Suggestions for revision or reasons for rejection (will be published if the paper is accepted for final publication)

This paper much improved. It is now ready for publication. If revision would be recommended, I made the following notes, which may be considered for further improving the manuscript.

RESPONSE: Thank you for your time and insights. They really improve the manuscript.

Lines 24-25: "… hydrologic model calibration for climate impact studies should be based on years that best approximate expected future conditions.": this is a strong statement. Also other types of validation may be done.

RESPONSE: This has been revised for better clarity "Hydrologic model calibration for climate impact studies could be based on years that closely approximate anticipated conditions to better constrain uncertainty in projecting extreme conditions in boreal and temperate regions". The word 'could' is now introduced to make the statement more of a suggestion for climate modelling community. See lines 22-27 for the changes

Line 63: extend sentence to: "… depends on the length of the time series used for calibration and validation"

RESPONSE: See lines 65-66 for the changes.

Line 67: it may be clarified that what is mainly tested in such split-sample approach is whether the model is not overtuned, often related to overparameterization

RESPONSE: See lines 70-71 for the changes

Lines 90-93: The sentence "Our expectation …" needs rephrasement. I suggest to rephrase sentences as follows: "Such uncertainty quantification will allow us to assess the limitations and uncertainties in hydrological model based climate change impact analysis related to the hydrological model calibration strategies and to compare these with the uncertainty related to the climate models."

RESPONSE: See lines 93-96 for the suggested changes

Section 2.2.: The old greenhouse gas scenarios (SRES based) were applied. These became outdated in the meantime; the new RCP based scenarios should be used in current climate change impact studies. However, because the focus of this paper lies on the methodology rather than on the impact results, it is acceptable to rely on the old SRES scenarios.

RESPONSE: See lines 122-127 for this new statement

Line 171: "NS .. dropped close to zero": Sure? The NS should be close to 1 (or at least above 0)

RESPONSE: See lines 177-179 for the changes

Lines 229: "The model performed better when calibrated to wet and dry conditions …": it may be clarified that this is logical because otherwise (e.g. using the NS) too much weight is given to the central part of the distribution (many more values in that part).

RESPONSE: See lines 239-240 for the changes

Section 3.5: This is an interesting approach and analysis!

RESPONSE: Thank you!

Line 292: I am not sure that "average" is the proper wording here

RESPONSE: This has been changed to 'mean'. We hope this clarifies point. See line 303 for the change

Line 331: “other metrics”: OK, but none of these focus on the (high) extremes. Another way to evaluate the model for its performance in describing extremes is the approach presented in Willems (2009) or the one by Van Steenbergen et al. (2012)

RESPONSE: See lines 344-346 for the changes

Line 344: “towards high flows”: not fully true (as discussed earlier in the paper as well): because the high flow durations are typically shorter than the low flow durations, they may receive lower weight rather than higher weight (because of the squared flow terms) in the NS computation

RESPONSE: See lines 358-360 for this insertion

Section 4.5: You may add that climate change in many places of the world leads to more extremes, both high and low flows.

RESPONSE: See line 417 for the change

Many references are rather old; it would be good to add some more recent ones. Let me suggest some.

Line 58: Other references next to Li et al., 2015, that may be added here are Breuer et al. (2009) and Vansteenkiste et al. (2014a)
RESPONSE: See lines 660-61 for the changes

Line 71: next to Butts et al., 2004, and Refsgaard et al., 2006, another more recent reference that can be added here is Vansteenkiste et al. (2014)
RESPONSE: See line 75 for the change

Line 85: next to Refsgaard et al., 2014, another recent reference is Andréassian et al. (2014).
RESPONSE: See line 88 for the change

Line 326: Another reference that may be added here is Willems (2009)
RESPONSE: See lines 336 for the change

Lines 347-348: That the uncertainty is higher in extrapolating low flows (compared to high flows) was also shown by others: Vansteenkiste et al. (2014b), Velázquez et al. (2012), Bae et al. (2011), Najafi et al. (2011), Maurer et al. (2010), …
RESPONSE: Thank you for this suggestion. See lines 260-362 for the changes

References:
Andréassian, V., F. Bourgin, L. Oudin, T. Mathevet, C. Perrin, J. Lerat, L. Coron, L. Berthet, (2014), ‘Seeking genericity in the selection of parameter sets: impact on hydrological model efficiency’, Water Resources Research, 50(10), 8356-8366.
Bae, D.-H., Jung, I.-W., Lettenmaier, D.-P. (2011), ‘Hydrologic uncertainties in climate change from IPCC AR4 GCM simulations of the Chungju Basin, Korea’. Journal of Hydrology, 401, 90–105.
Breuer L., Huisman J.A., Willems P., Bormann H., Bronstert A., Croke B.F.W., Frede H., Gräff T., Hubrechts L., Jakeman A.J., Kite G.W., Lanini J., Leavesley G., Lettenmaier D.P., Lindström G., Seibert J., Sivapalan M., Viney N.R. (2009), ‘Assessing the impact of land use change on hydrology by ensemble modelling (LUCHEM) I: Model intercomparison of current land use’, Advances in Water Resources, 32(2), 129-146.
Maurer, E.-P., Brekke, L.-D., Pruitt, T. (2010), ‘Contrasting lumped and distributed hydrology models for estimating climate change impacts on California watersheds’, J. Am. Water Resour. Assoc., 46(5), 1024-1035.
Najafi, M.-R., Moradkhani, H., Jung, I.-W. (2011), ‘Assessing the uncertainties of hydrologic model selection in climate change impact studies’. Hydrological Processes, 25(18), 2814–2826.
Van Steenbergen, N., Willems, P. (2012), ‘Method for testing the accuracy of rainfall-runoff models in predicting peak flow changes due to rainfall changes, in a climate changing context’, Journal of Hydrology, 414-415, 425-434
Vansteenkiste, Th., Tavakoli, M., Van Steenbergen, N., De Smedt, F., Batelaan, O., Pereira, F., Willems, P. (2014a), ‘Intercomparison of five lumped and distributed models for catchment runoff and extreme flow simulation’, Journal of Hydrology, 511C, 335-349
Vansteenkiste, Th., Tavakoli, M., Ntegeka, V., De Smedt, F., Batelaan, O., Pereira, F., Willems, P.

(2014b), 'Intercomparison of hydrological model structures and calibration approaches in climate scenario impact projections', Journal of Hydrology, 519, 743–755
Velázquez, J.-A., Schmid, J., Ricard, S., Muerth, M.-J., Gauvin St-Denis, B., Minville, M., Chaumont, D., Caya, D., Ludwig, R., Turcotte, R. (2012), 'An ensemble approach to assess hydrological models' contribution to uncertainties in the analysis of climate change impact on water resources'. Hydrology and Earth System Sciences, 17, 565-578.
Willems P. (2009), 'A time series tool to support the multi-criteria performance evaluation of rainfall-runoff models', Environmental Modelling & Software, 24(3), 311-321

RESPONSE: All the suggested references here have been added to the reference list in the revised manuscript. Please see lines 455-679 for the changes

Using dry and wet year hydroclimatic extremes to guide future hydrologic projections

Oni SK[1*], Futter MN[2], Ledesma JLJ[2], Teutschbein C[3], Buttle J[4] and Laudon H[1]

[1.] Department of Forest Ecology and Management, Swedish University of Agricultural Sciences, SE-901 83, Umeå, Sweden.

[2.] Department of Aquatic Sciences and Assessment, Swedish University of Agricultural Sciences, SE-750 07, Uppsala, Sweden.

[3.] Department of Earth Sciences, Uppsala University, Villavagen 16, SE-752 36, Uppsala, Sweden.

