limited to 10 by fit. Environmental variables averaged over an antecedent window of 21 days prior to biological sample collection. Explained cumulative variation is 34.3% with an adjusted explained variation of 26.1%.

Discussion

These three independent and contrasting sub-catchments of the River Eden, characterised by different levels of agricultural intensity, demonstrate robust and recurring seasonal patterns in stream benthic diatom assemblages over a six-year period. The resilience of the assemblage at this seasonal scale is greater than shown previously^{21,22} offering species-environment insights for catchment management programmes. Moreover, intensive instrumentation in these catchments has allowed antecedent environmental conditions occurring during bio-film formation to be captured²⁰. Our study demonstrates the crucial temporal perspective that is required to understand ecosystem variability, function and response to seasonality at inter- and intra-annual scales. The long-term diatom time-series displays a dynamic community of pioneer species, implying a constant resetting of the benthic community constrained within characteristic seasonal equilibria. The diatom taxa reflect the impact of both nutrient and stream conditions with a seasonal signature emerging through changes in relative species abundance. Changes in these underlying drivers, both anthropogenic (e.g. discharge associated changes in phosphorus concentration) and weather-related (water temperature, net radiation, and rainfall), combine to shift the ecosystem into clearly-defined alternative states under winter and summer conditions. Community variation left unexplained in the analysis can potentially be attributed to variables not measured in this study, including the influence of grazing²³. To increase our ability to predict ecological responses to changes in regional weather and climate patterns, it is necessary for future studies to integrate long-term, empirical survey data with models and manipulative experiments.

Multiple seasonal trends at regional scale provide a dynamic baseline against which land-use decisions are made, and this baseline itself is expected to change in the future. For example, projections for the study catchment region suggest a rise in mean annual temperature between 2.5 °C to 3.3 °C by the 2070 s under a 'medium' emissions scenario. The anticipated outcome is that more intense seasonal changes will be observed in rainfall, with drier summer and wetter winter conditions expected¹⁵. Across seasonal and multi-annual scales, direct weather variables (temperature, light) are the dominant factors dictating diatom assemblage, with seasonally-delivered nutrients providing further temporally-mediated regulation of diatom assemblages. An important consequence is that future climate changes could undermine catchment remediation programmes whose 'success' is typically measured over the short-term. Furthermore, current ecological monitoring frameworks are based largely on detecting responses to organic pollution under spring and autumn conditions. These programmes give little consideration to seasonal characteristics of diatom assemblages and will not capture this shifting seasonal baseline, nor how the health of a stream ecosystem over a full annual cycle may respond to land management strategies. Given climate predictions and suggested requirements for agricultural land-use change to mitigate phosphorus export²⁴, current assessments of 'ecological status' will increasingly become obsolete over time, as the baseline drifts to new warmer and wetter conditions. The different components of seasonal change (e.g. temperature, hydrology and atmospheric composition) not only affect multiple levels of biological organisation, but may also interact with other facets of anthropogenic stress to which rivers are exposed²⁵⁻²⁷. It is critical that future research addresses these potentially important synergies underpinned by a clear understanding of climate forcing and trajectories. Given the evidence presented here showing a strong and recurring seasonality in benthic diatoms in stream ecosystems, understanding future ecological variability is important to inform current and future conservation strategies. Capturing this dynamic will ensure that ecological function is maintained across all seasons, as well as allowing for better predictive models of how freshwater systems may respond to management under increasingly contrasting seasonal conditions.

Methods

Description of study sites. This study presents data for three low-order streams (Fig. 1 (1.1)); Newby Beck (Strahler second order: OSGB NY-59957,21249) which drains 12.5 km² of the headwaters of the Morland catchment, Pow Beck (Strahler third order: OSBG NY-38685,50074) which drains an area of 10.5 km² and Thackthwaite Beck (Strahler second order: OSBG NY-41190-25320) which drains 10.2 km² of the headwaters of the Dacre Beck catchment, within the larger River Eden catchment, NW England. The three sub-catchments form part of the River Eden UK Demonstration Test Catchment (DTC) programme, a catchment-scale research platform funded by the Department for the Environment, Food and Rural Affairs (2009–2018). Differences in bedrock, superficial geology, land use, rainfall and water quality characteristics are found among the three sub-catchments²⁸. Briefly, Pow Beck, with sandstone bedrock, supports intensive agricultural land use practice giving rise to higher diffuse N and P pressures within the catchment. Thackthwaite Beck, lying on hard rock of volcanic andesite sheets, is a flashy, oligotrophic, upland catchment. Newby Beck drains exposed steeply dipping, fractured carboniferous limestone, shale and sandstone units. Having improved grassland, with rough grazing and arable land as the predominant land use, Newby Beck is exposed to intermediate diffuse pollution pressures in comparison to the Pow Beck and Thackthwaite Beck catchments. All three catchments have a superficial layer of post-glacial till, separating much of the surface hydrological response from the direct influence of the bedrock geology.

Diatom preparation and identification. Diatoms were collected mid-monthly from March 2011 until August 2016 for all three streams. On each sampling occasion, five representative submerged cobbles were selected from typical riffle areas. Cobbles were scraped with a hard bristle brush and the diatom suspension collected. Samples were processed using 30% hydrogen peroxide and permanent slides were prepared using Naphrax²⁹. 300 valves were identified^{30–35} by the same analyst accredited within the UK Diatom Quality Assurance Scheme³⁶. Calculation of EQR was undertaken using DARLEQ II software. More details are given in Supplementary Information.

In situ environmental monitoring. Automatic weather stations installed in each catchment measured precipitation, air temperature, relative humidity and net solar radiation at 15 minute intervals²⁸. A weather station centrally located between the three catchments produced total sunshine hours (Penrith Weather Station, 2017). *In situ* environmental and ecological monitoring occurred at catchment outflows (Fig. 1 (1.1)). Bank-side monitoring stations adjacent to biological sampling areas provided *in situ* water quality measurements at a minimum resolution of 15 minutes. Hach Lange nutrient analysers provided total reactive phosphorus (TRP) and nitrate (NO₃-N) concentration data. Dissolved oxygen (electrochemical sensor, 0–500% ± 2%), pH (0–14 units ± 0.2 units), temperature (−5 to + 50 °C ± 0.15 °C), turbidity (0–1000 NTU, ± 2% of reading or 0.3 NTU) and conductivity (0–100 μS cm^{−1} ± 0.5%) were analysed every 15 minutes using a YSI 6600 sonde. Discharge calculations were made by applying a stage-discharge relationship to 15 minute water level readings recorded by a pressure transducer. The stage-discharge relationship was developed through the collection of manual current metering measurements, and extrapolated beyond the gauged range using standard assumptions for the stage-discharge relationship and the hydrological water balance³⁷. The *in situ* analysis was quality assured using monthly laboratory analysed samples to ensure accuracy in the measurements.

Data analysis. PCA was used to determine the importance of environmental controls on the benthic diatom assemblages. PCA relates species composition to measured environmental gradients by reducing the dimensionality of complex, multivariate data, finding a small number of linear combinations of variables which are visualized using an ordination diagram³⁸. All water chemistry data used within this analysis was averaged over the preceding 21 days to account for diatom community colonisation processes²⁰. Values of turbidity, rainfall and discharge along with the species matrix were Log-transformed (Ax + B). All ordinations were performed using CANOCO version 5.1 software³⁹.

Time-varying seasonal analysis (Mindham and Tych, 2018 *in submission*) of the data series was performed using ASDHR⁴⁰ and conducted within Matlab. The approach belongs to the Unobserved Components Modelling group of methods applying spectral decomposition to the observation series y_t isolating and estimating slow trends T_t as well as cyclic and seasonal components C_t and S_t respectively and the irregular component e_t . This is executed using a stochastic state-space approach with Kalman Filtering and Fixed Interval Smoothing⁴⁰ but placed within an irregular sampling framework. The arbitrary sampling, a novel approach in this context⁴¹, allows the Unobserved Components model to be estimated using irregularly sampled data.

$$y_t = T_t + C_t + S_t + e_t$$

In this case, no cyclic component was estimated and the seasonal component is expressed as:

$$S_t = \sum_{i=1}^{R_s} \{a_{i,t} \cos(\omega_i t) + b_{i,t} \sin(\omega_i t)\}$$

where $a_{i,t}$ and $b_{i,t}$ are stochastic Time-Varying Parameters (TVP) and ω_i are the fundamental and harmonic frequencies associated with seasonality in the series ($i = 1, 2, \dots, R_s$). The covariance parameters of the KF/FIS estimators are defined by the time scale of the estimated processes.

Data Availability

Data can be obtained by contacting the corresponding author.

References

- Ormerod, S. J. *et al.* Multiple stressors in freshwater ecosystems. *Freshwater Biology* **55**, 1–4 (2010).
- Hering, D. *et al.* Managing aquatic ecosystems and water resources under multiple stress—An introduction to the MARS project. *Science of the total environment* **503**, 10–21 (2015).
- Jackson, M. C. *et al.* Net effects of multiple stressors in freshwater ecosystems: a meta-analysis. *Global Change Biology* **22**(1), 180–189 (2016).
- Benda, L. *et al.* The network dynamics hypothesis: how channel networks structure riverine habitats. *AIBS Bulletin* **54**(5), 413–427 (2004).
- Lowe, W. H. & Likens, G. E. Moving headwater streams to the head of the class. *AIBS Bulletin* **55**(3), 196–197 (2005).
- Jones, J. B. & Stanley, E. *Stream Ecosystems in a Changing Environment*. Elsevier, London (2016).
- Alexander, R. B. *et al.* The role of headwater streams in downstream water quality. *JAWRA Journal of the American Water Resources Association* **43**(1), 41–59 (2007).
- Wipfli, M. S., Richardson, J. S. & Naiman, R. J. Ecological linkages between headwaters and downstream ecosystems: Transport of organic matter, invertebrates, and wood down headwater channels. *Journal of the American Water Resources Association* **43**(1), 72–85 (2007).
- Kelly, M. & Whitton, B. A. The trophic diatom index: a new index for monitoring eutrophication in rivers. *Journal of Applied Phycology* **7**(4), 433–444 (1995).
- Kelly, M. *et al.* Assessment of ecological status in UK rivers using diatoms. *Freshwater Biology* **53**(2), 403–422 (2008).
- Stevenson, R., Pan, Y. & van Dam, H. Assessing environmental conditions in rivers and streams with diatoms. *The Diatoms: Applications for the Environmental and Earth Sciences*, 2nd ed. Cambridge University Press, Cambridge (2010).
- European Commission. Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for community action in the field of water policy. Official Journal of the European Commission (2000). [cited 2018 22/03]; Available from: <http://jncc.defra.gov.uk/>.
- Anonymous, Monitoring under the Water Framework Directive. Common Implementation Strategy for the Water Framework Directive Guidance Document No 7, 153 (2003).
- Ockenden, M. *et al.* Changing climate and nutrient transfers: Evidence from high temporal resolution concentration-flow dynamics in headwater catchments. *Science of the Total Environment* **548**, 325–339 (2016).
- UKCP09. *Climate change projections*. 2014 [cited 2018 15/05]; Available from <http://ukclimateprojections.metoffice.gov.uk/23189>.
- Rimet, F. & Bouchez, A. Life-forms, cell-sizes and ecological guilds of diatoms in European rivers. *Knowledge and management of Aquatic Ecosystems* **406**, 01 (2012).
- McCormick, P. V. Resource competition and species coexistence in freshwater benthic algal assemblages, in *Algal ecology freshwater benthic ecosystems*, R.J. Stevenson, Bothwell, M. L. and Lowe, R. L., Editor. Academic Press: London (1996).
- Stelzer, R. S. & Lambert, G. A. Effects of N:P ratio and total nutrient concentration on stream periphyton community structure, biomass, and elemental composition. *Limnology and Oceanography* **46**(2), 356–367 (2001).
- Gottschalk, S. & Kahlert, M. Shifts in taxonomical and guild composition of littoral diatom assemblages along environmental gradients. *Hydrobiologia* **694**(1), 41–56 (2012).
- Snell, M. A. Headwater Stream Biofilm Structure and Function at High Resolution Spatial-Temporal Scales, PhD, Lancaster University Lancaster. Lancaster (2014).
- Soininen, J. & Eloranta, P. Seasonal persistence and stability of diatom communities in rivers: are there habitat specific differences? *European Journal of Phycology* **39**(2), 153–160 (2004).
- Wu, N. *et al.* Importance of sampling frequency when collecting diatoms. *Scientific reports* **6**, 36950 (2016).
- Lange, K. *et al.* Light, nutrients and grazing interact to determine stream diatom community composition and functional group structure. *Freshwater Biology* **56**(2), 264–278 (2011).
- Ockenden, M. C. *et al.* Major agricultural changes required to mitigate phosphorus losses under climate change. *Nature Communications* 1–9 (2017).
- Woodward, G., Perkins, D. M. & Brown, L. E. Climate change and freshwater ecosystems: impacts across multiple levels of organization. *Philosophical Transactions of the Royal Society B: Biological Sciences* **365**(1549), 2093–2106 (2010).
- Segner, H., Schmitt-Jansen, M. & Sabater, S. Assessing the impact of multiple stressors on aquatic biota: the receptor's side matters. *Environmental Science and Technology* **48**(14), 7690–7696 (2014).
- Pizarro, H. *et al.* Impact of multiple anthropogenic stressors on freshwater: how do glyphosate and the invasive mussel *Limnoperna fortunei* affect microbial communities and water quality? *Ecotoxicology* **25**(1), 56–68 (2016).
- Owen, G. J. *et al.* Monitoring agricultural diffuse pollution through a dense monitoring network in the River Eden Demonstration Test Catchment, Cumbria, UK. *Area* **44**(4), 443–453 (2012).
- WFD-UKTAG. River Assessment Method Macrophytes and Phytobenthos. Scotland. Water Framework Directive – United Kingdom Advisory Group (2014).
- Krammer, K. & Lange-Bertalot, H. *Süßwasserflora von Mitteleuropa, Bd 2/4. Bacillariophyceae. 4. Teil: Achnanthaceae Kritische Ergänzungen zu Navicula (Lineolatae) und Gomphonema*. Gustav Fischer Verlag: Stuttgart (1991).
- Krammer, K. & Lange-Bertalot, H. *Süßwasserflora von Mitteleuropa. Bacillariophyceae Teil iv: Achnanthaceae*. Gustav Fischer Verlag: Stuttgart (1991).
- Krammer, K. & Lange-Bertalot, H. *Süßwasserflora von Mitteleuropa. Bacillariophyceae, Band 2/3, 3. Teil: Centrales, Fragillariaceae, Eunotiaceae. 1–576*. Gustav Fischer Verlag: Stuttgart (1991).
- Krammer, K. & Lange-Bertalot, H. *Süßwasserflora von Mitteleuropa. Bacillariophyceae. 1. Teil: Naviculaceae, vol 2/1*. Gustav Fischer Verlag: Stuttgart (1986).
- Kelly, M. Identification of common benthic diatoms in rivers. *Field Studies Council* **9**, 583–700 (2000).
- Hofmann, G., Werum, M. & Lange-Bertalot, H. Diatomeen im Süßwasser-Benthos von Mitteleuropa: Bestimmungsflora Kieselalgen für die ökologische Praxis; über 700 der häufigsten Arten und ihrer Ökologie. Gantner (2011).
- Kelly, M. Building capacity for ecological assessment using diatoms in UK rivers. *Journal of Ecology and Environment* **36**(1), 89–94 (2013).
- Ewen, J. *et al.* Multiscale experimentation, monitoring and analysis of long-term land use changes and flood risk-SC060092: Final Science Report. Newcastle University, Newcastle upon Tyne, UK (2010).
- Šmilauer, P. & Lepš, J. *Multivariate analysis of ecological data using CANOCO 5*. Cambridge University Press. United Kingdom (2014).
- Braak, C. J. F. & Šmilauer, P. *CANOCO reference manual and user's guide: software for ordination (version 5.0)*. 2012: Biometris.
- Young, P. C., Pedregal, D. J. & Tych, W. Dynamic harmonic regression. *Journal of forecasting* **18**(6), 369–394 (1999).
- Mindham, D. A., Tych, W. & Chappell, N. A. Extended State Dependent Parameter modelling with a Data-Based Mechanistic approach to nonlinear model structure identification. *Environmental Modelling & Software* **104**, 81–93 (2018).

