An Information Theory Approach to Identifying a Representative Subset of Hydro-Climatic Simulations for Impact Modeling Studies

I.G. Pechlivanidis1, H. Gupta2, and T. Bosshard1

1Swedish Meteorological and Hydrological Institute, Norrköping, Sweden, 2Department of Hydrology and Water Resources, University of Arizona, Tucson, AZ, USA

Abstract

Uncertainties in hydro-climatic projections are (in part) related to various components of the production chain. An ensemble of numerous projections is usually considered to characterize the overall uncertainty; however in practice a small set of scenario combinations are constructed to provide users with a subset that is manageable for decision-making. Since projections are unavoidably uncertain, and multiple projections are typically informationally redundant to a considerable extent, it would be helpful to identify an informationally representative subset in a large model ensemble. Here a framework rooted in the information theoretic Maximum Information Minimum Redundancy concept is proposed for identifying a representative subset from an available ensemble of hydro-climatic projections. We analyze an ensemble of 16 precipitation and temperature projections for Sweden, and use these as inputs to the HBV hydrological model to project river discharge until the mid of this century. Representative subsets are judged in terms of different statistical properties of three essential climate variables (precipitation, temperature and discharge), whilst we further assess the sensitivity of the optimized subset for different seasons and future periods. Our results indicate that a quarter to a third of the available set of projections can represent more than 80% of the total information of hydro-climatic changes. We find that the representative subsets are sensitive to the regional hydro-climatic characteristics and the choice of variables, seasons and periods of interest. Therefore we recommend that any selection process should not be solely driven by climatic variables but, rather, should also consider variables of the impact model.

Plain Language Summary

The need for better understanding of climate change and its impact has led to an increasing number of climate models and consequently hydro-climatic projections. Using a large set of projections, we present the information content and the redundant information when an ensemble is considered, and hence demonstrate how representative projections can be identified in order to overcome artifacts/biases introduced by a common hand-picking approach. An identified subset reduced by almost 70% from the large ensemble is capable of representing more than 80% of the total information for precipitation, temperature and discharge; however the selected projections are sensitive to choice of variables, seasons, and period of interest. The size of the representative subset is also related to the regional hydro-climatic characteristics.

1. Introduction

The need to assess the impact of climate change on different sectors, e.g. water, has motivated the development of climate models capable of representing the complexity of natural processes in both space and time (Arheimer et al., 2017; Giorgi et al., 2009). However, uncertainties (epistemic and aleatoric) in such models (e.g. structural hypotheses, boundary conditions, and chaotic behavior of the climate system) can influence the future climatic projections and, further, their sectorial impact (Dégé et al., 2007; Knutti & Sedláček, 2012; Murphy et al., 2004; O’Gorman, 2014). Production of a large multi-model ensemble of climate simulations suitable for impact modeling studies is an attempt to enhance our knowledge about the associated uncertainties in future projections (Clark et al., 2016; Pechlivanidis et al., 2017; Samaniego et al., 2017). Nevertheless, practical designs of sectorial impact assessment studies are typically restricted to using small numbers of climate projections thereby providing users with a subset that is manageable for planning.
studies; accordingly communication of uncertainty between modelers and users can be challenging (Grafton et al., 2013). It would, therefore, be useful to have access to methodologies for the selection of representative subsets of climate/impact modeling projections from the large ensemble of available projections that may be available (Evans et al., 2014; Vautard et al., 2014).

Identifying representative subsets from a high-dimensional sample space is not a straightforward task. Some models share the same systematic errors and selecting two similar models would result in a double-counting problem where the same result is counted twice leading to bias in the overall result. Ideally, one would like the selected subset to capture as much of the full range in simulated future changes as possible (and hence maximize model diversity), and failure to incorporate relevant scenarios in the analysis could lead to inadequate assessments of uncertainty in the indicators of interest (Cannon, 2015; Christensen et al., 2010; Evans et al., 2013; McSweeney & Jones, 2016). Within the ENSEMBLES project, van der Linden and Mitchell (2009) highlighted that "Choosing a subset of the available RCMs based on the weights is not necessarily the best strategy. It might lead to an under-sampling of uncertainty. The minimum requirement would seem to be to use results based on two or more Regional Climate Models that are forced by at least two Global Climate Models". Similar conclusions were drawn by Weigel et al. (2010). Until now, such selection has often been made subjectively, e.g. by hand-picking simulations depending on availability, fidelity (representation of physical processes), resolution (the higher the better), performance (over the historical period), and/or context/representativeness (selecting the most extreme simulations of the climate change signals) (Knutti et al., 2013; McSweeney et al., 2012; Murdock & Spittlehouse, 2011; Salathe et al., 2007).

Recently, more diagnostic approaches have aimed to maximize preservation of information within the limitations dictated by impact studies (Olson et al., 2016; Wilcke & Bärring, 2016). Among the available methods for subset selection, clustering seems to be favored since it can be applied to various combinations of variables and climatic indices (Houle et al., 2012; Masson & Knutti, 2011; Mendlik & Gobiet, 2016). Clustering aims at maximizing the between group variance of an ensemble while minimizing the within-group variance, and can therefore help in selection of members that are representative of high-density regions of the projection space. However, in the case where certain models have contributed many simulations while others have contributed only a few, such "high-density" regions in the projection space may likely be caused by the poorly representative nature of the ensemble (Cannon, 2015). Finally, some clustering algorithms (i.e. k-means) for subset selection may not provide consistent results (with different initial partitions resulting in different final clusters) and so clustering may not produce an ordered sequence of solutions (Jain et al., 1999).

Information theory makes available several powerful tools that can help us in relating multiple interconnected data streams to process understanding (Koutsoyiannis, 2005; Pechlivanidis et al., 2014b; Ruddell & Kumar, 2009). In general, these tools make no assumptions about the underlying system dynamics or relationships among the system variables (e.g. they measure any-order correlations/dependencies among time-series/data sets). The fundamental basis for these tools is the mathematical concept of "entropy" as a quantitative measure of the information content (or conversely, the lack of precision) associated with a signal (in our case hydro-climatic simulations). Information-based metrics have been used in a wide spectrum of topical areas (including climatology, hydrology, environmental and water resources) (Molini et al., 2006; Pechlivanidis et al., 2016b; Peters-Lidard et al., 2017; Singh, 2000; Weijs et al., 2013). Among the various applications, information theory has been used as an avenue to identify where relevant information is present and/or conflicting; e.g., in the optimization, design and management of hydro-meteorological monitoring networks (Li et al., 2012; Ridolfi et al., 2014; Xu et al., 2015). Such problems are addressed via a multi-objective optimization approach, in which redundant information (which in information theory terms is described by the total correlation) is to be minimized whilst total information (described by the joint entropy) is to be maximized. This concept has been termed "Maximum Information Minimum Redundancy" (MIMR) (Li et al., 2012).

