

How well do large-scale models reproduce regional hydrological extremes in Europe?

Christel Prudhomme^{1a}, Simon Parry^a, Jamie Hannaford^a, Douglas B. Clark^a, Stefan Hagemann^b, Frank Voss^c

- 1 ^a, Centre for Ecology and Hydrology, Wallingford, Oxfordshire, OX10 8BB, United
- 2 Kingdom
- 3 ^b, Max Planck Institute for Meteorology, Bundesstrasse 53, 20146 Hamburg, Germany
- 4 ^c, Universität-Kassel, CESR - Center for Environmental Systems Research, Kurt-Wolters-
- 5 Str. 3, D-34109 Kassel, Germany

Revised version
Wednesday, 16 March 2011

Prepared for the Journal of Hydrometeorology for the 'Water and Global Change special collection'

¹ *Corresponding author address*: Christel Prudhomme, Centre for Ecology and Hydrology, Wallingford, Oxfordshire, UK OX10 8BB
E-mail: chrp@ceh.ac.uk

Abstract

This paper presents a new methodology for assessing the ability of gridded hydrological models to reproduce large-scale hydrological high and low flow events (as a proxy for hydrological extremes) as described by catalogues of historical droughts (using the Regional Deficiency Index, RDI) and high flows (Regional Flood Index, RFI) previously derived from river flow measurements across Europe. Using the same methods, total runoff simulated by three global hydrological models from WaterMIP (JULES, WaterGAP and MPI-HM) run with the same meteorological input (Watch Forcing Data) at the same spatial 0.5° grid was used to calculate simulated RDI and RFI for the period 1963-2001 in the same European regions, directly comparable with the observed catalogues. Observed and simulated RDI and RFI time series were compared using three performance measures: the relative mean error, the ratio between the standard deviation of simulated over observed series, and the Spearman correlation coefficient. Results show that all models can broadly reproduce the spatio-temporal evolution of hydrological extremes in Europe to varying degrees. JULES tends to produce prolonged, highly spatially coherent events for both high and low flows, with events developing more slowly, and reaching and sustaining greater spatial coherence than observed – this could be due to runoff being dominated by slow-responding sub-surface flow. In contrast, MPI-HM shows very high variability in the simulated RDI and RFI time series and a more rapid onset of extreme events than observed, in particular for regions with significant water storage capacity – this could be due to possible under-representation of infiltration and groundwater storage, with soil saturation reached too quickly. WaterGAP shares some of the issues of variability with MPI-HM - also attributed to insufficient soil

storage capacity and surplus effective precipitation being generated as surface runoff – and some strong spatial coherence of simulated events with JULES, but neither of these are dominant. Of the three global models considered here, WaterGAP is arguably best suited to reproduce most regional characteristics of large-scale high and low flow events in Europe. Some systematic weaknesses emerge in all models, in particular for high flows, which could be a product of poor spatial resolution of the input climate data (e.g. where extreme precipitation is driven by local convective storms) or topography. Overall, this study has demonstrated that RDI and RFI are powerful tools which can be used to assess how well large-scale hydrological models reproduce large-scale hydrological extremes – an exercise rarely undertaken in model inter-comparisons.

1. Introduction and background

There is growing evidence that the hydrological cycle is intensifying (e.g. Huntington, 2006; Stott *et al.*, 2010) as a result of anthropogenically-forced climatic change.

Generally speaking, at regional to continental scales, two contrasting approaches are used to examine the influence of climate change on the hydrological cycle:

(i) through analysis of historical data, to detect emerging trends (e.g. in Europe, by Stahl *et al.*, 2010; in North America by Douglas *et al.*, 2000, Khaliq *et al.*, 2008; globally by Dai *et al.*, 2009)

(ii) through modelling approaches (e.g. European analyses of future flood risk, Dankers and Feyen, 2008, 2009; global analysis of droughts by Burke *et al.* 2006).

The latter attempts to reproduce observed variability in hydrological characteristics using process-driven models, which can then be used in combination with climate prediction to provide assessments of potential hydrological changes. Wilby *et al.* (2008) observe that the two approaches often yield conflicting results, but that this “conceptual controversy” is in fact partly due to the different philosophies and specific methodologies underlying these two methods. Achieving greater reconciliation between models and observations, in order to better understand the influence of climate change on the hydrological cycle, is therefore one of the most pressing challenges of contemporary hydrology.

Whilst there is inevitably much attention on aspects of the hydrological cycle which are directly influenced by anthropogenic forcing (e.g. increased evaporation through higher

1 temperatures; increased rainfall as a result of higher moisture holding capacity in a
2 warmer atmosphere), one of the most important potential impacts of climate change is on
3 hydrological extremes, i.e. drought and flooding. Extremes are likely to be sensitive to
4 climate change, raising the possibility that changes in the extremes of hydrological
5 parameters may be more detectable than changes in mean conditions (Wilby et al. 2008;
6 Sheffield and Wood, 2008a). Furthermore, changes to extremes are of significant
7 importance to society; in reality, researchers, policy-makers and planners are more
8 interested in hydrological extremes than mean conditions. Droughts and floods have had
9 significant impacts in Europe in the recent past: the cost of the 2003 drought in Europe to
10 the agricultural and forestry sectors was estimated at 13.1 billion Euros, and the drought
11 and heat-wave caused an estimated 15,000 excess deaths (Fink et al. 2004); the 2007
12 floods which affected impacted the UK caused 15 fatalities (Marsh and Hannaford, 2008)
13 and cost £3.2billion (Environment Agency, 2010) while the 2002 floods on the Elbe river
14 resulted in 19 fatalities and cost 9.2 billion Euros (BMI, 2002). Clearly, any future
15 increase in the severity and frequency of hydrological extremes could have potentially
16 far-reaching socio-economic and environmental consequences.

17

18 Climate change influences on the hydrological cycle are often studied using large-scale
19 gridded models, which enable elements of the hydrological cycle (e.g. rainfall,
20 evapotranspiration, runoff) and potential perturbations to the cycle as a result of climate
21 change to be studied on regional (e.g. across Europe) to global scales. Such studies have
22 predominantly analysed climatic variables such as temperature and precipitation (e.g.
23 Burke et al., 2006) or land surface variables such as soil moisture (e.g. Sheffield and

1 Wood, 2007, 2008b; Sheffield et al., 2009). However, whilst some of the most pressing
2 direct impacts of droughts and floods are primarily related to runoff and river flows,
3 comparatively little effort has focused on examining runoff extremes in large-scale
4 models. A number of studies have examined runoff from global models at continental
5 scales, but these studies have tended to focus on mean runoff (e.g. Milly et al., 2005;
6 Gedney et al., 2006; Hamlet et al., 2007). Two notable exceptions are research by Milly
7 et al. (2002) who compared the frequency of large floods, globally, in the 20th century to
8 future scenarios, and Lehner et al. (2006) who examined future flood and drought
9 occurrence for Europe using a global hydrological model. At a finer spatial scale, Feyen
10 and Dankers (2009) and Dankers and Feyen (2008) investigated future changes in high
11 flows and drought in Europe using a hydrological model driven by a Regional Climate
12 Model (RCM), followed by a multi-model ensemble (Dankers and Feyern, 2009).
13 Generally, however, high or low flow extremes have been under-examined.

