Assessment of the Internal Dynamics of the Australian Water Balance Model under Different Calibration Regimes

Li Li a, Martin F Lambert a, Holger R Maier a, Daniel Partington b, and Craig T Simmons b

a. School of Civil, Environmental and Mining Engineering, The University of Adelaide, Adelaide, South Australia, AUSTRALIA 5005

l.li@alumni.adelaide.edu.au Tel: +61 0406745484

holger.maier@adelaide.edu.au Tel: +61 (0)8 8313 4139

martin.lambert@adelaide.edu.au Tel: +61 (0)8 8313 4320

Fax: +61 (0)8 83034359

b. National Centre for Groundwater Research and Training and School of the Environment, Flinders University, GPO Box 2100, Adelaide, South Australia, AUSTRALIA 5001

daniel.partington@flinders.edu.au Tel: +61 (0)8 8201 2648

craig.simmons@flinders.edu.au Tel: +61 (0)8 8201 5509

Fax: (+61 8) 8201 7906

Key Words: Australian Water Balance Model, calibration, baseflow, regression models, HydroGeoSphere, virtual laboratory, integrated surface water/groundwater model

This paper has been published in Environmental Modelling and Software:

Abstract

Conceptual rainfall runoff models are used extensively in practice, as they provide a good balance between transparency and computational and data requirements. However, the degree to which they are able to represent underlying physical processes is poorly understood. This is because the performance of such models is generally assessed based on their ability to match total streamflow, rather than component processes. In this paper, the ability of the Australian Water Balance Model (AWBM) to represent baseflow and quickflow is assessed for 66 synthetic catchments with different physical characteristics and hydrological inputs under seven calibration regimes utilising a shuffled complex evolution (SCE) algorithm. The “observed” total-, base- and quick-flow hydrographs for these catchments are generated using HydroGeoSphere. The results indicate that while AWBM is generally able to match total streamflow well, the same does not apply to baseflow and quickflow, suggesting that these processes are not represented well by AWBM.
1. Introduction

Modelling approaches for estimating runoff from rainfall and evapotranspiration (ET) can be traditionally classified into three main groups: black box models, physical process based models and conceptual rainfall runoff (CRR) models (Beven, 2005). While all of these approaches have been shown to be able to predict total streamflow successfully, the degree to which they are able to represent underlying streamflow generating mechanisms is highly variable (Chen and Adams, 2006; Ferket et al., 2010; Refsgaard and Knudsen, 1996).

Black box modelling approaches, such as artificial neural networks (Abrahart et al., 2012; Maier et al., 2010; Wu et al., 2014), are at one end of the spectrum, while physically based approaches are at the other end. Black box models produce streamflow outputs solely as a function of their inputs and transfer characteristics, without any knowledge or understanding of the underlying physical processes. However, they are generally computationally efficient and can be developed using limited data. Physically based approaches attempt to simulate the detailed mechanisms of the component physical processes within the hydrologic cycle using well-established physical laws, with numerical solutions of the mathematical representation of these processes (Jayatilaka et al., 1998). Such approaches include fully integrated surface water/ground water (SW/GW) models, such as InHM (VanderKwaak and Loague, 2001), MODHMS (HydroGeoLogic, 2000), HydroGeoSphere (HGS) (Therrien et al., 2009) and SHE (Abbott et al., 1986). However, there are problems with the application of these models in practice due to the difficulties and expense associated with obtaining the data required (e.g. due to limitations of existing instrumentation and intrinsic uncertainty in measurements), as well as their high computational demands.

CRR models represent a compromise between the high data and computational requirements of physical process based models and the lack of transparency of black-box models. They are more computationally efficient and less data intensive than process based models, as they do not attempt to represent all physical processes explicitly. However, they are more transparent than black-box models, as they represent the assumed underlying physical processes in a conceptual manner, generally in the form of a number of interconnected storages that are linked with empirical mathematical equations to conceptualise the movement of water into, between the storages of and out of a catchment. Many different CRR models have been proposed in the literature, such as the Australian Water Balance Model (AWBM) (Boughton, 1993, 2004), the Soil Moisture and Accounting Model (SMAR) (Tuteja and Cunnane, 1999), SIMHYD (Chiew et al., 2002) and GR4J (Oudin et al., 2005), for example. These models have been able to capture enough of the dynamics of rainfall runoff simulation time series to be useful in water resources assessments.

Among the different rainfall-runoff modelling approaches mentioned above, CRR models are the most widely utilized in practice, due to their relatively simple structure, small number of parameters and production of generally acceptable results. A common feature of CRR models is that some of their model parameters have limited physical interpretation (Delleur, 1982; Troutman, 1985), due to the fact that many complex catchment physical processes are lumped together. In addition, although some of the CRR model parameters, representing the physical properties of the catchment (e.g. catchment area, surface slope), are usually measurable, there are some physical parameters, such as hydraulic conductivity and porosity, which are measurable in theory, but difficult to measure in practice. Therefore, such CRR model parameters are generally estimated by calibration, by comparing the modelled total streamflow time series with the corresponding observed data until an acceptable fit to the objective function, or an acceptable trade-off between objective functions in cases where multi-objective optimisation is used (e.g. Ahmadi et al., 2014; Gibbs et al., 2012), has been obtained. Significant research effort has been directed towards obtaining a well-defined optimal parameter set, including local-type direct search optimisation methods.
and globally based optimisation methods (Duan et al., 1992). However, because CRR models are usually calibrated using only observed total streamflow time series, while internally they calculate a number of additional states and fluxes, such as baseflow and quickflow, there may be many combinations of parameter values that give similar objective function values. This phenomenon is called ‘equifinality’ (Beven, 1993), and is caused by problems such as over-parameterisation of models, data limitations and structural faults in the model. As a result, even though the structure of CRR models is based on a conceptual representation of underlying physical processes, how well these processes are represented by calibrated models is generally unknown, as a good match to total streamflow does not necessarily mean that the component processes are modelled accurately. For example, similar total streamflow time series can be obtained with very different combinations of baseflow and quickflow, without consideration of the appropriateness of their quantities and dynamics.

