



Filling the white space on maps of European runoff trends: estimates from a multi-model ensemble

K. Stahl^{1,2}, L. M. Tallaksen², J. Hannaford³, and H. A. J. van Lanen⁴

¹Institute of Hydrology, University of Freiburg, Freiburg, Germany

²Department of Geosciences, University of Oslo, Oslo, Norway

³Centre for Ecology and Hydrology, Wallingford, UK

⁴Hydrology and Quantitative Water Management Group, Wageningen University, Wageningen, The Netherlands

Correspondence to: K. Stahl (kerstin.stahl@hydrology.uni-freiburg.de)

Received: 3 February 2012 – Published in Hydrol. Earth Syst. Sci. Discuss.: 14 February 2012

Revised: 23 May 2012 – Accepted: 12 June 2012 – Published: 11 July 2012

Abstract. An overall appraisal of runoff changes at the European scale has been hindered by “white space” on maps of observed trends due to a paucity of readily-available streamflow data. This study tested whether this white space can be filled using estimates of trends derived from model simulations of European runoff. The simulations stem from an ensemble of eight global hydrological models that were forced with the same climate input for the period 1963–2000. The derived trends were validated for 293 grid cells across the European domain with observation-based trend estimates. The ensemble mean overall provided the best representation of trends in the observations. Maps of trends in annual runoff based on the ensemble mean demonstrated a pronounced continental dipole pattern of positive trends in western and northern Europe and negative trends in southern and parts of eastern Europe, which has not previously been demonstrated and discussed in comparable detail. Overall, positive trends in annual streamflow appear to reflect the marked wetting trends of the winter months, whereas negative annual trends result primarily from a widespread decrease in streamflow in spring and summer months, consistent with a decrease in summer low flow in large parts of Europe. High flow appears to have increased in rain-dominated hydrological regimes, whereas an inconsistent or decreasing signal was found in snow-dominated regimes. The different models agreed on the predominant continental-scale pattern of trends, but in some areas disagreed on the magnitude and even the direction of trends, particularly in transition zones between regions with increasing and decreasing runoff trends, in complex terrain with a high spatial variability, and in snow-dominated

regimes. Model estimates appeared most reliable in reproducing observed trends in annual runoff, winter runoff, and 7-day high flow. Modelled trends in runoff during the summer months, spring (for snow influenced regions) and autumn, and trends in summer low flow were more variable – both among models and in the spatial patterns of agreement between models and the observations. The use of models to display changes in these hydrological characteristics should therefore be viewed with caution due to higher uncertainty.

1 Introduction

Europe’s climate is changing and with it the spatial and temporal characteristics of its hydrology. In recent decades, precipitation has decreased around the Mediterranean and increased in parts of northern Europe (e.g. Zhang et al., 2007; Klein Tank et al., 2002). Climate assessments and environmental reports for Europe often contain detailed, high-resolution maps of observed changes in precipitation and temperature over recent decades, whereas the assessment of observed changes in streamflow, flood, and drought are largely based on a selection of regional and national case studies (e.g. Bates et al., 2008; EEA-JRC-WHO, 2008; EEA, 2010). A consistent mapping of observed changes in hydrological variables on large regional and continental scales is therefore required to enable a better understanding of global and regional changes in the hydrological cycle and related impacts on water availability and management.

The main obstacles to achieving this goal are the availability and quality of streamflow observations, both of which vary significantly in time and space. In Europe, recent hydrological change has prompted numerous national studies as well as some regional and transnational trend studies (e.g. Wilson et al., 2010; Bard et al., 2011). There have been few attempts to examine trends in streamflow for the whole of Europe. To this end, Stahl et al. (2010) assembled a dataset of European streamflow records from 11 countries and conducted trend analyses at the continental scale. While these studies have given a very detailed account of changes in large parts of Europe, “white space” remains on the maps, where observations are unavailable or sparse due to issues of data accessibility and quality (Hannah et al., 2011; Viglione et al., 2010).

Model simulations offer one possible approach for filling such white space. At the global scale, modelling studies have enabled past runoff changes in time and space to be mapped continuously, albeit at very coarse resolution (e.g. Milly et al., 2005; Gerten et al., 2008; Dai et al., 2009). They provide a reference against which future scenario projections can be compared and “hot spots” of change identified. Similar model-based analyses of past transient changes at regional scales (and thus, finer resolutions) are lacking for Europe, as indeed they are for other continents. Europe-based studies have largely focused on projecting future changes in runoff, including floods and droughts, as a response to climate change scenarios (e.g. Lehner et al., 2006; Feyen and Dankers, 2009; Dankers and Feyen, 2009). Model simulations in these studies were used to compare the relative change in averages and other summary statistics for 30-yr time periods in the past and future. Models used for such time-slice projection studies are commonly calibrated to represent average conditions over a period of time and are rarely tested for their ability to simulate transient changes in time. However, strong recent changes suggest that transient time trends appear to provide useful benchmarks against which simulated runoff from large-scale models can be tested (e.g. Stahl et al., 2011; McCabe and Wolock, 2011).

Furthermore, as with climate models, model intercomparison studies have shown that different models (land surface and hydrological models) show considerable variability in the magnitude and timing of the hydrological variables simulated (e.g. LUCHEM: Breuer et al., 2009; PILPS: Cornwell and Harvey, 2007; GSWP: Dirmeyer et al., 2006; WaterMIP: Haddeland et al., 2011; Prudhomme et al., 2011; Gudmundsson et al., 2012a). Some studies suggest that the ensemble mean (of all models) provides a more accurate estimate than any single model (Guo et al., 2007; Gudmundsson et al., 2012a).

A key objective of this study was to assess the ability of a multi-model ensemble of eight large-scale hydrological models to simulate relatively detailed spatial patterns (0.5° grid scale) of runoff trends in Europe. Hence, the study included a validation of modelled trends against previously published

trends from observed streamflow records from a substantial number of small rivers with near-natural flow across Europe (i.e. from the same dataset used in the observation-based trend study by Stahl et al., 2010). As described in detail in Sect. 2, the models were forced with the same input data and the model runs reflect “naturalized” conditions; i.e. human impacts, such as water storage in man-made reservoirs and agricultural water withdrawals, were not included in the model runs. This setup allowed an assessment of (i) the sensitivity of trend estimates to the hydrological model, and (ii) the validation of spatial patterns of modelled runoff trends against trends in observed streamflow records. This validation complements and adds to modelling studies that have assessed past streamflow trends at the mouth of a few continental river basins (e.g. Dai et al., 2009). In these large basins, anthropogenic disturbances (e.g. impoundments and water withdrawals) may impact streamflow magnitude and transient trends to a degree often not represented by the hydrological models. In contrast, given the high spatial resolution of the mapped trends, small catchments enable the focus to be on processes (in both the climate and hydrological system) controlling changes in runoff generation in space and time.

In addition to the immediate, obvious scientific benefit of an improved assessment of the capability of large-scale hydrological models to reproduce historical runoff changes, a major benefit of this validation exercise is that, contingent on the utility of the models in reproducing observed trends, the model outputs can be used to “fill the white space” on current maps of recent runoff changes on a European scale (e.g. Stahl et al., 2010). The notion of “filling space”, however, herein refers to the filling of gaps in the knowledge of the spatial changes rather than to the actual merging of observed and modelled hydrological information. This paper thus presents and discusses, for the first time, maps of detailed modelled runoff changes for the whole European continent, along with a consideration of the uncertainties resulting from differences among models, and between models and observations.