14

15 Results from all modelling studies are sensitive to the models used, and model outputs
16 must be corroborated against observed data if they are to be used with confidence. Most
17 of the above studies perform some validation of model outputs against observed data,
18 often based on comparing modelled flow at a limited number of catchment outlets against
19 equivalent observed data from relatively large basins. Haddeland et al. (accepted) detail a
20 model intercomparison study comparing the performance of eleven models that simulate
21 hydrological processes on a global scale. Their study, in common with previous model
22 intercomparison studies (see references therein), focused on comparing mean and median
23 runoff. A focus on the means also underpins the study undertaken for North America by

1 Lohmann et al. (2004). In general, however, to the authors' knowledge, there have been
2 very few attempts to systematically validate runoff extremes from large-scale models
3 against observed data in Europe or elsewhere. However, an appraisal of the performance
4 of large-scale models in replicating historical hydrological extremes is a necessary
5 precursor to assessing the suitability of such models for projecting characteristics of
6 hydrological extremes into the twenty-first century.

7

8 This paper complements other model intercomparison studies by examining the
9 performance of three large-scale hydrological models in reproducing observed
10 streamflow in Europe, with a focus on the extreme (high and low) flows. Drought and
11 flood characteristics vary spatially and temporally throughout Europe, and it is important
12 to consider this variability when characterising observed hydrological extremes at the
13 European scale. This study therefore uses published drought and flood catalogues based
14 on the Regional Deficiency Index (RDI) and the Regional Flood Index (RFI), time series
15 indicating the regional extent of hydrological extremes. Hannaford et al. (2010) applied
16 the RDI methodology to produce "drought catalogues" for 23 European regions. For
17 high flows, Parry et al. (2010) applied the same conceptual approach to propose the RFI
18 methodology as a proxy for large-scale flood events. In addition to the existence of
19 published historical European catalogues, the RDI and RFI methodologies have several
20 distinct advantages for evaluating global model performance in terms of hydrological
21 extremes: (i) the RDI and RFI historical catalogues are based on small, undisturbed
22 catchments. As argued by Stahl et al. (2010), who used the same dataset for a pan-
23 European trend assessment, this dataset provides a natural response and a level of detail

1 which is not normally captured by large basins, against which global model performance
2 is conventionally judged; (ii) the method focuses on regional analysis, meaning that local-
3 scale extremes (which could be driven by specific combinations of local meteorological
4 events and catchment properties that are not necessarily representative of the whole
5 region) are smoothed out, while large-scale generating mechanisms are captured;
6 furthermore the spatial scale of the regional time series is consistent with that of the
7 global models; (iii) RDI and RFI are derived from time series anomalies: when applied
8 to global models, the anomalies are based on the internal variability of each model-
9 simulated flow, hence eliminating the effects of systematic biases in the models.

10

11 The analyses of simulated RDI and RFI thus concentrate on evaluating whether large-
12 scale high and low flow events are reproduced by the global models at the right time, and
13 with the right intensity, duration and variability, compared to the observed data. Note,
14 however, that whilst historical RDI and RFI have been shown to be powerful indicators
15 of regional extremes, (i) they are not extremes *per se*, as they are defined here as 10%
16 anomalies; (ii) they are subject to limitations (for discussion see Hannaford et al., 2010),
17 due to the uneven distribution of hydrological records used to derive the drought and
18 flood catalogues. Due to these limitations, the RDI and RFI from even a ‘perfect’ model
19 might not be identical to the regional observed catalogues, but the comparison of
20 simulated indices with observed indices should provide a solid basis on which to
21 consistently evaluate performance of multiple global models.

22

1 This paper compares outputs from three contrasting large scale hydrological models, all
2 driven by the same meteorological forcing data, with observed regional high and low
3 flow catalogues over the period 1963 – 2000 in an attempt to assess the validity of their
4 reproduction of regional runoff ‘extremes’ in Europe. These global models are part of the
5 model inter-comparison project WaterMIP (Haddeland et al., accepted). The current
6 study uses the European RDI and RFI historical catalogues as a benchmark, to assess and
7 compare how well large-scale hydrological models reproduce large-scale hydrological
8 phenomena in Europe. A complete explanation of the results in terms of the formulation
9 and process descriptions of the models, and extending to other models, are beyond the
10 scope of this paper. Future work, which is already underway, will examine these issues
11 in more detail.

12

13 2. Data and Methodology

14 *Observed Data*

15 The streamflow dataset used to validate model outputs is based on an updated version of
16 the European Water Archive, a dataset of daily streamflow records for ten countries
17 across Europe, the derivation of which is discussed by Stahl et al. (2010). The dataset
18 comprises catchments with minimal anthropogenic disturbances on flow regimes,
19 including gauging stations regarded to have good hydrometric performance and records
20 from 1961 - 2005. An additional selection criterion was that all catchments should be
21 smaller than 1000 km² (to ensure a likelihood of minimal disturbance), although in a
22 small number of cases larger catchments were included if influences could be proved to

1 be negligible. Additional data from undisturbed catchments were sourced from the UK
2 benchmark catchments (Hannaford & Marsh, 2008) and from Banque Hydro in France
3 (Prudhomme & Sauquet, 2006). The total dataset consists of 579 gauging stations. The
4 distribution of stations over Europe is somewhat uneven (see Figure 1) with high
5 densities of stations in some areas (e.g. Germany) and limited data in areas which are
6 heavily affected by anthropogenic disturbances (e.g. northern France and the Benelux
7 countries). No data were available across the majority of southern or eastern European
8 countries.

9 *Simulated Data*

10 For this study, outputs from three of the WaterMIP models were used (JULES,
11 WaterGAP and MPI-HM) covering the range of model types discussed in Haddeland et al
12 (accepted). JULES is the land surface component of a climate model, while WaterGAP
13 has been developed to study water resources, and MPI-HM is an intermediate type. Their
14 main characteristics, including input meteorological variables and the schemes used for
15 evapotranspiration, runoff generation and snow melt processes are summarised in
16 Table 1, while brief descriptions of the models are provided below.

17

18 All global models have the same 0.5° spatial resolution and were run over the period
19 1963-2001 using the same meteorological input data, the Watch Forcing Data (WFD,
20 Weedon et al., 2010). As the models have originally been developed for global
21 simulation, it is difficult to accurately assess the minimum scale they can work on. This
22 is likely to depend on the accuracy of the input data, but also on the catchment
23 characteristics and dominant processes in the catchment (which can vary in time even at

1 the same location). In particular, processes with substantial variability at smaller scales
2 than the 0.5° spatial resolution of the input data used here may not be well captured –
3 perhaps more significantly for processes dependant on orography (e.g. in the Alps) and/or
4 meteorology (e.g. convective storms) than for the energy balance processes, for which all
5 models have a representation of subgrid variability. For the study, our main assumption
6 is that runoff simulations at 0.5° are generally reasonable. The simulations considered
7 ‘naturalised’ conditions, when direct anthropogenic effects such as dams and water
8 abstraction were not included in the models. This is consistent with the use of
9 observations from undisturbed catchments. For comparison with the observed data, total
10 runoff (the sum of surface and subsurface flows) was used as reference data to generate
11 simulated RDI and RFI daily time series. Runoff, rather than routed discharge, was
12 considered so that the individual anomalies of all grids in the regions, and hence their
13 spatial pattern, were accounted for without any smoothing added by the hydrological
14 routing. WaterGAP applies a correction factor on runoff to match observed river
15 discharge in major rivers across the globe and evapotranspiration is adjusted accordingly.
16 Neither JULES nor MPI-HM were calibrated for this exercise, although they may have
17 been calibrated for previous studies.

