Simulating discharge time series in regions with contrasting seasons using duration curves

VLADIMIR SMAKHTIN & NISHADI ERIYAGAMA
International Water Management Institute, PO Box 2075, Colombo, Sri Lanka
n.eriyagama@cgiar.org

Abstract Continuous discharge time series in ungauged basins where winter and summer flow generation mechanisms are distinctly different are simulated from limited observed meteorological data (rainfall, snow, temperature). Duration curves are used to convert the precipitation data from source gauges into a continuous hydrograph at an ungauged destination site. Temperature data is used as a control variable which determines whether precipitation is in a liquid (rainfall) or solid (snow) state, and whether the catchment is currently "active" to generate flow. The method is tested in several small catchments in Ontario, Canada, and is designed primarily for application at ungauged sites in data poor regions where the use of more complex and information consuming techniques of data generation may be difficult to justify.

Key words flow time series; flow duration curve; spatial interpolation; observed records; active storage; passive storage; precipitation index; ungauged basins

INTRODUCTION

Daily streamflow time series are required for a variety of hydrological analyses and engineering applications. In data poor regions and/or cases where flow time series *only* are sought, either the use of very simple deterministic models, or application of observed limited data transfer techniques may be justified. One such technique – a non-linear spatial interpolation of observed flow time series – was suggested by Hughes & Smakhtin (1996). Its key characteristic is a flow duration curve (FDC) which gives a summary of flow variability at a site and is interpreted as a relationship between any discharge value and the percentage of time that this discharge is equalled or exceeded. The underlying principle in this technique is that flows occurring simultaneously at sites in reasonably close proximity to each other correspond to similar percentage points on their respective FDCs. The site at which a streamflow time series is generated is referred to as a *destination site*. The site (or sites) with available time series, which is used for generation, is called a *source site.* In essence, the procedure is to transfer the streamflow time series from the location where data are available to the destination site.

The method was originally developed only for patching or extension of observed flow data. Subsequently, Smakhtin *et al.* (1997) illustrated how a FDC may be established at ungauged sites and translated into a complete flow time series. Smakhtin & Masse (2000) developed a modification of the method that allowed rainfall data to be utilized in the frequent cases when no source flow records are available. The modification entailed defining a current precipitation index (CPI), a continuous function of daily rainfall which would abruptly increase on rainy days and exponentially decay during dry periods, thus mimicking the general pattern of streamflow variability. In the modified algorithm, both source flow time series and source FDC were replaced by CPI time series and its duration curve, respectively. The method, in its various forms, was used by Metcalfe *et al.* (2005) to simulate non-regulated flow regimes at numerous hydropower facilities' sites in Ontario, Canada; by Lee *et al.* (2007) to simulate daily flow at ungauged locations in South Korea through the use of a GIS interface; by Archfield *et al.* (2010) to assess water availability at ungauged stream locations in Massachusetts, USA; and in some other applications.

One problem that remains unresolved to date was the use of the CPI based technique in regions, where precipitation falls in the form of both rain and snow, and where snowmelt dominates the spring flow. This paper examines one possible pragmatic solution to fill this gap.

METHOD

The dynamics of the accumulated catchment wetness at any precipitation observation point (*source site*) can be described by the CPI (Smakhtin & Masse, 2000). The CPI for any day is calculated as:

$$\mathrm{CPI}_t = \mathrm{CPI}_{t-1} \times \mathrm{REC} + R_t \quad (1)$$

where CPI_t is the current precipitation index (mm) on day t; R_t is the catchment precipitation for day t and REC is the daily recession coefficient. CPI is also referred to in this paper as "active catchment storage/wetness" – accumulated wetness that determines the current discharge from the catchment. On any day with no rain (R_t = 0) the CPI is equal to the CPI of the previous day multiplied by REC. However, if it rains, the daily rainfall depth too is added to the present day's CPI. CPI is therefore similar to the antecedent precipitation index, reflecting the rate of soil moisture depletion during a period of no rainfall, but it also represents the effects of the current precipitation as well. The range of the REC value is the same as that of the baseflow recession constant, and can be estimated by means of regional regression models with catchment characteristics. Transferring CPI time series at source sites into flow time series at destination sites includes several steps:

