

A Canadian viewpoint on data, information and uncertainty in the context of prediction in ungauged basins

C. Spence, D. H. Burn, B. Davison, D. Hutchinson, T. B. M. J. Ouarda, A. St-Hilaire, F. Weber and P. H. Whitfield

ABSTRACT

The quality (i.e. the degree of uncertainty that results from the interpretation and analysis) of information dictates its value for decision making. There has been much progress towards improving information on the water budgets of ungauged basins by improving knowledge, tools and techniques during the Prediction in Ungauged Basins (PUB) initiative. These improvements, at least in Canada, have come through efforts in both hydrological process and statistical hydrology research. This paper is a review of some recent Canadian PUB efforts to use data to generate information and reduce uncertainty about the hydrological regimes of ungauged basins. The focus is on the Canadian context and the problems it presents, but the lessons learned are applicable to other countries with similar challenges. With a large land mass that is relatively poorly gauged, novel approaches have had to be developed to extract the most information from the available data. It can be difficult in Canada to find gauged or research basins sufficiently similar to ungauged sites of interest that contain the data required to force either statistical or deterministic models. Many statistical studies have improved information or at least an understanding of the quality of that information, of ungauged basin streamflow regimes using innovative regression-based approaches and pooled frequency analysis. Hydrological process research has reduced knowledge uncertainty, particularly in regard to cold regions processes, and this situation has led to the development of new algorithms that are reducing predictive uncertainty. There remains much to do. Current progress has created an opportunity to better integrate statistical and deterministic models via data assimilation of regionalization model estimates and those from coupled atmospheric-hydrological models. Aspects of such a modelling system could also provide more robust uncertainty analyses than traditional approaches.

Key words | Canada, hydrological processes, prediction, statistical hydrology, streamflow, uncertainty

INTRODUCTION

The aim of the Prediction in Ungauged Basins (PUB) initiative is to reduce uncertainty in predictions and improve the quality of information on ungauged basins through improved and more robust methodologies (Sivapalan *et al.* 2003). As not all aspects of all watersheds are measured to the same degree, it is common for many watersheds to be missing the required data and information of the phenomena of interest. Insight from experience and physical

process knowledge can be a powerful tool for hydrologists and water resource managers. More quantitative estimates must be derived using data collected in gauged areas. There are three necessary components to generate information about an ungauged basin using data from a gauged area. These are forcing variables, parameters and a model (Sivapalan *et al.* 2003). Uncertainty associated with each comprises the predictive uncertainty and in turn the quality

C. Spence (corresponding author)

B. Davison

Environment Canada,
Saskatoon,
Canada

E-mail: chris.spence@ec.gc.ca

D. H. Burn

University of Waterloo,
Waterloo,
Canada

D. Hutchinson

P. H. Whitfield

Environment Canada,
Vancouver,
Canada

T. B. M. J. Ouarda

A. St-Hilaire

Institut National de la Recherche Scientifique,
Ville de Québec,
Canada

F. Weber

BC Hydro,
Vancouver,
Canada

P. H. Whitfield

Simon Fraser University,
Burnaby,
Canada

of the information about the ungauged basin. Uncertainty has also been classified as either of a statistical or knowledge nature; where the former can be quantified with likelihood methods but not the latter (Beven 2008). Knowledge uncertainty exists because of differences between what is known about the system and processes of interest, and what it actually is. The result is deficiency in model structure. While knowledge uncertainty is a component of predictive uncertainty, its reduction does not necessarily lead to enhanced predictability (Liu & Gupta 2007). This circumstance arises for instance, when there is a gap between the model complexity and the availability of appropriate forcing and parameterization data. This is a particular problem in data scarce and data poor areas (Hughes 2010).

Hydrologists depend upon data, information, and knowledge. Data, defined simply as a series of observations or measurements, are an important component of the foundation of predictive techniques. Information is acquired through the study of data in context. Just as the quality of data depends on sound collection, so too does the quality of information depend on sound interpretation and analysis. Knowledge is the expertise and skills acquired by a person through experience or education; the theoretical or practical understanding of a subject; what is known in a particular field or in total; facts and information; or awareness or familiarity gained by experience of a fact or situation.

The quality of the information, as measured by the degree of uncertainty that results from the interpretation and analysis, dictates its value for decision making. The most recent discussions about uncertainty in hydrological prediction have focused with how to best estimate the uncertainties and how to present uncertainty to decision makers. While it is important that users have the best tools to understand the quality of information provided to them so that they can understand its value, perhaps the focus has followed this tangent too much, as debates over the minutia of *measuring* uncertainty (Beven 2008; Sivapalan 2009) has pulled the focus of some PUB research away from the original goal to find ways to *reduce* uncertainty. Reducing uncertainty can be achieved through increasing data, information or knowledge.

Data can contain a variety of information. The means by which information is gleaned from data needs to be relevant to the perspective (Gupta *et al.* 2008) in order to be

successful. For example, hydrologists often need information on the states, patterns and trends of hydrological regimes and extremes. The models used to acquire this information on ungauged basins using data from gauged areas need to represent the physical processes that could be active in producing hydrological regimes. Not doing so increases uncertainty and decreases the value of the information, whether the user understands the uncertainties or not.

