Journal of Hydrology 517 (2014) 1166–1175

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

'As simple as possible but not simpler': What is useful in a temperature-based snow-accounting routine? Part 1 – Comparison of six snow accounting routines on 380 catchments

Audrey Valéry [a], Vazken Andréassian [b,*], Charles Perrin [b]

[a] EDF-DTG, Operational Forecasting Unit, Grenoble, France
[b] IRSTEA, Hydrosystems and Bioprocesses Research Unit (HBAN), Antony, France

ARTICLE INFO

Article history:
Received 16 September 2013
Received in revised form 8 April 2014
Accepted 23 April 2014
Available online 10 May 2014
This manuscript was handled by Andras Bardossy, Editor-in-Chief, with the assistance of Axel Bronstert

Keywords:
Snow accounting routine
Snowmelt
Snow accumulation
Precipitation-runoff models
Degree-day approach
Snowpack variability

SUMMARY

This paper analyzes the behavior of hydrological snow accounting routines (SARs) used in combination with hydrological models to simulate streamflow at the catchment scale. To reach conclusions as general as possible, we compare the performance of six existing SARs combined with two different precipitation-runoff models. The SARs are temperature-based, have different levels of complexity (understood here as the number of optimized parameters and model functions), include various processes and also differ by the way they account for the spatial heterogeneity of snow cover. The SARs were tested on a set of 380 catchments significantly affected by snow and located in four countries (France, Switzerland, Sweden and Canada), showing different climatic conditions and altitude ranges. The value of each SAR is evaluated solely in terms of flow simulation quality at the catchment outlet. Several efficiency criteria are used, some of them specifically focusing on the time periods affected by snow accumulation and melt.

As expected, the use of a snow accounting routine on snow-affected catchments significantly improves model efficiency, and this is true even for the simplest SARs. More interestingly, our results show that the most complex SAR does not yield the highest performance. Surprisingly, a lumped routine (i.e. without distribution in elevation bands) appears to be the most efficient on average on the whole catchment set. Results seem particularly sensitive to the spatial variability of processes in the snowpack and to the determination of the precipitation phase (solid or liquid). One critical point remains the identification of the solid precipitation correction factor necessary to compensate for snowfall measurement errors. In the companion article, we further investigate the sensitivity of model results to the description of snow processes in the SAR and try to identify the most important components of a parsimonious and general SAR.

1. Introduction

1.1. Temperature index vs energy budget methods

Existing methods to account for snow accumulation and melt can either be classified as *temperature index* or *energy budget* methods: the former use temperature as an integrating indirect measurement of energy fluxes, while the latter aim at detailing the different components of a complete energy budget. Several papers have discussed the pro and cons of each of these two classes (Braun and Lang, 1986; Charbonneau et al., 1981; Franz, 2006). The question whether air temperature is a sufficient indicator to model all snow-processes has been raised repeatedly in the past (Ohmura, 2001). A hybrid approach called *restricted degree-day method* has even been proposed (Brubaker et al., 1996; Cazorzi and Dalla Fontana, 1996; Kustas et al., 1994): it combines a radiation term with the degree-day relation (Hock, 2003). The authors argued that the energy budget method is efficient at the point scale, for well-instrumented plots, and for short time steps of computation. However, they acknowledged that the degree-day method is a good and efficient approach for forecasting purposes at catchment scale and larger time steps.

In this paper, our purpose is not to contest the theoretical superiority of the energy budget methods. But it is a fact that temperature-based methods are still widely used in operational hydrology, mostly for input data availability reasons: in many mountainous catchments of the world, meteorological information is extremely scarce, the spatial variability of temperature and

* Corresponding author. Tel.: +33 1 40 96 62 58.
E-mail address: vazken.andreassian@irstea.fr (V. Andréassian).

http://dx.doi.org/10.1016/j.jhydrol.2014.04.059
0022-1694/

precipitation is large. Some authors have already tried to discriminate between different types of SARs independently of this question of input data availability: for instance, Walter et al. (2005) compared a simplified process-oriented model (that extrapolates unknown energy components) with temperature-index approaches, all SARs being fed with the same inputs.

Thus, regarding the important literature on this subject, one should not try to give absolute judgements regarding alternative solutions but rather try to better understand the advantages, sensitivity and limits of each of them.

1.2. Comparative studies are still needed

If we are to aim for a simplified representation of the hydrological model, we should remember Alfred Einstein's words: a model should be kept 'as simple as possible but not simpler'. Thus, we should compare the different alternatives before deciding which one to use. As it unfortunately often happens in hydrology, most of the comparisons presented in the literature have been based on a single catchment, or in the best case on a small number of them, under rather similar climatic conditions. Two noteworthy exceptions exist:

- The intercomparison of snowmelt runoff models organized by the World Meteorological Organization (WMO, 1986), which compared eleven snow accounting routines (SARs), each used with a specific precipitation-runoff model, on six catchments located in different parts of the world. The objective of this study, which only looked at temperature-index approaches, was to provide guidance to modellers wishing to use SARs.
- The Snow Model Intercomparison Project (SMIP) launched in 2001 (Essery and Yang, 2001), included more than 20 snow models, all of them adopting energy and mass budgets approaches. Some of these snow models were meant to address a wide range of applications (i.e. not only hydrological forecasting, but also avalanche forecasting, climate modelling or fundamental studies of snow physics). Most of the models had large input data requirements; consequently they were only tested on a few sites (Etchevers et al., 2004).

1.3. How to account for snow-melt and snow-accumulation processes in a hydrological model?

The operational application of any hydrological model on a snow-affected catchment requires the operational hydrologist to consider several practical questions, relative to the four components essential to any catchment-scale snow accounting routine (see e.g. Ferguson, 1999):

a. *Meteorological extrapolation*: operational observation networks provide point measurements, which must be extrapolated to the entire catchment area, taking into account the particular variability and heterogeneity of mountain environments (see Charbonneau et al., 1981; WMO, 1986; Klemeš, 1990; Braun et al., 1994).
b. *Computation of snow-melt at a local scale*: based on available measurements, several computational approaches can be used, usually classified as temperature index or energy budget methods.
c. *Integration of snow-melt on the entire snow-covered area*: this point raises the question of distribution in elevation zones (Blöschl et al., 1991; WMO, 1986), as well as the question of snow-covered area's estimation which can either be an input data (Anderson, 1973; Martinec and Rango, 1981) or an internal state of the snow accounting routine.
d. *Snowmelt routing*: since the objective of hydrological computations is to simulate a hydrograph at the outlet of a catchment, the issue of routing the simulated melt must also be addressed.

In this paper, we will discuss in more detail points b and c, which have attracted the most attention in the hydrological literature, although from a practical point of view, the other points are probably equally important.

1.4. Scope of the paper

In this paper, our objective is to compare existing catchment-scale snow accounting routines (SARs), in order to evaluate their respective merits, and to identify the processes which seem necessary (unavoidable) in a SAR. Our aim being to draw as general conclusions as possible, we based our analysis on a large set of catchments specifically assembled for this purpose. Because only a few observed variables were available for all catchments, it was not possible to deal here with the purely physically-based energy balance methods, and we had to limit our analysis to temperature-index methods.