2 Data

This study used simulated daily runoff from eight large-scale hydrological models that were part of the model intercomparison experiments within the EU funded WATER and global CHANGE (WATCH) project (www.eu-watch.org). Details of the models included (i.e. GWAVA, HTESSSEL, JULES, LPJml, MATSIRO, MPI-HM, Orchidee, and WaterGAP) can be found in Haddeland et al. (2011) and Gudmundsson et al. (2012a), including an overview of the schemes used for simulation of evapotranspiration, runoff generation and snowmelt. All eight models were run for the period 1958–2000 on a global 0.5° grid and forced by the WATCH Forcing Data (WFD). The WFD were derived

from the ERA-40 reanalysis product, interpolated to half-degree resolution and bias-corrected based on Climate Research Unit (CRU) and Global Precipitation Climatology Centre (GPCC) data (Weedon et al., 2011). The size of a grid cell varies depending on the latitude, between 1065 km² (at 70° N) and 2387 km² (at 39.5° N). The first five years of the ERA-40 period (starting in 1958) were used for model spin-up and disregarded from the analysis. With the exception of WaterGAP, the models were not calibrated specifically for this experiment and models use their default soil and vegetation information (Haddeland et al., 2011). The variable used in this study is daily total runoff (sum of fast and slow component) simulated for each grid cell in Europe (4425 land cells).

Streamflow observations from across Europe were available from the combined dataset of the European Water Archive of the UNESCO IHP FRIEND programme (<http://www.unesco.org/new/en/natural-sciences/environment/water/ihp/ihp-programmes>) and the WATCH project. This dataset contains over 400 near-natural streamflow records for the period 1962 to 2004. The catchments span a range of European climates from less than 100 mm of annual runoff in Spain to over 3000 mm of annual runoff in Norway. Although there is a general bias towards headwater catchments, which tend to be uninfluenced by extensive regulation, the catchments span a range of mean elevations from 100 m a.s.l. in Denmark, the UK and northern Germany to over 2000 m a.s.l. in the Alps. Further details on the derivation of the dataset are provided in Stahl et al. (2010), whereas details about the distribution of elevation and catchment area can be found in Stahl et al. (2011) and Gudmundsson et al. (2012a).

Most of the catchments are between 100 and 1000 km² and hence subscale to the size of the model grid cells (ranging from around 1000–2400 km² depending on latitude), particularly in the lower latitudes of the domain and in mountainous regions. Therefore, each gauged catchment was first assigned to the model grid cell in which its centroid lies. For model grids with more than one catchment, only the record from the largest catchment within a grid cell was kept for the analysis, resulting in a dataset of 293 daily streamflow records that could be paired with model simulations of the runoff from the corresponding grid cells.

All daily simulated runoff and observed streamflow data were converted to mm of runoff per unit area. Grid cell-simulated runoff was used directly and not routed to larger river basins as the focus was on spatial patterns of trend in runoff generation as represented by the catchments. The finest temporal resolution used was a 7-day average, which is longer than the typical runoff concentration time in small catchments.

3 Methods

Trends in annual and monthly runoff, annual 7-day maxima (high flow) and minima (low flow) were computed for the modelled runoff from each model and grid cell in Europe and for the 293 observation-based runoff records. The low flow values were derived for the summer period (May to November) to exclude low flow periods caused by snow and ice. The trend magnitude was estimated from the slope of the Kendall-Theil robust line (Theil, 1950). Based on the median of all individual slopes within a time series, this trend estimate is robust to outliers and has been used previously to describe trend magnitudes in observed runoff (Stahl et al., 2010; Déry et al., 2009). Similar to these previous studies and as discussed there in detail, only trend magnitudes are presented and no significance test was carried out. This procedure is the result of many years of debate over the violation of statistical assumptions such as independence of data in time and the power of trend tests as well as over the nature of a trend. For the objective of this study, it is not important whether a trend is monotonic or part of a long-term cycle, or whether a trend is statistically significant. Rather, to facilitate a relative comparison, the trend T (%) for each time series was expressed as the percent change over the period of record of n years relative to the mean \bar{x} for the period, where m (mm yr⁻¹) is the slope:

$$T = \frac{m \cdot n}{\bar{x}} \cdot 100. \quad (1)$$

Three performance measures were derived to compare modelled and observed trends T of the 293 paired model grid cell and catchment runoff. The first measure is the cumulative distribution of trend magnitudes (observed and modelled). The distributions were compared visually to assess potential systematic over- or under-estimation of particular trends, and by the Kolmogorov-Smirnov (KS) test with a significance level of 5% (Smirnov, 1948) to compare the equality of two samples (drawn from the same distribution or not). The second and third measures relate to trend pattern and magnitude. The Pearson correlation coefficient r with its significance at 5% was computed to assess the similarity of the trend pattern across the 293 paired trend values for each model. The mean absolute error e was chosen as a measure of the magnitude of the difference between modelled and observed trends. The performance measures were derived for the trends in annual and monthly runoff, high and low flow for the eight individual models, as well as for the model ensemble mean (i.e. the mean T of individual models).

Spatial patterns in runoff trends were then visualized on maps. A grid-by-grid comparison of the direction of trends indicates areas where the uncertainty is high: grid cells where less than six out of eight (< 75%) of the models agree on the direction of the trend were highlighted. In addition, grid cells where the trend direction of the ensemble mean is opposite to the observed trend were highlighted.

This study does not aim to discuss the performance of the individual models and hence does not reveal the identity of individual models in the results. Previous studies have analysed individual model performance and provide ample reference (Haddeland et al. 2011; Gudmundsson et al., 2012a,b; Prudhomme et al., 2011; and various Technical Reports on www.eu-watch.org). Overall, apart from the snow accumulation and melt model component employed, these studies found little systematic relation between model performance and specific models or model properties, such as model structure, process representation, and parameterization. A ranking of model performance has to be interpreted with caution and can only be thought of as guidance for the careful inspection of the performance metrics themselves (Gudmundsson et al., 2012a).

4 Results

4.1 Trend validation

The first performance measure compared the cumulative distribution of trend magnitude in the paired observed and modelled annual runoff trends (Fig. 1). The hypothesis of the KS test for similar distributions was rejected for all but one model for the annual trends, all but one model and the ensemble mean for the high-flow trends, and all models for the low flow trends. Around 60 % of the observed trends in the sample were found to be positive (wetter), whereas more of the corresponding modelled trends were positive (Fig. 1, upper panel). About 10 % of the observed trends had magnitudes lower than -30% , but less than 5 % of the simulated trends with all models, and of the ensemble mean, were lower than -30% . Hence, the model simulations underestimated the number and magnitude of negative annual trends and overestimated the positive trends.

The distribution of trends in the observed high flow was found to be similar in shape to the annual runoff (Fig. 1, middle panel). The spread among models was wider for high flow trends than for annual runoff trends. The shape of the distribution of the ensemble mean resembles that of the observations, in particular for the positive high flow trends. The widest spread among models was found for low flow trends (Fig. 1, lower panel). While the ensemble mean in this case captures the proportion of negative and positive trends correctly, modelled trends were weaker than in the observations for both positive and negative trends.

The distributions of trends in monthly runoff overall had a wider spread among the models than those of the annual runoff trends, particularly for the summer months (not shown). The models were able to capture the general shift from predominantly positive trends from October to March, to predominantly negative trends in May, June and August, with similar tendencies for over- and underestimation of the magnitudes as found for trends in annual runoff. The KS test