Hydrologists can improve information value in three general ways. The three main sources of uncertainty can be reduced with efforts to: (1) improve forcing data; (2) improve parameterization; and (3) improve models. It has been argued that hydrological science has far more available data than before and that models have significantly improved, so that the most productive use of resources would be to develop means to ensure the data can provide the best information (Liu & Gupta 2007). This approach may be a bit premature in Canada, where portions of the country are relatively data scarce (Spence *et al.* 2005), and land use and climate change is impacting the hydrological regimes (Brannen & Bielak 2004). Canada has a strong foundation in both hydrological process and statistical hydrology research (Beltaos *et al.* 2005; Pomeroy & Moore 2009), and during the PUB decade this has resulted in improvements to both deterministic and statistical models that can reduce the uncertainty of predictions in ungauged basins. Deterministic models are often described as bottom-up models within the PUB initiative, as these incorporate fundamental physical principles. Top-down models, conversely, utilize the presence of relationships and patterns evident at larger scales that dictate the entire behaviour of hydrological systems. Statistical models are good examples of top-down models, and there are benefits in designing deterministic models in a top-down fashion (Littlewood *et al.* 2003, 2006). There is recognition in the community that these two approaches need not be mutually exclusive (Kinsey-Henderson & Post 2006; Post *et al.* 2006). This paper is a review of recent Canadian PUB efforts to use data to generate information and reduce uncertainty about the hydrological regimes of ungauged basins. The paper includes a discussion of statistical approaches to estimation of flows in ungauged basins as well as a discussion of deterministic approaches for PUB. The paper also includes a discussion of methods to integrate

statistical and deterministic approaches to enhance the estimation of hydrological parameters for ungauged basins.

STATISTICAL METHODOLOGIES

Regression-type approaches

Regression analysis can be used to relate a flow quantile to catchment physiographic, geomorphologic and climatic characteristics (Thomas & Benson 1970), typically using an equation of the form Equation (1):

$$Q_T = ax_1^{\beta_1} x_2^{\beta_2} x_3^{\beta_3} \dots x_p^{\beta_p} \quad (1)$$

where Q_T is the flow quantile of interest, a is a constant, x_i is the i th catchment characteristic, β_i is the i th model parameter and p is the number of catchment characteristics. McLean & Watt (2005) applied multiple linear regression to estimate low-flow quantiles for ungauged catchments in Ontario, while Pandey & Nguyen (1999) used non-linear regression to directly estimate the parameters in Equation (1) for flood quantile estimation.

Artificial neural networks (ANNs) have been used to estimate extreme flow quantiles for ungauged catchments (Shu & Burn 2004; Shu & Ouarda 2008). The use of ANNs can result in more accurate estimates of quantiles than those obtained from non-linear regression approaches (Shu & Ouarda 2007). ANNs have also been used in conjunction with canonical correlation analysis (CCA), in which CCA (Ouarda *et al.* 2001) has been used to define the relationship between physiographic variables and flow quantiles (Shu & Ouarda 2007).

The selection of gauged stations to be used for the estimation of the model parameters in Equation (1) is important: the selected stations should be driven by the same hydrologic processes and similar catchment characteristics to the ungauged sites where quantile estimates are required so as to minimize extrapolation. However, there is also a need for a sufficiently large sample of gauged sites to properly define the parameters in the model. In practice, regional regression often involves the development of different models for different regions. There have been

several applications of CCA to form hydrological neighbourhoods within which regression-based approaches can be used to estimate quantiles for ungauged catchments (Cavadias *et al.* 2001; Ouarda *et al.* 2001, 2008a).

Pooled frequency analysis

Pooled (or regional) frequency analysis entails using information from a group of sites to estimate quantiles at the site of interest. Pooled frequency analysis can be employed at ungauged locations, where information from similar sites that are gauged is used to estimate the flow quantile magnitude at the ungauged site. Pooled frequency analysis involves the formation of a pooling group (region or neighbourhood) that is required to achieve the spatial transfer of information. A pooling group has come to mean a collection of catchments, not necessarily geographically contiguous, that can be considered to be similar in terms of hydrologic response (Burn 1990). A focused pooling group is a site-specific collection of gauged catchments that are useful for the transfer of extreme flow information for the estimation of flood quantiles for a specific ungauged catchment of interest. The goal of the pooling process is the identification of groupings of catchments that are sufficiently similar to warrant the combination of extreme flow information from sites within the pooling group. Several researchers have addressed the definition of focused pooling techniques, which specifically design the homogeneous group of catchments for a site of interest. A focused pooling group is generally identified using catchment or climate similarity measures (Acreman & Wiltshire 1989; Burn 1990, 1997; FEH 1999; Castellarin *et al.* 2001; Cunderlik & Burn 2002; Kjeldsen & Jones 2009) but geographical location (Burn & Goel 2000) can be successful sometimes as well.

Pooled frequency analysis involves estimating the growth curve by combining (pooling) information from all sites in the pooling group. The information that is combined from the pooling group can be the probability weighted moments (PWMs), or similarly the L-moments, calculated for each gauged site (Hosking & Wallis 1997). The parameters for the distribution function are then estimated for the pooled growth curve. The second requirement for pooled frequency analysis is an estimate for the index event magnitude. The index event typically represents the

mean, or sometimes the median, of the data series for the site of interest. For an ungauged site, the index event can be estimated from information available at sites for which data records exist, using regression-type approaches (see Grover *et al.* 2002) or ANNs (Shu & Burn 2004).

An important issue with regard to pooled estimation of low flows is the multivariate nature of low-flow events, wherein there is often an interest in both the magnitude and the duration of low flows, typically in the context of flow events below a defined threshold. Farid & Burn (2006), Ouarda (2007) and Sadri & Burn (2011) proposed multivariate pooled procedures for the joint estimation of several low-flow characteristics based on the use of copulas, which can efficiently estimate the bivariate (or multivariate) distribution for dependent variables. Software exists to implement pooled frequency analysis, for either high or low flows. An example is the REGIONS software (Ouarda *et al.* 2002, 2008b), which includes modules for pooled frequency analysis in both stationary and non-stationary frameworks.