[4.] Department of Geography, Trent University, 1600 West Bank Drive, Peterborough, ON, K9J 7B8, Canada.

*Corresponding Author: stephen.oni@slu.se

Abstract

There are growing numbers of studies on climate change impacts on forest hydrology but limited attempts have been made to use current hydroclimatic variabilities to constrain projections of future climatic conditions. Here we used historical wet and dry years as a proxy for expected future extreme conditions in a boreal catchment. We showed that runoff could be underestimated by at least 35% when dry year parameterizations were used for wet year conditions. Uncertainty analysis showed that behavioural parameter sets from wet and dry years separated mainly on precipitation related parameters and to a lesser extent on parameters related to landscape processes. While uncertainties inherent in climate models (as opposed to differences in calibration or performance metrics) appeared to drive the overall uncertainty in runoff projections under dry and wet hydroclimatic conditions. Hydrologic model calibration for climate impact studies could be based on years that closely approximate anticipated conditions to better constrain uncertainty in projecting extreme conditions in boreal and temperate regions.

Keyword: Boreal forest, boreal hydrology, climate change, uncertainty assessment, hydroclimatic extremes

1 Introduction

There are growing numbers of studies on climate change impacts on watershed hydrology but these are usually based on long-time series that depict average system behaviour (Bonan, 2008; Lindner et al., 2010: Tetzlaff et al., 2013). As a result, limited attempts have been made to use extreme dry and wet conditions to assess plausible future conditions. Increasing numbers of studies are showing the importance of ensemble projections to create a matrix of possible futures, where the mean provides a statistically more reliable estimate than can be obtained from a single realization of possible future conditions (Bosshard et al., 2013; Dosio and Paruolo, 2011; Oni et al., 2014a; Raty et al., 2014). However, the predictive uncertainty of precipitation projections is still larger than that for temperature (Teutschbein and Siebert, 2012). This inherent uncertainty might further increase in the warmer future as precipitation dynamics become less consistent due to a shift in winter precipitation patterns toward rainfall dominance (Berghuijs et al., 2014; Dore, 2005).

It is unequivocally believed that climate is a first order control on watershed hydrology (Oni et al., 2015a, b; Vörösmarty et al., 2000). Although climate change is a global phenomenon (IPCC, 2007), it will likely also alter local catchment water balances (Oni et al., 2014b; Porporato et al., 2004). Prolongation of drought regimes or increasing frequency of storm events observed in different parts of the world (Dai, 2011; Trenberth, 2012) calls for greater attention on how to constrain uncertainty in predicting extreme dry and wet conditions. While the frequency of hydroclimatic extremes might be low under present day conditions (Wellen et al., 2014), there could be intensification of precipitation events globally as climate changes (Chou et al., 2013). Otherwise, preparations for the future could be undermined by our inability to properly simulate or project new conditions outside our current modelling conditions.

Models are useful tools in hydrology and runoff has become a central feature in the modelling community to assess cumulative impacts (Futter et al., 2014; Lindström et al., 2010). Hydrological modelling has benefitted immensely from the use of long term runoff series from monitoring programs to gain insights on change in fundamental system behaviour (Karlsson et al., 2013) and to aid our understanding of watershed responses to both short and long term environmental changes (Wellen et al., 2014). While conceptualization of many of these hydrologic models is based on average natural rainfall-runoff processes derived from long term series, both simple and complex models still performed well in simulating long term dynamics at the watershed scale (Breuer et al., 2009; Li et al., 2015; Vansteenkiste et al., (2014a). Growing complexity in hydrologic models has led to increasing equifinality (Beven, 2006) due to multi-dimensionality of compensatory parameter spaces. However, extensive explorations of parameter spaces in complex models have also helped to gain further insights on system behaviour beyond simple models.

Uncertainty in model predictions depends on the length of time series used for calibration and validation (Larssen et al., 2007). Despite strong arguments against the use of the term "validation" (Oreskes et al., 1994), it is still a norm in the hydrologic modelling community to calibrate to one condition and reevaluate the model on different conditions (Cao et al., 2006; Donigiang, 2002; Wilby, 2005). This has made split-sample testing a popular way of assessing the internal working process of a model in hydrologic study (Klemeš, 1986) to ensure that model is not over-tuned or over-parameterized before embarking on future projections. While modelling staged under this framework is usually based on average system conditions depicted by long term series, it may not fully reflect processes operating under very dry and wet hydroclimatic conditions. This can also be due in part to inherent structural uncertainties in models (Butts et al., 2004; Refsgaard et al., 2006, Vansteenkiste et al., 2014b) that can stem from conceptualization, scaling and connectivity of processes between the landscape mosaic patches of a watershed that the models are representing (Tetzlaff et al., 2008; Ren and Henderson-Seller, 2006). This is the case of Karlson et al. (2013) that showed increasingly large predictive uncertainty when their model was tested on over a century long record due to non-stationarity of the historical series. It is therefore inevitable that this level of uncertainty will be amplified when projected into the unknown future where, unlike at present, we have no data to confirm our findings (Refsgaard et al., 2014). However, no consensus has yet been reached regarding whether the uncertainty due to differences in hydrologic model structures and/or calibration strategies would be greater than the unresolved uncertainty inherent in climate models when projecting hydrologic conditions in boreal or temperate ecozones.

One way to constrain the uncertainty in hydroclimatic projections is to utilize historical wet and dry years as a proxy for the future conditions expected as climate changes. This is analogous to differential split-sample test previously used (Coron et al., 2012; Klemeš, 1986; Seibert, 2003; Refsgaard and Knudsen, 1996) but is less commonly used in hydrology (Andreassian et al., 2014; Refsgaard et al., 2014). Here we used hydrological and meteorological observations in dry and wet years in a long term monitored headwater catchment in northern Sweden. The objectives of this study were to: 1) utilize long term field observations in Svartberget to gain insights into hydroclimatic behaviour in dry and wet years as a proxy to future climate extremes and 2) quantify the uncertainty in our current predictive practices that is based on such long term series. Such uncertainty quantification will allow us to assess the limitations and uncertainties in hydrological model based climate change impact analysis related to the hydrological model calibration strategies and to compare these with the uncertainty related to the climate models.

2 Data and method

2.1 Study site

This modeling exercise was carried out in Svartberget (64° 16′N, 19° 46′ E), a 50 ha headwater boreal catchment within the Krycklan experimental research infrastructure in northern Sweden (Fig. 1) (Laudon et al., 2013). Modelling results presented here were based on the long-time series of precipitation, air temperature and runoff (1981-2012) from a weather and flow monitoring station at the outlet of Svartberget. Svartberget has two headwater streams, one of which drains a completely forest landscape while the other drains a headwater mire. The catchment has a long term mean annual temperature of about 1.8°C with minimum (January) and maximum (July) mean monthly temperatures of -9.5°C and 14.5°C. The catchment receives a mean annual precipitation of 610 ± 109 mm with more than 30% falling as snow (Laudon and Ottosson-Löfvenius, 2015). Snow cover usually lasts from November to May (Oni et al., 2013). The catchment has a long term mean annual runoff of 320 ± 97 mm with subsurface pathways dominating runoff delivery to streams. Spring melt represents the dominant runoff event in the catchment and lasts 4 to 6 weeks. Forest cover includes a century old Norway spruce (*Picea abies*) and Scot pine (*Pinus sylvestris*) with some deciduous birch species (*Betula spp*). *Sphagnum sp* dominates the mire landscape and riparian zones (Ledesma et al., 2016). Svartberget has gneissic bedrock overlain by compact till of about 30 m thickness to the bedrock. The catchment elevation ranges from 114-405 m above sea level and was delineated using DEM and LIDAR (Laudon et al., 2013).