While there have been many studies comparing the performance of CRR models using total streamflow time series (Ferket et al., 2010; Knapp et al., 1991; Post et al., 2007; Ranatunga et al., 2008), very few attempts have been made to use baseflow or quickflow estimates for CRR model internal dynamic performance assessment, due to the difficulty of accurately measuring baseflow or quickflow in the field (Dukic, 2006; McCallum et al., 2010). Recently, Ferket et al. (2010) used baseflow estimated from a physically-based digital filter (Furey and Gupta, 2001) to validate the internal dynamics of two CRR models (HBV and PDM) for a subcatchment of the Dender catchment in Belgium. As part of the study, two optimisation algorithms (SCE-UA and MWARPE) are used to calibrate the models by matching total streamflow to observations. They conclude that no clear picture emerges of which model produces the best results of simulating total streamflow, but that the MWARPE calibration algorithm and the HBV model leads to the best baseflow estimates, giving the best internal model dynamics, at least when compared with the results obtained using the Furey and Gupta filter (Furey and Gupta, 2001).

This study builds on the research by Ferket et al. (2010) by assessing (i) how well the Australian Water Balance Model, which is a commonly used CRR, is able to represent total-, base- and quick-flow for 66 synthetic catchments with different catchment characteristics and hydrological inputs and (ii) the impact of seven different calibration regimes that take internal model dynamics into account in different ways on the accuracy of total-, base- and quick-flow hydrograph prediction. While the methodology is illustrated for a particular case study, its generic nature means it could easily be adapted and applied to other CRR models around the world. The remainder of this paper is organized as follows. The methodology is given in Section 2, followed by the results and discussion in Section 3 and summary and conclusions in Section 4.

2. Methodology

The underlying premise of the proposed methodology for assessing the internal dynamics of CRR models is that fully integrated SW/GW models can be used to obtain reasonably accurate estimates of actual total streamflow, quickflow and baseflow (see Partington et al., 2012; Li et al., 2013; Li et al., 2014), thereby providing a benchmark against which the internal dynamics of CRR models can be assessed (e.g. whether flow components that make up total streamflow are predicted accurately). This is a reasonable assumption, as fully integrated SW/GW models provide a rigorous representation of the underlying physical processes of hydrologic systems (Brookfield et al., 2009; Furman, 2008; Partington et al., 2012; Sulis et al., 2010; Therrien and Sudicky, 1996). They typically represent 3D variably saturated subsurface flow with the Richards’ equations, and 1D and 2D surface flow with the diffusion wave approximation to the St. Venant equations. A unique feature is that such models can simulate the partitioning of rainfall into different components, including overland flow, streamflow, evaporation, infiltration and recharge, as well as subsurface discharge to surface water features (e.g. lakes and streams), in a physically realistic fashion (Therrien et al., 2009). All of the governing flow equations implemented by fully integrated SW/GW
models are solved simultaneously to obtain total streamflow, baseflow and quickflow, making them ideal candidates for assessing the internal dynamics of CRR models.

While it is acknowledged that fully integrated SW/GW models are in themselves an approximation of the actual processes in real catchments, they provide the best means of quantifying the absolute volume of the flow components (e.g. baseflow and quickflow) currently available (see Partington et al., 2012; Li et al., 2013; Li et al., 2014). In addition, they can be used to obtain estimates of different flow components for catchments with different characteristics (see Partington et al., 2013). Therefore they are able to provide the first step towards being able to assess the internal dynamics of CRR models under a range of physical conditions in a controlled manner.

The steps in the methodology adopted for assessing the internal dynamic performance of AWBM are given in Fig. 1. As shown, initially synthetic total streamflow (\(q_{obs}^{Tq}\)), baseflow (\(q_{obs}^{Bq}\)) and quickflow (\(q_{obs}^{Qq}\)) hydrographs are generated using a fully integrated SW/GW model for a number of catchments with different physical properties and hydrological inputs in order to ensure the results are as generic as possible. Next, AWBMs are developed for the same catchments by using the same hydrological inputs, but different calibration methods. Finally, the performance of the AWBMs calibrated using the different methods is compared in terms of the ability to predict total-, base- and quick-flow hydrographs accurately, thereby providing a means of assessing the performance of the internal dynamics of AWBM (e.g. whether flow components that make up total streamflow are predicted accurately). Details of each step in the methodology are given in subsequent sections. It should be noted that while the AWBM is used as the CRR model in this study, the same approach should also be used to test the internal dynamic performance of other CRR models.
2.1 Catchment Characteristics and Hydrological Inputs

The 66 synthetic catchments with different physical characteristics and hydrological inputs developed by Li et al. (2014) are used. These catchments have drainage areas ranging from 6 to 192 km² and are loosely based on a benchmarked integrated surface-subsurface hydrology problem, the V-catchment test case, as shown in Fig. 2 (Panday and Huyakorn, 2004). A detailed description of these catchments can be found in Li et al. (2014), thus only a brief overview is provided here. Different physical catchment characteristics are represented using seven variables: catchment area (A), catchment hill slope (S1, which is perpendicular to the channel), catchment channel slope (S2, which is parallel to the channel), catchment aspect ratio (AR), and soil type, which includes Ks and van Genuchten parameters α and β. The hydrological inputs are represented using the ratio of daily rainfall to ET from five Australian cities (Li et al., 2014), which are obtained from the Australian Bureau of Meteorology National Climate Centre. Details of the different values of the physical catchment characteristics considered are given in Table 1 and the characteristics of the hydrological inputs used are given in Table 2. Li et al. (2014) used Latin Hypercube Sampling (LHS) to generate the 66 catchments with different physical catchment characteristics and hydrological inputs, which is also used in this study by sampling from these catchment characteristics and hydrological inputs.

Fig. 2 Schematic representation of tilted V-catchment flow problem (adopted from Panday and Huyakorn (2004))
Table 1 Catchment characteristics considered (adopted from (Li et al., 2014))

<table>
<thead>
<tr>
<th>Catchment Characteristic</th>
<th>Unit</th>
<th>Explanation</th>
<th>Values Considered</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_s$</td>
<td>m/s</td>
<td>Saturated hydraulic conductivity</td>
<td>$2.44 \times 10^{-05}$, $3.99 \times 10^{-05}$, $1.12 \times 10^{-04}$, $2.11 \times 10^{-04}$, $9.70 \times 10^{-04}$</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>-</td>
<td>van Genuchten parameter</td>
<td>0.572, 3.370, 6.161, 8.955, 11.75</td>
</tr>
<tr>
<td>$\beta$</td>
<td>-</td>
<td>van Genuchten parameter</td>
<td>1.32, 1.556, 1.793, 2.029, 2.270</td>
</tr>
<tr>
<td>$A$</td>
<td>$\text{km}^2$</td>
<td>Catchment area</td>
<td>6, 48, 80, 120, 192</td>
</tr>
<tr>
<td>$S_1$</td>
<td>-</td>
<td>Hill slope (perpendicular to the channel)</td>
<td>0.005, 0.008, 0.012, 0.016, 0.02</td>
</tr>
<tr>
<td>$S_2$</td>
<td>-</td>
<td>Channel slope (along the channel)</td>
<td>0.0025, 0.004, 0.006, 0.008, 0.01</td>
</tr>
<tr>
<td>$AR$</td>
<td>-</td>
<td>Ratio of catchment width to length (x/y)</td>
<td>0.5, 0.75, 1.0, 1.25, 1.5</td>
</tr>
</tbody>
</table>