Acknowledgements

We thank our funding project: Defra FFG0909 Design and implementation of monitoring approach at catchment scale, and development of the catchment conceptual (WQ0210). We would also like to thank other Eden Demonstration Test Catchment colleagues who have contributed to this research platform and the wider Demonstration Test Catchment team.

Author Contributions

M.A. Snell: Collection, processing and analyses of diatom samples with supporting environmental datasets available within the Eden Demonstration Test Catchment platform. Conceived concept for paper and wrote the manuscript with support from P.A. Barker. Discussed the results with all co-authors and edited the final manuscript. P.A. Barker: Contribution to the establishment, development and management of the Demonstration Test Catchment monitoring platform. Conceived concept for paper, contribution to formulation and writing of the manuscript. Discussed the results with M.A. Snell and contributed to editing of the final manuscript. Supervised the findings of this work. B.W.J. Surridge: Contributed to the design, implementation and supervision of diatom work and its findings. Contribution to the Eden Demonstration Test Catchment monitoring platform. Discussed the results and contributed to the final manuscript. C.McW.H. Benskin: Contribution to the collection of diatom and environmental datasets within the Eden Demonstration Test Catchment. Discussed the results and contributed to the editing of the final manuscript. N. Barber: Contribution to data collection and QC/QA analysis of environmental datasets within the Eden Demonstration Test Catchment. Discussed the results and contributed to the final manuscript. S.M. Reaney: Contribution to the establishment and development of the Eden Demonstration Test Catchment monitoring platform. Discussed the results and contributed to the final manuscript. W. Tych: Contribution to data analysis, specifically Arbitrary Sampled Dynamic Harmonic Regression modelling and discussed the results. Contributed to the final manuscript. D. Mindham: Contribution to data analysis, specifically Arbitrary Sampled Dynamic Harmonic Regression modelling and discussion of results. Contributed to the final manuscript via assistance with modelling graphics and editing. A.R.G. Large: Contributed to the design, implementation and supervision of diatom work and its findings. Discussed the results and contributed to the final manuscript and editorial. S. Burke: Contribution to the establishment and development of the Eden Demonstration Test Catchment monitoring platform. Discussed the results and contributed to the final manuscript. P.M. Haygarth: Contribution to the establishment, development and management of the Eden Demonstration Test Catchment monitoring platform. Discussed the results and overarching concept. Contributed to the final manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-37831-w>.

Competing Interests: I declare that there are no financial or non-financial interests that could directly undermine, or be perceived to undermine the objectivity, integrity and value of this publication, through a potential influence on the judgements and actions of authors with regard to objective data presentation, analysis and interpretation.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019

Polycyclic Aromatic Hydrocarbons Not Declining in Arctic Air Despite Global Emission Reduction

Yong Yu,^{*,†,‡} Athanasios Katsoyiannis,[§] Pernilla Bohlin-Nizzetto,[§] Eva Brorström-Lundén,^{||} Jianmin Ma,[⊥] Yuan Zhao,[⊥] Zhiyong Wu,[†] Włodzimierz Tych,[#] David Mindham,[#] Ed Sverko,[▽] Enzo Barresi,[○] Helena Dryfhout-Clark,[†] Phil Fellin,[◆] and Hayley Hung^{*,†,⊕}

[†]Air Quality Processes Research Section, Environment and Climate Change Canada, Toronto, M3H 5T4, Canada

[‡]Key Laboratory of Wetland Ecology and Environment, Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Changchun 130102, China

[§]Norwegian Institute for Air Research (NILU), NO-2007 Kjeller, Norway

^{||}Swedish Environmental Research Institute (IVL), 411 33 Göteborg, Sweden

[⊥]College of Urban and Environmental Sciences, Peking University, Beijing 100871, China

[#]Lancaster Environment Centre, Lancaster University, Lancaster, LA1 4YQ, United Kingdom

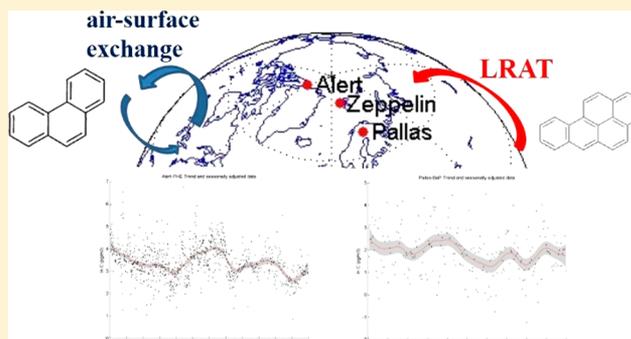
[▽]State Key Laboratory of Urban Water Resource and Environment, Harbin Institute of Technology, Harbin 150090, China

[○]National Laboratory for Environmental Testing (NLET), Canada Centre for Inland Waters, Environment and Climate Change Canada, Burlington, L7R 4A6, Canada

[◆]AirZone One Ltd., Mississauga, L4Z 1X1, Canada

Supporting Information

ABSTRACT: Two decades of atmospheric measurements of polycyclic aromatic hydrocarbons (PAHs) were conducted at three Arctic sites, i.e., Alert, Canada; Zeppelin, Svalbard; and Pallas, Finland. PAH concentrations decrease with increasing latitude in the order of Pallas > Zeppelin > Alert. Forest fire was identified as an important contributing source. Three representative PAHs, phenanthrene (PHE), pyrene (PYR), and benzo[*a*]pyrene (BaP) were selected for the assessment of their long-term trends. Significant decline of these PAHs was not observed contradicting the expected decline due to PAH emission reductions. A global 3-D transport model was employed to simulate the concentrations of these three PAHs at the three sites. The model predicted that warming in the Arctic would cause the air concentrations of PHE and PYR to increase in the Arctic atmosphere, while that of BaP, which tends to be particle-bound, is less affected by temperature. The expected decline due to the reduction of global PAH emissions is offset by the increment of volatilization caused by warming. This work shows that this phenomenon may affect the environmental occurrence of other anthropogenic substances, such as more volatile flame retardants and pesticides.



INTRODUCTION

Polycyclic aromatic hydrocarbons (PAHs) are an important class of organic pollutants released into the environment primarily through incomplete combustion of fossil fuels and biomass. They are of great public concern due to their toxicity and potential carcinogenicity. PAHs can undergo long-range atmospheric transport (LRAT) to remote locations and are listed for regulation under the United Nations Economic Commission for Europe (UNECE) Aarhus Protocol on Persistent Organic Pollutants (POPs) in the Convention on Long Range Transboundary Air Pollution (CLRTAP).^{1–3} Due to their tendency to travel over long distances, PAHs are regularly detected in very remote areas, far away from primary

sources. As such, PAHs are ubiquitous in the Arctic environment. Levels of most regulated POPs are declining over time in the Arctic environment, reflecting the effectiveness of national and international emission control initiatives such as CLRTAP and the Stockholm Convention on POPs.^{4,5} However, temporal trends of PAHs measured in various Arctic media did not show consistent declining trends,^{6–8} despite the estimated reduction in global PAH emissions from 592000 to

Received: September 21, 2018

Revised: February 7, 2019

Accepted: February 12, 2019

Published: February 12, 2019

499000 tons between 1995 and 2008.⁹ Studies have shown that PAHs measured in Arctic marine waters and sediments mainly originate from natural underwater hydrocarbon seeps,^{10–12} while those measured in air originate from atmospherically derived sources, making the air the most suitable medium for PAH monitoring to verify the outcome of regulations on PAH emissions. In addition, their occurrence in the Arctic atmosphere is of particular importance because what is detected is the combined result of LRAT, continuous regional emissions (heating, industrial activities), and temporary local emissions due to commercial and other activities (maritime traffic, coal mining etc.). In a warming Arctic, PAH emissions due to increased human activities within the Arctic may also increase.

PAHs are characterized as “seasonal contaminants” as space heating is one of the most important sources, suggesting that their concentrations in winter are much higher than their respective levels during the warmer months. In addition, some PAHs are subject to photodegradation. With the extended winter darkness in the Arctic, PAHs can behave differently in the Arctic atmosphere than regions that experience a regular diurnal cycle of daylight.¹³ For some of the above reasons, PAHs have been identified as emerging contaminants in the Arctic.⁶

The Arctic Monitoring and Assessment Programme (AMAP) investigated the occurrence of POPs in the Arctic atmosphere for three decades.^{4,14} Here, we use long-term measurement data sets (~20 years) of PAHs in Arctic air collected at three AMAP sites, i.e., Alert, Canada; Zeppelin, Svalbard; and Pallas, Finland, to assess sources, transport pathways, and cycling of PAHs in Arctic air, and from their temporal trends, try to evaluate the outcome of reduction in PAH global emissions.

The Arctic is recognized as a key area for studying climate effect on pollutants due to its sensitivity to climate change. Climate change-driven processes, e.g., temperature, precipitations, and winds, may affect the LRAT and temporal trends of POPs in the Arctic atmosphere.^{15,16} One approach to test the relationship between climate change and the variation of POPs is by comparing modeling results and long-term monitoring data.^{17,18} These large data sets may help to provide greater insight on the influence of climate change on temporal trends of PAHs.

METHODS

Sampling. The locations of the sampling sites are shown in Figure S1. A super high-volume air sampler (SHV) was employed to collect air samples at the Alert Global Atmosphere Watch Observatory, Nunavut, Canada (82.50°N, 62.33°W, 200 m a.s.l.) from 1992 to 2015. One 20 cm glass fiber filter (GFF) and two polyurethane foam (PUF) plugs (20 cm diameter, 4 cm thickness) were used to trap the particle and gas phase compounds separately in approximately 13500 m³ of air over a 7-day sampling period. In the Norwegian-operated Zeppelin Observatory, Ny-Ålesund, Svalbard (78.91° N, 11.88° E, 478 m a.s.l.), a high-volume air sampler (HV) was employed to sample from 1994 to 2015. Two-day integrated weekly samples with a total sample volume of about 1200 m³ of air were collected every week with one 11 cm GFF and two PUFs (11 cm in diameter and 5 cm in height). In Pallas, Finland (68.0°N, 24.24°E, 340 m a.s.l.), 7-day integrated samples (1996–2008) with a volume of ~4000 m³ and 30-day integrated samples (composite of 4 weekly samples) (2009–

2015) with a volume of ~16000 m³ air were collected with a HV. One 14 cm GFF and three PUFs (11 cm diameter, 4.5 cm thickness) were used to collect the particle and gaseous substances. The sampling frequency and extraction strategy may vary in different years (Table S1), subject to availability of funds. The details about the analytical procedure and breakthrough can be found in the Supporting Information.

Data and Analysis. PAH concentrations of Alert are entered into the Research Data Management and Quality Control (RDMQ) system which is a software system written in SAS for data management that includes an extensive flagging system to highlight anomalies and to perform data quality checks. Data from all three sites were reported to the EBAS database (ebas.nilu.no). To ensure data quality, two large-scale interlaboratory comparison studies were conducted for the analysis of trace organic chemicals in Arctic air and an air sample was shared among all participating laboratories.^{19,20}

A Positive Matrix Factorization (PMF, version 5.0) model recommended by the USEPA was used to quantify the contribution of various emission sources to PAHs in the Arctic. For all PAH congeners, the specificity of unknown sources is excluded, and the congeners with weak signal-to-noise ratio and poor linear relationship were removed based on the PMF guideline.²¹ PMF analysis was repeatedly run 20 times with 2–6 factors and then the results at the lowest Q value were output for analysis. The details about the PMF model can be found in the Supporting Information.

Venier et al.²² compared four time series models: a modified Clausius–Clapeyron equation, a multiple linear regression, digital filtration, and dynamic harmonic regression (DHR). Of these, the DHR model exhibited the best performance in fitting the data in the long term time trends. Moreover, DHR can handle extreme values and time series breaks,²³ e.g., the concentrations of PAHs in summer are much lower than those in winter. Therefore, DHR was applied to derive time trends here. A detailed description of the DHR can be found in the Supporting Information.

Model Description. A Canadian Model for Environmental Transport of Organochlorine Pesticides (CanMETOP) was used to simulate the transport, deposition, and degradation of air pollutants from the surface up to 11 km, which successfully simulated the transport of PAHs²⁴ and pesticides.^{25,26} The model is driven by assimilated Canadian Meteorological Centre (CMC) meteorological data, using a time step length of 30 min, spatial resolution of 1° latitude × 1° longitude, and 14 vertical levels to simulate daily mean concentrations of selected PAHs from 1992 to 2015. Estimated global atmospheric emissions of PAHs for 2004 were employed for the simulations.²⁷ By using an emission inventory (from 2004) in the middle of the time period between 1992 and 2015, we can eliminate any bias, which may be introduced by a changing emission that is unknown/uncertain for an individual PAH, and focus on the influence of climate change. The details about the CanMETOP, including air–surface flux and sensitivity analysis, can be found in the Supporting Information.

RESULTS AND DISCUSSION

Air Concentrations of PAHs. Details and data availability of PAHs at each site are provided in Table S2. The concentration ranges and detection frequencies of PAHs at the three Arctic sites are summarized in Table S3. In general, concentrations of most PAHs decreased from the most southerly site of Pallas to the most northerly site of Alert in

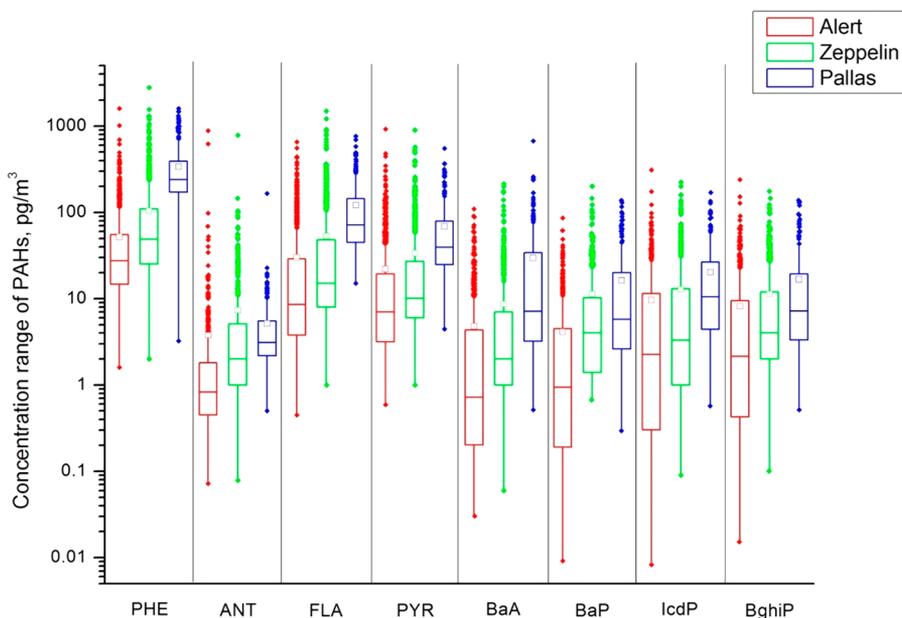


Figure 1. Box-and-whisker plots of 8 PAHs at the three sites during the sampling periods. The boxes represent the 25th and 75th percentiles of the data. The lines in the boxes and square symbols represent the median and the mean, respectively. All the outliers beyond the whiskers are shown individually. Nondetects are not reported in this figure.