2.2 Climate models

We used 15 different regional climate models (RCMs) from the ENSEMBLES project (Van der Linden and Mitchell, 2009, Table 1). All RCMs had a resolution of 25 km and were based on Special Report on Emission Scenario (SRES) A1B emission scenarios. The SRES A1B represents a balanced growth of economy and greenhouse gas emission in the future (IPCC, 2007). The old greenhouse gas scenario (SRES based) became outdated in the meantime; the new Representative Concentration Pathway (RCP) based scenarios could have been used in current climate change impact studies. However, because the focus of this paper lies on the methodology rather than on the impact results, it is acceptable to rely on old SRES scenario in line with our other recent studies in this region (Jungkvist et al., 2014; Oni et al., 2014, 2015b). Precipitation and temperature values (2061-2090) were obtained by averaging the values of the RCM grid cell with center coordinates closest to the center of the catchment and of its eight neighboring grid cells. Due to systematic biases in RCM data and the spatial disparity between RCM grid cell and small catchment like Svartberget, post processing of RCM data is required Teutschbein and Seibert, 2012; Ehret et al., 2012; Muerth et al., 2013). The distribution mapping method (Ines and Hansen, 2006; Boe et al., 2007) was used for bias-correction of the 15 RCM-simulated precipitation and air temperature series on monthly basis using data from a weather station (1981-2010) located within the Svartberget catchment. This was achieved by adjusting the theoretical cumulative distribution function (CDF) of RCM-simulated control runs (1981-2010) to match the observed CDF. The same transformation was then applied to adjust the RCM-simulated scenario runs for the future (2061-2090). As some RCMs tend to simulate a large number of days with low precipitation (e.g. drizzle) instead of dry conditions, we applied a specific precipitation threshold to prevent considerable alteration of the distribution. RCM bias corrections presented here were fully described in Jungqvist et al. (2014) and Oni et al. (2014, 2015b).

2.3 Modelling and analysis

The Precipitation, Evapotranspiraton and Runoff simulator for Solute transport (PERSiST) is a semi-distributed bucket type rainfall-runoff model with a flexibility that allows modelers to specify the routing of water following the perceptual understanding of their landscapes (Futter et al., 2014). This feature makes PERSiST a useful tool to simulate streamflow from landscape mosaic patches at a watershed scale. The model operates on a daily time scale with inputs of precipitation and air temperature. The spatial interface requires an estimate of area, land cover proportion and reach length/width of the hydrologic response units. In the PERSiST application presented here, we used three buckets to represent the hydrology of Svartberget. These include snow, upper soil and lower soil buckets. In the snow routine bucket, the model utilized a simple degree day evapotranspiration and degree day melt factor (Futter et al., 2014). Although the maximum rate of evapotranspiration could be independent of wet and dry years as used in this study, the actual rate of evapotranspiration could be influenced by the amount of water in the soil and by an evapotranspiration (ET) adjustment parameter. The latter is an exponent for limiting evapotranspiration that adjusts the rate of evapotranspiration (depending on water depth in the bucket or how much is evapotranspired). The snow threshold partitions precipitation as either rain or snow. The model also simulates canopy interception for snowfall and rainfall to the uppermost bucket. In the modelling analysis presented here, we used three buckets to generate runoff processes in Svartberget. The quick flow bucket simulates surface or direct runoff in response to the inputs of rainfall or snowfall depending on antecedent soil moisture status. The runoff generation process was partitioned between the quick flow and lower soil buckets (upper and lower) following the square matrix described in Table 2**.**

We utilized Monte Carlo analysis to explore parameter spaces using a range of parameter values listed in Table 3. The evapotranspiration adjustment parameter sets the rate at which ET can occur when the soil is no longer able to generate runoff and this was set to 1 in the upper soil box. Maximum capacity is the field capacity of the soil that determines the maximum soil water content held. The time constant specifies the rate of water drainage from a bucket and requires a value of at least 1 in PERSiST. The relative area index determines the fraction of area covered by the bucket and is also set to 1 for our simulations. Infiltration parameters in each bucket determine the rate of water movement through the soil matrix. The model is based on series of first order differential equations that are solved sequentially following the bucket order in the square matrix. More detailed information about PERSiST parameterization and equations is provided in Futter et al. (2014).

The model was calibrated against streamflow to generate present day runoff conditions. Initial manual calibration was performed on the entire time series to minimize the difference between the simulated and observed runoff based on Nash-Sutcliffe (NS) statistics. The manual calibration also helped to identify a suite of parameters ranges to be used in the Monte Carlo analysis by varying each parameter value following steps listed in Futter et al. (2014). The Monte Carlo tool works in such a way that the model was calibrated on NS-1 in line with ~~similar to~~ other works (Senatore et al., 2011; Mascaro et al., 2013), so that NS value for the overall period of simulation ~~dropped close to zero instead of~~tends toward 1~~similar to other works (Senatore et al., 2011; Mascaro et al., 2013)~~. This helped to determine the ranges to use in the subsequent Monte Carlo analysis for the wet and dry year simulations. Starting from a random point, we sampled each parameter space 500 times before jumping to the next space (depending on whether the model performance was better or worse). We specified 100 iterations during the initialization of Monte Carlo tool so that 100 ensemble of credible parameter sets could be generated. This resulted in 50,000 (500 x 100) runs. In addition to Nash-Sutcliffe statistics, the Monte Carlo tool also takes note of other metrics during sampling. The Monte Carlo tool utilizes the Metropolis-Hasting algorithm and its mode of operation was described in Futter et al. (2014).

The best parameter sets (100 in this case) were selected based on highest NS statistics from untransformed/log transformed data. The parameter sets were also analyzed for other metrics such as variance of modeled/observed series (Var), absolute volume difference (AD), root mean square error (RMSE) and coefficient of determination (R^2). These top parameter sets derived from the Monte Carlo tool are referred to as behavioural parameters henceforth. The behavioural parameters were subjected to further analyses to determine hydrologic behaviour in dry and wet years. These include the cumulative distribution function (CDF) of behavioural parameters to determine the sensitive parameters and discriminant function analysis (DFA) to determine the dominant parameter(s) that separate the hydrology of wet from dry years. Wet years were defined as hydrologic years with runoff exceeding 430 mm/yr or 40% higher than average annual runoff (1995, 2002, 2005 and 2010). Dry years were defined as hydrologic years with runoff less than 150 mm/yr or less than 50% of average annual runoff (1987, 1992, 2000 and 2001). Hydrologic year was September 1 of a year to August 31 of the following calendar year. The bias corrected future climate series from the ensemble of climate models (Table 1) were used to drive PERSiST so as to project future hydrologic conditions under long term, as well as dry and wet year conditions.

3 Results

3.1 Long term climate and hydrology series

Preliminary analysis showed that the Svartberget hydroclimate was highly variable and thus helped partition the long term series into dry and wet years as shown in Supplementary Information 1 (SI 1). As a result, dry and wet year conditions differed in terms of climate and cumulative runoff patterns. The cumulative distribution of the dry/wet year series (Fig 2a) showed that dry year precipitation (462 ± 102 mm) was only 64% of precipitation observed in wet years (716 ± 56 mm). Similar patterns were observed in runoff dynamics (Fig. 2b) where total runoff in dry years (129 ± 35 mm) was 29% of total runoff observed in wet years (449 ± 19 mm). Runoff response was 63% of total precipitation in wet years and 28% of precipitation in the dry year regime (Table 4). Mean annual temperature was 2.4 ^{o}C in wet versus 1.8 ^{o}C in dry years.