Table 2 Hydrological inputs considered (adopted from (Li et al., 2014))

<table>
<thead>
<tr>
<th>City</th>
<th>Gauge No.</th>
<th>Average annual rainfall (mm/a)</th>
<th>Rainfall data period for sampling</th>
<th>Average annual potential ET (mm/a)</th>
<th>R/ET (average annual rainfall/average annual potential ET)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adelaide</td>
<td>023011</td>
<td>510.16</td>
<td>1984-1994</td>
<td>1470.97</td>
<td>0.347</td>
</tr>
<tr>
<td>Melbourne</td>
<td>086071</td>
<td>525.33</td>
<td>1968-1978</td>
<td>911.33</td>
<td>0.576</td>
</tr>
<tr>
<td>Sydney</td>
<td>070062</td>
<td>1095.85</td>
<td>1982-1992</td>
<td>920.90</td>
<td>1.190</td>
</tr>
<tr>
<td>Brisbane</td>
<td>031011</td>
<td>2238.22</td>
<td>1997-2007</td>
<td>2077</td>
<td>1.078</td>
</tr>
<tr>
<td>Darwin</td>
<td>014015</td>
<td>1707.16</td>
<td>1999-2009</td>
<td>2056.37</td>
<td>0.811</td>
</tr>
</tbody>
</table>

2.2 Fully Integrated SW/GW Model

As stated above, fully integrated SW/GW models are used for assessing the internal dynamic performance of AWBM under a range of calibration approaches. This is because they provide the best possible approximation to the physical processes of water flow within catchments and can therefore be used as an approximation to such processes subject to a variety of physical characteristics and forcings. In this study, Hydrogeosphere (HGS) is used as the fully integrated SW/GW model to model the 66 synthetic catchments’ response to rainfall and ET inputs under different catchment characteristics and hydrological inputs, which are assumed to represent the “true” catchment physical processes. This is because HGS can be used to simulate hydrological processes within catchments in a physically based manner (Therrien et al., 2009). HGS has been applied successfully to various studies, such as the comparison of baseflow estimation methods (Li et al., 2013; Partington et al., 2011; Partington et al., 2012), SW/GW disconnection problems (Banks et al., 2011; Brunner et al., 2009), bank storage dynamic processes analysis (Doble et al., 2012) and the study of dual permeability systems (Schwartz et al., 2010). Further description of the code and its numerical formulation can be found in Therrien et al. (2009) and Brunner and Simmons (2012).

As shown in Fig. 2, the catchment used in this study is symmetrical. As a result, all simulations are conducted for only half of the catchment, as shown for one example of the catchment configurations considered in Fig. 3, and the reported fluxes are correspondingly half of those expected when accounting for both sides of the stream. Detailed information of the catchment model with the selected values of the HGS parameters can be found in Li et al. (2014).
As mentioned previously, the HGS models are used to obtain the ‘observed’ streamflow ($q_T^{\text{obs}}$), baseflow ($q_B^{\text{obs}}$) and quickflow ($q_Q^{\text{obs}}$) hydrographs for the 66 catchments. The baseflow hydrograph ($q_B^{\text{obs}}$) is extracted from the model using the Hydraulic Mixing-Cell (HMC) method (Partington et al., 2013; Partington et al., 2011).

The HMC method developed by Partington et al. (2011) utilises the spatiotemporal information of flow generation mechanisms to obtain the component hydrographs (i.e. baseflow and quickflow hydrographs). When combining the HMC method with the fully integrated SW/GW model, each node in the surface domain of the model is treated as a mixing-cell. The method uses the fluid mass balance from the fully integrated SW/GW model of each node at each model time step to calculate the fraction of water in the cell that comes from different streamflow generation mechanisms (e.g. groundwater discharge to the stream). For each component of streamflow, the fraction is determined using the modified mixing rule (Campana and Simpson, 1984). The equation for each fraction $f$ for each streamflow generation mechanism $k$ at time $N$ in cell $i$ is given by Partington et al. (2011) as:

$$f_{i(k)}^N = \left(\frac{V_{i}^{N-1}}{V_{i}^{N}} - \frac{\sum_{j=1}^{n} V_{i,j}^{N-1} f_{j(k)}^{N-1}}{V_{i}^{N}}\right) + \frac{\sum_{j=1}^{n} V_{j,i}^{N-1} f_{j(k)}^{N-1}}{V_{i}^{N}}$$

Where there are $n$ sources and $m$ sinks for cell $i$; $f_{j(k)}^{N-1}$ denotes fraction $k$ at time $N-1$ in the neighbouring cell $j$; $V$ denotes the volume with the superscript denoting time state and subscript $i$ denoting the cell, $ij$ denoting volume into cell $j$ from cell $i$ over the time step from $N-1$ to $N$ and $ji$ denoting volume from neighbor $j$ into $i$. 

Fig. 3 An example of the 3D catchment model for the case study (adopted from Li et al. (2014))
The HMC method extracts streamflow generation mechanisms using only hydraulic information and is able to capture the storage effects and time lags within catchments (see Partington et al., 2013), which provides a means of estimating baseflow from fully integrated SW/GW models. The quickflow \( q_{q}^{obs} \) is taken as the difference between the total streamflow \( q_{T}^{obs} \) and baseflow \( q_{B}^{obs} \).

The dataset of the 66 catchment characteristics and hydrological inputs, as well as the corresponding simulated total streamflow and baseflow hydrographs obtained using the integrated SW/GW model, can be downloaded as supplementary material.