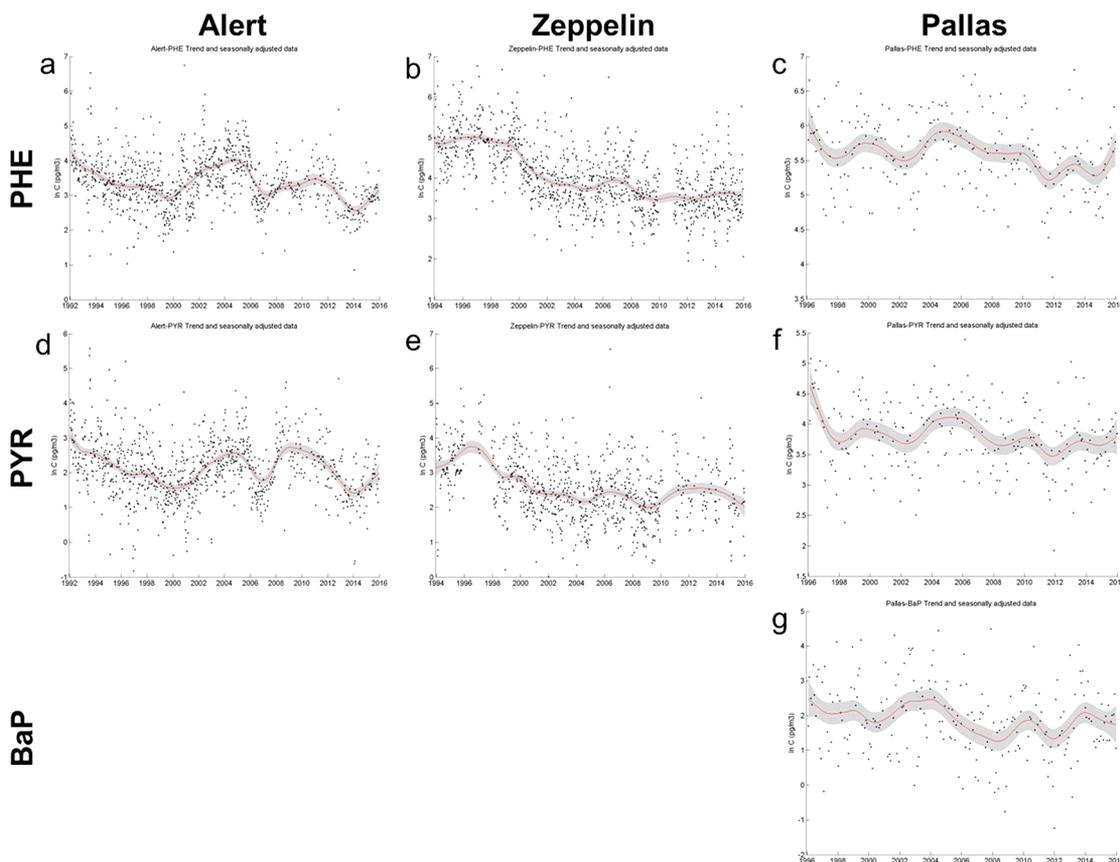


Figure 2. Long-term trends of (a, b, c) PHE and (d, e, f) PYR at three sites and (g) BaP at Pallas.

the order of Pallas > Zeppelin > Alert, especially for lighter PAHs, e.g., phenanthrene (PHE), anthracene (ANT), fluoranthene (FLA) and pyrene (PYR). The concentrations of eight of the targeted PAHs which were analyzed at all three sites during the sampling periods, i.e., PHE, ANT, FLA, PYR, benzo[*a*]anthracene (BaA), benzo[*a*]pyrene (BaP), indeno-

[1,2,3-*c,d*]pyrene (IcdP), and benzo[*g,h,i*]perylene (BghiP), are shown in Figure 1. The median concentrations of $\Sigma 8$ PAHs for Alert, Zeppelin and Pallas were 47.8, 76.0, and 404 pg/m^3 , respectively. PHE, FLA, and PYR were the most abundant PAHs at all sites, accounting for >85% of $\Sigma 8$ PAHs at Zeppelin and >91% of $\Sigma 8$ PAHs at the other two sites. The annual mean

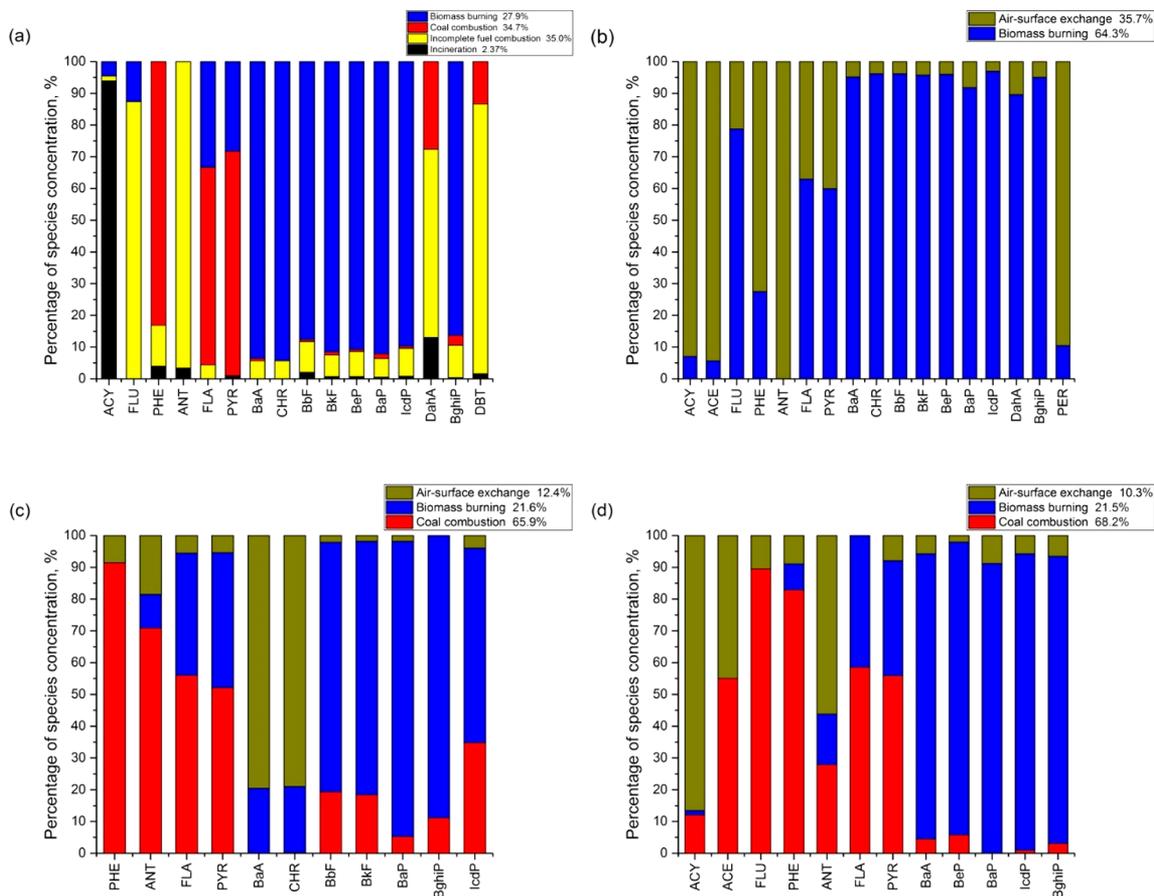


Figure 3. Factor Fingerprints for (a) Alert (1992–2003), (b) Alert (2004–2015), (c) Pallas (1996–2015), and (d) Zeppelin (1994–2015).

concentrations of Σ8PAHs ranged from 49.0 to 363, 91.7 to 523, and 346 to 817 pg/m³ for Alert, Zeppelin, and Pallas, respectively. Higher concentrations of PAHs at Pallas compared to those at Alert and Zeppelin are likely due to Pallas being closer to human settlements than the other two sites.

By separating the sampling periods into two parts, i.e., warm seasons (from May to October) and cold seasons (November to April), the seasonality of PAHs was assessed. As shown in Tables S4–S6 and Figure S2, the concentrations and detection frequencies of PAHs in cold, dark seasons are much higher than those in warm, sunny seasons. Specifically, high concentrations of PAHs were always measured in winter months, especially from December to February (Figure S3), which is consistent with previous studies.^{23,28,29} According to the back trajectory analysis, the air arriving at Alert in these three months includes more air masses originating from Russia than other months (Figure S4). The other two sites are more impacted by air masses coming from Northern Europe and Northwest Russia during December to February compared with the other nine months.

Temporal Trends. Two representative PAHs, namely PHE and PYR, were chosen for temporal trend assessment due to their high concentrations and detection frequencies compared with the other PAHs. BaP was also selected for trend assessment due to its toxicological importance and significantly different physical–chemical properties from PHE and PYR (Table S2) which would render differences in their atmospheric transport pathways. Data of BaP was sufficient only at Pallas for this assessment. Temporal trends of PAHs

were assessed with the Dynamic Harmonic Regression (DHR) model using full data sets collected at the three sites, i.e., from 1992 to 2015 for Alert, from 1994 to 2015 for Zeppelin, and 1996 to 2015 for Pallas (Figure 2 and Figure S5). Since PAHs are “seasonal contaminants”, the Seasonal Kendall test was performed for the statistical significance of the trends (Z_{SK}) of selected PAHs. Negative Z_{SK} values mean a decline and positive Z_{SK} values mean an increase, and p -values are used to confirm if the trends were statistically significant. Becker et al.²³ previously investigated the occurrence and trends of PAHs in the Canadian Arctic atmosphere from 1992 to 2000 and revealed a significant decrease of PAHs throughout the 1990s, e.g., the Z_{SK} for PHE and PYR were -3.39 and -2.97 ($p < 0.01$), respectively. In this study, with 15 more years of data, the temporal trends at Alert showed great variability since 2001. For PHE and PYR, the concentrations at Alert were found to increase significantly between 2001 and 2005 to levels similar to those observed in the early 1990s with Z_{SK} of 2.82 ($p < 0.01$) and 2.57 ($p = 0.01$), respectively. From 2006 to 2015, they increase and decline again to a lesser extent but generally have no significant trend ($Z_{SK} = -0.94$, $p > 0.05$; $Z_{SK} = -1.57$, $p > 0.05$). The complex trends since 2001 at Alert are mainly due to the much higher concentrations in summers during 2003 to 2005, which could be associated with the relatively more frequent active forest fire events in Canada, Alaska, and Greenland during these years (Figure S6). Similarly, higher PAH concentrations were observed in the summer of 2015 (Figure 2 and Figure S5) which coincide with more frequent forest fires during that year in the same regions (Figure S6). Retene (RET) is an ideal tracer of forest fire activity; high

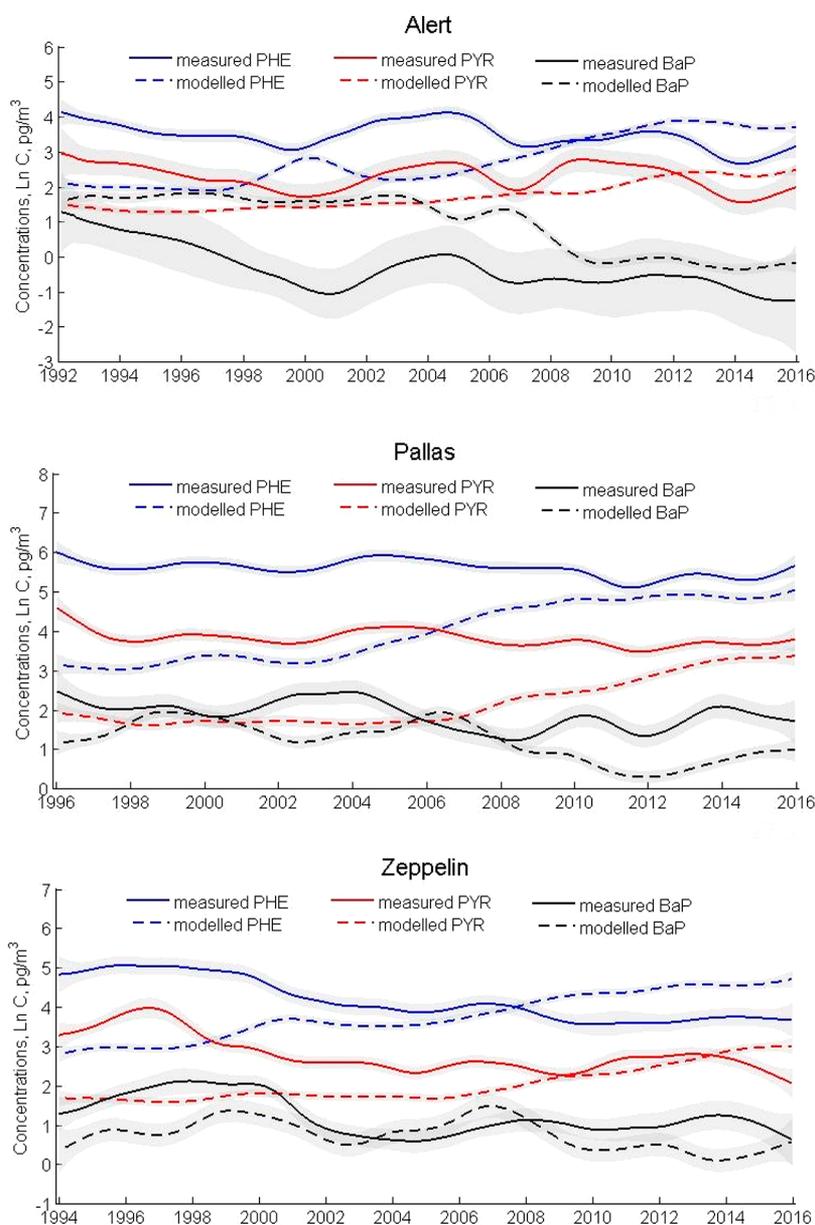


Figure 4. Trends of monthly mean concentrations of three PAHs (measured and simulated) at three sites. The shade represents the standard deviations of the trends.

levels of RET were found during 2003 to 2005 and 2015 (Figure S5), which confirm the contribution of forest fire events. At Zeppelin, significant declining trends were found for PHE and PYR between 1994 and 2001 ($Z_{SK} = -2.11$, $p < 0.05$; $Z_{SK} = -3.11$, $p < 0.01$), which is similar to Alert. Between 2002 and 2015, PHE and PYR seem to have reached a steady-state at Zeppelin ($Z_{SK} = -1.79$, $p > 0.05$; $Z_{SK} = 1.61$, $p > 0.05$). At Pallas, PHE and BaP were relatively stable over the two decades ($Z_{SK} = -1.72$, $p > 0.05$; $Z_{SK} = -1.38$, $p > 0.05$). Significant decline was found for PYR ($Z_{SK} = -2.37$, $p < 0.05$) which is mainly due to high concentrations in the first year of sampling; there is no significant trend for 1997 to 2015 ($Z_{SK} = -1.58$, $p > 0.05$). According to an estimation, the global emission of PAHs declined significantly from 1992 to 2015, especially in developed countries.⁹ However, PAHs found in air at Arctic sites did not seem to reflect this reduction in PAH emissions. In the past decade, human activities in the Arctic, for example, resource exploration, research, tourism, fisheries,

and maritime traffic, have increased substantially due to warming and the corresponding reduction of sea ice, opening up new shipping routes.³⁰ Such activities are potential new local sources of PAHs in the Arctic.

To better assess this, here, we examine whether the sources of PAHs changed during these years using a combination of PAH molecular diagnostic ratios and positive matrix factorization (PMF). Both techniques have been used widely over the years, and even though their limitations are known, when used in combination, it is possible to reduce their inherent weaknesses and strengthen the conclusions on potential sources and trends.^{3,13,31,32} The PAH ratios, e.g., PHE to the sum of PHE and ANT [$\text{PHE}/(\text{PHE}+\text{ANT})$] and FLA to the sum of FLA and PYR [$\text{FLA}/(\text{FLA}+\text{PYR})$], are widely used for source identification and, in particular, to understand if PAHs are mainly emitted from petroleum sources, or from combustion processes.^{33,34} Figure S7 shows the variations of the $\text{PHE}/(\text{PHE}+\text{ANT})$ and $\text{FLA}/(\text{FLA}+\text{PYR})$

+PYR) ratios over the sampling periods for the three sites. The calculated ratios at Pallas and Zeppelin were relatively constant, suggesting the sources of PAHs might not have changed significantly. Variations of PAH ratios were observed at Alert. The ratios of PHE/(PHE+ANT) were increasing and ratios of FLA/(FLA+PYR) were decreasing (especially after 2005), indicating that the sources shift from combustion of coal and wood to petroleum slowly, e.g., vehicle emissions. The ratios of FLA/(FLA+PYR) at Zeppelin and Pallas are significantly higher than those at Alert ($t < 0.01$), which may be reflective of the continued use of coal-fired power plants in Europe. This may also explain the nondeclining/slow declining trends for PHE and PYR at these two sites after the early 2000s.