Geostatistical approaches

Geostatistical methods can be used to estimate the value of a variable of interest at an ungauged location based on measured values for the variable of interest at other locations. Many of these approaches are based on different types of kriging. Kriging methods estimate the variable of interest as a weighted average of the neighbouring observations; the kriging problem consists of finding the optimal weights based on the spatial correlation structure of the data (Skøien & Blöschl 2007). There have been several applications of kriging to the problem of estimating flow quantiles at ungauged locations. Chokmani & Ouarda (2004) used kriging in physiographic space to estimate quantiles for ungauged catchments. They used both CCA and principal components (PCs) analysis to define the multivariate space in which kriging was applied.

Hydrological indices

The multivariate nature of low flow has also been recognized for more than a decade within the context of the natural flow paradigm, which is the basis for a purely

hydrological approach to assess environmental or instream flows. In this approach, the natural flow of a river is used to define a reference condition (Poff *et al.* 1997). Hydrologic indices (HIs) describing magnitude, frequency, duration and variability have been defined and calculated on gauged basins (e.g. Daigle *et al.* 2011) to be used as references. Following the work of Monk *et al.* (2007), Daigle *et al.* (2011) used PC analysis to decrease the dimensionality of the HI space and found that the first three PCs represented combinations of low-flow indices from the amplitude, duration, variability and frequency and timing categories in decreasing order of explained variance. Projections of scores in PC space showed that contiguous, relatively homogenous regions can easily be identified in Eastern Canada (Atlantic Canada and Quebec). In the Prairies, Beveridge (2010) found that physiographic and climate variables such as annual precipitation, elevation, slope, land use and drainage area can explain a meaningful percentage of variance for indices in each of the aforementioned categories.

Assessing quality of information

Several techniques can be used for quantifying the degree of information in estimates of ungauged flows. Jackknife and bootstrap methodologies for assessing uncertainty have been used extensively in Canada (e.g. Ouarda *et al.* 2001); often these studies demonstrate that the state of streamflow regimes is poorly known because of sparse observation networks and short records, especially in northern Canada (Spence & Burke 2008). The combination of relatively short observation records and environmental change can create a great deal of angst amongst practicing hydrologists in Canada. This uncertainty associated with non-stationary conditions comes from trends in forcing variables and parameters. This has prompted the development of new techniques for estimating extreme quantiles under non-stationary conditions. Most of this work relies on the use of covariates (time, or variables such as low frequency climate oscillation indices) to facilitate non-stationary estimation at the local and regional scale (El-Adlouni *et al.* 2007; Leclerc & Ouarda 2007; Ouarda *et al.* 2008b). A more difficult problem is that of uncertainty in knowledge. Model structure may need to change as predominant hydrological processes change with climate (e.g. more

groundwater flow in thawing permafrost), and there are few ways to confirm if existing relationships between parameters and streamflow will apply in the future. Improvements may be gained using tools such as fuzzy logic and Bayesian modelling (Ouarda *et al.* 2008b). Burn *et al.* (2007) explored several approaches for quantifying the uncertainty in ungauged estimates of high flows and concluded that further work is required to reach definitive conclusions as to the preferred approach for quantifying uncertainty in ungauged flow estimates.

DETERMINISTIC MODELLING

Reducing uncertainty in knowledge

While Canadian hydrological regimes and trends of ungauged basins tend to be estimated using statistical modelling, the application of deterministic modelling is not uncommon. For example, in a study by St-Hilaire *et al.* (2010) the Streamflow Simulation and Reservoir Regulation (SSARR) (United States Army Corps of Engineers 1987) and HSAMI (Bisson & Roberge 1983) models were both used to calculate low-flow HIs from simulated flow time series on Quebec's Romaine River calibrated for the period 1965–1976. Daily flows were simulated for the recent past (up to 1990) using precipitation and air temperature inputs originating from climate models. The same approach was used to generate future hydrological scenarios for a 30-year period centred on 2050. Ten scenarios were used. In each case, low-flow hydrological indices quantifying amplitude, duration, frequency, variability and timing of events, were calculated and compared for the entire period.

Outside of hydroelectrical utilities and flood forecasting offices, the primary use of deterministic models is to answer research questions. Recent PUB efforts in Canada with deterministic modelling have focused on cold regions and small basins as it is in this regime and scale at which practicing hydrologists have the most difficulty with prediction using statistical models (Spence *et al.* 2005). At this scale nuances among different physiographies and predominant hydrological processes within a watershed can have large impacts on catchment response; this hinders statistical approaches when a small sample size means the variability

tends not to be captured, especially in Canada. With small scale physical processes so influential to small scale catchment response, uncertainty in physical process knowledge can undermine total predictive uncertainty at this scale. Understanding the physical processes is thus crucial for sound prediction at this scale (Spence *et al.* 2005).

For instance, infiltration into frozen ground is a critical hydrological process across much of Canada, but often poorly represented mathematically in deterministic models. Some hydrological models use frozen ground infiltration schemes but the approaches vary widely and few have been widely tested for organic soils. Sound understanding of how water infiltrates into frozen organic soils, in particular, is important because of their wide distribution across Canada's cold regions. Zhang *et al.* (2010) reviewed infiltration algorithms and found the single most important factor controlling infiltration into frozen ground was the soil temperature and depth of thaw. Sensitivity tests showed that evapotranspiration does not affect infiltration rates into frozen ground. Additionally, preferential flow in organic soils can be represented with high hydraulic conductivity values and no special additional algorithm is needed. Zhang *et al.* (2010) increased knowledge by providing valuable information on the infiltration process and the implementation of appropriate algorithms, which will improve prediction.