In Section 2, we detail the data and models used for this comparative assessment: the set of 380 catchments, the six SARs and the two hydrological models. Then, Section 3 describes our model assessment methodology and discusses the criteria that can be specifically used to assess the impact of SARs on model performance. Last, we present results and draw some conclusions which will be used as a starting point for the more detailed investigations presented in the companion paper.

Note that we will not present here the long preliminary work necessary to regionalise precipitation and temperature data at the catchment scale, especially to account for altitudinal gradients. The reader can refer to Valéry (2010), Valéry et al. (2010) for full information about it. However, what must be kept in mind here is that all SARs were fed with exactly the same regionalized inputs.

2. Dataset and models

2.1. Dataset

We gathered for this study an international dataset of 380 catchments from four countries (France, Switzerland, Sweden and Canada). Although we do not claim that this dataset covers the whole range of snow cover types, we consider that the variety of snow covers represented here allows drawing general conclusions. Fig. 1 presents the location of the catchments used in this study, while Table 1 summarizes the dataset's main characteristics.

Snow amounts differ in the four countries:

In France, 159 catchments are moderately affected by snow (i.e. less than 10% of their total precipitation falls in the solid form). These catchments are located on moderate reliefs: the median range of altitude (ΔZ) is only 512 m (red[1] circles in Fig. 1a). Sixty more French catchments (blue triangles in Fig. 1a) are located in zones with much more pronounced reliefs: the French Alps (south-eastern France), the Pyrenees (south-western France) and the Jura (East, at the border with Switzerland).

In Switzerland, the 30 available catchments included in this dataset display strong mountainous influences, being mostly located at high elevations (the lowest catchment has its median altitude at 600 m a.s.l.). Their characteristics are quite similar to

[1] For interpretation of color in Fig. 1, the reader is referred to the web version of this article.

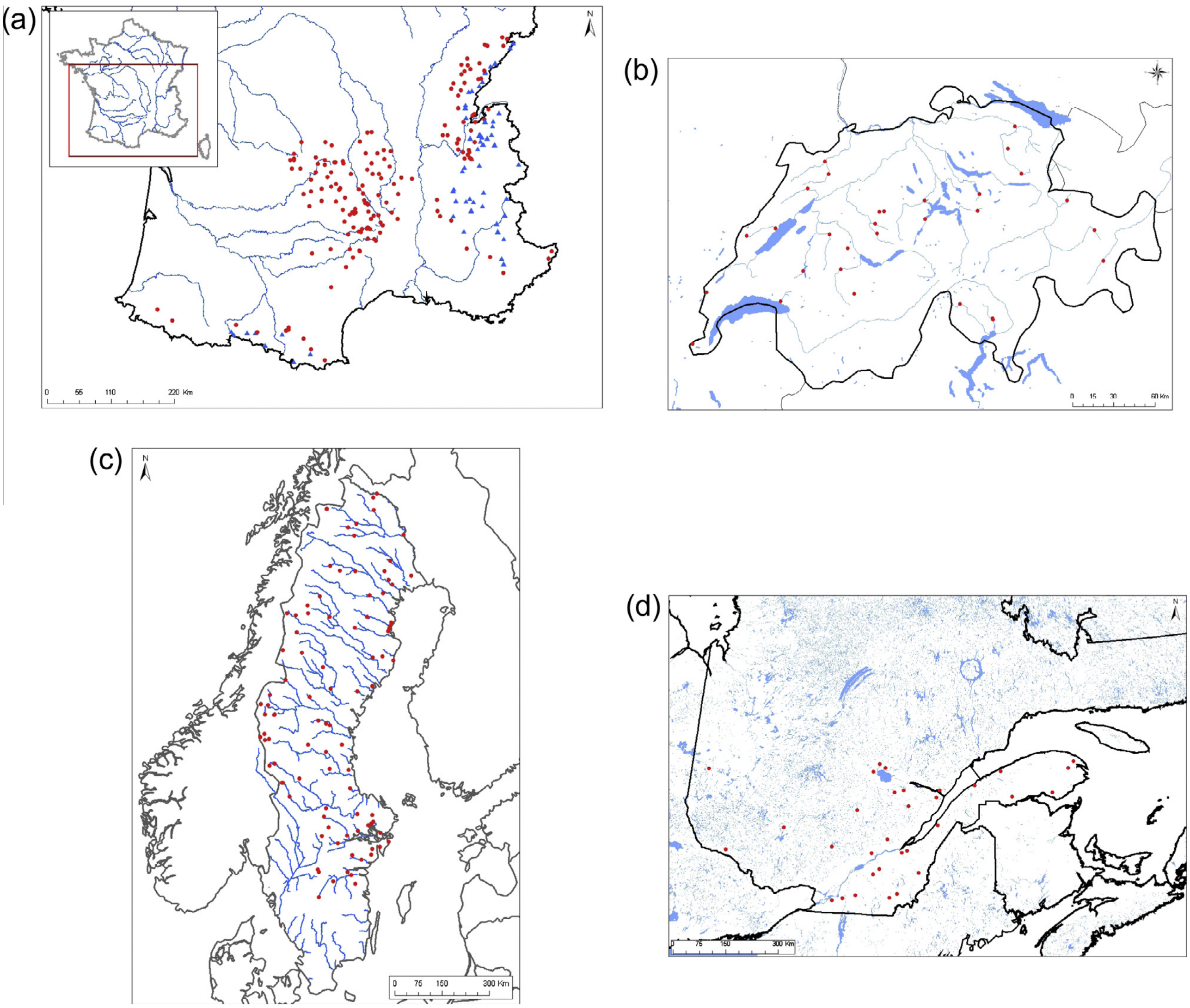

Fig. 1. Location of gauging stations in (a) France, (b) Switzerland, (c) Sweden and (d) Canada (Québec).

Table 1
Main characteristics of the 380 catchments form the four countries. Minimum and maximum values are given in each case.

Countries	France	Switzerland	Sweden	Canada
Number of catchments	219	30	94	36
Period of data availability	1995–2005	1995–2005	1995–2006	2002–2007
Catchment area (km^2)	5–3580	0.5–1085	1–14480	254–15,300
Mean annual precipitation P (mm/yr)	760–2290	1170–2580	560–1110	860–1190
Mean annual streamflow Q (mm/yr)	230–2070	540–2250	160–1330	430–1270
Mean annual potential evaporation PE (mm/yr)	280–720	240–630	220–630	420–620
Mean annual snowfall G (mm/yr)	10–740	50–730	50–510	200–790
Percentage of snowfall (%)	1–51	4–43	8–46	20–35
Catchment median altitude Z_{median} (m)	368–2688	600–2580	17–967	50–834
Catchment range of altitude ΔZ (m)	28–2575	110–2072	2–1053	43–792

the 60 high-elevation French catchments, so the results of these 90 (30 + 60) catchments will be analyzed together.