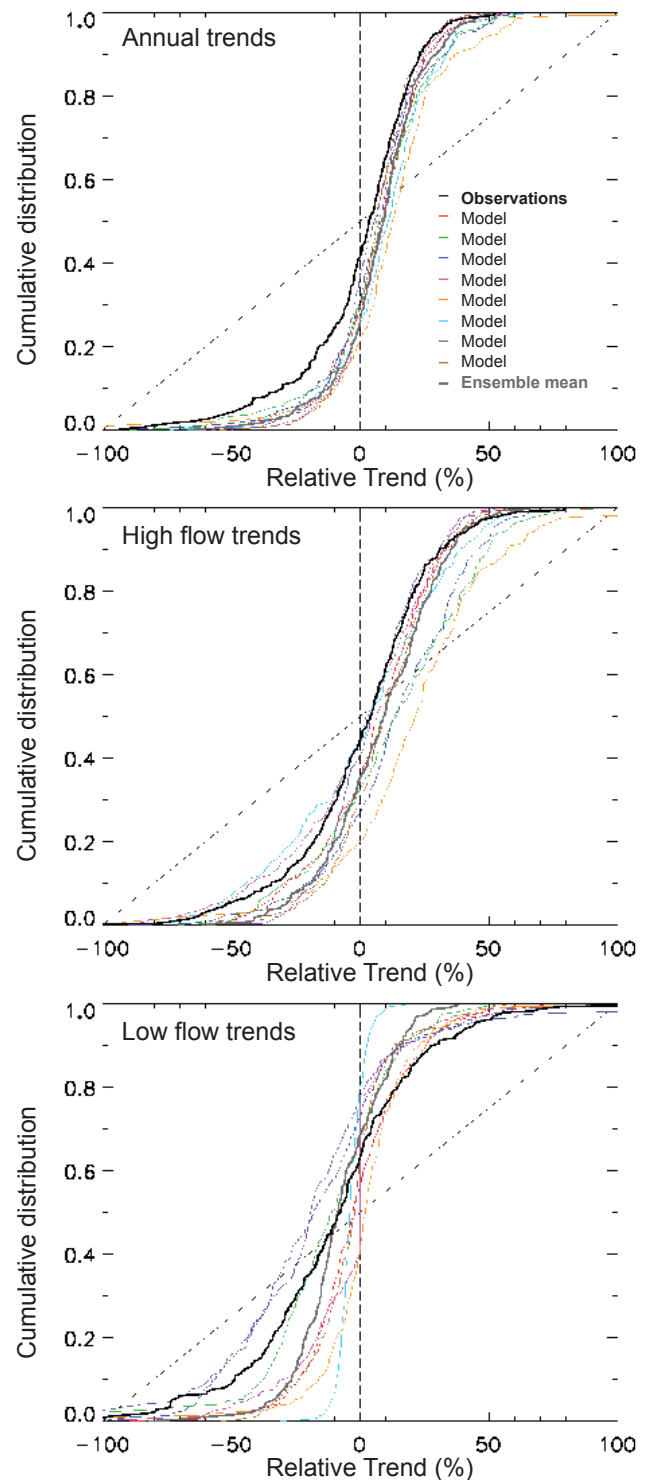


Fig. 1. Distribution of trend in annual runoff (upper panel), high flow (middle panel) and summer low flow (lower panel) from observations, individual models and the ensemble mean trend.

rejected the hypothesis for similar distributions of observed and simulated trends in the winter half-year for most models. Exceptions were two models in October, five in November, four in January, three in February and March, and the ensemble mean in November and March. Except for the month of August, where the hypothesis of similar distribution was only rejected for about half the models, rejection rates for trends in the summer were even higher. Exceptions were one model in April and May and two models in June.

The performance measures for trend pattern r (correlation coefficient) and magnitude e (mean absolute error) of the 293 paired runoff series varied considerably across the models, with values in the range $0.1 < r < 0.7$ and $12\% < e < 34\%$ (Fig. 2). Some of the correlations were weak, but they were all significant (5% level). The measures showed a decreasing agreement of observed and modelled trends from annual mean runoff, to high flow, and to summer low flow (Fig. 2). The ensemble mean performed better than any individual model for high and low flow trends and was among the best models for annual trends.

For the monthly runoff trends, the two measures (r and e) showed large seasonal differences and a large variation among the models (Fig. 3). Correlation coefficients were weak for some models and months, but significant with the exception of one model in April. Throughout the year, the ensemble mean trend ranked consistently high in the agreement with observed monthly trends. In some months, r or e was higher for an individual model's trend, although the best performing model tended to vary. Performance of the ensemble mean was best in the period December to April, and in June and October and worst in August and September (both r and e). The spread of errors among the models was lowest for February, October and November; the spread of correlations among the models was lowest from January to March. Overall highest (lowest) correlations were found in February and March (May, August and September), whereas lowest (highest) errors were observed in October and November (August and September).

4.2 Spatial patterns of trends

The maps in Fig. 4 present the spatial distribution of trends in annual runoff, high flow, and summer low flow for the observations (Fig. 4a, d, g), the model ensemble mean at the locations of the observations (Fig. 4b, e, h), and the ensemble mean trend for the whole European domain (Fig. 4c, f, i). The trends in *annual runoff* (Fig. 4a, b, c) were characterized by a prominent gradient from the south to the north-west: strong negative trends in Iberia, the Mediterranean and in eastern Europe (from the Black Sea in the south to nearly the Baltic Sea in the north), contrasting with predominantly positive trends in western to central Europe and in northern Europe. The trends in observations of annual flow (Fig. 4a) showed a broadly similar pattern to that of the modelled trends (Fig. 4b), but a higher local variability. Deviations

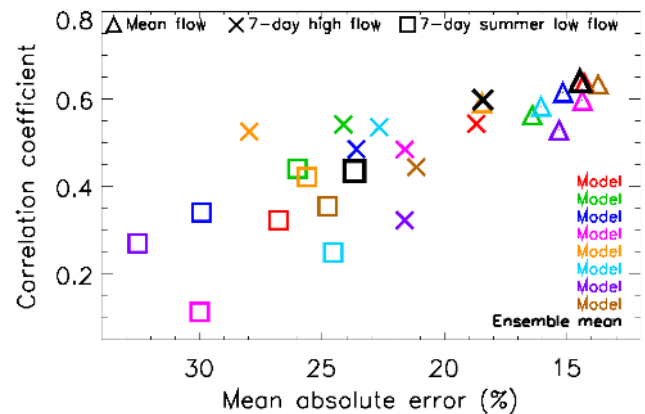


Fig. 2. Comparison of correlation r (pattern) and error e (difference in trend magnitude) between observed and modelled trends in annual flow, high flow, and low flow for all paired basins and grid cells (location in inset). Bold symbols are the ensemble means. Best performing models plot in the top right corner.

between observed and modelled trends were found in regions of predominantly positive modelled trends, specifically in the UK, Germany, the Alps, and Norway. The European pattern of annual runoff trends modelled by the ensemble mean is regionally very coherent (Fig. 4c). Areas where models disagreed on the trend direction were largely located in areas of weak trends – notably, in the transition areas between regions with consistent negative and positive trends.

The distribution of trends in *high flow* was found to be generally similar to that of annual runoff (i.e. negative trends in southern and eastern Europe and positive elsewhere), with more positive trends than for the annual flows, which is particularly visible in the observations and paired modelled grid cells (Fig. 4d, e versus a, b). Differences of the high flow trends to the annual trends were found in the Alps and Scandinavia, where high flow decreased in some areas despite an increased annual runoff. Hence, high flow appears to have increased in rain-dominated hydrological regimes, whereas an inconsistent or decreasing signal was found in snow-dominated regimes, which typically have a late spring maximum runoff generated by snowmelt. The maps of trends in high flow showed more notable differences between modelled trends and observations, as well as in the general continental pattern, than for annual runoff trends. Besides differences in snow-affected regions such as the Alps and Scandinavia, there were also selected catchments in the UK, Spain, the Czech Republic and Slovakia, where local observations show an opposite trend to the simulations. Such differences appear to be often located within areas of model disagreement. Generally, there was greater disagreement among models (shown as crosses on the map) for high flow trends than for annual runoff trends.