18

19 The Joint UK Land Environment Simulator (JULES, Best et al., submitted, Clark et al.,
20 submitted) is the land surface scheme used in the climate models of the UK Met Office
21 Hadley Centre. It includes mechanistic descriptions of the processes that control the
22 exchanges of energy, momentum, water and carbon between the land surface and the
23 atmosphere. The energy balance of the surface is calculated on a timestep of one hour or

1 less and there is a multi-layer snow model. Fluxes of water and heat in the soil are
2 represented using four soil layers with a total depth of 3m. In the configuration used in
3 this exercise, runoff can occur through two main mechanisms: infiltration excess surface
4 runoff and drainage through the bottom of the soil column. Drainage is calculated as a
5 Darcian flux assuming zero gradient of matric potential (Best et al, submitted).

6

7 MPI-HM was developed at the Max Plank Institute for Meteorology based on research
8 from Hagemann & Dümenil (1998) and Hagemann & Dümenil Gates (2003). In this
9 model, lateral flow movements are considered at the global scale. A simplified land-
10 surface scheme (SL) is used in combination with a hydrological discharge model (HD).
11 SL uses daily time series of precipitation and temperature input, and the soil is
12 represented as a single layer. HD describes overland flow, river flow and base flow
13 movement based on the spatial distribution of topography, slope, lakes and wetlands
14 within each grid box.

15

16 The hydrological modelling system WaterGAP (Water - Global Assessment and
17 Prognosis) has been developed to simulate the distribution and availability of water
18 resources at the global scale. WaterGAP combines a global hydrological model (Döll et
19 al. 2003) with several global water use models (Flörke and Alcamo 2004, Alcamo et al.
20 2003, Döll and Siebert 2002). The global hydrological model simulates the continental
21 water cycle incorporating spatially distributed physiographic characteristics and climate
22 factors. The global water use model estimates water withdrawals and water consumption
23 for households, manufacturing, energy production, livestock and irrigation taking into

1 account basic socio-economic factors. WaterGAP was calibrated against more than 1560
2 stations worldwide, based on observed discharge data provided by the GRDC 2009
3 (Global Runoff Data Centre, Koblenz, Germany) and the WFD data set. The minimum
4 drainage basin area considered for calibration was 9000 km² and intermediate catchment
5 areas were generally larger than 20,000 km². Depending on the data availability for each
6 gauging station the model was calibrated to discharge time series of thirty years. Due to
7 long computing times WaterGAP was calibrated for one parameter, the runoff coefficient
8 (see Table 1).

9

10 *Large-scale Deficiency and Exceedance Index time series*

11 The methodology used in this paper follows that of the European catalogue of regional
12 droughts in Europe (Lloyd-Hughes et al., 2010; Hannaford et al., 2010) and is based on
13 the RDI concept, first introduced by Stahl (2001) and modified to be applied to high
14 flows (RFI) by Parry et al. (2010).

15

16 The construction of the regional time series of RDI follows three steps (the alternative
17 methodology for RFI is given in parentheses):

18 1) Definition of daily Deficiency (Exceedance) Index time series, subsequently referred
19 to as DI and EI respectively. For each available river flow record, the flow on a given
20 day is compared to a daily-varying low (high) flow threshold, and is replaced by a
21 single index equal to 1 when the flow is less (greater) than or equal to the threshold,
22 and 0 otherwise (Equation 1). By construction, the resulting DI (EI) time series are
23 binary. The use of a local daily varying threshold removes the influence of

1 streamflow seasonality and geographical disparity between records. In this study, the
 2 flow exceeded 90% (10%) of the time, Q90 (Q10), is used as a threshold and is
 3 defined as follows: for a given Julian day d , the daily-varying Q90(d) (Q10(d)) is
 4 calculated by ranking all historical flow values across the 31 days centred on day d
 5 (i.e. 15 days either side). The 31-day window increases the size of the sample and
 6 ensures a more robust estimate of the flow duration curves (ranked flows) and thus
 7 Q90 (d) and Q10 (d).

$$\begin{array}{ll}
 8 \quad DI(d) = 1 & \text{if } Q(d) \leq Q90(d) \\
 DI(d) = 0 & \text{if } Q(d) > Q90(d)
 \end{array}
 \quad \text{And} \quad
 \begin{array}{ll}
 EI(d) = 1 & \text{if } Q(d) \geq Q10(d) \\
 EI(d) = 0 & \text{if } Q(d) < Q10(d)
 \end{array}
 \quad (1)$$

9 2) Sites experiencing a deficiency (exceedance) at the same time are grouped into
 10 homogeneous regions using cluster analysis (Stahl, 2001 and Prudhomme and
 11 Sauquet, 2006). The clustering was performed by applying an agglomerative
 12 (hierarchical) technique using the Ward method and a binary Euclidean distance
 13 measure. Tests in France showed that groups are relatively robust to the measurement
 14 period. Small manual adjustments were made to ensure that regions are spatially
 15 continuous in their extent (Hannaford et al. 2010; Prudhomme & Sauquet, 2006). Note
 16 that work by Gudmundsson et al. (2010) suggests that there is little difference in
 17 spatial structure of clusters grouped according to annual Q95 and Q90 (Q5 and Q10)
 18 for low (high) flow. The homogeneity of a cluster c $H(c)$ is empirically calculated
 19 (Equation 2) as the departure between the theoretical cumulative density of RDI(RFI)
 20 and the empirical cumulative distribution $F(x)$; by construction, 90% of the RDI (RFI)
 21 series should be equal to 0, and 10% should be equal to 1, if all stations had identical
 22 DI(EI) time series.

$$H(c) = \int_0^1 |F(x) - p| \quad (2)$$

With $H(c)$: homogeneity index, $F(x)$ cumulative distribution of the RDI(RFI) of cluster c ; p : 90%. $H(c)$ close to zero suggests high homogeneity.

3) Regional Deficiency (Flood) Index time series are defined for each homogeneous region as the arithmetic mean of the DI (EI) series of all sites within the regions for each day of record (Equation 3). The RDI (d) (RFI (d)) represents the proportion of catchments in the region that experience low (high flow) flows that fall below (rise above) the threshold on the same Julian day d . A maximum value of 1 represents a flow deficiency (exceedance) at all sites in the region, thus indicating a large-scale low (high) flow event.

$$RDI(d) = \frac{1}{M} \sum_{i=1}^M DI_i(d) \quad \text{and} \quad RFI(d) = \frac{1}{M} \sum_{i=1}^M EI_i(d) \quad (3)$$

with M , number of records in the region.

Historical RDI and RFI catalogues are derived from the 579 streamflow records discussed earlier and based on 23 homogenous regions modified from Hannaford et al. (2010) where South East Great Britain is a single geographic region instead of divided according to groundwater influence (Figure 1).