Ellis *et al.* (2010) tested the ability of the snow schemes in the Cold Regions Hydrological Model (CRHM) (Pomeroy *et al.* 2007) to simulate both the mean and maximum snow water equivalent states and the timing and quantity of accumulation and ablation at five paired forest and clearing sites in Canada, Switzerland, Finland, Japan and the USA. Comparisons with observations in coniferous forests suggested that the understanding and representation of many of the processes are generally sound. The snow energetics were well captured by the model, but not those internal to the snowpack during large snowfalls. Also, less appears to be known about canopy unloading, though progress is being made. What is most encouraging about this research is that the diagnostic and predictive tools are available to investigate the effect of forest covers on snow processes. This is important in much of western Canada where the impacts of climate variability and land use change on forest canopies and in turn, streamflow regimes are growing concerns.

Applying regional physioclimatic data to produce information

Deterministic models are one of the regionalization tools available to the hydrologist. One specific challenge of applying deterministic models to ungauged or poorly gauged basins is dealing with the uncertainties in model parameters that exist because of only a partial knowledge of actual physioclimatic basin data characteristics (e.g. geology and soils). For example, there were numerous issues with parameterization of the landscape heterogeneity in Fang *et al.*'s (2010) application of CRHM to the ~450 km² Smith Creek basin in the Canadian Prairies. Furthermore, the snowmelt process active in nival streams such as Smith Creek requires hydrological models to carry many more parameters to describe dominant hydrological process than their rainfall-runoff counterparts. To address this complexity, it is most common to reduce the number of parameters by using observations and calibrated model values from comparable areas which are then assumed to transfer to self-similar hydrological response units (HRUs) defined with satellite imagery, digital elevation models (DEMs) and geographic information systems (GIS). Hydrological process research has revealed that the topology of HRUs, or their relative location and distribution, has profound implications for runoff response across many Canadian landscapes such as the Boreal Plains (Devito *et al.* 2005), Canadian Shield (Allan & Roulet 1994) and prairie (Stichling & Blackwell 1958). However, HRU topology remained poorly represented in any Canadian deterministic model until Fang *et al.* (2010) addressed this situation by defining representative sub-basins which each had typical landscape sequences through which runoff moved. Their model could predict the timing and magnitude of peak streamflow, but experienced problems with recession. The oversimplification of a semi-distributed representative basin approach was suspected to remain a major source of uncertainty that could be addressed with a more fully distributed model. It also may have been that too few landscape types were selected, and therefore did not encapsulate all the variability. Nevertheless, the application showed that such models could be generally successful in predicting some aspects of the water budget if high spatial resolution data are available.

High spatial resolution data, however, are often not available in Canada, and this impacts the amount of information that can be derived from ungauged basins with regional physiographic data. Moore *et al.* (2010) used a simple spatially distributed water balance model to predict monthly hydrological regimes in British Columbia. Model parameters were established from published studies with minimal manual calibration. Spatial coverages of surficial geology, glaciers, lakes, and biogeoclimatic zones were used to define the site characteristics at each model grid cell. In general, the model was able to replicate hydroclimatic regimes, but was generally less successful in sparsely observed regions.

THE INTEGRATION OF STATISTICAL AND DETERMINISTIC MODELS

With relatively sparse monitoring data and many complicated and diverse hydrological processes, it is often difficult to produce sound information on certain Canadian hydrological regimes. Sometimes, it may prove prudent to reduce uncertainty and improve knowledge of fluxes by estimating them with several independent and separate methods. Figure 1 presents a conceptualization of the integration of statistical and deterministic models in hydrology. At different levels in this conceptualization, shown by levels (a), (b), and (c) in Figure 1, we can compare the degree of integration between these two approaches. At the lowest level is the comparison of results from different approaches without any integration of statistical and deterministic models (Figure 1 level (a)). At this level, the hydrologist compares predictions using separate methodologies. However, it is rare in Canada to predict hydrological fluxes in ungauged basins using both statistics and deterministic modelling with the intent of comparing results to reduce uncertainty. There are examples in which some studies have compared results with independent research in order to reduce a high level of uncertainty in results (e.g. Spence & Burke 2008), but this approach does not generate all possible synergies.

The successful implementation of models benefits from integration. A simple illustration is that of a statistical

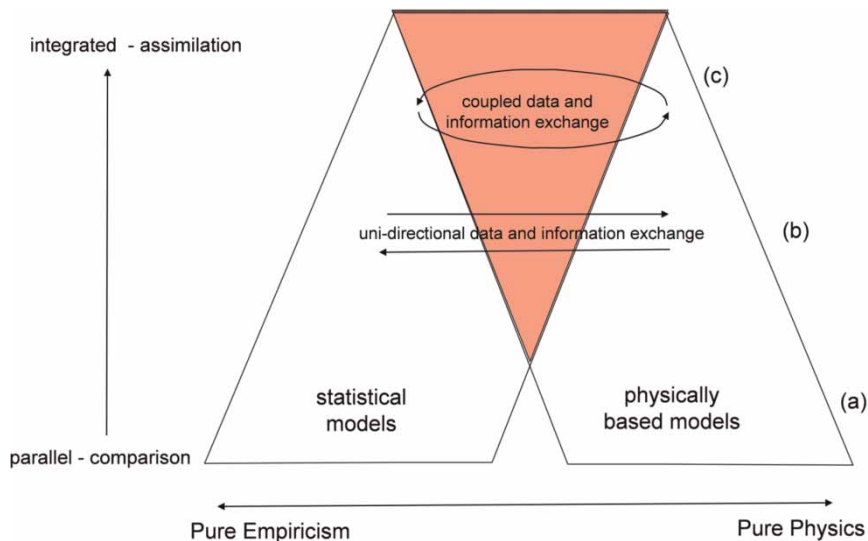


Figure 1 | The spectrum of integration of statistical and deterministic models. This conceptual model shows the range on integration in hydrology. At the lowest level (a), statistical models and physical models are completely separate and comparable to each other. At the highest level (c), the statistical and physical models are fully coupled. In between area (b), information generated by one type of model informs the other without feedback.

