Automating drainage direction and physiographic inputs to the CEQUEAU hydrological model: sensitivity testing on the lower Saint John river watershed, Canada

Stephen J. Dugdale, André St-Hilaire and R. Allen Curry

ABSTRACT

CEQUEAU is a process-based hydrological model capable of simulating river flows and temperatures. Despite an active user base, no facility yet exists for the automatic assembly and input of watershed data required for flow simulations. CEQUEAU can therefore be time-consuming to implement, particularly on large ($\geq 10^4$ km^2) watersheds. We detail a new MATLAB toolbox designed to remove this key limitation by automatically computing CEQUEAU's key drainage direction and physiographic inputs from GIS data. With the toolbox, model implementation can now be achieved extremely quickly (<1.5 hr) given suitable inputs. This time saving enabled us to assess CEQUEAU's sensitivity to changes in grid size by implementing the model on a large (14,990 km^2) watershed at successively decreasing resolution (2.5 km to 112 km), using a fixed calibration parameter set. Results of this analysis showed that despite some model strength fluctuations linked to variability in computed basin size/land-use, only a minor decrease in model strength (mean NSE reduction $= 0.03$) was observed at relatively fine resolutions (2.5 km to 20 km). Although results might change if the model was recalibrated at resolution step, findings indicate that CEQUEAU is able to provide realistic flow simulations at a wide range of resolutions.

Key words | CEQUEAU, drainage network analysis, GIS, hydrological model, sensitivity testing

Stephen J. Dugdale (corresponding author)
André St-Hilaire
R. Allen Curry
Canadian Rivers Institute, Department of Biology & Faculty of Forestry and Environmental Management,
University of New Brunswick,
Fredericton,
New Brunswick,
Canada
E-mail: *s.j.dugdale@bham.ac.uk*

André St-Hilaire
Institut National de la Recherche Scientfique,
Centre Eau Terre Environnement,
Québec,
Québec,
Canada

INTRODUCTION

Given the growing threat of anthropogenic pressures to stream ecosystems (e.g., Jonsson & Jonsson 2009; Tockner *et al.* 2010; van Vliet *et al.* 2013), hydrological models are increasingly important for the simulation and forecasting of river flows. Advances in computing power and storage mean that a huge variety of hydrolagical models now exists (Singh 1995; Devia *et al.* 2015). At the most basic level, hydrological models can be categorised as either statistical or process-based. While statistical models (which function by means of statistical linkages between model inputs and outputs (Yevjevich 1987)) are generally considered simpler than process-based models (which attempt to mimic real-world hydrological processes in order to simulate river flows, e.g., Clark *et al.* 2015), process-based models explicitly incorporate land-use information as model inputs, and are therefore especially useful to simulating flows in 'modified' environments (e.g., Kite 2001; Hart *et al.* 2002; Ouyang *et al.* 2011). Process-based models are generally data-intensive and highly parameterised, often requiring numerous inputs pertaining to the physiographic and hydrological processes of a given watershed (Beven 2001; Ajami *et al.* 2004; Khakbaz *et al.* 2012). Lumped process-based models attempt to minimise these data requirements by summarising model parameters into global values representative of the entire watershed (Khakbaz *et al.* 2012). However, they are therefore unable to encapsulate spatial variability in hydrological processes. Conversely, fully distributed process-based models provide simulations of flow throughout

Uncorrected Proof

2   S. J. Dugdale *et al.* | Automating physiographic inputs to the CEQUEAU hydrological model   **Journal of Hydroinformatics | in press | 2017**

a watershed by discretising model parameters across a user-defined grid (Carpenter & Georgakakos 2006; Khakbaz *et al.* 2012). However, they are often difficult to implement, requiring a good understanding of the variability in model parameters across the watershed (Beven 1989; Refsgaard 1997). Semi-distributed models aim to bridge the gap between lumped and full distribution by holding certain model parameters and data inputs as global values while discretising others at intermediate (e.g., sub-basin) scales (e.g., Schumann 1993; Ajami *et al.* 2004; Khakbaz *et al.* 2012). Semi-distributed models are therefore often seen as offering the best trade-off between ease of implementation and spatial discretisation (Marcé *et al.* 2008; Jajarmizadeh *et al.* 2012).

While recent studies on semi-distributed models have focused on optimising the ratio of global to discretised inputs in order to achieve an ideal balance between simulation quality and model parsimony (e.g., Khakbaz *et al.* 2012), semi-distributed models can nonetheless be demanding to implement because of the need to input spatially explicit meteorological, physiographic and stream network topology data. In order to address these limitations, researchers have increasingly looked to geographic information systems (GIS) to facilitate or automate data input (Sui & Maggio 1999). While GIS has long been used in a supportive capacity to assemble the data required to run hydrological models (DeVantier & Feldman 1993; Bhatt *et al.* 2014), more recent advances have enabled full integration of hydrological models within GIS environments (e.g., Anderson *et al.* 2002; Olivera *et al.* 2006; Hughes & Liu 2008; Merkel *et al.* 2008; Bhatt *et al.* 2014). The advent of open-source GIS packages means that model customisation and modification possibilities are now essentially limitless (e.g., Formetta *et al.* 2014). However, despite these key advances, a large number of widely used hydrological models still lack GIS integration, presumably due to their age and the relative difficulty of implementing them in a GIS interface. Even in the case of models that work well for a given application, the stagnation and eventual decline of their user base resulting from the difficulty of assembling data inputs may eventually lead to the model's obsolescence, as newer, GIS-interfaced systems are deemed preferable.

This is a potential weakness of CEQUEAU, a semi-distributed process-based hydrological and water temperature model originally developed by the Institut National de la Recherche Scientifique, Québec, Canada (Morin & Couillard 1990). Renewed interest in CEQUEAU for flow and water temperature simulations (particularly in regulated rivers) has spurred efforts to update and modernise the current version of the model. Rio Tinto, one of the model's largest users, has led the development of a new modular C++ implementation of CEQUEAU. This new version allows CEQUEAU to run on ×64 CPUs, removes previous limitations on the size and resolution of the model grid, and allows users familiar with C++ to develop alternative new snowmelt and evapotranspiration routines (St-Hilaire *et al.* 2015). Furthermore, the high degree of interoperability between MATLAB (MathWorks 2015) and C++ means that CEQUEAU routines in the form of MATLAB 'MEX' files can be called directly from MATLAB, allowing researchers and river managers with limited programming experience the ability to easily run and script CEQUEAU simulations. However, despite these advances, the new implementation of CEQUEAU currently has no facility for the automatic extraction of basin physiography or the computation of drainage direction between model cells. Because these inputs are still assembled manually, this represents a considerable limitation when working in large (i.e., $\geq 10^4$ km^2) or complex watersheds comprising many hundreds or thousands of grid squares. A GIS-like system capable of assembling the physiography data and computing drainage direction between model cells would, therefore, be a substantial boon to CEQUEAU users, speeding up model implementation and aiding research efforts.

In light of this, this article details the development of MATLAB toolbox designed to remove these key limitations to CEQUEAU, automating the computation of drainage direction and the extraction/input of physiographic data necessary to run the model. First, we give a brief overview of the CEQUEAU hydrological model. We then discuss the implementation and functionality of the CEQUEAU Physiography Toolbox, a series of MATLAB functions designed to automate data entry to the CEQUEAU model. The toolbox, as well as documentation explaining the various functions, is freely available at https://github.com/sjdugdale/cequeauPhysiography. Finally, we examine the sensitivity of the CEQUEAU model to changes in grid size by using the toolbox to implement CEQUEAU on a large

(14,990 km^2) watershed at successively decreasing resolutions until it essentially functions as a lumped model. Such an exercise would not previously have been possible owing to the large amount of time required to manually repeat the drainage direction and physiographic data input steps at each resolution iteration. The results of this exercise have the potential to improve the quality and decrease the computation time of future CEQUEAU flow simulations by providing important information regarding the effect of grid resolution on model performance.

THE CEQUEAU MODEL

A detailed description of CEQUEAU is provided in Morin & Couillard (1990). However, to facilitate understanding of the CEQUEAU Physiography Toolbox described in this article, the structure and functioning of the model is briefly summarised here. CEQUEAU simulates discharge across a grid of equally sized square cells superimposed on a watershed (hereafter referred to as elementary representative areas; ERAs) by calculating a hydrological budget on each ERA. Each ERA is divided into a maximum of 4 'partial squares' based on drainage divides that are present in the ERA (Figure 1). An ERA can contain between 1 and 4 partial squares whose sizes are determined by the presence of drainage divides. The volume of water present within each partial square is calculated as the fraction of the ERA's hydrological budget that corresponds to the partial square's area. Downstream transfer is achieved through routing streamflow from each partial square to its downstream neighbouring partial square based on the drainage directions input to CEQUEAU. Each ERA and partial square requires a series of physiographic data inputs which, in addition to meteorological data (daily solid and liquid precipitation, daily minimum and maximum temperature), govern the volume and residence time of water available for transfer from each partial square to the next. These comprise their altitude, percentage forest cover, percentage bare soil, percentage water body cover and percentage wetland.

The movement of water through the CEQUEAU grid is governed by two principal equations: the model's production function, which governs the vertical movement of
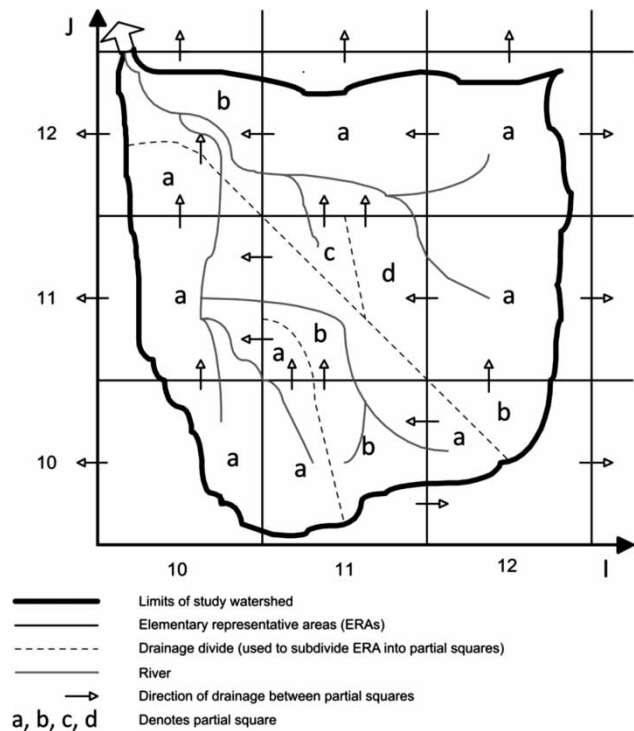
Figure 1 │ Subdivision of ERAs into partial squares based on drainage divides present within ERA (modified from Morin & Couillard (1990)). If ERA contains no drainage divides, one partial square is created of equal size to ERA. If ERA contains ≥1 drainage divides (e.g., ERA I = 11, J = 11), ERA is divided into multiple partial squares (denoted here by (a, b, c, d) notation. Direction of drainage between partial squares is given by arrows. These drainage directions are currently determined manually, but the CEQUEAU Physiography Toolbox now accomplishes this automatically.

water within each partial square, and the transfer function, which determines the volume of water transferred from a partial square to its downstream neighbour. The production function is conceptualised as a sequence of reservoirs which represent the storage of water within the saturated and unsaturated zones in the ground and within lakes and wetlands (Figure 2). For any given partial square, the movement of water into and between each of these reservoirs is represented by a series of equations (see Morin & Couillard (1990) for details) which simulate the partial square's hydrological budget in terms of snowpack formation/melt, evapotranspiration, infiltration to the unsaturated and saturated zones and lake/wetland storage. Model calibration is achieved through modifying the coefficients of these equations until calculated flows approach observed values as closely as possible. The production function thus calculates the volume of water available as streamflow for a
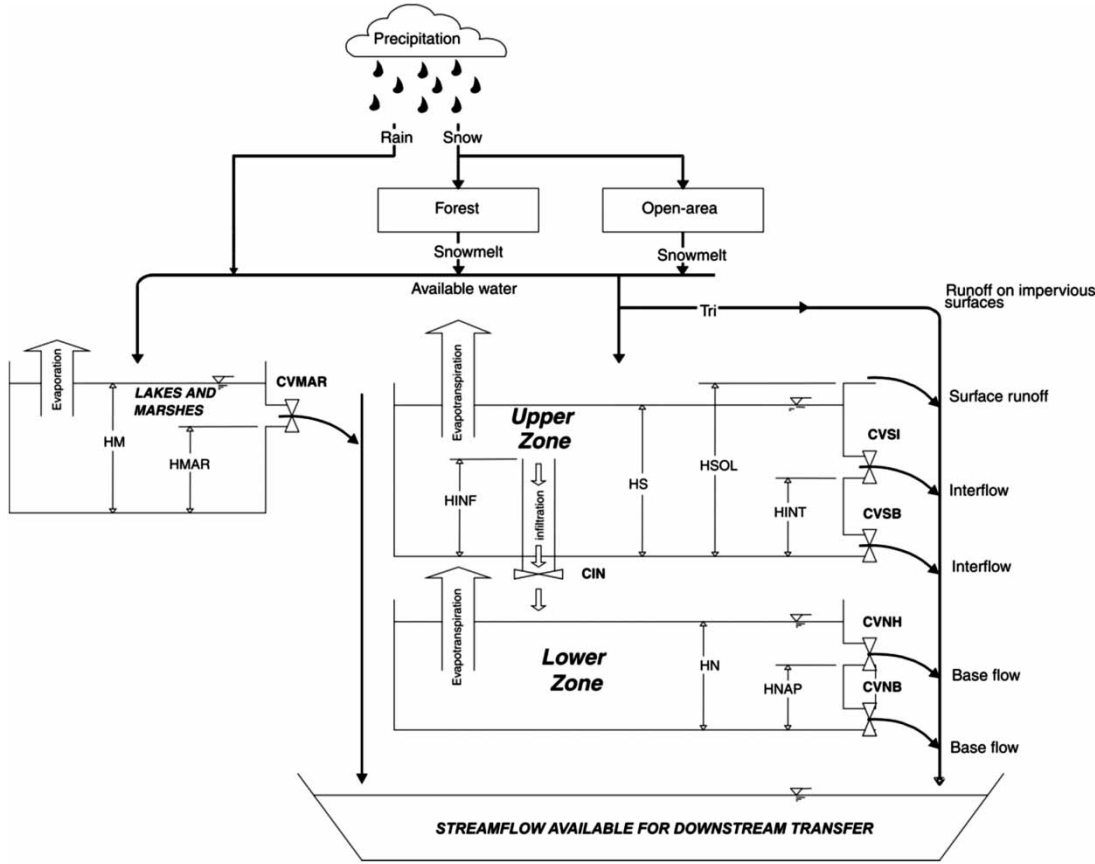
Uncorrected Proof