Simulated RDI and RFI time series are derived using an identical procedure as outlined above, but applied to the total runoff (the sum of surface and subsurface runoff) with thresholds (Q_{90} and Q_{10}) calculated for each grid cell. Once simulated DI and EI time series are generated for each grid cell, simulated RDI and RFI daily time series are

1 calculated (using eq. 3) by considering all cells whose centroid is within the geographical
2 boundaries of the 23 European regions.

3 It should thus be noted that the use of the RDI and RFI means that the procedure does not
4 assess the magnitude of runoff compared to observed discharge, but evaluates whether
5 the periods of simulated and observed high and low flows occur at the same time and
6 with the same spatial extent as observed.

7 3. Results

8 Three complementary analyses were undertaken to assess the ability of the large-scale
9 models to reproduce large-scale hydrological extremes:

- 10 • Statistical analysis of RDI and RFI time series derived from river flow observations
11 (the historical catalogues) against those derived from simulated runoff. Three goodness-
12 of-fit measures are used and shown in Table 2 for six case study regions:
 - 13 i. Relative mean error (RME, %) as a measure of the overall bias in the model;
 - 14 ii. Ratio between the standard deviation of simulated and observed RDI/RFI as a
15 measure of the overall reproduction of the spread in the series;
 - 16 iii. Spearman (ranked) correlation (ρ ; Spearman, 1944) between observed and
17 simulated RDI/RFI, as a measure of the overall reproduction of the timing of the
18 events. This measure is preferred to Pearson correlation analysis as it does not
19 assume a Gaussian distribution for the time series.
- 20 • Graphical analyses of the complete RDI and RFI time series for six contrasting
21 regions. This allows rapid identification of whether the major historical large-scale
22 events have been simulated;

1 • For two specific events (the high flows of 2000, and the drought of 1975-1976),
2 detailed comparisons of the RDI and RFI time series. This allows the temporal structure
3 of the simulated RDI and RFI to be highlighted at a very fine resolution and enables the
4 identification of differences between the three models during notable high/low flow
5 periods.

6 *3.1 Statistical measures*

7 Below we discuss each measure of goodness-of-fit, with values for six selected regions
8 being shown in Table 2.

9 *3.1.1 Relative Mean Error*

10 While, by construction, DI (EI) time series for each cell are populated by 10% of 1 and
11 90% of 0, only a region in which the DI(EI) time series are identical within all cells has
12 RDI (RFI) comprising 10% of 1 and 90% of 0. In reality, RDI (RFI) time series have a
13 value of 1 only when the whole region experiences an anomaly on the same day (i.e.
14 when DI/EI of all stations or all grid cells belonging to that region are equal to 1 for that
15 day). The magnitude of RDI (RFI) expresses the spatial coherence of the low (high) flow
16 event, with the duration (e.g. time with index above threshold) depicting the length of the
17 event. The mean relative error hence measures whether a model tends to overestimate the
18 spatial coherence (i.e. there are more high RDI/RFI values than observed; positive RME)
19 or underestimate it (negative RME) and by how much.

20

21 Generally, RME is positive for RDI and negative for RFI (Table 2), suggesting all models
22 tend to overestimate the spatial coherence and length of drought events but underestimate

1 high flow events. The model with overall RME of smallest magnitude was MPI-HM, for
2 both RDI (2.43% overall) and RFI (-2.42%). RME is of similar magnitude and sign for
3 JULES and WaterGAP, larger for RDI (overall +7.98 and +9.27 respectively) than for
4 RFI (overall -4.22 and -3.18 respectively). There is a marked difference in the
5 performance by regions (not shown), with S France, French Alps, High Alps and E
6 Germany amongst the regions with the largest RME (between 15% and 22% for RDI for
7 JULES and WaterGAP), while NW and SW Great Britain, N France and N Germany
8 show the lowest RME for JULES and WaterGAP (less than 10% for RDI) but not always
9 for MPI-HM.

10 *3.1.2. Reproduction of the spread*

11 The variability of RDI or RFI time series depicts the level of spatial coherence of
12 low/high flow events: as already discussed, the lowest variability would be achieved for a
13 region where low/high flow anomalies occur simultaneously. With spatial coherence of
14 events reducing, their number increases as different areas of the region experience a flow
15 anomaly at different time [by definition 10% of days have DI/EI associated with a flow
16 anomaly]. The larger the standard deviation, the less coherent, the more numerous, and,
17 by extension, the shorter the events. For RDI, all three models overestimate the standard
18 deviation of the time series in all regions but NW Great Britain and S Scandinavia, and
19 for WaterGAP French S Alps (not shown). The largest overestimation of the variability
20 is found for MPI-HM but exceeds 30% for all three models in four regions (albeit
21 different ones), and is greater than 10% for 14 (JULES and MPI-HM) or 15 regions
22 (WaterGAP). In contrast, overall time series variability is better reproduced for RFI, and
23 within +/- 10% for 7 (JULES), 9 (WaterGAP) and 10 (MPI-HM) regions. This could be

1 because high flow events are already shorter and less spatially coherent than low flow
2 episodes. High flow events are also often predominantly driven by precipitation, which is
3 a direct input to the models and so a good estimate of precipitation will tend to force
4 realistic variability of high flows. In general, all of the models (and in particular MPI-
5 HM) tend to generate high and low flow episodes that are shorter and not as spatially
6 coherent as those observed in Europe. Note that JULES is the only model which
7 underestimates RFI variability by over 10% in two regions (S France, western and central
8 France), suggesting it generates high flow episodes that are too long and over too large a
9 spatial domain in those regions.

10

11 *3.1.3. Correlation between observed and simulated data*

12 In addition to the relative magnitude and duration of high and low flow events, a global
13 model should be able to accurately simulate their timing (onset and decay). Spearman's
14 rank (Spearman, 1944), which does not require time series to follow a normal
15 distribution, was used to examine correlations between observed and simulated RDI and
16 RFI. Figures 2 and 3 map the correlation coefficients between historical catalogues and
17 simulated time series respectively for high flows (RFI) and low flows (RDI) for each of
18 the 23 European study regions, alongside background information on station density used
19 to derive the historical catalogues (number of stations divided by the number of grid cells
20 within a given region) and region homogeneity (H, called RDI_{area} in Hannaford et al.,
21 2010). All plotted correlations are significant at $\alpha=0.05$.

22

1 Correlation coefficients are generally greater for low flows than for high flows across
2 most regions and models. This is likely to reflect the difference in temporal structure of
3 RDI and RFI, as droughts and low flow episodes tend to develop slowly and persist for
4 longer than floods and high flow events, which are more frequent but quicker to subside.
5 This large day-to-day variability of RFI compared to RDI will tend to result in smaller
6 correlation coefficients as a result of any errors in modelled timing of the event.