(a) *Source site selection for data transfer* In data poor regions, source site selection is limited/ obvious, e.g. meteorological stations within the basin or immediately adjacent to it shall be used. If more than one source site is identified, weights may be assigned to each of them.
(b) *Generation of tables of CPI values* These discrete tables are produced for each source site and each month of the year for fixed percentage points on the CPI duration curve, i.e. if one source site is used, the total number of required tables is 12. The tables use 17 arbitrarily-selected fixed percentage points, which cover the entire range of probabilities. Another alternative is to generate an annual 1-day CPI duration table using the entire CPI record instead of one for each month, especially in cases where only an annual regionalised FDC is available/can be established at the destination site. The differences between the two alternatives do not lead to significant differences in results (Metcalf *et al*., 2005).
(c) *Calculation of FDC tables at ungauged destination site* The set of FDCs for each month of the year (or entire year only) should be established prior to the simulation of the actual time series. A FDC at an ungauged site obviously cannot be calculated from an observed record, as the latter does not exist. It has to be established using various procedures of hydrological regionalisation (e.g. Smakhtin *et al*., 1997; Castellarin *et al*., 2004; Natural Resources Canada: http://www.retscreen.net/ang/home.php). The destination FDCs are approximated by tables of values, similarly to CPI above.
(d) *Data transfer from individual source sites and final estimate* This is the main computational step during which the percentage point of each day's CPI at each source site is identified and the flow value for the equivalent percentage point from the destination site's FDC is read off. The discharge tables are used to "locate" the CPIs and flows on corresponding curves. L-interpolation is used between fixed percentage points. The procedure is repeated for each source site. Finally, a weighted averaging of all flow estimates for the destination site (obtained using individual source CPIs) is performed.

The last step is repeated for each day during the calculation period. The beginning of the period corresponds to the earliest start date of the selected meteorological records, and the end date to the latest of all end dates.

In regions with pronounced differences between seasonal precipitation types, adjustments need to be made to the way that CPI is calculated. To account for these differences, the additional variable – temperature – has to be included. Changes in precipitation type can then be related to temperature; this approach would help to avoid seeking observed snow data, which often simply do not exist.

During the warm part of the year, when daily mean temperatures are continuously positive ($T_t > 0$), the CPI calculation is the same as in equation (1). In the late autumn and winter periods, when the daily temperature becomes and continuously stays negative ($T_t < 0$), precipitation is assumed to fall as snow. This snow does not immediately contribute to "active" catchment storage, which is equivalent to CPI, but accumulates until temperatures become positive again. Therefore, precipitation which falls during negative temperature days is assumed to increment what is referred to here as "passive catchment storage" (PS, mm). PS is incremented by the amount of daily precipitation and no water is released from it at negative temperatures. Active catchment storage, CPI, however, recedes in line with REC:

$$\text{If } T_t < 0: \ \text{CPI}_t = \text{CPI}_{t-1} \times \text{REC} \ \text{ and } \ \text{PS}_t = \text{PS}_{t-1} + R_t \tag{2}$$

When the temperatures become positive again in spring and PS is non-zero, melting occurs:

$$\text{If } T_t > 0 \text{ and } \text{PS}_{t-1} > 0: \ \text{MELT}_t = \text{DD} \times T_t \tag{3}$$

where MELT_t is snowmelt on day t (mm/d) and DD is the degree-day factor (mm/°C per day). The "daily" check is made to ensure that there is still "water" in passive storage. If precipitation occurs during the melting stage, it is considered to be in the form of rain. The required CPI values during the melting period are then calculated as:

$$\text{CPI}_t = \text{CPI}_{t-1} \times \text{REC} + R_t + \text{MELT}_t \tag{4}$$

The DD parameter is the amount of melting which occurs per one degree of positive air temperature within 24 hours – per one degree-day. The DD varies depending on the month, terrain and weather conditions and snow density. In the current study, the DD parameter value has been fixed at 3.5 mm/°C per day as per Kuusisto (1984). Region-specific DD values may be obtained from the literature. Once the continuous CPI time series and their corresponding duration curves are calculated for each selected source meteorological station, they may be used as described above in steps (c) and (d) of the algorithm to calculate the destination flow time series.

APPLICATION TO ONTARIO CATCHMENTS

The approach described above is generic. Therefore, the choice of a "testing ground" is determined by (naturally) the climate of the area and ease of access to input data on which to test the method. The Canadian province of Ontario satisfies both criteria. The catchments selected have been drawn mainly from the southern part of Ontario (Fig. 1). The exact geographical location of these catchments is, however, largely irrelevant, but the amount and quality of available observed source data are of primary importance. To test the algorithm under the "harshest" possible conditions, only one source meteorological station, nearest to each selected test catchment outlet, was used in the present study.