When assessed on a seasonal scale, both precipitation and runoff were higher in almost all months in wet compared to dry year conditions (Fig. 3) but differed in terms of seasonal patterns. While runoff peaked in May in both wet and dry years reflecting spring snowmelt dynamics that characterize Svartberget, runoff magnitude differed. Peak precipitation events occurred in summer months with additional autumn peaks in wet year. However, there was a shift in precipitation patterns with lowest precipitation in February/March in dry years compared to April in wet years. Winter months were generally slightly warmer during wet years and summers slightly warmer in dry years (Fig 3c).

3.2 Future climate projections

There was less agreement between the observed series and uncorrected individual RCMs (SI 2a, b). However, bias correction helped to reduce the uncertainty on the historical time scale by providing a better match for the ensemble mean of the air temperature and precipitation with their corresponding observed series (SI 2c, d). The ensemble mean performed better in fitting observed air temperature than precipitation. There is also a possible increase in air temperature by 2.8-5^{o}C (median of 3.7^{o}C) and possible increase in precipitation by 2-27% (median of 17%). Although precipitation and temperature were projected to increase throughout the year, the temperature changes would be more pronounced during winter months irrespective of whether it was a dry or wet year (Fig. 3c). However, projected changes in precipitation followed similar patterns to historical wet years with more precipitation expected between late winter months through spring (Fig. 3a). Result also showed that the winter period with temperature below 0°C could be shortened as climate warms in the future (SI 2).

3.3 Model calibrations and performance statistics

Model behavioural performance followed similar patterns when metrics such as R^2, NS and log NS were used (SI 3a-c) and metrics could be used interchangeably to measure model performances. The model performed better when calibrated to wet and dry conditions (compared to long term) using NS metrics (SI 3b, c). It may be clarified that this is logical because otherwise (using the NS) too much weight is given to the central part of the distribution (due to many more values in that part). Although no major improvements to model efficiency above NS of 0.79 and 0.81 were obtained in dry and wet years, respectively, we obtained a wider range of model performances in wet relative to dry year. The patterns of other performance metrics were different as we observed the highest RMSE in dry years and lowest RMSE in wet year condition (SI 3d). There was minimum AD range in the long term record and maximum range in dry years (SI 3e). Model performances based on the Var metric also showed the largest variability in dry years compared to the long term record and least Var in the wet year (SI 3f).

3.4 Runoff simulations and behavioural prediction range

Using the best performing parameter sets based on the NS statistic as an example, the model performed well in simulating interannual runoff patterns but underestimated the peaks (SI 4). When resolved to their respective dry and wet year components, the model performed better in simulating runoff conditions in wet years despite its larger data spread and higher spring peaks than the dry year regime (SI 5). When parameterization for dry years was used for runoff prediction in wet years, runoff was underestimated by 35% due to significant uncertainty that stemmed from the growing season months (Fig. 4). Modelling analysis also showed that no single metric can be an effective measure of model performance under dry and wet year conditions (Fig 5a- c). However, utilizing a behavioural mean of these different performance metrics (Fig. 5d-f) appeared to be a more effective way of calibrating to extremely dry and wet hydroclimatic conditions. While the behavioural mean performed better in simulating runoff dynamics in winter through spring in the long term record and significantly reduced the uncertainty in dry and wet years, larger uncertainty existed in summer through autumn months in dry and wet years compared to the long term record.

3.5 Parameter uncertainty assessments

While we observed a wide prediction range from behavioural parameter sets (Fig. 5), we have limited information on the underlining processes. Therefore, we subjected the behavioural parameter sets to further analysis to identify sensitive parameters and plausible patterns of hydrologic processes that differentiate dry and wet years (Fig. 6). The cumulative distribution function (CDF) of behavioural parameter sets showed that both rain and flow multipliers were sensitive parameters in dry years. The rain multiplier was less sensitive in wet years unlike the flow multiplier. Long term simulations showed no sensitivity to the rain multiplier but were sensitive to the flow multiplier. We observed similar patterns of response to the flow multiplier in all three hydrologic regimes (Fig. 6b). Result also pointed to the sensitivity of interception in wet years but all the three hydrologic regimes showed similar patterns for the time constant (water residence time) in lower soil.

We subjected the pool of behavioural parameters in dry and wet year regimes to discriminant function analysis (DFA) to identify the key parameters that separate the extreme hydroclimatic conditions (Fig. 7). Results showed that both dry and wet years separated well in canonical space. However, the separation was driven mainly on quantitative parameters related to precipitation, interception and evapotranspiration on canonical axis 1 (Rmult, Int and DDE). The parameters separated to a lesser extent on processes related to snow parameters on canonical axis 2 (Smult, SM and DDM).

3.6 Quantification of uncertainty in hydrologic projections

We compared the effects of different performance metrics in wet and dry year regimes to constrain uncertainty in runoff projections under future hydroclimatic extremes in Svartberget catchment (SI 6). Results showed that differences in model representation of present day conditions might be minimal (compared to the observed) but a wide range of runoff regimes were projected in the future. We also observed small difference in the range of runoff projections (derived from minimum and maximum of behavioural parameter sets) using different model performance metrics. Uncertainties inherent in climate models (as opposed to differences in calibration or performance metrics) appeared to drive the overall uncertainty in runoff projections under dry and wet hydroclimatic conditions**.** Wet year is the closest to plausible projections of future condition expected in the boreal ecozone. However, model results suggested that the uncertainty in present day long term simulations is mostly driven by dry years. We compared the runoff predictions using dry year parameterization to parameterization based on wet years to quantify our current predictive uncertainty. Results showed that future runoff could be under predicted by up to 40% (relative to wet year ensemble mean) if the projections are based on dry year parameterization alone (Fig. 8). Both parameterizations projected a shift in spring melt from May to April in the future. However, ensemble projections showed that summer months could be a lot wetter (based on wet year parameterization compared to dry year) and wet year spring peak could be up to 43% more compared to projections based on the wet year ensemble mean.

4 Discussion

4.1 Insights from long term hydroclimatic series

Several studies have evaluated the impact of climate change on surface water resources (Berghuijs et al., 2014; Chou et al., 2013; Dore, 2005 among the others) but most of these were based on long term series that depict mean system behaviour. However, present day hydroclimatic extremes, such as those derived from historical wet and dry years, can be used as simple proxies to gain insights that will aid our understanding of future hydroclimatic conditions. Using this approach we found that standard calibrations can result in underestimation of runoff by up to 35% due to high variability of hydroclimate series in northern boreal catchments. Several explanations can be offered for the high variability in the long term hydroclimate series at the study site. First, snowmelt hydrology is important in understanding the boreal water balances due to their location in the northern hemisphere (Euskirchen et al., 2007; Dore, 2005; Tetzlaff et al., 2011, 2013). As a result, northern headwater catchments tend to show high variability (Brown and Robinson, 2011; Burn, 2008).

We observed annual runoff yield to be 63% of total precipitation in the wet years compared to 28% of total precipitation in dry year. More runoff yield in the wet year regime could be seen as a result of near field capacity of the soils throughout the year, leading to greater propensity for runoff generation because hydrological conductivity increases towards soil surface in the catchment (Nyberg et al., 2001). This can also imply more winter snow accumulation during the long winter period, resulting in higher spring melt that drives the overall water fluxes (Laudon et al., 2004). Less runoff yield in dry years could be attributed to higher soil moisture deficit and relatively more important evapotranspiration rates (Dai, 2013).