4   S. J. Dugdale *et al.* | Automating physiographic inputs to the CEQUEAU hydrological model   **Journal of Hydroinformatics** | in press | 2017

Figure 2 | Conceptual diagram of CEQUEAU's production function (modified from Morin & Couillard (1990)).

given partial square at time t by means of the equation:

$$Q_t = P_t - ETP_t + (HU_t - HU_{t-1}) + (HL_t - HL_{t-1}) \qquad (1)$$

where Q is the total streamflow (mm), P is the liquid precipitation or snowmelt from the accumulated snowpack, ETP is evapotranspiration (mm), HU is the amount of water stored in the upper (unsaturated) zone (mm) and HL is the amount of water in the lower (saturated) zone (mm). The model timestep (t) is usually daily.

The model's transfer function is subsequently applied to the volume yielded by the production function to determine the volume of water available for transfer from one partial square to its downstream neighbour:

$$v_i = XKT_i \cdot V_i \qquad (2)$$

where v is the volume of water yielded by the production function for partial square i as a function of that partial square's total water volume V and a routing coefficient XKT. Given that the magnitude of transfer between one partial square and its downstream neighbour is governed by the partial square surface storage capacity in lakes or marshlands, XKT essentially describes this storage capacity:

$$XKT_i = 1 - exp\left(\frac{EXXKT \cdot SA_i}{SL_i} \cdot \frac{100}{CEKM2}\right) \qquad (3)$$

where SA is the area of the watershed upstream of partial square i (km^2), SL is the area of lakes and wetlands in partial square i (km^2), CEKM2 is the area of the ERA (km^2) and EXXKT is a fitting parameter determined during model calibration.

Given suitable physiography, drainage directions and meteorological data, and provided a reasonable calibration has been achieved, the production and transfer functions

Uncorrected Proof

5   S. J. Dugdale *et al.* │ Automating physiographic inputs to the CEQUEAU hydrological model   **Journal of Hydroinformatics** │ **in press** │ **2017**

of CEQUEAU allow for the simulation of streamflow within each partial square comprising the watershed at the chosen model timestep.
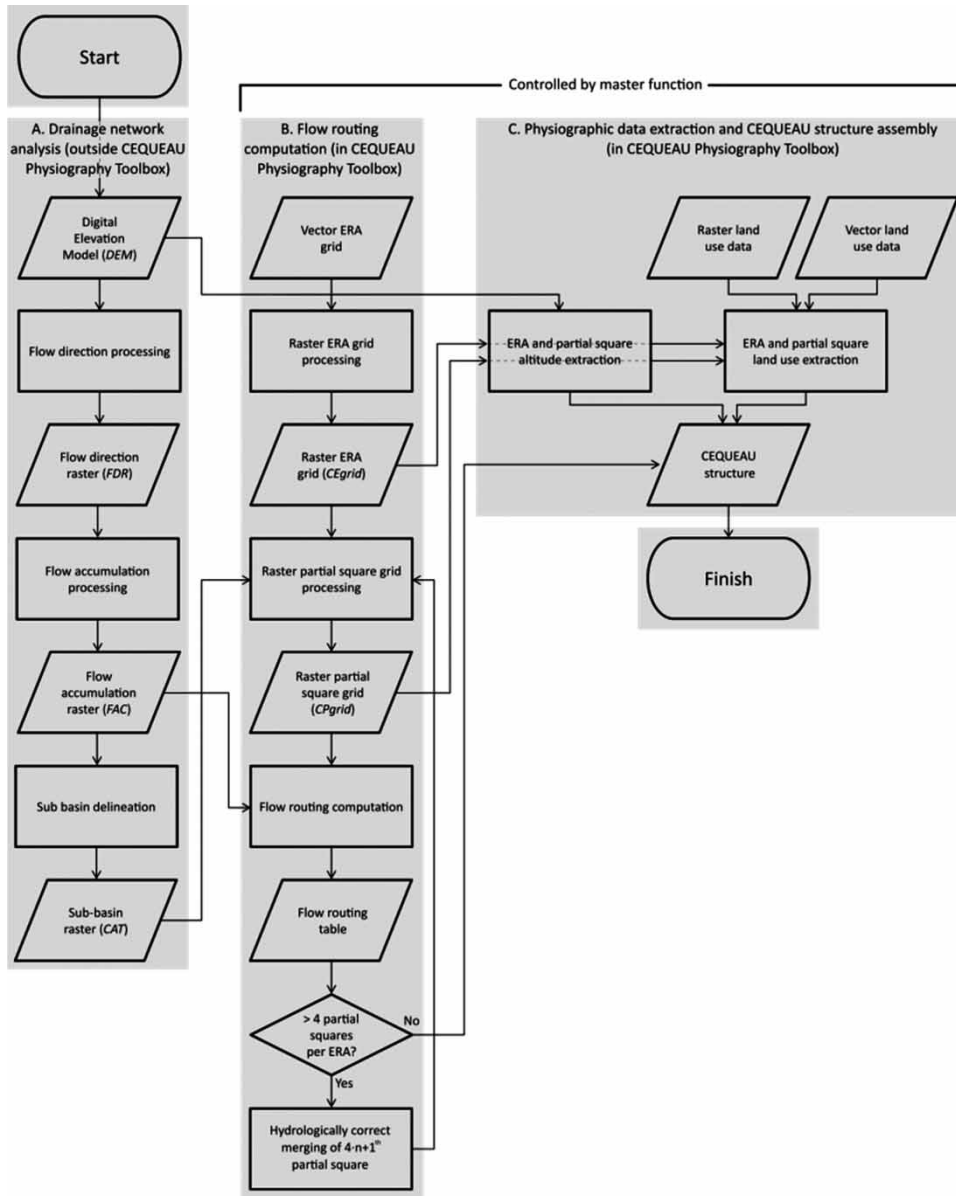
THE CEQUEAU PHYSIOGRAPHY TOOLBOX

In spite of recent updates to CEQUEAU, the principal limitation of the model is the necessity of determining the direction of flow from each partial square to its neighbour (Figure 1). Although digital elevation model-based (DEM) drainage network analysis packages such as Arc Hydro Tools (Maidment 2002) or TopoToolbox (Schwanghart & Kuhn 2010) can support the subdivision of each ERA into its constituent partial squares through the delineation of watershed sub-basins, no tool currently exists for automatically determining the drainage direction between

CEQUEAU's partial squares. This key stage of model implementation is currently accomplished manually by visually identifying drainage direction from DEMs or contour maps. However, modern DEM-based drainage network analysis techniques mean that this process can now be automated.

This is the function of the CEQUEAU Physiography Toolbox, a series of MATLAB routines (Table 1) that use the outputs of drainage network analysis software (e.g., Maidment 2002; Schwanghart & Kuhn 2010) to automatically assemble the drainage direction and physiographic data required by CEQUEAU. The heart of the CEQUEAU Physiography Toolbox is a master function which controls data input/output and calls sub-routines responsible for computing drainage direction and extracting basin physiography (Figure 3). In order to compute drainage direction, the master function requires raster inputs comprising the

Table 1 │ Matlab functions comprising CEQUEAU Physiography Toolbox

Function name	Description
IO/control	
CEQUEAUphysiography.m	Master function controlling IO and assembly of data into a CEQUEAU Matlab structure
createCEgrid.m	Creates raster ERA grid from ArcGIS fishnet
createCPgrid.m	Creates raster partial square grid by intersecting ERA grid with sub-basin raster
populateStructs.m	Populates CEQUEAU structure with data generated by drainage direction/physiography functions
Drainage direction computation	
doCProuting.m	Computes drainage direction between partial squares
doRoutingTable.m	Re-organises drainage direction table to conform with CEQUEAU input requirements
removeCPsegments.m	Removes and merges partial squares <1% of ERA size
removeFACzeroCPs.m	Removes and merges partial squares with maximum flow accumulation of zero
do4CPs.m	Ensures that each ERA contains no more than 4 partial squares
outletRoutes.m	Traces downstream path from each partial square to watershed outlet
redoCPgrid.m	Re-draws raster partial square grid based on computed drainage direction
redoCEgrid.m	Re-draws raster ERA grid based on computed drainage direction
doCEcoordinates.m	Assigns (I,J) coordinates necessary for CEQUEAU to ERAs and partial squares
Physiographic data extraction	
getCPareas.m	Computes area of each partial square as % of ERA
getCumulCPareas.m	Computes cumulative basin area upstream of each partial square
getAltitudes.m	Extracts ERA and partial square altitudes from DEM
doRasterLandCover.m	Extracts ERA and partial square land cover (forest/bare soil cover) from raster
doVectorLandCover.m	Extracts ERA and partial square land cover (waterbodies/wetlands) from shapefile
getCumulLandCover.m	Computes cumulative land cover areas upstream of each partial square

Figure 3 │ Diagram of CEQUEAU Physiography Toolbox functionality.

watershed DEM, a raster containing watershed flow accumulation values (hereafter referred to as *FAC*) and a raster delineating sub-basins present within the watershed (hereafter referred to as *CAT*). These flow accumulation and sub-basin rasters are readily available as outputs of drainage network analysis software; as such, a detailed description of the methods used to generate them is beyond the remit of the article. However, for further information, we refer the reader to Tarboton (1997),

Schwanghart & Kuhn (2010), Shahzad & Gloaguen (2011), O'Callaghan & Mark (1984), Seibert & McGlynn (2007) and Costa-Cabral & Burges (1994). In addition to these raster inputs, the master function also requires a vector grid representing the layout of the model's ERAs. To facilitate GIS-model interoperability, the function accepts shapefile grids created using ArcGIS's 'Create Fishnet' function, although it is nevertheless possible to construct the ERA grid natively in MATLAB.

The master function also requires inputs pertaining to the key physiographic data required to run CEQUEAU. The toolbox therefore contains functions for extracting the physiographic data (catchment land use in terms of forest cover, bare soil, waterbodies and wetlands) from raster or vector GIS files compatible with ESRI's ArcGIS package (ESRI 2014). A raster land cover file is used to compute the percentage forest and bare soil coverage of each ERA and partial square; percentage cover of waterbodies and wetlands are calculated from ArcGIS polygon shapefiles. All input raster and vector data sources must be projected in the same (metric) coordinate system. Furthermore, all input data must be accompanied by GIS world files allowing MATLAB to translate between raster pixel coordinates and real world positions (or vice versa) using the affine transformation:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} A & B \\ D & E \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} C \\ F \end{bmatrix} \tag{4}$$

where x and y are the column and row indices of each raster pixel, A and E are the pixel dimensions (m), B and D are the cosine and sine of image rotation angle θ, C and F are the raster translation terms (coordinates of the raster's upper left corner in map projection) and x' and y' are the transformed pixel coordinates in map projection.