regional estimation model applied to ungauged sites that requires a set of explanatory covariates to function. These covariates represent the physiographical and meteorological characteristics of gauged basins. At this conceptual level of integration (Figure 1 level (b)), an understanding of the physical processes involved and the physical characteristics of the basins informs statistical models, so at least some uni-directional exchange of data and information are required. For example, *Cavadias et al. (2001)* selected covariates such as drainage area, mean annual precipitation, percentage of wetland cover and slope of the channel based on *a priori* knowledge of the basic physical processes that drive streamflow generation. This level of integration is both essential and ubiquitous, and it could be argued it is so common that hydrologists do not even notice its occurrence.

Any lack of an understanding of how physical processes should be parameterized in deterministic models results in much uncertainty in the selection and estimation of model parameters. Hydrologists use calibration procedures that perturb model parameter values as a means of dealing with the uncertainty associated with the physical processes. This calibration is almost always accomplished with statistical techniques (*Soulis et al. 2005*) that compare observed and simulated hydrographs. Unfortunately, from the perspective of better integrating predictive approaches, the calibration procedures are usually applied as part of a

hindcasting exercise to minimize the difference between the modelled and observed streamflows rather than as part of an interactive procedure that assists in selecting and estimating values of the 'best' parameters for forecasting purposes. Furthermore, most of the time the normal calibration procedure fails to produce probabilities of parameter values, despite the possibility to do so. At this level of integration, research is often directed by either statistical hydrologists or experimentalists, and hence is not optimal. Canadian hydrologists could learn from Australian research (*Post et al. 2006*) to develop new improved model approaches that apply process-based understanding gained from deterministic (bottom-up) models to modify the representation of processes in statistical (top-down) models. Furthermore, it is recommended that progress should be made in not just developing models that *link*, but effort should be made to progress towards iterative models that *couple* the two approaches, to extract the strengths of both, while at the same time eliminating the weaknesses.

Although currently rare in practice, hydrological ensemble predictions (H-EPS) and data assimilation together represent an opportunity to fully couple statistical (top-down) and deterministic (bottom-up) approaches (Figure 1 level (c)). Such a fully coupled system could be used to assist in dealing with design, flood warning and resource management issues in gauged and ungauged basins.

Atmospheric ensemble prediction is far more common than hydrologic ensemble prediction, possibly because basic atmospheric physics are relatively well understood, and initial condition representation is considered as the largest source of uncertainty. There has been a systematic attempt in Canada under a variety of research programmes, including some within the PUB initiative, to join atmospheric and hydrological models into hydrology–land surface schemes (HLSSs) using soil–vegetation–atmospheric transfer (SVAT) schemes as the common link. These deterministic models have developed to the point at which they are suitable for hydrological process research purposes, but their use in hydrometeorological forecasting has been limited.

To address this gap, Pietroniro *et al.* (2007) describe the application of a coupled modelling system, MESH (Modélisation Environnementale Communautaire – Surface and Hydrology), to hydrological ensemble forecasting of the water budgets of the North American Laurentian Great Lakes. In this example, which was meant as a proof-of-concept for using H-EPS for forecasting, the uncertainty in ‘future’ Lake Ontario water levels was predicted, as well as ensemble snowpack predictions. The first step in considering how H-EPS can be improved is to enhance the ‘traditional’ approach in two ways. The first enhancement is to assimilate data. Assimilated data could include meteorological ensemble prediction system precipitation fields at the temporal and spatial scales of the hydrological model. This has been identified as a requirement for post-processing and downscaling of forcings for input to hydrological models by Schaake *et al.* (2010) and is being pursued in Canada by the realization of products such as the Canadian Precipitation Assimilation (CaPA; Mahfouf *et al.* 2007). CaPA is an assimilation of a regional deterministic precipitation model and observed data. Using products like CaPA and other numerical weather model forcing fields, in the calibration of the reference deterministic hydrologic model run, ensures that the climatological characteristics between the calibrated run and the ensemble members are the same. Testing and assimilation of regionalized streamflow estimates from statistical methods into the H-EPS also show great promise for improving prediction.

The second enhancement in considering how short-term H-EPS can be improved is to more fully explore the sources of uncertainty in the modelled streamflow, which is an

improvement over the traditional approach of simply calibrating the hydrological model and then forcing the calibrated hydrological model with output from each member of a meteorological ensemble prediction system. The uncertainty analysis could be accomplished by running each member of the meteorological ensemble prediction system with a variety of hydrological models. In a linked mode, the perturbations for the hydrological model will be in the model forcing and initial conditions, as well as the model physics and parameters. The HLSS model parameters and physics that should be perturbed would attempt to encompass the relevant physical processes being represented in the model. Linked runs could also be compared with a fully coupled atmospheric–land-surface hydrological model. In the coupled case, various aspects of the modelling system could be tested by perturbing the atmospheric and land-surface initial conditions, parameters, and physics. This approach comes much closer to a full uncertainty analysis than a traditional approach.