We also observed differences in dry/wet year peak summer precipitation and a shift in the lowest precipitation in late winter/early spring. Despite the differences in precipitation, we observed similar patterns of runoff responses that only differ in terms of magnitude. This suggested that there was more effective rainfall (net available water) available to infiltrate, continuously recharge groundwater systems and generate runoff from upstream sources in wet year. Slightly warmer temperatures in summer months could drive more of growing season evapotranspiration in dry year. Small differences in temperature regime between wet and dry year, unlike precipitation, also explained why larger uncertainty and biases still exist during post-processing of precipitation series in using any scenario-based GCMs as observed in SI 2.

4.2 Multi-criteria calibration of hydrological models

There has been considerable discussion about the calibrating procedure in the hydrological modelling community (Andreassian et al., 2012; Boij and Krol, 2010; Efstratiadis and Koutyiannis, 2010; Oreskes et al., 1994; Price et al., 2012). One of the key reasons for this is the difference in goodness-of-fit measures utilized in each model (Krause et al., 2005; Pushpathala et al., 2012). The most common strategy is to calibrate hydrologic models using the Nash-Sutcliffe (NS) statistic (Nash and Sutcliffe, 1970). However, many modelers believe that the NS-based method alone tends to underestimate variance in modelled time series as this metric could be biased toward high or low flow periods (Futter et al., 2014; Jain and Sudheer, 2008; Pushpalatha et al., 2012; Willens, 2009). This is promoting our use of multi-criteria statistics in model calibrations to constrain predictive uncertainty in hydrologic projections to extreme dry and wet hydroclimatic conditions. Therefore, multi-criteria calibration objectives that assessed model performances using different goodness-of-fit metrics could aid our understanding of hydrologic behaviour in boreal catchments. Our observation of differences in model performances in terms of NS and other metrics presented here is expected as a three box model proposed by Seibert and McDonnell (2002) similarly showed good fit for NS but poor fit using other metrics. However none of these focus on the extremes. Another way to evaluate model for its performance in describing extremes is the approach presented in Willems (2009) or the one by Van Steenberger and Willems, (2012). However, lower model performance (based on NS) for the long term record is explainable as most hydrologic models are based on mean system behaviour represented by long term rainfall-runoff processes (Futter et al., 2014; Oni et al., 2014b; Wellen et al., 2014).

The lower range of model performances in calibrating to the observed runoff in dry years is an indication of variable runoff generation processes associated with this wetness regime. Dry years cause drought-like conditions (Dai, 2011; Mishra and Singh, 2010) as a result of less water availability that reduces hydrologic connectivity within the catchment. However, the model performed better when applied to wet and dry years individually compared to the long term record based on NS statistics. This suggested that the mechanisms driving hydrologic processes in dry and wet years might be similar but their relative magnitude differs from long term average conditions (Grayson et al., 1997). Better performance under dry ~~and wet~~ conditions (compared to average long term) can also be attributed to the bias of NS ~~and log NS~~ towards ~~high flows and~~ baseflow~~, respectively~~ (Futter et al., 2014; Jain and Sudheer, 2008; Pushpalatha et al., 2012). Durations of high flows associated with wet years are typically shorter than the low flow durations; as a result, higher flows receive lower weight because of the squared flow terms in the NS computation. Therefore the uncertainty is higher in extrapolating low flows (compared to high flows) and was also shown by others (Bae et al., 2011; Najarafi et al., 2011; Maurer et al., 2010; Vansteenkiste et al., 2014b; Velazquez et al., 2013).

However, NS statistics alone are not enough to assess model performances in climate-sensitive boreal headwater streams such as Svartberget. Other metrics such as the RMSE showed that dry years could be a major driver of the uncertainty we observed in simulating the long term record. A possible explanation could be that the soil moisture deficit is larger in dry year, leading to soil matrix or vertical flow (Grayson et al., 1997) that can only generate runoff after filling soil pore spaces (McDonnell, 1990). For example, soil pore spaces are usually not close to saturation under dry condition due to 1) intermittent precipitation events throughout the year and 2) several patchy source areas of high water convergence that are characterized by local landscape terrain or soil properties (Fang and Pomeroy, 2008; Jencso et al., 2009). Also higher rates of evapotranspiration coupled with low precipitation can contribute to more spatially decoupled antecedent soil moisture conditions and thus lower runoff in dry years (Dai, 2013; Vicente-Serrano et al., 2010). Therefore, no single model performance metric can be effective in simulating the hydrology of dry and wet year conditions, as our results showed that the mean of behavioural metrics outperformed any individual metric in dry and wet years under present day conditions.

4.3 Parameter sensitivity in dry and wet year regimes

The robust uncertainty assessment conducted here showed that extensive exploration of model parameter spaces suggests how hydrologic behaviour differs between wet and dry year regimes. A possible explanation for the non-sensitivity of the rain multiplier in wet years could be attributed to 1) a more consistent or stable precipitation feeding the system throughout the year compared to intermittent precipitation in dry years (Fang and Pomeroy, 2008; McNamara et al., 2005) or 2) the effect of rain water collector missing proportionally more rain in dry than wet years. This can explain the smaller spring peak that characterizes the dry year regime or its non-sensitivity to interception unlike its role in wet year regimes.

We observed that sensitivity of the lower soil time constant followed similar patterns in dry and wet years unlike the upper soil box. Therefore, we could expect faster flow and higher runoff ratio in the wet years due to rapid response to precipitation events and more macropore flow (Peralta-Tapia et al., 2015). This can lead to steady runoff generation due to 1) near saturation of soils and 2) greater connectivity between stream channels and upland areas (Bracken et al., 2013; Ocampo et al., 2006) that become disconnected in dry years. The patterns of the flow multiplier parameter showed that both dry and wet year conditions followed similar runoff generation processes. These suggested that the main physical mechanisms to explain parameter sensitivity and hydroclimatic behaviour to dry/wet conditions were related to differences in their precipitation patterns rather than landscape-driven hydrologic processes.

4.4 Drivers of hydrologic behaviour in dry and wet year regimes

Even though equifinality limits the use of CDFs alone in identifying all sensitive parameters, DFA of behavioural parameters gave further holistic insights into plausible differences in wet/dry hydrologic behaviour when projected on canonical space. This suggested that hydrological model parameterizations calibrated to high flow associated with wet years differ from parameterizations for long term or dry conditions. Therefore, parameter separation primarily on quantitative parameters (Rmult, Int and DDE) related to rainfall and evapotranspiration on canonical axis 1 suggested that climate is still a first order control of dry and wet year hydroclimatic regimes in the boreal forest. This is consistent with Wellen et al. (2014), who showed that extreme conditions could be triggered in a watershed when precipitation reaches a threshold that can initiate saturation overland flow. This is because soils are always near saturation capacity under prolonged wet conditions (Grayson et al., 1997). This can explain the increase in hydrologic model uncertainty in capturing the peak runoff events in wet years unless parameter ranges that combined different performance metrics are considered. Unfortunately, we might face a new challenge of increased precipitation ranges in the future as climate changes (Chou et al., 2013; Dore, 2005). The separations of wet and dry years on snow process-related parameters (Smult, SM and DDM) to a lesser extent on canonical axis 2 suggested that indirect landscape influences on snow processes could be important but are a second order control on runoff response to dry and wet conditions. This agrees with Jencso et al. (2009), who showed that landscape mosaic structures with their unique source contribution areas control the overall watershed response.