The spatial distribution of trends in *summer low flows* (Fig. 4g, h, i) differs, with more prominent negative trends

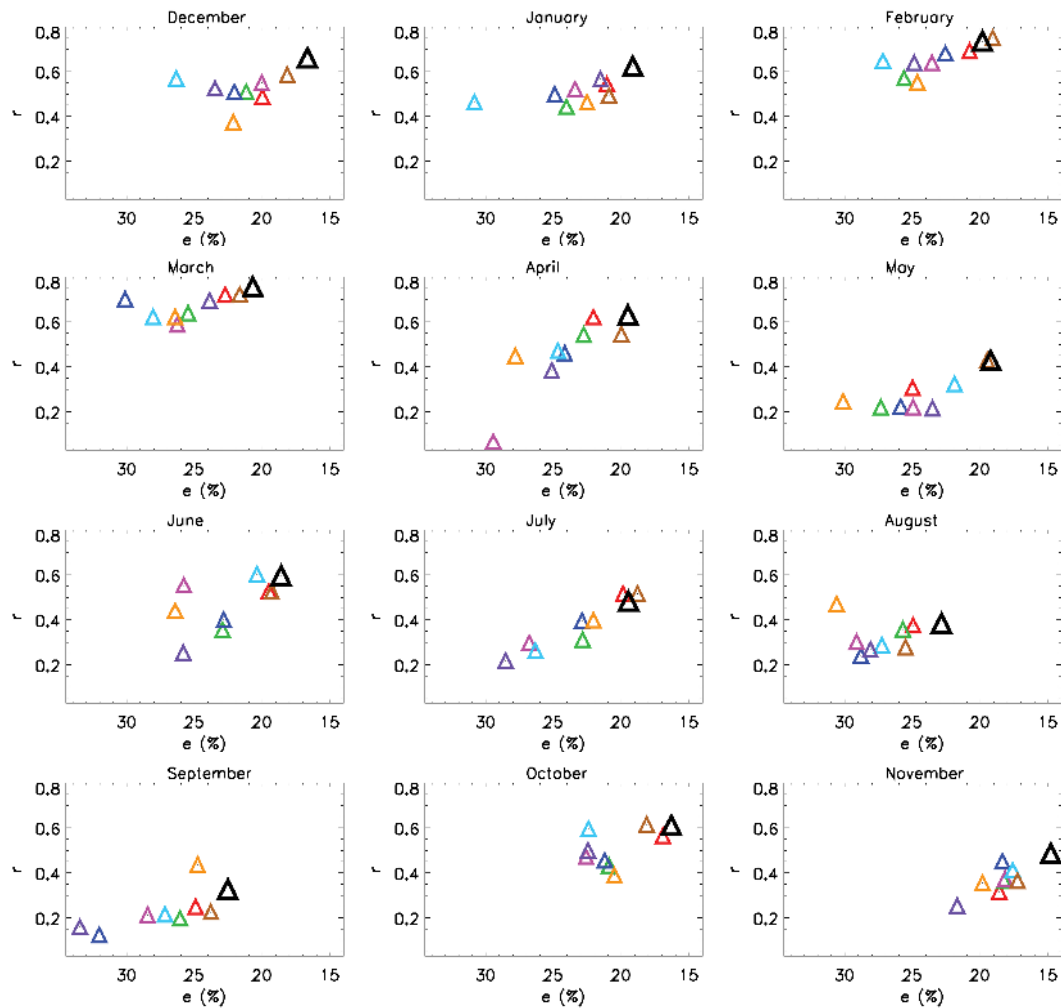


Fig. 3. As Fig. 2 for trends in monthly runoff.

across larger parts of Europe than for annual and high flows. Decreasing low flow trends were most pronounced in the Mediterranean. Exceptions (i.e. increasing or no trends) were the northeast (the Baltic countries and Scandinavia, except the southwestern coasts) and some regions in western Europe. Low flow trends in the observations were generally stronger and more locally variable than the modelled low flow trends in the corresponding grid cells (Fig. 4g and h). In comparison with annual mean and high flow trends, regions where low flow trends in observations and models disagreed are more widespread across Europe. The disagreement among models for low flow trends (Fig. 4i) was similar to the high flow trends, and larger than for the trends in annual flow. Disagreement was mainly found in regions with weak trends and along the transition between areas with positive and negative low flow trends.

Figure 5 shows maps of *monthly runoff* trends based on the ensemble mean. The pronounced dipole pattern found for the annual flow trends (Fig. 4) appears to reflect the wetting trend

pattern of the winter period (ca. December to April) in the north and northwest and the widespread drying trend pattern from late winter to late summer (ca. February–August) in southern and parts of eastern Europe. Months with stronger trends, such as the distinct trend patterns from December to April, resulted in a higher agreement with the observations (ref. Fig. 3). From December to March, different trend directions in the paired observed and modelled grid cells occurred mainly along the boundary between areas of positive and negative trends, in northern Scandinavia (in December only), and for other, mainly isolated, locations throughout Europe. From April to July, however, when negative trends started to increasingly dominate in the observations (as discussed in detail in Stahl et al., 2010 for the observations), many modelled trends in central Europe point in the opposite direction. Negative trends, which dominated the results for the summer months, were generally less reliably modelled, both according to the large differences among models and when compared to observations. Trends in the autumn

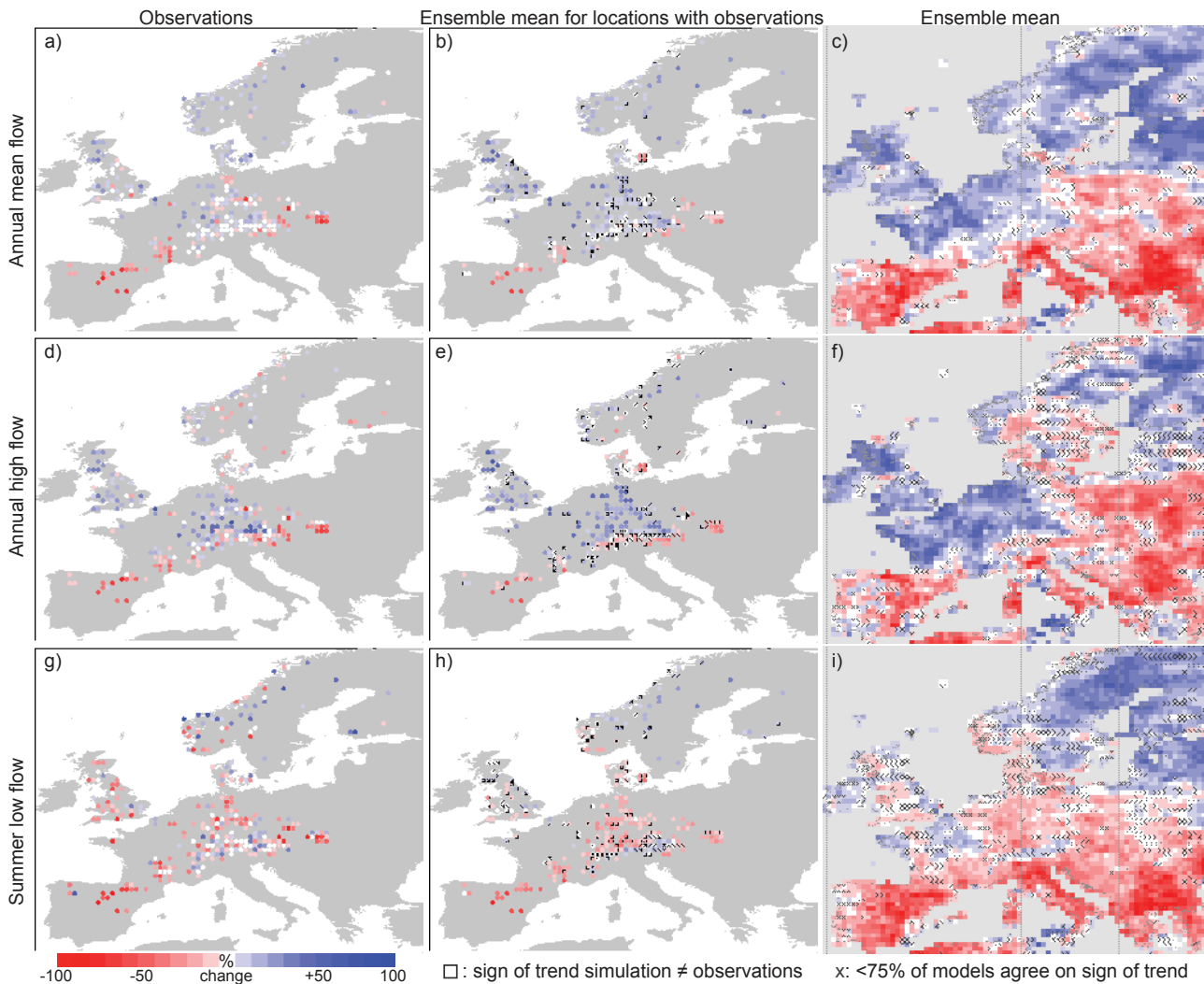


Fig. 4. Spatial distribution of the ensemble mean trend in annual runoff (upper panel), high flow (middle panel) and summer low flow (lower panel).

months September to November, being the weakest of all seasons, showed the largest disagreement in trend direction among models, and also had the largest disagreements with observations.