In Sweden, 94 catchments present strong Nordic features. They are located above latitude of 61.9° North, where the winter season is well-defined, the catchments being snow-covered most of the time during this season. Snow influence increases with latitude and with elevation (the highest catchments are at the border with Norway).

In Canada, 36 catchments were available in the Province of Québec, which is affected by the Labrador oceanic current and by a marked continental climate because of its location in the Eastern part of the continent. Important contrasts in temperatures characterize the selected Canadian catchments with generally very cold winter and hot summer. Although there is less relief in Québec than in the other three countries, catchments are strongly snow-affected: between 20% and 35% of precipitation falls in solid form

(Table 1). As a consequence of the absence of important relief in Québec, these catchments are probably the most homogeneous in terms of climatic forcing.

During the catchment selection process, we tried to minimize possible interactions and combination with non-snow related processes that could also influence streamflow. Therefore, we avoided glacierized basins (especially in Switzerland), basins with large lakes (especially in Sweden), basins with known intercatchment groundwater flows, and catchments with documented flow diversions.

For most catchments, daily data were available over the 1995–2005 period, except for Canada where only shorter six-year series (2002–2007) were available. For each catchment, series of precipitation (P), temperature (T) and flow (Q) data were collected. Data are of generally good quality. Visual inspection of data was made to check for major mistakes. Catchment areal precipitation and temperature were calculated after regionalizing altitudinal gradients, as detailed by Valéry (2010) and Valéry et al. (2010). Potential evapotranspiration (PE) data were estimated using the temperature-based formulation proposed by Oudin et al. (2005).

2.2. Snow accumulation and melt routines

Six different SARs were tested and compared in this study, each of them having specific features and particularities (see Table 2). The reader can refer to the references provided in Table 2 for full details on these SARs. Further information is also provided by Valéry (2010).

The six SARs selected for this comparison represent a wide sample of existing temperature-based SARs. Some, like MOHYSE, are very simple whereas others, like M_SNE or CEQUEau, are more complex taking into account in their structures a large number of snow-processes and different degrees of freedom. These six SARs have been developed for different environments: CEQUeau is widely used in Québec (Canada) and sometimes in France, NAM and HBV are applied in European high-latitude countries and MORD4 and M_SNE are especially developed for French mountainous catchments.

Different spatial distributions are considered such as lumped structures (MOHYSE and MORD4) whereas the majority of the SARs are distributed by altitudinal bands. As different distributions can be considered in a SAR, it is important to remind that in this study every catchment was distributed by altitudinal bands of equal area. Thus, if five bands of altitude are considered, each represents 20% of the total catchment area. In the same time, they have different ranges of altitude.

Comparing to the original versions of these six SARs, the number of free parameters indicated in Table 2 is not necessarily the same: some of the less sensitive parameters were fixed after a sensitivity analysis. For instance, the original version of MORD4 considers 10 free parameters: 6 have been fixed with a minimal loss of performances over the 380 catchments of the sample set (see Valéry, 2010).

2.3. Precipitation-runoff models used

In this paper, the evaluation of snow accounting routines will exclusively be based on their ability to contribute to streamflow simulation at the catchment outlet. Indeed, no snow cover measurements were available to evaluate the relevance of the snowpack simulation. Therefore, the six SARs presented above must necessarily be combined with a hydrological model to simulate streamflow. To avoid model-specific results, we chose to use two different precipitation-runoff models (each hydrological model being successively associated with one of the six SARs). As hydrological models, we used (see structures in Fig. 2):

- The four-parameter GR4J presented by Perrin et al. (2003).
- A nine-parameter lumped version of the HBV model (Bergström, 1995), which is called here HBV9 to avoid confusion with the original version.

The two models were run at the daily time step and used in lumped mode, even if some of the snow accounting routines were distributed in altitudinal bands. The structure and the number of

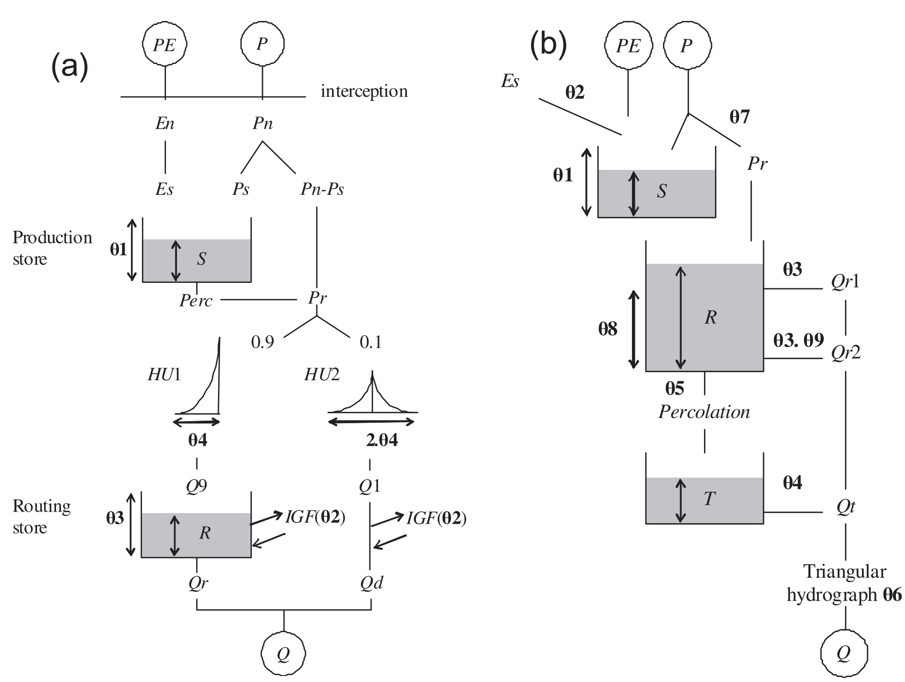

Fig. 2. Diagrams of the two precipitation-runoff models used in this study: (a) GR4J and (b) HBV9 (θ_i are calibrated parameters).

Table 2
Characteristics of the versions of the six snow accounting routines tested here.