4.5 Implications for future climate projections

Climate change in many places of the world leads to more extremes, both high and low flows. This study is not an exception as all 15 RCMs considered here~~in this study~~ projected a range of plausible futures in the Swedish boreal forest. Irrespective of the model performance metrics, results suggested that the future could be substantially wetter and could make drought conditions less severe in boreal ecozones. This could explain the large uncertainty in projecting runoff under wet conditions. For example, dry year and long term parameterizations were similar and runoff was under-predicted by 35% under the present day condition when parameterization in dry years was used for wet years. This was due to large predictive uncertainty in runoff dynamics (Fig. 4) that resulted from high evapotranspiration rates during the snow free growing seasons in dry year. This suggests that wet year calibration could give more credible projections of the future in the boreal ecozone as the distribution of precipitation in wet years is closer to the precipitation pattern expected in the future. While our modelling results suggested negligible differences in runoff projections based on either dry year or long term parameterization, wetter conditions could become a more dominant feature in the boreal ecozone.

These have implications for future climate change as both dry and wet year parametrization showed a consistent shift in spring melt patterns from May to April (Fig. 8). This temporal advance in spring melt patterns could result from altered distribution of snowfall and rainfall patterns in the winter (Berghuijs et al., 2014; Dore, 2005), and may likely have effects on soil frost in the upper layer (Jungkvist et al., 2014) or change in evapotranspiration rates (Jung et al., 2010; Vicente-Serrano et al., 2010). Therefore, intensification of hydroclimatic regimes as climate changes in the future (Kunkel et al., 2013) could drive water quality issues to a new level in the boreal forest due to changes in the flux of organic carbon and aquatic pollutants. Furthermore, precipitation has been shown to have much larger biogeochemical implications for the boreal carbon balance than previously anticipated (Öquist et al., 2014).

The large spread of mean annual runoff projected by each RCM in wet years is an indication of less agreement between RCMs when predicting future conditions. This suggested that inherent uncertainty in climate models, rather than differences in model calibrations, drive the overall uncertainty in runoff projections. However, hydrologic model calibration for climate impact studies should be based on years that closely approximate anticipated conditions to better constrain uncertainty in projecting extremely dry and wet conditions in boreal and temperate regions.

Acknowledgement

This project was funded by two larger projects ForWater and Future Forest, studying the effect of climate and forest management on boreal water resources. Funding for KCS came from the Swedish Science Council, Formas, SKB, MISTRA and Kempe Foundation. The ENSEMBLES data used in this work were funded by the EU FP6 Integrated Project ENSEMBLES (Contract number 505539) whose support is gratefully acknowledged. We also thank Prof. Patrick Willems of KU Leuven, Belgium and an anonymous reviewer for their insightful comments that greatly improved the manuscript.

References

Andréassian, V., Le Moine, N., Perrin, C., Ramos, M. H., Oudin, L., Mathevet, T., Lerat, J., and Berthet, L.: All that glitters is not gold: the case of calibrating hydrological models, Hydrological Processes, 26, 2206-2210, 2012.

Andréassian, V., Bourgin, F., Oudin, L., Mathevet, T., Perrin, C., Lerat, J., Coron, L. and Berthet, L.: Seeking genericity in the selection of parameter sets: Impact on hydrological model efficiency, Water Resources Research, 50(10), 8356-8366, 2014.

Bae, D.H., Jung, I.W. and Lettenmaier, D.P.: Hydrologic uncertainties in climate change from IPCC AR4 GCM simulations of the Chungju Basin, Korea, Journal of Hydrology, 401(1), 90-105, 2011.

Berghuijs, W., Woods, R., and Hrachowitz, M.: A precipitation shift from snow towards rain leads to a decrease in streamflow, Nature Climate Change, 4, 583-586, 2014.

Beven, K.: A manifesto for the equifinality thesis, Journal of hydrology, 320, 18-36, 2006.

Boe, J., Terray, L., Habets, F. and Martin, E.: Statistical and dynamical downscaling of the Seine basin climate for hydro-meteorological studies, Int J Climatol, 27(12), 1643–1655, doi:10.1002/joc.1602, 2007.

Bonan, G. B.: Forests and climate change: forcings, feedbacks, and the climate benefits of forests, Science, 320, 1444-1449, 2008.

Booij, M. J., and Krol, M. S.: Balance between calibration objectives in a conceptual hydrological model, Hydrological Sciences Journal, 55, 1017-1032, 2010.

Bosshard, T., Carambia, M., Goergen, K., Kotlarski, S., Krahe, P., Zappa, M., and Schär, C.: Quantifying uncertainty sources in an ensemble of hydrological climate-impact projections, Water Resources Research, 49, 1523-1536, 2013.

Breuer, L., Huisman, J.A., Willems, P., Bormann, H., Bronstert, A., Croke, B.F.W., Frede, H.G., Gräff, T., Hubrechts, L., Jakeman, A.J. and Kite, G.: Assessing the impact of land use change on hydrology by ensemble modeling (LUCHEM). I: Model intercomparison with current land use. Advances in Water Resources, 32(2), 129-146, 2009.

Bracken, L., Wainwright, J., Ali, G., Tetzlaff, D., Smith, M., Reaney, S., and Roy, A.: Concepts of hydrological connectivity: Research approaches, pathways and future agendas, Earth-Science Reviews, 119, 17-34, 2013.

Brown, R., and Robinson, D.: Northern Hemisphere spring snow cover variability and change over 1922–2010 including an assessment of uncertainty, The Cryosphere, 5, 219-229, 2011.Burn, D. H.: Climatic influences on streamflow timing in the headwaters of the Mackenzie River Basin, Journal of Hydrology, 352, 225-238, 2008.

Butts, M.B., Payne, J.T., Kristensen, M. and Madsen, H.: An evaluation of the impact of model structure on hydrological modelling uncertainty for streamflow simulation. *Journal of Hydrology*, *298*(1), pp.242-266, 2004.

Cao, W., Bowden, W.B., Davie, T. and Fenemor, A.: 2006. Multi-variable and multi-site calibration and validation of SWAT in a large mountainous catchment with high spatial variability. *Hydrological Processes*, *20*(5), 1057-1073, 2006.

Chou, C., Chiang, J. C., Lan, C.-W., Chung, C.-H., Liao, Y.-C., and Lee, C.-J.: Increase in the range between wet and dry season precipitation, Nature Geoscience, 6, 263-267, 2013.Coron, L., Andreassian, V., Perrin, C., Lerat, J., Vaze, J., Bourqui, M. and Hendrickx, F.: Crash testing hydrological models in contrasted climate conditions: An experiment on 216 Australian catchments. *Water Resources Research*, *48*(5), 2012.

Dai, A.: Drought under global warming: a review, Wiley Interdisciplinary Reviews: Climate Change, 2, 45-65, 2011.

Dai, A.: Increasing drought under global warming in observations and models, Nature Climate Change, 3, 52-58, 2013.

Dore, M. H.: Climate change and changes in global precipitation patterns: what do we know?, Environment International, 31, 1167-1181, 2005.

Donigian, A.S.: Watershed model calibration and validation: The HSPF experience. *Proceedings of the Water Environment Federation*, *2002*(8), 44-73, 2002.

Dosio, A., and Paruolo, P.: Bias correction of the ENSEMBLES high-resolution climate change projections for use by impact models: Evaluation on the present climate, Journal of Geophysical Research: Atmospheres (1984–2012), 116, 2011.

Efstratiadis, A., and Koutsoyiannis, D.: One decade of multi-objective calibration approaches in hydrological modelling: a review, Hydrological Sciences Journal, 55, 58-78, 2010.

Ehret, U., Zehe, E., Wulfmeyer, V., Warrach-Sagi, K. and Liebert, J.: HESS Opinions "Should we apply bias correction to global and regional climate model data?," Hydrol Earth Syst Sci, 16(9), 3391–3404, doi:10.5194/hess-16-3391-2012, 2012.

Euskirchen, E., McGuire, A., and Chapin, F. S.: Energy feedbacks of northern high-latitude ecosystems to the climate system due to reduced snow cover during 20th century warming, Global Change Biology, 13, 2425-2438, 2007.