5 Discussion

5.1 Spatially distributed trend validation

Key systematic differences that emerged from the comparison of modelled trends with paired grid cell observations from catchments include (i) a shifted distribution in the trend magnitude with an overestimation of the number and magnitude of increasing (wetter) trends and an underestimation of the number and magnitude of decreasing (drier) trends, and (ii) a considerably higher local spatial variability of trends in the observed than in the modelled trends. These differences

can have various sources, including errors in the forcing data, limitations related to model resolution and model concepts or physics, quality of the input data to derive model parameters and quality and availability (spatial coverage) of streamflow gauging stations (observed runoff).

The good agreement between modelled and observed trends in annual runoff, high flow and winter month runoff in rain-dominated hydrological regimes implies that the WATCH forcing data (WFD) are reliable with respect to the forcing that these runoff characteristics are sensitive to. Most internationally available precipitation records have been used in the construction of the WFD. Independent validations have been carried out in the framework of specific catchment studies (Van Huijgevoort et al., 2010) and with FLUXNET data from six sites in Europe and North America (Weedon et al., 2011). These time series, which cover a range of climatic regimes, land-cover types and elevations, show a good match

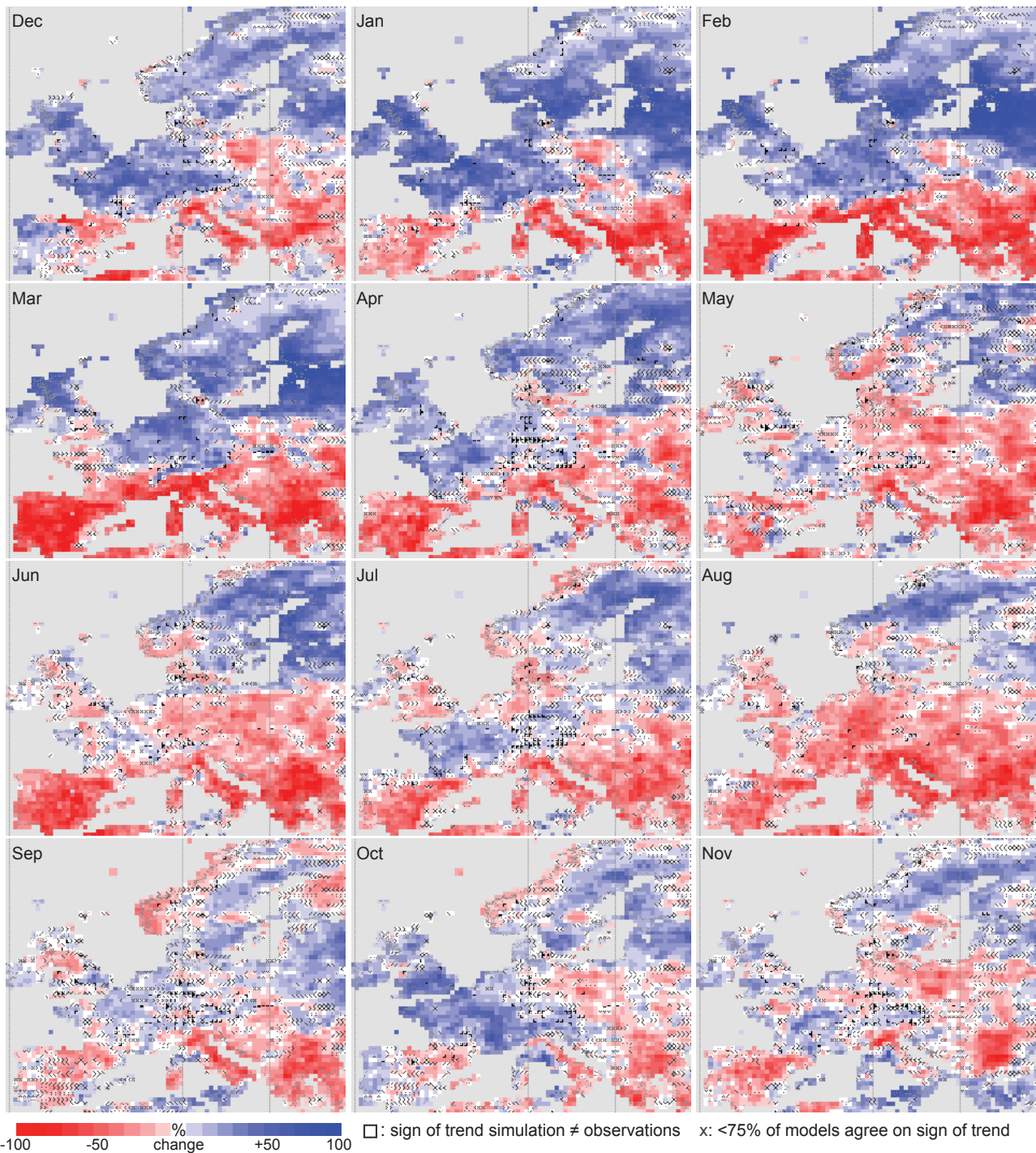


Fig. 5. Spatial distribution of the ensemble mean trend in monthly runoff.

in the occurrence and intensity of daily precipitation; however, time trends were not specifically compared. The increasing trends found in high flow are located in areas that coincide with areas of increasing rainfall (Klein Tank et al., 2002; Zhang et al., 2007) and with increasing wet spell length (Zolina et al., 2010), which is probably associated with the

enhanced runoff generation reported herein. The comparison with observations in this study shows that the models likely exaggerate this effect at the grid scale, causing steeper trends than observed in the catchments. Within the dataset used, no scaling effect was evident, but further studies may want to consider nested subcatchments within larger river basins (not

available in the EWA) to explore scaling effects. However, potential human influences and the necessity to introduce a routing model will then add considerable uncertainty to the attribution of errors to runoff generation (e.g. Balsamo et al., 2009).

Larger differences between modelled and observed trends were found in regions with snow influence. This is consistent with the results of Gudmundsson et al. (2012b), who found a systematic lower model performance in simulating the mean annual cycle for snow-dominated regimes in Europe. Similarly, Haddeland et al. (2011) highlighted that significant differences in simulated monthly flow between land surface and global hydrological models are related to the relative partitioning between rainfall and snowfall and the snow accumulation and melt scheme employed. The present study shows that these differences have strong implications for the detection of trends in monthly runoff and in annual 7-day high-flow, which occur during snowmelt in some regions. In parts of Scandinavia and northeastern Europe, trends in the high flow observations are negative, but the modelled trends are predominantly positive. In areas with complex topography like in Norway, the modelling of snow processes is challenging due to a sub-grid scale elevation dependence of the climate variables. Around the Baltic Sea, where there is snow but little topographic relief, available observations were, unfortunately, sparse.

In addition to the differences between models in representing snow processes, a further reason for model disagreement with observations in mountain areas is the localised small-scale variability of terrain, geology and climate. Unsurprisingly, observed trends in relatively small mountain headwater catchments appear to be more variable than in other regions. Heterogeneity of observed trends in mountainous and snow-influenced areas has been reported previously, with patterns being highly dependent on the dominant process controls of the hydrological regimes. In the Nordic countries, Wilson et al. (2010) found no coherent pattern of high flow trends and seasonal flow, partly due to temperature-driven changes affecting the timing of snowmelt and the seasonal distribution of flow. A study of trends in snowmelt season flow in the Alps found a similar heterogeneity except for glacial regimes (Bard et al., 2011). Rivers with differing degrees of rain and snow dominance on runoff occur side-by-side in mountainous regions. Model grid cell sizes of 0.5° cannot resolve such differences and therefore may provide, at best, a rather generalized picture of reality of hydrological change in mountain regions.

