

## Application of bivariate mapping for hydrological classification and analysis of temporal change and scale effects in Switzerland

Matthias J.R. Speich <sup>a,\*</sup>, Luzi Bernhard <sup>a</sup>, Adriaan J. Teuling <sup>b</sup>, Massimiliano Zappa <sup>a</sup>

<sup>a</sup> Hydrological Forecasts, Swiss Federal Research Institute WSL, Zürcherstrasse 111, 8903 Birmensdorf, Switzerland

<sup>b</sup> Hydrology and Quantitative Water Management Group, Wageningen University, The Netherlands

### ARTICLE INFO

#### Article history:

Received 9 September 2014

Received in revised form 20 January 2015

Accepted 31 January 2015

Available online 7 February 2015

This manuscript was handled by Konstantine P. Georgakakos, Editor-in-Chief, with the assistance of Ashish Sharma, Associate Editor

#### Keywords:

Bivariate mapping  
Hydrological classification  
Map similarity  
Climate impact  
Hydrological regime  
Hydrological modeling

### ABSTRACT

Hydrological classification schemes are important tools for assessing the impacts of a changing climate on the hydrology of a region. In this paper, we present bivariate mapping as a simple means of classifying hydrological data for a quantitative and qualitative assessment of temporal change. Bivariate mapping consists of classifying map objects into discrete classes based on the values of two variables. We demonstrate the application of bivariate mapping to distributed hydro-climatic model outputs for the whole of Switzerland with a cell size of 200 m and compared the resulting bivariate maps with an existing classification of Swiss river regimes. The effects of scale were investigated by comparing these raster maps with a map showing the same variables aggregated to sub-basins with a mean area of 40 km<sup>2</sup>. Finally, maps of the current state were compared with predictions for future periods based on various model chains and greenhouse gas emission scenarios. For the map comparisons, four measures of association and two measures of agreement were used.

Of all the variable pairs tested, a bivariate map combining runoff and snowmelt contribution to runoff obtained the highest similarity scores with the map of river regimes, which suggests a strong link between the combination of these variables and intra-annual streamflow variations. Also, this classification offers new insights, as it includes absolute values of runoff, which are often ignored in classification schemes. Comparing current-state maps with future predictions indicated that the magnitude of change is reflected in the patterns of bivariate maps, with lower agreement scores for predictions further away in time or when higher greenhouse gas emissions are assumed. Furthermore, a visualization of the spatial distribution of agreement scores allows a qualitative assessment of the magnitude of change for different regions, and an analysis of the differences in spatial patterns of predictions based on different model chains or emission scenarios.

© 2015 Elsevier B.V. All rights reserved.

### 1. Introduction

To date, numerous classification schemes for hydrological data have been developed, using different methodologies and addressing a wide range of needs and research questions, usually with the aim to establish a framework to be used in further hydrological studies. Gottschalk (1985) notes that, as for any classification scheme for physiogeographic data, establishing such a system facilitates research so that “things can be named, generalizations developed and information exchanged”. In addition, the classification should have a predictive character, i.e. the criteria used for classification should be closely connected to the hydrological behavior of the region in question. Wagener et al. (2008) also stress

the need for predictive power in hydrological classification schemes, especially concerning insights into the impacts of land use and climate change on hydrology. Krasovskia (1997) states that such frameworks can be used in climate change impact studies both as a ground truth for control of model output and as a means to assess the magnitude of change. To fulfill this function, classifications should be based on strictly quantitative criteria. This has the further advantage of permitting frequent updating. A recent example for a climate impact study based on hydrological regime classification can be found in Coopersmith et al. (2014). In recent years, an important aim of hydrological classification has been to extend parameters or process information to ungauged basins (Sivapalan et al., 2003; Hrachowitz et al., 2013).

The fundamental unit of most classification schemes used in hydrology is the catchment or basin. Some schemes are based on streamflow seasonality, such as those developed by Aschwanden and Weingartner (1985) for Switzerland, Krasovskia (1997) for

\* Corresponding author.

E-mail address: [matthias.speich@wsl.ch](mailto:matthias.speich@wsl.ch) (M.J.R. Speich).

Scandinavia or Coopersmith et al. (2012) for the United States. In addition to the seasonality criterion, the classification scheme of Aschwanden and Weingartner (1985) contains an a priori regionalization based on the hydro-climatic region of Switzerland in which the catchment is located. For Alpine catchments, this classification scheme also includes catchment elevation and degree of glaciation. This scheme has been widely used in hydrological studies in Switzerland as well as in several climate impact studies (e.g. Horton et al., 2006; Weingartner et al., 2013; Köplin et al., 2014). Other classification schemes rely on long-term hydro-climatic values, such as the regions defined by Budyko (1974), or combine both approaches, like the global river regime classifications by Beckinsale (1969) and L'vovich (1979). However, more gridded climate, terrain and land cover data is available today and spatially distributed models are much more advanced, which allows hydrological studies at a much finer resolution. These developments have so far been largely ignored in the development of hydrological classification schemes.

We present bivariate mapping as a way of representing hydrological similarity in the era of distributed model applications. Bivariate mapping consists of classifying map objects, such as grid cells or polygons, into discrete classes based on the values of two variables. Bivariate classifications and the resulting maps have the advantage of being simple to understand and communicate (Olson, 1981). Such classifications can be used as tools for assessing temporal change, taking the roles described by Krasovskaja (1997) provided that they show a similar spatial pattern to those of river regime classifications. This is especially useful if the hydrological regimes in a region have not been classified, or if no streamflow data is available to generate such a classification. A classification based on bivariate mapping can easily be applied to different datasets, such as outputs of hydrological models for different periods or using different parameter values. Unlike classification schemes that focus on catchment classification, bivariate mapping can also be applied to other types of map units, such as grid cells. It is important to note, however, that such a classification is likely to be strongly affected by the level of aggregation of the data used. The setting of the number of classes and the position of class breaks, which are both arbitrary choices, may also affect the outcome of the classification. The effects of scale and classification methods on bivariate maps therefore need to be quantitatively and qualitatively investigated.

Only a limited amount of information can be input in such a map. We therefore use long-term average values describing the climatic conditions or the hydrological response of a region. To make the classification more useful, it is sensible to choose pairs of variables that integrate a wide range of catchment characteristics. Unlike many other classification schemes (e.g. Aschwanden and Weingartner, 1985; Weiskel et al., 2014), which are based on relative indices of streamflow seasonality or ratio variables, we propose classifications that include absolute values as criteria, to account for the high spatial heterogeneity of water balance terms in mountainous regions.

Bivariate mapping is a technique that has been applied in various earth system (e.g. UNEP, 1997; Teuling, 2011) and social science studies (e.g. Hall et al., 2008). To our knowledge, it has rarely been applied to hydrological data. In the only example known to us, Weiskel et al. (2014) propose a classification of hydrological units according to two ratio variables, namely the ratio of evapotranspiration to precipitation, and that of land-atmosphere water fluxes (green water) to horizontal flow (blue water). While that study was carried out at the continental scale and at a rather coarse resolution (mean basin area 138 km<sup>2</sup>), we focus on data at a finer resolution (gridded data with a cell size of 200 m) and more regionally specific processes.

While much of the research on bivariate mapping has focused on visualization issues (e.g. Trumbo, 1980; Leonowicz, 2006), we focus on the potential of classification schemes including only

two variables. Teuling (2011) produced a global map combining long-term average values for temperature and relative humidity to replicate the spatial patterns of the Köppen–Geiger classification (Köppen, 1918). Although this classification used a different methodology, both maps present similar spatial patterns. Teuling et al. (2011) note that bivariate mapping is particularly suitable in cases where two variables are physically linked, but with spatial variation in their correlation.

Various authors have identified precipitation and potential evapotranspiration (PET) as the main climate controls on hydrology (Budyko, 1974; Chapman, 1989; Wagener et al., 2007; Yadav et al., 2007). Together with runoff, which is the main descriptor of hydrological responses, these variables have been used for hydro-climatological classification. Budyko (1974) proposes the aridity index, i.e. the ratio of precipitation to PET, as a criterion to classify hydro-climatological regions.

In this paper, we examined the possible applications of bivariate mapping to hydrology, focusing on the quality of the information content of a classification based on two variables. Several combinations of variables were tested, selected after a review of the literature on hydrological classification. The spatial patterns of the bivariate maps were compared to those of an existing, well-known classification based on streamflow seasonality. A high degree of similarity indicates that the two variables chosen for the map are closely connected to intra-annual streamflow variations, and that the resulting bivariate classification can be considered an approximation for classifications based on hydrological regimes. We not only tested the application of bivariate mapping, but also the effects of varying the number of classes.

As noted before, much of the research on hydrological classification focused on catchments, i.e. on a different spatial scale than the one used in this study. As other authors have noted (e.g. Blöschl and Sivapalan, 1995), results and processes are not always transferrable between spatial scales. Representation of hydrological variables at different scales leads to different patterns in the spatial distribution of these variables, such as snow (Blöschl, 1999) or of processes, such as transpiration and interception by vegetation (Wood, 1998) or runoff generation (Cerdan et al., 2004). This is especially the case in mountainous regions, where spatial heterogeneity is particularly high (e.g. García-Ruiz et al., 2010). It is expected that the outcome of a classification will be different depending on the resolution of the input data. Therefore, this paper discusses the differences between a bivariate raster map with a cell size of 200 m and a map representing the same variables, aggregated to the level of sub-basins with a mean area of 40 km<sup>2</sup>.

Bivariate mapping also has the potential for assessing temporal change, as we demonstrated by using two measures of map agreement to compare maps of the current conditions with predictions made for future periods, based on various climatological and hydrological model chains and various greenhouse gas emission scenarios. These results provided the basis for a qualitative analysis of the effects of climate change on the hydrology of Switzerland, which we related to previous findings (Zappa and Bernhard, 2012; CH2014-Impacts, 2014; Köplin et al., 2014).

Various methods are available for quantifying the degree of similarity between two thematic maps. While some methods measure association, i.e. the similarity in the spatial patterns of two maps that use different classifications, other methods measure agreement, i.e. the degree of concordance of two maps that use the same classification. As for bivariate mapping, such measures have been successfully applied in other fields (e.g. remote sensing (Rees, 2008; Foody, 2008; Schimel et al., 2010), ecological modeling (Couto, 2003) or land-use change assessment (Chaudhuri and Clarke, 2014)), but only on few occasions in hydrology, for purposes other than those presented in this paper. For example, Wealands et al. (2005) review the potential of several measures of map agree-

ment for validating spatial predictions of hydrological models, while [Güntner et al. \(2004\)](#) and [Grabs et al. \(2009\)](#) use such measures to compare different wetness indices. In this paper, four different measures of association (Goodman and Kruskal's  $\lambda$  ([Goodman and Kruskal, 1954](#)), Theil's  $U$  ([Finn, 1993](#)), Cramér's  $V$  ([Cramér, 1946](#)) and Mapcurves ([Hargrove et al., 2006](#))) were used for assessing the similarity of bivariate maps with another classification and for assessing the effect of varying the number of classes. For assessing scale effects and temporal change, two measures of agreement (Cohen's  $\kappa$  ([Cohen, 1960](#)) and Fuzzy  $\kappa$  ([Hagen-Zanker, 2009](#))) were used. Using more than one measure in each case has the added benefit of providing a comparison of these methods in a framework in which they have rarely been applied.

## 2. Data

The input data for the maps generated in this paper was a grid data set of hydro-climatic variables modeled for the whole territory of Switzerland, including tributary catchments in neighboring countries at a resolution of 200 m. For this paper, the grids for precipitation, PET, snowmelt and runoff were used. This data set was originally used to assess the impact of climate change on the water balance of Switzerland ([CH2011, 2011; Zappa and Bernhard, 2012; Pflugshaupt, 2013](#)). The gridded values were calculated using the rainfall-runoff model PREVAH ([Viviroli et al., 2009](#)) in its spatially explicit version, as applied by [Schattan et al. \(2013\)](#) and represent long-term average values. The data set includes values for a reference period (1980–2009) as well as predictions for the periods 2021–2050 and 2070–2099, based on ten different climate model chains from the ENSEMBLES project ([van der Linden and Mitchell, 2009](#)), and using scenario A1B from the Special Report on Emissions Scenarios ([SRES; Nakicenovic and Swart, 2000](#)) (moderate increase in greenhouse gas emissions). The climate change signal was transferred to the temperature and precipitation data using the delta-change methodology ([Bosshard et al., 2011](#)). Similarly, predictions were calculated for the periods 2021–2035, 2045–2060 and 2070–2085 using the same reference data and three different greenhouse gas emission scenarios: the SRES scenarios A1B (increase of greenhouse gas emissions until the middle of the 21st century) and A2 (until the end of the century), as well as the scenario RCP3PD (strong mitigation; stabilization of the greenhouse gas concentration in the atmosphere) developed for the IPCC Fifth Assessment report (AR5; [IPCC, 2013](#)). The emission scenarios chosen are the same as those used in a report on climate change scenarios for Switzerland ([CH2011, 2011](#)). As the focus in this paper is on grouping hydrologically similar regions, the grid cells that are part of lakes were not considered in the classification.