Fang, X., and Pomeroy, J. W.: Drought impacts on Canadian prairie wetland snow hydrology, Hydrological Processes, 22, 2858-2873, 2008.

Futter, M., Erlandsson, M., Butterfield, D., Whitehead, P., Oni, S., and Wade, A.: PERSiST: a flexible rainfall-runoff modelling toolkit for use with the INCA family of models, Hydrology and Earth System Sciences 10, 8635-8681, 2014.

Grayson, R. B., Western, A. W., Chiew, F. H., and Blöschl, G.: Preferred states in spatial soil moisture patterns: Local and nonlocal controls, Water Resources Research, 33, 2897-2908, 1997.

Ines, A. V. M. and Hansen, J. W.: Bias correction of daily GCM rainfall for crop simulation studies, Agr Forest Meteorol, 138(1-4), 44–53, doi:10.1016/j.agrformet.2006.03.009, 2006.

IPCC: The physical science basis. contribution of working group I to the fourth assessment report of the intergovernmental panel on climate change, in: Climate Change 2007: The Physical Science Basis, edited by: Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor and H.L. Miller Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 996, 2007.

Jain, S. K., and Sudheer, K.: Fitting of hydrologic models: a close look at the Nash–Sutcliffe index, Journal of Hydrologic Engineering, 13, 981-986, 2008.

Jencso, K. G., McGlynn, B. L., Gooseff, M. N., Wondzell, S. M., Bencala, K. E., and Marshall, L. A.: Hydrologic connectivity between landscapes and streams: Transferring reach-and plot-scale understanding to the catchment scale, Water Resources Research, 45, 2009.

Jung, M., Reichstein, M., Ciais, P., Seneviratne, S. I., Sheffield, J., Goulden, M. L., Bonan, G., Cescatti, A., Chen, J., and De Jeu, R.: Recent decline in the global land evapotranspiration trend due to limited moisture supply, Nature, 467, 951-954, 2010.

Jungqvist, G., Oni, S. K., Teutschbein, C., and Futter, M. N.: Effect of climate change on soil temperature in Swedish boreal forests, PLoS ONE. doi, 10, 1371, 2014.

Karlsson, I.B., Sonnenborg, T.O., Jensen, K.H. and Refsgaard, J.C.: Evaluating the influence of long term historical climate change on catchment hydrology–using drought and flood indices. *Hydrol Earth Syst Sci Discuss*, *10*, 2373-2428, 2013.

Klemeš, V.: Operational testing of hydrological simulation models. *Hydrological Sciences Journal*, *31*(1), 13-24, 1986.

Krause, P., Boyle, D., and Bäse, F.: Comparison of different efficiency criteria for hydrological model assessment, Advances in Geosciences, 5, 89-97, 2005.

Kunkel, K. E., Karl, T. R., Easterling, D. R., Redmond, K., Young, J., Yin, X., and Hennon, P.: Probable maximum precipitation and climate change, Geophysical Research Letters, 40, 1402-1408, 2013.

Larssen, T., Høgåsen, T. and Cosby, B.J.: Impact of time series data on calibration and prediction uncertainty for a deterministic hydrogeochemical model. *Ecological Modelling*, *207*(1), 22-33, 2007.

Laudon, H., Seibert, J., Köhler, S., and Bishop, K.: Hydrological flow paths during snowmelt: Congruence between hydrometric measurements and oxygen 18 in meltwater, soil water, and runoff, Water Resources Research, 40, 2004.

Laudon, H., Taberman, I., Ågren, A., Futter, M., Ottosson-Löfvenius, M., and Bishop, K.: The Krycklan Catchment Study—a flagship infrastructure for hydrology, biogeochemistry, and climate research in the boreal landscape, Water Resources Research, 49, 7154-7158, 2013.

Laudon, H., and Ottosson Löfvenius, M.: Adding snow to the picture–providing complementary winter precipitation data to the Krycklan catchment study database, Hydrological Processes, Doi: 10.1002/hyp.10753, 2015.

Ledesma, J. L. J., Futter, M. N., Laudon, H., Evans, C. D., and Köhler, S. J: Boreal forest riparian zones regulate stream sulfate and dissolved organic carbon, Science of the Total Environment, 560-561, 110-122, doi: 10.1016/j.scitotenv.2016.03.230, 2016.

Li, H., Xu, C.-Y., and Beldring, S.: How much can we gain with increasing model complexity with the same model concepts?, Journal of Hydrology, 527, 858-871, 2015.

Lindner, M., Maroschek, M., Netherer, S., Kremer, A., Barbati, A., Garcia-Gonzalo, J., Seidl, R., Delzon, S., Corona, P., and Kolström, M.: Climate change impacts, adaptive capacity, and vulnerability of European forest ecosystems, Forest Ecology and Management, 259, 698-709, 2010.

Lindstrom, G., Pers, C., Rosberg, J., Stromqvist, J., and Arheimer, B.: Development and testing of the HYPE (Hydrological Predictions for the Environment) water quality model for different spatial scales, Hydrology Research, 41, 295-319, 2010.

Mascaro, G., Piras, M., Deidda, R. and Vivoni, E.R.: Distributed hydrologic modeling of a sparsely monitored basin in Sardinia, Italy, through hydrometeorological downscaling. *Hydrology and Earth System Sciences*, *17*(10), 4143-4158, 2013.

Maurer, E.P., Brekke, L.D. and Pruitt, T.: Contrasting lumped and distributed hydrology models for estimating climate change impacts on California Watersheds, Journal of American Water Resources Association, 4685), 1024-1035, 2010

McDonnell, J. J.: A rationale for old water discharge through macropores in a steep, humid catchment, Water Resour. Res, 26, 2821-2832, 1990.

McNamara, J. P., Chandler, D., Seyfried, M., and Achet, S.: Soil moisture states, lateral flow, and streamflow generation in a semi-arid, snowmelt-driven catchment, Hydrological Processes, 19, 4023-4038, 2005.

Mishra, A. K., and Singh, V. P.: A review of drought concepts, Journal of Hydrology, 391, 202-216, 2010.

Muerth, M. J., Gauvin St-Denis, B., Ricard, S., VelÃ¡zquez, J. A., Schmid, J., Minville, M., Caya, D., Chaumont, D., Ludwig, R. and Turcotte, R.: On the need for bias correction in regional climate scenarios to assess climate change impacts on river runoff, Hydrol Earth Syst Sci, 17(3), 1189–1204, doi:10.5194/hess-17-1189-2013, 2013.

Najafi, M.R., Moradkhani, H. and Jung, I.W.: Assessing the uncertainties of hydrologic model selection in climate change impact studies, Hydrological Processes, 25(18), 2814-2826, 2011.

Nash, J. E., and Sutcliffe, J.: River flow forecasting through conceptual models part I—A discussion of principles, Journal of hydrology, 10, 282-290, 1970.

Nyberg, L., Stähli, M., Mellander, P. E., Bishop, K. H.: Soil frost effects on soil water and runoff dynamics along a boreal forest transect: 1. Field investigations, Hydrological Processes 15, 909-926, 2001.

Ocampo, C. J., Sivapalan, M., and Oldham, C.: Hydrological connectivity of upland-riparian zones in agricultural catchments: Implications for runoff generation and nitrate transport, Journal of Hydrology, 331, 643-658, 2006.

Oni, S., Futter, M., Bishop, K., Köhler, S., Ottosson-Löfvenius, M., and Laudon, H.: Long-term patterns in dissolved organic carbon, major elements and trace metals in boreal headwater catchments: trends, mechanisms and heterogeneity, Biogeosciences, 10, 2315-2330, 2013.