All input data are referred to in the coding system followed by the Canadian National Climate Data Archive (CDCD: http://climate.weatheroffice.gc.ca/prods_servs/index_e.html) and the Canadian National Water Data Archive (HYDAT: http://www.ec.gc.ca/rhc-wsc/default.asp?lang=En&n=9018B5EC-1, Table 1). The data selection procedure took into account such factors as the length of record, amount of missing data, existence of concurrent precipitation and temperature data, closeness of meteorological station to the catchment, catchment size, and flow regulation (unregulated flow records must be selected).

All catchments are gauged; yet two different scenarios of observed data availability were considered. The first scenario represents the case where limited historical flow time series may be available to construct a FDC for a destination site. The second scenario is that the catchment is completely ungauged. For the first scenario, 12 FDC tables were generated (one for each calendar month) and it was assumed that they did not change with time and remain a representative flow "signature" of each selected catchment. In the second scenario use was made of Regional Normalized FDCs and specific runoff maps (runoff/km^2) developed by Natural Resources Canada

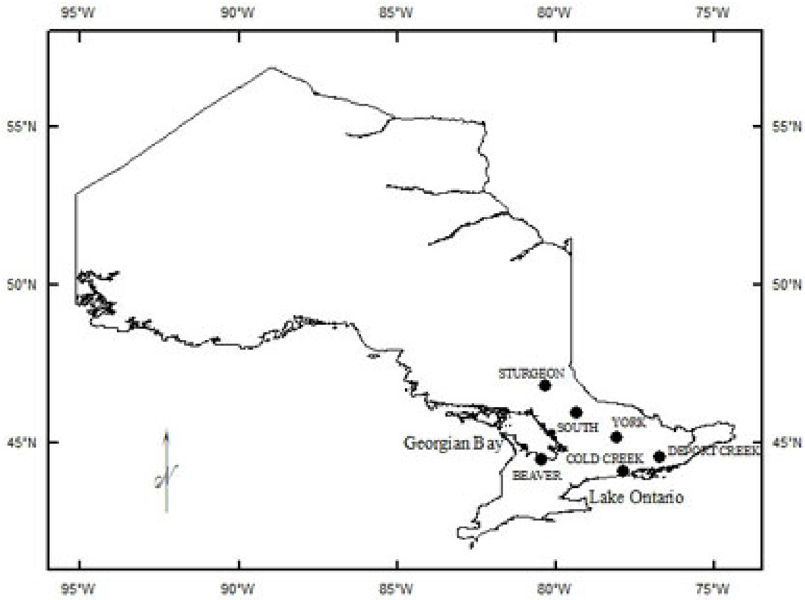

Fig. 1 Locations of the study catchments (black circles) on the map of Ontario Province.

for the entire country as part of the RETScreen software (a software for evaluating renewable energy projects: http://www.retscreen.net/ang/hydrology_data.php). Within each FDC region and specific runoff region, normalized FDCs have been provided for several HYDAT gauge locations (but not for the ones simulated in this study). Three types of normalized FDCs were derived by averaging the nearest available FDCs (from the same FDC region and specific runoff region) to each simulated gauge location. Annual FDCs for each site were then estimated by multiplying each averaged non-dimensional FDC by catchment area and specific runoff. The corresponding annual CPI duration curve for source sites were used in the simulation in this case.

The purpose of the first scenario is to examine how the method performs in conditions when an accurate FDC can be established for the ungauged location. In this case, most, if not all, estimation uncertainty will be related to the amount, quality and temporal pattern of the source meteorological data. The second scenario then represents the case where uncertainty is associated with both the estimation of regional duration curves and source meteorological data.

DISCUSSION AND CONCLUSIONS

Table 1 presents the results of simulations in all catchments considered so far, and Fig. 2 gives a snapshot of some individual, arbitrarily selected simulations. No attempt has been made to "calibrate" the proposed "model". On the other hand, calibration options of this approach are limited to changing the recession parameter value, the degree-day factor and the meteorological stations' number and weights. Therefore the results illustrate the performance of the approach under very stringent conditions. The recession REC value for each catchment was assumed to be equal to the median recession ratio value of the flow time-series (FREND, 1989), but generally, recession characteristics of streams may be estimated from regional relationships with catchment parameters (Tallaksen, 1995).