SARs	Free parameters	Internal state variables	Spatial discretisation	Determination of the form of precipitation (liquid/solid)	Melt factor	Liquid water retention and refreezing	Rainfall through snowpack	Ground-melt	Specific additional features
MOHYSE (Fortin and Turcotte, 2007)	1 – Snowmelt factor	1 – Snowpack	Lumped	Temperature threshold	Calibrated	No	No	No	Intentionally simple for teaching purposes
CEQUeau (Morin, 1997)	1 – Snowmelt factor	1 – Snowpack	Distributed by altitudinal bands	Range of temperature	Calibrated + sunshine	No	Yes	No	Used in Québec +
	2 – Snowpack ripening temperature	2 – Snowpack cold content							Subdivision according to the land cover (forest, open areas) in the original version (not accounted for here)
	3 – Cold content factor	3 – Snowpack ripening index (min value equal to zero)							
HBV (Bergström, 1975)	1 – Snowmelt factor	1 – Snowpack	Distributed by altitudinal bands	Range of temperature	Calibrated	Yes	No	No	Used in Sweden +
	2 – Snow correction factor	2 – Liquid water content in the snowpack							Correction of snow undercatch with a free parameter
	3 – Cold content factor								
NAM (DHI, 2009)	1 – Snowmelt factor	1 – Snowpack	Distributed by altitudinal bands	Temperature threshold	Calibrated	Yes	Yes	No	Used in Norway +
	2 – Liquid water retention capacity of the snowpack	2 – Liquid water quantity in the snowpack							Estimation of the percentage of zone covered by snow; no PE acting on snow-covered areas
	3 – Snow-covered area threshold	3 – Snow-covered area percentage							
MORD4 (Garçon, 1999)	1 – Snowmelt factor	1 – Snowpack	Lumped	Range of temperature	Calibrated	No	No	Yes	Used in French Alps +
	2 – Temperature correction before the determination of the precipitation form	2 – Snowpack cold content							Estimation of the percentage of zone covered by snow; use of hypsometric curve
	3 – Temperature correction before the snowmelt	3 – Snowmelt cold content							
	4 – Previous temperature correction for cold content computation	4 – Snow-covered area percentage							
M_SNE (Paquet, 2004)	1 – Snowmelt factor	1 – Snowpack	Distributed by altitudinal bands	Range of temperature	Calibrated	Yes	No	Yes	Used in French Alps +
	2 – Snowmelt increase index	2 – Snowpack cold content							Snowmelt increases when the snowpack temperature reaches 0 °C
	3 – Cold content factor								
	4 – Refreezing content factor								
	5 – Refreezing factor								
	6 – Factor for snow precipitation snow on cold content								
	7 – Groundmelt factor								

degrees of freedom differ between GR4J and HBV9, and this should confer to our results more generality.

In this study, SARs were used on top of precipitation-runoff models. *P*, *PE* and *T* data are used as inputs to the SARs which provide intermediate outputs (rainfall, snowmelt, and remaining *PE*). Rainfall and snowmelt are summed to represent the new liquid precipitation quantity which enters the lumped hydrological model at every time step.

3. Model assessment methodology

3.1. General assessment methodology

We used the split sample test procedure, a classical calibration/validation approach (Klemeš, 1986). The whole available period of record was split into two sub-periods alternatively used for calibration and validation. It means that for each catchment, two calibration and two validation tests were performed, thus providing results in validation mode on all available data. However, the first year of each period was used for model warm-up, i.e. the results on this year were not considered in performance computation.

The parameters of the SAR and the precipitation-runoff model were optimized simultaneously, using the same local search algorithm as in Edijatno et al. (1999). As the number of parameters could be quite large in some cases, we made a pre-screening of the parameter space to identify the most likely zone of convergence, thus providing the initial value for the local search. This approach allows avoiding most of the secondary optima pitfalls that the local search alone would be subject to. This was found by Mathevet (2005) to provide results comparable to more sophisticated global search algorithms for this type of models.

The objective function used was the Nash and Sutcliffe (1970) efficiency. As we were interested in both high and low flows, the Nash–Sutcliffe criterion was computed on root-square transformed streamflow which represents a good trade-off to provide acceptable results in both cases (Oudin et al., 2006):

$$NS_{rQ} = 1 - \frac{\sum_{j=1}^{N}\left(\sqrt{Q_{obs}(j)} - \sqrt{Q_{sim}(j)}\right)^2}{\sum_{j=1}^{N}\left(\sqrt{Q_{obs}(j)} - \overline{\sqrt{Q_{obs}}}\right)^2} \quad (1)$$

where Q_{obs} is the observed daily runoff for the day *j* (m^3/s), Q_{sim} is the simulated daily runoff for the day *j* (m^3/s), and $\overline{\sqrt{Q_{obs}}}$ is the mean of transformed (root square) interannual daily value of runoff for the julian day *j*, computed on the whole available period of observation.

All SARs were tested on the whole dataset and with the two hydrological models, each using exactly the same input data, the same subdivision in elevation zones (when required) and the same objective function. In the following, only the results obtained in validation mode will be discussed, as they are the most representative of the actual model efficiency. Note that the significance of the differences in performance was evaluated based on our past experience in model testing, but statistical tests could also have been applied (see e.g. Pushpalatha et al., 2011).

3.2. Specific criteria to judge of the efficiency of the SARs

To evaluate model efficiency, we will use the classical Nash–Sutcliffe criterion computed on all the time steps of the test period considering the root-square transformed streamflow. It will be noted NS_{year} hereafter. However, the influence of snow on streamflow is significant only over a specific period of time in the year. Therefore the evaluation of SARs may not be well reflected by using this general criterion. Thus we looked for criteria better adapted to evaluate streamflow simulation in snow-affected periods. We propose to distinguish two additional sub-periods of computation:

- The first snow-specific criterion is computed on the half-year the most influenced by snow. The aim is to assess performance during both snow-accumulation and snowmelt periods. As all our catchments are located in the Northern hemisphere, winter and spring seasons are quite similar between countries. Therefore the criterion, noted NS_{snow}, was computed only over the six-month period from December 1st to May 31st. We checked on our data set that only very few days influenced by snow lied outside this period.
- The second snow-specific criterion, noted NS_{melt}, focuses on the snowmelt period, which is often considered to be the most critical period in terms of modelling as well as in an operational perspective. Indeed, it is usually shorter than the snow-accumulation period since it seldom lasts more than one month for the majority of catchments, and can even represent just a few days. Snowmelt period also greatly varies from one year to another for a given basin, according to the air temperature variability. A sub-period of two months was defined for every catchment depending on its climatic peculiarity. For French catchments moderately affected by snow, the criterion NS_{melt} is computed from February 1st, to March 31st; for the other catchments, it is computed from April 1st to May 31st. We checked on our data set that these periods included the majority of snowmelt events.

Consequently, we worked with three criteria in validation: the classical Nash–Sutcliffe criterion calculated over the whole year and the two snow-specific ones described above. These two last criteria should help discriminating between SARs.

Note that we could have worked with NS_{snow} and NS_{melt} criteria calculated only on periods where snow actually has an influence on streamflow. As we did not have snowpack measurements, it means that we should have relied on snowpack estimates simulated by the SAR. However we preferred not to do so, because the evaluation period would depend on the tested SAR, which would end-up with different evaluation periods between SARs and thus not comparable results.

As we work on a large dataset, we used a bounded version of the previously defined *NS*-type criteria, as proposed by Mathevet et al. (2006). It is defined by:

$$C = \frac{NS}{2 - NS} \quad (2)$$

The transformed values vary between −1 and 1 instead of $-\infty$ and 1 for the original criterion. This allows analysing mean values over the catchment set without being overly sensitive to a few highly negative performance values for catchments where the model fails. Note that the bounded version provides lower positive values than the original ones (e.g. a *NS* value of 0.8 corresponds to a bounded value of 0.67).