The hydrological model was calibrated and validated using the runoff time series from 65 gauges, representing basins with a median area of 1085 km<sup>2</sup>. The gauges were selected to ensure that each of the eleven large-scale basins of Switzerland (see [Fig. 1](#)) is represented at least once. The calibration period covers 13 years (1984–1996), while the verification period consists of the years 1980–1983 and 1997–2009 (see also [Schattan et al., 2013; Kobierska et al., 2013; Jörg-Hess et al., 2014](#)). In order to quantify model performance, two measures, benchmark efficiency (BE) and volumetric deviation (DV), were used, as shown in [Fig. 1](#). BE, proposed by [Schaeefli and Gupta \(2007\)](#), is an extension of the Nash–Sutcliffe efficiency (NSE) ([Nash and Sutcliffe, 1970](#)) that compares the mean square error of a model simulation to the expected value of the squared deviations between the observed runoff and a selected benchmark model runoff. BE is calculated as follows:

$$BE = \frac{\sum_{t=1}^N [q_{obs}(t) - q_{sim}(t)]^2}{\sum_{t=1}^N [q_{obs}(t) - q_b(t)]^2} \quad (1)$$

where  $q_{obs}(t)$ ,  $q_{sim}(t)$  and  $q_b(t)$  represent the observed, simulated and benchmark model discharge, respectively, at time step  $t$ , for a simulation of length  $N$ . The benchmark model selected here, from [Jörg-Hess et al. \(2014\)](#), is the interannual mean discharge for every calendar day within a window of 30 days around the current day. This allows the strong seasonality of the time series to be taken into account, as NSE tends to overestimate model performance in such cases ([Schaeefli and Gupta, 2007](#)). Like NSE, BE is a normalized measure ranging from  $-\infty$  to 1, with 1 indicating a perfect simulation, and 0 indicating that the simulation does not perform better than the benchmark. DV is the volumetric deviation in percent between the simulated and observed runoff ([Viviroli et al., 2007; Zappa and Kan, 2007](#)). It is calculated as follows:

$$DV = \frac{\left| \frac{\sum_{t=1}^N q_{sim}(t)}{q_{obs}(t)} - 1 \right|}{N} \quad (2)$$

[Fig. 1](#) shows that, for most catchments, the BE score is between 0.5 and 1. Poorer scores are obtained in catchments of small or intermediate size that are subject to strong anthropogenic influences, such as hydropeaking (e.g. the Rhone basin) or artificial lake level regulation (Lake Geneva). Scores are particularly high in the lower part of the Rhine basin, where the catchment size is in the order of magnitude of 10,000 km<sup>2</sup>. The lakes upstream of these catchments buffer runoff, which helps to improve the model quality. For more details on calibration and verification of this simulation, see [Zappa et al. \(2012\)](#).

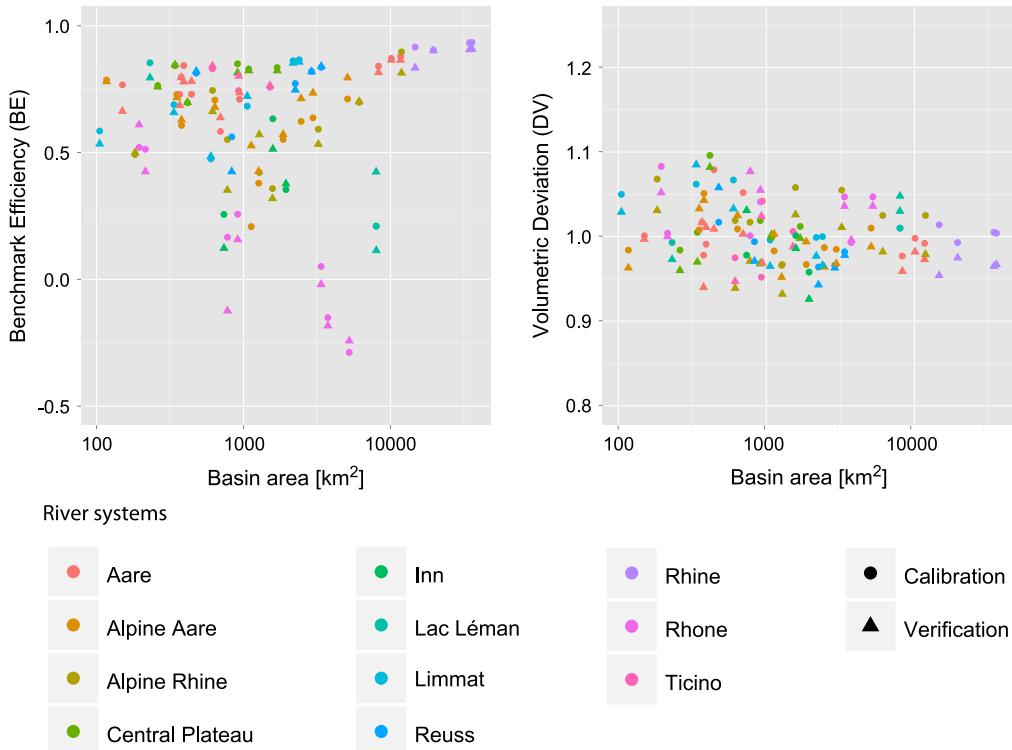
To investigate the effects of scale, the maps generated with these grid data sets were compared to maps where the values were averaged over the area of sub-basins with a mean area of 40 km<sup>2</sup>. The vector data of these sub-basins is part of a data set provided by the Swiss Federal Office of Environment ([Swiss Federal Office of Environment, 2012](#)). The same data set also contains the shapes of catchments with a mean area of 1.8 km<sup>2</sup>, together with their river regime class, according to the classification proposed by [Aschwanden and Weingartner \(1985\)](#). This data was used as a reference map to assess the similarity of bivariate maps with another classification scheme obtained with a different methodology. [Fig. 6\(a\)](#) shows a map of Swiss catchments (with a mean area of 1.8 km<sup>2</sup>) classified according to this method.

## 3. Methods

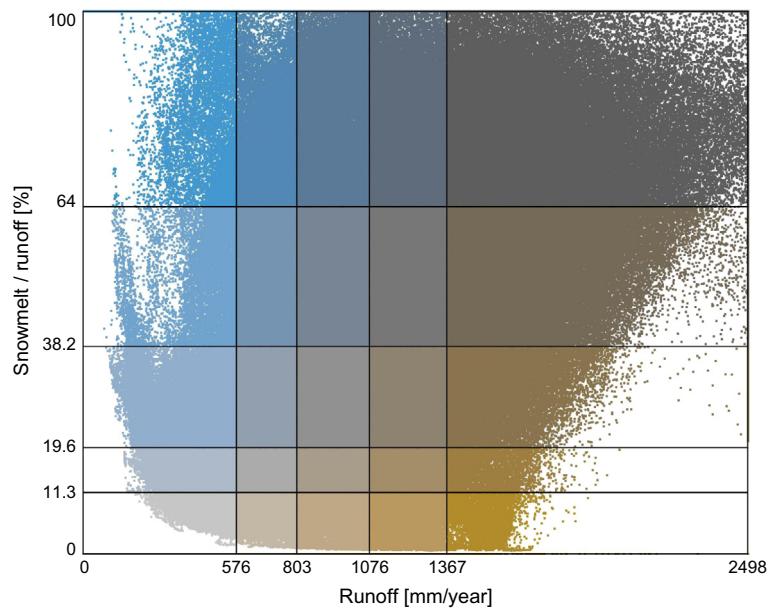
### 3.1. Bivariate mapping

Bivariate mapping involves assigning a category to spatial objects (such as grid cells or area polygons) based on the values of two variables. The color key of a bivariate map can be thought of as a scatter plot of the two variables in question, divided into discrete classes. The classification is made by specifying class breaks based on the univariate distribution of the observations. Here, quantile breaks were used (see [Fig. 2](#) for an example of such a classification). Each dot represents a grid cell and its position on the plot shows the values of runoff and snowmelt/runoff ratio for that cell. The horizontal and vertical lines are quantiles of each distribution. The colors in each square are used to represent the corresponding points on the map. The color schemes used for the bivariate maps here follow the recommendations of [Trumbo \(1980\)](#). See [Fig. 4\(c\)](#) for the resulting bivariate map. The numbers in the boxes of the color key indicate the percentage of grid cells that belong to each category. To make them more legible, all boxes were given the same size. The spacing of the class breaks does therefore not represent the actual interval between the quantiles.

The maps for future periods were all generated using the same class breaks as the corresponding map of the reference period, to make them directly comparable. The same applies to the maps at



**Fig. 1.** Benchmark efficiency and volumetric deviation scores obtained for the calibration and verification of PREVAH.



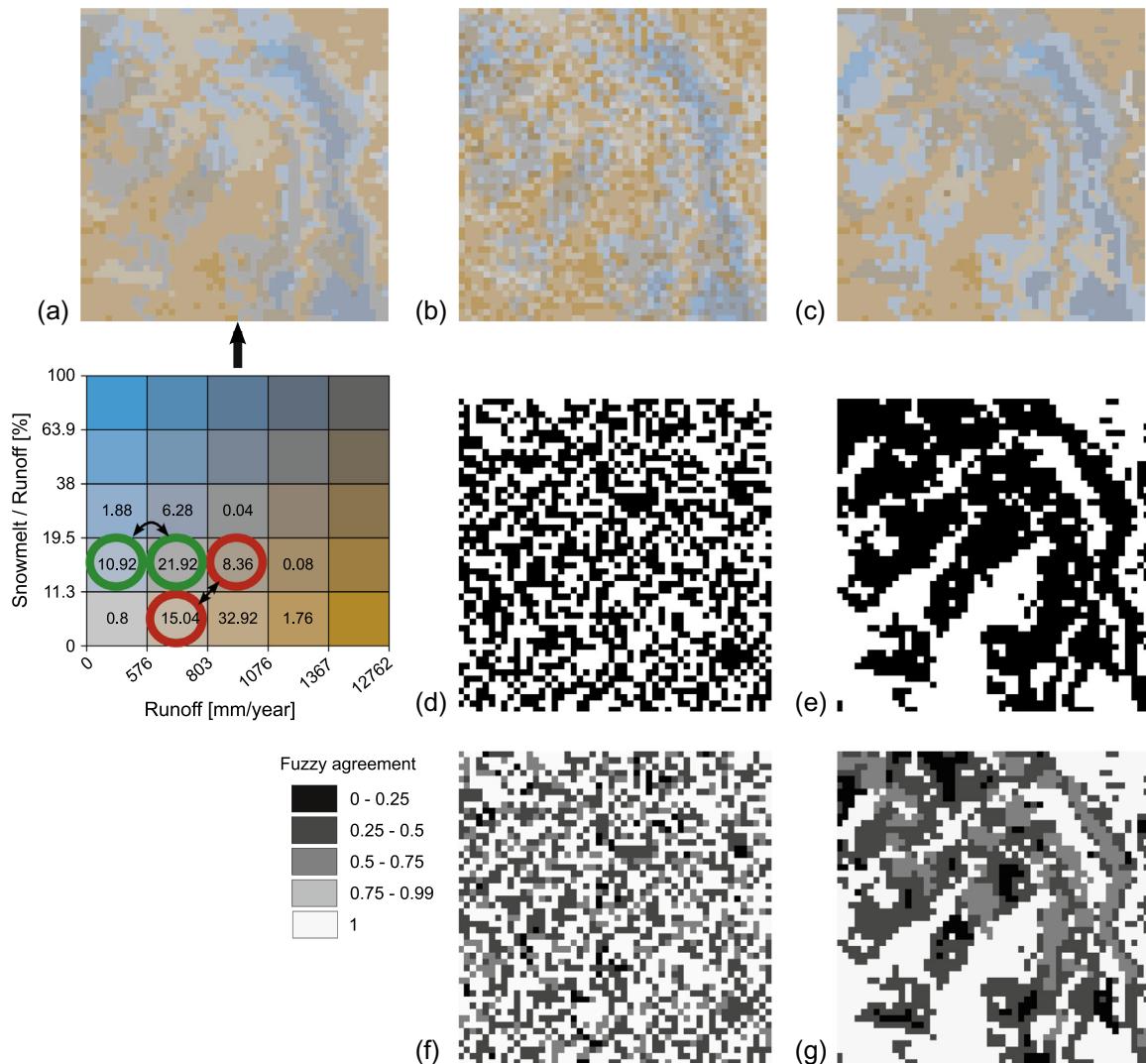
**Fig. 2.** Scatter plot of runoff and snowmelt percentage for the period 1980–2009. The horizontal and vertical lines represent quantiles of the two data sets and form the bivariate class breaks.

sub-basin scale, where the same class breaks were used as for the corresponding grid map.

### 3.2. Map similarity measures

To quantify the degree of similarity between categorical maps, various measures have been proposed. These measures typically return a score between 0 and 1, where 0 indicates complete dissimilarity and 1 perfect similarity (i.e. the two maps are identical).

Since there are relatively few examples of applications of these measures, it is difficult to interpret a single value. Furthermore, the score depends on the coarseness of the maps (e.g. Hargrove et al., 2006), and probably on the nature of the classifications, so that the similarity scores obtained cannot universally be compared with each other. Much of the research on map similarity comes from the field of remote sensing (e.g. Rees, 2008). At a given spatial resolution, a higher level of agreement can be expected if the classifications compared are from the same satellite picture than for



**Fig. 3.** Extract of a bivariate map (a) with random perturbations (b) and class swap (c), together with maps of crisp agreement (d and e) and fuzzy agreement (f and g). The numbers in the legend show the percentage of cells corresponding to each class in the original map (a). The colored circles show which classes are swapped in the second perturbed map (c).

more abstract classifications, such as those in this paper (see also [Foody, 2008](#)). To interpret such similarity scores better, several measures of associations were used and the results compared.