2.3 Australian Water Balance Model (AWBM)

The AWBM is a saturation overland flow model developed by Boughton (1993, 2004) and is now one of the most widely used CRR models in Australia (Marshall et al., 2004; Ranatunga et al., 2008). The AWBM is a typical lumped CRR model, with interconnected storages and algorithms that mimic the underlying hydrological processes used to describe the movement of water into and out of storages. AWBM uses three surface stores, representing the impacts of antecedent wetness and spatial variability of the abstractions, for modelling rainfall-runoff relationships. It also adopts the structure of transferring a fraction of the generated runoff direct to the baseflow store at the same time as the residual is transferred to the surface attenuation store. AWBM can be operated at either daily or hourly time steps, although a daily time step is used in this study. At each time step, rainfall is added to each of the surface stores and evapotranspiration is subtracted. If there is any excess from any store, it becomes runoff and is divided between surface runoff (quickflow) and baseflow, and total streamflow is the sum of surface runoff and baseflow (Boughton et al., 2004). As the aim of the AWBM is to represent the underlying physical processes, it is a reasonable assumption that it should be able to reproduce the dynamics of the rainfall-runoff relationships generated synthetically using HGS.

The detailed structure of the AWBM is given in Fig. 4. As can be seen, the AWBM uses rainfall and actual evapotranspiration as inputs and principally consists of a configuration of three different surface storages. The depths of these storages correspond to the parameters C1, C2 and C3. A fraction of the total area is associated with each surface storage, as represented by the parameters A1, A2 and A3. Moisture capacity variation over the catchment is described by the combination of the surface storages and the related fractional areas. An important feature of AWBM is the ability to account for baseflow when predicting streamflow by using a baseflow index (BFI), which is the ratio of the amount of baseflow to the total amount of streamflow and determines the proportion of excess moisture at each time step that is returned to the baseflow. The daily baseflow recession constant (KBase) and daily routed surface runoff recession constant (KSurf) are used to describe the daily discharge from baseflow and surface runoff (quickflow) storage. A summary of the AWBM parameters and their ranges used is given in Table 3. A number of software frameworks have been developed for hydrological modelling and are in active use, such as the Rainfall Runoff Library (RRL) and hydromad (Hydrological Model Assessment and Development) package (Andrews et al., 2011). In this study, AWBM is implemented using the RRL developed by the Cooperative Research Centre on Catchment Hydrology (www.toolkit.net.au/rrl).
**Table 3 AWBM parameter description and ranges**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Parameter range</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1*</td>
<td>Partial Area</td>
<td>0-1</td>
</tr>
<tr>
<td>A2*</td>
<td>Partial Area</td>
<td>0-1</td>
</tr>
<tr>
<td>BFI</td>
<td>Baseflow Index</td>
<td>0-1</td>
</tr>
<tr>
<td>C1</td>
<td>Surface Storage Capacity</td>
<td>0-50</td>
</tr>
<tr>
<td>C2</td>
<td>Surface Storage Capacity</td>
<td>0-200</td>
</tr>
<tr>
<td>C3</td>
<td>Surface Storage Capacity</td>
<td>0-500</td>
</tr>
<tr>
<td>KBase</td>
<td>Daily Baseflow Recession Constant</td>
<td>0-1</td>
</tr>
<tr>
<td>KSurf</td>
<td>Daily Surface Flow Recession</td>
<td>0-1</td>
</tr>
</tbody>
</table>

*A1+A2 must be less than or equal to 1.0.

### 2.4 Calibration of Australian Water Balance Model (AWBM)

As stated previously, the effectiveness of calibration methods that take internal model dynamics into account in different ways is investigated in this study. The calibration methods are summarised in Section 2.4.1. Details of the optimisation method and error measure used in the calibration methods are given Sections 2.4.2 and 2.4.3, respectively.
2.4.1 Calibration methods

Values of all of the eight parameters listed in Table 3 are obtained as part of the different calibration methods investigated. Five of these parameters (A1, A2, C1, C2 and C3) are related to the moisture capacity of the catchment, two (BFI and KBase) affect the separation of rainfall excess into quickflow/baseflow storage and the baseflow component of total streamflow, and KSurf affects the routed surface runoff, which is referred to as quickflow in this study. Details of the calibration methods are given in Table 4 and Fig. 5.

<table>
<thead>
<tr>
<th>Calibration method</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Calibrate all parameters simultaneously to total streamflow.</td>
</tr>
<tr>
<td>2</td>
<td>First calibrate BFI and KBase to baseflow obtained using LH filter with default filter parameter (0.925) and then calibrate remaining parameters to total streamflow.</td>
</tr>
<tr>
<td>3</td>
<td>First calibrate BFI and KBase to baseflow obtained using LH filter with optimal filter parameters and then calibrate remaining parameters to total streamflow.</td>
</tr>
<tr>
<td>4</td>
<td>First calibrate BFI and KBase to baseflow obtained using Eckhart filter with optimal filter parameters and then calibrate remaining parameters to total streamflow.</td>
</tr>
<tr>
<td>5</td>
<td>Estimate BFI and KBase based on regression relationships with catchment characteristics and hydrological inputs and then calibrate remaining parameters to total streamflow.</td>
</tr>
<tr>
<td>6</td>
<td>Use the regression relationship developed by Li et al. (2014) to determine whether the LH filter can provide satisfactory estimates of baseflow. If so, use Method 3. If not, use calibration Method 5.</td>
</tr>
<tr>
<td>7</td>
<td>Identical to Method 6, except that Method 4 is used instead of Method 3.</td>
</tr>
</tbody>
</table>
In order to be able to assess the impact of calibration methods that take internal model dynamics into account in different ways on the ability to predict total-, base- and quick-flow hydrographs, a calibration approach that does not take internal model dynamics into account explicitly is used to provide a basis of comparison (Method 1).

**Method 1:**
In Method 1, all parameters are calibrated simultaneously so as to minimise the selected error measure between the total streamflow obtained using AWBM \( q_{\text{sim}} \) and that obtained using HGS \( q_{\text{obs}} \) using the selected optimisation method, as shown in Fig. 5. The aim of this calibration method is to provide the best possible fit to total streamflow, without any consideration of internal model dynamics.

As can be seen from Fig. 5, in order to take internal model dynamics into account explicitly during calibration, six different methods (Methods 2 to 7) are used. In each of these methods, a two-step process is adopted as follows:

1. In the first step, values of two of the model parameters, BFI and KBase, are estimated. BFI and KBase are selected for this step as they control the separation of rainfall excess into quickflow/baseflow storage and the baseflow component of total streamflow, respectively, and are therefore the two parameters that have a direct effect on the baseflow hydrograph produced by AWBM. Consequently, by estimating values of these parameters first, an attempt is made to obtain the best possible match to the baseflow hydrograph.
2. In the second step, BFI and KBase are fixed at the values obtained in the first step and the remaining parameters are calibrated simultaneously so as to minimise the selected error measure between the total streamflow obtained using AWBM \(q_T^{\text{sim}}\) and that obtained using HGS \(q_T^{\text{obs}}\) using the selected optimisation method.