4. Results

4.1. Overall comparison

Table 3 presents the mean efficiencies over the 380 catchments obtained in validation by the six SARs combined with the two different hydrological models. As reference, we have also included the performance of GR4J and HBV9 without any snow accounting routine. Several general comments can be made:

- Introducing a snow accounting routine yields a large improvement in modelling efficiency: there is a jump in C_{year} criteria

Table 3
Mean performance over the 380 catchments obtained in validation mode by the two precipitation–runoff models (GR4J and HBV9) without any snow accounting routine and with one of the six SARs assessed in this paper (bold values indicate best SAR performance).

Hydrological models	Assessment criteria	SAR option (number of optimized parameters)						
		No snow routine (–)	MOHYSE (1)	CEQUeau (3)	HBV (3)	NAM (3)	MORD4 (4)	M_SNE (7)
GR4J	C_{year}	0.415	0.640	0.657	0.671	0.668	**0.692**	0.681
	C_{snow}	0.285	0.580	0.606	0.615	0.633	**0.652**	0.634
	C_{melt}	0.157	0.481	0.504	0.535	**0.576**	**0.576**	0.547
HBV9	C_{year}	0.348	0.560	0.590	0.600	0.543	**0.607**	0.598
	C_{snow}	0.221	0.504	0.545	0.561	0.516	**0.567**	0.549
	C_{melt}	0.122	0.425	0.470	0.493	0.462	**0.500**	0.485

from 0.415 to at least 0.640 for GR4J (this corresponds to a change from 0.587 to 0.780 in terms of *NS*), and from 0.348 to at least 0.543 for HBV9 (from 0.516 to 0.704 in terms of *NS*). Although this jump could be expected, it is true even for the simplest SAR (MOHYSE) that has a basic structure (snow accumulation below a fixed threshold temperature and melt above this threshold using a calibrated degree-day factor). The differences in performance between the SARs are much lower than between with and without SAR. Therefore including a SAR in the modelling process seems to have a first-order role on results while changing the formulation of the SAR has a second-order role. However, the difference between the poorest and best SARs remains significant, indicating that it is still worth working on SAR formulation.

- The improvement of performance shows a general positive trend with increasing complexity of the snow accounting routine. However, the most parameterized SAR does not yield the best mean efficiency. The MORD4 SAR (four free parameters) is the most efficient whatever the hydrological model and the criterion. M_SNE, with seven parameters, is only the second and third most efficient SAR for GR4J and HBV9 respectively.
- The results are generally coherent between criteria: when a SAR is better than another, this is generally true for all the criteria. Although this could be expected as calculation periods overlap for the three criteria, it means that the improvement gained on some periods is not cancelled by degradation on others, which stresses the coherence of SARs throughout the year.
- When ranking SARs by increasing performance, the differences between two consecutive SARs are not that large (though most of the time it is significant since we are working on a large dataset). These limited differences may partly explain why the ranking is not exactly the same when using the two different hydrological models. This may also be partly due to the interactions between the SAR and the model itself. It is likely that the two parts interact to some extent during the calibration process. Although model-dependent to some extent, the results show

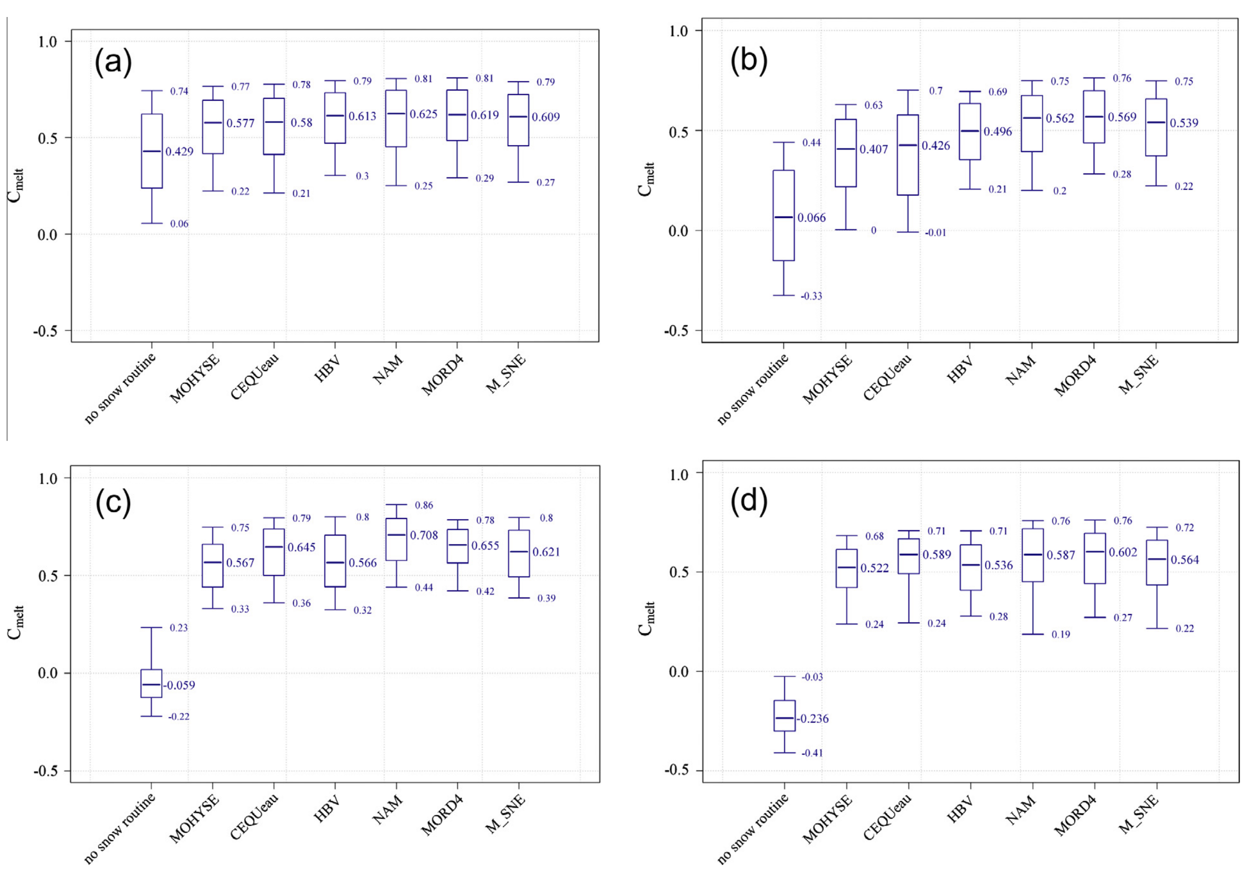

Fig. 3. Box plots (showing 0.10, 0.25, 0.50, 0.75 and 0.90 percentiles) of the distributions of C_{melt} criteria obtained in validation by the GR4J model without any SAR and with one of the six SARs on (a) the 159 moderately snow-affected French catchments, (b) the 90 Swiss and French Alpine catchments, (c) the 94 Swedish catchments and (d) the 36 Canadian catchments.

that the three best SARs are the same for the two models (MORD4, HBV and M_SNE). The added value of these SARs may therefore be quite general. Note that results are the same for C_{melt} for NAM and MORD4 SARs when combined with GR4J. It means that these two SARs provide equivalent results for the melting period. However, MORD4 provides better results with C_{snow}, showing that it better simulates the accumulation period than NAM.