### 3.2.1. Measures of association

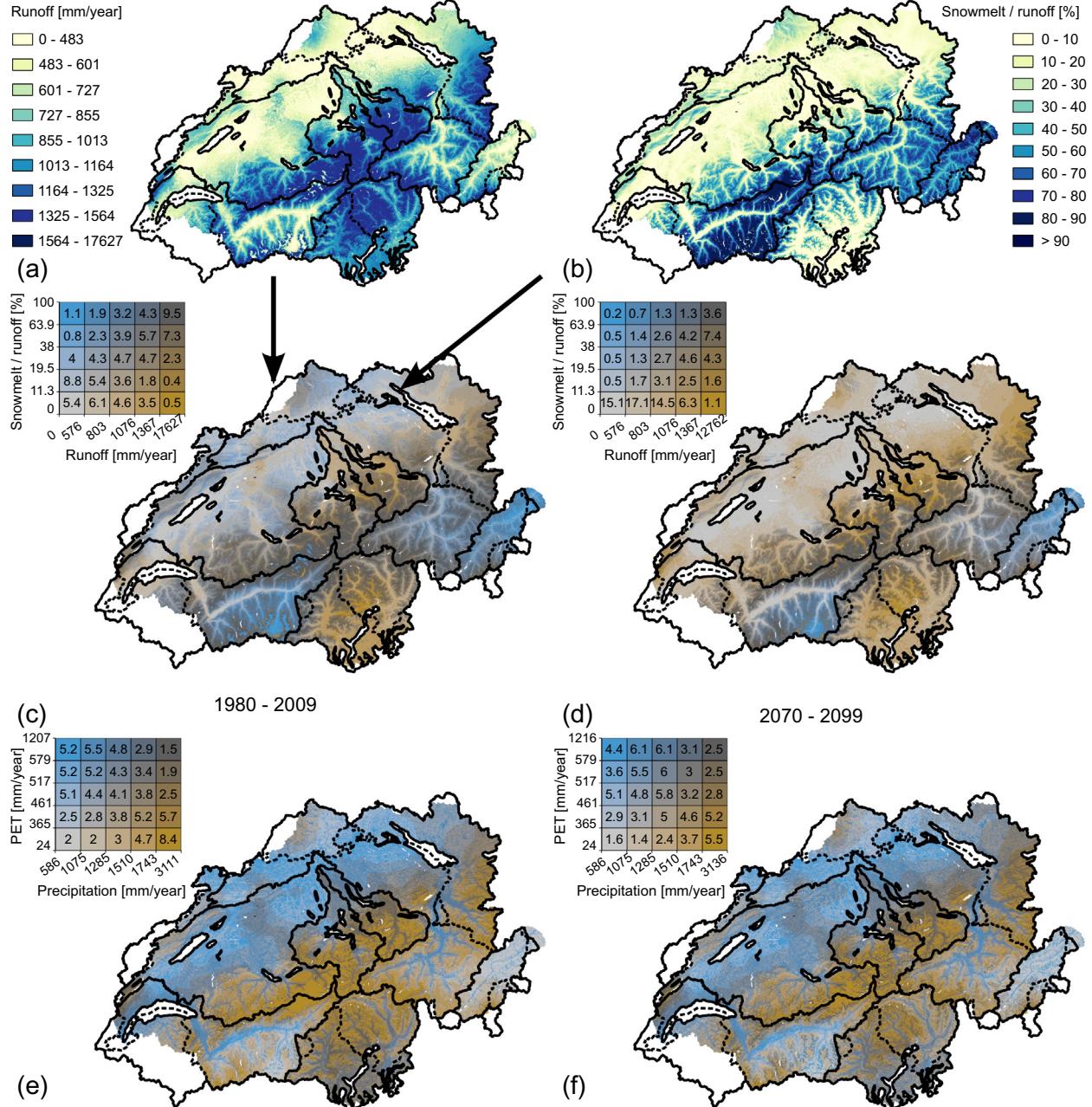
[Rees \(2008\)](#) applied and compared three different measures of association, all based on contingency tables: Cramér's V ( $V$ ), Theil's Uncertainty coefficient ( $U$ ) and Goodman and Kruskal's lambda ( $\lambda$ ). This comparison showed that all three measures are highly correlated and that they all provide a significant degree of discrimination between similar maps (in this case, different reclassifications of the same image) and less similar maps (reclassifications of images from the same area, but taken at an interval of ten years). All three measures were used in this paper.  $\lambda$  (see also [Rees, 2008; Schimel et al., 2010](#)), originally proposed by [Goodman and Kruskal \(1954\)](#), measures the reduction in probability of error in predicting the category of map A of any given map element if its category on map B is known, and vice-versa. In its symmetrical form,  $\lambda$  is calculated using the following equation:

$$\lambda = \frac{\sum_{i=1}^m \max_j(c_{ij}) + \sum_{j=1}^n \max_i(c_{ij}) - \max_j(c_{+j}) - \max_i(c_{i+})}{2N - \max_j(c_{+j}) - \max_i(c_{i+})} \quad (3)$$

where  $N$  corresponds to the total number of elements;  $m$  and  $n$  are the numbers of classes in map A and map B, respectively;  $i$  and  $j$  are indices iterating over the classes of map A and map B, respectively;  $c_{ij}$  is the number of cells belonging to class  $i$  in map A and to class  $j$  on map B;  $c_{i+}$  is the number of cells contained in category  $i$  on map A;  $c_{+j}$  is the number of cells contained in category  $j$  on map B;  $\max_i(c_{i+})$  is the number of elements in the modal class of map A, i. e. the class with the largest number of elements; and  $\max_j(c_{ij})$  the number of elements in the class of map B that has most elements in common with a given class of map A. The same notation conventions apply to all of the following equations.

Theil's  $U$  (see also [Rees, 2008; Schimel et al., 2010](#)) is another percent reduction in error measure, which, unlike  $\lambda$ , takes into account the whole distribution of the data, rather than using only the modal classes.  $U$  is based on the concepts of entropy and joint entropy as proposed by [Shannon \(1948\)](#) and of average mutual information as proposed by [Finn \(1993\)](#), (see also [Couto, 2003; Chen et al., 2008](#)). The entropy of a map is a way to quantify its information content, and is computed as follows for map A (in hartleys):

$$H(A) = -\sum_{i=1}^m \frac{c_{i+}}{N} \log \left( \frac{c_{i+}}{N} \right) \quad (4)$$



**Fig. 4.** Bivariate maps showing the distribution of runoff and snowmelt percentage (c and d) and precipitation and PET (e and f) for a reference period and a prediction in the far future. The univariate distributions of the two variables represented in (c) are shown in (a) and (b). Swiss GIS elements reproduced with the authorization of swisstopo (JA100118).

and likewise for map B:

$$H(B) = - \sum_{j=1}^n \frac{c_{+j}}{N} \log \left( \frac{c_{+j}}{N} \right). \quad (5)$$

The joint entropy  $H(A, B)$  measures the information content of the joint distribution of the classes on both maps, and is computed as follows:

$$H(A, B) = - \sum_{i=1}^m \sum_{j=1}^n \frac{c_{ij}}{N} \log \left( \frac{c_{ij}}{N} \right). \quad (6)$$

The average mutual information  $H(A; B)$ , quantifying the amount of information shared between both maps, is calculated based on the entropy of each map and the joint entropy:

$$H(A; B) = H(A) + H(B) - H(A, B) \quad (7)$$

The uncertainty coefficient is calculated using the following equation:

$$U = 2 \times \frac{H(A; B)}{H(A) + H(B)}. \quad (8)$$

Cramér's V (see also Agresti, 2002; Rees, 2008; Schimel et al., 2010), originally proposed by Cramér (1946), is a measure based on Pearson's  $\chi^2$  statistic, which is calculated as follows for the contingency table of the two maps:

$$\chi^2 = \sum_{i=1}^m \sum_{j=1}^n \frac{(c_{ij} - c_{i+} c_{+j} / N)^2}{c_{i+} c_{+j} / N}. \quad (9)$$

$V$  is a transformation of  $\chi^2$  that makes it independent of the number of categories and the number of map elements, and is defined as:

$$V = \sqrt{\frac{\chi^2}{N \times (\min(m, n) - 1)}} \quad (10)$$

The Mapcurves algorithm (see also Williams et al., 2008; Moore and Messina, 2010), proposed by Hargrove et al. (2006), is based on the degree of spatial overlap between the classes of two maps. For each pair of classes  $(i, j)$ , each from one of the two maps being compared, the algorithm calculates a goodness-of-fit measure according to the following equation:

$$GOF_{ij} = \frac{c_{ij}}{c_{i+}} \times \frac{c_{ij}}{c_{j+}} \quad (11)$$

The GOF values are then summed for each class of both maps:

$$G_{A,i} = \sum_{j=1}^n G_{ij} \quad (12)$$

for the classes of map A, and likewise for the classes of map B. The classwise GOF values are then sorted in ascendent order to obtain an array  $G'_A$ . To be able to integrate the function shown below, the values 0 and 1 are added at the beginning and the end of  $G'_A$ , respectively, so that the length of  $G'_A$  is now equal to  $m + 2$ . For each GOF value  $i \in G'_A$ , the Mapcurves function can be expressed as the fraction of classes  $1 \dots m$  that have a GOF value greater or equal to  $i$ :

$$f_A(i) = \frac{\sum_{k=1}^m [G_{A,k} \geq i]}{m}; i \in G'_A \quad (13)$$

and likewise for the classes of map B. The final Mapcurves score is obtained by integrating  $f(x)$  between zero and one. An exact value for the area under the curve for each interval can be obtained by applying the trapezoid rule:

$$MC_A = \sum_{i=1}^n (G'_{A,i+1} - G'_{A,i}) \times f_A(x+1) + \frac{(G'_{A,i+1} - G'_{A,i}) \times (f_A(x+1) - f_A(x))}{2} \quad (14)$$

In case of perfect fit, this returns a score of one. This procedure is illustrated graphically in Fig. 5. As the diagram shows, the result differs depending on whether the classes of one or the other map are plotted. Hargrove et al. (2006) postulated that the higher score is the one that describes the degree of association between the maps.

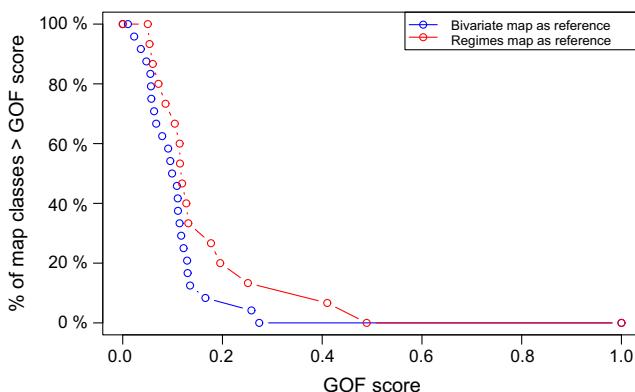


Fig. 5. Mapcurves diagram between Fig. 4(c) and a map of river regimes. Each dot on the curves represents the total GOF score of one map class.

### 3.2.2. Measures of agreement

A commonly used measure of agreement between two thematic maps is Cohen's  $\kappa$  (Cohen, 1960). It is based on the notion of overall accuracy, i.e. the proportion of elements that belong to the same class on both maps. Note that, unlike the measures of association described above, measures of agreement such as  $\kappa$  require an identical definition of classes on both maps. As a result, the contingency table must be square, i.e.  $m = n$ . The overall accuracy is computed as follows:

$$A = \frac{1}{N} \sum_{i=1}^n c_{ii} \quad (15)$$

where  $i$  represents classes of both maps. The overall accuracy is weighted by the expected agreement, defined as:

$$E = \frac{1}{N^2} \sum_{i=1}^n c_{i+} c_{+i}. \quad (16)$$

The  $\kappa$  statistic is calculated as follows:

$$\kappa = \frac{A - E}{1 - E} \quad (17)$$

Hagen-Zanker (2009) proposed an extension of Cohen's  $\kappa$  incorporating fuzzy logic, called Fuzzy  $\kappa$ , implemented in the software package Map Comparison Kit (Visser and de Nijs, 2006). More precisely, it incorporates the concepts of fuzziness of category (the idea that some pairs of map categories are semantically closer than others) and fuzziness of location (the idea that a cell belongs partially to a class, according to its distance from the closest cell of that class). Fuzziness of category is defined through a similarity matrix in which the degree of similarity is specified for each pair of classes, and fuzziness of location is defined through a distance decay function. The choices of the similarity values, the form of the distance decay function and the value of its parameters (neighborhood size and halving distance) are arbitrary. In this paper, Fuzzy  $\kappa$  was only used to compare bivariate maps of the same variables with each other. This means that the two maps compared have the same number of classes. It is in theory possible to specify a similarity matrix for two maps with a different set of classes (e.g. for a bivariate map and the map of river regimes), but the choice of similarity scores between the classes involves a much higher degree of arbitrariness. For the use of Fuzzy  $\kappa$  in this paper, the similarity matrix was defined so that immediately contiguous classes on the bivariate color key have a degree of similarity of 0.5. This means that, for example, a cell belonging to the class on the lower left corner of the color key also "belongs" to a degree of 50% to the classes to the right and above that class. Diagonally contiguous classes were defined as having a degree of similarity of 0.25. Fuzziness of location is described by an exponential decay function with a halving distance of two cells. The level of agreement is calculated for each grid cell  $c$  in map A with a category  $B_j$  from map B, using the following equation:

$$\mu_{c,B_j}^A = \max_{i=0}^m [M_{i,B_j} \times f(d_{ci}^A)] \quad (18)$$

and vice-versa for the other direction.  $M$  represents the similarity matrix,  $i$  and  $j$  indices for the classes of map A and B, respectively,  $f$  the distance decay function, and  $d_{ci}^A$  the distance of the current cell from a cell of category  $i$ . The level of agreement for each cell, called fuzzy agreement, is determined by the minimum of the scores  $\mu_{c,B_j}^A$  and  $\mu_{c,A_i}^B$ :

$$p_c^{A,B} = \min(\mu_{c,B_j}^A, \mu_{c,A_i}^B). \quad (19)$$

The mean agreement, analogous to the overall accuracy in Cohen's  $\kappa$ , is obtained by averaging the fuzzy agreement over all cells:

$$A_{fuzzy} = \frac{\sum_{c=1}^N p_c^{A,B}}{N}. \quad (20)$$

If fuzziness of categories is taken into account, the expected agreement needs to consider the agreement for each combination  $(i,j)$  of categories of map A and map B:

$$E_{fuzzy} = \sum_{i=1}^n \sum_{j=1}^n E(p | i,j) \times \frac{c_i + c_j}{N} \quad (21)$$

where  $E(p | i,j)$  is the expected agreement conditional to the presence of categories  $i$  and  $j$ . For a more detailed description of the expected agreement, we refer to Hagen-Zanker (2009). The  $\kappa_{fuzzy}$  statistic is computed analogously to Cohen's  $\kappa$ :

$$\kappa_{fuzzy} = \frac{A_{fuzzy} - E_{fuzzy}}{1 - E_{fuzzy}} \quad (22)$$

Unlike the other comparison methods, Fuzzy  $\kappa$  has the advantage of returning a value for each cell, which permits a precise mapping of the sources of agreement and disagreement between the maps. The other methods only return a value for each pair of classes, which makes a mapping of the results less precise. A requirement of all map comparison methods in the form in which they are presented here is that the two maps need to be co-registered, i.e. to cover the same area at the same resolution.

### 3.2.3. Example comparison

To clarify the purpose and the theoretical properties of the map comparison methods presented above, they were applied to an example map, shown in Fig. 3. The original map, in Fig. 3(a), is a 50 by 50 cells extract of a bivariate map generated as described in Section 3.1. As the legend shows, this extract contains only eleven of the 25 classes that can be found on the entire map. The meaning of the classes is not important for now. Two altered versions of this map were produced: to produce the map in Fig. 3 (b), half of the cells were assigned to another of the eleven present classes. The shift was done so that, with a few exceptions, the new class of the cell is immediately contiguous to its class on the original map. On Fig. 3(c), the cells belonging to four of the most frequent classes were swapped, so that the cells that belong to the category that comes second on the x-axis and first on the y-axis (noted hereafter as  $(X2, Y1)$ ) were assigned the class  $(X3, Y2)$ , and vice-versa. The same was done for the classes  $(X1, Y2)$  and  $(X2, Y2)$ .