The methods used for obtaining values of BFI and KBase for Methods 2 – 7 from the above step 1 can be divided into three broad categories. Methods 2 to 4 belong to the first category, in which BFI and KBase are calibrated to baseflow extracted from total HGS streamflow using recursive digital filters (RDFs). In order to do this, KSurf, which determines the magnitude of the quickflow component within total streamflow (Fig. 4), is set as 1.0, so that the total streamflow produced by AWBM is comprised solely of baseflow. Method 5 belongs to the second category, in which BFI and KBase are estimated using regression relationships with catchment characteristics and hydrological inputs. Methods 6 and 7 belong to the third category, in which BFI and KBase are either estimated by calibration to baseflow extracted using RDFs, as in Methods 2 to 4, or by regression, as in Method 5, based on the predicted accuracy of the RDF. Details of the different methods for estimating BFI and KBase in the first step of the two-step process described above are given below.

**Method 2:**
In Method 2, BFI and KBase are calibrated simultaneously so as to minimise the selected error measure between the baseflow obtained using AWBM \(q_B^{\text{sim}}\) and that obtained using the Lyne and Hollick (LH) RDF (Nathan and McMahon, 1990), with the filter parameter set to its default value of 0.925 \((q_B^{\text{obs-LH (0.925)}})\), using the selected optimisation method. The LH filter is used to obtain the baseflow hydrograph as it is one of the most widely used RDFs and has been found to give good results in a number of case studies (Arnold and Allen, 1999; Li et al., 2014; Murphy et al., 2009). It should be noted that the LH filter with a filter parameter value of 0.925 is adopted by Nathan and McMahon (1990) based on comparison of the results of this filter with manual methods, and this value has been widely adopted by a number of researchers since. When people have actually tried to measure baseflow (although in only a few catchments), a value of the LH filter parameter of 0.98 is found to produce a better fit (CSIRO and SKM, 2010). Although LH filter parameter values in the range of 0.9-0.98 are found to be reasonable for real catchments, 0.925 is the most commonly accepted value in most published studies. During the calibration process using Method 2, the remaining parameters are set to the optimal values obtained using Method 1.

**Method 3:**
Method 3 is identical to Method 2, except that the LH filter parameter is estimated using the regression relationship developed by Li et al (2014), which provides an estimate of the optimal filter parameter as a function of catchment characteristics and hydrological inputs, rather than using the default value of 0.925. Consequently, the “observed” baseflow is given by \(q_B^{\text{obs-LH (opt)}}\).

**Method 4:**
Method 4 is identical to Method 3, except that the Eckhart filter (Eckhardt, 2005) is used to obtain the hydrograph of “observed” baseflow \(q_B^{\text{obs-Eck (opt)}}\), rather than the LH filter. The Eckhart filter is used as it is mathematically identical to the Boughton filter (Boughton, 1993), which has been recommended for the purposes of estimating BFI for AWBM by Boughton (2004), but has a filter parameter \(BFI_{\text{max}}\) that can be estimated more easily from catchment characteristics and hydrological inputs than the corresponding filter parameter in the Boughton filter (Li et al., 2014).

**Method 5:**
In Method 5, values of BFI and KBase are obtained directly (i.e. without calibration) by developing regression models predicting KBase and BFI as a function of the catchment characteristics (Table 1) and hydrologic inputs (Table 2) investigated, similar to the regression relationships for predicting filter performance and optimal filter parameters developed by Li et al. (2014) used in Methods 6 and 7. Values of BFI are calculated as the ratio of the volume of baseflow to the volume of total streamflow obtained from HGS and KBase is obtained by performing recession analysis on the total streamflow hydrographs obtained from HGS. The resulting regression equations and scatter plots are shown in Fig. 6. As can be seen, BFI can be predicted very well and KBase can be predicted reasonably well. These results also indicate that BFI and KBase are closely related to catchment characteristics and hydrological inputs.

![Nonlinear regression models for the prediction of BFI and KBase](image)

For Method 6, a modified version of the procedure for improving baseflow estimation using RDFs suggested by Li et al. (2014) is used in order to improve the baseflow estimates obtained using the LH filter in Method 5. First, the regression relationship developed by Li et al. (2014) is used to determine whether the LH filter can provide satisfactory estimates of baseflow for a particular catchment based on catchment properties and hydrological inputs. If the performance of the LH filter is predicted to be acceptable, Method 3 is used to obtain estimates of BFI and KBase. However, if this is not the case, Method 5 is used.

For Method 7, Method 7 is identical to Method 6, except that Method 4 (Eckhart filter) is used instead of Method 3 (LH Filter).
In order to be able to assess the absolute performance of the different calibration methods in terms of their
ability to predict base- and quick-flow hydrographs accurately, and to be able to assess the degree to
which the structure of AWBM is able to represent base- and quick-flow processes, two benchmarks are
developed. As part of the benchmark development process, the parameters that affect baseflow and
quickflow are calibrated to the “observed” base- and quick-flow hydrographs produced by HGS, which are
also used to assess the performance of the different calibration methods. A two-step calibration process
similar to that used for Methods 2-4 is used in order to obtain the benchmark base- and quick-flow
hydrographs. The only difference is in the way the parameters affecting baseflow (i.e. BFI and KBase)
are estimated in step 1 of the procedure for obtaining the baseflow benchmark (Benchmark 1) and the
way the parameters affecting quickflow (BFI and KSurf) are estimated in step 1 of the procedure for
obtaining the quickflow benchmark (Benchmark 2), as detailed below:

**Benchmark 1 (Baseflow):**
BFI and KBase are calibrated simultaneously so as to minimise the selected error measure between the
baseflow obtained using AWBM \( q_{B}^{\text{sim}} \) and that obtained with HGS \( q_{B}^{\text{obs, HGS}} \), using the selected
optimisation method.