Model disagreements with observations in the winter months are concentrated along the transition areas between the large continental regions of predominately positive and negative trends. This result, which may be related to model resolution or to a spatial offset between a model parameterization based on coarse thematic maps of land properties and reality, may be expected. The result suggests that modelling of hydrological change in these areas is particularly

challenging and additional downscaling and improved representation of local hydrological processes are required to determine changes reliably. However, for the trends in the spring–autumn, as well as for derived indices such as high and low flow, the disagreements are not limited to these transition areas. This suggests that, rather than model resolution, potentially systematic errors either in the forcing data, parameter values or in the model concepts or physics may cause the deviation. The drier the conditions, the more important become evapotranspiration, catchment storage and release. The models apparently differ considerably in how they model these processes (Gudmundsson et al., 2012a), and this affects also their derived estimates of hydrological change.

A few isolated observations exist that seem to have trends different from the regional signal. Examples are two stations in Denmark and northern France and some in the UK. These catchments are all located in groundwater-dominated systems with possibly large storage carry-over between seasons or even longer. Numerous recent studies have demonstrated the role of groundwater storage in modulating climatic signals (e.g. Laizé and Hannah, 2010 in the UK; Fleig et al., 2011, in UK and Denmark; Van Loon and Van Lanen, 2012 for contrasting catchments in Europe, and in a global analysis by Van Lanen et al., 2012). The nature and magnitude of such storages are highly complex and variable, and dependent on aquifer characteristics, which are likely to be poorly replicated in the simplified storage schemes of large-scale hydrological models. Poorly reproduced September and October trends, i.e. a time when storages are being replenished after summer depletion, suggest that this may be a general model weakness that influences the representation of trends.

Every effort has been made to ensure validation against the best data available, using catchments with good quality data where streamflow is unaffected by alterations and abstractions. However, there may still be issues with individual catchments, which may affect these fine-scale differences. In Europe, in practice it is not possible to rule out local, unknown anthropogenic effects such as a reduction in low flow by nearby groundwater abstractions or an augmentation of low flow by discharges from water treatment plants or return flow from irrigation. Such indirect influences may be very local and, for the greater part of the year, negligible and unknown to data providers – further illustrating the importance of well-documented data and metadata on artificial influences on river flow regimes (e.g. Hannah et al., 2011).

5.2 Regional trend patterns and uncertainty

The large-scale hydrological models used in this study have shown considerable variability in their representation of the runoff trends. The variability is largest in the magnitudes of transient runoff trends, but there are also large areas where the trend direction differs for many of the indices, suggesting a high uncertainty even in the hindcasting of hydrological change. The generally larger variability among the models

found in the summer season and for low flow trends in regions where a validation was possible suggests a higher uncertainty in simulating trends under dry conditions. Overall, the model validation confirms the findings of other studies that the ensemble mean tends to outperform individual models, potential explanations for which include the averaging of model errors (e.g. Gao and Dirmeyer, 2006; Guo et al., 2007). Examples also include the evaluation of continental-scale summaries of high and low flow with the same datasets (Gudmundsson et al., 2012a), and numerous studies on runoff from large basins (e.g. Hagemann and Jacob, 2007; Matera et al., 2010) and other variables of the terrestrial water cycle (e.g. Guo et al., 2007).

Notwithstanding the inherent uncertainties, the maps of modelled trends elucidate spatial details not available from previously published observation-based trend maps from specific European analyses of modelled changes. Key “white space” now mapped with modelled trends includes persistent negative trends throughout the year in the southeast, the Iberian Peninsula and Italy. For the Iberian Peninsula, the few observations in the north show some deviations for the winter months, but otherwise agree with the modelled trends. Several other studies from the Iberian Peninsula have previously documented widespread negative runoff trends in which the climate component is still discernable despite a considerable additional impact of water management in the region (Lorenzo-Lacruz et al., 2012). For Italy and south-eastern Europe, no streamflow observations were available to confirm the modelled patterns, although the findings accord with climate trends, which have shown a long-term drying in these areas of the Mediterranean (Sousa et al., 2011). This further underlines the need for future extension of the observed streamflow dataset into southern regions of Europe in particular.

It should also be kept in mind that the trends were derived for the period 1963–2000, and any trend calculation depends strongly on the period of record, as highlighted in numerous previous studies (e.g. Chen and Grasby, 2009). Further work is underway by the authors to establish how representative this shorter period is of longer-term variability, using a selection of long (> 90 yr) records within the dataset (Hannaford et al., 2011). Whether the trends shown here are due to long-term variability or recent climate change (or a combination) remains to be investigated.

However, the spatial patterns throughout the year with more positive trends in winter and more negative trends in summer agree with other hydroclimatological studies as well as with future climate projections of a drier southern and wetter northern Europe (e.g. Bates et al., 2008; EEA-JRC-WHO, 2008). In addition, the north–south gradient may reflect the increased prevalence of the positive phases of the North Atlantic Oscillation (NAO) over the period considered by this study, which has previously been associated with increased rainfall, mean runoff and high flow (particularly in winter) in northern Europe (Shorthouse and Arnell, 1997;

Hannaford and Marsh, 2008), and concomitant decreases in rainfall and river flow in the Mediterranean (e.g. Lopez-Moreno and Vicente-Serrano, 2008; Lorenzo-Lacruz et al., 2011). The two factors are not mutually exclusive, as NAO variability may itself be heavily influenced by anthropogenic warming (Dong et al., 2011). A detailed attribution was beyond the scope of this paper; however, hopefully it will spark interest in the study of atmospheric causes for the specific patterns of change and variability found herein.

In a recent discussion paper on flood trends, Merz et al. (2012) argue that greater scientific rigour is needed in the attribution of streamflow changes. They note that many studies detect hydrological change in observed datasets, but fall short of proving and quantifying the relation to the drivers. Comparing climate and hydrology trends is not sufficient, and empirical approaches to establishing such a quantitative link, e.g. trend analysis with covariates (e.g. Stahl and Moore, 2006), suffer for example from the lack of data on land-cover and land-use changes (LULC). Modelling experiments have thus become a tool for attribution at different scales by using various model formulations based on different forcings and determining how well they reproduce observed historical streamflow patterns. Such an approach requires confidence into the model, and a study such as this one, which tests whether the model is able to reproduce observed trends in the first place, is an important first step. Attribution of transient changes would be a major challenge but could represent a significant advance in hydrological science.

Finally, it must be emphasised that this study attempts to “fill the white space” of knowledge on runoff changes by using modelled data for all of Europe, including both areas with and without observations. As such, this study presents a significant advance in spatial detail compared to previous trend assessments, but it is still only indicative of regional change due to the model inaccuracies and uncertainties discussed. Once confidence in such model estimates is high enough, a better way to fill the white space might be to combine observations and modelling, i.e. to merge both datasets to derive a composite map based on the spatially complete model outputs, combined with sparser, more reliable, observations. There are precedents to this approach; an analogy would be the composite datasets formed by the merging of rain-gauge observations and radar data (e.g. Severino and Alpuim, 2005), although clearly there are fundamental differences between continuous rainfall fields and trend statistics based on catchment streamflow. New methods would have to be developed to support the derivation of such composite datasets in the future.

6 Conclusions

Previous continental-scale assessments of recent hydrological change in Europe have had to rely on a sparse coverage of streamflow observations and disparate regional studies, which used very different methodologies and study periods. As such, there is typically “white space” on previously published maps of streamflow change, where observations are unavailable (e.g. southeastern and northeastern Europe), and limited consistency in published trend results across Europe. This study evaluated the potential for creating useful, high-resolution, continuous and regionally consistent maps of runoff changes for the whole of Europe by extrapolating in space using large-scale models, and thus demonstrated the current limitations of the approach. The results suggest that care should be taken when interpreting these maps for seasons and regions where temporary or long-term storage processes influence the propagation of climatic trends and for regions where the spatial variability is higher than that resolved by the models.

Generally, a higher confidence in model simulations should be sought through validation. Consequently, additional observation networks are encouraged to contribute to this important task of model validation at different scales to improve future confidence in areas where data availability or accessibility is currently limited – many of which are apparent “hot spots” where model simulations suggest that recent changes are strong and regionally consistent (for example, the general tendency towards strong drying trends in southeastern Europe).