The maps on Fig. 3(d) and (e) show the areas on which the original map agrees with the altered maps (b) and (c), respectively. White means that a cell belongs to the same class on both maps, and black means that they belong to different classes. This gives a first impression of the degree of agreement between the maps, and the spatial distribution of their sources of agreement and disagreement.

The fuzzy agreement, i.e. the cell-by-cell measure of agreement obtained during the calculation of fuzzy  $\kappa$ , is shown for both comparisons in Fig. 3(f) and (g). As described above, the similarity matrix defines a degree of similarity between map classes. Therefore, the spatial distribution of fuzzy agreement depends not only on crisp agreement or disagreement, but also on the defined degree of similarity between classes, as well as the presence of a class the neighborhood of a cell. For the second map, for example,  $(X2, Y1)$  and  $(X3, Y2)$  are diagonally contiguous on the bivariate legend, so they have a degree of similarity of 0.25. For  $(X1, Y2)$  and  $(X2, Y2)$ , who are immediately contiguous, the degree of similarity is 0.5. Therefore, most cells have a degree of fuzzy agreement of at least 0.25, even though overall crisp agreement is rather low.

Table 1 shows the values obtained by applying the various measures of association and agreement used in this paper to this example. For the first map, all measures show a mediocre fit, which is

indeed the case, since half of the cells have been assigned another class. Also, the random distribution of the class shifts disrupts the spatial pattern, which leads to low association as well as low agreement. For the second map, all measures of association indicate a perfect fit, while  $\kappa$  and fuzzy  $\kappa$  indicate a significant degree of disagreement. This is due to the fact that measures of association do not consider the definition of classes. Therefore, class  $(X2, Y1)$  on the original map shows a perfect fit with class  $(X3, Y2)$  on the second map, and vice-versa. The measures of agreement, on the other hand, are sensitive to class definition. Note that in both cases, fuzzy  $\kappa$  is significantly higher than  $\kappa$ . This is due to the fact that in both cases, the classes were replaced with semantically similar ones, and there is some degree of spatial correlation. Since fuzzy  $\kappa$  takes into account fuzziness of categories and of location, this contributed to increase the overall agreement score.

### 3.2.4. Comparison with random maps and regression analysis against mean difference

To be able to situate the results obtained on a scale, these measures were first used with randomly generated data sets. A bivariate map with 25 classes was compared to a series of 41 maps randomly and uniformly populated with 25 classes. To eliminate the effect of using different distributions, the same map was compared to another series of 41 randomly populated maps, on which the classes occur with the same frequency as on the reference map. Also, to investigate the dependency of association measures on the number of classes, the same bivariate map was compared with a series of maps randomly populated with 16 classes. Such a map is usually much finer than the bivariate map it is compared with, in the sense that it has few patches of contiguous cells of the same class, as the randomly populated cells are not spatially autocorrelated. Therefore, very low similarity scores are expected. The same bivariate map was also compared to a series of maps of the sub-basins described above, on which the sub-basins were assigned random values (i.e. all cells in each sub-basin were assigned the same random number). The polygon maps were rasterized, so that the comparisons could be done on a grid cell basis. Table 2 shows the expected values and variance for each series of map comparisons. The scores are all very low, but with considerable spread between the different measures. For example, the scores for  $V$  and Mapcurves are often one or more orders of magnitude higher than those for the other measures of association. The variance is very low in all cases. This suggests that, for maps of this size, scores higher than those shown here are unlikely to have been obtained by chance.

To determine the suitability of the proposed method for assessing temporal change, instead of using random maps, the agreement scores ( $\kappa$  or fuzzy  $\kappa$ ) obtained between current-state and future maps were compared to the mean values of the absolute difference of each variable by regression analysis. A linear regression analysis was performed, with the agreement score as the predicted variable and the mean value of the absolute difference of each variable between the two maps, calculated over all grid cells, as predictors. A high coefficient of determination ( $R^2$ ) indicates that the processes are correctly represented by the classification scheme and the method for quantification of map agreement. The results of this regression analysis are presented and discussed in Section 4.

**Table 1**  
Degree of association and agreement for the example maps in Fig. 3.

Example map	$\lambda$	$U$	$V$	MC	$\kappa$	Fuzzy $\kappa$
Gradual change	0.44	0.57	0.57	0.38	0.42	0.74
Class swap	1	1	1	1	0.31	0.71

**Table 2**

Expected value and variance of similarity scores obtained by comparing a bivariate map with randomly populated maps.

	$\lambda$	$U$	$V$	Mapcurves	$\kappa$	Fuzzy $\kappa$
25 Classes (uniform distribution)	7E-4 $\pm 5.5E - 9$	7.2E – 5 $\pm 1.8E - 11$	4.3E – 3 $\pm 1.7E - 8$	4E – 3 $\pm 7.3E - 12$	1.2E – 5 $\pm 2.8E - 8$	4.9E – 5 $\pm 4.8E - 8$
25 Classes (distribution of bivariate map)	2.6E – 6 $\pm 3.1E - 11$	7.5E – 5 $\pm 2.3E - 11$	4.3E – 3 $\pm 1.9E - 8$	4E – 2 $\pm 1.6E - 11$	3.4E – 5 $\pm 4.6E - 8$	9.8E – 5 $\pm 1.4E - 7$
16 Classes (uniform distribution)	8.1E – 4 $\pm 6.9E - 9$	4.9E – 5 $\pm 1.5E - 11$	4.3E – 3 $\pm 2.9E - 8$	6.2E – 2 $\pm 2.21E - 11$	–	–
25 Classes (sub-basins)	2.4E – 2 $\pm 1.1E - 5$	2.4E – 2 $\pm 3.8E - 6$	8.1E – 2 $\pm 3.3E - 7$	4.6E – 2 $\pm 1.9E - 5$	3.9E – 4 $\pm 4.9E - 5$	4.9E – 5 $\pm 5.8E - 5$

## 4. Results

The example of a bivariate map shown in Fig. 4(c) shows the distribution of runoff (on the  $x$ -axis of the color key) and snowmelt/runoff ratio (on the  $y$ -axis) in Switzerland. The high-mountain areas, with high runoff and high percentage of snowmelt, are distinguished from the inner-alpine valleys, where the runoff is lower but snowmelt percentage still high. The prealpine regions and the Jura mountains have medium values, whereas the valleys in the Southern and Central Alps have high runoff, but with less snowmelt contribution. The Central Plateau has rather low runoff with a low snowmelt percentage, but varies greatly. In the alpine regions, the accumulation and ablation zones of glaciers can be clearly seen on the map. Fig. 4(a) and (b) shows the univariate distributions of the two variables represented in panel c).

All bivariate maps presented in this paper were divided into 25 classes, with five classes on each axis. In order to assess the role of the number of classes, the map on Fig. 4(c) was compared to three maps with a different number of class breaks, chosen to illustrate the effects of this choice. The different measures of association obtained for these maps are shown in Table 3. All results are significantly higher than those obtained by comparison with random maps. The map with  $3 \times 5$  classes shows the highest level of similarity by all measures, as the univariate snowmelt/runoff classes are the same on both maps. The rather low degree of association for all these maps shows that the spatial patterns yielded by the classification are heavily dependent on the choice of class breaks and the number of classes.

Comparing the bivariate maps with different variable pairs and the map derived from Aschwanden and Weingartner (1985)'s classification (see Table 4) resulted in all cases to significantly higher scores than those obtained with random maps. However, the low scores indicate that all four maps are rather different from the distribution of river regimes. For the distribution of the Mapcurves GOF scores obtained by comparing the bivariate map in Fig. 4(c) and the river regimes map, see Fig. 6(b). Note that, unlike the maps of fuzzy agreement, the values shown do not apply to each cell separately, but to each pair of classes of the two maps. The highest scores were obtained with the pluvial regimes, mainly on the Central Plateau, but also at the bottom of the Rhone valley and ranged from 0.07 to 0.18. The scores for the nival and glacial regimes in the high-mountain regions were up to 0.08, and in the Prealpine valleys, the Jura mountains and the Southern alpine valleys up to 0.05. The ellipses on the graph in Fig. 6 give an indication of the

**Table 4**

Association scores obtained with the map of river regimes after Aschwanden and Weingartner (1985).

x-axis	y-axis	$\lambda$	$U$	$V$	Mapcurves
Runoff	Snowmelt percentage	0.2	0.26	0.32	0.15
Precipitation	PET	0.13	0.17	0.25	0.12
Aridity index	Runoff	0.19	0.18	0.25	0.12
Aridity index	Runoff ratio	0.2	0.19	0.25	0.11

distribution of the regime types in the runoff-snowmelt contribution space (see also Teuling et al., 2011). Each ellipse represents the bivariate normal density of the grid cells belonging to a regime type in the space defined in Fig. 2 (i.e. the runoff-snowmelt percentage space), with a confidence level of 0.68. While this figure shows a certain degree of association, there is still too much scatter to permit a meaningful visualization of these distributions with a higher confidence level. Three large clusters can be distinguished: The Northern pluvial regimes on the bottom left, the Southern regimes on the bottom right, and the glacial and nival regimes, with considerable scatter and overlap, in the upper half of the plot. Scatter plots for each individual regime type, including the corresponding density ellipse, are given in Appendix B.

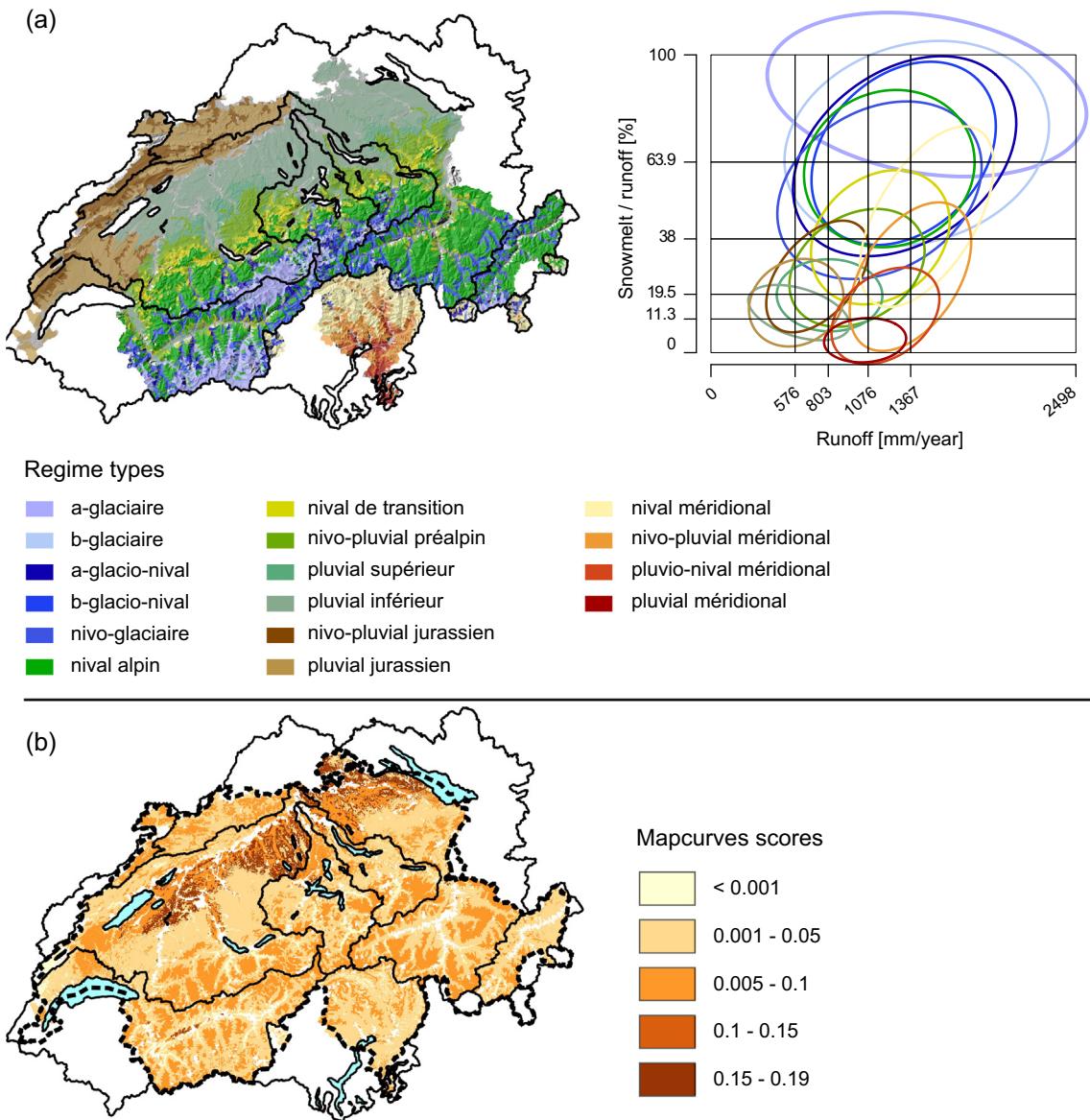
The levels of agreement obtained by comparing between gridded bivariate maps and maps at the sub-basin scale are shown in Table 5 for two variable pairs. Fig. 7 shows a detail of the map resulting from a comparison of the bivariate map in Fig. 4(c) with the corresponding polygon map using fuzzy  $\kappa$ . The outlines of sub-basins (median area  $40 \text{ km}^2$ ) are shown in blue. The map shows the region surrounding the Aletsch glacier in the Swiss Alps. The upper and lower parts of the glacier have very low fuzzy agreement, indicating a low level of similarity. On the other hand, the zone situated between 2500 and 3500 m mostly contains values of 1. Indeed, averaging the values for the single variables places this sub-basin in the top right category (very high runoff and very high snowmelt contribution). The upper parts of the sub-basin, however, are characterized by accumulation (low runoff), whereas the part of the glacier below 2500 m represents the ablation zone (high runoff, but low snowmelt contribution).

The distribution of the agreement scores obtained by comparing the current-state bivariate maps with future predictions is shown in Fig. 8 for the different model chains, and in Fig. 9 for the three emission scenarios. For the table with agreement scores for all scenarios and GCM–RCM model chains, see Appendix A. For all model chains, the agreement is lower for distant predictions than for those in the near future. The combination of runoff and snowmelt percentage appears to be more sensitive than the combination of precipitation and PET, as the scores are lower for the former in all cases. Concerning the scenarios, the scores for scenarios A1B and A2 both decrease through the three periods, while those for RCP3PD increase, suggesting a shift again towards current conditions from the middle of the 21st century.