Figure 3 shows the factor fingerprint obtained from PMF for the three sites. Detailed explanations for the factor source attribution are given in the Supporting Information. PMF was applied to the Alert data set separately for two time periods; four factors were identified for 1992–2003, but only two factors were found for 2004–2015 confirming a source shift at this site. After 2004, biomass burning, with high loads of 5–6 ring PAHs including IcdP and BghiP,³⁵ become Factor 1 [representing 64.3% of the sum of the measured PAHs (Σ PAHs)] which coincides with the increase in active forest fire events in 2003 to 2005. Factor 2 (35.7% of Σ PAHs) is dominated by more volatile 3–4 ring PAHs with very slight contributions from heavier 5–6 ring PAHs which seems to reflect air–surface exchange. Factor 3 for Zeppelin (10.3% of Σ PAHs) can also be attributed as air–surface exchange. Volatilization of PAHs from surfaces, such as ocean, snow, ice, permafrost and soil, may have become more important in recent years due to retreating sea ice and general warming in the arctic region. Coal combustion was no longer identified as a factor after 2004 at Alert. Coal combustion, which is usually identified by high loadings of PHE, ANT, FLU, FLA³⁶ and moderate loadings of 5–6 ring PAHs BbF, BkF, BaP, BghiP, and IcdP³⁵ was identified as Factor 1 for both Pallas and Zeppelin and Factor 3 for Alert before 2004, which accounted for 65.9%, 68.2%, and 34.7% of Σ PAHs, respectively. This observation reaffirmed that coal-fire power plants were still major sources of PAHs for the two European sites. Sofowote et al. identified petroleum/petrogenic emissions as a major source of PAHs at a sub-Arctic site, attributed to local oil/gas exploration and LRAT.³⁷ Alert is located at the highest latitude among the three sites and receives air mass more evenly from North America and Eurasia (Figure S4) and therefore may provide a more general reflection of changes in energy usage pattern.

Model Simulation. The changing energy usage pattern is not able to interpret atmospheric trends of PAHs in the Arctic completely. We used a global numerical model, the CanMETOP, to simulate the transport and occurrences of the three selected PAHs, PHE, PYR, and BaP. Figure 4 shows the trends of the monthly mean concentrations of the measured and modeled PAHs derived by DHR at the three Arctic sites during their sampling times. As shown in Figure 4, it underestimates the concentrations of PHE and PYR in the 1990s by less than 1 order of magnitude for Alert and Zeppelin but by 13–17 times for Pallas. A modeling sensitivity analysis (Table S7) indicates that the uncertainty in emission inventory affects modeling results at the Pallas site slightly more than the other two sites. The great discrepancy between the modeled and measured PAHs at Pallas could be associated

with the actual emissions in these years being higher than the emissions of 2004²⁷ which we employed as input for the model. The simulated concentrations of PHE increased significantly over the ~20 years at the three sites, whereas simulated PYR exhibited different trends; increasing concentrations only became apparent since 2008. This difference is probably due to the higher volatility of PHE (Henry's Law constant, $H = 4.78 \text{ Pa}\cdot\text{m}^3\cdot\text{mol}^{-1}$ at 25 °C) compared to PYR ($H = 2.13 \text{ Pa}\cdot\text{m}^3\cdot\text{mol}^{-1}$ at 25 °C) (Table S2). Warming within and around the Arctic region would enhance volatilization from surfaces and render higher simulated concentrations of PHE and PYR in the Arctic atmosphere. Figure S8 shows the annual mean of modeled net air–surface flux. The air–surface exchange switched from net deposition to net volatilization around 2005, and the net air–surface flux increased rapidly since then, which accounted for the increment of the simulated concentrations. This observation also corresponds well with the PMF result which identified air–surface exchange as Factor 2 after 2004 at Alert and Factor 3 for Zeppelin. Meanwhile, it explains that the measured concentrations of PHE and PYR did not decrease substantially even though the global emissions declined significantly during the last two decades.⁹ The expected decline is offset by the increment caused by greater volatilization due to warming, especially in recent years. Casal et al. measured PAH air–surface exchange from 2013 to 2014 at a coastal Arctic site (Tromsø, 69° N) and they found that PAHs were in a volatilization zone in warm seasons and in an equilibrium zone in cold seasons.³⁸ The overall simulated concentrations of BaP match the observations well and showed a decreasing trend (Figure 4). Heavier PAHs like BaP are less influenced by temperature due to their relatively low volatility ($H = 0.20 \text{ Pa}\cdot\text{m}^3\cdot\text{mol}^{-1}$ at 25 °C), but these particle-bound chemicals are subject to LRAT as they can be protected by organic coatings on aerosols,³⁹ and thus the simulation of BaP is more accurate than PHE and PYR. Our conclusions are consistent with Friedman et al.¹⁵ where they evaluated the influence of 2000–2050 climate and emission changes on three atmospheric PAHs transport to the Arctic.

Simulated concentrations of PHE at two high-latitude sites, Alert and Zeppelin, increased from 1998 to 2000, likely due to the warming phase during those years and the modeling sensitivity to temperature (Figure S9, Table S7), whereas a slight increase was found at Pallas (Table S8), which is consistent with the conclusion by Ding et al.⁴⁰ that the most prominent warming in the Arctic occurred in northeastern Canada and Greenland. This observation shows that Arctic sites are useful in investigating the influence of climate change on the occurrence of PAHs and other contaminants. The highest annual mean temperatures over the ~20 years at Alert were observed in 1998 and 2001 (Figure S9), when the model predicted a period of rising concentrations of PHE in air at Alert. Moreover, simulated concentrations of PYR slightly increased but no such increase was found for BaP (Table S8) due to its lower H (Table S2) resulting in a lower tendency to volatilize from oceans and ice-covered surfaces. The second rising period, for both PHE and PYR, were found after 2008. The Arctic sea ice was drastically reduced in this century; specifically, it reached a minimum record in 2007 and 2012.^{41–43} The more volatile PAHs, PHE, and PYR, can be released from the melting Arctic ice and the recently opened ocean (Table S7). Simulated concentration ranges of PHE and PYR are smaller than those observed at Alert and Zeppelin, indicating that there might be some local sources, such as the

military base at Alert, science stations at Ny-Ålesund, and shipping emissions.

In summary, two decades of measurements of PAHs in air at Arctic sites did not show a significant decreasing trend that one would have expected to occur relating to a global reduction of PAH emissions. Model simulation indicates that climate change may enhance the volatilization of lighter PAHs and thus alter the expected decline. In view of increasing PAH emissions due to human activities in the North as a result of warming, e.g. increased shipping, tourism, and resource development activities, both direct and indirect climate change impacts of contaminant cycling in the Arctic environment must be considered in assessing environmental and health risks. While efforts should continue to reduce PAH emissions, monitoring of PAHs and other chemicals of emerging arctic concern is essential to better understand climate change influence on the occurrence and transport of these contaminants to and within the Arctic. To simulate the concentrations of PAHs more accurately, scientists should continue to update the emission inventory bearing in mind the effect of climate change, in particular, with forest fire as an increasingly important source.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: [10.1021/acs.est.8b05353](https://doi.org/10.1021/acs.est.8b05353).

Potential in sample breakthrough, DHR, the CanMETOP, air–surface exchange, model evaluation, and PMF results. Tables and figures summarize concentrations, trends, seasonality, ratios, and factor profiles of PAHs at the three sites. (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

*Phone: 86-431-85542296. E-mail: yuyong@iga.ac.cn.

*Phone: 1-416-739-5944. E-mail: hayley.hung@canada.ca.

ORCID

Jianmin Ma: [0000-0002-6593-570X](https://orcid.org/0000-0002-6593-570X)

Hayley Hung: [0000-0003-0719-8948](https://orcid.org/0000-0003-0719-8948)

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank all site and laboratory operators and students at the three stations. Financial support for the monitoring programs was provided by the Northern Contaminants Program (NCP), Crown-Indigenous Relations and Northern Affairs Canada (Alert); the Swedish Environmental Protection Agency and Finnish Meteorological Institute (FMI) (Pallas); and the Norwegian Environment Agency (Zeppelin). We thank the Canadian Forces Station Alert for supporting data collection at Alert. We appreciate Dr. Marta Venier's assistance on DHR, Dr. Philip Cheung's assistance on assimilation of Canadian Meteorological Center (CMC) meteorological data, Prof. Tao Huang's assistance on sensitivity analysis, and Prof. Xingming Zheng's assistance on ArcGIS.

■ REFERENCES

- (1) Keyte, I. J.; Harrison, R. M.; Lammel, G. Chemical reactivity and long-range transport potential of polycyclic aromatic hydrocarbons—a review. *Chem. Soc. Rev.* **2013**, *42* (24), 9333–9391.
- (2) Schuster, J. K.; Harner, T.; Su, K.; Mihele, C.; Eng, A. First Results from the Oil Sands Passive Air Monitoring Network for Polycyclic Aromatic Compounds. *Environ. Sci. Technol.* **2015**, *49* (5), 2991–2998.
- (3) Katsoyiannis, A.; Breivik, K. Model-based evaluation of the use of polycyclic aromatic hydrocarbons molecular diagnostic ratios as a source identification tool. *Environ. Pollut.* **2014**, *184*, 488–494.
- (4) Hung, H.; Katsoyiannis, A. A.; Brorström-Lundén, E.; Olafsdottir, K.; Aas, W.; Breivik, K.; Bohlin-Nizzetto, P.; Sigurdsson, A.; Hakola, H.; Bossi, R. Temporal trends of Persistent Organic Pollutants (POPs) in arctic air: 20 years of monitoring under the Arctic Monitoring and Assessment Programme (AMAP). *Environ. Pollut.* **2016**, *217*, 52–61.
- (5) Li, W.-L.; Liu, L.-Y.; Song, W.-W.; Zhang, Z.-F.; Qiao, L.-N.; Ma, W.-L.; Li, Y.-F. Five-year trends of selected halogenated flame retardants in the atmosphere of Northeast China. *Sci. Total Environ.* **2016**, *539*, 286–293.
- (6) De Laender, F.; Hammer, J.; Hendriks, A. J.; Soetaert, K.; Janssen, C. Combining monitoring data and modeling identifies PAHs as emerging contaminants in the Arctic. *Environ. Sci. Technol.* **2011**, *45* (20), 9024–9029.
- (7) Dahle, S.; Savinov, V.; Petrova, V.; Klungsøyr, J.; Savinova, T.; Batova, G.; Kursheva, A. Polycyclic aromatic hydrocarbons (PAHs) in Norwegian and Russian Arctic marine sediments: Concentrations, geographical distribution and sources. *Norw. J. Geol.* **2006**, *86* (1), 41–50.
- (8) Usenko, S.; Simonich, S. L. M.; Hageman, K. J.; Schrlau, J. E.; Geiser, L.; Campbell, D. H.; Appleby, P. G.; Landers, D. H. Sources and Deposition of Polycyclic Aromatic Hydrocarbons to Western U.S. National Parks. *Environ. Sci. Technol.* **2010**, *44* (12), 4512–4518.
- (9) Shen, H.; Huang, Y.; Wang, R.; Zhu, D.; Li, W.; Shen, G.; Wang, B.; Zhang, Y.; Chen, Y.; Lu, Y. Global atmospheric emissions of polycyclic aromatic hydrocarbons from 1960 to 2008 and future predictions. *Environ. Sci. Technol.* **2013**, *47* (12), 6415–6424.
- (10) Harvey, H. R.; Taylor, K. A.; Pie, H. V.; Mitchelmore, C. L. Polycyclic aromatic and aliphatic hydrocarbons in Chukchi Sea biota and sediments and their toxicological response in the Arctic cod, *Boreogadus saida*. *Deep Sea Res., Part II* **2014**, *102*, 32–55.
- (11) Yunker, M. B.; Macdonald, R. W.; Snowdon, L. R.; Fowler, B. R. Alkane and PAH biomarkers as tracers of terrigenous organic carbon in Arctic Ocean sediments. *Org. Geochem.* **2011**, *42* (9), 1109–1146.
- (12) Foster, K. L.; Stern, G. A.; Carrie, J.; Bailey, J. N. L.; Outridge, P. M.; Sanei, H.; Macdonald, R. W. Spatial, temporal, and source variations of hydrocarbons in marine sediments from Baffin Bay, Eastern Canadian Arctic. *Sci. Total Environ.* **2015**, *506–507*, 430–443.
- (13) Katsoyiannis, A.; Sweetman, A. J.; Jones, K. C. PAH Molecular Diagnostic Ratios Applied to Atmospheric Sources: A Critical Evaluation Using Two Decades of Source Inventory and Air Concentration Data from the UK. *Environ. Sci. Technol.* **2011**, *45* (20), 8897–8906.
- (14) Hung, H.; Kallenborn, R.; Breivik, K.; Su, Y.; Brorström-Lundén, E.; Olafsdottir, K.; Thorlacius, J. M.; Leppänen, S.; Bossi, R.; Skov, H. Atmospheric monitoring of organic pollutants in the Arctic under the Arctic Monitoring and Assessment Programme (AMAP): 1993–2006. *Sci. Total Environ.* **2010**, *408* (15), 2854–2873.
- (15) Friedman, C. L.; Zhang, Y.; Selin, N. E. Climate Change and Emissions Impacts on Atmospheric PAH Transport to the Arctic. *Environ. Sci. Technol.* **2014**, *48* (1), 429–437.
- (16) Ma, J.; Hung, H.; Tian, C.; Kallenborn, R. Revolatilization of persistent organic pollutants in the Arctic induced by climate change. *Nat. Clim. Change* **2011**, *1* (5), 255–260.
- (17) Kong, D.; MacLeod, M.; Hung, H.; Cousins, I. T. Statistical Analysis of Long-Term Monitoring Data for Persistent Organic

Pollutants in the Atmosphere at 20 Monitoring Stations Broadly Indicates Declining Concentrations. *Environ. Sci. Technol.* **2014**, *48* (21), 12492–12499.

(18) Macleod, M.; Riley, W. J.; McKone, T. E. Assessing the influence of climate variability on atmospheric concentrations of polychlorinated biphenyls using a global-scale mass balance model (BETR-global). *Environ. Sci. Technol.* **2005**, *39* (17), 6749–56.

(19) Schlabach, M.; Farag-Clement, R.; Hung, H.; Kallenborn, R.; Su, Y.; Aas, W. *AMAP/EMEP/NCP inter-laboratory study for POP analysis 2010. EMEP/CCC-Report 7/2011*; Norwegian Institute for Air Research: Kjeller, Norway, 2012; p133.

(20) Su, Y.; Hung, H. Inter-laboratory comparison study on measuring semi-volatile organic chemicals in standards and air samples. *Environ. Pollut.* **2010**, *158* (11), 3365–3371.

(21) Paatero, P.; Hopke, P. K.; Begum, B. A.; Biswas, S. K. A graphical diagnostic method for assessing the rotation in factor analytical models of atmospheric pollution. *Atmos. Environ.* **2005**, *39* (1), 193–201.

(22) Venier, M.; Hung, H.; Tych, W.; Hites, R. A. Temporal Trends of Persistent Organic Pollutants: A Comparison of Different Time Series Models. *Environ. Sci. Technol.* **2012**, *46* (7), 3928–3934.