Drainage direction computation

Prior to the computation of drainage direction, the appropriate rasters (*DEM, FAC, CAT*) are generated using drainage network analysis software (e.g., Maidment 2002; Schwanghart & Kuhn 2010). These data, along with the ERA vector grid, are then input to the CEQUEAU Physiography Toolbox. The toolbox subsequently generates a raster ERA grid (*CEgrid*) of identical dimensions to the DEM using the ERA vector grid. This raster is then intersected with the sub-basin raster *CAT* to create a new raster grid of partial squares (*CPgrid*). Following this step, the drainage direction between successive partial squares is computed. This is achieved by first identifying the downstream-most point of each partial square as given by the highest flow accumulation pixel within the corresponding area of *FAC* (Figure 4(a) and 4(b)). Next, the 8-connected neighbourhood
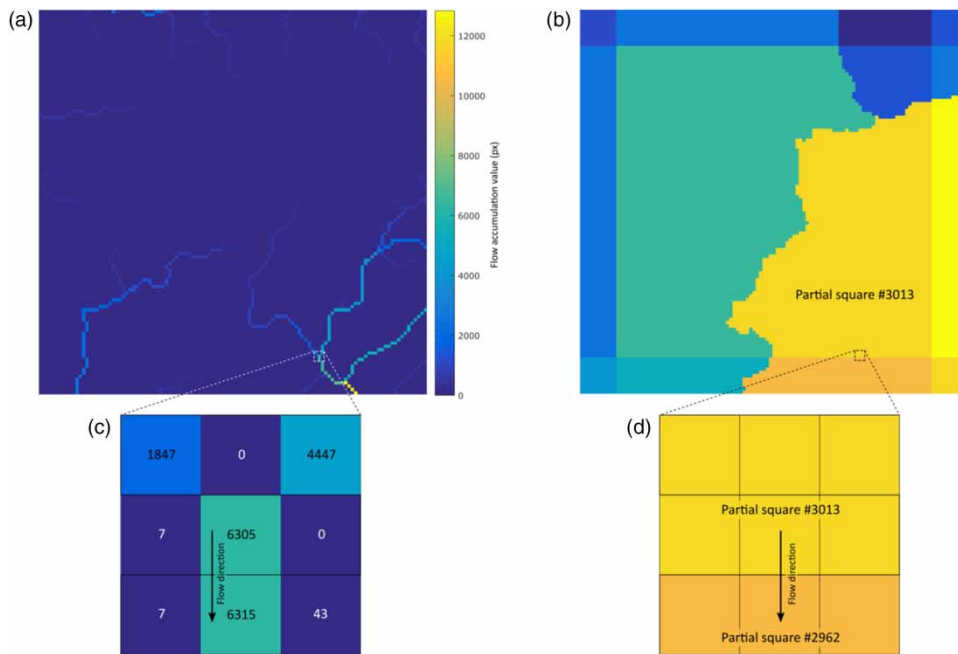
Figure 4 | (a, b) Highest valued pixel in area of flow accumulation raster (*FAC*) corresponding to partial square #3013 is partial square's outlet. (c) 8-connected neighbourhood of pixel extracted, allowing for identification of its downstream neighbour (highest valued pixel in neighbourhood). (d) Downstream neighbour corresponds to the inlet of the next partial square, thus allowing the nth partial square's downstream neighbour to be tabulated.

Uncorrected Proof

8   S. J. Dugdale *et al.* | Automating physiographic inputs to the CEQUEAU hydrological model   **Journal of Hydroinformatics** | in press | 2017

around this pixel is extracted and the new highest flow accumulation pixel within this neighbourhood identified. The partial square in which this new highest flow accumulation pixel lies will intrinsically correspond to the original partial square's downstream neighbour (Figure 4(c) and 4(d)). The index of this downstream partial square is subsequently tabulated, and the process repeated for all partial squares.

Although the resulting drainage direction table should in theory be suitable for input to CEQUEAU, it must be checked to ensure that each ERA contains no more than 4 partial squares, a key input requirement of CEQUEAU. This can generally be avoided when preparing the drainage network analysis inputs to the CEQUEAU Physiography Toolbox by taking care to ensure that the sub-basins in *CAT* are sufficiently large that they will not produce more than 4 partial squares when intersected with the ERA grid (as required by CEQUEAU). This is achieved by modifying the flow accumulation threshold used to delineate sub-basins in *CAT*. Although preliminary investigations indicate that a flow accumulation threshold equivalent to 20 times the area of the ERA will generally yield no more than 4 partial squares per ERA, this is unlikely to represent a universal rule. The CEQUEAU Physiography Toolbox therefore contains three functions to correct ERAs containing >4 partial squares. The first function loops through each ERA within *CEgrid* and detects any partial squares within the corresponding area of *CPgrid* that comprise ≤1% total area of the ERA. These partial square 'fragments' are subsequently merged with neighbouring partial squares within the ERA. The second function acts in a similar way, identifying and merging partial squares in which the corresponding area of *FAC* only contains flow accumulation values equal to zero. Finally, the third of these functions loops through each ERA, counting the number of partial squares contained within it. Should an ERA contain more than 4 partial squares, the smallest of these is again merged with a neighbouring partial square. In order to ensure that the downstream drainage direction between partial squares remains hydrologically correct, the newly created drainage direction table is used to ensure that partial squares are only merged with those into which their flow will subsequently converge in the next ERA downstream. The

$4 \cdot n + 1^{th}$ partial squares are thus merged without causing circular drainage direction problems.

Proceeding in this manner, the algorithm parses the drainage direction table, ensuring that CEQUEAU's input requirements are met. Once complete, the drainage direction table is used by the CEQUEAU Physiography Toolbox to assemble the remaining flow network topology data required by the C++ implementation of CEQUEAU. The index values of each partial square's upstream neighbours and the cumulative basin area upstream of each partial square are computed by recursively looping through the list of neighbouring partial squares. Additionally, the partial square and ERA index values are re-numbered in a reverse streamwise direction (downstream-most partial square/ERA is denoted as the 1st model square), in order to satisfy CEQUEAU's input requirements.

Physiographic data extraction and CEQUEAU structure assembly

After computing partial square drainage direction, the CEQUEAU Physiography Toolbox assembles the physiographic data necessary to run the hydrological model (Figure 3(c)). First, ERA and partial square altitudes are extracted directly from the watershed DEM. As per the requirements of CEQUEAU, ERA altitude is defined as the DEM elevation at the south-west most corner of each ERA, whereas partial square altitude is calculated as the mean elevation for all pixels bounded by the area of each partial square. Next, percentage forest and bare soil cover is attributed to each ERA and partial square by intersecting a raster land cover map with the ERA/partial square locations. The raster land cover routine is encoded to recognise land cover values corresponding to the North American Land-Change Monitoring System classes (Rasim *et al.* 2012) whereby areas of forest cover are defined as NALCMS classes 1–6 (various combinations of needleleaf, broadleaf evergreen, broadleaf deciduous and mixed forest cover) and bare soil is defined as NALCMS classes 7–13 (shrubland, grassland) and 15–17 (cropland, barren land and urban environments). However, simple modifications to the lines of code pertaining to the class definitions would enable the routine to be applied to raster land cover data from other sources (e.g., GeoBase 2009; Morton *et al.*

Uncorrected Proof

9   S. J. Dugdale *et al.* | Automating physiographic inputs to the CEQUEAU hydrological model   **Journal of Hydroinformatics | in press | 2017**

2011). Although raster land cover sources such as these also comprise information regarding surface waters and marshlands, such data are often of lower resolution than vector maps that are commonly available from national cartographic agencies. The CEQUEAU Physiography Toolbox is therefore designed to accept waterbodies and wetlands inputs formatted as ArcGIS polygon shapefiles. The toolbox's vector land cover routines function by looping though each ERA/partial square in *CEgrid/CPgrid* and finding their perimeter coordinates using a Moore-Neighbour contour tracing algorithm (Gonzalez *et al.* 2004). These perimeter vectors are subsequently transformed to map coordinates (see the section 'Physiographic data extraction and CEQUEAU structure assembly') and Boolean intersections performed to determine the areas of waterbody/wetland that are bounded by the nth ERA/partial square. The percentage of the ERA/partial square surface covered by waterbodies/wetlands can subsequently be tabulated.

Finally, all drainage direction and physiographic data necessary to run CEQUEAU are assembled into a MATLAB structure compatible with the C++ implementation of the model. The CEQUEAU Physiography Toolbox also outputs the final *CEgrid* and *CPgrid* rasters and a range of further variables which can also be displayed within a GIS environment to visualise the CEQUEAU grid topology. Provided with the appropriate meteorological variables (see Morin & Couillard (1990) and St-Hilaire *et al.* (2000) for details) and model parameters, the resulting MATLAB structure can now be executed using CEQUEAU in order to simulate discharge.
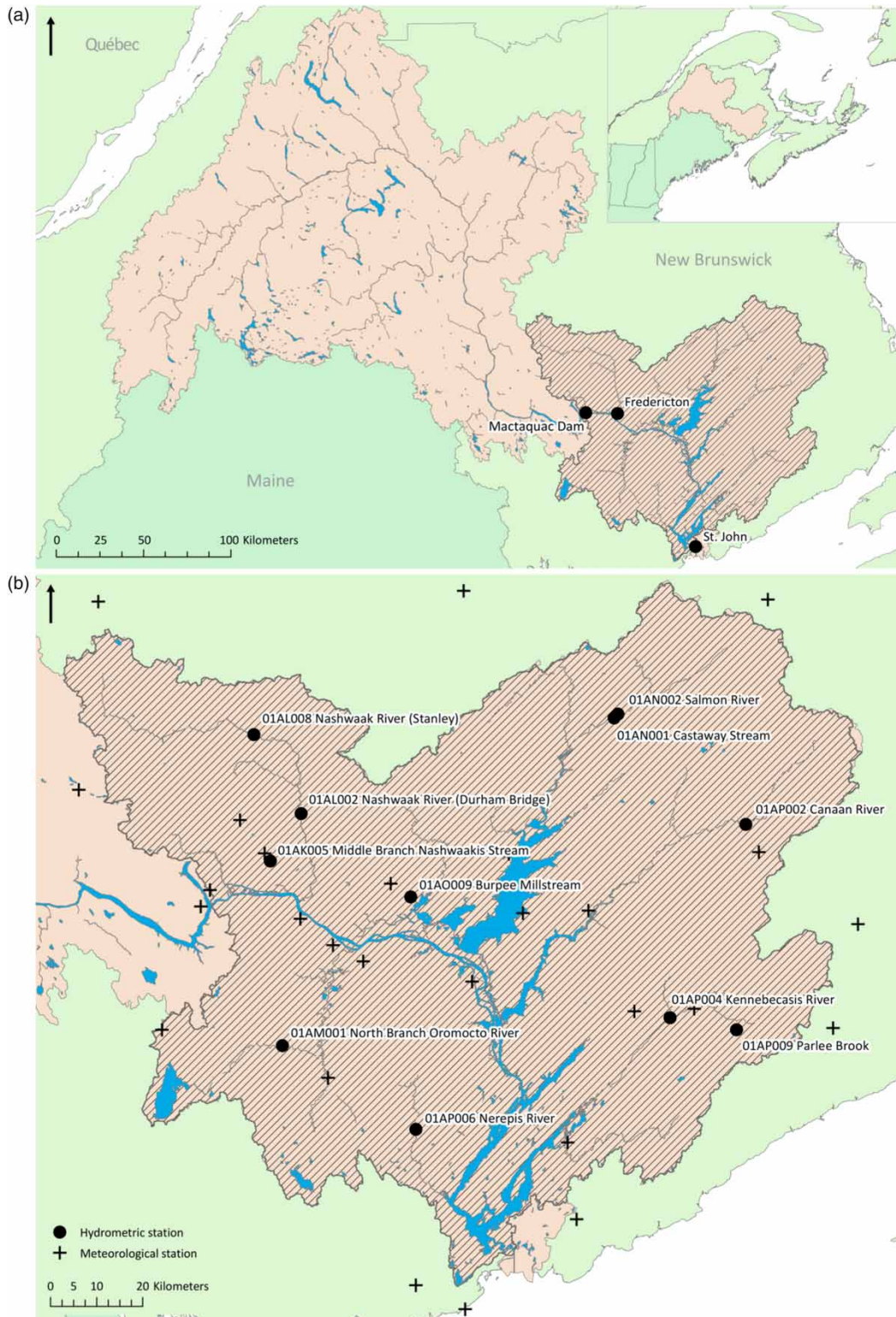
SENSITIVITY TESTING: THE LOWER SAINT JOHN RIVER BASIN

Although a range of previous studies (e.g., Charbonneau *et al.* 1977, 1979; Morin *et al.* 1986, 1987; Couillard *et al.* 1988; Ayadi & Bargaoui 1998; St-Hilaire *et al.* 2000, 2003; Turcotte *et al.* 2004; Dibike & Coulibaly 2005; Coulibaly 2008) detail the implementation of CEQUEAU on small and medium watersheds (10^0–10^3 km^2), the few that describe its use in larger systems (10^4–10^5 km^2) (Morin & Couillard 1990; Morin *et al.* 1994; Arsenault *et al.* 2014; Brisson *et al.* 2015) were compelled to use a relatively coarse

ERA grid (≥ 10 km^2), owing to the difficulty of manually defining the partial square drainage direction at finer resolutions. The use of the CEQUEAU Physiography Toolbox effectively removes this limitation, allowing for the implementation of CEQUEAU on large (i.e., $\geq 10^4$ km^2) watersheds at increased resolution (<10 km^2) with a minimum of user input. Indeed, the only limitations on watershed size and resolution are determined by the memory of the computer on which the toolbox is run. However, little is known concerning the relative performance of semi-distributed models such as CEQUEAU at different resolutions. In this section, we detail the use of the CEQUEAU Physiography Toolbox to implement CEQUEAU on a large eastern-Canadian watershed at successively decreasing ERA resolution, with a view to assessing variability in model performance at varying ERA size.