**Table 3**

Association scores obtained by comparing the bivariate map in Fig. 4(c) with bivariate maps with a different number of classes.

Number of classes	$\lambda$	$U$	$V$	Mapcurves
4 × 4	0.53	0.67	0.71	0.54
3 × 5	0.72	0.82	0.88	0.79
3 × 3	0.49	0.61	0.78	0.64



**Fig. 6.** Spatial distribution of Mapcurves scores between Fig. 4c) and the distribution of river regimes after Aschwanden and Weingartner (1985). Swiss GIS elements reproduced with the authorization of swisstopo (JA100118).

**Table 5**  
Agreement scores for the gridded bivariate maps and the maps at sub-basin scale.

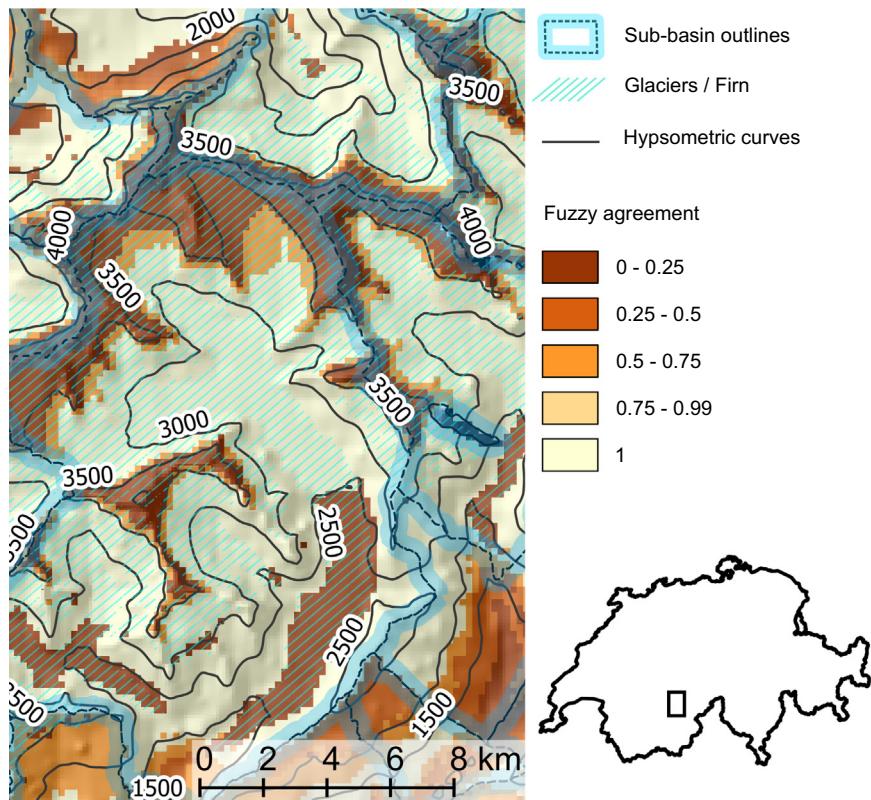
x-axis	y-axis	$\kappa$	Fuzzy $\kappa$
Runoff	Snowmelt percentage	0.35	0.54
Precipitation	PET	0.28	0.47

Table 6 shows the adjusted  $R^2$  values obtained by linear regression of agreement scores given mean absolute difference in one or both of the variables. For example, the linear least squares regression obtained with mean absolute difference in runoff being the only explanatory variable explains 68.02% of the variation of the fuzzy  $\kappa$  scores. When the snowmelt/runoff ratio is considered instead of runoff, 94.37% of the variation can be explained. When the mean absolute differences of both variables are used as explanatory variables, 97.84% of the variation can be explained.

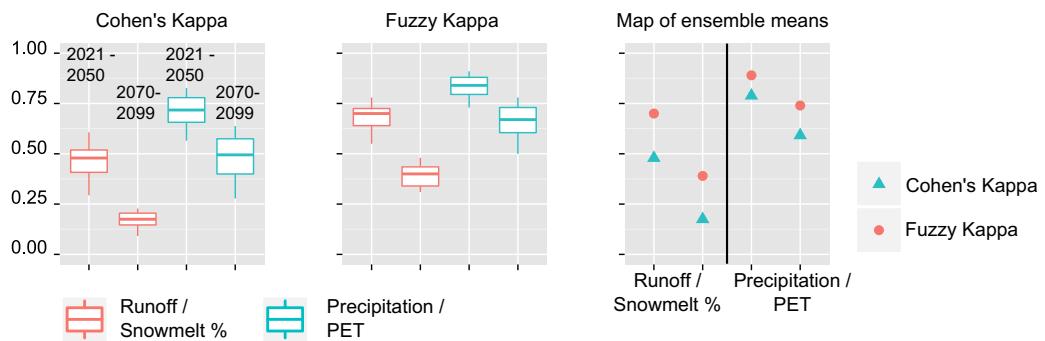
Fig. 10 shows the spatial distribution of the fuzzy agreement obtained by comparing bivariate maps for two future periods with the control period. The variable pairs chosen are the same as in

Fig. 4, and the future maps use the average values of all ten model chains. A fuzzy agreement score of 1 indicates that the classes on that cell are the same in the control period and the future. A lower score indicates a higher degree of disagreement between the maps, i.e. a more pronounced change. The visual impression confirms the tendency of the overall fuzzy  $\kappa$  scores for the intensity of change to be higher for the distant future than for the near future, as well as for the variable pair runoff/snowmelt percentage to be more sensitive than precipitation/PET. For the former, the lowest scores are found around the Alpine glaciers as well as the sub-alpine Prealps and Jura regions, while regions at lower elevations and at the bottom of valleys tend to keep their class. A visual comparison of the maps in Fig. 4(c) and (d) shows a progression of pluvially dominated classes towards higher regions.

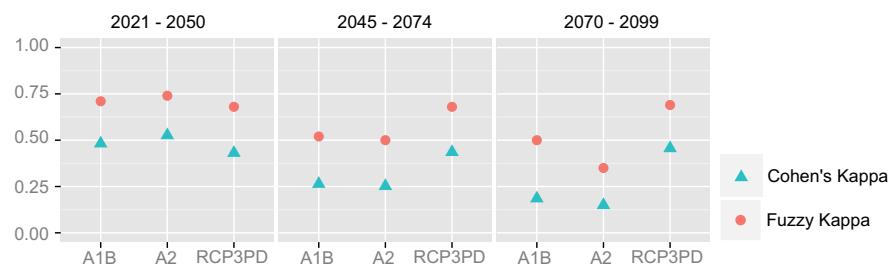
The similarity measures can also be used to compare maps classified according to different assumptions. For example, Fig. 11 shows two bivariate maps for the far future, based on two different emission scenarios, together with a fuzzy agreement map. A visual assessment of the two bivariate maps and the relative frequencies of the classes shows that the A2 map indicates a stronger change



**Fig. 7.** Detail of Fuzzy agreement map resulting from comparing gridded and sub-basin scale bivariate maps of runoff and snowmelt percentage. Swiss GIS elements reproduced with the authorization of swisstopo (JA100118).



**Fig. 8.** Distribution of agreement scores on comparing bivariate maps for the near (2021–2050) and far future (2070–2099) with the reference period (1980–2009).



**Fig. 9.** Distribution of agreement scores on comparing bivariate maps of runoff/snowmelt percentage for the near, intermediate and far future with the reference period (1980–2009).

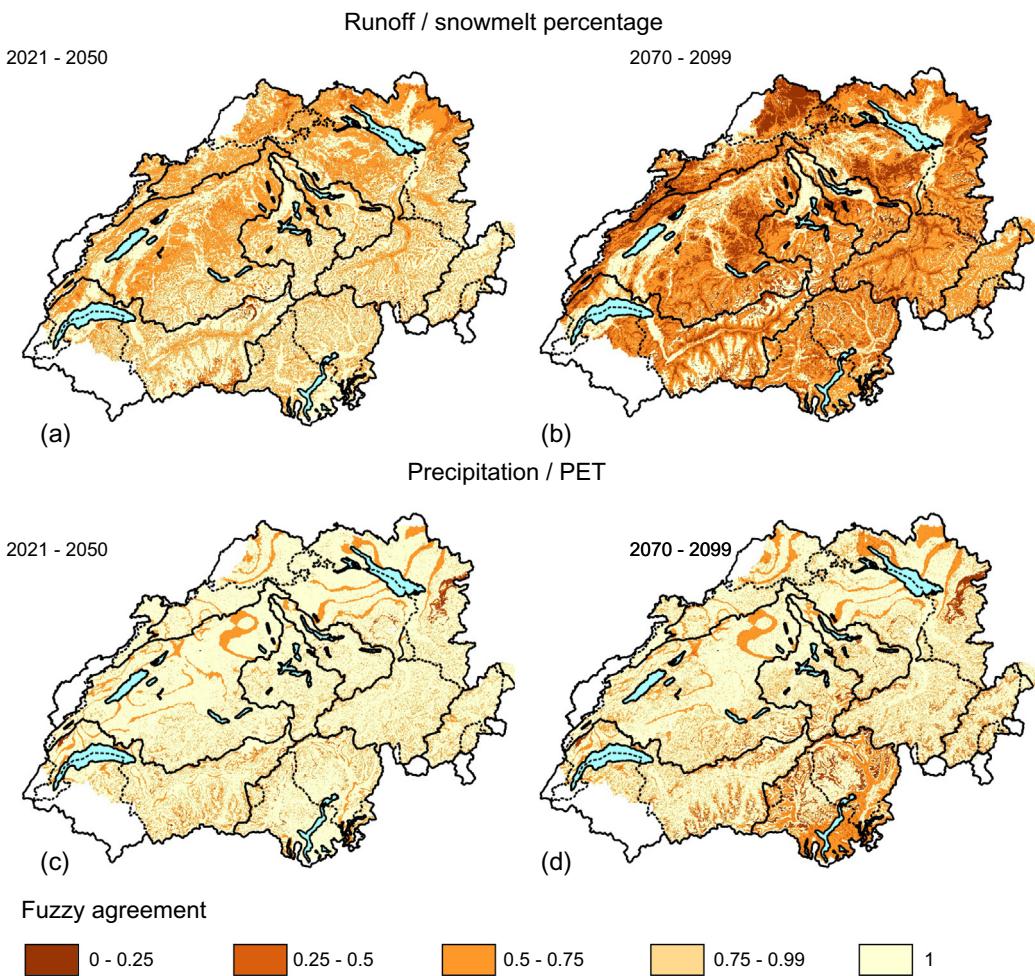
than the RCP3PD map, which has a Fuzzy  $\kappa$  value of 0.69 when compared to the control period, whereas the former returns a score of 0.35. A direct comparison of these two scenario maps returns an

overall score of 0.61. As the fuzzy agreement map in (c) illustrates, the maps agree well in some areas, especially the low regions on the Swiss Plateau and at the bottom of valleys, while they disagree,

**Table 6**

Adjusted coefficients of determination ( $R^2$ ) obtained by linear regression of similarity scores against mean absolute difference of both variables between current-state and future prediction maps.

Predictors (mean absolute difference)	$R^2$ for $\kappa$	$R^2$ for fuzzy $\kappa$
Runoff and Snowmelt/runoff	0.9685	0.9784
Runoff	0.7272	0.6802
Snowmelt/runoff	0.9088	0.9437
Precipitation and PET	0.995	0.9967
Precipitation	0.9597	0.9628
PET	0.7715	0.7695

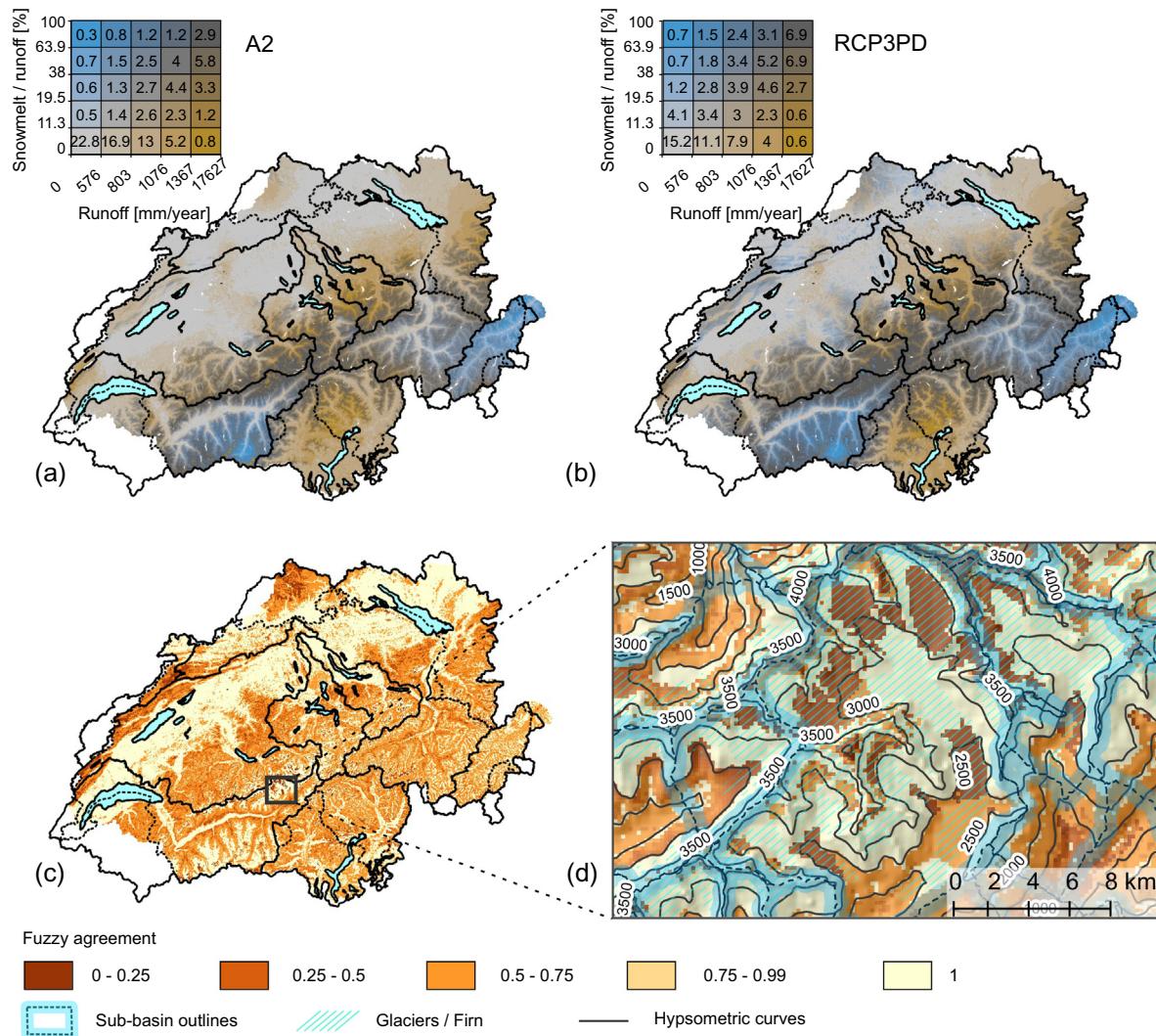


**Fig. 10.** Distribution of Fuzzy agreement obtained by comparing bivariate maps for future predictions with the control period (1980–2009). Swiss GIS elements reproduced with the authorization of swisstopo (JA100118).

much like in Fig. 10, in glaciated areas and sub-alpine regions (see the detail of the Aletsch region in Fig. 11(d)). Three classes are strongly represented in glaciated areas: low runoff and high snowmelt contribution (top left class on the color key) where accumulation is dominant, high runoff and low snowmelt contribution in the ablation zone (bottom right) and high runoff and high snowmelt contribution (top right) in the intermediate part of the glacier. In this part, both maps agree to a large extent, i.e. most cells between 2500 and 3000 m show perfect agreement (fuzzy agreement = 1). The extent of the accumulation and ablation areas, on the other hand, varies considerably between the two maps, since the A2 scenario leads to a much faster glacier melt. Since the classes corresponding to these areas are not contiguous on the classification scheme, the fuzzy agreement in these areas is very low.