7

8 The value and range of correlations for high flow and low flow events are of similar
9 magnitude for all three models. Correlations show clear regional patterns, with all three
10 models systematically reproducing large-scale hydrological extreme time series in
11 different regions to varying extents. For example, the spatio-temporal variation of both
12 extremes (as measured by the correlation) in the SE Great Britain region is consistently
13 well reproduced by the three models. Conversely, simulated RDI and RFI of the
14 Pyrenees and French Southern Alps regions are consistently poorly correlated with the
15 observed series, despite RME of approximately 10% and -5% for RDI and RFI,
16 respectively, for the Pyrenees. This suggests that all models had difficulty in reproducing
17 the hydrology of mountainous regions (and in particular the occurrence and duration of
18 large-scale low/high flow events), perhaps linked to the spatial resolution of the models
19 and the driving climate data (0.5°), which is relatively coarse for adequately describing
20 the very variable topography of the regions or the weather in mountainous areas. In other
21 high mountainous regions, such as High Alps, poor correlation is found for the high flows
22 (between 0.39 for WaterGAP and 0.47 for JULES) while performance is slightly better
23 for low flows (between 0.51 for MPI-HM and 0.55 for WaterGAP). Note, however, that

1 the larger RME for RDI than RFI suggests a systematic bias in simulating dry episodes
2 by all models. In contrast, both regional index time series for NW Scandinavia are
3 relatively well reproduced (correlations between 0.59 and 0.61 for RDI and 0.51 and 0.65
4 for RFI), suggesting that the hydrological processes underlying snow-melt dominated
5 hydrology can be relatively well reproduced by all models (also reflected by a RME of
6 less than 5%). Haddeland et al. (accepted) suggest that differences in the modelled
7 partitioning of precipitation into rainfall and snowfall (which is generally calculated using
8 a temperature threshold close to 0°C) is responsible for the large spread in the estimated
9 snowfall, particularly in regions where winter temperatures are closer to 0°C, and by
10 extension, for the timing of the simulated runoff. MPI-HM and WaterGAP both partition
11 precipitation depending on temperature, while JULES takes snowfall and rainfall inputs
12 directly, so it is possible that all three models used rather different partitioning. This
13 could explain the large differences in RDI and RFI simulated by the three models, and the
14 relatively poor performance of WaterGAP, JULES and MPI-HM in reproducing large-
15 scale low/high flow events in the Pyrenees and Alps compared to better performance in
16 Scandinavia (which is much colder and hence all models tend to agree on snowfall).

17

18 With the exception of two British regions, large-scale high flow events tend to be better
19 reproduced in western and northern European regions, influenced by a maritime climate,
20 than in the more continental regions of central Europe. This could reflect the regional
21 differences in flood-generating mechanisms, mainly associated with frontal precipitation
22 (hence with relatively large spatial coherence, and relatively large persistence) in
23 maritime climates, and convective precipitation (hence driven by local, short-lived

1 intense events) in continental climates. The spatial scale of the modelling grid (0.5°), and
2 uncertainty in the measurement of highly-variable convective rainfall and its timing in the
3 forcing data, all tend limit the reproducibility of such local precipitation events, and
4 hence, of the resulting simulated high flows.

5

6 Inter-model differences in model performance at both low and high flows are not
7 substantial in terms of the correlation coefficients. For low flows, for which the driving
8 meteorological conditions are more similar in different climatic regions, model
9 performances are relatively similar. Overall, WaterGAP shows the largest correlations,
10 followed by MPI-HM and then JULES, where correlations lower than 0.55 are found for
11 12 regions (this is the case for only 4 and 6 regions for WaterGAP and MPI-HM,
12 respectively). For high flows, WaterGAP and JULES have similar correlation
13 coefficients, both outperforming MPI-HM, but there is no model which is systematically
14 associated with the largest correlation coefficients in all regions (5, 7 and 4 regions show
15 correlation coefficients greater or equal than 0.55 for WaterGAP, JULES and MPI-HM,
16 respectively). Regionally, JULES seems to perform best in France (all regions except
17 Pyrenees and Alps), Scandinavia and parts of Germany, while MPI-HM performs best in
18 Great Britain, Germany and Scandinavia, outperforming JULES in those last two areas.
19 There are no regions where WaterGAP systematically performs well or poorly for both
20 high and low flows.

21

1 *3.1.4. Influence of regional characteristics on the simulated regional indices*

2 Station density reflects how accurately the historical RDI and RFI catalogues capture the
3 spatio-temporal development of regional hydrological extremes: the more stations and the
4 better their distribution in the region, the more representative the regional indices. H is a
5 measure of the region homogeneity, i.e. whether the entire region experiences a flow
6 anomaly at the same time as historical measurements. There is no strong pattern linking
7 model performance (regardless of the goodness-of-fit measure considered) with station
8 density or region homogeneity (Figure 4). This suggests that differences in regional
9 characteristics are unlikely to significantly affect the model intercomparison, and that the
10 historical catalogues defined by Hannaford et al. (2010) and Parry et al. (2010) can be
11 used as benchmark time series against which to compare large-scale hydrological model
12 performance.

13

14 *3.2. Comparing drought and flood catalogues for observed and simulated data*

15 The three goodness-of-fit measures provide an overall summary of similarity between the
16 historical catalogues and the simulated RDI and RFI time series, but they do not provide
17 an adequate indication of the temporal and spatial variability in model performance. For
18 a hydrological model to show skill in its simulation of extreme events, three criteria must
19 be fulfilled:

- 20 • Accurate timing of occurrence/absence of extreme events; this is mainly influenced by
21 the meteorological drivers;

- 1 • Flashiness/smoothness of simulated time series (and in particular of extreme events)
2 similar to those of the observed catalogues; this reflects the rainfall-runoff transformation
3 mechanisms;
- 4 • Spatial consistency of extreme events of similar magnitude with that observed; this
5 shows that the spatial pattern of runoff is well reproduced by the models.

6

7 Figures 5 – 8 present RDI and RFI time series obtained from historical observations (left)
8 and simulations for the period 1963-2000 for six contrasting regions: SE Great Britain,
9 NW Spain, western and central France, High Alps, E Germany/ Czech Republic, and NW
10 Scandinavia.

11

12 European drought occurrence is generally fairly well reproduced by the considered global
13 models across the six regions, although with varying degrees of success (Figure 5 and 6).
14 In general, the timing of all major drought events is well reproduced, and periods with no
15 major large-scale drought are also represented successfully. However, skill in
16 reproducing drought duration or severity (in terms of regional spatial coherence of the
17 events, as quantified by the RDI) differs between the global models.

18

19 JULES tends to overestimate the duration (and occasionally the severity) of droughts;
20 excessive drought length is particularly notable for NW Spain and E. Germany/Czech
21 Republic. This would be consistent with over-estimation of evaporation but poor
22 parameter values or inadequate representation of processes could also contribute. JULES
23 has more success in reproducing the drought characteristics of regions with long periods

1 of observed drought due to higher catchment storage capacities, such as the groundwater-
2 dominated SE Great Britain, where RDI is reasonably well reproduced by JULES. This
3 could be because the 3m soil column and slowly responding, drainage-dominated runoff
4 of JULES are closer to reality in this area than elsewhere. JULES also shows some
5 ability in modelling the drought characteristics of snowmelt influenced regions such as
6 the High Alps, where the typical short duration of droughts is well reproduced, albeit
7 with the duration of major events being slightly too long and with the peak intensity too
8 high.