Study area

The Saint John River basin is a large socio-economically important river system straddling the Canadian provinces of Québec and New Brunswick and the US state of Maine (Figure 5(a)). It drains an area of approximately 55,110 km^2 from the Appalachian highlands (northern Maine, Québec) through the New Brunswick lowlands into the Bay of Fundy (Cunjak & Newbury 2005) at 45.258–66.088. Forest cover predominates within the watershed, composing 86.6% of the total land cover, although natural shrubland and grassland make up an additional 4.1%. Agricultural land (principally potato farming in the Saint John River valley) comprises another 4.8%, while urban land use contributes 0.5%. Lakes and wetlands present in the watershed (particularly in southern New Brunswick) contribute the remaining 3.9%. Most of the Saint John River basin experiences a humid continental climate (Cunjak & Newbury 2005), with coolest temperatures and moderate precipitation found in the northern and central upland portions of the watershed (Zelazny 2008), and dryer, warmer conditions found in the south-central lowlands ecoregions (Zelazny 2008). Temperatures in the southern-most extrema of the watershed are moderated by the proximity of the Bay of Fundy, which also contributes to high levels of summer precipitation (Zelazny 2008).

Uncorrected Proof

10   S. J. Dugdale *et al.* │ Automating physiographic inputs to the CEQUEAU hydrological model   │   **Journal of Hydroinformatics** │ in press │ 2017

Figure 5 │ (a) Saint John River basin showing the location of major water courses. Shaded region corresponds to section of watershed downstream of the Mactaquac Dam on which CEQUEAU was implemented. (b) Modelled section of watershed showing location of meteorological and hydrometric stations used to run model/calibrate model.

Uncorrected Proof

11   S. J. Dugdale *et al.* │ Automating physiographic inputs to the CEQUEAU hydrological model   **Journal of Hydroinformatics** │ **in press** │ **2017**

The Saint John River is impounded at several locations for the purposes of hydroelectric generation. The most significant of these is the Mactaquac Dam (Figure 5(a)), a 670 MW run-of-the-river dam located approximately 18 km to the west of Fredericton, New Brunswick at (45.952–66.872) with a mean daily outflow of 745.5 $m^3 s^{-1}$. The concrete portions of the dam are currently experiencing an alkali–aggregate reaction causing unwarranted expansion of the dam structure (Stantec 2015), necessitating its removal or replacement by 2030. As a result, efforts are currently underway to understand the potential impacts of dam removal/replacement on the physio-chemical and biological regimes of the lower Saint John River. One aspect of this work is the implementation of a hydrological model of the watershed in order to quantify the likely impacts of dam replacement/removal on the lower Saint John River's flow regime. CEQUEAU was chosen for this role due to both its ability to simulate flows in impounded water courses and its capacity for modelling stream temperatures, another key requirement of potential impact assessments regarding the removal or replacement of the dam.

CEQUEAU model implementation

CEQUEAU was implemented on the portion of the watershed downstream of the Mactaquac Dam (approximately 14,990 km^2 comprising the lowland and coastal regions of the watershed; Figure 5(b)). The model presently functions by imposing recorded discharges from the Mactaquac Dam at the partial square corresponding to the location of the Mactaquac Dam's outflow. Thus, simulated discharges within the Saint John River itself will largely be a reflection of these data (in addition to contributions from downstream tributaries). However, the various sub-basins represented within the model contain no such limitation. Given the large number of gauged unregulated tributaries present within the lower portion of the Saint John River basin, the current model presents a prime opportunity to assess the effect of changing ERA size by comparing modelled discharges to real data recorded by hydrometric stations on several of the Saint John River's sub-basins.

The CEQUEAU Physiography Toolbox was used to extract and assemble the drainage direction and physiographic data necessary to run the model. The flow accumulation and sub-basin rasters required by the toolbox were generated by applying the Arc Hydro Tools package (ESRI 2011) to a 1 arc-second (~25 m) SRTM DEM (Farr *et al.* 2007) of the study area obtained from the USGS' Earth Explorer service (http://earthexplorer.usgs.gov/). Raster land use data necessary for the calculation of forest and bare soil cover were acquired from the North American Land-Change Monitoring System (Rasim *et al.* 2012). Vector shapefiles used to compute percentage cover of waterbodies and wetlands were obtained from the Government of New Brunswick's GeoNB geospatial data portal (http://www.snb.ca/geonb1/e/index-E.asp). Meteorological data necessary to run the model (daily precipitation, minimum temperature and maximum temperature) were assembled from records from 27 Environment Canada weather stations in or near the lower Saint John River basin (Figure 5(b)). Preliminary findings indicated that this was the optimal number of stations for maximising model performance against all hydrometric stations.

Model calibration/validation strategy

Initial model calibration/validation was performed on a CEQUEAU model implemented using ERAs of 5×5 km (25 km^2), comprising 689 ERAs subdivided into 1,271 partial squares. The only discharge gauge in the main stem Saint John River is situated directly below the Mactaquac Dam. It was therefore not logical to calibrate/validate the model on the Saint John River itself as observed flows will be very similar to those imposed in the model square corresponding to the Mactaquac Dam outflow. Instead, observed flows used to calibrate/validate the model were assembled from 11 Environment Canada hydrometric stations (https://wateroffice.ec.gc.ca/) located within the various tributaries of the lower Saint John River (Figure 5(b)). Principal calibration/validation efforts were centred on data from hydrometric station 01AL002 on the Nashwaak River, a major tributary of the lower Saint John River. This is because the hydromorphology of the Nashwaak River sub-basin provides a good analogue of the main Saint John River area around Fredericton where dam replacement/removal will have the largest potential impact. However, the model was calibrated to ensure that it was also able to provide reasonable discharge simulations at the remaining

Uncorrected Proof

12   S. J. Dugdale *et al.* | Automating physiographic inputs to the CEQUEAU hydrological model   **Journal of Hydroinformatics** | in press | 2017

ten gauging stations. Model strength was assessed at each gauging station to a) assess model performance across the watershed and b) understand whether the effects of varying ERA size on model performance are basin size-dependent.

Good model calibration is reliant on detailed meteorological records from multiple locations. The optimum overlap between concomitant meteorological data and available hydrometric records for the lower Saint John River basin (i.e., the period for which the greatest number of weather stations and hydrometric stations yielded gapless observations) was for the period 1983–1994; this period was therefore selected for model calibration/validation. Seven years of data were used for model calibration and the remaining four for validation. Model calibration was achieved using a two-stage process. First, the various model parameters (see St-Hilaire *et al.* (2000) for details) were adjusted manually to ensure that the components of the hydrological model (water levels in saturated and unsaturated zones, lake and marsh levels, evaporation) stayed within realistic values. Following this manual phase, the CMA-ES optimisation algorithm (Hansen & Ostermeier 2001) was applied (on each occasion using 1,000 optimisation cycles) to further refine the calibration. This process was repeated 11 times choosing different seven- and four-year windows (e.g., calibrated on 1983–1989, validated on 1990–1993; calibrated on 1984–1990, validated on 1991–1994, etc.) to ensure that the calibration provided an accurate estimate of river discharge for all times within the 1983–1994 time period and was not overly influenced by extreme weather events during one or more time periods. Model quality was assessed by calculating the Nash–Sutcliffe model efficiency coefficient (*NSE*, Equation (5); Nash & Sutcliffe 1970), per cent (or normalised) root-mean-square error (*%RMSE*, Equation (6)) and per cent bias (*%Bias*, Equation (7)) for each four-year validation period, subsequently taking the mean of all validation periods:

$$NSE = 1 - \frac{\sum_{i=1}^{n} \left(Q_{obs,i} - Q_{sim,i}\right)^2}{\sum_{i=1}^{n} \left(Q_{obs,i} - \bar{Q}_{obs}\right)^2} \qquad (5)$$

$$\%RMSE = \frac{\left(1/n \sum_{i=1}^{n} \left(Q_{obs,i} - Q_{sim,i}\right)^2\right)^{0.5}}{\bar{Q}_{obs}} \times 100 \qquad (6)$$

$$\%Bias = \left| \frac{\sum_{i=1}^{n} \left(Q_{obs,i} - Q_{sim,i}\right)}{\sum_{i=1}^{n} Q_{obs,i}} \times 100 \right| \qquad (7)$$

where $Q_{obs,i}$ and $Q_{sim,i}$ are the observed and simulated mean discharges respectively on day i and $\bar{Q}_{obs}$ is the mean observed discharge for the validation period n.

Sensitivity analysis

Once calibrated, the model was iteratively re-implemented at ERA sizes ranging from 2.5×2.5 km (6.25 km^2, 2,592 ERAs, 4,293 partial squares) to 112×112 km ($12,544$ km^2, 4 ERAs, 16 partial squares) using the optimum calibration parameter set. Drainage directions computed by the CEQUEAU Physiography Toolbox at each ERA size iteration were checked against real hydrographic data from the New Brunswick Hydrographic Network (http://www.snb.ca/geonb1/e/index-E.asp) to ensure that flow was routed in the correct direction. The same calibration was used for all resolution iterations to ensure that any variability in model performance was solely due to changes in ERA resolution and not the re-calibration procedure. Furthermore, given the time involved in achieving a reasonable calibration, it was not feasible to re-calibrate the model at each ERA size iteration. ERA sizes between 2.5×2.5 km^2 and 5×5 km^2 were iterated at 0.5 km increments. However, it was not feasible to continue incrementing the resolution at 0.5 km steps up to 112×112 km because the time involved in the Arc Hydro Tools data preparation stages would have been prohibitive. Instead, resolutions between 5×5 km^2 and 10×10 km^2 were implemented at 1 km increments, 10×10 km^2–20×20 km^2 at 2 km increments, 20×20 km^2–40×40 km^2 at 4 km increments, 40×40 km^2–80×80 km^2 at 8 km increments and 80×80 km^2–112×112 km^2 at 16 km increments. Exploratory analyses indicated that these increments were sufficient to encapsulate variability in model strength as a function of ERA size, and that minor variations in increment at larger ERA sizes ($>20 \times 20$ km) did not produce notably different results. At the largest ERA size iteration, the entire watershed was covered by only four ERAs, meaning that all gauged sub-basins within the watershed were essentially functioning as lumped models. It was not possible to increase the ERA size further,

Uncorrected Proof

13   S. J. Dugdale *et al.* | Automating physiographic inputs to the CEQUEAU hydrological model   Journal of Hydroinformatics | in press | 2017

as any simulated flows would have essentially reflected that of the entire watershed, far in excess of those produced by the various gauged sub-basins.

In order to conduct a sensitivity analysis, it was necessary to relax the rules governing the generation of partial squares within the CEQUEAU Physiography Toolbox. CEQUEAU generates discharge simulations for the outlet of each partial square within the model grid. However, as ERA size increases, the probability of a partial square's outlet coinciding with the real geographic location of a hydrometric station decreases. This means that for larger ERA sizes, simulated flows may grossly under- or over-represent observed data in cases when the partial square's outlet is substantially upstream or downstream of the hydrometric station's true location. In order to resolve this problem, an additional routine was added to the CEQUEAU Physiography Toolbox that forced the creation of a partial square whose outlet corresponded to the location of each hydrometric station. If this additional partial square violated the condition of a maximum of 4 partial squares per ERA, one or more of the other partial squares within the ERA were merged in a hydrologically correct manner, following the process described in 'Physiographic data extraction and CEQUEAU structure assembly'. The inclusion of this routine allowed for the generation of flow simulations that can be compared with observations for each gauged sub-basin, even at the largest ERA sizes tested here.

RESULTS

Hydrometeorological characterisation of calibration/validation period

Mean daily minimum and maximum temperature during the 1983–1994 calibration/validation period was 0.3 °C and 11.0 °C, respectively, compared to the 50-year means (1964–2013) of 0.1 °C and 10.8 °C, respectively. Absolute minimum temperature during the study period was −41.0 °C; the absolute maximum temperature observed was 37.5 °C. Mean annual precipitation (liquid and snow-equivalent) across the study area was 1,143.2 mm, close to the 50-year mean of 1,170.5 mm. Persistent snow cover was usually present between December and April, with mean annual snowfall of 2,399.7 mm. This represents a small reduction on the 50-year mean of 2,493.2 mm. Mean daily discharge at the principal calibration site (Nashwaak River) was 33.9 $m^3 s^{-1}$, again slightly reduced in comparison to the 50-year mean of 36.2 $m^3 s^{-1}$. Typically, the annual hydrograph for water courses in the Saint John River basin comprises a primary snowmelt-driven peak between April and May and a smaller secondary peak in October–December associated with late autumn/winter rainfall (see Figure 6 for example). Lower flows persist between late June and late September, although storm events occasionally produce high discharge events in the summer months.
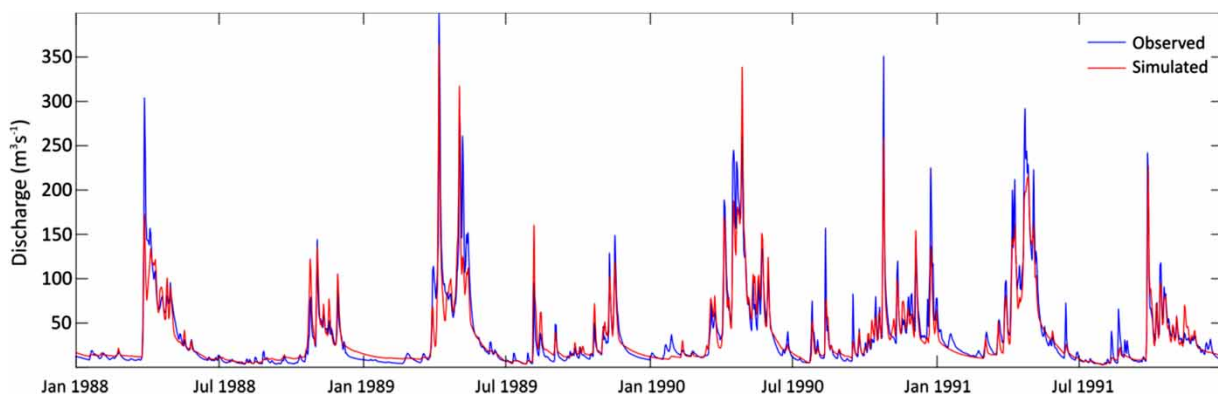
Figure 6 | Example of four-year model validation window (1988–1991) showing observed and simulated discharges for Nashwaak River (Durham Bridge gauge, 01AL002).