To date, there seems to have been no formal application of the MIMR concept to the identification of representative subsets from an ensemble of hydro-climatic projections. What an impact-modeler eventually would like to base the subset selection on is its representativeness with respect to some output variables of the impact model (e.g. discharge, soil moisture, evapotranspiration etc.). However, to fulfill the goal of reducing the number of impact model simulations, subset selection has to also be based on the input.
variables of an impact model. Meanwhile, choosing a subset based on the input data alone may not provide an optimal choice with respect to the output variable.

In this paper, we analyze the representativeness of selected subsets with respect to both the input and the output variables of hydrological impact models and pose the following scientific questions: 1) What is the information content in hydro-climatic projections? 2) If one has N projections, how can one quantify their representativeness with respect to a larger set of projections? 3) If one has to select N projections from a larger set, which projections should one choose?

The paper is organized as follows. In Section 2, the region and data are introduced. Basic entropy theory is briefly presented in Section 3 for ease of understanding the basis for the MIMR criterion. Section 4 presents results of an experiment in identification of a representative hydro-climatic subset. A discussion on the potential of the proposed framework, application over other hydro-climatic gradients, and comparison with other sub-selection methods is presented in Section 5. Finally, Section 6 states the conclusions.

2. Study Area, Data and Models

2.1. Study Area and Climatic Data

Sweden covers a surface area of about 450,000 km² with forest dominating the country’s land use (about 65%), yet major agricultural areas are present in the south. The country is drained by a large number of rivers that have their sources in the west (mountainous areas) and run eastwards to the Baltic Sea and southwest to the North Atlantic. The climatic patterns over the country can be clustered into four regions following analyses of long-term records and scenario modeling (Figure 1a) (Lindström & Alexandersson, 2004). The river basins in these regions also show similarities in morphology, but also represent the catchments of the marine basins (Arheimer & Lindström, 2015).

The large ensemble set consists of 16 available climate projections for Sweden (Table 1) (Bergström et al., 2012). It includes 3 different emission scenarios, 5 GCMs, 6 RCMs, and two spatial resolutions. Note that the ensemble does not sample all the uncertainty sources of the modeling chain in a systematic manner; the ECHAM5 GCM, RCA3 RCM, and the A1B emission scenario are dominating the ensemble. In addition, there are cases in which the model combination (modeling chain) is the same but projections are generated at

Figure 1. (a) The four climate regions in Sweden, and (b) the HBV model performance (in terms of NSE) at 69 test-basins over the country with long-term records.
different spatial resolutions (i.e. projection ID 3 and 4). Thus, it is rather an ensemble of opportunity constructed of readily available data, as it is often the case in real world impact studies. Nevertheless, we assume here that all available climate models are equally valuable for application in Sweden; i.e., that they do not contain any systematic errors and have acceptable historical performance.

In a recent study by Kjellström et al. (2016) the uncertainty of an ensemble downscaled by RCA4 (a newer version of RCA3) has been compared to a larger ensemble of EURO-CORDEX projections. The uncertainty in the RCA4 ensemble was shown to be substantially smaller than in the larger ensemble. Hence, it is very likely that the ensemble used here tends to underestimate the uncertainty in the climate projections.

The PTHBV 4x4 km gridded data set for observed precipitation and temperature over Sweden (Johansson, 2002) was used as a reference (period 1961–1990) for the bias-correction of the climate projections. The Distribution Based Scaling, DBS (Yang et al., 2010) method was applied to conduct the bias correction. DBS is a variant of the quantile-mapping bias-correction methods; see more details in Yang et al. (2010).

### 2.2. Hydrological Modeling

The HBV model (Lindström et al., 1997) is applied over Sweden with a spatial resolution of some 1,000 basins (average 450 km²). The model produces daily values for hydrological variables over historical periods and in forecasting (Pechlivanidis et al., 2014a). It includes conceptual numerical descriptions of hydrological processes at the catchment scale, i.e. snow accumulation and melting, evapotranspiration, soil moisture, discharge generation, groundwater recharge, and routing through rivers and lakes. For basins of considerable elevation range a subdivision into elevation zones is also made. Each elevation zone is further divided into different vegetation zones (e.g., forested and non-forested areas). The model is reasonably well calibrated with a Nash-Sutcliffe Efficiency (Nash & Sutcliffe, 1970) greater than 0.70 at all 69 selected indicator basins, for all of which real-time discharge observations are available and for which there is little or no upstream regulation (Figure 1b). This indicates that the HBV model structure and the identified parameters can appropriately represent the dominant hydrological processes over Sweden, avoiding unreliable portrayals of climate change impacts. The model was forced with daily precipitation and temperature from the bias-corrected climate model projections simulating the period 1981–2100 (apart from projections with ID 1, 3, 7 and 8 which are available until 2051).

### Table 1

| ID | Emission scenario | GCM     | RCM (Resolution) | Abbreviation                        |
|----|-------------------|---------|------------------|-------------------------------------|
| 1  | A1B               | ECHAM5r1| SMHI-RCA3 (*)    | A1B-ECHAM5r1-RCA3*                  |
| 2  | A1B               | ECHAM5r2| SMHI-RCA3 (*)    | A1B-ECHAM5r2-RCA3*                  |
| 3  | A1B               | ECHAM5r3| SMHI-RCA3 (*)    | A1B-ECHAM5r3-RCA3*                  |
| 4  | A1B               | ECHAM5r3| SMHI-RCA3 (**)   | A1B-ECHAM5r3-RCA3**                 |
| 5  | B1                | ECHAM5r1| SMHI-RCA3 (*)    | B1-ECHAM5r1-RCA3*                   |
| 6  | A1B               | ARPEGE  | SMHI-RCA3 (*)    | A1B-ARPEGE-RCA3*                    |
| 7  | A1B               | CCSM3   | SMHI-RCA3 (*)    | A1B-CCSM3-RCA3*                     |
| 8  | A1B               | ARPEGE  | CNRM-ALADIN (**) | A1B-ARPEGE-ALADIN**                 |
| 9  | A1B               | ECHAM5r3| KNMI-RACMO (**)  | A1B-ECHAM5r3-RACMO**                |
| 10 | A1B               | ECHAM5r3| MPI-REMO (**)    | A1B-ECHAM5r3-REMO**                 |
| 11 | A2                | ECHAM5r3| SMHI-RCA3 (**)   | A2-ECHAM5r3-RCA3**                  |
| 12 | A1B               | HadCM3Q0| HC-HadRM3 (**)   | A1B-HadCM3Q0-HadRM3**               |
| 13 | A1B               | HadCM3Q16| SMHI-RCA3 (**)  | A1B-HadCM3Q16-RCA3**                |
| 14 | A1B               | BCM     | DMI-HIRHAM (**)  | A1B-BM-HIRHAM**                     |
| 15 | A1B               | HadCM3Q0| DMI-HIRHAM (**)  | A1B-HadCM3Q0-HIRHAM**               |
| 16 | A1B               | ECHAM5r3| DMI-HIRHAM (**)  | A1B-ECHAM5r3-HIRHAM**               |