4.2. Can we identify physical/regional trends in the performance of the six snow accounting routines?

In order to check whether there might be regional differences in the ranking of SARs, we present in Fig. 3 plots on each of the catchment subsets identified in Section 2.1. We used here the C_{melt} criterion, but similar results would be obtained with the other criteria. Several comments can be made.

In Sweden (Fig. 3c), GR4J obtains the best efficiency when combined with the NAM SAR. The NAM routine is the only SAR which considers both subdivision into elevation zones and percentage of snow-covered area on each zone to simulate the spatial variation of the snowpack.

For the Swiss and French alpine catchments (Fig. 3b), GR4J presents the best distribution of C_{melt} values when using MORD4. NAM and M_SNE SARs also yield quite high efficiency criteria even if they are slightly below. It is quite unexpected to observe that a lumped approach is the most efficient SAR on mountainous catchments (Swiss and French alpine catchments have the most important reliefs). In the same time, it should be highlighted that MORD4 has been especially developed to be applied on French high mountainous catchments.

In Canada (Fig. 3d), MORD4 is again the SAR with the best distribution of C_{melt} values. NAM presents similar values for high percentiles (0.90 and 0.75) but seems a bit less robust. The lumped approach MORD4 is also very efficient for non-mountainous catchments, but which are strongly snow-affected.

On the moderately snow-affected French catchments (Fig. 3a), MORD4, NAM, HBV and M_SNE present similarly high performance distributions. MOHYSE and CEQUeau are not as efficient. Performances are clearly sensitive to the presence of a function accounting for snowpack maturing (MORD4, NAM, HBV and M_SNE have specific components in this aim). The poor efficiency of the CEQUeau routine is quite surprising since it is of complexity similar to the HBV or the NAM routines (three free parameters). CEQUeau seems more efficient on catchments where the snow processes dominate streamflow patterns (Sweden and Canada), even if it is not the most efficient snow accounting routine on them. It can be linked with the fact that CEQUeau has been first developed to be applied in such environment, in Québec.

Fig. 4 presents results obtained with the HBV9 hydrological model. It shows that:

- In Sweden, NAM is still one of the most efficient snow accounting routine in terms of distribution of performances. Nevertheless, MORD4 and CEQUeau's distributions become closer to the NAM's than with GR4J model.
- In Canada, CEQUeau shows the best distribution of performance among all the SARs. M_SNE and MORD4 present similar high values for their high percentiles (0.90 and 0.75).
- In Switzerland and France, results are quite different than is the case of the GR4J model. The HBV SAR shows an important improvement in terms of performance distributions, which

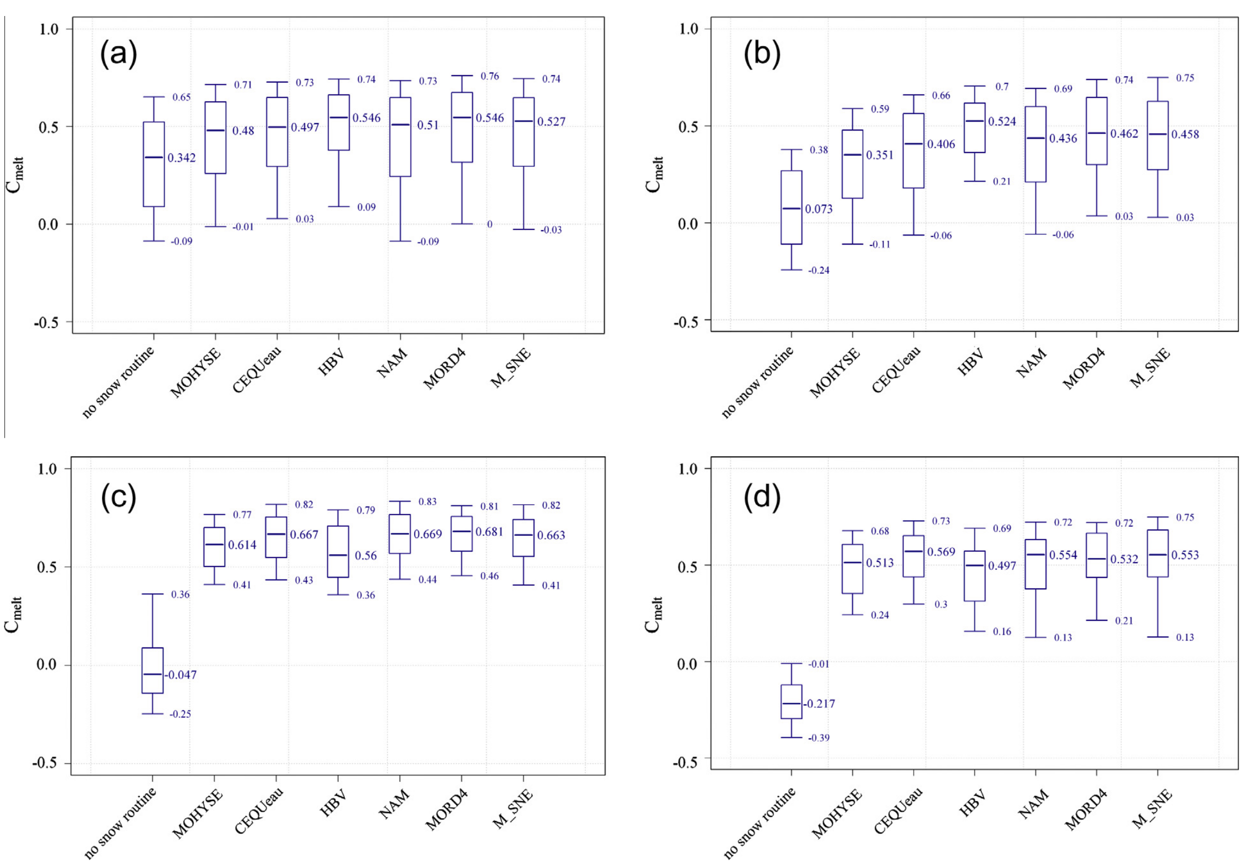

Fig. 4. Box plots (showing 0.10, 0.25, 0.50, 0.75 and 0.90 percentiles) of the distributions of C_{melt} criteria obtained in validation by the HBV9 model without any SAR and with one of the six SARs on (a) the 159 moderately snow-affected French catchments, (b) the 90 Swiss and French Alpine catchments, (c) the 94 Swedish catchments and (d) the 36 Canadian catchments.