The considerable variability among simulated trends for the different large-scale hydrological models is a strong reminder of the uncertainty of projected future changes in runoff if limited to only one such model. Specifically, where and when storage processes (including snow and groundwater) play a major role, and differences in how models store and release water can also cause large differences in the modelled transient hydrological response to the same climatic forcing signal. Thus, the study makes a case for multi-model approaches that include different land surface hydrology schemes, for the improvement of process conceptualization and resolution of large-scale hydrology models. The WATCH model output is available globally, and similar studies may be carried out on other continents. The results herein may potentially not only advance our understanding of hydrological change; they may further guide the assessment of the uncertainty of future scenario runs with those models, and thus assist future efforts to improve them.

Acknowledgements. Funding was provided by the EU WATCH project (WATER and global CHange), EC Priority Area “Global Change and Ecosystems”, contract number 036946. The study is also a contribution to UNESCO IHP-VII and the Euro-FRIEND project. The authors thank all data providers: national authorities for the observed streamflow records, the creators of WATCH

forcing data and hydrological multi-model output. The DFG (German Research Foundation) Fund for Open Access Publishing covered the publication costs.

Edited by: R. Merz

References

- Balsamo, G., Viterbo, P., Beljaars, A., van den Hurk, B., Hirschi, M., Betts, A. K., and Scipal, K.: A revised hydrology for the ECMWF model: verification from field site to terrestrial water storage and impact in the Integrated Forecast System, *J. Hydrometeorol.*, 10, 623–643, 2009.
- Bard, A., Renard, B., and Lang, M.: The AdaptAlp Dataset. Description, guidance and analyses, Final Report, UR HHLy, Hydrology-Hydraulics, Lyon, 15 pp., 2011.
- Bates, B. C., Kundzewicz, Z. W., Wu, S., and Palutikof, J. P. (Eds.): Climate Change and Water, Technical Paper of the Intergovernmental Panel on Climate Change, IPCC Secretariat, Geneva, Switzerland, 2008.
- Breuer, L., Huisman, J. A., Willems, P., Bormann, H., Bronstert, A., Croke, B. F. W., Frede, H., Gräff, T., Hubrechts, L., Jake-man, A. J., Kite, G., Lanini, J., Leavesley, G., Lettenmaier, D. P., Lindström, G., Seibert, J., Sivapalan, M., and Viney, N. R.: Assessing the impact of land use change on hydrology by ensemble modeling (LUCHEM) I: Model intercomparison of current land use, *Adv. Water Resour.*, 32, 129–146, doi:10.1016/j.advwatres.2008.10.003, 2009.
- Chen, Z. and Grasby, S. E.: Impact of decadal and century-scale oscillations on hydroclimate trend analyses, *J. Hydrol.*, 365, 122–133, 2009.
- Cornwell, A. and Harvey, L. D. D.: Soil moisture: a residual problem underlying AGCMs, *Climatic Change*, 84, 313–336, doi:10.1007/s10584-007-9273-0, 2007.
- Dai, A., Qian, T., Trenberth, K. E., and Milliman, J. D.: Changes in continental freshwater discharge from 1948 to 2004, *J. Climate*, 22, 2773–2791, 2009.
- Dankers, R. and Feyen, L.: Flood hazard in Europe in an ensemble of regional climate scenarios, *J. Geophys. Res.*, 114, D16108, doi:10.1029/2008JD011523, 2009.
- Déry, S. J., Stahl, K., Moore, R. D., Whitfield, P. H., Menounos, B., and Burford, J. E.: Detection of runoff timing changes in pluvial, nival and glacial rivers of Western Canada, *Water Resour. Res.*, 45, W04426, doi:10.1029/2008WR006975, 2009.
- Dirmeyer, P. A., Gao, X., Zhao, M., Guo, Z., Oki, T., and Hanasaki, N.: GSWP-2: Multimodel analysis and implications for our perception of the land surface, *B. Am. Meteorol. Soc.*, 87, 1381–1397, doi:10.1175/BAMS-87-10-1381, 2006.
- Dong, B. W., Sutton, R. T., and Woollings, T.: Changes of inter-annual NAO variability in response to greenhouse gases forcing, *Clim. Dynam.*, 37, 1621–1641, 2011.
- EEA: the European environment, State and outlook 2010, Water resources: quantity and flows, EEA Report, Copenhagen, Denmark, 2010.
- EEA-JRC-WHO: Impacts of Europe’s changing climate – 2008 indicator-based assessment, EEA Report No. 4/2008, European Environment Agency, Copenhagen, Denmark, 2008.