**Note.** Projections in italics are only available for the period 1981–2051. RCMs with (*) and (**) have a 50x50 and 25x25 km spatial resolution, respectively.
3. Methods

3.1. Information Theory Metrics

Entropy ($H$) is a quantitative measure of the amount of information content (or uncertainty) in a system, represented in the probability distribution of independent draws of a discrete non-negative random variable $X$ (Shannon, 1948), i.e.:

$$H(X) = -\sum_{i=1}^{N} p(x_i) \cdot \log_2[p(x_i)]$$

where $p(x_i)$ is the probability of occurrence of outcome $x_i$, $N$ is the number of possible outcomes. Entropy is measured in bits (average number of binary digits), when the base of the logarithm equals 2. Entropy takes values between 0 (complete information) and $\log_2(N)$ (no information); a low value of entropy indicates a high degree of structure and a low uncertainty.

In the case of an ensemble of $M$ (multivariate) discrete non-negative random variables, the joint entropy ($JH$) is defined as a measure of the overall information retained by the random variables, i.e.:

$$JH = H(X_1, \ldots, X_M) = -\sum_{i_1=1}^{N_1} \cdots \sum_{i_M=1}^{N_M} p(x_1, \ldots, x_M) \cdot \log_2[p(x_1, \ldots, x_M)]$$

where $p(x_1, \ldots, x_M)$ is the joint probability of the $M$ variables. Quantitatively, $JH$ is less than or equal to the sum of its one-dimensional marginal entropies; the equality holds only when the random variables are stochastically independent.

Within an ensemble of variables, it is probable that information regarding one random variable (e.g. $X_1$) can be inferred from knowledge of another (e.g. $X_2$). Mutual information (also known as transinformation, $T$) is a measure of dependence (linear and nonlinear) between two random variables.

$$T(X_1; X_2) = \sum_{x_1} \sum_{x_2} p(x_1, x_2) \cdot \log_2 \left( \frac{p(x_1, x_2)}{p(x_1) \cdot p(x_2)} \right)$$

Total correlation ($C$) measures the amount of redundant (duplicated) information among random variables, i.e.:

$$C(X_1, \ldots, X_M) = \sum_{i=1}^{M} H(X_i) - H(X_1, \ldots, X_M)$$

where $H(X_i)$ is the marginal entropy of the $i$th random variable and $H(X_1, \ldots, X_M)$ is the joint entropy of the $M$ random variables. The total correlation is non-negative, and takes on the value 0 when all random variables are independent.

3.2. Data Discretization

Most entropy-based metrics (i.e. $T$ and $C$) are based on the evaluation of probability density functions (PDFs), whose estimation (marginal and joint probabilities) is not trivial when the space is of high dimension (e.g., when there are a large number $M$ of ensemble members). This problem is typically tackled using data discretization methods, of which the most frequently used methods are the data-binning/histogram method and the mathematical floor function (Mishra and Coulibaly, 2010; Ridolfi et al., 2014). Herein, we use the first method since this avoids the need to select a mathematical form for the distribution to fit the continuous data; however a selection (by optimization) of the number of bins is required (note that entropy-based metrics can be sensitive to the bin size (Pechlivanidis et al., 2016b)). Through the discretization, a continuous value is converted to an integer and hence into entropy values, which are no longer measures of information of the continuous random variables, but the information of the newly discretized value (see an example in supporting information Figure S1). However, in hydro-climatic projections, where precise quantification of the information content is typically not critical, this approximation is reasonable and can be accepted. Nevertheless it is important that the relative dependency among projected changes is preserved in terms of information content. For evaluating further the joint PDFs from more than two variables, it is necessary due to computational challenges to agglomerate the values (usually rounded to the greatest
integer) and generate a new variable with similar properties. Consequently, no joint probabilities from a large ensemble of variables are needed for the calculation of the entropy metrics (Li et al., 2012).

3.3. The Maximum Entropy Minimum Redundancy (MIMR) Approach

The basic concept of the MIMR approach is to select and rank members from a large ensemble of a variable. MIMR aims to select sets that: 1) maximize the overall information content (joint information), 2) maximize the overall information transition ability (transinformation), and 3) minimize the redundant information (total correlation) (Li et al., 2012).

In the case of $M$ random members of a variable, there are several records of the variable ($X$) distributed in space over the domain. Let $S$ be the case of optimal selection of a set of members, whose elements are denoted as $X_{S1}, X_{S2}, \ldots, X_{SM}$. The multi-objective optimization problem is to select an optimal subset, which can convey effective information as much as possible, while retaining redundant information (if any) as little as possible. All objectives are needed because simply minimizing the total correlation could result in a set of non-redundant variables (members) that would not necessarily provide enough information content. Likewise, only maximizing joint entropy could result in a set of very informative but highly redundant variables. The multi-objective goal is mathematically expressed as:

$$\text{Max}(JH) = \text{Max} \{ H(X_{S1}, \ldots, X_{SM}) \}$$
where $JH$ is the joint entropy, $T$ is the transinformation, and $C$ is the total correlation of $M$ random variables.

### 3.4. Experimental Design

For this analysis, we aggregated all data to the four climatic regions (Region 1–4), and the entire domain (All) (Figure 1a). At that level of spatial aggregation, we derived climate change signals of: 1) seasonal (‘DJF’, ‘MAM’, ‘JJA’, ‘SON’) and annual (‘Annual’) scales, 2) mean and extreme (10th and 90th percentile) statistical values, and 3) several different variables (i.e. precipitation, temperature and discharge). In doing so, we masked out the results for the low extremes (10th percentile) of precipitation since these values correspond to 0 mm, and hence artifacts could be introduced into the relative change signal under climate change. The climate change signals were estimated as differences between the corresponding statistics of the future 30-year scenario periods 2021–2050 (early-century) and 2046–2075 (mid-century), and the reference period 1981–2010. We chose a reference period closer to the present than the one used in bias-correction. As we are studying the climate change signal only, the choice of reference period is not restricted to the same period that has been used for bias-correction. Note that internal variability will impact on the climate change signal and thus results differ depending on the choice of the reference period. This is even more so in the case of analyzing climate change signal of extremes (see for example Figure 5 in Arheimer and Lindström (2015)) Keeping this effect in mind, we restrict all further analysis here to the climate change signals with respect to only one reference period. Figure 2 shows an outline of the selection process.