CONCLUDING REMARKS

There has been much progress towards transforming data into information on the water budgets of Canadian ungauged basins by improving knowledge, tools and techniques through efforts in both hydrological process and statistical hydrology research. It can be difficult in Canada to find gauged or research basins similar to ungauged sites of interest that contain the data required to support either statistical or deterministic models. Several statistical studies that have improved information, or the quality of that information, for ungauged basin streamflow regimes were reviewed. Progress on addressing uncertainties associated with non-stationarity has been made, but more work is required to address the problem of uncertainty in how model structure may need to change as predominant hydrological processes change. Hydrological process research has reduced knowledge uncertainty, particularly in regards to cold regions processes, and this has led to the development of new algorithms that have reduced predictive uncertainty for some processes. The practitioner and researcher communities are increasingly well placed for better integration of statistical and deterministic approaches through many

avenues. These approaches could include stronger partnerships between the statistical and deterministic research and modelling communities to ensure the most knowledge can be generated from research that compares and links data and information about the hydrological regimes of ungauged basins. From a more technical perspective, the recent development of coupled models and assimilation methods for observations, and statistical (top-down) and deterministic (bottom-up) model data could also produce sound information about estimates for ungauged basins and the uncertainty associated with them.

REFERENCES

- Acreman, M. C. & Wiltshire, S. E. 1989 The regions are dead: long live the regions. Methods of identifying and dispensing with regions for flood frequency analysis. In: *FRIENDS in Hydrology* (L. Roald, K. Nordseth & K. A. Hassel, eds). IAHS Publications, Wallingford, UK, No. 187, pp. 175–188.
- Allan, C. & Roulet, N. 1994 [Runoff generation in zero order Precambrian shield catchments: the stormflow response of a heterogeneous landscape](#). *Hydrological Processes* **8**, 369–388.
- Beltaos, S., Martz, L. W., Moore, D. & Pomeroy, J. W. 2005 [Preface: Canadian geophysical union hydrology section special issue](#). *Hydrological Processes* **19**, 1.
- Beven, K. 2008 [On doing better hydrological science](#). *Hydrological Processes* **22**, 3549–3553.
- Beveridge, D. 2010 Hydrologic Indices for the Prairie Provinces. M.Sc. Thesis, University of New Brunswick, pp. 108.
- Bisson, J. L. & Roberge, F. 1983 [Prévision des apports naturels: Expérience d'Hydro-Québec](#) Workshop on Flow Predictions, Toronto, November 1983.
- Brannen, L. & Bielak, A. 2004 [Threats to Water Availability in Canada](#). Environment Canada, Burlington, NWRI Scientific Assessment Report Series #3 and ACSD Science Assessment Series #1, pp. 128.
- Burn, D. H. 1990 [Evaluation of regional flood frequency analysis with a region of influence approach](#). *Water Resources Research* **26**, 2257–2265.
- Burn, D. H. 1997 [Catchment similarity for regional flood frequency analysis using seasonality measures](#). *Journal of Hydrology* **202**, 212–230.
- Burn, D. H. & Goel, N. K. 2000 [The formation of groups for regional flood frequency analysis](#). *Hydrological Sciences Journal* **45**, 97–112.
- Burn, D. H., Ouarda, T. B. M. J. & Shu, C. 2007 [Estimation of extreme flow quantiles and quantile uncertainty for ungauged catchments, quantification and reduction of predictive uncertainty for sustainable water resources management](#). *Proceedings of Symposium HS2004 at IUGG2007*, Perugia, July 2007. IAHS Publication 313, pp. 417–424.
- Castellarin, A., Burn, D. H. & Brath, A. 2001 [Assessing the effectiveness of hydrological similarity measures for flood frequency analysis](#). *Journal of Hydrology* **241** (3–4), 270–285.
- Cavadias, G. S., Ouarda, T. B. M. J., Bobée, B. & Girard, C. 2001 [A canonical correlation approach to the determination of homogeneous regions for regional flood estimation of ungauged basins](#). *Hydrological Sciences Journal* **46**, 499–512.
- Chokmani, K. & Ouarda, T. B. M. J. 2004 [Physiographical space-based kriging for regional flood frequency estimation at ungauged sites](#). *Water Resources Research* **40**, W12514.
- Cunderlik, J. M. & Burn, D. H. 2002 [Analysis of the linkage between rain and flood regime and its application to regional flood frequency estimation](#). *Journal of Hydrology* **261** (1–4), 115–131.
- Daigle, A., St-Hilaire, A., Beveridge, D., Caissie, D. & Benyahya, L. 2011 [Multivariate analysis of low flow regimes in eastern Canadian rivers](#). *Hydrological Sciences Journal* **56**, 51–67.
- Devito, K., Creed, I., Gan, T., Mendoza, C., Petrone, R., Silins, U. & Smerdon, B. 2005 [A framework for broad scale classification of hydrologic response units on the boreal plain: is topography the last thing to consider?](#) *Hydrological Processes* **19**, 1705–1714.
- El-Adlouni, S., Ouarda, T. B. M. J., Zhang, X., Royand, R. & Bobée, B. 2007 [Generalized maximum likelihood estimators for the nonstationary Generalized Extreme Value model](#). *Water Resources Research* **43**, W03410.
- Ellis, C. R., Pomeroy, J. W., Brown, T. & MacDonald, J. 2010 [Simulation of snow accumulation and melt in needleleaf forest environments](#). *Hydrology and Earth System Sciences* **14**, 925–940.
- Fang, X., Pomeroy, J. W., Westbrook, C. J., Guo, X., Minke, A. G. & Brown, T. 2010 [Prediction of snowmelt derived streamflow in a wetland dominated prairie basin](#). *Hydrology and Earth System Sciences* **14**, 991–1006.
- Farid, R. & Burn, D. H. 2006 [Bivariate frequency analysis of drought characteristics](#). *Proceedings, Canadian Water Resources Association 59th Annual Conference*, Toronto, Ontario, Canada.
- FEH 1999 [Flood Estimation Handbook](#). Institute of Hydrology, Wallingford, UK.
- Grover, P. L., Burn, D. H. & Cunderlik, J. M. 2002 [A comparison of index flood estimation procedures for ungauged catchments](#). *Canadian Journal of Civil Engineering* **29**, 734–741.
- Gupta, H. V., Wagener, T. & Liu, Y. 2008 [Reconciling theory with observations: elements of a diagnostic approach to model evaluation](#). *Hydrologic Processes* **22**, 3802–3813.
- Hosking, J. R. M. & Wallis, J. R. 1997 [Regional Frequency Analysis: An Approach Based on L-Moments](#). Cambridge University Press, Cambridge, UK.
- Hughes, D. A. 2010 [Hydrological models: mathematics or science?](#) *Hydrological Processes* **24**, 2199–2201.