## 5. Discussion

Using bivariate mapping explicitly for classification is a rather uncommon approach, as much of the literature considers this technique mainly as a visualization tool (e.g. Trumbo, 1980; Olson, 1981; UNEP, 1997; Leonowicz, 2006). The examples described in this paper illustrate the advantages of such an approach in providing a transparent and easily understandable classification scheme which can easily be updated and adapted to other data sets (Krasovskia, 1997). The choice of variables should of course be adapted to the region in question and the purpose of the classification. In the case of Switzerland, the spatial distribution of snowmelt provides valuable information. It is closely tied to temperature and elevation, which means that a certain degree of



**Fig. 11.** Comparison of bivariate maps for the period 2070–2099 based on the scenarios A2 and RCP3PD, including a detail of the Aletsch region. Swiss GIS elements reproduced with the authorization of swisstopo (JA100118).

correlation with other important variables, such as total runoff or evapotranspiration, is likely. Snowmelt strongly influences intra-annual streamflow variability and defines the hydrological regime, as [Aschwanden and Weingartner \(1985\)](#) show. They claim the form of precipitation is an important classification criterion for hydrological regimes. Combining the snowmelt percentage with absolute values of runoff yields a sensible classification. In addition, this variable pair appears to be rather sensitive to climate change. This makes it suitable for comparing future predictions. The choice of the number of classes and the position of class breaks, however, is arbitrary. Quantile breaks were used for all bivariate maps discussed in this paper. Comparing these maps with a series of maps with a smaller number of classes revealed a high degree of dissimilarity, with a  $\lambda$  score as low as 0.49 for a map of nine classes. Therefore, the choice of classes should be consistent across all maps used within a study. The classification schemes can be easily adapted depending on the research question and the information available.

The low scores obtained by comparing bivariate maps with river regimes can partly be explained by the different scales of the maps, since the river regimes are defined at the scale of catchments, and hence do not show the same degree of spatial variability as the grid maps. The geographical dependence inherent in [Aschwanden and](#)

[Weingartner \(1985\)](#)'s classification scheme also plays a role. The density plots in Fig. 6 reveal both similarities and differences between the two classification schemes. On one hand, individual regime types, such as the *pluvial inférieur* and *pluvial méridional* regimes, have little spread, indicating that the two schemes offer a similar distinction for these cases. On the other hand, for the glacial and nival regimes, the spread is higher, and the overlap between these classes is considerable. This suggests that the two classification schemes give rather similar spatial patterns for pluvially dominated regime types at low elevations, and quite different patterns for glacial and nival regime types in Alpine regions. An advantage of bivariate classifications is that they do not rely on subjective criteria, such as the a priori classification based on geographical location. Another methodological difference between the classification scheme applied here and that of [Aschwanden and Weingartner \(1985\)](#) is that in the former, absolute values for water balance terms are included in the classification, while the latter is based on relative indices of streamflow seasonality. The high spatial heterogeneity of runoff, as shown in Fig. 4, suggests that absolute values of runoff and other water balance terms are an important descriptor of the hydrological properties of a region.

Comparing gridded maps with maps aggregated at catchment scale revealed an effect of classification at different scales: an

aggregation on a larger scale hides much of a catchment's spatial variability, with considerable effects on the spatial pattern of the classification. This effect is not peculiar to bivariate classification, but is a well acknowledged issue in hydrology (e.g. Blöschl and Sivapalan, 1995) and a problem inherent to all spatial analysis. Much of the research on scale issues in hydrology focused on modeling, suggesting that models may yield significantly different results when applied at different spatial scales (e.g. Wood, 1998). In this paper, this issue is illustrated by considering glaciated catchments, where the variability in runoff and snowmelt is particularly high. However, similar effects have been observed in other catchments, especially where there is a strong elevation gradient. While the example shown here considers the rather trivial example of catchment-scale values obtained by spatial aggregation of distributed model results, it illustrates the use of the proposed method for comparing hydrological data across spatial scales. This may be applied to comparisons of lumped and distributed model outputs, or to comparisons of classifications of distributed classifications, such as the one shown in this paper, with similar classifications done at a coarser resolution, such as that proposed by Weiskel et al. (2014).

The different measures of association used for comparing classification schemes all behaved in a similar way. While the number of comparisons does not permit a detailed statistical analysis, they seem broadly correlated, and the scores for comparisons of different classification schemes are all significantly higher than those obtained with randomly populated maps. It is also worth noting that Mapcurves and  $V$  returned significantly higher scores than two of the other measures for random maps, which might indicate that they are less reliable than the other methods.

The agreement scores obtained by comparing current-state maps with future predictions show that both  $\kappa$  and fuzzy  $\kappa$  are good measures for the magnitude of change. This is confirmed both intuitively and through statistical analysis. The high  $R^2$  values in Table 6 suggest that the intensity of the process is represented correctly in the methods of discrete classification and map comparison applied here. It is worth noting that the mean absolute difference of individual variables also give a high  $R^2$  score, suggesting that a similar assessment of temporal change based on the distribution of only one variable could also be valuable. However, for both variable pairs, using the mean absolute differences of both variables as predictors improves the adjusted  $R^2$  value. It is also worth considering that in both cases, the  $R^2$  values obtained for one of the variables are significantly higher than for the other. This suggests that the snowmelt/runoff ratio is a stronger driver of overall change than runoff, and likewise for precipitation and PET, respectively. A visual assessment of the maps in Fig. 4(c) and (d) confirms this impression, as the evolution of the spatial pattern is driven mostly by transition to classes with a smaller snowmelt/runoff ratio. However, the absolute value of runoff is important for assessing local effects, e.g. transition between accumulation and ablation zones of a glacier.

The scores of agreement obtained for  $\kappa$  and fuzzy  $\kappa$  are highly correlated ( $R = 0.98$ ), which is not surprising, considering their theoretical similarity. Fuzzy  $\kappa$  scores tend to be higher than  $\kappa$ , as it allows partial agreement in cases where  $\kappa$  only distinguishes between correct and incorrect classification. Indeed, since the classes are often replaced by similar ones, and there is often a certain degree of spatial correlation between the old and the new class, the consideration of fuzziness of categories and location leads to a higher overall agreement score. Fuzzy  $\kappa$  seems to be more suitable for the purpose of this study. Indeed, it is sensible to consider fuzziness of categories and of location. The definition of the similarity matrix, as well as the choice of a distance decay function, are themselves arbitrary, but can still help to reduce

the effect of separating elements into classes based on arbitrary class breaks. Furthermore, for assessment of temporal change, it is useful to consider that some pairs of classes are semantically closer than others. Fuzzy agreement maps are a useful means to describe the distribution of difference between two bivariate maps, but as the scores indicate only the degree of agreement, without any indication of which classes are present in each grid cell, it is also necessary to consider the two maps being compared when analyzing the result.

It is also interesting to note that, in this case, the measures of map association showed a similar development as the measures of agreement, i.e. decreasing scores for a higher intensity of change (see Appendix C for the corresponding box plots, and the correlation coefficients with fuzzy  $\kappa$ ). This would suggest that these measures are also appropriate for assessing temporal change. However, due to their properties, this effect cannot always be expected. For example, if all cells between two maps shift classes, the degree of association between the maps might still be quite high, while the degree of agreement is much lower (see also Rees, 2008). A possible explanation for the similar behavior of map association and agreement in this study is that the class shift investigated here is gradual in most cases, with a large proportion of cells keeping their original class. In other words, the change in spatial pattern in such cases is more likely to look like in Fig. 3(b) than in Fig. 3(c).

Maps based on different scenarios can, for example, be compared to analyze the agreement and disagreement between model outputs based on different assumptions. Both the bivariate maps as well as the maps of fuzzy agreement must be compared for a complete analysis. The region around the Aletsch glacier was chosen as an example, but the effects in other regions are different. For example, the expansion of classes with low values for both variables is less pronounced in middle elevation regions. This translates into a higher prevalence of blue classes in the Prealps and Jura mountains on the RCP3PD map, and lower fuzzy agreement in both regions.

The application of these techniques to hydrological data in Switzerland focused on the combinations runoff/snowmelt contribution and precipitation/PET. The former combination appears to be more sensitive to expected future changes in temperature and precipitation, as it returns lower similarity scores. The maps on Fig. 10(a) and (b), together with those on Fig. 4(c) and (d), show a strong decrease of the number of cells with high snowmelt contribution to runoff, and a strong increase of the number of cells with low runoff and low snowmelt contribution. On the maps, this translates into an expansion of the pluvially dominated regimes with low runoff, which occur only in some parts of the Plateau and at the bottom of large valleys during the control period, to higher regions of the Plateau, but also in the Jura and Prealpine mountains. In Fig. 4(d), this is illustrated by the quasi-total absence of classes with high snowmelt (blue to dark gray) outside the Alpine regions. Glacier melt is expressed by a reduced glaciated surface, but also a shift of the extent of the accumulation and ablation zones. On Fig. 4(c), classes with low runoff and a high snowmelt contribution are typical for inner-alpine valleys. This is still the case for the period 2070–2099, but to a lesser extent. Indeed, in the areas closer to the valley bottoms, these classes are also replaced by pluvial classes.

The maps of precipitation and PET show a slight increase in precipitation North of the Alps, and a slight decrease South of the Alps. The increase takes place mainly early in the century and results in the expansion of classes with higher precipitation, visible in Fig. 10 in the form of stripes with lower fuzzy agreement on the Plateau and in the Jura mountains. In most cases, the scores do not go below 0.5, which suggests a moderate shift, as classes are being replaced by directly contiguous ones. This also indicates that this shift occurs only in one dimension, with no class shift for PET. The decrease in precipitation South of the Alps is visible in the

**Table A.7**

Agreement scores obtained by comparing current state maps with future predictions based on different model chains.

Model chain	Period	$\kappa$	Fuzzy $\kappa$
<i>Runoff/Snowmelt contribution to runoff</i>			
CNRM ARPEGE ALADIN	2021–2050	0.53	0.73
	2070–2099	0.15	0.42
DMI ECHAM5 HIRHAM	2021–2050	0.60	0.78
	2070–2099	0.21	0.44
ETHZ HadCM3Q0 CLM	2021–2050	0.29	0.55
	2070–2099	0.09	0.31
HC HadCM3Q0 HadRM3Q0	2021–2050	0.35	0.6
	2070–2099	0.15	0.34
ICTP ECHAM5 REGCM	2021–2050	0.51	0.72
	2070–2099	0.23	0.45
KNMI ECHAM5 RACMO	2021–2050	0.44	0.67
	2070–2099	0.13	0.34
MPI ECHAM5 REMO	2021–2050	0.48	0.70
	2070–2099	0.17	0.39
SMHI BCM RCA	2021–2050	0.56	0.75
	2070–2099	0.22	0.48
SMHI ECHAM5 RCA	2021–2050	0.47	0.69
	2070–2099	0.19	0.40
SMHI HadCM3Q3 RCA	2021–2050	0.37	0.61
	2070–2099	0.15	0.33
<i>Map of ensemble means</i>	2021–2050	0.49	0.71
	2070–2099	0.20	0.43
<i>Precipitation/PET</i>			
CNRM ARPEGE ALADIN	2021–2050	0.83	0.91
	2070–2099	0.28	0.5
DMI ECHAM5 HIRHAM	2021–2050	0.76	0.86
	2070–2099	0.45	0.65
ETHZ HadCM3Q0 CLM	2021–2050	0.58	0.73
	2070–2099	0.34	0.55
HC HadCM3Q0 HadRM3Q0	2021–2050	0.66	0.80
	2070–2099	0.49	0.67
ICTP ECHAM5 REGCM	2021–2050	0.72	0.84
	2070–2099	0.61	0.76
KNMI ECHAM5 RACMO	2021–2050	0.65	0.79
	2070–2099	0.37	0.58
MPI ECHAM5 REMO	2021–2050	0.79	0.89
	2070–2099	0.59	0.74
SMHI BCM RCA	2021–2050	0.77	0.87
	2070–2099	0.54	0.70
SMHI ECHAM5 RCA	2021–2050	0.71	0.83
	2070–2099	0.56	0.72
SMHI HadCM3Q3 RCA	2021–2050	0.57	0.73
	2070–2099	0.43	0.63
<i>Map of ensemble means</i>	2021–2050	0.79	0.89
	2070–2099	0.64	0.78

far future (2070–2099), mostly at the bottom of valleys. Fig. 10(c) and (d) also show low fuzzy agreement along mountain ridges. They appear on south-oriented slopes, where PET increases strongly. Cells with a score below 0.5, where a shift occurred in both dimensions or towards a class that is not immediately contiguous to the original one, are quite rare, unlike in the map of runoff and snowmelt percentage.