### 4. Results

#### 4.1. Future Changes in Long-Term Seasonal Means

The patterns of changes in precipitation and temperature were analyzed for each climate model and the percentiles of changes were computed (supporting information Figures S2 and S3). As expected, projected changes depend on the climate model, region and season; in general, however, changes in precipitation and temperature seem to be more prone to seasonal variations than to geographical locations. Overall, the results indicate an increase in precipitation over most of the country, including areas where precipitation amounts are already quite high (e.g., the mountain range at the borders with Norway). The signal of increased precipitation is stronger during the winter in comparison to the remaining seasons; however, the deviation in the direction of trends increases during the warm months (MAM and JJA). A spatial North-South gradient is visible with smaller precipitation increases (or even decrease during summer) in the South. Precipitation increases are strongest in the mountain regions in the North-West. Temperature increases in all seasons and over the whole domain. Increases are strongest during winter. In general, there is also a North-South gradient with larger increases in the North.

These climatic changes also affect discharge and its spatiotemporal regime (supporting information Figure S4). In particular, the projections show remarkably different results between the northern-central and the southern regions. In the south, the increasing temperature lead to higher evapotranspiration which over-compensates for the small increase in precipitation and leads to a decrease of mean discharge in all seasons except winter. In summer, projections indicate decreasing discharge in the whole domain and the opposite is projected for winter. It is clear that projected changes in discharge are determined by a combination of temperature and precipitation changes.

#### 4.2. Information Content (Marginal and Joint) in Hydro-Climatic Projections

We next address the question “What is the information content in hydro-climatic projections?” For this, we analyze the information content (described by marginal entropy ($H$)) from each projection and the information properties (mutual information ($MI$) and joint entropy ($JH$)) when projections are inter-compared and/or combined in a pairwise manner. Figure 3 shows the information matrix ($H$, $MI$, and $JH$) for the three variables (mean annual statistics) averaged over the entire Sweden. Results highlight the presence of highly informative projections, however any given projection is not always highly informative for all variables (e.g. ID10 (A1B-ECHAM5r3-REMO**)).
Note that the information content of a single projection or the joint information from two projections is higher for precipitation than for temperature. This is due to precipitation’s spatiotemporal patterns being generally characterized by high variance (thus containing a lot of information). Climate models represent this high variance fairly well, which results in a small range for MI (MI for precipitation and temperature are as large as 34% and 43% respectively).

Further, although precipitation drives the hydrological response, dampening mechanisms at the basin scale act to filter the information content. Results show MI values between projections to be as large as 40%, which indicates the presence of redundant information in a potential combination, and hence suggests that a random sub-selection could introduce bias into a resulting analysis.

In particular, of the 16 projections, ID5 (B1-ECHAM5r1-RCA3*) has the highest H (information content) and generally high MI with all remaining projections, particularly for precipitation. Consequently, ID5 can be used as a starting projection for subset selection. Other highly informative projections include ID1 (A1B-ECHAM5r1-RCA3*), ID12 (A1B-HadCM3Q0-HadRM3**) and ID14 (A1B-BCM-HIRHAM**), which share the same emission scenario (A1B; which generally predicts a fast rising mean temperature in the early-century and a smooth rise in the middle of the century), but different GCM and RCM.

4.3. Information Gain in Hydro-Climatic Projections
Here, we address the question “if one has N projections, how can one quantify their representativeness with respect to a larger set of projections?” For this, we investigate the cumulative distribution of the joint information where projections are introduced into the subset in a step-wise manner (Figure 4). This is done for all variables and their statistical properties, future periods and regions of the domain.

As expected, the slope of the distributions differs between the variables; interestingly, however, they also differ between regions. This could be due to the different impact (relative change) that climate change can have in the four regions and overall. Precipitation and temperature cumulative distributions are characterized by sharp rises, in comparison to the smoother increase in total information associated with discharge.

An important characteristic of this analysis lies in the diagnostic ability of MIMR to quantify the minimum number of projections needed to represent a certain fraction of the total information. For instance, if 80% of the total information from the large ensemble is sufficient to represent the changes due to climate, our results indicate that one could select as few as 4 (or even 3) climatic projections and 5 (or 4) hydrological projections (see also supporting information Figure S5). This is a key finding in regards to climate and impact studies, and even climate services, since any additional projections (with extra associated computational costs) can only increase the total available information by about 20%.

This is particularly true in results for the mid-century where, in general, fewer projections are needed to represent 80% of the total information. Although the general pattern is that the ensemble spread gets wider.
with increasing lead time, here many of the models project a similar pattern of climatic changes for the mid-century period. This is probably because we are still within the time horizon when internal variability (i.e. decadal variability) is dominant, and hence only a smaller subset of projections is needed when compared with the corresponding subset for the early-century. The homogeneity of mid-century climate change and its impact over the entire country results into a more consistent (less spread) distribution of total information between the regions.

The size of the optimal representative subset also varies with regional hydro-climatic characteristic. A larger subset is generally required for the representation of changes in precipitation in region 1 during winter and spring compared to the other regions and seasons. Precipitation over region 1 shows high spatial variability that is strongly related to the high elevation differences. Temperature is more variable over regions 1 and 2, resulting in the need for a larger subset of projections. Elevation is also a factor controlling temperature’s spatial variability in these two regions, yet Sweden is characterized by a temperature gradient from north to south. It is interesting to note that temperature seems to require a larger subset than precipitation when

Figure 4. Cumulative joint information with increasing number of projections for different statistics of the annual hydro-climatic variables and for two future periods (early- and mid-century).

Figure 5. Ranking (from 1 to 16 with 1 being the most important) of projection members for the mean statistic of the hydro-climatic variables, and for different seasons, future periods, and regions in Sweden. Boxes with a red dot indicate membership to a subset representing 80% of the total information from the large ensemble.
mean changes are assessed. However, if one focuses on high extremes, the stochastic nature of precipitation results in values with high spatiotemporal variability, and hence representativeness of a subset requires more projections for precipitation than for temperature.

The results further show that discharge occasionally needs the largest subset of projections to adequately represent the information content from the climate change signal compared to the other two variables. The spatiotemporal variability in discharge is subject to effects of joint precipitation and temperature spatiotemporal patterns, resulting into a regime with snow-driven and rain-driven high flows in northern–central and southern Sweden, respectively; snow melt results in spring floods in the northern-central region. However, the river systems in the northern-central parts of the country are characterized by long memory (following climatology), which has a dampening effect on variability in the precipitation-discharge relationship. Regions 3 and 4 respond quickly to precipitation and are more sensitive to variability in its patterns and therefore a representative subset consists of more projections compared to the subsets for regions 1 and 2.

4.4. Identification of an Optimal Subset

We finally address the question “if one had to select N projections from a larger set, which projections should one choose?” We address this as a function of variable, future period, season, and region. The MIMR approach allows ranking of the projections based on the fraction of information they add to the total information in a selected subset. These results are presented in Figure 5 for the mean statistics for all regions, variables, seasons and future periods.