- Feyen, L. and Dankers, R.: Impact of global warming on streamflow drought in Europe, *J. Geophys. Res.*, 114, D17116, doi:10.1029/2008JD011438, 2009.
- Fleig, A. K., Tallaksen, L. M., Hisdal, H., and Hannah, D. M.: Regional hydrological drought in north-western Europe: linking a new Regional Drought Area Index with weather types, *Hydrol. Process.*, 24, 1163–1179, doi:10.1002/hyp.7644, 2011.
- Gao, X. and Dirmeyer, P. A.: A Multimodel Analysis, Validation, and Transferability Study of Global Soil Wetness Products, *J. Hydrometeorol.*, 7, 1218–1236, 2006.
- Gerten, D., Rost, S., von Bloh, W., and Lucht, W.: Causes of change in 20th century global river discharge, *Geophys. Res. Lett.*, 35, L20405, doi:10.1029/2008GL035258, 2008.
- Gudmundsson, L., Tallaksen, L. M., Stahl, K., Clark, D. B., Dumont, E., Hagemann, S., Bertrand, N., Gerten, D., Heinke, J., Hanasaki, N., Voß, F., and Koirala, S.: Comparing Large-scale Hydrological Model Simulations to Observed Runoff Percentiles in Europe, *J. Hydrometeorol.*, 13, 604–620, doi:10.1175/JHM-D-11-083.1, 2012a.
- Gudmundsson, L., Wagener, T., Tallaksen, L. M., and Engeland, K.: Evaluation of nine Large-Scale Hydrological Models with Respect to Seasonal Runoff Climatology in Europe, *Water Resour. Res.*, in revision, 2012b.
- Guo, Z., Dirmeyer, P. A., Gao, X., and Zhao, M.: Improving the quality of simulated soil moisture with a multi-model ensemble approach, *Q. J. Roy. Meteorol. Soc.*, 133, 731–747, doi:10.1002/qj.48, 2007.
- Haddeland, I., Clark, D. B., Franssen, W., Ludwig, F., Voß, F., Arnell, N. W., Bertrand, N., Best, M., Folwell, S., Gerten, D., Gomes, S., Gosling, S. N., Hagemann, S., Hanasaki, N., Harding, R., Heinke, J., Kabat, P., Koirala, S., Oki, T., Polcher, J., Stacke, T., Viterbo, P., Weedon, G. P., and Yeh, P.: Multi-model estimate of the global water balance: setup and first results, *J. Hydrometeorol.*, 12, 869–884, doi:10.1175/2011JHM1324.1, 2011.
- Hagemann, S. and Jacob, D.: Gradient in the climate change signal of European discharge predicted by a multi-model ensemble, *Climatic Change*, 81, 309–327, doi:10.1007/s10584-006-9225-0, 2007.
- Hannaford, J. and Marsh, T. J.: High flow and flood trends in a network of undisturbed catchments in the UK, *Int. J. Climatol.*, 28, 1325–1338, 2008.
- Hannaford, J., Newton, G., Stahl, K., and Tallaksen, L. M.: Decadal-scale river flow variability in Europe: evidence from long hydrometric records spanning 1900–2004, *Geophysical Research Abstracts*, 13, EGU2011-12053, <http://meetingorganizer.copernicus.org/EGU2011/EGU2011-12053.pdf> (last access: 7 July 2012), 2011.
- Hannah, D. M., Demuth, S., van Lanen, H. A. J., Looser, U., Prudhomme, C., Rees, G., Stahl, K., and Tallaksen, L. M.: Large-scale river flow archives: importance, current status and future needs, Invited Commentary, *Hydrol. Process.*, 25, 1191–1200, doi:10.1002/hyp.7794, 2011.
- Klein Tank, A. M. G., Wijngaard, J. B., Konnen, G. P., Bohm, R., Demaree, G., Gocheva, A., Mileta, M., Pashiardis, S., Hejkrlik, L., Kern-Hansen, C., Heino, R., Bessemoulin, P., Muller-Westermeier, G., Tzanakou, M., Szalai, S., Palsdottir, T., Fitzgerald, D., Rubin, S., Capaldo, M., Maugeri, M., Leitass, A., Bukantis, A., Aberfeld, R., Van Engelen, A. F. V., Forland, E., Miletus, M., Coelho, F., Mares, C., Razuvaev, V., Nieplova, E., Cegnar, T., López, J. A., Dahlstrom, B., Moberg, A., Kirchhofer, W., Ceylan, A., Pachaliuk, O., Alexander, L. V., and Petrovic, P.: Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment, *Int. J. Climatol.*, 22, 1441–1453, 2002.
- Laizé, C. R. L. and Hannah, D. M.: Modification of climate-river flow associations by basin properties, *J. Hydrol.*, 389, 186–204, 2010.
- Lehner, B., Döll, P., Alcom, J., Henrichs, T., and Kaspar, F.: Estimating the impact of global change on flood and drought risk in Europe: a continental integrated analysis, *Climatic Change*, 75, 273–299, 2006.
- Lopez-Moreno, J. and Vicente-Serrano, S. M.: Positive and negative phases of the wintertime North Atlantic Oscillation and drought occurrence over Europe, A multitemporal-scale approach, *J. Climate*, 21, 1220–1243, 2008.
- Lorenzo-Lacruz, J., Vicente-Serrano, S. M., López-Moreno, J. I., González-Hidalgo, J. C., and Morán-Tejeda, E.: The response of Iberian rivers to the North Atlantic Oscillation, *Hydrol. Earth Syst. Sci.*, 15, 2581–2597, doi:10.5194/hess-15-2581-2011, 2011.
- Lorenzo-Lacruz, J., Vicente-Serrano, S. M., Lopez-Moreno, J. J., Moran-Tejeda, E., and Zabalza, J.: Recent trends in Iberian streamflows, 1945–2005, *J. Hydrol.*, 414–415, 463–475, doi:10.1016/j.jhydrol.2011.11.023, 2012.
- Materia, S., Dirmeyer, P. A., Guo, Z., Alessandri, A., and Navarra, A.: The sensitivity of simulated river discharge to land surface representation and meteorological forcings, *J. Hydrometeorol.*, 11, 334–351, doi:10.1175/2009JHM1162.1, 2010.
- McCabe, G. J. and Wolock, D. M.: Independent Effects of Temperature and Precipitation on Modeled Runoff in the Conterminous United States, *Water Resour. Res.*, 47, W11522, doi:10.1029/2011WR010630, 2011.
- Merz, B., Vorogushyn, S., Uhlemann, S., Delgado, J., and Huncheda, Y.: *HESS Opinions* “More efforts and scientific rigour are needed to attribute trends in flood time series”, *Hydrol. Earth Syst. Sci.*, 16, 1379–1387, doi:10.5194/hess-16-1379-2012, 2012.
- Milly, P. C. D., Dunne, K. A., and Vecchia, A. V.: Global pattern of trends in streamflow and water availability in a changing climate, *Nature*, 438, 347–350, 2005.
- Prudhomme, C., Parry, S., Hannaford, J., Clark, D. B., Hagemann, S., and Voss, F.: How well do large-scale models reproduce regional hydrological extremes in Europe?, *J. Hydrometeorol.*, 12, 1181–1204, 2011.
- Severino, E. and Alpuim, T.: Spatiotemporal models in the estimation of area precipitation, *Environmetrics*, 16, 773–802, 2005.
- Shorthouse, C. A. and Arnell, N. W.: Spatial and temporal variability on European river flows and the North Atlantic Oscillation, *FRIEND’97-Regional Hydrology: Concepts and Models for Sustainable Water Resource Management*, IAHS. Publ., 246, 77–85, 1997.
- Smirnov, N. V.: Tables for estimating the goodness of fit of empirical distributions, *Ann. Math. Stat.*, 19, 279–281, 1948.
- Sousa, P. M., Trigo, R. M., Aizpurua, P., Nieto, R., Gimeno, L., and Garcia-Herrera, R.: Trends and extremes of drought indices throughout the 20th century in the Mediterranean, *Nat. Hazards Earth Syst. Sci.*, 11, 33–51, doi:10.5194/nhess-11-33-2011, 2011.

- Stahl, K. and Moore, R. D.: Influence of watershed glacier coverage on summer streamflow in British Columbia, Canada, *Water Resour. Res.*, 42, W06201, doi:10.1029/2006WR005022, 2006.
- Stahl, K., Hisdal, H., Hannaford, J., Tallaksen, L. M., van Lanen, H. A. J., Sauquet, E., Demuth, S., Fendekova, M., and Jódar, J.: Streamflow trends in Europe: evidence from a dataset of near-natural catchments, *Hydrol. Earth Syst. Sci.*, 14, 2367–2382, doi:10.5194/hess-14-2367-2010, 2010.
- Stahl, K., Tallaksen, L. M., Gudmundsson, L., and Christensen, J. H.: Streamflow data from small basins: a challenging test to high resolution regional climate modeling, *J. Hydrometeorol.*, 12, 900–912, doi:10.1175/2011JHM1356.1, 2011.
- Theil, H.: A rank-invariant method of linear and polynomial regression analysis, *Indag. Math.*, 12, 85–91, 1950.
- Van Huijgevoort, M. H. J., Van Loon, A. F., Rakovec, O., Hadeland, I., Horáček, S., and Van Lanen, H. A. J.: Drought assessment using local and large-scale forcing data in small catchments, in: *Global Change: Facing Risks and Threats to Water Resources*, edited by: Servat, E., Demuth, S., Dezetter, A., Daniell, T., Ferrari, E., Ijjaali, M., Jabrane, R., Van Lanen, H., and Huang, Y., IAHS Publ. No. 340, 77–85, 2010.
- Van Lanen, H. A. J., Wanders, N., Tallaksen, L. M., and Van Loon, A. F.: Hydrological drought across the world: impact of hydro-climatology and physical catchment structure, *Int. J. Climatol.*, in review, 2012.
- Van Loon, A. F. and Van Lanen, H. A. J.: A process-based typology of hydrological drought, *Hydrol. Earth Syst. Sci.*, 16, 1915–1946, doi:10.5194/hessd-8-11413-2011, 2011.
- Viglione, A., Borga, M., Balanbanis, P., and Blöschl, G.: Barriers to the exchange of hydrometeorological data in Europe: Results from a survey and implications for data policy, *J. Hydrol.*, 394, 63–77, doi:10.1016/j.jhydrol.2010.03.023, 2010.
- Weedon, G. P., Gomes, S., Viterbo, P., Shuttleworth, W. J., Blyth, E., Sterle, H., Adam, J. C., Bellouin, N., Boucher, O., and Best, M.: Creation of the WATCH forcing data and its use to assess global and regional reference crop evaporation over land during the twentieth century, *J. Hydrometeorol.*, 12, 823–848, doi:10.1175/2011JHM1369.1, 2011.
- Wilson, D., Hisdal, H., and Lawrence, D.: Has streamflow changed in the Nordic countries? – Recent trends and comparisons to hydrological projections, *J. Hydrol.*, 394, 334–346, 2010.
- Zhang, X., Zwiers, F. W., Hegerl, G. C., Lambert, F. H., Gillett, N. P., Solomon, S., Stott, P. A., and Nozawa, T.: Detection of human influence on twentieth-century precipitation trends, *Nature*, 448, 461–465, 2007.
- Zolina, O., Zimmer, C., Gulev, S. K., and Kollet, S.: Changing structure of European precipitation: Longer wet periods leading to more abundant rainfalls. *Geophys. Res. Lett.*, 37, L06704, doi:10.1029/2010GL042468, 2010.