**Benchmark 2 (Quickflow):**
BFI and KSurf are calibrated simultaneously so as to minimise the selected error measure between the
quickflow obtained using AWBM \( q_{Q}^{\text{sim}} \) and that obtained using HGS \( q_{Q}^{\text{obs, HGS}} \), using the selected
optimisation method. In order to do this, KBase, which determines the magnitude of the baseflow
component within total streamflow (Fig. 4), is set as 1.0, so that the total streamflow produced by AWBM
is comprised solely of quickflow.

### 2.4.2 Optimisation method

Model calibration is conducted using the shuffled complex evolution (SCE-UA) algorithm, because it has
been proven to be both accurate and efficient in previous studies (Hapuarachchi et al., 2001). In addition,
SCE-UA is widely recognized as being one of the best search procedures for use in CRR modelling
applications (Ajami et al., 2004; Franchini et al., 1998). The SCE-UA algorithm is based on the strengths
of several existing search procedures, including genetic algorithms (GAs) (Goldberg, 1989) and the
Nelder & Mead Simplex downhill search scheme (Nelder and Mead, 1965), but also introduces the
concept of complex shuffling (Duan et al., 1992). A detailed description of this method can be found in
Duan et al. (1992).

In this study, SCE-UA is implemented using the Rainfall Runoff Library (www.toolkit.net.au/rrl). The
number of complexes is set equal to the number of calibration parameters to reduce the chance of
premature termination of the search algorithm, as suggested by Kuczera (1997). All of the other
parameters are set to the recommended values in Duan et al. (1994). In order to check whether parameter
equifinality is a potential problem and to ensure near globally optimal solutions are obtained, each
 calibration run is repeated ten times. Another important aspect in the application of SCE-UA is that a
parameter space needs to be defined, in which the algorithm searches for the optimal parameter
combination. The ranges of all of the eight AWBM parameters used are based on the suggestions in the
Rainfall Runoff Library (www.toolkit.net.au/rrl) and are listed in

Table 3.

The length of streamflow data available for AWBM calibration is 10 years (Li et al., 2014). AWBM runs
on a daily time step, and therefore, is calibrated against daily streamflow. The calibration period requires
a warm up period, which enables the models to account for residual water in the catchment by partially
filling their storages, which primes the models for the calibration period. Initial testing demonstrates that a 1 year warm up period is sufficient for model calibration. Chronologically, the warm up period directly precedes the calibration period.

2.4.3 Error measure

The choice of an appropriate error measure is very important (e.g. see (Bennett et al., 2013)). In this study, the Nash-Sutcliffe coefficient (Ef) (Nash and Sutcliffe, 1970) is chosen for this purpose, as it is one of the most highly used performance measures in hydrology. Ef values are calculated by comparing the difference between the ‘observed’ (e.g. outputs from HGS simulations) and simulated time series for each time step for total streamflow, baseflow and quickflow from AWBM. The Ef value ranges from $-\infty$ (very bad model performance) to 1.0 (perfect model performance), with 0.0 indicating a model that can make predication of the same quality as the mean of the observations. An exact criterion of ranges of Ef values indicating model performance does not exist in the literature, although various empirical guidelines are given when implementing Ef as the metric. Chiew and McMahon (1993) and Ladson (2008) suggest that Ef values above about 0.7 to 0.8 indicate ‘acceptable’ model performance, while Moriasi et al. (2007) suggest, in general, the performance of the model can be judged as satisfactory when Ef values are greater than 0.5. In this study, Ef values between 0.5 and 1.0 correspond to ‘good’ model performance; Ef values between 0.0 and 0.5 show ‘acceptable’ model performance and ‘poor’ model performance is represented by negative values of Ef. However, Gupta and Kling (2011) indicate that high values of Ef can give poor model performance. Consequently, although more subjective than the use of statistical measures of goodness-of-fit, plots of simulated and observed hydrographs are also inspected following optimisation.

2.5 Evaluation of Model Performance

As mentioned previously, the performance of AWBM models calibrated using the different methods outlined in Section 2.4.1 is compared in terms of overall model predictive performance (i.e. how well the total streamflow generated using AWBM ($q_T^{\text{sim}}$) matches the corresponding streamflow generated using HGS ($q_T^{\text{obs}}$)) and the accuracy of the resulting internal model dynamics (i.e. how well the AWBM generated baseflow ($q_B^{\text{sim}}$) and quickflow ($q_Q^{\text{sim}}$) hydrographs match the corresponding hydrographs obtained using HGS (i.e. ($q_B^{\text{obs}}$ and $q_Q^{\text{obs}}$)) (Fig. 1). The performance of models developed using the different calibration methods is assessed using Ef and by visual inspection.

3. Results and Discussion

The performance of the AWBMs calibrated with the seven different methods investigated, as well as that of the two benchmarks, is summarised in Fig. 7. As can be seen, the results are presented in terms of the percentage of models developed for the 66 catchments resulting in “good” (Ef $\geq$ 0.5), “acceptable” (0 $\leq$ Ef $<$ 0.5) and “poor” (Ef $<$ 0) performance, for each of the total-, base- and quick-flow hydrographs. It can also be seen that two sets of results are presented for Method 1. This is because there are two distinct sets of model parameters that result in similar model performance in terms of total streamflow during the 10 calibration trials conducted for some of the 66 catchments. This is because when Method 1 is used, there is no control on internal model dynamics and the only objective is to find a set of model parameters that provides the best match to the “observed” total streamflow hydrograph, as discussed previously. However, as shown in Fig. 4, the AWBM total streamflow is the sum of routed surface runoff and baseflow, both of which are modelled in an identical fashion, each with a single parameter (KBase for baseflow and KSurf for routed surface runoff). Consequently, similar total streamflow can be obtained by
exchanging parameter values for KBase and KSurf, resulting in model equifinality, as observed in the calibration results for Method 1.

An example of this is given in Fig. 8. As can be seen in Fig. 8(a) and (b), good overall model performance is obtained for both parameter sets in terms of matching total streamflow. However, the internal dynamics are much better when parameter set 1 is used, as indicated by significantly better matches to the base-and quick- flow hydrographs. It is clearly evident that the baseflow and quickflow patterns are reversed for the models with the different parameter sets (e.g. the pattern of baseflow obtained with parameter set 1 is very similar to the pattern of quickflow obtained with parameter set 2). In Fig. 7, all of the results with parameters that result in better internal model dynamics (e.g. as in Fig. 8(a)) are represented by Set 1, whereas the results with parameters that result in poorer internal model dynamics (e.g. as in Fig. 8(b)) are represented by Set 2. However, in the practical application of AWBM, the parameters of Set 2 can be avoided. This is because experienced modellers would usually subjectively assess the values of the KBase and KSurf parameters based on the calibration procedure (i.e. Method 1) and only accept values that seem reasonable.