9

10 In contrast to JULES, MPI-HM tends to underestimate drought duration and the RDI
11 plots are 'noisier', with periods of deficiency often broken by very short periods of low
12 RDI. This could be explained by limited storage within the model, leading to runoff
13 being generated due to saturated soils following minor precipitation events. This is
14 consistent with the poor reproduction of drought for the SE Great Britain region, where
15 storage is primarily within groundwater aquifers, and where MPI-HM fails to reproduce
16 the sustained multi-season deficits that are often observed. In contrast, in NW
17 Scandinavia - where storage is in the form of snow pack and ice but the resulting RDI
18 time series is relatively flashy - MPI-HM reproduces drought characteristics reasonably
19 well. In other regions with low storage capacity and fast-responding catchments, MPI-
20 HM accurately reproduces longer periods of streamflow deficiency, most notably in the
21 western and central France region. In E Germany/Czech Republic and NW Spain, typical
22 drought duration is reasonably simulated, but the model generally overestimates the
23 number and severity of major droughts.

1
2 WaterGAP shows substantial and generally realistic variation in drought characteristics
3 across the six regions, suggesting that internal processes within the model are perhaps
4 most appropriate for reproducing drought characteristics. Where droughts are
5 predominantly sustained periods of low flows (e.g. in SE Great Britain) WaterGAP
6 simulates both duration and severity of such episodes well; where the drought regime is
7 dominated by short and frequent low flow periods, e.g. in the High Alps and E
8 Germany/Czech Republic, WaterGAP is also able to reproduce the drought
9 characteristics. Note, however, that for the 1976 episode, WaterGAP overestimates the
10 length of the drought in these regions (see next section for detailed analysis). Finally, as
11 with JULES, WaterGAP simulates long summer droughts in NW Spain which are not
12 present in the observed catalogue.

13
14 Investigating regionally important hydrological processes also yields some notable
15 observations of model performance. The observed RDI for the NW Scandinavia region
16 shows moderate streamflow deficiencies in the January - April period, which abruptly
17 terminate in the following months when snowmelt runoff dominates (Hannaford et al.
18 2010), followed by more sporadic deficiencies. All three models are able to reproduce
19 this process to some extent, with JULES and WaterGAP perhaps the most accurate.
20 Observed RDI for the January - March period for the High Alps shows distinct periods of
21 drought (in the 1960s and early 1970s) or a lack of droughts (through the 1970s and
22 1980s), which again are likely to reflect inter-annual differences in storage as ice and
23 snow. This is reasonably well reflected in all three models, but spring/summer droughts

1 in this region (which may be caused by the delayed onset of snowmelt) are simulated to
2 varying extents; JULES does not appear to capture the seasonality, whereas MPI-HM
3 captures a tendency for short, frequent deficiencies in summer. This is consistent with
4 the earlier description of JULES as having relatively large storage and slowly responding
5 runoff, whereas runoff in MPI-HM is much more responsive. Again we note that snow
6 simulation (and by extension, its subsequent melt) was identified by Haddeland et al.
7 (accepted) as one of the major differences between the models of WaterMIP.

8

9 The general patterns of model performance found for high flows are similar to those
10 described for low flows. Note it is more difficult to assess how well the models
11 reproduce the development of large-scale high flow events because such events are
12 usually short (i.e. with RFI time series very variable in time) and there are no sustained
13 periods of high flows across regions (Figures 7 and 8). This ‘flashiness’/noisiness in RFI
14 is generally reproduced by all the global models, but they all tend to simulate occurrence
15 of extremes even when no large-scale high flow events have been historically observed.
16 MPI-HM exhibits the most significant temporal variability in RFI, and to a lesser extent
17 WaterGAP. However, they both show generally good agreement for the most significant
18 high flow events, as well as a good degree of reproducibility of the distribution of ‘high
19 flow rich’ episodes in the period of record.

20

21 As is the case for low flows, the modelled high flow periods generated by JULES
22 typically overestimate duration and spatial coherence compared to the corresponding
23 observed RFI; this is particularly notable for the High Alps. Again, it is suggested that

1 this is a result of runoff in JULES being dominated by relatively slowly responding
2 subsurface flow. JULES performs well in other regions, such as E Germany/Czech
3 Republic (other than overestimation of high flows in 1987) and western and central
4 France, where long and coherent high flow episodes are observed. In NW Spain and SE
5 Great Britain, JULES also simulates the development of large-scale high flow episodes
6 reasonably well, particularly the damaging 2000/1 floods, but still tends to overestimate
7 high flow magnitude for most events. In Scandinavia, JULES is also able to mimic the
8 ‘flashy’ observed RFI, despite overestimating the persistence of the large-scale events.
9 However, JULES also simulates short-duration high flows in other regions (such as SE
10 Great Britain, NW Spain and western and central France), predominantly clustered in
11 June - October, even when no such temporal pattern has been observed nor simulated by
12 WaterGAP and MPI-HM in those regions. A thorough investigation of model
13 parameterisation and model structure would be necessary to understand the simulated
14 response of JULES, which is beyond the scope of this paper.

15

16 Akin to the patterns simulated at low flows, the simulated RFI generated from MPI-HM
17 is noisier than the observed RFI across all the regions presented in Figures 7 and 8 as it
18 generally simulates short events developing very rapidly in all regions. Whilst observed
19 periods of high flows are often reproduced as increased oscillations of the noisy signal
20 (e.g. in E Germany/Czech Republic in Dec 1974), for most regions high flows are almost
21 randomly distributed throughout the period of record, with few discernable periods of
22 sustained high flows, other than the periods of summer high flow in NW Spain which are
23 relatively accurately reproduced compared to other regions. For regions with flashy RFI

1 such as High Alps and NW Scandinavia, the spatio-temporal pattern simulated by MPI-
2 HM is in agreement with observations, but this is not the case for regions with prolonged
3 development of regional hydrological extremes (e.g. SE Great Britain). Further
4 investigation of model parameterisation and model structure, and in particular the
5 infiltration/saturation processes, would help to explain the systematic variability of
6 regional indices simulated by MPI-HM.

7

8 For WaterGAP, simulated RFI shows large seasonal variation not apparent in the
9 observed data. Occurrence of summer large-scale high flows, sustained over several
10 months, is usually correctly simulated for SE Great Britain, western and central France
11 and NW Spain, but simulated events in autumn, winter and early spring tend to be more
12 frequent and shorter than observed. The cause of this seasonal duality is unclear but is
13 unlikely to be due to the modelling of groundwater influence as the resulting smoothing
14 would not be restricted to summer months only, but occur in all seasons. This could be a
15 product of insufficient soil depth, so that only a ‘small’ part of the effective precipitation
16 can be stored in the soil. When the soil column is full of water, surface runoff is
17 generated, simulating ‘flashy’ events in winter, autumn and spring. Another factor might
18 be lower potential evapotranspiration estimated by WaterGAP than most other WaterMIP
19 models, which might tend to further increase surface runoff. In general, however,
20 WaterGAP is successful at reproducing the development of large-scale high flow events
21 in the six regions presented here, with temporal patterns of both flashy (High Alps, NW
22 Scandinavia) and more persistent (groundwater-dominated SE Great Britain) regimes
23 captured by the simulations.

1

2 As was witnessed in mountainous and snowmelt-influenced regions for modelled low
3 flow occurrence, the models have shown an ability to reproduce seasonal snowmelt-
4 driven streamflow responses and, in some years, the onset of high flow events in late
5 spring, a response to the introduction of melt water into river systems. These effects are
6 particularly evident at high flows in NW Scandinavia, with observed data showing a
7 relative scarcity of high flow events in the January – April period, particularly prior to
8 1988, which reflects the predominantly frozen nature of those catchments at that time of
9 year, with river flow recessions mitigating potential high flows.