Uncorrected Proof

14   S. J. Dugdale *et al.* | Automating physiographic inputs to the CEQUEAU hydrological model   **Journal of Hydroinformatics | in press | 2017**

Initial model calibration/validation results

Following calibration, the initial 5×5 km CEQUEAU model of the lower Saint John River basin was found to simulate discharge with reasonable accuracy in most sub-basins (Table 2). Taking the mean of all four-year validation periods, it performed especially well at the Nashwaak (Durham Bridge) hydrometric station (01AL002; NSE = 0.86), unsurprising given that principal calibration efforts were focused on this sub-basin. A similarly high Nash–Sutcliffe value was achieved at another station located further upstream on the Nashwaak River (01AL008; NSE = 0.84), indicating that the model is able to simulate flows in the upper reaches of the Saint John River basin with a good degree of accuracy. %RMSE was lowest at the two hydrometric stations situated on the Nashwaak River, again suggesting that simulation quality was highest in this section of the watershed. Inspection of the simulated hydrograph (Figure 6) indicates that the model underestimates some peak flow events (especially during spring snowmelt); however, summer low flows are generally well simulated. This was deemed acceptable as the principal purpose of the model was the simulation of summer low flows that could engender high water temperature events. Model strength diminishes as a function of both distance from the principal calibration site and sub-basin size, with more distant hydrometric stations (e.g., 01AP004, 01AP006) and smaller sub-basins (i.e., 01AK005, 01AN001, 01AP009) reporting Nash–Sutcliffe values between 0.51 and 0.65. However, no significant geographic or sub-basin size trends in model performance were observed among the other validation sites. Furthermore, no significant correlation was observed between the various model strength criteria and sub-basin land-use (Table 2).
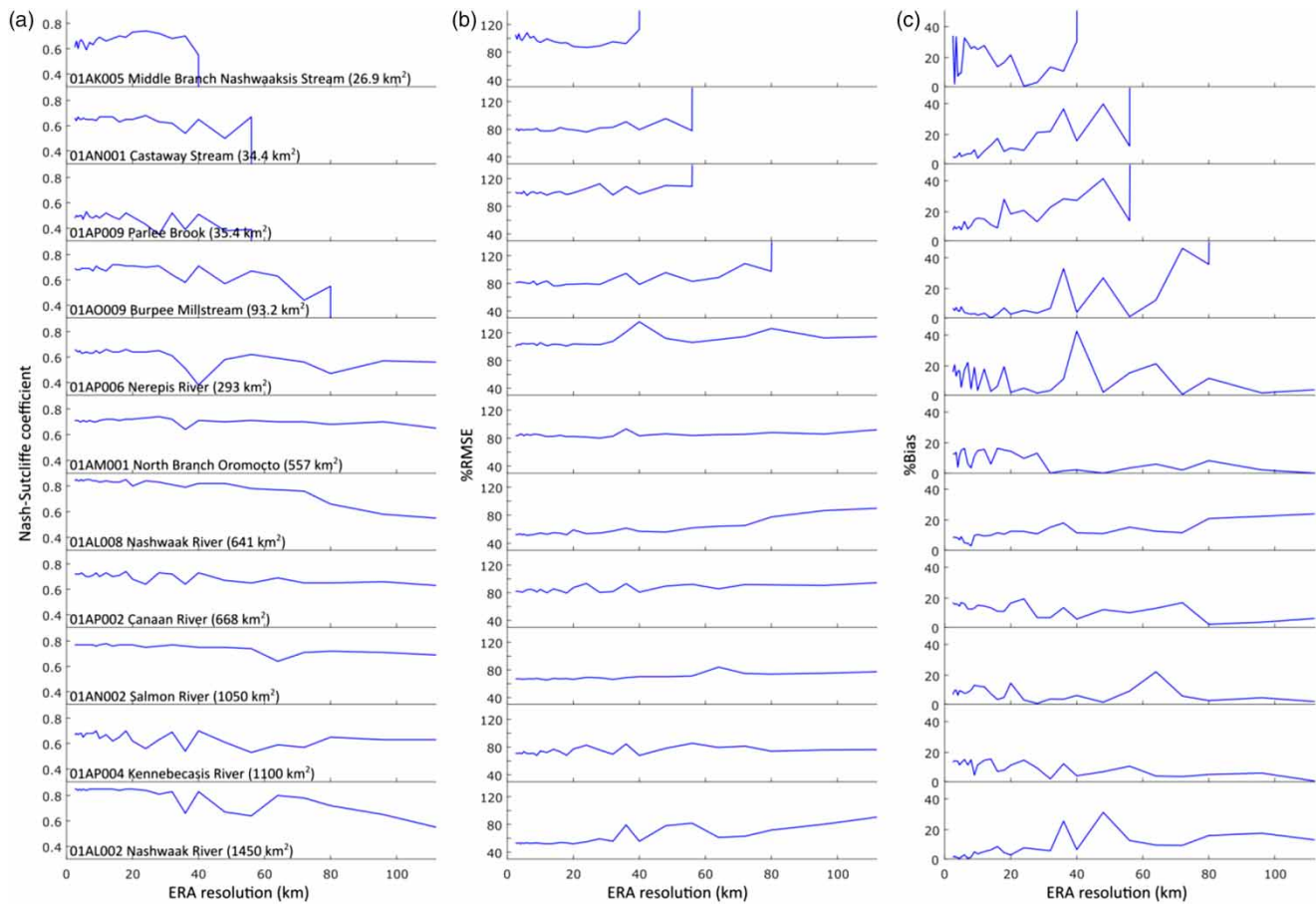
Sensitivity analysis results

Model performance

Results of the sensitivity analysis show that CEQUEAU is able to produce reasonable simulations of discharge even at relatively coarse ERA resolutions (Figure 7(a)–7(c)), exhibiting only a minor reduction in NSE compared to the original model (mean NSE decrease between original model and coarsest ERA resolution = 0.20). For ERA sizes $\lesssim 20 \times 20$ km, this reduction is an order of magnitude smaller (mean NSE decrease = 0.03). While NSE and %RMSE are generally stable at low ERA sizes (e.g., $\lesssim 20 \times 20$ km^2), % Bias fluctuates by a larger amount, especially in sub-basins 01AK005, 01AP006 and 01AM001. Model strength at most stations subsequently decreases as a function of increasing ERA size, with a reduction in NSE and an

Table 2 | Calibration results of initial 5×5 km resolution CEQUEAU model implemented on lower Saint John River basin

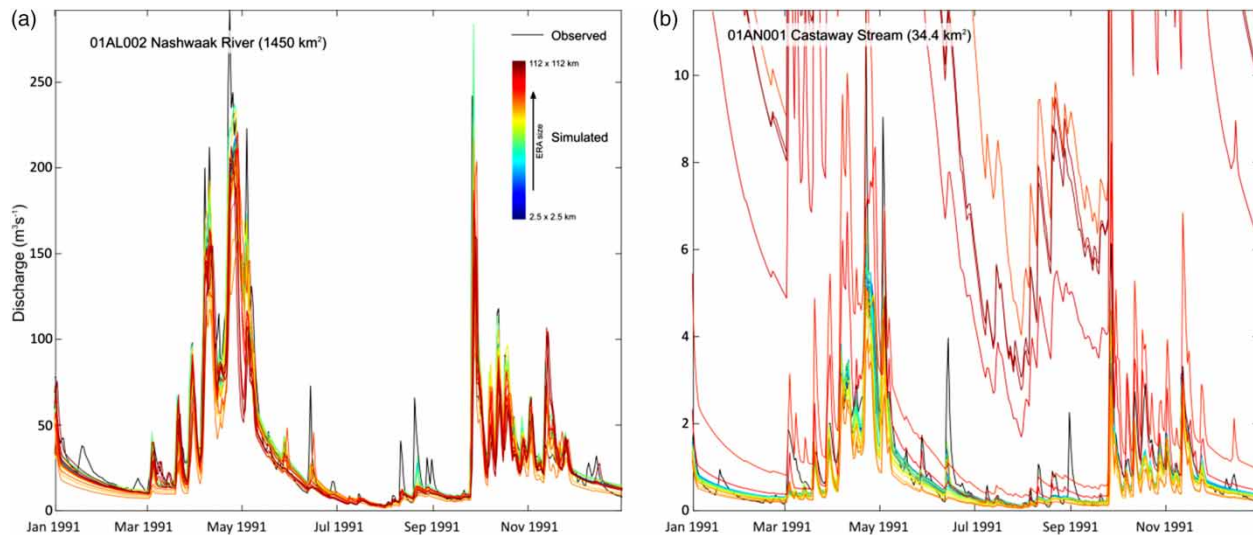| Station number	Station name	Sub-basin size (km^2)	% water-bodies	% forest	% marshland	% bare soil	NSE	% RMSE	% Bias
01AK005	Middle Branch Nashwaaksis Stream	26.9	0.23	95.47	3.42	0.85	0.60	101.05	39.06
01AN001	Castaway Stream	34.4	0.02	82.62	13.94	3.80	0.65	80.22	6.60
01AP009	Parlee Brook Below	35.4	0.06	96.86	0.33	3.22	0.51	88.79	3.70
01AO009	Burpee Millstream	93.2	0.09	82.42	14.64	2.90	0.67	83.37	4.22
01AP006	Nerepis River	293	0.31	82.92	5.39	11.39	0.64	102.52	5.99
01AM001	North Branch Oromocto River	557	8.02	76.58	11.62	3.79	0.70	85.19	3.06
01AL008	Nashwaak River (Stanley)	641	0.75	93.84	3.55	1.89	0.84	49.82	7.93
01AP002	Canaan River	668	0.26	80.01	12.70	6.99	0.70	84.98	12.39
01AN002	Salmon River	1,050	0.40	82.88	14.02	2.73	0.76	69.04	4.82
01AP004	Kennebecasis River	1,100	0.28	83.58	1.65	14.47	0.63	76.46	13.57
01AL002	Nashwaak River (Durham Bridge)	1,450	0.51	94.07	3.36	2.07	0.86	48.48	1.02

Figure 7 │ Sensitivity analysis showing variation in NSE (a), %RMSE (b) and %Bias (c) as a function of ERA resolution.

associated increase in %RMSE and %Bias, although the opposite trend is observed at station 01AK005 (the smallest gauged watershed). Although the reduction in model strength is generally steady, several sub-basins exhibit fluctuations to decreased NSE values at ERA sizes $\cong 36 \times 36$ km and 40×40 km. Following these fluctuations, model strength at most stations continues to decrease gradually. However, stations 01AL008, 01AP004 and 01AL002 show several further fluctuations, with 01AL008 showing instability at $\sim 80 \times 80$ km while 01AP004 and 01AL002 experience variability (reduced NSE, increased %RMSE/%Bias) between 48×48 km and 64×64 km.

The overall trend highlights a general pattern of gradually decreasing model strength as a function of increasing ERA size. However, while there is no statistically significant relationship between basin size and the net reduction in NSE from the smallest ERA size to the largest, the four

smallest gauged sub-basins in the lower Saint John River basin (01AK005, 01AN001, 01AP009 and 01AO009) all show a large drop in NSE to extreme negative values (and associated increases in %RMSE and %Bias) when ERA size increases past a certain threshold. This result points to the existence of a fundamental threshold at which the ERA size is too large to allow for flow simulations in smaller sub-basins. While the ERA size threshold at which this NSE drop-off occurs is likely a function of the sub-basin size, insufficient data exist to allow for statistical inferences to be drawn. However, results do show that the smallest of the four sub-basins (01AK005) exhibits the earliest drop-off in model performance ($\sim 40 \times 40$ km), while the largest of the four (01AO009) continues to yield valid model predictions at larger ERA sizes, stopping at $\sim 80 \times 80$ km. Simulated hydrographs associated with the sensitivity testing shed light on this NSE drop-off (Figure 8). In the

Uncorrected Proof

16   S. J. Dugdale *et al.* | Automating physiographic inputs to the CEQUEAU hydrological model   **Journal of Hydroinformatics** | in press | 2017

Figure 8 | Examples of simulated hydrographs produced by sensitivity testing for (a) larger sub-basin (Nashwaak River) and (b) smaller sub-basin (Castaway Stream).

larger sub-basins (e.g., 01AL002; Figure 8(a)), results of the sensitivity testing indicate that simulated flows most closely replicate observed data at smaller ERA sizes. As ERA size increases, low flows are generally underestimated and high flows are both over- and underestimated, with no clear trend regarding snowmelt or precipitation driven events. However, simulated flows nevertheless stay relatively close to observed values, even at coarse ERA resolutions. Conversely, in the smaller sub-basins (e.g., 01AN001; Figure 8(b)), the same pattern (under-prediction of low flows and under/over prediction of high flows with increasing ERA size) is initially present, but simulated flows increase by an order of magnitude once ERA size passes a given threshold, reaffirming the observed drop-off in NSE.