The results show that the identification of a consistent subset of projections that is highly informative (or optimum) for all of the factors investigated (region, variables, seasons and future periods) is not straightforward. As expected, the optimum number of projections to represent 80% of the total information from the large ensemble varies between different combinations. These results highlight the need for the investigation to be tailored to the objective of the user, given for instance that the optimum subset for winter differs from the optimum subset for summer in a region. Similar conclusions were drawn when other statistics (e.g. extremes) were used.

It is also important to note that the overall pattern of ranking for precipitation and temperature differs to the pattern for discharge. This is a key finding, and it contradicts the traditional strategy of identifying hydrological projections based only on information from climate projections. Despite the computational requirements, a robust selection of representative hydro-climatic projections requires both hydrological and climate modeling, and detailed analysis of these results.

Although, in general, spatial patterns are difficult to decompose, an emerging pattern (here defined from the count of red points in a selected period) seems to be identifiable and be consistent for all factors. Projections ID4 (A1B-ECHAM5r3-RCA3**) or ID5 (B1-ECHAM5r1-RCA3*) have been selected in most investigations. This is particularly the case for the later scenario period for which differences between emissions scenarios become clearer (see for e.g. Hawkins & Sutton 2009). Other most highly selected projections include ID12 (A1B-HadCM3Q0-HadRM3**), and ID14 (A1B-BCM-HIRHAM**). ID12 and ID14 have modeling chains with differing GCM and RCM from ID4 or ID5 and that is also believed to be an important factor in selecting independent models (Déqué et al., 2007). In addition, ID1 (A1B-ECHAM5r1-RCA3*) is also listed among the selected projections, however interestingly in most cases ID1 is not selected if ID4 is selected, which could indicate interchangeability (both projections have the same model chain).

It is also interesting to note which projections were not often selected. ID11 (A2-ECHAM5r3-RCA3) gets rarely selected despite a differing emission scenario. The emission scenario A2 does not differ much from A1B until mid-century (Nakicenovic & Swart, 2000), which may partly explain its exclusion from the selected subset. Also, ID15 and ID16 are rarely selected. They belong to the DMI-HIRHAM sub-ensemble which is already represented by ID14 in the selected subset. In general, the overrepresentation of ECHAM5 in the ensemble is efficiently eliminated from the selected subset.

The final sub-selection resulted in a subset that consists of 2 emission scenarios, 3 GCMs, and 3 RCMs (out of 3, 5, and 6 respectively); hence respecting to a wide extent the divergence of the climate probability distribution function from the large ensemble. Note that this is not the optimum combination for all investigations but a generally informative set of projections. This means that despite the different structure of the climate models, similar information is present within their setups.
5. Discussion

5.1. Sub-Selection Using the MIMR Concept

Uncertainty characterization is not straightforward because the uncertainties are not explicitly encapsulated in the differences among the models. The set of projections used here could offer a biased sample of the range of possible futures, but this is commonly the “real case” where the ensemble set-up is based on availability. Regardless, we assume here that: 1) all available climate models do not contain any systematic errors and are equally valuable (in terms of process representation) for application in our geographic domain, and 2) all climate models are equally performing and hence suited to drive the hydrological model.

Our results indicate that an ensemble of hydro-climatic projections used in climate change impact studies is characterized by significant redundant information. Our exploration of the information content in hydro-climatic projections suggests that the identification of the optimum representative subset should be characterized by maximum total information and transition ability, and minimum redundant information. Our findings demonstrate that the identification (and ranking) of representative projections must be treated as a set and not individually, whilst the differentiaion of the hydrological from the climatological analysis can again lead to biased subsets. A marginal and/or joint information content analysis provides diagnostic value, however even at this stage a subset selection is subjective (for example, the selection of the most marginally informative projections may not result into the most highly informative subset (in terms of total information) due to the potential of introducing redundant information) and can further lead to artifacts. The MIMR algorithm has the potential to quantify the added information value when a new member is introduced in a pre-selected subset, which further allows structuring the optimum subset.

It is unfortunate that results do not lead to general rules for subset selection; this is because identifiability of optimum subsets depends on a number of factors, e.g. variables, regions, seasons, future periods, and acceptable level of representativeness (here, 80% of total information from the large ensemble). However, a list of solutions is provided that are optimal for different objectives, explicitly considering projection uncertainty and information content. This is important from a practical perspective providing users of climate information with a manageable subset of projections for decision-making.

5.2. Applicability Over a Global Hydro-Climatic Gradient

In the context of the relevance of our approach to the global scale, we note that Sweden’s characteristics undoubtedly describe only a portion of the global or even other regional hydro-climatic gradients (for instance, India experiences stronger spatiotemporal variability in atmospheric and hydrologic fluxes than Sweden) (Berghuis et al., 2014; Pechlivanidis et al., 2016a). Here, we found a dependency between the identified representative subset and the regional hydro-climatic characteristics, but this link is dependent on the changes in the fluxes as these are derived from hydro-climatic projections.

Although, climate change and its impact at the global and regional scale have been assessed using a large ensemble of projections (Krysanova et al., 2017; Prudhomme et al., 2014), the potential for reduction of the number of ensembles has heretofore not being investigated. Although, our demonstration focuses on Sweden (due to the data availability), our proposed approach is not limited to a geographic location. Our results show that average increases in temperature in the mountainous regions affect snowpack seasonality more than changes in precipitation (Destouni et al., 2013; Molini et al., 2011). Snow/ice and lake processes can strongly influence the river response (as in Sweden and other regions around the globe), however the various human impacts (reservoir regulation, irrigation etc.) on the current future regimes are uncertain (Arheimer et al., 2017). We therefore expect that such anthropogenic influences may mask potential patterns in change and further bias the representative subsets.

5.3. Comparison With Other Sub-Selection Methods

As indicated by our results, the common ‘hand-picking’ approach can easily introduce artifacts into the analysis of expected impact, including e.g. underestimation of internal variability and model uncertainty. Recent state of the art sub-selection methods instead aim to fulfill certain objective criteria for the indicators of interest; i.e. to respect the spread of change signals, accounting for model independency, and ensure model performance.
Approaches to respect the spread of change signals commonly rely on clustering methods, which are indeed a step forward and can diagnostically guide towards a representative subset. The basic concept is to group the ensemble members for a set of climate impact indicators and then either select representative members out of each group (Mendlik & Gobiet, 2016; Wilcke & Bärring, 2016) or select the most extreme projections and hence maximize model diversity (Cannon, 2015). However, clustering methods only aim, non-systematically due to the random initial partition, at sampling from the defined clusters, without quantifying the information gain with respect to the large set (something capable with the MIMR method), and typically without being capable of respecting higher-order moments (clustering is usually dependent on first and second order moments). Consequently, this can lead to overestimation of the optimum number of members within the subset. Moreover, in the case where certain models have contributed several simulations to the ensemble, high-density regions in the projection space are formed, affecting the properties of the cluster groups.