Fit statistics in Table 1 suggest that flow hydrographs obtained under Scenario 1 for all catchments are of better quality than those under scenario 2. This indicates that the results of the simulation depend on the accuracy of the established FDC for the destination site. In conditions where there is a reliable FDC, the algorithm yields good results.

The use of just one meteorological station (as done in this study intentionally to represent the most limited data condition) is not normally recommended (although it might be unavoidable in

Table 1 Statistics of fit between observed daily flows, and flows simulated by the proposed model for the simulation period.

Catchment and simulation period	Gauged area (km^2)	Data	Mean (m^3/s)	SD (m^3/s)	Max (m^3/s)	Min (m^3/s)	R^2	CE
South: 02DD005	787	Obs	14.8	13.3	121	0.52		
1981–2006		Scenario1	11.8	12.4	89.7	0.55	0.72	0.67
		Scenario2	14.2	21.4	179	0.03	0.53	–0.88
Sturgeon: 02DC004	2980	Obs	37.4	34.6	271	6.12		
1990–2002		Scenario1	38.1	35.2	345	6.07	0.70	0.67
		Scenario2	40.6	59.5	410	3.01	0.46	–0.63
Beaver: 02FB009	572	Obs	8.85	7.70	62.0	2.13		
1967–1969		Scenario1	7.92	6.98	57.9	1.19	0.60	0.56
		Scenario2	8.62	20.1	166	2.13	0.16	–4.79
York: 02KD002	839	Obs	11.9	13.9	102	0.58		
1958–1983		Scenario1	11.8	12.7	104	0.05	0.60	0.58
		Scenario2	11.4	16.7	116	0.83	0.37	0.01
Cold Creek:	159	Obs	1.95	1.91	24.7	0.45		
02HK007		Scenario1	2.02	1.97	28.6	0.47	0.44	0.30
1991–2006		Scenario2	2.14	3.10	21.8	0.16	0.25	–1.02
Deport Creek:	189	Obs	2.11	1.63	11.2	0.02		
02HM002		Scenario1	1.99	1.92	20.6	0.00	0.48	0.25
1986–2006		Scenario2	2.54	3.64	26.0	0.19	0.18	–3.15

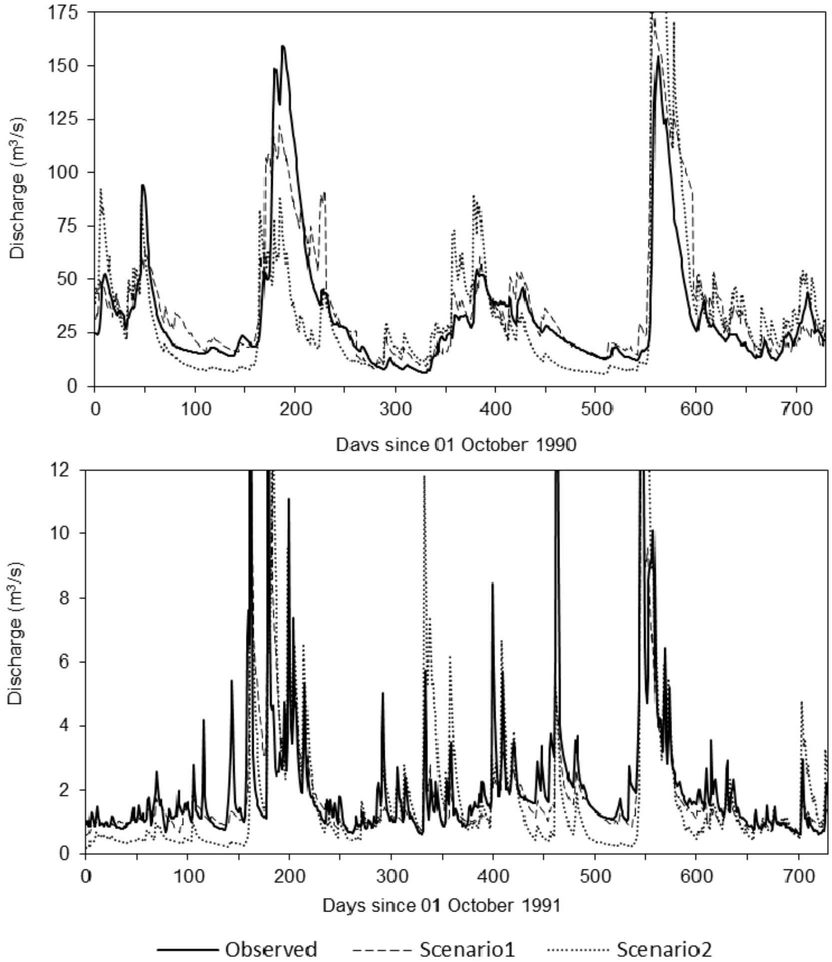

Fig. 2 Observed and simulated hydrographs for Sturgeon River (top) and Cold Creek (bottom).