(23) Becker, S.; Halsall, C. J.; Tych, W.; Hung, H.; Attewell, S.; Blanchard, P.; Li, H.; Fellin, P.; Stern, G.; Billeck, B. Resolving the long-term trends of polycyclic aromatic hydrocarbons in the Canadian Arctic atmosphere. *Environ. Sci. Technol.* **2006**, *40* (10), 3217–3222.

(24) Zhang, Y.; Shen, H.; Tao, S.; Ma, J. Modeling the atmospheric transport and outflow of polycyclic aromatic hydrocarbons emitted from China. *Atmos. Environ.* **2011**, *45* (17), 2820–2827.

(25) Zhang, L.; Ma, J.; Venkatesh, S.; Li, Y.-F.; Cheung, P. Modeling Evidence of Episodic Intercontinental Long-Range Transport of Lindane. *Environ. Sci. Technol.* **2008**, *42* (23), 8791–8797.

(26) Zhang, L.; Ma, J.; Tian, C.; Li, Y.; Hung, H. Atmospheric transport of persistent semi-volatile organic chemicals to the Arctic and cold condensation in the mid-troposphere – Part 2: 3-D modeling of episodic atmospheric transport. *Atmos. Chem. Phys.* **2010**, *10* (15), 7315–7324.

(27) Zhang, Y.; Tao, S. Global atmospheric emission inventory of polycyclic aromatic hydrocarbons (PAHs) for 2004. *Atmos. Environ.* **2009**, *43* (4), 812–819.

(28) Anttila, P.; Brorström-Lundén, E.; Hansson, K.; Hakola, H.; Vestenius, M. Assessment of the spatial and temporal distribution of persistent organic pollutants (POPs) in the Nordic atmosphere. *Atmos. Environ.* **2016**, *140*, 22–33.

(29) Halsall, C. J.; Barrie, L. A.; Fellin, P.; Muir, D.; Billeck, B.; Lockhart, L.; Rovinsky, F. Y.; Kononov, E. Y.; Pastukhov, B. Spatial and temporal variation of polycyclic aromatic hydrocarbons in the Arctic atmosphere. *Environ. Sci. Technol.* **1997**, *31* (12), 3593–3599.

(30) Jorundsdottir, H. O.; Jensen, S.; Hylland, K.; Holth, T. F.; Gunnlaugsdottir, H.; Svavarsson, J.; Olafsdottir, A.; El-Taliawy, H.; Riget, F.; Strand, J.; Nyberg, E.; Bignert, A.; Hoydal, K. S.; Halldorsson, H. P. Pristine Arctic: Background mapping of PAHs, PAH metabolites and inorganic trace elements in the North-Atlantic Arctic and sub-Arctic coastal environment. *Sci. Total Environ.* **2014**, *493*, 719–728.

(31) Wang, C.; Wu, S.; Zhou, S.; Sill, Y.; Song, J. Characteristics and Source Identification of Polycyclic Aromatic Hydrocarbons (PAHs) in Urban Soils: A Review. *Pedosphere* **2017**, *27* (1), 17–26.

(32) Khairy, M. A.; Lohmann, R. Source apportionment and risk assessment of polycyclic aromatic hydrocarbons in the atmospheric environment of Alexandria, Egypt. *Chemosphere* **2013**, *91* (7), 895–903.

(33) Yunker, M. B.; Macdonald, R. W.; Vingarzan, R.; Mitchell, R. H.; Goyette, D.; Sylvestre, S. PAHs in the Fraser River basin: a critical appraisal of PAH ratios as indicators of PAH source and composition. *Org. Geochem.* **2002**, *33* (4), 489–515.

(34) Galarneau, E. Source specificity and atmospheric processing of airborne PAHs: Implications for source apportionment. *Atmos. Environ.* **2008**, *42* (35), 8139–8149.

(35) Chen, P.; Li, C.; Kang, S.; Yan, F.; Zhang, Q.; Ji, Z.; Tripathee, L.; Rupakheti, D.; Rupakheti, M.; Qu, B.; Sillanpaa, M. Source apportionment of particle-bound polycyclic aromatic hydrocarbons in Lumbini, Nepal by using the positive matrix factorization receptor model. *Atmos. Res.* **2016**, *182*, 46–53.

(36) Wang, K.; Shen, Y.; Zhang, S.; Ye, Y.; Shen, Q.; Hu, J.; Wang, X. Application of spatial analysis and multivariate analysis techniques in distribution and source study of polycyclic aromatic hydrocarbons in the topsoil of Beijing, China. *Environ. Geol.* **2009**, *56* (6), 1041–1050.

(37) Sofowote, U. M.; Hung, H.; Rastogi, A. K.; Westgate, J. N.; Deluca, P. F.; Su, Y.; McCarry, B. E. Assessing the long-range transport of PAH to a sub-Arctic site using positive matrix factorization and potential source contribution function. *Atmos. Environ.* **2011**, *45* (4), 967–976.

(38) Casal, P.; Castro-Jimenez, J.; Pizarro, M.; Katsoyiannis, A.; Dachs, J. Seasonal soil/snow-air exchange of semivolatile organic pollutants at a coastal arctic site (Tromsø, 69 degrees N). *Sci. Total Environ.* **2018**, *636*, 1109–1116.

(39) Zhou, S.; Lee, A. K. Y.; McWhinney, R. D.; Abbatt, J. P. D. Burial Effects of Organic Coatings on the Heterogeneous Reactivity of Particle-Borne Benzo(a)pyrene (BaP) toward Ozone. *J. Phys. Chem. A* **2012**, *116* (26), 7050–7056.

(40) Ding, Q.; Wallace, J. M.; Battisti, D. S.; Steig, E. J.; Gallant, A. J. E.; Kim, H.-J.; Geng, L. Tropical forcing of the recent rapid Arctic warming in northeastern Canada and Greenland. *Nature* **2014**, *509*, 209–212.

(41) Comiso, J. C.; Parkinson, C. L.; Gersten, R.; Stock, L. Accelerated decline in the Arctic Sea ice cover. *Geophys. Res. Lett.* **2008**, *35* (1), 179–210.

(42) Comiso, J. C. Large Decadal Decline of the Arctic Multiyear Ice Cover. *J. Clim.* **2012**, *25* (4), 1176–1193.

(43) Liu, Y.; Key, J. R. Less winter cloud aids summer 2013 Arctic sea ice return from 2012 minimum. *Environ. Res. Lett.* **2014**, *9* (4), 044002.

Hydrological functioning of cattle ranching impoundments in the Dry Chaco rangelands of Argentina

Patricio N. Magliano, David Mindham, Wlodek Tych, Francisco Murray, Marcelo D. Noretto, Esteban G. Jobbágy, Marcos J. Niborski, Mariana C. Rufino and Nick A. Chappell

ABSTRACT

Rainwater harvesting and associated storage is essential for cattle ranching in the drylands of Argentina and elsewhere. This is the first study to attempt to quantify the hydrological inflows and losses from rainwater harvesting impoundments. To address the direct effect of cattle within impoundments, a typical cattle-affected impoundment was instrumented and compared with that of a similar impoundment but without cattle access. Analysis of the storage dynamics with reference to the controlling variables demonstrated the highly episodic nature of the generation of infiltration-excess overland flow that recharged the impoundments. The impoundments experienced 43 and 35% of storage loss to open-water-evaporation for the cattle-affected and control impoundments, respectively. Critically, the cattle-affected impoundment lost only 15% of storage to leakage (after cattle consumption was taken into account), while the control lost 65% of its water to basal leakage. Indeed systems modelling of the rainfall-storage dynamics showed that the cattle-affected impoundment, despite consumption by 300 cows, maintained water in the impoundment (per a unit input of rainfall) for longer than the control (a 65- versus 25-day residence time). These results highlight the unintended beneficial effect of cattle trampling on the floor of the impoundment reducing leakage losses.

Key words | arid, ecohydrology, grazing, livestock, rainwater harvesting, runoff

Patricio N. Magliano (corresponding author)

Francisco Murray

Marcelo D. Noretto

Esteban G. Jobbágy

Grupo de Estudios Ambientales – IMASL,
Universidad Nacional de San Luis & CONICET,
Ejército de los Andes 950, D5700HHW San Luis,
Argentina
E-mail: pnmagliano@gmail.com

Patricio N. Magliano

Departamento de Bioquímica y Ciencias
Biológicas, Facultad de Química, Bioquímica y
Farmacia,
Universidad Nacional de San Luis (UNSL),
Ejército de los Andes 950, D5700HHW San Luis,
Argentina

Patricio N. Magliano

David Mindham

Wlodek Tych

Mariana C. Rufino

Nick A. Chappell

Lancaster Environment Centre,
Lancaster University,
Lancaster, UK

Francisco Murray

Instituto Nacional de Tecnología Agropecuaria
(INTA),

AER San Luis, Ruta 20 km 4.5, D5700HHW San Luis,
Argentina

Marcelo D. Noretto

Cátedra de Climatología,

Facultad de Ciencias Agropecuarias (UNER),
Entre Ríos, Argentina

Marcos J. Niborski

Cátedra de Manejo y Conservación de Suelos,
Facultad de Agronomía,
Universidad de Buenos Aires (UBA),
Avda. San Martín 4453, C1417DSE Ciudad
Autónoma de Buenos Aires, Argentina

INTRODUCTION

Extensive livestock production occupies one-third of the global land area and represents a key economic activity in low- and middle-income countries (Schimel 2010; Robinson *et al.* 2014). Given that most livestock systems are located

in arid and semi-arid regions (hereafter drylands), water is the main constraint to production due to the shortage of palatable forage and limited access to drinking water for the cattle (Evenari *et al.* 1971; UNEP 2009). In the case of

forage shortage, there are several management practices to consider, such as adjusting stocking rates (Ganskopp & Bohnert 2009; Fernández *et al.* 2015). However, water supply represents a crucial limitation for cattle survival (Ganskopp 2001; Bailey 2005) and it has been historically overcome by capturing infiltration-excess overland flow during rain events, so called 'rainwater harvesting' (Lavee *et al.* 1997; Oweis & Hachum 2009; Denison & Wotshela 2012). Huxman *et al.* (2004), Newman *et al.* (2006) and Jobbágy *et al.* (2008) have shown that this is feasible in dry and flat sedimentary landscapes where infiltration-excess overland flow may represent less than 5% of rainfall inputs. Infiltration-excess overland flow is surface flow on slopes (rather than channels) created when short-term rainfall intensities locally (in time and space) exceed the infiltration capacity of the soil (equivalent to the saturated hydraulic conductivity of the soil surface).

By definition, rainwater harvesting comprises of capturing, storing and then using infiltration-excess overland flows for water supply (Critchley *et al.* 1991; Fox & Rockström 2000). Most rainwater harvesting systems have a well-defined area upslope of an impoundment, where infiltration-excess overland flow is generated (Boers & Ben-Asher 1982; Pandey *et al.* 2003). Each impoundment may be a natural topographic depression or man-made excavation with embankments (Glendenning *et al.* 2012; Adham *et al.* 2016). In very flat landscapes it can be difficult to determine the exact contributory area of an impoundment because it is hard to define topographic divide and because of the enhanced role of preferential pathways of vehicle roads or cattle trails that may connect distal areas with the impoundment (Pandey *et al.* 2003; Magliano *et al.* 2015c). The generation of infiltration-excess overland flow is, however, enhanced by the presence of the high-intensity convective rainfall that characterizes semi-arid regions.

The Dry Chaco rangelands of northern Argentina (also known as Arid Chaco) is ~10 million hectares of sedimentary plain dominated by native dry woodlands (Cabrerá 1976; Oyarzabal *et al.* 2018). In this region, rainwater harvesting systems used for cattle ranching are typically composed of a man-made impoundment associated with a contributory area crossed by cattle trails and vehicle roads (Karlin *et al.* 2013; Magliano *et al.* 2015c). Impoundments, with typical dimensions of approximately 50 × 80 m, consist of arched

embankments built 2–4 m above the ground with an open section that allows entry of the infiltration-excess overland flow. Cattle drink water directly from within the impoundment so trampling the floor of the impoundment (in addition to trampling of the access points and contributory catchment) may be important. Cattle trails include networks of tracks characterised by bare and compacted ground that converges onto the open section of the impoundment. They are created by cattle as they move daily from grazing sites to the impoundments (George *et al.* 2004; Ganskopp & Bohnert 2009; Richardson & Rejmánek 2011; Karlin *et al.* 2013). These trails are often associated with an overgrazed surface zone close to the open side of the impoundment that is sometimes called the 'píosphere' or zone of focused impact of grazing animals on soils and vegetation (Macchi & Grau 2012; Chillo *et al.* 2015). Vehicle tracks are also present in the contributory catchments and are typically unpaved, slightly depressed roadways. There is some evidence that undisturbed Dry Chaco woodlands do not generate infiltration-excess overland flow that is able to travel more than a few metres (Aguilera *et al.* 2003; Kunst *et al.* 2003; Magliano *et al.* 2016; Magliano *et al.* 2015a), highlighting the importance of the cattle trails and vehicle tracks to generate the infiltration-excess overland flow that fills the impoundments built for cattle water supply.

Despite its great importance to cattle production in Dry Chaco rangelands, there is little scientific knowledge about the hydrological functioning of these impoundments and the biophysical controls on their water fluxes. Typically, there is one rainwater harvesting system every 1,000 ha within the cattle-grazed rangeland, that each collect only a tiny fraction of the rain falling in their contributory areas (Magliano *et al.* 2015c). Although rainwater harvesting systems make cattle production possible in the region, they are not considered efficient systems because of their perceived large water losses. During dry years, or at the end of the dry season, impoundments often run out of water generating serious problems for cattle ranchers. While local ranchers believe that water depletion in impoundments results from a combination of consumption by cattle and evaporation, the preliminary study of Magliano *et al.* (2015c) suggested that leakage through the impoundment floor may be an important water loss. Detailed understanding of the hydrological functioning of example

impoundments is, therefore, a potential first step to improving the design and management of these rainwater harvesting systems in the Dry Chaco of Argentina and similar systems in other rangelands worldwide.

This study explored the hydrological functioning of two instrumented impoundments built to hold harvested rainwater in the Dry Chaco rangelands of Argentina (Figure 1). Two particular impoundments were selected for this first pioneering study because of the potential importance of the combined effects of cattle consumption of the impoundment water and the effects of trampling of the impoundment floor during cattle access for drinking. One impoundment ('cattle impoundment') was selected, as cattle accessed that impoundment to drink (Figure 1). The second impoundment did not have cattle access or water consumption (though it did have compacted tracks within its contributory area) and was selected as a reference or 'control' to permit the study of the effects of cattle presence/absence within an impoundment.

The first objective of the study was to quantify the storage dynamics of the two instrumented impoundments over the water year. The second objective was to estimate, for the first time: (1) the temporal dynamics of the additions to storage from infiltration-excess overland flow from the local rainfall, and (2) a decomposition of the losses from storage attributed separately to open water evaporation, cattle

drinking and leakage through the impoundment floor. As the most critical attribute of the cattle-drinking impoundments is maintenance of storage for the longest period possible (i.e. security from water supply failure), the dynamic relationship between the unique synchronous rainfall and impoundment storage data were analysed using a systems modelling approach. This final objective being to quantify the residence time of a unit input of rainfall within the two impoundments (i.e. how long storage is maintained for standardised rainfall inputs), one with cattle consumption and trampling, one without.