Agreement between models may arise due to the similarity of process representation, featuring further similar errors (Bishop &Abramowitz, 2013; Knutti et al., 2010). Consequently, models in a large ensemble commonly do not represent independent information, and in the case of double-counting, they could introduce biases in the statistical analysis (Pirtle et al., 2010). Pairwise distance metrics have been applied to define similarity between models in an ensemble leading to “family trees” in which the spatial fields (seasonal cycle, interannual variations, spatial correlation) of the variables of interest are similar. However, in that approach one needs to analyze (for each projection) the metric and variables that are important and further how the distance in the present relates to the distance in predicted changes (Knutti et al., 2013; Evans et al., 2013). Here, our results show that the proposed MIMR approach down-weights the modeling chains that have similar control biases to avoid biased projections from near duplicate models leading in a robust solution for a specific set of objectives (e.g. variables, seasons, future periods).

Sub-selection can also be based on model performance evaluation (for the past), with or without regard to the spread in the change signal (Knutti, 2010; Wilby, 2010). However it is widely argued that approaches to sub-selection of models should not depend solely on model performance (Weigel et al., 2010). Historical model performance is not strongly correlated to future hydro-climatic change signals, which means that the “best” historically performing models might not be the most realistic with regard to change signals (Knutti et al., 2010). Further, performance is subjectively defined; e.g. “best” performance in terms of agreement with observations and if so for which variables and which data sets? “best” in terms of physical representation? Finally, although climate models are increasingly documented in the literature in terms of process representation and performance, they commonly still remain as massive and complex black boxes to many impact modelers.

**6. Conclusions**

Herein, we have investigated the information content present in a number of hydro-climatic projections and, consequently, the factors affecting the identification of a representative subset in an available ensemble. Because large model ensembles will typically contain redundant information, we propose that a subset be selected that maximizes independence and minimizes redundancy. Via implementation of this multi-objective information-based approach, representative hydro-climatic projections can be identified and artifacts/biases introduced by hand-picking projections for climate change impact assessments can be reduced.

Over Sweden’s hydro-climatic gradient, a subset of 20–35% of the total available projections can represent a large fraction (>80% of total information) of the ensemble range for changes in precipitation, temperature and discharge. The selection of subset is sensitive to choice of variables and seasons. The spatiotemporal pattern of precipitation is more variable compared to temperature and discharge, and this results in higher information content provided by a single precipitation projection. We find that the information content in the hydro-climatic projections changes with time, and hence the representative set of projections can vary depending on the period of interest. Results show that a representative subset for the mid-century has fewer members compared to the identified subset for the early-century, due to the similarity of change pattern (and hence information content) during mid-century.
The size of the representative subset is also related to the hydro-climatic characteristics of the river system. High variability in precipitation and temperature changes are observed in the northern-central Sweden, resulting in a larger subset size compared to the southern part of the country. However, this part of the country characterized from rain-driven flows is more prone to change and hence a representative subset is larger than the subset for the snow-fed northern-central river basins. We finally conclude that identified representative subsets of projections vary in both in space and time (i.e. regions, variables, seasons and future periods). Accordingly climate change impact assessments should target the objective of the user and/or investigation, and identification of subsets should not be solely based on climatic information but should consider hydrological information as well.

Acknowledgments
This study was partially funded by the EU FP7-funded project SWITCH-ON (grant agreement 603587), which explores the untapped potential of Open Data to tackle changes in the Hydrosphere. This work was also partially funded by the project AQUACLEW, which is part of ERA4CS, an ERA-NET initiated by JPI Climate, and funded by FORMAS (SE), DLR (DE), BMWPFW (AT), IPF (DK), MNCEO (ES), ANR (FR) with co-funding by the European Commission (grant agreement 690462). Funding was also received from the Swedish Environmental Protection Agency (Naturvårdsverket) through the Climate Change and Environmental Objectives (CLEO) project (contract no. 802-0115-09). The investigation was performed at the SMHI Hydrological Research unit, where much work benefits from joint efforts in developing models and concepts by the whole team. Thanks go to Dr C. Li for sharing the MIRR scripts. The scientific findings will contribute to the decadal research initiative “Panta Rhei - changes in hydrology and society” by the International Association of Hydrological Sciences (IAHS). The authors would finally like to express their sincere gratitude to Dr Martyn Clark, Dr Ole Rössler and an anonymous reviewer for their constructive comments. The data that support the findings of this study are available in Zenodo (https://doi.org/10.5281/zenodo.1298987). I.G.P. contributed with the basic idea, the overall study design, result analysis, figures and writing the manuscript; H.G. contributed with complementary ideas, interpretation of results, and commenting on the manuscript; T.B. contributed with the basic idea, compilation of model outputs, supporting information figures, and commenting on the manuscript.