Oni, S., Futter, M., Teutschbein, C., and Laudon, H.: Cross-scale ensemble projections of dissolved organic carbon dynamics in boreal forest streams, Climate Dynamics 42, 2305-2321, 10.1007/s00382-014-2124-6, 2014a.

Oni, S., Futter, M., Molot, L., Dillon, P., and Crossman, J.: Uncertainty assessments and hydrological implications of climate change in two adjacent agricultural catchments of a rapidly urbanizing watershed, Science of the Total Environment, 473, 326-337, 2014b.

Oni, S. K., Futter, M. N., Buttle, J., and Dillon, P. J.: Hydrological footprints of urban developments in the Lake Simcoe watershed, Canada: a combined paired-catchment and change detection modelling approach, Hydrological Processes, 29, 1829-1843, 2015a.

Oni, S.K., Tiwari, T., Ledesma, J.L., Ågren, A.M., Teutschbein, C., Schelker, J., Laudon, H. and Futter, M.N.: Local-and landscape-scale impacts of clear-cuts and climate change on surface water dissolved organic carbon in boreal forests. *Journal of Geophysical Research: Biogeosciences*, *120*(11), pp.2402-2426, 2015b.

Oreskes, N., Shrader-Frechette, K. and Belitz, K., 1994. Verification, validation, and confirmation of numerical models in the earth sciences. *Science*, *263*(5147), pp.641-646.

Öquist, M., Bishop, K., Grelle, A., Klemedtsson, L., Köhler, S., Laudon, H., Lindroth, A., Ottosson Löfvenius, M., Wallin, M. B., and Nilsson, M. B.: The full annual carbon balance of boreal forests is highly sensitive to precipitation, Environmental Science & Technology Letters, 1, 315-319, 2014.

Peralta-Tapia, A., Sponseller, R. A., Tetzlaff, D., Soulsby, C., and Laudon, H.: Connecting precipitation inputs and soil flow pathways to stream water in contrasting boreal catchments, Hydrological Processes, 29, 3546-3555, 2015.

Porporato, A., Daly, E., and Rodriguez-Iturbe, I.: Soil water balance and ecosystem response to climate change, The American Naturalist, 164, 625-632, 2004.

Price, K., Purucker, S. T., Kraemer, S. R., and Babendreier, J. E.: Tradeoffs among watershed model calibration targets for parameter estimation, Water Resources Research, 48, 2012.

Pushpalatha, R., Perrin, C., Le Moine, N., and Andréassian, V.: A review of efficiency criteria suitable for evaluating low-flow simulations, Journal of Hydrology, 420, 171-182, 2012.

Räty, O., Räisänen, J., and Ylhäisi, J. S.: Evaluation of delta change and bias correction methods for future daily precipitation: intermodel cross-validation using ENSEMBLES simulations, Climate dynamics, 42, 2287-2303, 2014.Refsgaard, J.C. and Knudsen, J.: Operational validation and intercomparison of different types of hydrological models. *Water Resources Research*, *32*(7), 2189-2202, 1996.

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

Refsgaard, J.C., Van der Sluijs, J.P., Brown, J. and Van der Keur, P.: A framework for dealing with uncertainty due to model structure error. *Advances in Water Resources*, *29*(11), pp.1586-1597, 2006.

Refsgaard, J.C., Madsen, H., Andréassian, V., Arnbjerg-Nielsen, K., Davidson, T.A., Drews, M., Hamilton, D.P., Jeppesen, E., Kjellström, E., Olesen, J.E. and Sonnenborg, T.O.: A framework for testing the ability of models to project climate change and its impacts. *Climatic change*, *122*(1-2), 271-282, 2014.

Ren, D., and Henderson-Sellers, A.: An analytical hydrological model for the study of scaling issues in land surface modeling, Earth Interactions, 10, 1-24, 2006.

Seibert, J., and McDonnell, J. J.: On the dialog between experimentalist and modeler in catchment hydrology: Use of soft data for multicriteria model calibration, Water Resources Research, 38, 23-21-23-14, 2002.Seibert, J.: Reliability of model predictions outside calibration conditions. *Hydrology Research*, *34*(5), 477-492, 2003.

Senatore, A., Mendicino, G., Smiatek, G. and Kunstmann, H.: Regional climate change projections and hydrological impact analysis for a Mediterranean basin in Southern Italy. *Journal of Hydrology*, *399*(1), 70-92, 2011.

Tetzlaff, D., McDonnell, J., Uhlenbrook, S., McGuire, K., Bogaart, P., Naef, F., Baird, A., Dunn, S., and Soulsby, C.: Conceptualizing catchment processes: simply too complex?, Hydrological Processes, 22, 1727, 2008.

Tetzlaff, D., Soulsby, C., Hrachowitz, M., and Speed, M.: Relative influence of upland and lowland headwaters on the isotope hydrology and transit times of larger catchments, Journal of Hydrology, 400, 438-447, 2011.

Tetzlaff, D., Soulsby, C., Buttle, J., Capell, R., Carey, S., Laudon, H., McDonnell, J., McGuire, K., Seibert, S., and Shanley, J.: Catchments on the cusp? Structural and functional change in northern ecohydrology, Hydrological Processes, 27, 766-774, 10.1002/hyp.9700, 2013.

Teutschbein, C., and Seibert, J.: Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods, Journal of Hydrology, 456-457, 12-29, 2012.

Trenberth, K. E.: Framing the way to relate climate extremes to climate change, Climatic Change, 115, 283-290, 2012.

Van der Linden, P., and Mitchell, J. F. B.: ENSEMBLE: Climate change and its impacts: Summary of research and results from the ENSEMBLES project: http://ensembles-eu.metoffice.com/docs/Ensembles_final_report_Nov09.pdf, 2009.

Van Steenbergen, N. and Willems, P.: Method for testing the accuracy of rainfall–runoff models in predicting peak flow changes due to rainfall changes, in a climate changing context, Journal of Hydrology, 414, 425-434, 2012.

Vansteenkiste, T., Tavakoli, M., Van Steenbergen, N., De Smedt, F., Batelaan, O., Pereira, F. and Willems, P.: Intercomparison of five lumped and distributed models for catchment runoff and extreme flow simulation, Journal of Hydrology, 511, 335-349, 2014a.

Vansteenkiste, T., Tavakoli, M., Ntegeka, V., De Smedt, F., Batelaan, O., Pereira, F. and Willems, P.: Intercomparison of hydrological model structures and calibration approaches in climate scenario impact projections, Journal of Hydrology, 519, 743-755, 2014b.

Velázquez, J.A., Schmid, J., Ricard, S., Muerth, M.J., Gauvin St-Denis, B., Minville, M., Chaumont, D., Caya, D., Ludwig, R. and Turcotte, R.: An ensemble approach to assess hydrological models' contribution to uncertainties in the analysis of climate change impact on water resources. *Hydrology and Earth System Sciences*, *17*(2), 565-578, 2013.

Vicente-Serrano, S. M., Beguería, S., and López-Moreno, J. I.: A multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index, Journal of Climate, 23, 1696-1718, 2010.

Vörösmarty, C. J., Green, P., Salisbury, J., and Lammers, R. B.: Global water resources: vulnerability from climate change and population growth, Science, 289, 284-288, 2000.

Wellen, C., Arhonditsis, G. B., Long, T., and Boyd, D.: Accommodating environmental thresholds and extreme events in hydrological models: a Bayesian approach, Journal of Great Lakes Research, 40, 102-116, 2014.

Wilby, R.L.: Uncertainty in water resource model parameters used for climate change impact assessment. *Hydrological Processes*, *19*(16), 3201-3219, 2005.