METHODS

Study site

Extensive cattle ranching represents the most important economic activity of the Dry Chaco rangelands of Argentina. Drinking water for cattle is supplied by rainwater harvesting systems, while native grasses and woody species provide the forage (Aguilera 2003). This study was undertaken within a 20,000 ha farm located in the centre of the San Luis province of Argentina (33.5°S, 66.5°W; Figure 1) and was managed by Ser Beef S.A. Typical stocking rates

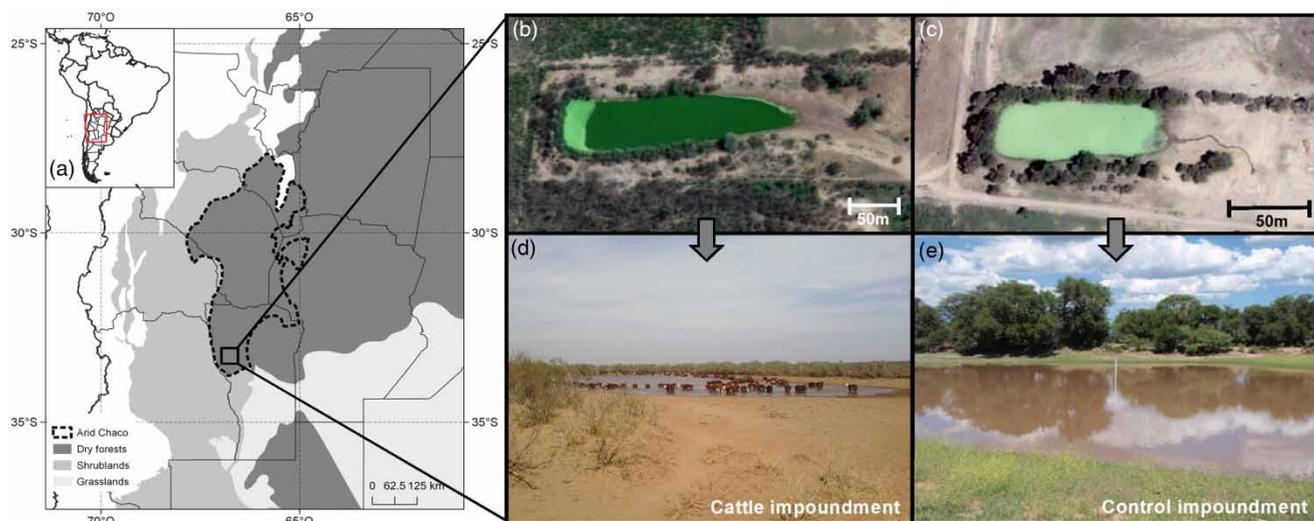


Figure 1 | Study site in central Argentina. (a) The map on the left shows the Dry Chaco rangelands (Arid Chaco, ~10 million of hectares), while (b) and (d) are images of the instrumented impoundment used for cattle water supply (cattle impoundment), and (c) and (e) are the control impoundment. Images (b) and (c) are publicly available 'Quickbird' images from Google Earth, while the lead author took photographs (d) and (e).

were 0.15 calving units per ha (Ser Beef S.A. personal communication).

The vegetation is dominated by native dry woodlands with isolated stands of altered vegetation that include rainfed pastures of *Cenchrus ciliaris* and *Eragrostis curvula* and some localised irrigated plots of maize, sorghum and soybean (Marchesini et al. 2015; Steinaker et al. 2016). The native dry woodlands are dominated by two tree species: *Prosopis flexuosa* and *Aspidosperma quebracho-blanco*. The understorey includes several shrub species such as *Larrea divaricata* and *Senna aphylla*, and perennial grasses such as *Stipa eriostachya* and *Aristida mendocina* (García et al. 2017). The mean leaf area index of the dry woodland approaches 1 and mean incident radiation at the ground level is 47.5% (García et al. 2017).

The soils are well drained and derived from fine loessic sediments deposited throughout the Holocene with some alluvial reworking (Iriondo 1993; Tripaldi et al. 2013). Soils are Typic Torriorthents with 53% sand, 15% clay, and 1.4% organic matter in the top 10 cm of the profile (Peña Zubiate et al. 1998; Pennington et al. 2000). Topography is gentle with slopes of <1.5%. Saturated hydraulic conductivity values around 140 mm h⁻¹ have been measured for the surface soil in nearby undisturbed, native dry woodlands, with dry bulk density and ‘water holding capacity’ values of approximately 1.2 g cm⁻³ and 20.6%, respectively (Magliano et al. 2017b). The regional water table is approximately 30 m deep at the study site (Ser Beef S.A. unpublished data).

The climate is semiarid with a long-term mean annual rainfall of only 429 ± 83 mm (mean and standard deviation), distributed in 40 ± 12 events per year (Magliano et al. 2015b). Rainfall has a strong seasonal pattern. Some 85% of annual rainfall occurs in the wet season of September–March. During the wet season, there are more rainfall events of larger size and so more likely to generate infiltration-excess overland flow. The mean annual potential evapotranspiration is approximately 1,500 mm year⁻¹ (New et al. 2002) and reaches maximum daily values of approximately 8.5 mm day⁻¹ in the summer (Magliano et al. 2017a). The mean annual temperature is 17.8 °C. The hottest and coldest months have mean temperatures of 24.8 (January) and 10.3 °C (July), respectively. Some 38 ground frost events are observed each year, mostly in the cold dry season (New et al. 2002).

Field monitoring of control and cattle-affected impoundments

Imagery of the ‘cattle impoundment’ is given in Figure 1(b) and 1(d), and for the ‘control impoundment’ (with no cattle ingress) in Figure 1(c) and 1(e). At each impoundment, the water level was monitored using a data-logged pressure transducer (depth resolution ±2 mm; Hobo, Onset Corporation, Bourne, MA, USA). As the impoundments have a bowl shape, sensors were installed at the base of the deepest point, and the water level was recorded every 30 min. All monitoring was undertaken between 10th November 2010 and 4th January 2012, totalling 421 days of measurements. A detailed bathymetric survey of both impoundments was undertaken so that the daily-average water levels could be converted into values of daily water volume (Equations (1) and (2)) and daily water surface area (Equations (3) and (4)). Further details of the water volume calculations using an adapted fitting curve method applied to the high-resolution digital elevation model are presented in Magliano et al. (2015c). The derived level to volume relationships for the cattle and control impoundments are given in Equations (1) and (2), respectively:

$$V = 1335.5 h^2 - 125.8 h \quad (1)$$

$$V = 1033.9 h^2 - 181.5 h \quad (2)$$

where V is the water volume (m³) and h is the water level (m), and level to water surface area relations are given in Equations (3) and (4), respectively:

$$S = 2586.7 h \quad (3)$$

$$S = 1993.3 h \quad (4)$$

where S is water surface (m²) and h is water level (m)

Rainfall was measured with a tipping-bucket gauge installed 1 km from the impoundments (resolution = 0.1 mm per tip; Vantage Pro2 weather station, Davis Instruments, USA), and accumulated tips recorded every 30 min. The annual average rainfall recorded by the tipping-bucket gauge was checked against those data from a nearby storage rain gauge. Meteorological records of solar radiation,

temperature, humidity and wind speed from the same Vantage Pro2 weather station were used to estimate open water evaporation using an adapted Penman method (following Allen et al. 1998).

Derivation of infiltration-excess inputs and evaporation, consumption and leakage outputs of the impoundments from the storage dynamics and associated data

Water loss from the two impoundments by the combined effect of open water evaporation, drainage through the impoundment floor and consumption by cattle, was derived on a daily time-step:

$$Y = (x_n - x_{n-1}), \text{ where } (x_n - x_{n-1}) < 0 \quad (5)$$

where a negative Y is daily water loss ($\text{m}^3 \text{ day}^{-1}$), x is water volume at n day. In the case of $Y > 0$, this equates to very short episodes of recharge of the impoundment (m^3) by infiltration-excess overland flow from the surrounding 'rainwater harvesting' catchment.

Daily water loss was decomposed into estimates of the three component fluxes (evaporation, cattle consumption and drainage through the impoundment floor), using Equations (6)–(8), respectively:

$$y_e = x_1^* S, \text{ where } Y < 0 \quad (6)$$

where y_e is daily estimate of the volume of water lost by evaporation from the impoundment (m^3), x_1 is daily evaporation depth estimate by the adapted Penman method of Allen et al. (1998; m) and S is water surface area of the impoundment (m^2):

$$y_c = x_2^* n, \text{ where } Y < 0 \quad (7)$$

where y_c is daily cattle water consumption volume from the 'cattle impoundment' (m^3), x_2 is the daily rate of consumption estimated to be 0.05 m^3 per cow for dryland sites (Schlink et al. 2010) and n is the number of cattle accessing the 'cattle impoundment' each day (300 cows: Ser Beef S.A. personal communication). The cattle access the studied

impoundment eight months per year; in this study, the access started on 15th February 2011 (Ser Beef S.A. personal communication). Consequently, the estimate of the drainage through the floor of the impoundment is the total water loss minus the evaporation and consumption:

$$y_i = Y - y_e - y_c, \text{ where } Y < 0 \quad (8)$$

where y_i is daily volume of infiltration into the floor of the wetted area of the impoundment (m^3), Y is daily volume losses from the rate of change of water level (m^3), y_e is daily evaporation volume (m^3) and y_c is daily cattle consumed water volume (m^3).

The overall input-output water balance of each impoundment over the 421-day period, taking into account the infiltration-excess overland flow inputs of the rainwater harvesting system (m^3) and the losses (evaporation, cattle consumption and drainage through the impoundment floor, m^3), was then computed with Equation (9):

$$y = \sum a - \sum b \quad (9)$$

where y is the daily volume balance, a is an impoundment input and b is an impoundment loss. To explore the dynamics these daily values were expressed as a percentage of the maximum observed storage (6,140 and $4,102 \text{ m}^3$ for cattle and control impoundment, respectively).

Systems analysis of the input-storage dynamics of the impoundments

Rainwater harvesting is the dominant strategy for supplying drinking water to cattle in the Dry Chaco rangelands of Argentina. The main concern of ranchers is to have enough water at the end of the dry season, i.e. August (Figure 2(a)). Consequently, a key hydrological characteristic of the impoundments constructed for cattle water supply is how long a unit input of rainfall can be maintained in an impoundment.

To quantify how long storage can be maintained following a unit input of rainfall, a transfer-function approach has been used to identify the relationship between rainfall intensity and impoundment storage. The approach used was the

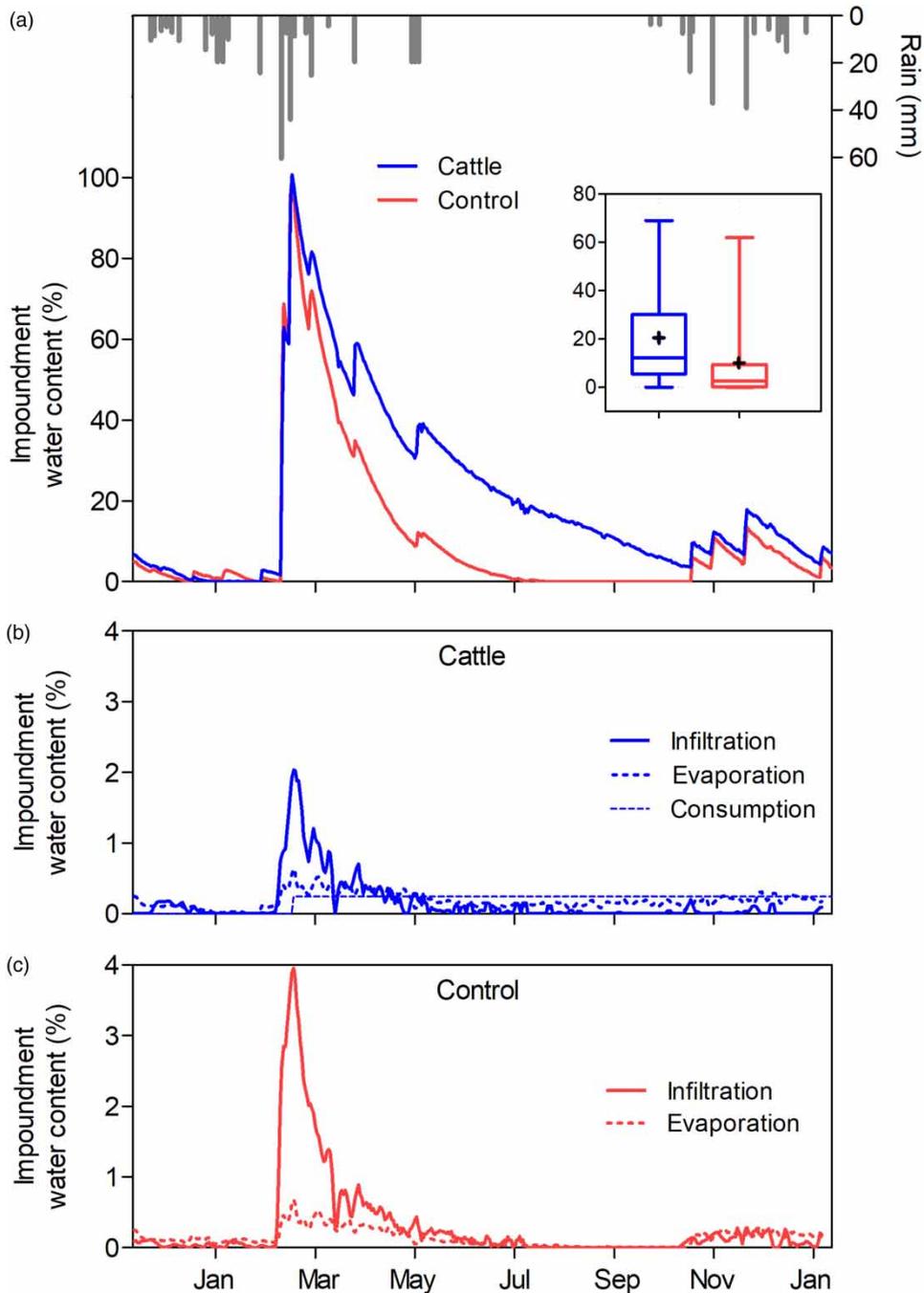


Figure 2 | Temporal dynamics of water content and derived losses for cattle and control impoundments, where: (a) water content dynamics, with the inset showing the distribution of the data; (b) dynamics of the derived estimates of evaporation, infiltration/leakage and consumption in the cattle impoundment; and (c) dynamics of the same variables for control impoundment.

Refined Instrumental Variable in Continuous-time (RIVC) algorithm of Young (2015) which is one of the algorithms that form the CAPTAIN Toolbox for Matlab™ (Taylor et al. 2007).

In summary, the RIVC algorithm implements an iterative instrumental variable method for estimation of general *Transfer Functions* that capture the dynamic relationship between an input (here rainfall, R) and an output variable

(here impoundment water volume, V) that can be written as:

$$V = \frac{\beta}{s - \alpha} e^{-sT} R; s \sim \frac{d}{dt} \quad (10)$$

where V is the impoundment water volume ($\text{m}^3 \text{ day}^{-1}$), R is the rainfall (mm day^{-1}), T is the pure time delay between R and an initial V response, where present (given in number of day intervals), α is the parameter capturing the rate of impoundment depletion or residence time of response (day^{-1}), β is the parameter capturing the magnitude of volume gain ($\text{m}^3 V$, $\text{mm}^{-1} R$), t is time in 1 day periods, and s is the Laplace operator. Full details of this method are given in Young (2015).

Physical interpretation of the behaviour is made after first calculating the dynamic response characteristics (DRCs) of the observed dynamics derived from transfer function parameters α and β . These DRCs include the time constant (TC) or residence time of response (here residence time of the storage following a unit input of rainfall):

$$TC = \frac{\Delta t}{\alpha} \quad (11)$$

where Δt is the time-step (days).

RESULTS AND DISCUSSION

Analysis of infiltration-excess overland-flow inputs to each impoundment

The accumulated infiltration-excess overland flow input to each impoundment (from the surrounding 'rainwater harvesting' catchment) over the 421 days of measurements was 10,690 and 6,630 m^3 for the cattle and control impoundments, respectively (Table 1). This is 1.7 and 1.6 times the maximum water storage capacity of the respective impoundments.