References
Arheimer, B., Donnelly, C., & Lindström, G. (2017). Regulation of snow-fed rivers affects flow regimes more than climate change. Nature Communications, 8(62), 1–9. https://doi.org/10.1038/s41467-017-00992-8
Arheimer, B., & Lindström, G. (2015). Climate impact on floods – changes of high-flows in Sweden for the past and future (1911–2100). Hydrology and Earth System Sciences, 19, 771–784. https://doi.org/10.5194/hess-19-771-2015
Berghuis, W. R., Sivapalan, M., Woods, R. A., & Savenije, H. H. G. (2014). Patterns of similarity of seasonal water balances: A window into streamflow variability over a range of time scales. Water Resources Research, 50, 5638–5661. https://doi.org/10.1002/2014WR015692
Bergström, S., Andréasson, J., & Graham, L. P. (2012). Climate adaptation of the Swedish guidelines for design floods for dams. Proceedings of Cold 24th Congress Kyoto 2012, Q94 (International Commission on large dams (ICOLD), 2012), Kyoto, Japan.
Bishop, C. H., & Abramowitz, G. (2013). Climate model dependence and the replicate Earth paradigm. Climate Dynamics, 41(3–4), 885–900. https://doi.org/10.1007/s00382-012-1610-y
Cannon, A. J. (2015). Selecting GCMs that span the range of changes in a multimodel ensemble: Application to CMIP5 climate extremes indices. Journal of Climate, 28(3), 1260–1267. https://doi.org/10.1175/JCLI-D-14-00636.1
Christensen, J. H., Kjellström, E., Giorgi, F., Lenderink, G., & Rummukainen, M. (2010). Weight assignment in regional climate models. Climate Research, 44(2–3), 179–194. https://doi.org/10.3354/cr00916
Clark, M. P., Wilby, R. L., Gutmann, E. D., Vano, J. A., Gangopadhyay, S., Wood, A. W., et al. (2016). Characterizing Uncertainty of the Hydrologic Impacts of Climate Change. Current Climate Change Reports, 2(2), 55–64. https://doi.org/10.1007/s40641-016-0034-x
Déqué, M., Rowell, D., Lüthi, D., Giorgi, F., Christensen, J. H., Rockel, B., et al. (2007). An intercomparison of regional climate simulations for Europe: Assessing uncertainties in model projections. Climatic Change, 81, 53–70. https://doi.org/10.1007/s10584-006-9226-e
Destouni, G., Jaramillo, F., & Prieto, C. (2013). Hydroclimatic shifts driven by human water use for food and energy production. Climate Change, 2, 213–217. https://doi.org/10.1038/nclimate1719
Evans, J. P., Ji, A., Abramowitz, G., & Ekström, M. (2013). Optimally choosing small ensemble members to produce robust climate simulations. Environmental Research Letters, 8(4). https://doi.org/10.1088/1748-9326/8/4/044050
Evans, J. P., Ji, A., Lee, C., Smith, P., Argueso, D., & Fita, L. (2014). Design of a regional climate modelling projection ensemble experiment - NARCIM. Geoscientific Model Development, 7(2), 621–629. https://doi.org/10.5194/gmd-7-621-2014
Giorgi, F., Jones, C., & Asrar, G. R. (2009). Addressing climate information needs at the regional level: The CORDEX framework. WMO Bulletin, 58(3), 175–183.
Graffon, R. Q. et al. (2013). Global insights into water resources, climate change and governance. Nature Climate Change, 3(4), 315–321. https://doi.org/10.1038/nclimate1746
Hawkins, E., & Sutton, R. (2009). The potential to narrow uncertain in regional climate predictions. Bulletin of the American Meteorological Society, 90(8), 1095–1107. https://doi.org/10.1175/2009BAMS2607.16
Houle, D., Bouffard, A., Duchesne, L., Logan, T., & Harvey, R. (2012). Projections of future soil temperature and water content for three Southern Quebec forested sites. Journal of Climate, 25(21), 7690–7701. https://doi.org/10.1175/JCLI-D-11-00440.1
Jain, A. K., Muñoz, M. N., & Flynn, P. J. (1999). Data clustering: A review. ACM Computer Survey, 31(3), 264–323. https://doi.org/10.1145/331499.331504
Johansson, B. (2002). Estimation of areal precipitation for hydrological modelling in Sweden (PhD thesis, Report No. A76), Göteborg, Sweden: Earth Sciences Centre, Göteborg University.
Kjellström, E., Bärring, L., Nikulin, G., Nilsson, C., Persson, G., & Strandberg, G. (2016). Production and use of regional climate model projections – A Swedish perspective on building climate services. Climate Services, 2–3, 15–29. https://doi.org/10.1016/j.cliser.2016.06.004
Knutti, R. (2010). The end of model democracy? Climate Change, 102(3), 395–404. https://doi.org/10.1007/s10584-010-9800-2
Knutti, R., Furrer, T., Tebaldi, C., Cermak, J., & Meehl, G. A. (2010). Challenges in combining projections from multiple climate models. Journal of Climate, 23(10), 2739–2758. https://doi.org/10.1175/2009JCLI3361.1
Knutti, R., Masson, D., & Gettelman, A. (2013). Climate model genealogy: Generation CMIPs and how we got there. Geophysical Research Letters, 40, 1194–1199. https://doi.org/10.1029/2012GL052056
Knutti, R., & Sedláček, J. (2012). Robustness and uncertainties in the new CMIP3 climate model projections. Nature Climate Change, 3(4), 369–373. https://doi.org/10.1038/nclimate1716
Koutsouyannis, D. (2005). Uncertainty, entropy, scaling and hydrological stochastic. 2. Time dependence of hydrological processes and time scaling/Incertitude, entropie, effet d'echelle et proprietes stochastiques hydrologiques. 2. Dependance temporelle des processus hydrologiques. Hydrological Sciences, 50(3), 1–426. https://doi.org/10.2166/hys.2005.65028
Krysanova, V., Vetter, T., Esiner, S., Huang, S., Peclivandis, I., Strauch, M., et al. (2017). Intercomparison of regional-scale hydrological models in the present and future climate for 12 large river basins worldwide - A synthesis. Environmental Research Letters, 12, 105002. https://doi.org/10.1088/1748-9326/a8359
Li, C., Singh, V. P., & Mishra, A. K. (2012). Entropy theory-based criterion for hydrometric network evaluation and design: Maximum information to minimum redundancy. Water Resources Research, 48, WR011251. https://doi.org/10.1029/2011WR011251
Lindström, G., & Alexandersson, H. (2004). Recent mild and wet years in relation to long observation records and climate change in Sweden. Ambio, 33(4), 183–186. https://doi.org/10.1579/0044-7447-33.4.183
Lindström, G., Johansson, B., Persson, M., Gardelin, M., & Bergström, S. (1997). Development and test of the distributed HBV-96 hydrological model. Journal of Hydrology, 201, 272–288.
Masson, D., & Knutti, R. (2011). Climate model genealogy. Geophysical Research Letters, 38, L08703. https://doi.org/10.1029/2011GL046864
McSweeney, C. F., & Jones, R. G. (2016). How representative is the spread of climate projections from the 5 CMIP5 GCMs used in ISI-MIP? Climate, 24, 29–29. https://doi.org/10.1002/2016JD025895
McSweeney, C. F., Jones, R. G., & Booth, B. B. B. (2012). Selecting ensemble members to provide regional climate change information. Journal of Climate, 25(20), 7100–7121. https://doi.org/10.1175/JCLI-D-11-00526.1
Mendlik, T., & Gobiet, A. (2016). Selecting climate simulations for impact studies based on multivariate patterns of climate change. Climate Change, 135(3–4), 381–393. https://doi.org/10.1007/s10584-015-1582-0