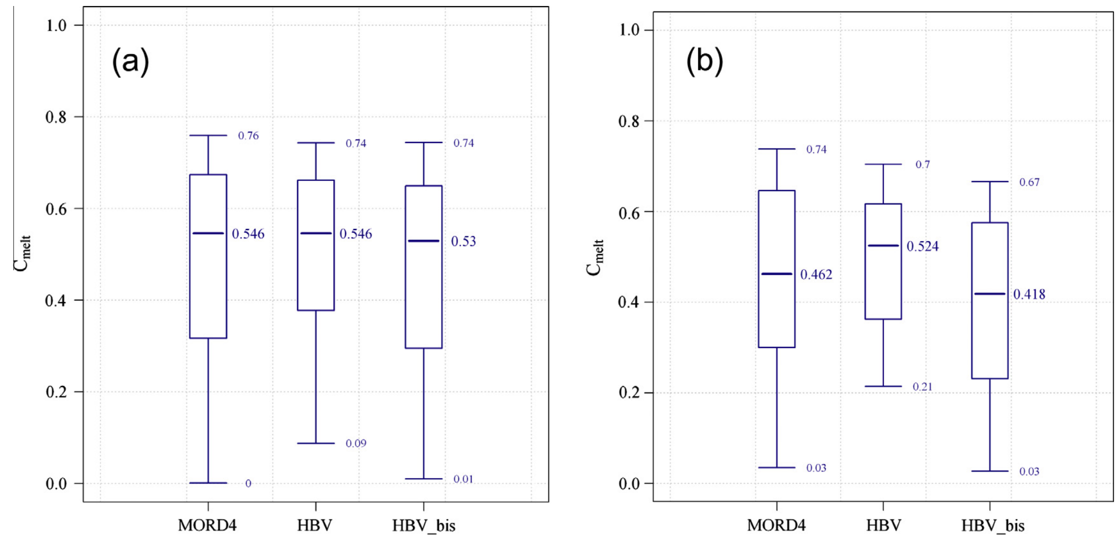

Fig. 5. 0.10, 0.25, 0.50, 0.75 and 0.90 percentiles of C_{melt} obtained by three SARs, MORD4, HBV and HBV_bis (where the solid precipitation correction has been neutralized), with the HBV9 model on (a) French moderately snow-affected catchments (159 basins) and (b) Swiss and French alpine catchments (30 + 60 basins).

Table 4
Number of catchments for which C_{melt} has the highest value for each of the six tested SARs.

	MOHYSE	CEQUeau	HBV	NAM	MORD4	M_SNE	Total
GR4J	6	24	67	**135**	**96**	52	380
HBV9	14	46	**93**	78	**89**	60	380

are even better than other routines for low percentiles and median values. The HBV SAR is the only one which allows a correction of solid precipitation, and so a modification of the precipitation input. HBV9 model can use this parameter to solve water balance issues: it is a way to increase or decrease total input precipitation, and try to fulfill the water balance equation. This additional function is not so important in Sweden and Canada where the water balance is less problematic. In order to better understand how the HBV9 interacts with the snow accounting routine, we tested a version of the HBV SAR without correction of solid precipitation (HBV_bis). Fig. 5 shows that there is no longer improvement in performance compared to MORD4-SAR.

5. Conclusions

5.1. Synthesis

In this paper, we tested six different snow accounting routines, over a dataset of 380 catchments spread over four countries. None of the six SARs tested presents the best performances on the four sub-datasets. Nevertheless, some of them often appear efficient (and robust) in a majority of climatic and hydrological conditions, like MORD4 and, in a second rank, NAM and HBV (Table 4).

From the analysis of the results of this large-scale comparison, we can identify several key-issues:

- It appears particularly important to consider a range of temperatures to determine the form of precipitation rather than a single temperature threshold (like in MOHYSE). Several solutions exist for this: introducing an absolute range of temperature determined *a priori* (CEQUeau, MORD4, M_SNE) or calibrating a free parameter(s) (HBV), or using minimal and maximal air temperature rather than daily mean temperature (Leavesley and Stannard, 1995; Turcotte et al., 2007).
- The question whether a subdivision in elevation zones and/or an introduction of the snowpack variability (percentage of snow-covered area) is needed requires a more detailed assessment, because we found surprising results. Indeed, it is commonly accepted that, as soon as we work on snow-affected catchments with high reliefs, distributed approaches become superior (Ferguson, 1999; WMO, 1986). Nevertheless, MORD4 is a lumped SAR which has most of the time better mean performances than distributed SARs, especially in our catchments with the most important mountainous features (Switzerland and French alpine catchments). Two snow accounting routines introduce a computation of snow-covered percentage (MORD4 and NAM) and show very good efficiency (especially for C_{melt} criterion, when associated with GR4J hydrological model).
- The issue of the level of complexity required in order to account for internal snowpack processes is not clear. Except MOHYSE which is a very simplistic snow accounting routine, other routines choose different and specific approaches to account for delay in snowmelt due to snowpack inertia. Some of them consider liquid water retention (HBV, NAM), others only accept refreezing of snowmelt without any liquid storage through time steps (M_SNE). Others "play" with temperatures index and comparisons between air temperature and snowpack temperature index (MORD4).
- Our results clearly show the potentially dangerous aspects of a free parameter such as the solid precipitation correction factor in the HBV-SAR. Some models can use this function in a different way to correct the whole water balance, and not only the snow undercatch. Modellers should keep in mind the risk of possible compensation and have to be very careful about the degree of freedom they finally allow for automatic calibration.

5.2. Perspectives

There are many existing snow accounting routines in the hydrologic literature, with very different degrees of complexity. Our results show that a selection of six SARs leads to complex results in terms of performance: none of them is better than the others in every situation (for all catchments) and with all hydrological models. In the contrary, it seems that some SARs are good in some specific environments and inversely worse than others in different conditions.

Although we believe that this intercomparison was very instructive, it is not enough to identify the predominant components which would have to be taken into account even in a minimalist snow accounting routine. To reach this goal, we will set up, in a companion paper, an exhaustive and systematic assessment of increasingly complex SARs, which we will assess on the data set gathered for this study.

A complementary analysis (not shown here) indicates that the similarity between models' behavior in terms of flow simulation does not imply similarity of internal state variables of snow accounting processes. This means that the various internal model states can compensate to provide similar simulations between models. It would be very useful to test these SARs using auxiliary observations of snow cover, like those provided by MODIS observations (see e.g. Thirel et al., 2013), which would give additional insights on the reliability of the modelled snow processes.

Acknowledgements

The authors wish to thank the following organizations which accepted to share with us large amounts of data, thus making possible the set-up of our very large catchment data set: Météo–France, Electricité de France (DTG, Grenoble), and SCHAPI (Banque Hydro) for French catchments, Office Fédéral de l'Environnement (OFEV) and Météo Suisse for Swiss catchments, Swedish Meteorological and Hydrological Institute (SMHI) for Swedish catchments, Centre d'Expertise Hydrique du Québec (CEHQ), Hydro-Québec and Environnement Canada for Canadian catchments. The authors also thank the Associate Editor, as well as Dr. Harald Kling and an anonymous reviewer for their constructive comments which helped improving the manuscript.