Representation of basin physiography and meteorology

CEQUEAU's physiographic representation of the lower Saint John River basin varies as a function of increasing ERA size (Figure 9). This is because any change in the size of ERAs will propagate through as changes in the size, shape and position of partial squares computed by the CEQUEAU Physiography Toolbox. In terms of sub-basin size, results show that for larger sub-basins ($\gtrsim 500$ km^2), the ratio of real to computed sub-basin size is relatively stable, even at moderately coarse ERA resolutions ($\cong 60 \times 60$ km), although minor fluctuations in computed basin

size ($\sim 5\%$) are occasionally present. However, the smaller sub-basins are considerably less stable, even at relatively fine ERA resolutions. Figure 9(a) shows the presence of repeated sub-basin size fluctuations, sometimes by as much as 20%. Similar to that noted for the model performance metrics, sub-basins 01AK005, 01AN001, 01AP009 and 01AO009 also demonstrate the existence of a threshold at which computed basin size jumps by several orders of magnitude when ERA size passes a certain point.

Figure 9(b) shows that computed land-use follows similar trends to that observed for sub-basin size. However, the magnitude of land-use fluctuation is dependent on land-use type. Results show that bare soil cover is the most variable in response to increasing ERA size. However, this is presumably due in part to the spatially variable nature of bare soil cover present within the watershed; a minor change in the size or shape of a sub-basin could result in a relatively large difference in computed bare soil coverage should the boundary of the sub-basin lie close to an area of bare soil. Conversely, the forest/marshland classes more closely mirror trends in computed sub-basin size, probably because the distribution of these classes remains relatively constant throughout the lower Saint John River basin.

CEQUEAU interpolates meteorological inputs across the ERA grid using a nearest neighbours' approach. Changes in grid resolution will therefore lead to variability
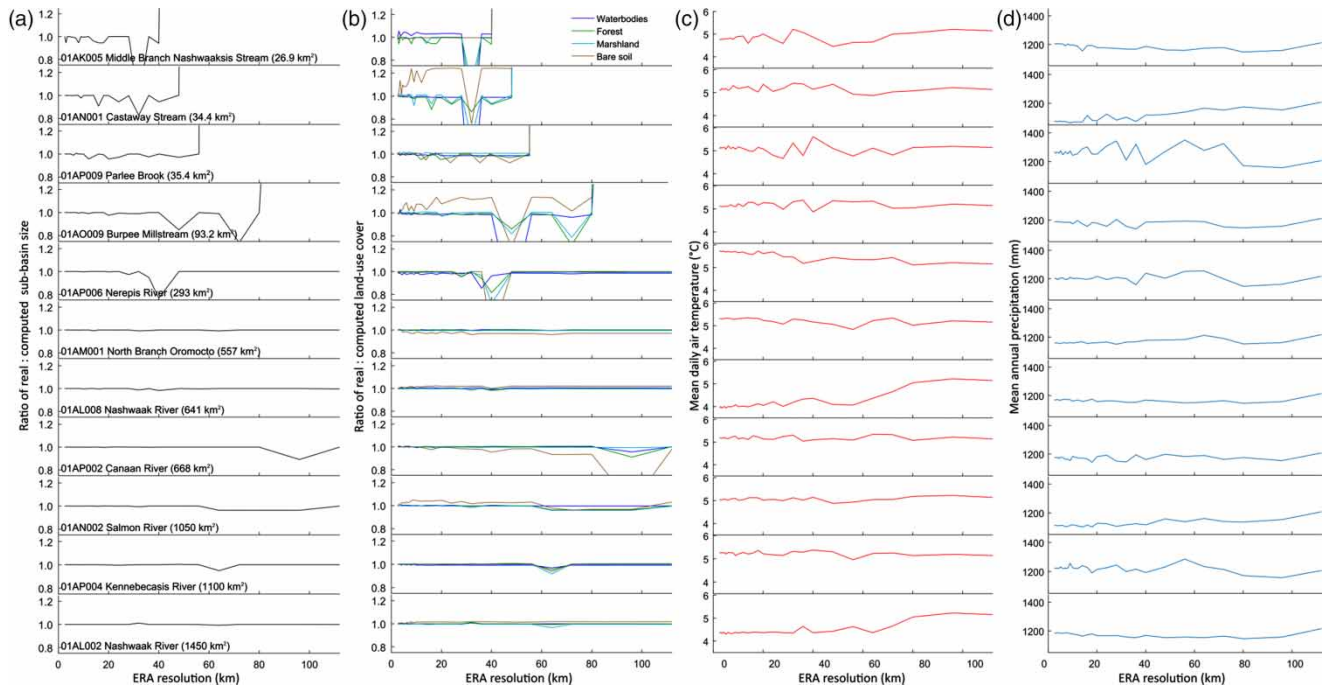
Figure 9 │ Variation in (a) ratio of real to computed sub-basin size, (b) ratio of real to computed land-use, (c) interpolated mean daily air temperature and (d) interpolated mean annual precipitation as a function of ERA resolution.
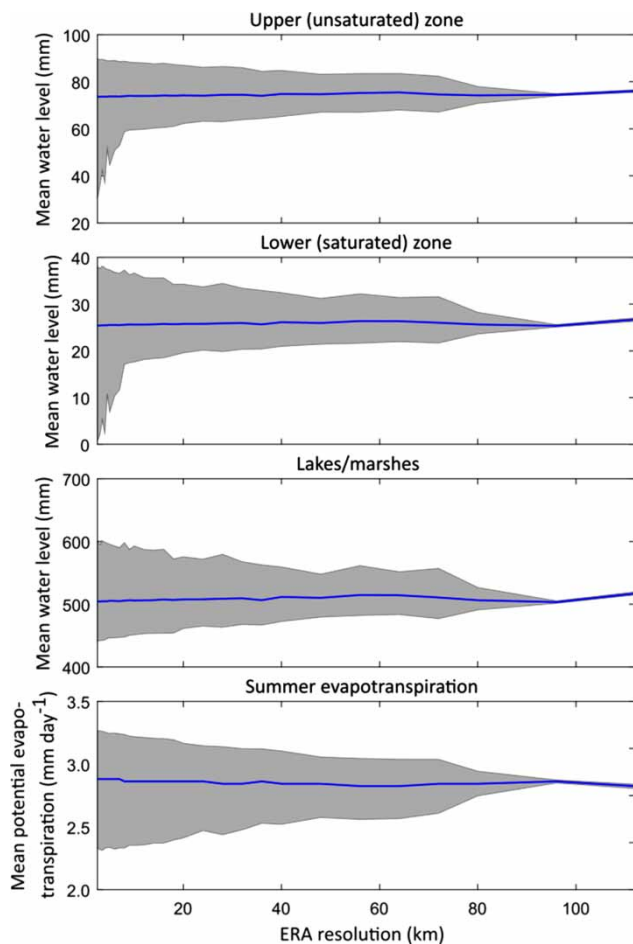
in the meteorology applied to a sub-basin, as the nearest meteorological observations to a given ERA will change depending upon the size and layout of the ERA grid. Figure 9(c) and 9(d) demonstrate how variability in ERA resolution drives changes in sub-basin air temperature and precipitation. Results indicate that interpolated meteorology in the larger sub-basins ($\gtrsim$500 km^2; with the exception of 01AP004) is relatively stable at lower ERA sizes, although a gradual increase in precipitation and temperature is apparent as ERA size approaches its maximum. However, the smaller sub-basins are less stable, exhibiting periodic air temperature and precipitation fluctuations of $\pm\sim$0.5 °C and $\sim$100 mm, respectively. Despite these initial fluctuations, meteorology within all sub-basins eventually stabilises as spatial variability in meteorology across the lower Saint John River basin is lost as a function of ERA size increase.

Inspection of the physiographic and meteorological trends (Figure 9) indicates that some sub-basins appear to show small but sustained increases in sub-basin size, land-use, precipitation or temperature as a function of ERA size. However, this increase was not observed to be significant ($p < 0.01$) for the vast majority of sub-basins. Furthermore, there appears to be no consistency regarding which sub-basins demonstrate significant correlations between ERA size and a given physiography/meteorological variable. Therefore, there does not appear to be any clear ERA-size dependency effect with regards to the physiographic and meteorological data computed with the CEQUEAU Physiography Toolbox.

Inspection of the values calculated by CEQUEAU's production function (water storage within the conceptual reservoirs; see the section 'The CEQUEAU model') shows that despite the fluctuations in basin area/land-use, CEQUEAU's production function is relatively stable (Figure 10). While the range of values describing storage within the upper and lower zones, lake/marshes and potential evapotranspiration clearly narrows as a function of increasing ERA size, this is simply a reflection of the decrease in basin heterogeneity as the ERA grid resolution coarsens. Conversely, each mean varies very little as ERA size increases, indicating that the model continues to function correctly (i.e., producing realistic water volumes) even at extremely coarse ERA resolutions.

Uncorrected Proof

18   S. J. Dugdale *et al.* │ Automating physiographic inputs to the CEQUEAU hydrological model   Journal of Hydroinformatics │ in press │ 2017

Figure 10 │ Variation in modelled water volume and potential evapotranspiration in CEQUEAU's conceptual reservoirs as a function of increasing ERA size. The shaded area depicts the range of variability in water volume/potential evapotranspiration across the CEQUEAU model grid.

DISCUSSION

CEQUEAU physiography toolbox computation times and ease of implementation

Using a 2014 specification high-powered laptop computer (Intel Core i7-4700MQ 2.4 GHz quad-core CPU, 16Gb RAM, solid-state drive with 520 MB/s read and 400 MB/s write speed), implementation of CEQUEAU on the lower Saint John River basin took approximately 1 h 24 min at the smallest ERA size tested in the present study (2.5×2.5 km) using the CEQUEAU Physiography Toolbox. At coarser ERA resolutions, the time needed to implement the model is even further reduced (on the order of

5–10 min). This represents a very substantial time-saving over previous manual model implementation methods; indeed, conversations with researchers who have previously manually implemented CEQUEAU on similar sized watersheds (i.e., 10^4–10^5 km^2) indicate that the time required to compute the partial square drainage direction and input the required physiography data into CEQUEAU is on the order of 2–6 weeks (Boisvert, J., Ouellet-Proulx, S., personal communications).

In addition, the CEQUEAU Physiography Toolbox requires relatively low levels of GIS and MATLAB expertise to use. The appropriate GIS inputs can be prepared simply by following established Arc Hydro Tools procedures (e.g., Maidment 2002), while execution of the toolbox within MATLAB consists simply of calling the master CEQUEAU Physiography Toolbox function. A simple user guide to assembling the necessary data inputs and executing the CEQUEAU Physiography Toolbox is provided with the MATLAB .m files for download. Given that most CEQUEAU users working with the C++ implementation of the model will already be familiar with MATLAB, there is considerable potential for further toolbox development, as most users will possess the knowledge to modify and add to the code detailed here. The CEQUEAU Physiography Toolbox therefore represents a substantial benefit in terms of time saved and ease of implementation when compared to previous manual methodologies.

Hydrological model of lower Saint John river basin

In the initial 5×5 km CEQUEAU model, model strength was observed to vary across the watershed, diminishing as a function of both distance from the principal calibration site and sub-basin size. This spatial trend is likely a function of variability in basin physiography and meteorology. Although no significant correlation was observed between model strength and sub-basin land-use, CEQUEAU does not account for more detailed physiographic variability such as soil type or watershed geology. Minor spatial variability in model strength among the more well-conditioned areas of the watershed therefore presumably results from localised hydrogeological differences between sub-basins. Furthermore, the lower calibration/validation NSE towards the southern extent of the watershed (i.e., hydrometric

Uncorrected Proof

19   S. J. Dugdale *et al.* | Automating physiographic inputs to the CEQUEAU hydrological model   **Journal of Hydroinformatics | in press | 2017**

stations 01AP009, 01AP006 and 01AP004) is presumably due to the increasingly different hydrogeology (e.g., Zelazny 2008) of the northern and southern sections of the watershed. The maritime climate of the southern regions of the watershed (characterised by increased precipitation and cooler summer temperatures; Cunjak & Newbury 2005; Zelazny 2008) is also likely to contribute to reduction in NSE reduction compared to the drier and warmer areas where principal calibration took place. Nevertheless, the generally high NSE and relatively low %RMSE and %Bias metrics obtained in the upper sections of the modelled watershed indicate that the model has good potential for simulating and understanding flows in the regions of most interest for understanding the impacts of the removal of the Mactaquac Dam (i.e., the upstream sections of the watershed below the Mactaquac Dam).