Overall, the results show that total streamflow is predicted well using all calibration methods investigated, with “poor” model performance for fewer than 10% of the catchments, except when Method 5 (Regression) is used, in which case 15% of catchments result in poor model performance (Fig. 7a). For most of the catchments (54%-68%) “good” performance is obtained. However, the internal model dynamics are not represented as well, especially for baseflow, where “poor” performance is obtained for more than half (58%-74% (ignoring Method 1 with parameter set 2)) of the catchments and “good” performance for fewer than 20% of the catchments (Fig. 7b). The ability of AWBM to represent quickflow is slightly better, with good performance for 32%-41% of the catchments when the calibration methods that consider internal model dynamics explicitly (i.e. Methods 2 to 7) are used and poor performance for fewer than 30% of the catchments (12%-29%) (Fig. 7c). It should be noted that when Method 1 is used, “good” performance is only obtained for 12% of catchments, but “acceptable” performance is achieved for the vast majority of the remaining catchments (71%), with poor performance for only 17% of the catchments.

Method 1 is the best-performing calibration approach in terms of matching total streamflow (Fig. 7a). This is not surprising, as the method calibrates all of the model parameters simultaneously in order to obtain the best match to total streamflow. However, this is at the expense of internal model dynamics. As discussed above, there is a potential problem with equifinality as a result of the structure of AWBM. As shown in Fig. 7, even if the results for Set 1 are considered, the method results in the second highest percentage of catchments with “poor” model performance for baseflow estimation and the lowest percentage of catchments with “good” model performance for quickflow estimation. Nevertheless, the performance of Method 1 is comparable with that of the other methods in terms of baseflow estimation, as baseflow is estimated poorly, irrespective of the method used. In relation to quickflow estimation, there is only a small percentage of catchments with “poor” model performance, although Method 1 results in a significantly smaller number of catchments with “good” performance compared with the other calibration methods. Consequently, the overall performance of the method is best in terms of total streamflow prediction and reasonable compared with that of the other methods in terms of baseflow and quickflow prediction. However, the equifinality problem requires careful attention, although in practice, experienced modellers could avoid an inappropriate parameter set (i.e. Set 2) by accepting the parameter set that seems reasonable, as discussed above. This is because very poor internal model dynamics can be obtained (e.g. “poor” performance for 100% of the catchments in terms of baseflow prediction and for 50% of the catchments in terms of quickflow prediction) if the inappropriate parameter set (Method 1 (Set 2)) is used.

The methods that calibrate the BFI and KBase to the baseflow extracted using RDFs (i.e. Methods 2-4)
result in the best match to baseflow. This indicates that the RDFs are able to produce reasonably accurate estimates of baseflow, which is evidenced by the fact that the performance of Methods 2-4 is only slightly worse than that of Benchmark 1 (i.e. where BFI and KBase are calibrated to the baseflow produced by HGS, which is the same baseflow used for performance assessment). It also indicates that there is some benefit in terms of improving internal model dynamics by constraining BFI and KBase to ensure that baseflow is matched as well as possible.

As stated previously, the benchmarks are used to test how much the changes in performance are due to the calibration method or the model structure. The relatively poor performance for Benchmark 1 (i.e. “poor” performance for 57% of the catchments) tends to suggest that the improvement that is possible by adopting this approach is rather limited, due to problems imposed by the constraints of the model structure in AWBM preventing simulation of the whole range of baseflow dynamics resulting from the complex underlying physical processes associated with baseflow generation and depletion.
Fig. 7 Performance of AWBM for the different calibration methods investigated.
Of the three methods considered, Method 2, which uses the LH filter with its default filter parameter value of 0.925, performs worst, while the performance of the other two methods (i.e. LH and Eckhart filters with optimal filter parameters) is very similar. The Eckhart filter performs slightly better, with a slightly larger percentage of catchments with “good” performance and a slightly smaller percentage with “poor” model performance. This supports the suggestion by Boughton (2004) that the Boughton filter, which is mathematically identical to the Eckhart filter, as mentioned previously, is well suited to use with AWBM.

While use of Methods 2-4 is able to produce the best results in terms of baseflow estimation, there are some trade-offs in terms of the ability to match total streamflow and quickflow. Even though use of Methods 2-4 results in a significant increase in the percentage of catchments with “good” performance in
terms of quickflow prediction compared with Method 1, there is also a slight increase in the percentage of catchments with “bad” performance. In relation to total streamflow, use of Methods 2-4 also results in a slight reduction in performance compared with Method 1. Overall, the LH method with optimal filter parameter values produces the best trade-offs in performance between matching total-, base- and quick-flow among Methods 2-4. However, a disadvantage of this method compared with Method 2 is that information on catchment properties and hydrological inputs is needed in order to apply the regression equations used to obtain optimal filter parameter values, making it more difficult to apply.

The regression method (Method 5) results in the best performance in terms of matching quickflow, performing only slightly worse than the quickflow benchmark (Benchmark 2). The method results in the smallest percentage of catchments with “poor” performance and second highest percentage of catchments with “good” performance. However, it is also the worst-performing method in terms of matching total- and quick-flow. An advantage of the method is that it does not require estimates of baseflow hydrographs (only total streamflow hydrographs are needed), but a disadvantage is that information on catchment properties and hydrological inputs is needed in order to apply the regression equations used to obtain values of BFI and KBase as part of the calibration process.

The two methods combining aspects of the filter (Methods 2-4) and regression (Method 5) methods (Methods 6-7) provide a good compromise in terms of performing reasonably well on all three hydrographs (i.e. total-, base- and quick-flow). While they are not the best-performing methods for any of the three hydrograph components, their performance is more consistent than that of any of the other methods and not far from that of the best-performing method in each case. Of the two methods, the method using the LH filter (Method 6) performs better than the method using the Eckhart filter (Method 7) overall. A disadvantage of these methods is that they are more complex to apply than the other methods, as they require baseflow extraction using a RDF, as well as data on catchment characteristics and hydrological inputs in order to apply the regression equations for predicting filter performance, obtaining the optimal filter parameters and obtaining direct estimates of BFI and KBase in the case where predicted filter performance is poor.