References

Anderson, E.A., 1973. National Weather Service River Forecast System – Snow Accumulation and Ablation Model. NOAA Technical Memorandum, NWS HYDRO-17.
Bergström, S., 1975. Development of a snow routine for the HBV-2 model. Nordic Hydrol. 6 (2), 73–92.
Bergström, S., 1995. The HBV model. In: Singh, V.P. (Ed.), Computer Models of Watershed Hydrology. Water Resources Publications, pp. 443–476, Chapter 13.
Blöschl, G., Kirnbauer, R., Gutknecht, D., 1991. Distributed snowmelt simulations in an alpine catchment I. Model evaluation on the basis of snow cover patterns. Water Resour. Res. 27 (12), 3171–3179.
Braun, L.N., Brun, E., Durand, Y., Martin, E., Tourasse, P., 1994. Simulation of discharge using different methods of meteorological data distribution, basin discretization and snow modelling. Nordic Hydrol. 25 (1–2), 129–144.
Braun, L.N., Lang, H., 1986. Simulation of Snowmelt Runoff in Lowland and Lower Alpine Regions of Switzerland, vol. 155. IAHS-AISH Publication, pp. 125–140.
Brubaker, K., Rango, A., Kustas, W., 1996. Incorporating radiation inputs into the snowmelt runoff model. Hydrol. Process. 10 (10), 1329–1343.
Cazorzi, F., Dalla Fontana, G., 1996. Snowmelt modelling by combining air temperature and a distributed radiation index. J. Hydrol. 181 (1–4), 169–187.
Charbonneau, R., Lardeau, J.P., Obled, C., 1981. Problems of modelling a high mountainous drainage basin with predominant snow yields. Hydrol. Sci. Bull. 26 (4), 345–361.
DHI, 2009. MIKE 11, A Modelling System for Rivers and Channels, Reference Manual.
Edijatno, De Oliveira, Nascimento, N.D., Yang, X., Makhlouf, Z., Michel, C., 1999. GR3J: a daily watershed model with three free parameters. Hydrol. Sci. J. 44 (2), 263–277.
Essery, R., Yang, Z.L., 2001. An overview of models participating in the snow model intercomparison project (SnowMIP). In: SnowMIP Workshop, 8th Scientific Assembly of IAMAS, Innsbruck.
Etchevers, P. et al., 2004. Validation of the energy budget of an alpine snowpack simulated by several snow models (SnowMIP project). Ann. Glaciol. 38, 150–158.
Ferguson, R.I., 1999. Snowmelt runoff models. Prog. Phys. Geogr. 23 (2), 205–227.
Fortin, V., Turcotte, R., 2007. Le modèle hydrologique MOHYSE, Note de cours pour SCA7420. Université du Québec à Montréal, Département des sciences de la terre et de l'atmosphère.
Franz, K.J., 2006. Characterization of the comparative skill of conceptual and physically-based snow models for streamflow prediction. PhD Thesis Thesis, University of California, Irvine, p. 223.
Garçon, R., 1999. Modèle global pluie-débit pour la prévision et la prédétermination des crues. La Houille Blanche 7 (8), 88–95.
Hock, R., 2003. Temperature index melt modelling in mountain areas. J. Hydrol. 282 (1–4), 104–115.
Klemeš, V., 1986. Operational testing of hydrological simulation models. Hydrol. Sci. J. 31 (1), 13–24.
Klemeš, V., 1990. The Modelling of Mountain Hydrology: The Ultimate Challenge. IAHS-AISH Publication 190. IAHS, pp. 29–43.
Kustas, W.P., Rango, A., Uijlenhoet, R., 1994. A simple energy budget algorithm for the snowmelt runoff model. Water Resour. Res. 30 (5), 1515–1527.
Leavesley, G.H., Stannard, L.G., 1995. The precipitation-runoff modeling system – PRMS. In: Singh, V.P. (Ed.), Computer Models of Watershed Hydrology. Water Resources Publications, Highlands Ranch, pp. 281–310.
Martinec, J., Rango, A., 1981. Areal distribution of snow water equivalent evaluated by snow cover monitoring. Water Resour. Res. 17 (5), 1480–1488.
Mathevet, T., 2005. Quels modèles pluie-débit globaux au pas de temps horaire? Développements empiriques et intercomparaison de modèles sur un large échantillon de bassins versants. PhD Thesis, ENGREF-Cemagref, Paris, p. 463.
Mathevet, T., Michel, C., Andréassian, V., Perrin, C., 2006. A bounded version of the Nash–Sutcliffe criterion for better model assessment on large sets of basins. In: Andréassian, V., Hall, A., Chahinian, N., Schaake, J. (Eds.), Large Sample Basin Experiments for Hydrological Model Parameterisation: Results of the Model Parameter Experiment – MOPEX. IAHS Red Books Series no 307, pp. 211–219.
Morin, G., 1997. CEQUeau, INRS-Eau, Terre & Environnement.
Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models. Part I – A discussion of principles. J. Hydrol. 27 (3), 282–290.
Ohmura, A., 2001. Physical basis for the temperature-based melt-index method. J. Appl. Meteorol. 40 (4), 753–761.
Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andréassian, V., Anctil, F., Loumagne, C., 2005. Which potential evapotranspiration input for a rainfall-runoff model? Part 2 – Towards a simple and efficient PE model for rainfall-runoff modelling. J. Hydrol. 303 (1–4), 290–306. http://dx.doi.org/10.1016/j.jhydrol.2004.08.026.
Oudin, L., Andréassian, V., Mathevet, T., Perrin, C., Michel, C., 2006. Dynamic averaging of rainfall-runoff model simulations from complementary model parameterizations. Water Resour. Res. 42 (7).
Paquet, E., 2004. Evolution du modèle hydrologique MORDOR: modélisation du stock nival à différentes altitudes. La Houille Blanche 2, 75–82.
Perrin, C., Michel, C., Andréassian, V., 2003. Improvement of a parsimonious model for streamflow simulation. J. Hydrol. 279, 275–289.
Pushpalatha, R., Perrin, C., Le Moine, N., Mathevet, T., Andréassian, V., 2011. A downward structural sensitivity analysis of hydrological models to improve low-flow simulation. J. Hydrol. 411 (1–2), 66–76.
Thirel, G., Salamon, P., Burek, P., Kalas, M., 2013. Assimilation of MODIS snow cover area data in a distributed hydrological model using the particle filter. Remote Sens. 5 (11), 5825–5850.
Turcotte, R., Fortin, L.G., Fortin, V., Fortin, J.P., Villeneuve, J.P., 2007. Operational analysis of the spatial distribution and the temporal evolution of the snowpack water equivalent in southern Québec, Canada. Nordic Hydrol. 38 (3), 211–234.
Valéry, A., 2010. Modélisation précipitations – débit sous influence nivale. Élaboration d'un module neige et évaluation sur 380 bassins versants. PhD Thesis, Cemagref-AgroParisTech, Antony & Paris, p. 405.
Valéry, A., Andréassian, V., Perrin, C., 2010. Regionalization of precipitation and air temperature over high-altitude catchments – learning from outliers. Hydrol. Sci. J. 55 (6), 928–940.
Walter, T. et al., 2005. Process-based snowmelt modeling: does it require more input data than temperature-index modeling? J. Hydrol. 300, 65–75.
WMO, 1986. Intercomparison of models of snowmelt runoff. Operational Hydrology Report No 23. World Meteorological Organisation, Genève.