Sensitivity analysis: understanding and implications

Results indicate that CEQUEAU is able to provide realistic flow simulations at a wide range of resolutions. This is particularly the case at finer resolutions ($\lesssim 20 \times 20$ km) where partial squares are small enough to preserve maximum spatial variability in basin physiography. The slight but persistent reduction in model strength as ERA size increases results from the loss of spatial variability in the input physiographic and meteorological data. As ERA size grows, landuse is progressively averaged across the sub-basin and the meteorological interpolation progressively incorporates data from more distant stations (see the section 'Sensitivity analysis results'). This spatial 'averaging' explains the reduction in %Bias as ERA size increases. Indeed, as ERA sizes approach the maximum (112×112 km), all gauged sub-basins within the lower Saint John River basin begin to function as lumped models (i.e., each sub-basin is represented by a single partial square and all meteorological data are averaged). This indicates that, at least in this study, CEQUEAU is capable of reasonable discharge simulations even with an extremely coarse representation of basin physiography. Although lumped hydrological models have previously been shown to perform with a similar degree of accuracy to their semi-distributed or distributed counterparts (e.g., Beven 1989; Reed *et al.* 2004; Carpenter & Georgakakos 2006; Khakbaz *et al.* 2012), it is perhaps

surprising that CEQUEAU performs well as a lumped model, given that it was not designed to function in such a way. This may be a 'hidden' benefit of CEQUEAU's 1970s-era programming; the limited computing power of this time meant that model grids were generally coarse, requiring CEQUEAU to produce robust results even in sub-optimal circumstances.

Aside from the general trend of decreasing model strength as a function of ERA size, results of the sensitivity analysis show occasional fluctuations in NSE, %RMSE and %Bias. These fluctuations are caused by both a) variability in sub-basin size and land-use computed by the CEQUEAU Physiography Toolbox (Figure 9(a) and 9(b)) and b) variability in sub-basin meteorology resulting from changes in ERA size (Figure 9(c) and 9(d)). Variability in computed sub-basin size and land-use primarily results from the inclusion in the CEQUEAU Physiography Toolbox of routines that merge small partial squares (see the section 'Physiographic data extraction and CEQUEAU structure assembly'). These routines (essential for the production of hydrologically correct CEQUEAU model grids) mean that the computed area of a sub-basin will increase or decrease depending upon whether partial squares are merged into or away from a sub-basin. Such variability in sub-basin area and land-use will inevitably produce changes in modelled flows, explaining the model strength fluctuations observed in this study. Variability in sub-basin meteorology occurs when changes in the size and position of ERAs alters the selection of meteorological stations used in the nearest-neighbour interpolation. Changes to the meteorological interpolation will intrinsically alter the sub-basin's hydrological budget. Any variability in sub-basin meteorology will therefore be mirrored in the simulated flow data, leading to NSE/%RMSE fluctuations like those observed here.

Some of the model strength fluctuations occurring at larger ERA sizes (e.g., 36×36 km to 80×80 km) were observed to persist across several sub-basins. This persistence is the result of two processes. First, the partial square merging process intrinsically results in the simultaneous increase in size of one sub-basin and decrease in size of another, meaning that both sub-basins will exhibit a model strength fluctuation. Second, at coarser ERA resolutions, meteorology in neighbouring ERAs will become

Uncorrected Proof

20   S. J. Dugdale *et al.* | Automating physiographic inputs to the CEQUEAU hydrological model   **Journal of Hydroinformatics | in press | 2017**

increasingly similar, meaning that any variability in the meteorology will produce model strength fluctuations across multiple sub-basins. While NSE/%RMSE only fluctuates by relatively small amounts, considerably more variability was noted in the computed %Bias values. This can again be explained as a result of variability in sub-basin size and meteorology; while a small change in sub-basin size or meteorology will not necessarily drive a substantive change in the precision of the model, it will generally result in a systematic over- or underestimation of flows. Such an occurrence will therefore engender a larger change in %Bias than in NSE/%RMSE. Normally, these biases are minimised during model calibration. However, because the same parameter set was used for the entire sensitivity analysis, the calibration is not optimised for each ERA size iteration, meaning that bias was larger than during CEQUEAU's 'normal' usage.

That CEQUEAU ceased to produce reasonable discharge simulations for the smaller sub-basins (01AK005, 01AN001, 01AP009 and 01AO009) at larger ERA sizes is unsurprising. Rather, it is in fact noteworthy that they continued to produce viable discharge simulations at all. This is presumably due to the addition of the routine in the CEQUEAU Physiography Toolbox that acts to force the creation of a partial square corresponding to the area of each gauged sub-basin (see the section 'Sensitivity analysis'), which enabled CEQUEAU to simulate flows within an area corresponding to a very small fraction of the total ERA size. The fact that CEQUEAU eventually failed to simulate flows in these small sub-basins at larger ERA sizes is because the routine which forces the creation of a partial square is specifically programmed not to preserve this new partial square, should its inclusion preclude the hydrologically correct drainage direction computation. At such large ERA sizes, the inclusion of a small partial square in place of a legitimate partial square could result in a circular drainage loop, so any such small partial squares are removed. Such occurrences will intrinsically result in that sub-basin being merged with a much larger partial square, leading to the gross overestimation of discharge observed in smaller sub-basins once ERA size passes a given point (Figure 8).

Taken together, these results indicate that provided appropriate steps are taken to verify that sub-basin size and land-use computed using the CEQUEAU Physiography Toolbox are within acceptable limits, CEQUEAU is able to simulate realistic flows at a wide range of ERA resolutions without introducing substantial additional error into the model. Although lower resolution models are faster to implement using the CEQUEAU Physiography Toolbox, the speed of implementation even for increased resolution models (e.g., 2.5×2.5 km) is such that we advise the use of smaller ERAs in order to produce fine scale flow simulations across the entire watershed. Even in the case of a relatively large and complex watershed like that detailed here (comprising 2,592 ERAs divided into 4,293 partial squares), the simulation of daily flows over a 6,240 day period requires little time to compute ($\sim$55 s using a 2014 specification high-powered laptop computer). In cases where this computation time precludes the use of an iterative optimisation routine (e.g., CMA-ES) for model calibration, we advocate a two-step process whereby a lower resolution model is used for initial model calibration and optimisation, but final simulations are run using its higher resolution counterpart. Given that the *Riverscapes* concept (Fausch *et al.* 2002) advocates the study of fluvial process–biota interactions at a scale of reference amenable to that used by fluvial organisms, the ability to model flows at the increased resolutions demonstrated here could be of considerable benefit to river scientists and managers.

Limitations and future improvements

Despite the promising results detailed here, it is pertinent to discuss potential limitations regarding the use of the CEQUEAU Physiography Toolbox. Of particular interest is the resolution of the GIS inputs to the CEQUEAU Physiography Toolbox. Although the DEM and land cover data used here were found to be sufficient for implementing CEQUEAU on a large watershed, their resolution posed problems when attempting to apply the CEQUEAU Physiography Toolbox to a much smaller watershed at higher resolution than that discussed here (0.25×0.25 km; Ouellet-Proulx, S; personal communication). Here, the particularly fine ERA size meant that many partial squares only covered $\sim10^1$ pixels in the *FAC* raster, and therefore sometimes only contained flow accumulation values equal to zero. Although the CEQUEAU Physiography Toolbox

Uncorrected Proof

21   S. J. Dugdale *et al.* | Automating physiographic inputs to the CEQUEAU hydrological model   **Journal of Hydroinformatics | in press | 2017**

contains a routine that addresses this issue (see the section 'Physiographic data extraction and CEQUEAU structure assembly'), the presence of multiple zero-accumulation partial squares can nonetheless cause the CEQUEAU Physiography Toolbox to produce erroneous results. This issue can be addressed simply by choosing a higher resolution DEM product, indicating that efforts should be made to acquire the highest possible resolution input data when desired model grid resolution is particularly fine.

In this study, we used Arc Hydro Tools (Maidment 2002) to generate the input *CAT/FAC* rasters used by the CEQUEAU Physiography Toolbox. However, the algorithm used by Arc Hydro Tools to compute raster flow directions (termed the D8 flow direction algorithm; see Tarboton 1997) has been shown to produce an overly simplistic representation of flow, particularly in flat terrain (Tarboton 1997; Schwanghart & Kuhn 2010). This algorithm is likely partially responsible for the problems linked to DEM resolution described above. Furthermore, while the use of Arc Hydro Tools to assemble the necessary data for the CEQUEAU Physiography Toolbox is useful for visualisation purposes, it is a proprietary software package, and may not be available to some researchers. Initial attempts to resolve these issues by implementing more advanced drainage network analysis algorithms (e.g., Tarboton 1997) directly within the CEQUEAU Physiography Toolbox proved computationally intensive. It was therefore decided that separating the drainage network analysis from the toolbox and allowing inputs from other software to be used provided the best compromise between flexibility, processing speed and ease of GIS data manipulation. However, future development of the CEQUEAU Physiography Toolbox should focus on integrating these algorithms within the toolbox itself.

In conducting the sensitivity analysis, we chose not to recalibrate the model at each ERA size iteration. This choice was one of expediency. Even using optimisation routines (e.g., Hansen & Ostermeier 2001), proper calibration of CEQUEAU can be time-consuming and would not have been feasible due to the number of re-calibrations necessitated by the ERA size iterations. We do not believe that this decision has made a substantial difference to the results of this study, primarily because most of CEQUEAU's model parameters are reasonably scale-independent. This means that the effects of variations in sub-basin size, land-use and meteorology (discussed in 'Sensitivity analysis: Understanding and implications') will likely have had a much greater impact on simulated water volumes than minor changes in model calibration. This conclusion is supported by the extent to which the computed model performance criteria mirror variations in physiography/meteorology (Figure 9). Nevertheless, it is possible that non-recalibration might account for some of the minor model strength fluctuations that do not correspond to variations in computed physiography/meteorology. Furthermore, given the fact that calibration efforts were focused on achieving maximum model strength in a relatively large sub-basin (Nashwaak River), the effect of this non-recalibration (particularly at larger ERA sizes) was likely greatest in smaller sub-basins, potentially accounting for some of the reduction in model strength observed at these sites. It is therefore possible that the results of this study may indeed change slightly were the model to be re-calibrated, particularly in smaller sub-basins. We therefore suggest that any future sensitivity testing using either CEQUEAU or the CEQUEAU Physiography Toolbox should aim to determine the relative important of re-calibration at different resolutions in order to address any uncertainties raised by this study.

CONCLUSION

In this article, we have documented the development of a GIS-like MATLAB toolbox for the computation of flow routing and assembly of physiographic data necessary to implement the CEQUEAU hydrological model on a given watershed. Results indicate that the toolbox is both fast and relatively simple to execute and has considerable potential for reducing the time needed to implement CEQUEAU in a new location. The toolbox was used to test CEQUEAU's sensitivity to model grid resolution by successively increasing ERA size. Our findings show that CEQUEAU is highly robust to changes in ERA resolution, suggesting that model implementation efforts are best focused on achieving an optimum model parameter set rather than targeting the finest possible resolution. However, given the ease of implementation of the CEQUEAU Physiography Toolbox, it should now be possible to achieve both, using the toolbox

Uncorrected Proof

22   S. J. Dugdale *et al.* | Automating physiographic inputs to the CEQUEAU hydrological model   **Journal of Hydroinformatics** | in press | 2017

to implement CEQUEAU at the highest resolution possible and using the time saved by the toolbox to improve model conditioning. It is hoped that the CEQUEAU Physiography Toolbox will both expose CEQUEAU to new users and spur further development of similar routines for GIS-model integration in order to generate high quality hydrological simulations. Such advances are invaluable to both applied and fundamental fluvial research and could help generate mitigation strategies to the effects of anthropogenic pressures on river ecosystems.

ACKNOWLEDGEMENTS

REFERENCES

Ajami, N. K., Gupta, H., Wagener, T. & Sorooshian, S. 2004 Calibration of a semi-distributed hydrologic model for streamflow estimation along a river system. *Journal of Hydrology* **298**, 112–135.

Anderson, M., Chen, Z., Kavvas, M. & Feldman, A. 2002 Coupling HEC-HMS with atmospheric models for prediction of watershed runoff. *Journal of Hydrologic Engineering* **7**, 312–318.

Arnold, J., Moriasi, D., Gassman, P., Abbaspour, K., White, M., Srinivasan, R., Santhi, C., Harmel, R., Van Griensven, A. & Van Liew, M. 2012 SWAT: model use, calibration, and validation. *Transactions of the ASABE* **55**, 1491–1508.

Arsenault, R., Poulin, A., Côté, P. & Brissette, F. 2014 Comparison of stochastic optimization algorithms in hydrological model calibration. *Journal of Hydrologic Engineering* **19**, 1374–1384.

Ayadi, M. & Bargaoui, Z. 1998 Modelling of flow of the Miliane River using the CEQUEAU model. *Hydrological Sciences Journal* **43**, 741–758.

Beaujouan, V., Durand, P., Ruiz, L., Aurousseau, P. & Cotteret, G. 2002 A hydrological model dedicated to topography-based simulation of nitrogen transfer and transformation: rationale and application to the geomorphology–denitrification relationship. *Hydrological Processes* **16**, 493–507.

Beven, K. 1989 Changing ideas in hydrology – the case of physically-based models. *Journal of Hydrology* **105**, 157–172.

Beven, K. 2001 How far can we go in distributed hydrological modelling? *Hydrology and Earth Systems Science* **5**, 1–12.

Bhatt, G., Kumar, M. & Duffy, C. J. 2014 A tightly coupled GIS and distributed hydrologic modeling framework. *Environmental Modelling & Software* **62**, 70–84.

Brisson, C., Boucher, M.-A. & Latraverse, M. 2015 Illustration of the added value of using a multi-site calibration and correction approach to reconstruct natural inflows and inter-catchment transfer flow: a case study. *Canadian Journal of Civil Engineering* **42**, 342–352.