10

11 *3.3. Simulation of notable high and low flow periods*

12 This section investigates two major episodes of large-scale hydrological extremes in
13 detail: the drought of 1975-76 (Zaidman et al., 2001) and the high flows of 2000 (Marsh
14 and Dale, 2002). The observed catalogues and simulated RDI and RFI time series are
15 presented for each model for the six contrasting regions, along with the correlation
16 coefficients corresponding to the two time periods (Figures 9 and 10, respectively). Two
17 questions are posed: (i) are the onset and termination of events reproduced at the correct
18 time?; and (ii) is the temporal development of a large-scale event (spatial coherence,
19 temporal variability) reproduced accurately?

20

21 *3.3.1. The drought of 1975-76*

22 Although comparatively short in duration, the 1975-76 drought was the most spatially
23 coherent and hydrologically intense drought on a European scale in the last 50 years.

1 Drought developed slowly from late summer/autumn, becoming coherent from winter
2 1975/76 (UK) to late spring (France and Germany), with the majority of European
3 regions exhibiting their most significant deficits in summer 1976. The drought terminated
4 in September in most regions, although it extended into the autumn for most southern
5 European regions (Parry et al., 2010).

6

7 Generally, WaterGAP captures the onset of the drought well, but not in all regions
8 (Figure 9). In contrast, MPI-HM simulates an early onset of the drought in SE Great
9 Britain, Germany, High Alps, and JULES simulates a delayed and prolonged onset in SE
10 Great Britain, western and central France and the High Alps. The end of the drought is
11 better captured by MPI-HM than by the other two models; simulated deficiencies are too
12 prolonged in WaterGAP and JULES, especially in NW Spain (minor drought), France
13 and East Germany.

14

15 In terms of drought development, MPI-HM reproduces temporal patterns of drought
16 development reasonably well in most regions despite generally exhibiting too much
17 variability (SE Great Britain in particular, but also western and central France, High Alps
18 and E Germany). WaterGAP tends to overestimate the magnitude of droughts (RDI too
19 high) whilst JULES tends to underestimate temporal variability while overestimating
20 RDI, producing very long, intense simulated droughts. Spain, where the 1975-76 drought
21 was the least intense, was accurately simulated by all models as the region with the
22 lowest drought intensity, but RDI was still largely overestimated by both WaterGAP and
23 JULES. In Scandinavia, where only minor drought conditions were observed, low RDI

1 was accurately simulated by all three models. In the High Alps, no model could
2 reproduce the spatio-temporal development of the drought satisfactorily, although the
3 shape of the peak of the drought in May and June 1976 is reasonably well captured by
4 WaterGAP and JULES.

5

6 *3.3.2. High flows in 2000*

7 During 2000, there were two major high flow events in Europe: in the spring, Germany
8 and central Europe experienced flooding, but the rest of Europe remained unaffected
9 (Figure 10). In the autumn, persistent and extensive precipitation lead to widespread
10 flooding in Great Britain (with total discharges for October-December the highest in a
11 record dating back to 1940), and to a certain extent in France. As previously mentioned,
12 the ability to reproduce the development of large-scale high flow episodes by models is
13 difficult to assess due to the high variability of the high flow episodes. Nevertheless,
14 Figure 10 suggests that both WaterGAP and MPI-HM reproduce the start of large-scale
15 floods (RFI) well, whilst they generally start later than observed in JULES (Spain, SE
16 Great Britain, High Alps, E Germany). All models struggle to simulate the termination
17 of a flood, but this is also difficult to identify from observed RFI catalogues.

18 In terms of development of the event, JULES shows the least similarity with observed
19 RFI. It simulates episodes which are slow-developing, spatially coherent (i.e. high RFI)
20 and prolonged, with a temporal variability much lower than that observed: once the
21 model simulates an extreme, the whole region rapidly experiences the same extreme
22 condition, with low spatial variation generating high intensity RFI for sustained periods.
23 WaterGAP and MPI-HM, which both tend to produce time series that are more variable

1 (i.e. less spatial coherence for flow anomalies, large day-to-day variability), simulate
2 more realistic events. But while the general shape of the time series is reproduced
3 reasonably well, the timing and magnitude of RFI peaks is not always synchronous with
4 that of the observed catalogues.

5

6 4. Discussion and conclusion

7 For the first time, global models are assessed for their ability to reproduce large-scale
8 hydrological extreme events using a new methodology to quantify the spatio-temporal
9 development of low and high flow events, as a proxy for hydrological extremes.

10 Historical European catalogues of droughts (RDI) and high flows (RFI) were derived
11 using river flow measurements across Europe. Using the same method, total runoff
12 simulated by large-scale hydrological models was used to generate simulated RDI and
13 RFI time series for the same European regions, which were directly comparable with the
14 observed European catalogues. Since it is based on anomalies, the method evaluates to
15 what extent the high and low flow events are simulated at the right time (onset and
16 cessation of events) with the same duration and with the same temporal variability and
17 spatial consistency as observed. This is a new way for modellers to evaluate performance
18 of their model, taking into account the intensity and spatial and temporal variability of
19 runoff in evaluating events at the large-scale. Note, however, that the analysis is not
20 designed to evaluate the skill of global models in reproducing the magnitude (in absolute
21 terms) of hydrological extremes.

22

1 The method was implemented for Europe using three global models with quite different
2 structures from the WaterMIP project: JULES, WaterGAP and MPI-HM. All were run
3 with the Watch Forcing Data at the same 0.5° grid resolution across Europe. Results
4 showed that all three models have broadly comparable performance in terms of goodness-
5 of-fit measures (RME, ratio between simulated and observed SD, and Spearman
6 correlation) between observed and simulated European RDI and RFI time series, with
7 WaterGAP performing best in terms of correlation for both high and low flows. Whilst
8 RDI magnitude and variability is systematically overestimated (positive RME and SD
9 ratio greater than 1), the shorter nature of high flow episodes is well captured by all
10 global models (small RME and relatively good reproduction of the variability). In
11 contrast, the timing of the high flow events and their relative magnitude is less well
12 simulated than that of low flow events, possibly owing to the shorter nature and generally
13 lower spatial coherence of floods, which are also often a more direct response to
14 meteorological conditions. Correlation coefficients greater than 0.5 were found for most
15 European regions for low flows and a majority were over 0.4 for high flows (all
16 significant at 95% level), with the worst-performing regions consistently showing weak
17 correlations across all models. The generally good associations for the time series are
18 likely to be influenced by the prevalence of periods with low or zero RDI (RFI), with
19 zeros constituting 90% of a homogenous time series by definition, but they also
20 demonstrate that the non-occurrence of extremes is generally simulated by all models
21 during the correct periods. Future work could address this issue by employing statistical
22 analyses which assess magnitude of correlation for only non-zero values (e.g. Fox-Maule
23 et al., submitted).

1

2 For low flows, all three global models generally capture the broad-scale characteristics of
3 droughts in six contrasting European regions (SE Great Britain, western and central
4 France, NW Spain, High Alps, E Germany/Czech Republic, NW Scandinavia) with
5 differing degrees of success. For high flows, there is some congruency between the
6 models and observations, but this is dependent on both the model and the region under
7 consideration. The differences in historical spatio-temporal development of floods and
8 droughts are consistently well captured by the three considered models. However, some
9 general weaknesses emerge, in particular for high flow periods.