An alternative use of bivariate mapping for assessing temporal change has been proposed by Teuling et al. (2011). This method involves plotting the magnitude of change between two periods for both variables, using a divergent color scheme. An advantage of this method is that the resulting maps provide more information about the underlying processes, but unlike the method presented in this paper it does not offer a single score for quantifying the overall magnitude of change, which makes it more difficult to compare different periods.

## 6. Conclusion

Bivariate mapping can be used to produce a simple and easily understandable classification scheme. For the maps in this paper,

variable pairs were chosen that are well adapted to describe the water balance in Switzerland. To choose appropriate variable pairs, the relevant processes in the region being studied must be known, as well as the scale of the data. While much of the research on hydrological classification focused on catchments, bivariate classification can be applied to open or closed hydrological units of any size. Also, many classification schemes rely on relative indices of streamflow seasonality or ratio variables. However, the bivariate classifications presented in this paper show that the use of absolute values as a criterion for classification is sensible in a region with high spatial heterogeneity such as Switzerland. It is important to keep in mind that the number of classes and the position of class breaks are arbitrary choices. The scale of the data has a considerable influence on the resulting map, as the local variability of hydrological data can be very high, especially in mountainous regions. Further research is necessary to determine class breaks based on objective criteria for different spatial scales to reduce the degree of arbitrariness of the proposed classification scheme.

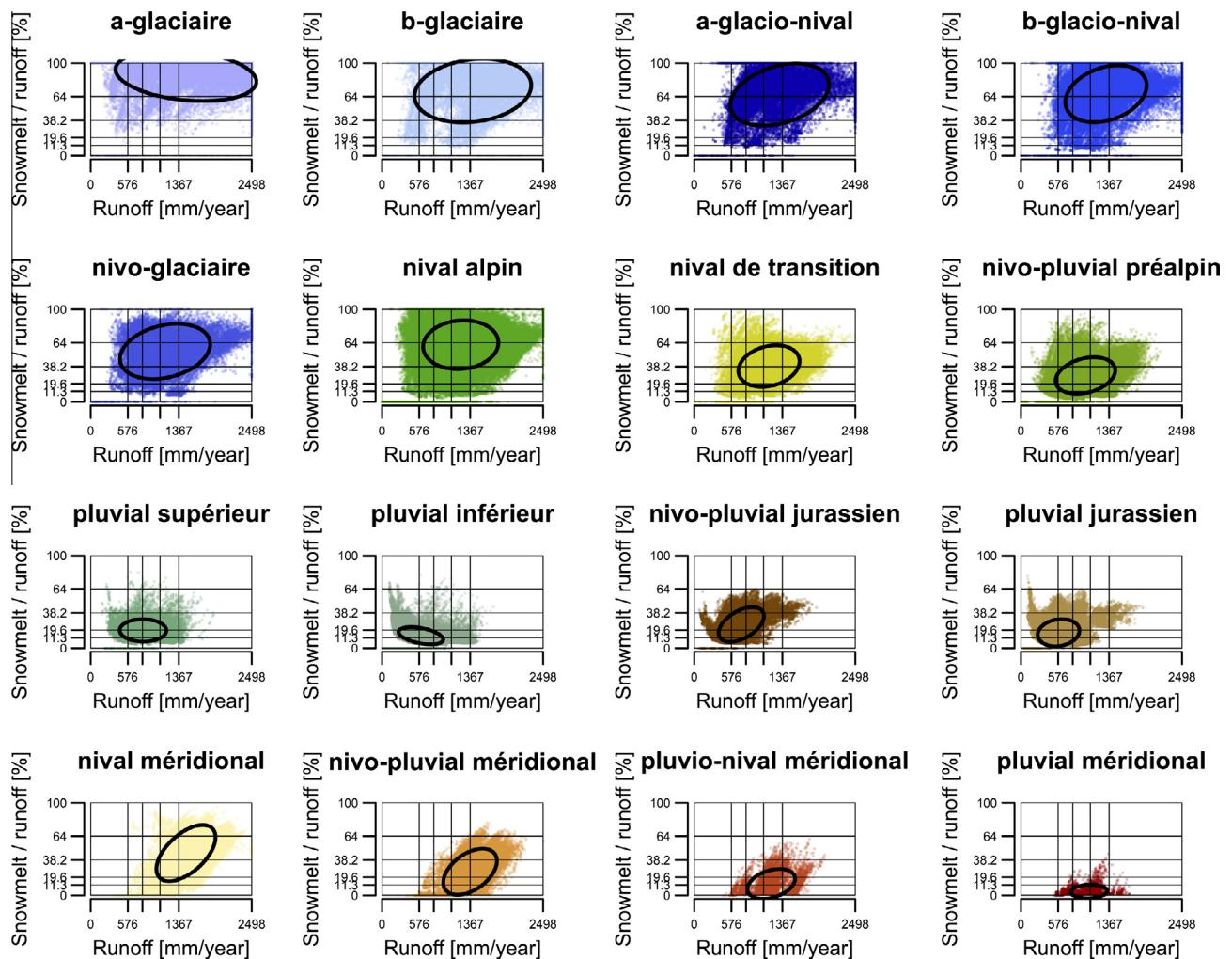
The approach described here can be used to assess the differences between maps representing two different periods, or model outputs based on different assumptions, with the help of map agreement measures. The results of these comparisons matched a priori expectations, i.e. maps that are expected to be more distinct obtained lower similarity scores. Both measures show a high degree of correlation, so that any one of them could be used for comparing different periods or different scenarios. Fuzzy  $\kappa$  allows to consider the similarity of classes with each other as well as spatial autocorrelation, which is why this measure was preferred for the analysis of maps in this paper.

The analysis of current-state bivariate maps together with maps of future predictions revealed differences between different periods, models and scenarios. These correspond with some of the effects of climate change on hydrology in Switzerland forecast in previous climate impact studies, based primarily on streamflow seasonality (e.g. CH2014-Impacts, 2014; Köplin et al., 2014), namely that regime types in lower and intermediate elevation ranges will become more pluvially dominated, and those North of the Alps more like those in the South under current conditions. Also, the choice of emission scenario has a strong impact on the degree of map agreement, indicating that the method is suitable for assessing the intensity of change. The results obtained with the method presented here are in accordance with those in previous studies and with a priori expectations. This indicates that

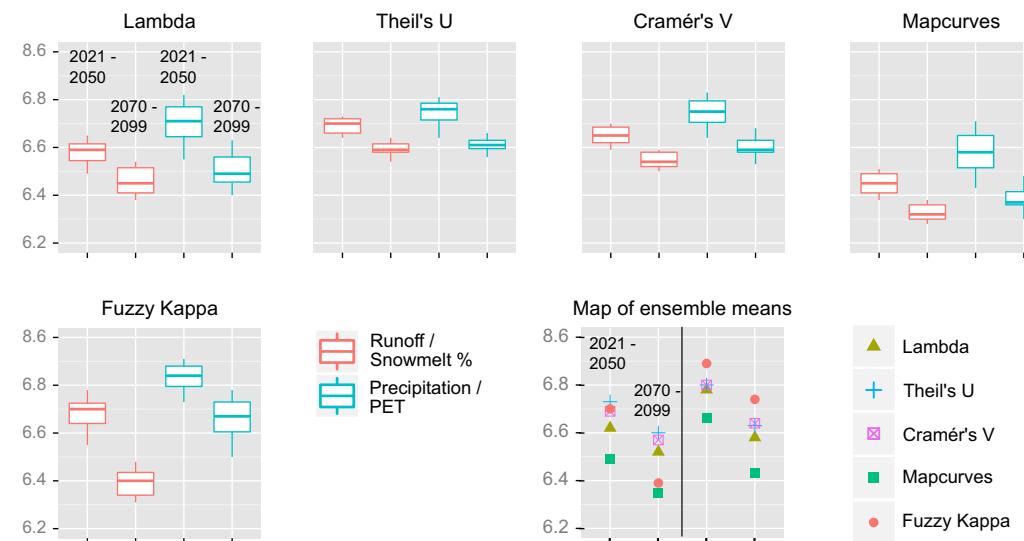
**Table A.8**

Agreement scores obtained by comparing current state maps with future predictions based on different greenhouse gas emission scenarios.

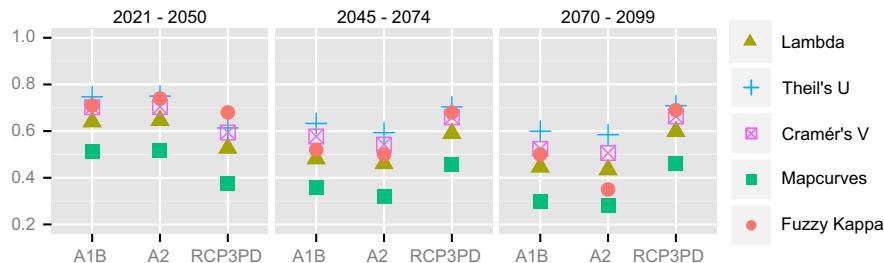
Model chain	Period	$\kappa$	Fuzzy $\kappa$
<i>Runoff/Snowmelt contribution to runoff</i>			
A1B	2021–2050	0.48	0.71
	2045–2075	0.26	0.52
	2070–2099	0.18	0.51
A2	2021–2050	0.53	0.74
	2045–2075	0.25	0.50
	2070–2099	0.15	0.35
RCP3PD	2021–2050	0.43	0.68
	2045–2075	0.44	0.68
	2070–2099	0.46	0.69
<i>Precipitation/PET</i>			
A1B	2021–2050	0.83	0.92
	2045–2075	0.62	0.78
	2070–2099	0.52	0.70
A2	2021–2050	0.85	0.93
	2045–2075	0.58	0.74
	2070–2099	0.46	0.65
RCP3PD	2021–2050	0.84	0.92
	2045–2075	0.76	0.87
	2070–2099	0.77	0.87



**Fig. B.1.** Density of individual hydrological regimes after Aschwanden and Weingartner (1985) in the scatter plot of runoff and snowmelt contribution.



**Fig. C.1.** Distribution of similarity measures on comparing bivariate maps for the near (2021–2050) and far future (2070–2099) with the reference period (1980–2009).



**Fig. C.2.** Distribution of similarity scores on comparing bivariate maps of runoff/snowmelt percentage for the near, intermediate and far future with the reference period (1980–2009).

the use of bivariate mapping together with map comparison measures provides a sound framework for assessing temporal changes in hydrological data, both quantitatively and qualitatively.

## Acknowledgments

This study is part of the CCHydro project suite, financed by the Swiss Federal Office for Environment FOEN. The meteorological data used as input was provided by the Swiss Federal Office of Meteorology and Climatology MeteoSwiss. The data for the different GCM/RCM model chains is based on the ENSEMBLES project (contract 505539), funded by the European Commission. This research was originally carried out for a Master Thesis in the Department of Geography, University of Zurich, supervised by Prof. Dr. Jan Seibert.

A script for displaying bivariate maps was provided by Dr. Felix Fundel (WSL). For the calculation of Mapcurves scores, the implementation made publicly available online by Dr. Emiel van Loon (University of Amsterdam) was used.

Silvia Dingwall proofread this paper and provided valuable suggestions for editing.

We thank the Editor, Dr. Bettina Schaeffli and an anonymous reviewer for their well addressed comments and suggestions to the original manuscript.

## Appendix A. Tables of similarity scores for different model chains and scenarios

See Tables A.7, A.8.

## Appendix B. Density of individual hydrological regimes in the scatter plot of runoff and snowmelt contribution

See Fig. B.1.

## Appendix C. Results of comparing maps across periods with measures of association

See Figs. C.1, C.2.