The inputs to the two impoundments were characterized by very short, discrete events of infiltration-excess overland flow, contrasting with the continuous losses to evaporation, leakage and cattle drinking (Figure 2(a)). The 421 days of the study period received 616.5 mm of rainfall. This was distributed in daily events ranging in size from

Table 1 | Total water inflows by infiltration-excess overland flow ('water harvested'), total water losses and their components (including percent of total loss) of open water evaporation from the impoundment, leakage through the floor of the impoundment and direct consumption of impoundment water by cattle

	Cattle		Control	
	m^3	%	m^3	%
Water harvested	10,690		6,630	
Water losses	10,673		6,835	
Storage change	17		-205	
Evaporation	4,666	43	2,392	35
Basal pond leakage	1,549	15	4,443	65
Consumption	4,458	42		

0.1 to 60 mm day^{-1} . The largest 10% of rainfall events accounted for 42.9% of total rainfall amount, reflecting the typical asymmetric distribution of rainfall in drylands (Knapp et al. 2015; Magliano et al. 2015b).

The smaller rainfall events ($<20 \text{ mm day}^{-1}$) seemed to generate much less infiltration-excess overland flow and so an increase in storage compared to the larger events ($>30 \text{ mm day}^{-1}$). Indeed both impoundments reached their maximum storage capacity only once in the study period, after a 60 mm rainfall event that was swiftly followed by a 45 mm event. Both impoundments were completely filled once on 13th February 2011. On this day, the maximum water depth was 2.19 m in the cattle impoundment, while water volume and water surface area were 6,140 m^3 and 6,134 m^2 , respectively. In the control impoundment, the maximum water depth was 2.08 m, maximum water volume was 4,102 m^3 and maximum water surface area was 4,543 m^2 .

To demonstrate the role of the event characteristics on preferential impoundment filling, the overland inflows into the impoundments were explored at a higher temporal resolution (30 min) during two example rainfall events of similar total event size but different short term intensity characteristics (Figure 3). In the cattle impoundment, the high-intensity event generated five times more overland flow than the low-intensity longer duration event (3,742 vs. 710 m^3); and 10 times more in the control impoundment (2,937 vs. 281 m^3 ; Figure 3).

Examining the full record of daily events, the cattle impoundment had 16 discrete input events while the control

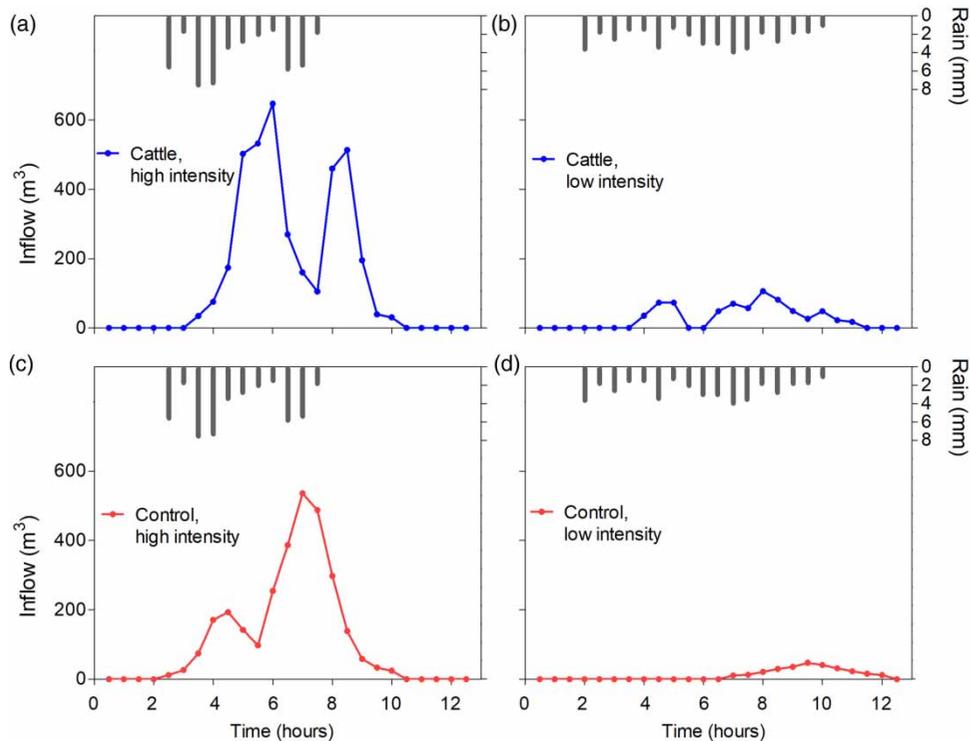


Figure 3 | Inflows recorded every 30 min in the cattle- (a) and (b) and control- (c) and (d) impoundments during example high-intensity (a) and (c) and low-intensity (b) and (d) rainfall events. High-intensity rainfall corresponds to a 45 mm event size with a weighted average intensity of 5.3 mm h⁻¹. Low-intensity rainfall corresponds to a 40 mm event size with a weighted average intensity of 2.8 mm h⁻¹.

had 13. Critically, the minimum daily rainfall event able to generate an infiltration-excess overland flow input to an impoundment was only 4 mm day⁻¹ in the cattle impoundment and a larger 20 mm day⁻¹ in the control. These observations may indicate that the surface of the catchment draining to the cattle impoundment was naturally less permeable than that of the control impoundment or that the cattle in the former catchment have more extensively compacted the ground enhancing the proportion of infiltration-excess overland flow relative to catchment-wide infiltration. This is consistent with the observations of long compacted cattle trails directed towards drinking impoundments seen in other regions (Ganskopp *et al.* 2000; Ganskopp & Bohnert 2009; Chillo *et al.* 2015).

Analysis of consumption, evaporation and leakage losses from each impoundment

The analysis of the storage dynamics showed that total water losses from the impoundments to all processes

over the 421-day monitoring period were 10,673 and 6,835 m³ from the cattle and control impoundments, respectively (Table 1). The difference between impoundment storage at the start and end of the 421-day period, for both impoundments, was negligible. Consequently, application of Equations (6)–(8) indicates that consumption by cattle represents 42% of the water input to the impoundment where cattle walk inside the impoundment to drink, with a similar amount lost to evaporation (43%), plus 15% lost through infiltration ('leakage') into the impoundment floor. In marked contrast, infiltration through the impoundment floor of the impoundment with no cattle access (or consumption) amounted to 65% of the water entering the impoundment, with a further 35% lost to evaporation.

Depletion of impoundment storage due to the combined losses (expressed by reductions in impoundment storage within Figure 2(a)), followed a clear exponential relationship ($R^2 = 0.82$ and 0.96 for cattle and control impoundment, respectively; Figure 4). The control

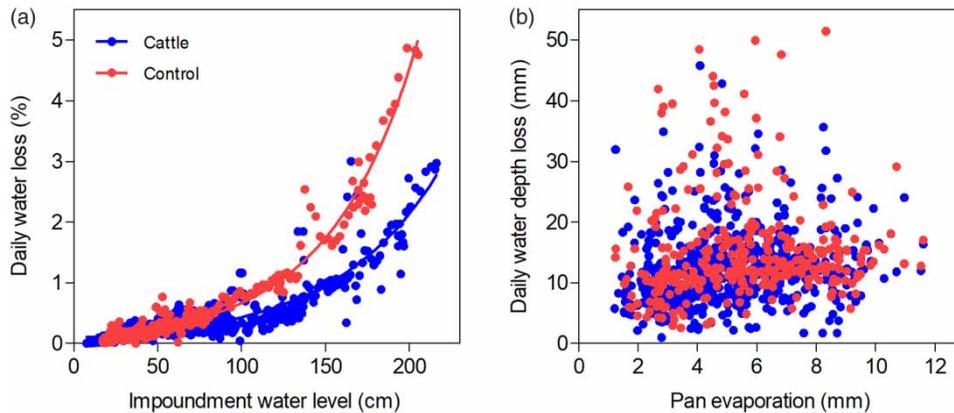


Figure 4 | Relationship between daily water loss (by all processes affecting that impoundment), expressed as a percentage of impoundments maximum storage capacity versus water level of both impoundments.

impoundment lost its water at a higher rate. For example, when the impoundments were full of water, daily losses were 4–5 and 2–3% of stored volume, for the control and cattle impoundments, respectively. Expressed differently, the cattle impoundment showed smaller variations in storage that were maintained closer to maximum storage on average (i.e. $20.4 \pm 50.9\%$ coefficient of variation, compared to $9.8 \pm 102.0\%$ for the control impoundment; Figure 2(a) inset).

At both impoundments the estimate of the evaporation was at its highest when the storage was at its greatest (Figure 2(a) and 2(b)), as the lateral extent of the impoundment surface was at its greatest at these times. Similarly infiltration (or ‘leakage’) into the impoundment floor was greatest at times of deepest water, as these were times of greatest hydraulic head driving the water into the impoundment floor. The rate of water loss into the soil below the control impoundment (relative to observed max storage) was, however, twice that beneath the impoundment where cattle were allowed to access. Indeed, 62% of the storage in the control impoundment was lost by infiltration in just one month (i.e. 13th February–15th March 2011). Either the impoundment floor of the ‘cattle impoundment’ has naturally less permeable soils, or trampling of the impoundment floor by cattle when they access to drink has reduced the permeability of the impoundment floor (so reducing leakage). Indeed wet soils beneath impoundments are likely to be particularly sensitive to cattle trampling (Belsky *et al.* 1999; Lebron *et al.* 2007; Fernández *et al.* 2015; Magliano *et al.* 2017b).

Systems (transfer function) modelling of input-storage dynamics

Application of the RIVC identification algorithm to the rainfall to storage dynamics of the two impoundments was shown to be able to capture more than 90% of the variance in the daily impoundment volume, m^3 (Figure 5; Table 2), thereby permitting comparison of the derived response characteristics between the two impoundments. Purely linear, first-order transfer function model parameters (see Equation (10)) were first identified for the cattle impoundment and a Nash–Sutcliffe simulation efficiency (R_f^2) of 0.875 was observed. The generation of infiltration-excess overland flow is sensitive to the wetness of the surface of the catchment. This may be represented by a single ‘memory term’ based on the rainfall history. Within this study, the memory term of the ‘Bedford-Ouse Sub-Model’, BOSM (see Equations (5) and (6) in (Chappell *et al.* 2006) was applied to the rainfall data, with R_{eff} replacing R in Equation (10). By including this one additional term, simulation efficiency increased to 0.945 (Table 2). It is worth mentioning that adding additional parameters can introduce uncertainty to simulation where the information content of the observations is insufficient to warrant the addition of further parameters. This was evaluated with the calculation of the Young Information Criterion, YIC (Young 2015). In the case of the cattle-affected impoundment, the addition of the single BOSM term shifted the value of the YIC performance measure from -8.584 to -10.391 (Table 2). A more negative number indicates

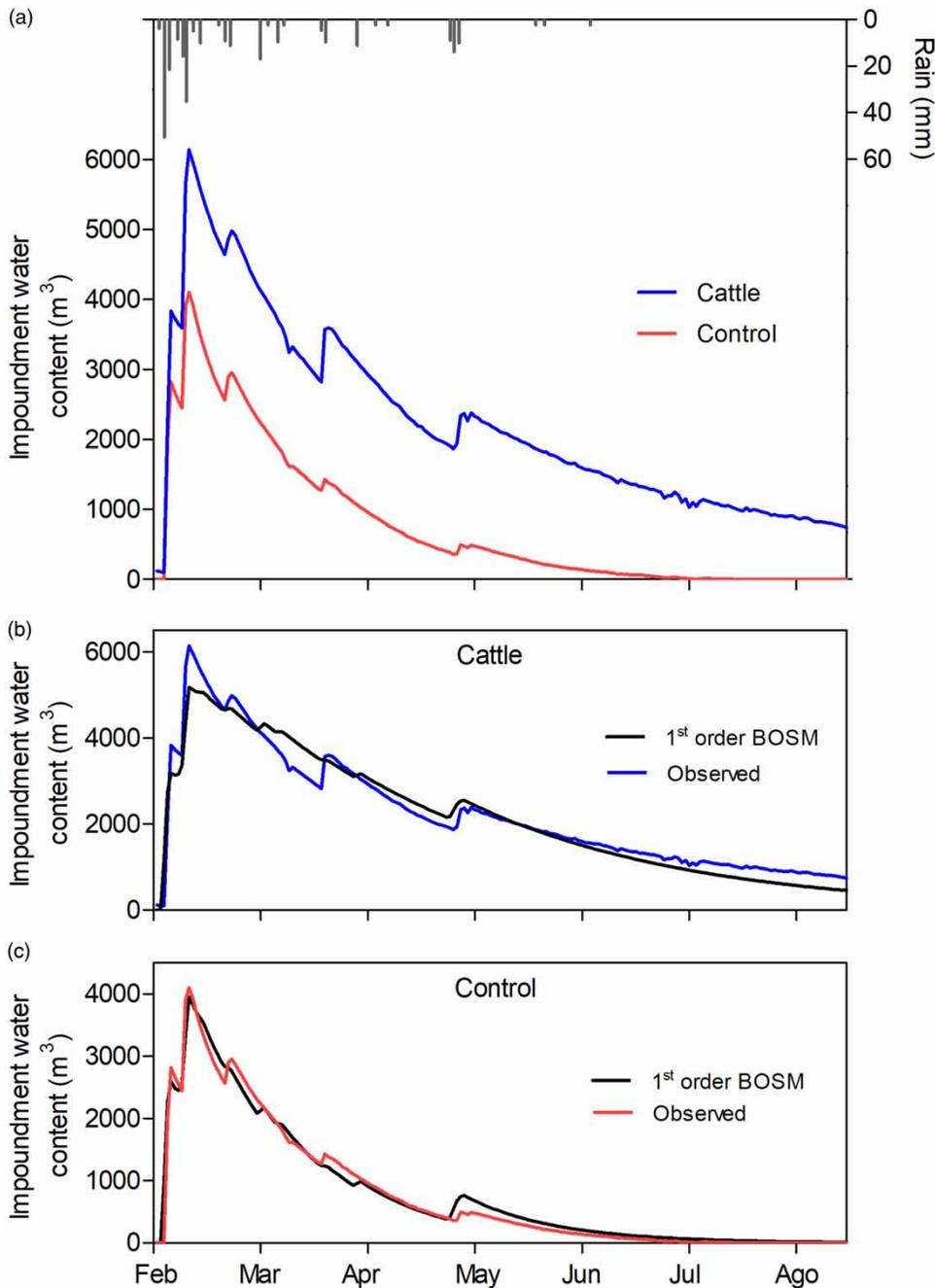


Figure 5 | Observed rainfall and impoundment volume: (a) together with the RIVC-simulated cattle- (b) and control- (c) impoundment volume.

that the use of a more slightly complex structure (i.e. adding one additional term) is warranted by the dynamics in the observed data.

The derived residence time of the response of the storage for a unit input of rainfall to the catchment (Equation (11)) was 62 days for the impoundment affected by cattle

access. In contrast, the residence time for the storage in control impoundment was only 25 days (Figure 5(c) and Table 2). Thus, despite the consumption of water by cattle drinking, the less permeable bed of the cattle-affected impoundment retained the water in the impoundment for longer. This greater retention may have been an unintended

Table 2 | The identified transfer function parameters of BOSM nonlinearity term and pure time delay with the derived dynamic response characteristic of the time constant or residence time of response for each impoundment, together with measures of simulation performance (R_t^2 , YIC)

	Cattle	Control
BOSM rainfall memory (days)	2	2
Pure time delay, τ (days)	0	0
<i>Measures of simulation performance</i>		
R_t^2	0.94469	0.98272
YIC	-10.3911	-11.888
Time constant or residence time of response (days)	62.1183	25.0948

The identification routine of the RIVC (Refined Instrumental Variable in Continuous-time) algorithm of Young (2015) was applied to the 421 days of daily rainfall and impoundment volume observations.

beneficial effect of the cattle trampling the floor of the impoundment when accessing to drink.