Overall, the results quantitatively demonstrate that while total streamflow can be predicted very well over a wide range of catchment characteristics and hydrological inputs using AWBM, the component hydrographs are not modeled very well, particularly baseflow. This raises questions about the way the processes associated with these streamflow components are conceptualized in AWBM. The pertinence of such questions is dependent on the purpose of the model though. If the model purpose is only predicting total streamflow over short time periods (days to weeks) within the bounds (minimum and maximum flow) for which it has been calibrated, then internal model dynamics are likely a non-issue and it is just a curve fitting exercise. However, in such instances, black-box models, such as artificial neural networks, might perform better. If the model purpose requires the model to be “right for the right reasons”, then process representation is important. This highly difficult challenge is strived for where the model use is, for example: aiding catchment behavior understanding and/or simulating flows outside of the calibration range.

Of the methods investigated, Methods 3 and 4, as part of which BFI and KBase are calibrated to the baseflow obtained using the LH and Eckhart filters with optimal filter parameters, provide the best performance in terms of baseflow hydrograph prediction. The regression method (Method 5) provides the best match to the quickflow hydrographs and the hybrid method using the LH filter with optimal filter parameters and the regression approach (Method 6) provides the best overall trade-offs in terms of matching all three hydrographs.
4. Summary and Conclusions

In this paper, the impact of seven different calibration methods on the ability of AWBM to predict total-, base- and quick-flow hydrographs is assessed for 66 synthetic catchments with different physical properties and hydrological inputs. The results indicate that total streamflow can generally be predicted to an acceptable level for more than 90% of the 66 synthetic catchments. In contrast, baseflow is predicted poorly, with acceptable performance levels ranging from 26% to 41% for the different calibration methods investigated. However, prediction of quickflow is much better, with acceptable performance levels ranging from 71% to 88%. This disparity in performance between baseflow and quickflow prediction is despite the fact that the hydrographs for the 66 synthetic catchments consist of a wide range of BFI values (0.006<BFI<0.997, Median BFI=0.53), suggesting that the way AWBM represents internal physical processes could be improved by using additional information about the catchment, as shown in calibration Methods 2 to 7.

Use of the calibration methods that take internal model dynamics into account explicitly (Methods 2 – 7) results in improved prediction of the component hydrographs. This improvement is particularly significant in terms of producing “good” estimates of quickflow, with the percentage of catchments for which good performance is obtained increasing from ~10% to ~40% for most methods. However, as AWBM is unable to provide good predictions of baseflow in general, as discussed above, the improvements obtained by using Methods 2 – 7 are only small (generally < 5%).

This study offers an analysis of seven calibration methods under which model internal consistency is assessed. Which calibration method should be used is based which component of the flow is of most interest. Generally, total streamflow prediction is the most important, and therefore, method 1 (i.e. simultaneously calibrating all parameters to total streamflow) is likely to be the most appropriate. However, in order to obtain reasonable internal model dynamics, the problem of equifinality needs to be addressed, although experienced modellers usually have an idea of which parameter sets are reasonable. For situations where baseflow is of most interest, such as low flow forecasting, identification of source areas and recharge estimation, the methods that calibrate BFI and KBase to baseflow extracted using recursive digital filters (RDFs) (Methods 2-4) should be used. Of these methods, the method that uses the Eckhardt filter with the optimal filter parameters obtained using the regression equations developed by Li et al. (2014) (Methods 4) performs best. If quickflow prediction is the primary objective (i.e. flood prediction), the regression method (Method 5) should be used. When all three hydrographs are important for water management, the hybrid method using the Lyne and Hollick (LH) filter with optimal filter parameter and the regression method (Method 6) is likely to give the best performance.

It should be noted that most of the methods that perform best require information on catchment characteristics and hydrological inputs in order to apply the regression equations for obtaining optimal filter parameters (Methods 3, 4 and 6), predicting BFI and KBase directly (Methods 5 and 6) and estimating the level of performance of the LH filter with optimal filter parameters (Method 6), which makes them difficult to apply if the required information is not readily available. However, use of Method 2 (LH filter with default filter parameter of 0.925) can be used in order to achieve reasonable improvements in internal model dynamics in such cases.

It is also important to highlight a number of limitations of this study that provide avenues for future studies. Firstly, it is worth pursuing other calibration methods that aim to take the internal model dynamics into account explicitly. For example, changes could be made to the objective function to ensure known features of the total streamflow hydrograph are captured, or different transformations could be applied to the data (e.g. use a Box-Cox transformation to shift the focus from high flows to low flows (Romanowicz, 2010)). In addition, uncertainty could be considered during the AWBM calibration process (e.g. (Ahmadi et al., 2014; Fonseca et al., 2014; Zuidema, 2011)). In addition, the fact that
synthetic data are used in this study limits its complexity and realism. However, there is a trade-off between realism and the ability to assess model dynamics accurately. For real catchments, the actual base- and quick-flow hydrographs are generally unknown and have to be estimated using a variety of indirect methods (e.g. tracers, RDFs). However, the approach adopted in this study enables the base- and quick-flow hydrographs to be known with certainty, enabling an accurate assessment of the internal dynamics of AWBM to be obtained in a simplified setting. Lastly, while the computational experiments are conducted for a range of catchment characteristics and hydrological inputs, these values are also represented in a simplified manner (e.g. uniform rainfall, homogeneous soils, uniform slopes etc.) and the effect of additional complexity (see e.g. (Frei and Fleckenstein, 2014)) on the results should be investigated in future studies. In addition, it would be interesting to compare the results obtained using the approach presented in this study with alternative approaches to assessing the internal dynamics of conceptual rainfall-runoff models, such as sensitivity analysis (e.g. (Shin et al., 2013)).

Finally, while the focus of this paper is on the assessment of the internal model dynamics of the AWBM, the generic approach presented is equally applicable to other CRR models. As mentioned in the Introduction, the methodology presented in this paper provides a first step towards the more rigorous assessment of the internal dynamics of CRR models, providing insight as to whether they primarily provide a curve-fitting service, similar to that of black-box models, or whether they are able to provide a reasonable representation of the underlying processes, as suggested by their structure. If the latter applies, there can be greater confidence in the results obtained from CRR models, particularly when they are applied in settings outside those used for model calibration. Consequently, the methodology presented in this paper should be applied to other CRR models in order to assess their internal dynamics.

5. Acknowledgements

This research was supported by the Chinese Scholarship Council (CSC) and the University of Adelaide. The authors also gratefully thank the three anonymous reviewers for their detailed comments, which have greatly improved the quality of the paper.

6. References