Carpenter, T. M. & Georgakakos, K. P. 2006 Intercomparison of lumped versus distributed hydrologic model ensemble simulations on operational forecast scales. *Journal of Hydrology* **329**, 174–185.

Charbonneau, R., Fortin, J. P. & Morin, G. 1977 The CEQUEAU model: description and examples of its use in problems related to water resource management/Le modèle CEQUEAU: description et exemples d'utilisation dans le cadre de problèmes reliés à l'aménagement. *Hydrological Sciences Bulletin* **22**, 193–202.

Charbonneau, R., Morin, G. & Fortin, J. P. 1979 Effet du pourcentage et de la distribution des surfaces boisées sur les crues de fonte de neige. *Journal of Hydrology* **41**, 93–103.

Clark, M. P., Nijssen, B., Lundquist, J. D., Kavetski, D., Rupp, D. E., Woods, R. A., Freer, J. E., Gutmann, E. D., Wood, A. W., Brekke, L. D., Arnold, J. R., Gochis, D. J. & Rasmussen, R. M. 2015 A unified approach for process-based hydrologic modeling: 1. Modeling concept. Water Resources Research **51**, 2498–2514.

Costa-Cabral, M. C. & Burges, S. J. 1994 Digital elevation model networks (DEMON): a model of flow over hillslopes for computation of contributing and dispersal areas. *Water Resources Research* **30**, 1681–1692.

Couillard, D., Cluis, D. & Morin, G. 1988 An extension of the grid-based hydrological model CEQUEAU to suspended sediment movement through drainage basins. *Water Research* **22**, 991–999.

Coulibaly, P. 2008 Multi-model approach to hydrologic impact of climate change. In: *From Headwaters to the Ocean: Hydrological Change and Water Management* (M. Taniguchi, W. Burnett, Y. Fukushima, M. Haigh & Y. Umezawa, eds). CRC Press, Boca Raton, FL, USA, pp. 249–255.

Cunjak, R. A. & Newbury, R. W. 2005 Atlantic coast rivers of Canada. In: *Rivers of North America* (A. C. Cushing & C. E. Benke, eds). Academic Press, Burlington, MA, USA, pp. 938–980.

DeVantier, B. & Feldman, A. 1993 Review of GIS applications in hydrologic modeling. *Journal of Water Resources Planning and Management* **119**, 246–261.

Devia, G. K., Ganasri, B. P. & Dwarakish, G. S. 2015 A review on hydrological models. *Aquatic Procedia* **4**, 1001–1007.

Dibike, Y. B. & Coulibaly, P. 2005 Hydrologic impact of climate change in the Saguenay watershed: comparison of downscaling methods and hydrologic models. *Journal of Hydrology* **307**, 145–163.

Döll, P., Kaspar, F. & Lehner, B. 2003 A global hydrological model for deriving water availability indicators: model tuning and validation. *Journal of Hydrology* **270**, 105–134.

ESRI 2011 *Arc Hydro Tools Tutorial.* Environmental Systems Research Institute, Redlands, CA, USA, p. 184.

ESRI 2014 *ArcMap 10.3.* Environmental Systems Resource Institute, Redlands, CA, USA.

Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D. & Alsdorf, D. 2007 The shuttle radar topography mission. *Reviews of Geophysics* **45**, RG2004.

Fausch, K. D., Torgersen, C. E., Baxter, C. V. & Li, H. W. 2002 Landscapes to riverscapes: bridging the gap between research and conservation of stream fishes. *BioScience* **52**, 483–498.

Formetta, G., Antonello, A., Franceschi, S., David, O. & Rigon, R. 2014 Hydrological modelling with components: a GIS-based open-source framework. *Environmental Modelling & Software* **55**, 190–200.

GeoBase 2009 *Land Cover, Circa 2000 - Vector,* Government of Canada, Natural Resources Canada, Earth Sciences Sector, Centre for Topographic Information.

Gonzalez, R. C., Woods, R. E. & Eddins, S. L. 2004 *Digital Image Processing Using MATLAB.* Pearson Prentice Hall, Upper Saddle River, NJ, USA, p. 624.

Hansen, N. & Ostermeier, A. 2001 Completely derandomized self-adaptation in evolution strategies. *Evolutionary Computation* **9**, 159–195.

Hart, D. D., Johnson, T. E., Bushaw-Newton, K. L., Horwitz, R. J., Bednarek, A. T., Charles, D. F., Kreeger, D. A. & Velinsky, D. J. 2002 Dam removal: challenges and opportunities for ecological research and river restoration. *BioScience* **52**, 669–682.

Hughes, J. D. & Liu, J. 2008 MIKE SHE: software for integrated surface water/ground water modeling. *Ground Water* **46**, 797–802.

Jajarmizadeh, M., Harun, S. & Salarpour, M. 2012 A review on theoretical consideration and types of models in hydrology. *Journal of Environmental Science and Technology* **5**, 249–261.

Jonsson, B. & Jonsson, N. 2009 A review of the likely effects of climate change on anadromous Atlantic salmon Salmo salar and brown trout Salmo trutta, with particular reference to water temperature and flow. *Journal of Fish Biology* **75**, 2381–2447.

Kauffeldt, A., Wetterhall, F., Pappenberger, F., Salamon, P. & Thielen, J. 2016 Technical review of large-scale hydrological models for implementation in operational flood forecasting schemes on continental level. *Environmental Modelling & Software* **75**, 68–76.

Khakbaz, B., Imam, B., Hsu, K. & Sorooshian, S. 2012 From lumped to distributed via semi-distributed: calibration strategies for semi-distributed hydrologic models. *Journal of Hydrology* **418–419**, 61–77.

Kite, G. 2001 Modelling the Mekong: hydrological simulation for environmental impact studies. *Journal of Hydrology* **253**, 1–13.

Latifovic, R., Homer, C., Ressl, R., Pouliot, D., Hossain, S. N., Colditz, R. R., Olthof, I., Giri, C. P. & Victoria, A. 2012 North American land-change monitoring system. In: *Remote Sensing of Land Use and Land Cover: Principles and Applications* (C. P. Giri, ed.). CRC Press, Boca Raton, FL, USA, pp. 303–324.

Maidment, D. R. 2002 *Arc Hydro: GIS for Water Resources.* ESRI Press, Redlands, CA, USA, p. 203.

Marcé, R., Ruiz, C. E. & Armengol, J. 2008 Using spatially distributed parameters and multi-response objective functions to solve parameterization of complex applications of semi-distributed hydrological models. *Water Resources Research* **44**.

MathWorks 2015 *MATLAB R2015a.* The MathWorks, Natick, MA, USA.

Merkel, W. H., Kaushika, R. M. & Gorman, E. 2008 NRCS Geohydro – a GIS interface for hydrologic modeling. *Computers & Geosciences* **34**, 918–930.

Morin, G. & Couillard, D. 1990 Chapter 5: Predicting river temperatures with a hydrological model. In: *Encyclopedia of Fluid Mechanics: Surface and Groundwater Flow Phenomena* (N. Cheremisinoff, ed.). Gulf Publishing, Houston, TX, USA, pp. 171–209.

Morin, G., Couillard, D., Cluis, D., Jones, H. G. & Gauthier, J.-M. 1986 Modélisation des solides dissous en rivière à l'aide des composantes de l'écoulement. *Canadian Journal of Civil Engineering* **13**, 196–202.

Morin, G. U. Y., Couillard, D., Cluis, D., Jones, H. G. & Gauthier, J.-M. 1987 Prediction of temperatures in rivers using a conceptual model. *Hydrological Sciences Journal* **32**, 31–41.

Morin, G., Nzakimuena, T.-J. & Sochanski, W. 1994 Prévision des températures de l'eau en rivières à l'aide d'un modèle conceptuel: le cas de la rivière Moisie. *Canadian Journal of Civil Engineering* **21**, 63–75.

Morton, D., Rowland, C., Wood, C., Meek, L., Marston, C., Smith, G., Wadsworth, R. & Simpson, I. C. 2011 *Countryside Survey: Final Report for LCM2007 - the new UK Land Cover Map.* Centre for Ecology & Hydrology, Wallingford, UK.

Nash, J. E. & Sutcliffe, J. V. 1970 River flow forecasting through conceptual models part I – a discussion of principles. *Journal of Hydrology* **10**, 282–290.

O'Callaghan, J. F. & Mark, D. M. 1984 The extraction of drainage networks from digital elevation data. *Computer Vision, Graphics, and Image Processing* **28**, 323–344.

Olivera, F., Valenzuela, M., Srinivasan, R., Choi, J., Cho, H., Koka, S. & Agrawal, A. 2006 ArcGIS-SWAT: a geodata model and GIS interface for SWAT. *JAWRA Journal of the American Water Resources Association* **42**, 295–309.

Uncorrected Proof

24   S. J. Dugdale et al. | Automating physiographic inputs to the CEQUEAU hydrological model   Journal of Hydroinformatics | in press | 2017

Ouyang, W., Hao, F., Song, K. & Zhang, X. 2011 Cascade dam-induced hydrological disturbance and environmental impact in the upper stream of the Yellow River. *Water Resources Management* **25**, 913–927.

Poff, N. L., Richter, B. D., Arthington, A. H., Bunn, S. E., Naiman, R. J., Kendy, E., Acreman, M., Apse, C., Bledsoe, B. P., Freeman, M. C., Henriksen, J., Jacobson, R. B., Kennen, J. G., Merritt, D. M., O'Keeffe, J. H., Olden, J. D., Rogers, K., Tharme, R. E. & Warner, A. 2010 The ecological limits of hydrologic alteration (ELOHA): a new framework for developing regional environmental flow standards. *Freshwater Biology* **55**, 147–170.

Reed, S., Koren, V., Smith, M., Zhang, Z., Moreda, F., Seo, D.-J. & DMIP Participants 2004 Overall distributed model intercomparison project results. *Journal of Hydrology* **298**, 27–60.

Refsgaard, J. C. 1997 Parameterisation, calibration and validation of distributed hydrological models. *Journal of Hydrology* **198**, 69–97.

Schumann, A. H. 1993 Development of conceptual semi-distributed hydrological models and estimation of their parameters with the aid of GIS. *Hydrological Sciences Journal* **38**, 519–528.

Schwanghart, W. & Kuhn, N. J. 2010 Topotoolbox: a set of Matlab functions for topographic analysis. *Environmental Modelling & Software* **25**, 770–781.

Seibert, J. & McGlynn, B. L. 2007 A new triangular multiple flow direction algorithm for computing upslope areas from gridded digital elevation models. *Water Resources Research* **43** (4), W04501.

Shahzad, F. & Gloaguen, R. 2011 TecDEM: A MATLAB based toolbox for tectonic geomorphology, Part 1: drainage network preprocessing and stream profile analysis. *Computers & Geosciences* **37**, 250–260.

Singh, V. P. 1995 Computer Models of Watershed Hydrology. Water Resources Publications, Highlands Ranch, CO, USA, p. 1130.

Stantec 2015 *Mactaquac Project: Comparative Environmental Review (CER) Report*. Stantec Consulting Ltd, Fredericton, NB, Canada, p. 504.

St-Hilaire, A., Morin, G., El-Jabi, N. & Caissie, D. 2000 Water temperature modelling in a small forested stream: implication of forest canopy and soil temperature. *Canadian Journal of Civil Engineering* **27**, 1095–1108.

St-Hilaire, A., El-Jabi, N., Caissie, D. & Morin, G. 2003 Sensitivity analysis of a deterministic water temperature model to forest canopy and soil temperature in Catamaran Brook (New Brunswick, Canada). *Hydrological Processes* **17**, 2033–2047.

St-Hilaire, A., Boucher, M.-A., Chebana, F., Ouellet-Proulx, S., Zhou, Q. X., Larabi, S., Dugdale, S. J. & Latraverse, M. 2015 Breathing new life to an older model: the Cequeau tool for flow and water temperature simulations and forecasting. In: *Proceedings of the 22nd Canadian Hydrotechnical Conference*, Montréal, Québec, Canada, April 29–May 2.

Sui, D. Z. & Maggio, R. C. 1999 Integrating GIS with hydrological modeling: practices, problems, and prospects. *Computers, Environment and Urban Systems* **23**, 33–51.

Tarboton, D. G. 1997 A new method for the determination of flow directions and upslope areas in grid digital elevation models. *Water Resources Research* **33**, 309–319.

Tockner, K., Pusch, M., Borchardt, D. & Lorang, M. S. 2010 Multiple stressors in coupled river–floodplain ecosystems. *Freshwater Biology* **55**, 135–151.

Turcotte, R., Lacombe, P., Dimnik, C. & Villeneuve, J.-P. 2004 Prévision hydrologique distribuée pour la gestion des barrages publics du Québec. *Canadian Journal of Civil Engineering* **31**, 308–320.

van Vliet, M. T. H., Franssen, W. H. P., Yearsley, J. R., Ludwig, F., Haddeland, I., Lettenmaier, D. P. & Kabat, P. 2013 Global river discharge and water temperature under climate change. *Global Environmental Change* **23**, 450–464.

Yevjevich, V. 1987 Stochastic models in hydrology. *Stochastic Hydrology and Hydraulics* **1**, 17–36.

Zelazny, V. F. 2008 *Our Landscape Heritage: The Story of Ecological Land Classification in New Brunswick*. New Brunswick Department of Natural Resources, Fredericton, NB, Canada, p. 359.