- Kinsey-Henderson, A. E. & Post, D. A. 2006 Evaluation of the scale dependence of a spatially explicit hillslope sediment delivery ratio model. iEMSs 2006 Summit on Environmental Modelling and Software, Burlington, USA, 9–13 July 2006. <http://www.iemss.org/iemss2006/papers/s2/Post.pdf>.
- Kjeldsen, T. R. & Jones, D. A. 2009 A formal statistical model for pooled analysis of extreme floods. *Hydrology Research* **40**, 465–480.
- Leclerc, M. & Ouarda, T. B. M. J. 2007 Non-stationary regional flood frequency analysis at ungauged sites. *Journal of Hydrology* **343**, 254–265.
- Littlewood, I. G., Croke, B. F. W., Jakeman, A. J. & Sivapalan, M. 2003 The role of ‘top-down’ modeling for prediction in ungauged basins (PUB). *Hydrological Processes* **17**, 1673–1679.
- Littlewood, I. G., Croke, B. F. & Post, D. A. 2006 The PUB Top-Down Modelling Working Group (TDWG): initial development and activities. In: *Prediction in Ungauged Basins: Promise, and Progress. Proceedings of a symposium held during the Seventh IAHS Assembly*, IAHS Publications, Wallingford, UK, Vol. 303, pp. 443–453.
- Liu, Y. & Gupta, H. V. 2007 Uncertainty in hydrologic modeling: toward an integrated data assimilation framework. *Water Resources Research* **43**, W07401.
- Mahfouf, J.-F., Brasnett, B. & Gagnon, S. 2007 A Canadian precipitation analysis (CaPA) project: description and preliminary results. *Atmosphere-Ocean* **45**, 1–17.
- McLean, R. K. & Watt, W. E. 2005 Regional low flow frequency relations for central Ontario. *Canadian Water Resources Journal* **30**, 179–196.
- Monk, W., Wood, P., Hannah, D. & Wilson, D. 2007 Selection of river flow indices for the assessment of hydroecological change. *River Research and Applications* **23**, 113–122.
- Moore, R. D., Trubilowicz, J. & Buttle, J. 2010 Hydroclimatic classification and annual runoff prediction for ungauged basins using a spatially distributed climatic water balance model. *Poster presented at the 63rd CWRA National Conference*, 15–18 June 2010, Vancouver, BC, Canada.
- Ouarda, T. B. M. J. 2007 Statistical estimation of low flow characteristics at ungauged basins, Atelier de l’ACRH ‘prediction in ungauged basins, low flows’, Québec, Canada, 12–13 avril 2007.
- Ouarda, T. B. M. J., Bâ, K. M., Diaz-Delgado, C., Cârsteanu, C., Chokmani, K., Gingras, H., Quentin, E., Trujillo, E. & Bobée, B. 2008a Intercomparison of regional flood frequency estimation methods at ungauged sites for a Mexican case study. *Journal of Hydrology* **348**, 40–58.
- Ouarda, T. B. M. J., Charron, C. & St-Hilaire, A. 2008b Statistical models and the estimation of low flows. *Canadian Water Resources Journal* **33**, 195–206.
- Ouarda, T. B. M. J., Girard, C., Cavadias, G. S. & Bobée, B. 2001 Regional flood frequency estimation with canonical correlation analysis. *Journal of Hydrology* **254**, 157–173.
- Ouarda, T. B. M. J., Gingras, H., Kouider, A., Rudolf, Z. R., Haché, M., Barbé, M. & Bobée, B. 2002 REGIONS, a general and automatic model for flood and drought regional estimation at ungauged basins. First Biennial Workshop of the Statistical Hydrology Committee of the CGU-HS, 2002 Annual Scientific Meeting of the CGU, Banff, AB, Canada, 18–21 May 2002.
- Pandey, G. R. & Nguyen, V.-T.-V. 1999 A comparative study of regression based methods in regional flood frequency analysis. *Journal of Hydrology* **225**, 92–101.
- Pietroniro, A., Fortin, V. F., Kouwen, N., Neal, C., Turcotte, R., Davison, B., Verseghy, D., Soulis, E. D., Caldwell, R., Evora, N. & Pellerin, P. 2007 Development of the MESH modeling system for hydrological ensemble forecasting of the Laurentian Great Lakes at the regional scale. *Hydrology and Earth System Sciences* **11**, 1279–1294.
- Poff, N. L., Allan, J. D., Bain, M. B., Karr, J. R., Prestegard, K. L., Richter, B. D., Sparks, R. E. & Stromberg, J. C. 1997 The natural flow regime. A paradigm for river conservation and restoration. *BioScience* **47**, 769–784.
- Pomeroy, J. W. & Moore, R. D. 2009 Introduction to the special issue on recent advances in Canadian hydrology, 2003–2007. *Canadian Water Resources Journal* **34**, 109–112.
- Pomeroy, J. W., Gray, D. M., Brown, T., Hedstrom, N. R., Quinton, W., Granger, R. J. & Carey, S. 2007 The cold regions hydrological model: a platform for basing process representation and model structure on physical evidence. *Hydrological Processes* **21**, 2650–2667.
- Post, D. A., Kinsey-Henderson, A. E., Bartley, R. & Hawdon, A. 2006 Deriving a spatially-explicit hillslope sediment delivery ratio model based on the travel time of water across a hillslope. iEMSs 2006 Summit on Environmental Modelling and Software, Burlington, USA, 9–13 July 2006. http://www.iemss.org/iemss2006/papers/s12/115_Post_1.pdf.
- Sadri, S. & Burn, D. H. 2011 A fuzzy C-means approach for regionalization using a bivariate homogeneity and discordancy approach. *Journal of Hydrology* **401**, 231–239.
- Schaake, J., Pailleux, J., Thielen, J., Arritt, R., Hamill, T., Lifeng, L., Martin, E., McCollor, D. & Pappenberger, F. 2010 Summary of recommendations of the first workshop on postprocessing and downscaling atmospheric forecasts for hydrologic applications held at Meteo-France, Toulouse, France, 15–18 June 2009. *Atmospheric Science Letters* **11**, 59–63.
- Shu, C. & Burn, D. H. 2004 Artificial neural network ensembles and their application in pooled flood frequency analysis. *Water Resources Research* **40**, W09301.
- Shu, C. & Ouarda, T. B. M. J. 2007 Flood frequency analysis at ungauged sites using artificial neural networks in canonical correlation analysis physiographic space. *Water Resources Research* **43**, W07458.
- Shu, C. & Ouarda, T. B. M. J. 2008 Regional flood frequency analysis at ungauged sites using the adaptive neuro-fuzzy inference system. *Journal of Hydrology* **349**, 31–43.
- Sivapalan, M. 2009 The secret to ‘doing better hydrological science’: change the question! *Hydrological Processes* **23**, 1391–1396.
- Sivapalan, M., Takeuchi, K., Franks, S. W., Gupta, V. K., Karambiri, H., Lakshmi, V., Liang, X., McDonnell, J. J.,