CONCLUSIONS

Rainwater harvesting remains essential for cattle production in the Dry Chaco rangelands of Argentina. This pioneering study into the hydrological functioning of a typical harvesting-impoundment system used for watering cattle against a similar physical system, except for cattle exclusion from the impoundment, produced the following key conclusions:

- The impoundment functioning (cattle-affected and control) is characterised by very short filling events (from infrequent rainfall events) with continuous losses of storage to evaporation and leakage (and also consumption in the cattle-affected impoundment);
- Most inflow is attributed to rainfall events of greater short-term intensity, due to the greater generation of infiltration-excess overland flow rather than infiltration;
- Some 43% of the storage was lost to open water evaporation in the cattle affected impoundment, and a similar 35% from the control impoundment;
- Using the value of 0.05 m^3 per cow multiplied by 300 cows for water consumption from within the cattle-affected impoundment, this left only 15% of the water balance to loss by leakage through the impoundment floor. In considerable contrast, leakage through the impoundment floor amounted to 65% of the

stored water in the impoundment without cattle. This greater leakage may be attributed tentatively to the absence of the beneficial effect of cattle trampling reducing the saturated hydraulic conductivity of the impoundment floor;

- Calculation of the water balance components involves uncertainty, particularly the estimates of cattle water consumption and open water evaporation, with the resultant impact on the estimate of impoundment leakage. Future studies on the hydrological functioning of these rainwater harvesting and storage systems should focus on quantifying and reducing the uncertainties in these three key variables;
- The systems modelling allowed quantification of the residence time of storage to unit inputs of rainfall, showing that the impoundment used for cattle drinking counterintuitively maintained storage longer (62 days residence time) than that of the impoundment without cattle ingress (25 days residence time). The beneficial effect of cattle trampling on the saturated impoundment floor to reducing leakage being a plausible first interpretation.

If further studies (with greater constraints on uncertainty in comparison to this pioneering study) similarly show the beneficial effect of cattle trampling on reducing leakage through the impoundment floor, this would have management implications for cattle ranchers in the Dry Chaco rangelands of Argentina. Cattle acting unconsciously as beneficial 'ecosystem engineers' in reducing impoundment leakage, in addition to their acknowledged impact of enhancing impoundment inflows by trampling effects in the contributory catchment, is a finding that cattle ranchers should be aware of. Semi-arid regions across the world are increasingly affected by livestock grazing (Schlesinger *et al.* 1990; Reynolds *et al.* 2007). It is critical that hydrologists undertake experimental studies to quantify both beneficial and negative impacts of these grazers on the sustainability of the natural environment and the food producing systems.

ACKNOWLEDGEMENTS

We thank Ann Kretzschmar, Keith Beven and Raul Giménez for their valuable suggestions. We also thank

George Castellanos and Ricardo Páez for their help with field measurements and analysis. We also thank Martin Ibarra, Alejandro Sanchez and Nicolas Rios Centeno (Ser Beef S.A.) for providing field information and for allowing us to work in the field. This work was funded by the ANPCyT-Argentina (PICT 2016-4136, PICT 2014-2790 and PICT 2013-2973), the CONICET-Argentina (PIP 112-201501-00609) and the Universities UK International (UUKi) Rutherford Fund Strategic Partner Grants (award reference RF-2018-78). Support from NERC NE/R004722/1 covered Nick Chappell's input to the project. San Luis University (Argentina), Lancaster University (United Kingdom) and San Luis Applied Maths Institute (CONICET; Argentina) provided laboratory spaces and workplaces.

REFERENCES

- Adham, A., Riksen, M., Ouassar, M. & Ritsema, C. J. 2016 [A methodology to assess and evaluate rainwater harvesting techniques in \(semi-\) arid regions](#). *Water* **8**, 1–23.
- Aguilera, M. 2003 With clear goals. The agricultural station of San Luis. In: *Livestock Use of the Natural Grasslands of San Luis* (M. Aguilera & J. Panigatti, eds). pp. 40.
- Aguilera, M. O., Steinaker, D. F. & Demaria, M. R. 2003 [Runoff and soil loss in undisturbed and roller-seeded shrublands of semiarid Argentina](#). *J. Range Manag.* **56**, 227–233.
- Allen, R. G., Pereira, L. S., Raes, D. & Smith, M. 1998 Crop evapotranspiration-Guidelines for computing crop water requirements. FAO Irrigation and Drainage Paper 56. FAO, Rome.
- Bailey, D. W. 2005 [Identification and creation of optimum habitat conditions for livestock](#). *Rangel. Ecol. Manag.* **58**, 109–118.
- Belsky, A. J., Matzke, A. & Uselman, S. 1999 Survey of livestock influences on stream and riparian ecosystems in the western United States. *J. Soil Water Conserv.* **54**, 419–431.
- Boers, T. M. & Ben-Asher, J. 1982 [A review of rainwater harvesting](#). *Agric. Water Manag.* **5**, 145–158.
- Cabrera, A. L. 1976 *Argentine Encyclopedia of Agriculture and Gardening: Argentine Phytogeographic Regions*. Acme, Buenos Aires, Argentina.
- Chappell, N. A., Tych, W., Chotai, A., Bidin, K., Sinun, W. & Chiew, T. H. 2006 [BARUMODEL: combined data based mechanistic models of runoff response in a managed rainforest catchment](#). *For. Ecol. Manag.* **224**, 58–80.
- Chillo, V., Ojeda, R. A., Anand, M. & Reynolds, J. F. 2015 [A novel approach to assess livestock management effects on biodiversity of drylands](#). *Ecol. Indic.* **50**, 69–78.
- Critchley, W., Siegert, K., Chapman, C. & Finkel, M. 1991 *Water Harvesting. A Manual for the Design and Construction of Water Harvesting Schemes for Plant Production*. Scientific Publishers, Rome, Italy.
- Denison, J. A. & Wotshela, L. 2012 [An overview of indigenous, indigenised and contemporary water harvesting and conservation practices in South Africa](#). *Irrig. Drain.* **61**, 7–23.
- Evenari, M., Shanan, L. & Tadmor, N. 1971 *The Negev. The Challenge of a Desert*, 2nd edn. Harvard University Press, USA.
- Fernández, P. L., Alvarez, C. R. & Taboada, M. A. 2015 [Topsoil compaction and recovery in integrated no-tilled crop–livestock systems of Argentina](#). *Soil Till. Res.* **153**, 86–94.
- Fox, P. & Rockström, J. 2000 [Water-harvesting for supplementary irrigation of cereal crops to overcome intra-seasonal dry-spells in the Sahel](#). *Phys. Chem. Earth B Hydrol. Oceans Atmos.* **25**, 289–296.
- Ganskopp, D. 2001 [Manipulating cattle distribution with salt and water in large arid-land pastures: a GPS/GIS assessment](#). *Appl. Anim. Behav. Sci.* **73**, 251–262.
- Ganskopp, D. C. & Bohnert, D. W. 2009 [Landscape nutritional patterns and cattle distribution in rangeland pastures](#). *Appl. Anim. Behav. Sci.* **116**, 110–119.
- Ganskopp, D., Cruz, R. & Johnson, D. 2000 [Least-effort pathways?: a GIS analysis of livestock trails in rugged terrain](#). *Appl. Anim. Behav. Sci.* **68**, 179–190.
- García, A. G., Di Bella, C. M., Houspanossian, J., Magliano, P. N., Jobbágy, E. G., Posse, G., Fernández, R. J. & Nosetto, M. D. 2017 [Patterns and controls of carbon dioxide and water vapor fluxes in a dry forest of central Argentina](#). *Agric. For. Meteorol.* **247**, 520–532.
- George, M. R., Larsen, R. E., McDougald, N. K., Tate, K. W., Gerlach, J. D. & Fulgham, K. O. 2004 [Cattle grazing has varying impacts on stream-channel erosion in oak woodlands](#). *Calif. Agric.* **58**, 138–143.
- Glendenning, C. J., Van Ogtrop, F. F., Mishra, A. K. & Vervoort, R. W. 2012 [Balancing watershed and local scale impacts of rain water harvesting in India – a review](#). *Agric. Water Manag.* **107**, 1–13.
- Huxman, T. E., Snyder, K. A., Tissue, D., Leffler, A. J., Ogle, K., Pockman, W. T., Sandquist, D. R., Potts, D. L. & Schwinning, S. 2004 [Precipitation pulses and carbon fluxes in semiarid and arid ecosystems](#). *Oecologia* **141**, 254–268.
- Iriondo, M. 1993 [Geomorphology and late quaternary of the Chaco \(South America\)](#). *Geomorphology* **7**, 289–303.
- Jobbágy, E. G., Nosetto, M. D., Santoni, C. S. & Baldi, G. 2008 The ecohydrological challenge of woody-herbaceous transitions in the Chaco-Pampas plains. *Ecol. Austral.* **18**, 305–322.
- Karlin, M. S., Karlin, U. O., Coirini, R. O., Reati, G. J. & Zapata, R. M. 2013 *The Arid Chaco*. Universidad Nacional de Córdoba, Córdoba, Argentina, 420 pp.
- Knapp, A. K., Hoover, D. L., Wilcox, K. R., Avolio, M. L., Koerner, S. E., La Pierre, K. J., Loik, M. E., Luo, Y., Sala, O. E. & Smith, M. D. 2015 [Characterizing differences in precipitation regimes of extreme wet and dry years: implications for climate change experiments](#). *Glob. Change Biol.* **21** (7), 2624–2633.
- Kunst, C. R., Ledesma, M., Basan Nickish, G., Angella, D. & Godoy, J. P. 2003 Roller chopping and water infiltration into

- the soil in the western Chaco (Argentina). *Rev. Invest. Agr.* **32**, 105–126.
- Lavee, H., Poesen, J. & Yair, A. 1997 Evidence of high efficiency water-harvesting by ancient farmers in the Negev Desert, Israel. *J. Arid Environ.* **35**, 341–348.
- Lebron, I., Madsen, M., Chandler, D., Robinson, D., Wendroth, O. & Belnap, J. 2007 Ecohydrological controls on soil moisture and hydraulic conductivity within a pinyon-juniper woodland. *Water Resour. Res.* **43**, 1–15.
- Macchi, L. & Grau, H. R. 2012 Piospheres in the dry Chaco. Contrasting effects of livestock puestos on forest vegetation and bird communities. *J. Arid Environ.* **87**, 176–187.
- Magliano, P. N., Breshears, D. D., Fernández, R. J. & Jobbágy, E. G. 2015a Rainfall intensity switches ecohydrological runoff/runon redistribution patterns in dryland vegetation patches. *Ecol. Appl.* **25**, 2094–2100.
- Magliano, P. N., Fernández, R. J., Mercau, J. L. & Jobbágy, E. G. 2015b Precipitation event distribution in central Argentina: spatial and temporal patterns. *Ecohydrology* **8**, 94–104.
- Magliano, P. N., Murray, F., Baldi, G., Aurand, S., Páez, R. A., Harder, W. & Jobbágy, E. G. 2015c Rainwater harvesting in Dry Chaco: regional distribution and local water balance. *J. Arid Environ.* **123**, 93–102.
- Magliano, P. N., Fernández, R. J., Giménez, R., Marchesini, V. A., Páez, R. A. & Jobbágy, E. G. 2016 Changes in water fluxes partition in the Arid Chaco caused by the replacement of forest by pastures. *Ecolog. Austral.* **26**, 95–106.
- Magliano, P., Giménez, R., Houspanossian, J., Páez, R., Nosetto, M., Fernández, R. & Jobbágy, E. 2017a Litter is more effective than forest canopy reducing soil evaporation in Dry Chaco rangelands. *Ecohydrology* **10**, e(1879).
- Magliano, P. N., Fernández, R. J., Florio, E. L., Murray, F. & Jobbágy, E. G. 2017b Soil physical changes after conversion of woodlands to pastures in Dry Chaco rangelands (Argentina). *Rangel. Ecol. Manage.* **70**, 225–229.
- Marchesini, V. A., Fernández, R. J., Reynolds, J. F., Sobrino, J. A. & Di Bella, C. M. 2015 Changes in evapotranspiration and phenology as consequences of shrub removal in dry forests of central Argentina. *Ecohydrology* **8**, 1304–1311.
- New, M., Lister, D., Hulme, M. & Makin, I. 2002 A high-resolution data set of surface climate over global land areas. *Clim. Res.* **21**, 1–25.
- Newman, B. D., Wilcox, B. P., Archer, S. R., Breshears, D. D., Dahm, C. N., Duffy, C. J., McDowell, N. G., Phillips, F. M., Scanlon, B. R. & Vivoni, E. R. 2006 Ecohydrology of water-limited environments: a scientific vision. *Water Resour. Res.* **42**, 1–15.
- Oweis, T. & Hachum, A. 2009 Water harvesting for improved rainfed agriculture in the dry environments. In: *Rainfed Agriculture: Unlocking the Potential* (S. P. Wani, ed.). CAB International, London, UK, pp. 164–182.
- Oyarzabal, M., Clavijo, J., Oakley, J., Biganzoli, F., Tognetti, P., Barberis, I., Maturo, H., Aragon, R., Campanello, P., Prado, D., Oesterheld, M. & Leon, R. J. C. 2018 Vegetation units of Argentina. *Ecolog. Austral.* **28**, 40–63.
- Pandey, D. N., Gupta, A. K. & Anderson, D. M. 2003 Rainwater harvesting as an adaptation to climate change. *Curr. Sci.* **85**, 46–59.
- Peña Zubiate, C. A., Anderson, D. L., Demmi, M. A., Saenz, J. L. & D'Hiriart, A. 1998 *Soils and Vegetation of San Luis Province*. Instituto Nacional de Tecnología Agropecuaria, San Luis, Argentina.
- Pennington, T. R., Prado, D. E. & Pendry, C. A. 2000 Neotropical seasonally dry forests and quaternary vegetation changes. *J. Biogeogr.* **27**, 261–275.
- Reynolds, J. F., Stafford Smith, D. M., Lambin, E. F., Turner II, B. L., Mortimore, M., Batterbury, S. P. J., Downing, T. E., Dowlatabadi, H., Fernández, R. J., Herrick, J. E., Huber-Sannwald, E., Jiang, H., Leemans, R., Lynam, T., Maestre, F. T., Ayarza, M. & Walker, B. 2007 Ecology: global desertification: building a science for dryland development. *Science* **316**, 847–851.
- Richardson, D. M. & Rejmánek, M. 2011 Trees and shrubs as invasive alien species – a global review. *Divers. Distrib.* **17** (5), 788–809.
- Robinson, T. P., Wint, G. W., Conchedda, G., Van Boeckel, T. P., Ercoli, V., Palamara, E., Cinardi, G., D'Aiotti, D., Hay, S. I. & Gilbert, M. 2014 Mapping the global distribution of livestock. *PLoS one* **9**, e96084.
- Schimel, D. S. 2010 Drylands in the earth system. *Science* **327**, 418–419.
- Schlesinger, W. H., Reynolds, J. F., Cunningham, G. L., Huenneke, L. F., Jarrell, W. M., Virginia, R. A. & Whitford, W. G. 1990 Biological feedbacks in global desertification. *Science* **247**, 1043–1048.
- Schlink, A., Nguyen, M. & Viljoen, G. 2010 Water requirements for livestock production: a global perspective. *Rev. Sci. Tech.* **29**, 603–619.
- Steinaker, D. F., Jobbágy, E. G., Martini, J. P., Arroyo, D. N., Pacheco, J. L. & Marchesini, V. A. 2016 Vegetation composition and structure changes following roller-chopping deforestation in central Argentina woodlands. *J. Arid Environ.* **133**, 19–24.
- Taylor, C. J., Pedregal, D. J., Young, P. C. & Tych, W. 2007 Environmental time series analysis and forecasting with the Captain toolbox. *Environ. Model. Softw.* **22**, 797–814.
- Tripaldi, A., Zárate, M. A., Forman, S. L., Badger, T., Doyle, M. E. & Ciccio, P. 2013 Geological evidence for a drought episode in the western Pampas (Argentina, South America) during the early–mid 20th century. *Holocene* **23**, 1731–1746.
- UNEP 2009 *Rainwater Harvesting: A Lifeline for Human Well-Being*. United Nations Environment Programme, Stockholm, Sweden.
- Young, P. C. 2015 Refined instrumental variable estimation: maximum likelihood optimization of a unified Box–Jenkins model. *Automatica* **52**, 35–46.