## References

- Agresti, A., 2002. *Categorical Analysis*. Wiley-Interscience, New York.
- Aschwanden, H., Weingartner, R. Die Abflussregimes der Schweiz. Publikation Gewässerkunde, 1985. Geographisches Institut der Universität Bern, Abt. Physikalische Geographie-Gewässerkunde.
- Swiss Federal Office of Environment/Bundesamt für Umwelt (BAFU), Einzugsgebietsgliederung Schweiz EZGG-CH, Release 2012.
- Beckinsale, R.P. 1969. River Regimes. In: Chorley, R.J. (Ed.), Water, Earth and Man. London: Methuen.b0025, pp. 176–192.
- Blöschl, G., Sivapalan, M., 1995. Scale issues in hydrological modelling: a review. Hydrol. Process. 9, 251–290. <http://dx.doi.org/10.1002/hyp.3360090305>.
- Blöschl, G., 1999. Scaling issues in snow hydrology. Hydrol. Process. 13, 2149–2175. [http://dx.doi.org/10.1002/\(SICI\)1099-1085\(199910\)13:14/15<2149::AID-HYP847>3.0.CO;2-8](http://dx.doi.org/10.1002/(SICI)1099-1085(199910)13:14/15<2149::AID-HYP847>3.0.CO;2-8).
- Bosshard, T., Kotlarski, S., Ewen, T., Schär, C., 2011. Spectral representation of the annual cycle in the climate change signal. Hydrol. Earth Syst. Sci. 15, 1161–1192. <http://dx.doi.org/10.5194/hess-8-1161-2011>.
- Budyko, M.I., 1974. *Climate and Life*. Academic Press, New York, English edition.
- Cerdan, O., Le Bissonnais, Y., Govers, G., Lecomte, V., van Oost, K., Couturier, A., King, C., Dubreuil, N., 2004. Scale effect on runoff from experimental plots to catchments in agricultural areas in Normandy. J. Hydrol. 299, 4–14. <http://dx.doi.org/10.1016/j.jhydrol.2004.02.017>.
- CH2011, 2011. Swiss Climate Change Scenarios CH2011, published by C2SM, MeteoSwiss, ETH, NCCR Climate, and OcCC, Zurich, Switzerland, 88 pp. ISBN: 978-3-033-03065-7.
- CH2014-Impacts, 2014. Toward Quantitative Scenarios of Climate Change Impacts in Switzerland, published by OCCR, FOEN, MeteoSwiss, C2SM, Agroscope, and ProClim, Bern, Switzerland, 136 pp. ISBN 978-3-033-04406-7.
- Chapman, T., 1989. Classification of regions. In: Falkenmark, M., Chapman, T. (Eds.), *Comparative Hydrology: An Ecological Approach to Land and Water Resources*. UNESCO, Paris, pp. 67–74.
- Chaudhury, G., Clarke, K.C., 2014. Temporal accuracy in urban growth forecasting: a study using the SLEUTH Model. Trans. GIS 18, 302–320. <http://dx.doi.org/10.1111/tgis.12047>.
- Chen, Y., Lin, Z., You, J., 2008. Information-theoretical comparison between actual and potential natural vegetation. In: Zhang, J., Goodchild, M.F. (Eds.), *Proceedings of the 8th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*. Shanghai, P.R. China. pp. 206–212.
- Cohen, J., 1960. A coefficient of agreement for nominal scales. Educ. Psychol. Meas. 20, 37–46. <http://dx.doi.org/10.1177/001316446002000104>.
- Coopersmith, E., Yaeger, M.A., Ye, S., Cheng, L., Sivapalan, M., 2012. Exploring the physical controls of regional patterns of flow duration curves – Part 3: a catchment classification system based on regime curve indicators. Hydrol. Earth Syst. Sci. 16, 4467–4482. <http://dx.doi.org/10.5194/hess-16-4467-2012>.
- Coopersmith, E.J., Minster, B.S., Sivapalan, M., 2014. Patterns of regional hydroclimatic shifts: an analysis of changing hydrologic regimes. Water Resour. Res. 50, 1960–1983. <http://dx.doi.org/10.1002/2012WR013320>.
- Couto, P., 2003. Assessing the accuracy of spatial simulation models. Ecol. Model. 167, 181–198. [http://dx.doi.org/10.1016/S0304-3800\(03\)00176-5](http://dx.doi.org/10.1016/S0304-3800(03)00176-5).
- Cramér, H., 1946. *Mathematical Methods of Statistics*. Princeton University Press.
- Finn, J.T., 1993. Use of the average mutual information index in evaluating classification error and consistency. Int. J. Geogr. Inf. Syst. 7, 349–366. <http://dx.doi.org/10.1080/02693799308901966>.
- Foody, G.M., 2008. Harshness in image classification accuracy assessment. Int. J. Remote Sens. 29, 3137–3158. <http://dx.doi.org/10.1080/01431160701442120b0110>.
- García-Ruiz, J.M., Lana-Reinault, N., Beguería, S., Lasanta, T., Regués, D., Nadal-Romero, E., Serrano-Muela, P., López-Moreno, J.I., Alvera, B., Martí-Bono, C., Alatorre, L.C., 2010. From plot to regional scales: interactions of slope and catchment hydrological and geomorphic processes in the Spanish Pyrenees. J. Geomorph. 120, 248–257. <http://dx.doi.org/10.1016/j.geomorph.2010.03.038>.
- Goodman, L.A., Kruskal, W.H., 1954. Measures of association for cross classifications. J. Am. Stat. Assoc. 49, 732–764. <http://dx.doi.org/10.1080/01621459.1954.10501231>.
- Gottschalk, L., 1985. Hydrological regionalization of Sweden. Hydrol. Sci. J. 30, 65–83. <http://dx.doi.org/10.1080/02626668509490972>.
- Grabs, T., Seibert, J., Bishop, K., Landon, H., 2009. Modeling spatial patterns of saturated areas: a comparison of the topographic wetness index and a dynamic distributed model. J. Hydrol. 373, 15–23. <http://dx.doi.org/10.1016/j.jhydrol.2009.03.031>.
- Güntner, A., Seibert, J., Uhlenbrook, S., 2004. Modeling spatial patterns of saturated areas: An evaluation of different terrain indices. Water Resour. Res. 40. <http://dx.doi.org/10.1029/2003WR002864>.
- Hagen-Zanker, A., 2009. An improved Fuzzy Kappa statistic that accounts for spatial autocorrelation. Int. J. Geogr. Inf. Sci. 23, 61–73. <http://dx.doi.org/10.1080/13658810802570317>.
- Hall, O., Duit, A., Caballero, L.N.C., 2008. World poverty, environmental vulnerability and population at risk for natural hazards. J. Maps 4 (1), 151–160. <http://dx.doi.org/10.4113/jom.2008.95>.
- Hargrove, W.W., Hoffman, F.M., Hessburg, P.F., 2006. Mapcurves: a quantitative method for comparing categorical maps. J. Geograph. Syst. 8, 187–208. <http://dx.doi.org/10.1007/s10109-006-0025-x>.

- Horton, P., Schaeafi, B., Mezghani, A., Hingray, B., Musy, A., 2006. Assessment of climate-change impacts on alpine discharge regimes with climate model uncertainty. *Hydrol. Process.* 20, 2091–2109.
- HRachowitz, M., Savenije, H.H.G., Blöschl, G., McDonnell, J.J., Sivapalan, M., Pomeroy, J.W., Arheimer, B., Blume, T., Clark, M.P., Ehret, U., Fenicia, F., Freer, J.E., Gelfan, A., Gupta, H.V., Hughes, D.A., Hut, R.W., Montanari, A., Pande, S., Tetzlaff, D., Troch, P.A., Uhlenbrook, S., Wagener, T., Winsemius, H.C., Woods, R.A., Zehe, E., Cudennec, C., 2013. A decade of Predictions in Ungauged Basins (PUB)-a review. *Hydrol. Sci. J.* 58, 1–58. <http://dx.doi.org/10.1080/02626667.2013.803183>.
- IPCC, 2013. In: Stocker, T.F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S.K., Boschung, J., Nauels, A., Xia, Y., Bex, V., Midgley, P.M. (Eds.), *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, p. 1535.
- Jörg-Hess, S., Kempf, S.B., Fundel, F., Zappa, M., 2014. The benefit of climatological and calibrated reforecast data for simulating hydrological droughts in Switzerland. *Met. Apps.* <http://dx.doi.org/10.1002/met.1474>.
- Kobierska, F., Jonas, T., Zappa, M., Bavay, M., Magnusson, J., Bernasconi, S.M., 2013. Future runoff from a partly glaciated watershed in Central Switzerland: a two-model approach. *Adv. Water Resour.* 55, 204–214. <http://dx.doi.org/10.1016/j.advwatres.2012.07.024>.
- Köplin, N., Schädler, B., Vivirol, D., Weingartner, R., 2014. Seasonality and magnitude of floods in Switzerland under future climate change. *Hydrol. Process.* 28, 2567–2578. <http://dx.doi.org/10.1002/hyp.9757>.
- Köppen, W., 1918. *Klassifikation der Klimate nach Temperatur, Niederschlag und Jahresablauf* (Classification of climates according to temperature, precipitation and seasonal cycle). Petermanns Geogr. Mitt. 64, 193–203, 243–248.
- Krasovskaja, I., 1997. Entropy-based grouping of river flow regimes. *J. Hydrol.* 202, 173–191. [http://dx.doi.org/10.1016/S0022-1694\(97\)00065-6](http://dx.doi.org/10.1016/S0022-1694(97)00065-6).
- Leonowicz, A., 2006. Two-variable choropleth maps as a useful tool for visualization of geographical relationships. *Geografija* 42, 33–37.
- L'vovich, M.I., 1979. *World Water Resources and their Future*. American Geophysical Union, Washington, DC, English Edition.
- Moore, N., Messina, J., 2010. A landscape and climate data logistic model of tsetse distribution in Kenya. *PLOS ONE* 5 (7). <http://dx.doi.org/10.1371/journal.pone.0011809>.
- Nakicenovic, N., Swart, R. IPCC Special Report on Emissions Scenarios Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, UK, 570 pp.
- Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models. Part I: a discussion of principles. *J. Hydrol.* 10, 282–290. [http://dx.doi.org/10.1016/0022-1694\(70\)90255-6](http://dx.doi.org/10.1016/0022-1694(70)90255-6).
- Olson, J.M., 1981. Spectrally encoded two-variable maps. *Ann. Assoc. Am. Geogr.* 71, 259–276. <http://dx.doi.org/10.1111/j.1467-8306.1981.tb01352.x>.
- Pflugshaupt, C., 2013. Future water resources in Switzerland: an assessment based on the Swiss Climate Change Scenarios CH2011. Unpublished Master Thesis, ETH Zurich/Swiss Federal Research Institute WSL.
- Rees, W.G., 2008. Comparing the spatial content of thematic maps. *Int. J. Remote Sens.* 29, 3833–3844. <http://dx.doi.org/10.1080/01431160701852088>.
- Schaeafi, B., Gupta, H.V., 2007. Do Nash values have value? *Hydrol. Process.* 21, 2075–2080. <http://dx.doi.org/10.1002/hyp.6825>.
- Schattan, P., Zappa, M., Lischke, H., Bernhard, L., Thürig, E., and Diekkrüger, B., 2013. An approach for transient consideration of forest change in hydrological impact studies. In: Climate and Land Surface Changes in Hydrology, Proceedings of H01, IAHS-IAPSO-IASPEI Assembly, Gothenburg, Sweden, July (IAHS Publ. 359), pp. 311–319.
- Schimel, A.C.G., Healy, T.R., Johnson, D., Immenga, D., 2010. Quantitative experimental comparison of single-beam, sidescan, and multibeam benthic habitat maps. *ICES J. Mar. Sci.* 67, 1766–1779. <http://dx.doi.org/10.1093/icesjms/fsq102>.
- Shannon, C.E., 1948. A mathematical theory of communication. *Bell.* <http://dx.doi.org/10.1145/584091.584093>.
- Sivapalan, M., Takeuchi, K., Franks, S.W., Gupta, V.K., Karambiri, H., Lakshmi, V., Liang, X., McDonnell, J.J., Mendiondo, E.M., O'Connell, P.E., Oki, T., Pomeroy, J.W., Schertzer, D., Uhlenbrook, S., Zehe, E., 2003. IAHS Decade on Predictions in Ungauged Basins (PUB), 2003–2012: shaping an exciting future for the hydrological sciences. *Hydrol. Sci. J.* 48, 857–880. <http://dx.doi.org/10.1623/hysj.48.6.857.51421>.
- Teuling, A.J., Stöckli, R., Seneviratne, S., 2011. Bivariate colour maps for visualizing climate data. *Int. J. Climatol.* 31, 1408–1412. <http://dx.doi.org/10.1002/joc.2153>.
- Teuling, A.J., 2011. Technical note: towards a continuous classification of climate using bivariate colour mapping. *Hydrol. Earth Syst. Sci.* 15, 3071–3075. <http://dx.doi.org/10.5194/hess-15-3071-2011>.
- Trumbo, B.E.A. 1980. Theory for Coloring Bivariate Statistical Maps. Tech. rept. 44. Department of Statistics, Stanford University, Stanford. doi.org/10.1080/00031305.1981.10479360
- United Nations Environmental Programme (UNEP), World Atlas of Desertification, 1997. London.
- van der Linden, P., Mitchell, J.F.B., 2009. ENSEMBLES: Climate Change and its Impacts: Summary of research and results from the ENSEMBLES project. Met Office Hadley Centre, Exeter, UK, 160 pp.
- Vivirol, D., Gurtz, J., Zappa, M., 2007. The hydrological modelling system PREVAH. Part II: Physical model description. *Geographica Bernensia*, Institute of Geography, University of Bern.
- Vivirol, D., Zappa, M., Gurtz, J., Weingartner, R., 2009. An introduction to the hydrological modelling system PREVAH and its pre- and post-processing tools. *Environ. Model. Softw.* 24, 1209–1222. <http://dx.doi.org/10.1016/j.envsoft.2009.04.001>.
- Visser, H., de Nijs, T., 2006. The map comparison kit. *Environ. Model. Softw.* 21, 346–358. <http://dx.doi.org/10.1016/j.envsoft.2004.11.013>.
- Wagener, T., Sivapalan, M., Troch, P., Woods, R., 2007. Catchment classification and hydrologic similarity. *Geography Compass* 4, 901–931.
- Wagener, T., Sivapalan, M., McGlynn, B., 2008. Catchment classification and services – toward a new paradigm for catchment hydrology driven by societal needs. *Encycl. Hydrolog. Sci.* <http://dx.doi.org/10.1111/j.1749-8198.2007.00039.x>.
- Wealands, S.R., Grayson, R.B., Walker, J.P., 2005. Quantitative comparison of spatial fields for hydrological model assessment-some promising approaches. *Adv. Water Resour.* 28, 15–32. <http://dx.doi.org/10.1016/j.advwatres.2004.10.001>.
- Weingartner, R., Schädler, B., Hänggi, P., 2013. Auswirkungen der Klimaänderung auf die schweizerische Wasserkraftnutzung. *Geogr. Helv.* 68, 239–248. <http://dx.doi.org/10.5194/gh-68-239-2013>.
- Weiskel, P.K., Wolock, D.M., Zarriello, P.J., Vogel, R.M., Levin, S.B., Lent, R.M., 2014. Hydroclimatic regimes: a distributed water-balance framework for hydrologic assessment, classification, and management. *Hydrol. Earth Syst. Sci.* 18, 3855–3872. <http://dx.doi.org/10.5194/hess-18-3855-2014>.
- Williams, C.L., Hargrove, W.W., Liebman, M., James, D.E., 2008. Agro-ecoregionalization of Iowa using multivariate geographical clustering. *Agr. Ecosyst. Environ.* 123, 161–174. <http://dx.doi.org/10.1016/j.agee.2007.06.006>.
- Wood, E., 1998. *Scale analyses for land-surface hydrology*. In: Sposito, G. (Ed.), *Scale Dependence and Scale Invariance in Hydrology*. Cambridge University Press.
- Yadav, M., Wagener, T., Gupta, H., 2007. Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins. *Adv. Water Resour.* 30, 1756–1774. <http://dx.doi.org/10.1016/j.advwatres.2007.01.005>.
- Zappa, M., Bernhard, L. Klimaänderung und natürlicher Wasserhaushalt der Grosseinzuflussegebiete der Schweiz, 2012. Schlussbericht zum Projekt Klimaänderung und Hydrologie in der Schweiz (CCHydro). Birmensdorf, Eidg. Forschungsanstalt WSL
- Zappa, M., Bernhard, L., Fundel, F., Jörg-Hess, S., 2012. Vorhersage und Szenarien von Schnee- und Wasserressourcen im Alpenraum. *Forum für Wissen*, 19–27.
- Zappa, M., Kan, C., 2007. Extreme heat and runoff extremes in the Swiss Alps. *Nat. Hazards Earth Syst. Sci.* 7, 375–389. <http://dx.doi.org/10.5194/nhess-7-375-2007>.