some cases), because all the deficiencies of its data are effectively transferred to the simulated destination flow time-series. At the same time, one gauge may not always be representative of the spatial variability of catchment precipitation and temperature, even in small catchments, especially when it is located outside the catchment boundary. This is one of the reasons for the poor R^2 and CE values (Table 1) obtained for Cold Creek and Deport Creek, where the source meteorological station is a considerable distance outside the catchment (12–25 km) when compared with catchment dimensions (159 and 189 km^2, respectively).

Overall, comparison of observed and generated hydrographs, as well as fit statistics, shows that the approach is capable of reproducing the general pattern of daily streamflow variability, although low flows tend to be underestimated in Scenario 2 (Fig. 2). However, this can be explained by the low accuracy of the destination sites' FDCs estimation – from a rather coarse nation-wide regional study.

The paper has illustrated the application of the method only for the generation of *daily* streamflow hydrographs. At the same time a similar approach may be applied to generate monthly flow time-series from monthly precipitation data. In this case, the actual monthly step precipitation data may be used instead of the CPI time-series, and the recession coefficient will need a new interpretation.

The performance of the method has been illustrated using examples of predominantly small catchments. However, it may be anticipated that the method, in principle, is applicable to large river catchments as well. In this case, the number of initially available meteorological stations with suitable data will be large and it may be appropriate to generate the weighted average catchment wetness first. Depending on the type of data used, it could be daily or monthly weighted average wetness. Also it should be possible to split the large catchment into a set of "homogeneous" subareas and calculate weighted average time-series for each of them. Subareas then will replace the source meteorological sites. On the other hand, the large catchments are more likely to have suitable streamflow source gauges that can be used for data transfer between the sites and the original version of the spatial interpolation algorithm (which uses streamflow data *only*) may apply. The significant amount of remotely sensed data presently available, for a variety of hydrological analyses, may also be used for such simulations in the future.

REFERENCES

Archfield, S. A., Vogel, R. M., Steeves, P. A., Brandt, S. L., Weiskel, P. K. & Garabedian, S. P. (2010) The Massachusetts Sustainable-yield Estimator: a decision support tool to assess water availability at ungaged stream locations in Massachusetts. *Scientific Investigations Report 2009-5227*. US Department of the Interior, US Geological Survey.

FREND (1989) *Flow Regimes from Experimental and Network Data. I: Hydrological Studies; II: Hydrological Data*. Institute of Hydrology, Wallingford, UK. 345 pp

Hughes, D. A. & Smakhtin, V. Y. (1996) Daily flow time series patching or extension: a spatial interpolation approach based on flow duration curves. *Hydrol. Sci. J.* **41**, 851–871.

Kuusisto, E. (1984) Snow Accumulation and snowmelt in Finland. *Water Research Institute Publ. no. 55*. National Board of Waters, Finland, Helsinki, 120 pp.

Lee, J. U., Yang, J. & Choi, J. (2007) Development of Integrated GIS Interface for characteristics of regional daily flow. *World Academy of Science, Engineering and Technology* **28**, 180–185.

Metcalfe, R. A., Chang C. & Smakhtin V. (2005) Tools to support the implementation of environmentally sustainable flow regimes at Ontario's waterpower facilities. *Canadian Water Resources J.* **30**(2), 97–110.

Natural Resources Canada. Retscreen Clean Energy Project Analysis Software Website (http://www.retscreen.net/ang/home.php). Accessed on 17 December 2010.

Smakhtin, V. U. & Masse, B. (2000) Continuous daily hydrograph simulation using duration curves of a precipitation index. *Hydrol. Processes*. **14**, 1083–1100.

Tallaksen, L. M. (1995) A review of baseflow recession analysis. *J. Hydrol*. **165**, 349–370.