Mishra, A. K., & Coulibaly, P. (2010). Hydrometric network evaluation for Canadian watersheds. Journal of Hydrology, 380(3–4), 420–437. https://doi.org/10.1016/j.jhydrol.2009.11.015
Molini, A., Katul, G. G., & Porporato, A. (2011). Maximum discharge from snowmelt in a changing climate. Geophysical Research Letters, 38, 1–5. https://doi.org/10.1029/2010GL046477
Molini, A., Labarbera, P., & Lanza, L. (2006). Correlation patterns and information flows in rainfall fields. Journal of Hydrology, 322(1–4), 89–104. https://doi.org/10.1016/j.jhydrol.2005.02.041
Murdoch, T. Q., & Spittlehouse, D. L. (2011). Selecting and using climate change scenarios for British Columbia. Victoria, BC: University of Victoria.
Murphy, J. M., Sexton, D. M., Barnett, D. N., Jones, G. S., Webb, M. J., Collins, M., & Staniforth, D. A. (2004). Quantification of modelling uncertainties in a large ensemble of climate change simulations. Nature, 430, 768–772.
Nakicenovic, N., & Swart, R. (2000). IPCC Special Report on Emission Scenarios. Cambridge, UK: Cambridge University Press.
Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models. Journal of Hydrology, 10, 282–290.
O’Gorman, P. (2014). Contrasting responses of mean and extreme snowfall to climate change. Nature, 512, 416–418. https://doi.org/10.1038/nature13625
Olson, R., Fan, Y., & Evans, J. P. (2016). A Simple Method for Bayesian Model Averaging of Regional Climate Model Projections: Application to South-East Australian Temperatures. Geophysical Research Letters. https://doi.org/10.1002/2016GL069704
Pechlivanidis, I. G., Bosshard, T., Spângmyr, H., Lindström, G., Gustafsson, D., & Arheimer, B. (2014a). Uncertainty in the Swedish operational hydrological forecasting systems. In Beer, M., Au, S. K., & Hall, J. M. (Eds.), Vulnerability, uncertainty, and risk: Quantification, mitigation and management (pp. 253–262). Liverpool, UK: American Society of Civil Engineers.
Pechlivanidis, I. G., Jackson, B., McMillan, H., & Gupta, H. (2014b). Use of an entropy-based metric in multiobjective calibration to improve model performance. Water Resources Research, 50, 8066–8083. https://doi.org/10.1002/2013WR014537
Pechlivanidis, I. G., Jackson, B., McMillan, H., & Gupta, H. (2014v). Robust informational entropy-based descriptors of flow in catchment hydrology. Hydrological Science Journal, 59(1), 1–18. https://doi.org/10.1080/02626667.2014.983516
Pechlivanidis, I. G., Olsson, J., Bosshard, T., Sharma, D., & Sharma, K. (2016a). Multi-basin modelling of future hydrological fluxes in the Indian subcontinent. Water, 8(177), 1–21. https://doi.org/10.3390/w8050177
Pechlivanidis, I. G., Arheimer, B., Donnelly, C., Hundecha, Y., Huang, S., Aich, V., Samaniego, L., et al. (2017). Propagation of forcing and model uncertainty into hydrological drought characteristics in a multi-model century-long experiment in continental river basins. Climate Change, 141(3), 467–481. https://doi.org/10.1007/s10584-016-1723-0
Peters-Lidard, C. D., Clark, M., Samaniego, L., Verhoest, N. E. C., van Emmerik, T., Uijlenhoet, R., et al. (2017). Scaling, Similarity, and the Fourth Paradigm for Hydrology. Hydrology and Earth System Sciences, 21, 3701–3713. https://doi.org/10.5194/hess-21-3701-2017
Pirte, Z., Meyera, R., & Hamilton, A. (2010). What does it mean when climate models agree? A case for assessing independence among general circulation models. Environmental Science Policy, 13(5), 351–361. https://doi.org/10.1016/j.envsci.2010.04.004
Prudhomme, C., Giuntoli, I., Robinson, E. L., Clark, D. B., Arnell, N. W., Dankers, R., et al. (2014). Hydrological droughts in the 21st century, hotspots and uncertainties from a global multimodel ensemble experiment. Proceedings of National Academy of Sciences of United States of America, 111(9), 3262–3267. https://doi.org/10.1073/pnas.1222473110
Rudolf, E., Alfonsio, L., Di Baldassarre, G., Dottori, F., Russo, F., & Napoli, L. (2014). An entropy approach for the optimization of cross-section space for river modeling. Hydrological Sciences, 59(1), 126–137. https://doi.org/10.1002/hys.12268
Ruddell, B. L., & Kumar, P. (2009). Eohydrologic process networks: 1. Identification. Water Resources Research, 45, 1–22. https://doi.org/10.1029/2009WR007279
Salathe, E. P., Mote, P. W., & Wiley, M. W. (2007). Review of scenario selection and downscaling methods for the assessment of climate change impacts on hydrology in the United States pacific northwest. Journal of Climate, 20(7), 1611–1621. https://doi.org/10.1080/joc.1540
Samaniego, L., Kumar, R., Breuer L., Chamorro A., Flörke M., Pechlivanidis I. G., et al. (2017). Propagation of forcing and model uncertainty into hydrological drought characteristics in a multi-model century-long experiment in continental river basins. Climate Change, 141(3), 435–449. https://doi.org/10.1007/s10584-016-1778-y.
Shannon, C. E. (1948). A mathematical theory of communication. The Bell System Technical Journal, 27, 379–423. Part I and 623–656 Part II.
Singh, V. (2000). The entropy theory as a tool for modelling and decision-making in environmental and water resources. Water, 24(11), 1–12.
van der Linden, P., & Mitchell, J. F. B. F. B. (2009). ENSEMBLES: Climate change and its impacts: Summary of research and results from the ENSEMBLES project. Exeter, UK: Met Office Hadley Centre.
Vautard, R., Gobiet, A., Sobolowski, S., Kjellstrom, E., Stegehuis, A., Watkiss, P., et al. (2014). The European climate under a 2°C global warming. Environmental Research Letters, 9(3), 34006. https://doi.org/10.1088/1748-9326/9/3/034006
Weigel, A. P., Knutti, R., Liniger, M. A., & Appenzeller, C. (2010). Risks of model weighting in multimodel climate projections. Journal of Climate, 23(15), 4175–4191. https://doi.org/10.1175/2010JCLI3594.1
Weijs, S. V., van de Giesen, N., & Parlange, M. B. (2013). Data compression to define information content of hydrological time series. Hydrology and Earth System Sciences, 17(8), 3171–3187. https://doi.org/10.5194/hess-17-3171-2013
Wilby, R. L. (2010). Evaluating climate model outputs for hydrological applications. Hydrological Science Journal, 55(7), 1090–1093. https://doi.org/10.1080/02626667.2010.513212
Wilcke, R., & Baring, L. (2016). Selecting regional climate scenario for impact modelling studies. Environmental Modelling Software, 78, 191–201. https://doi.org/10.1016/j.envsoft.2016.01.002
Xu, H., Xu, C.-Y., Saitthun, N. R., Xu, Y., Zhou, B., & Chen, H. (2015). Entropy theory based multi-criteria resampling of rain gauge networks for hydrological modelling – a case study of humid area in southern China. Journal of Hydrology, 525, 138–151. https://doi.org/10.1016/j.jhydrol.2015.03.034
Yang, W., Andréasson, J., Graham, P. L., Olsson, J., Rosberg, J., & Wetterhall, F. (2010). Distribution-based scaling to improve usability of regional climate model projections for hydrological climate change impact studies. Hydrological Research, 41(3–4), 211–229. https://doi.org/10.2166/nh.2010.004