- Mendiondo, E. M., O'Connell, P. E., Oki, T., Pomeroy, J. W., Schertzer, D., Uhlenbrook, S. & Zehe, E. 2003 [IAHS decade on predictions in ungauged basins \(PUB\), 2003–2012: shaping an exciting future for the hydrological sciences](#). *Hydrological Sciences—Journal des Sciences Hydrologiques* **48**, 857–880.
- Skøien, J. O. & Blöschl, G. 2007 [Spatiotemporal topological kriging of runoff time series](#). *Water Resources Research* **43**, W09419.
- Soulis, E. D., Kouwen, N., Pietroniro, A., Seglenieks, F. R., Snelgrove, K. R., Pellerin, P., Shaw, D. W. & Martz, L. W. 2005 A framework for hydrological modelling in MAGS. In: *Prediction in Ungauged Basins: Approaches from Canada's Cold Regions* (C. Spence, J. W. Pomeroy & A. Pietroniro, eds). Canadian Water Resources Association, Ottawa, pp. 119–138.
- Spence, C. & Burke, A. 2008 [Estimates of Canadian Arctic archipelago runoff from observed hydrometric data](#). *Journal of Hydrology* **362**, 247–259.
- Spence, C., Pomeroy, J. W., Pietroniro, A., Shook, K. & Marks, D. 2005 Introduction. In: *Prediction in Ungauged Basins: Approaches from Canada's Cold Regions* (C. Spence, J. W. Pomeroy & A. Pietroniro, eds). Canadian Water Resources Association, Ottawa, pp. 1–20.
- St-Hilaire, A., Daigle, A., Thiémonge, N., Pion, C., Roy, L., Ouarda, T. B. M. J. & Diatezua, J. 2010 [Analyses des étiages dans un contexte de changements climatiques](#). Rapport final du projet A3.2. Rapport de recherche R-1133 confidentiel. viii+100 pp.
- Stichling, W. & Blackwell, S. R. 1958 Drainage area as a hydrologic factor on the glaciated Canadian Prairies. In: *General Assembly of Toronto, IAHS Publication 45*, pp. 365–376.
- Thomas, D. M. & Benson, M. A. 1970 Generalization of streamflow characteristics from drainage-basin characteristics. US Geological Survey Water Supply Paper 1975. US Govt. Printing Office, pp. 55.
- US Army Corps of Engineers 1987 *User Manual. SSARR Model. Streamflow Synthesis and Reservoir Regulation*, pp. 62, six appendices. US Army Corps of Engineers North Pacific Division, Portland, OR.
- Zhang, Y., Carey, S. K., Quinton, W. L., Janowicz, J. R., Pomeroy, J. W. & Flerchinger, G. N. 2010 [Comparison of algorithms and parameterisations for infiltration into organic-covered permafrost soils](#). *Hydrology and Earth System Sciences* **14**, 729–750.

First received 23 March 2011; accepted in revised form 13 September 2012. Available online 17 December 2012