10

11 Three factors can impact on regional differences between observed and simulated RDI
12 and RFI: (i) the station density influences the observed RDI and RFI to a degree, and low
13 station density could lead to an observed index not truly representative of the regional
14 hydrological extremes; (ii) the gridded precipitation input (WFD) might not accurately
15 reflect the precipitation events driving hydrological extremes. This would be the case
16 particularly for regions with large spatial variability of precipitation and weather, e.g.
17 convective storms in continental regions, snowmelt in mountainous regions; (iii) the
18 rainfall-runoff processes are complex and strongly influenced by the catchment
19 properties, and in particular by infiltration and groundwater processes which are not
20 necessarily described adequately in the models. While the first two factors are
21 independent of the global models and the resulting errors are shared by all simulated RDI
22 and RFI, the third factor will be reflected by the performance measured for the different

1 global models, as the simulated RDI and RFI time series will highlight those differences
2 in modelling the hydrological processes.

3 When the simulated RDI and RFI from all three models are compared, some clear
4 similarities and differences emerge. JULES tends to produce prolonged, highly spatially
5 coherent events, both for high and low flows. The events develop more slowly than the
6 historical catalogues have shown, and reach and sustain spatial coherence greater than
7 observed. This suggests a strong spatial coherence of the model parameterisation,
8 perhaps with insufficient variability in the processes in nearby grid cells. [Note, however,
9 that small regions such as the High Alps are described by very few grid cells and hence,
10 by construction, would inherit low spatial variability in the simulated RDI/RFI]. This is
11 consistent with the fact that, for the configuration of JULES used here, runoff is
12 dominated by subsurface drainage, which gives a relatively slow response. In contrast,
13 MPI-HM shows very high variability in the simulated RDI and RFI time series,
14 regardless of the region, with a more rapid onset of extreme events than observed. This is
15 particularly notable for regions with significant water storage capacity (e.g. SE Great
16 Britain). The attenuated hydrological regime, which leads to a prolonged development of
17 droughts, and to a lesser extent of high flow events, is not reproduced by MPI-HM, from
18 which simulations remain flashy. This would suggest that infiltration and groundwater
19 storage are perhaps under represented in MPI-HM, and saturation is reached quicker for
20 the grid cells than is observed. WaterGAP shares some of the issues relating to high
21 variability with MPI-HM, and some strong spatial coherence of simulated events with
22 JULES, but neither is dominant. Of the three considered global models, WaterGAP is

1 arguably the best suited to reproduce most regional characteristics of large-scale
2 hydrological extremes, but it remains deficient in adequately reproducing their onset and,
3 to a lesser extent, spatio-temporal variation.

4 It is encouraging to see the reasonable similarity between observed and simulated
5 European RDI, suggesting that simulated total runoff gives a credible indication of
6 regional scale drought response. Whilst differences between the various model outputs
7 and observed data have been highlighted and should be accounted for in interpreting the
8 results of any future runs, the results presented herein suggest the global models could be
9 expected to provide relatively realistic simulations of future drought characteristics, in
10 terms of seasonality, duration and magnitude in Europe. This is a useful finding, as it
11 complements the range of other studies which have shown the utility of large-scale
12 models in reproducing average runoff conditions (e.g. Haddeland et al. accepted; Lohman
13 et al., 2004) and suggests that low runoff extremes can also be modelled with some
14 degree of confidence. The regional high flows shown in the RFI catalogues suggest that
15 more caution would be required in interpreting the outputs of high runoff extremes from
16 large-scale models. Whilst the results suggest that the global models often capture the
17 overall characteristics of regional high flows in Europe, the model outputs fail to
18 realistically capture key characteristics of regionally coherent high flow periods for most
19 European regions.

20 This paper has demonstrated how two regional indices, the Regional Deficiency Index
21 (RDI) and the Regional Flood Index (RFI) can be used to assess how well large-scale

1 hydrological models can reproduce the spatio-temporal development of large-scale
2 hydrological extremes. This preliminary study has demonstrated the utility of observed
3 RDI and RFI time series (and corresponding drought and high flow catalogues) for
4 providing a ‘baseline’ against which to test large-scale models. It also showed how RDI
5 and RFI can help to highlight regional differences in hydrological regimes of global
6 models and to evaluate their ability to reproduce specific regimes, such as groundwater-
7 dominated or alpine hydrology, and both extremes of the hydrological regime in a
8 temperate climate such as Europe. However, it is important to keep in mind the
9 limitations of the RDI/RFI approach, as identified by Hannaford et al. (2010), particularly
10 in terms of regional homogeneity and the extent to which individual catchments are
11 representative of large regions. Future work will employ a more up-to-date dataset and
12 will also consider observed rainfall data.

13 This study has highlighted some key differences in how JULES, MPI-HM and
14 WaterGAP simulate the spatio-temporal development of hydrological extremes in
15 Europe, which are likely to reflect differences in model formulation and process
16 representation. Whilst it was outside the scope of this paper to entirely explain the
17 differences between models in terms of the different parameterisation schemes used and
18 to provide full physical interpretation of the results, this will be an important next step to
19 help identify geographical areas or processes where some model improvement might be
20 possible, and to test any new parameterisation. The methodology used here could also be
21 used to analyse other model variables, such as surface runoff or evapotranspiration,
22 although the lack of suitable observations for comparison would mean that such analyses

1 would be aimed at better understanding inter-model differences in spatial and temporal
2 variability. Whilst the framework developed here has proven successful and suggests that
3 there is scope for application in other regions of the world, results regarding model
4 performance only apply to Europe. Model behaviour and the ability to reproduce
5 hydrological processes may be very different in different climate regimes. In this vein,
6 Haddeland et al. (accepted) show that large differences in the schemes used for
7 estimating evapotranspiration can be important when explaining differences between
8 models, and between arid and wet regions. Furthermore, other indices might be better
9 suited than RDI and RFI in terms of their ability to capture certain aspects of complex
10 extremes, or to describe the spatial evolution of extreme events as suggested by Sheffield
11 et al. (2009) or Lloyd-Hughes (2011).

12

13 5. Acknowledgements

14 The authors would like to thank Dr. George Goodsell for coding the RDI and RFI
15 calculations, and three anonymous reviewers for their suggestions. This research was
16 undertaken as part of the European Union (FP6) funded Integrated Project called
17 WATCH (contract 036946).

18

19 6. References

20 Alcamo J., Döll P., Henrichs T., Kaspar F., Lehner B., Rösch T. and Siebert S., 2003. De-
21 velopment and Testing of the WaterGAP 2 Global Model of Water Use and
22 Availability, *Hydrological Sciences Journal* **48** (3): 317-337.

- 1 Best, M., Pryor, M., Clark, D., Rooney, G., Essery, R., M'énard, C., Edwards, J., Hendry,
2 M., Porson, A., Gedney, N., Mercado, L., Sitch, S., Blyth, E., Boucher, O., Cox, P.,
3 Grimmond, C., and Harding, R., submitted: The Joint UK Land Environment
4 Simulator (JULES), Model description, Part 1: Energy and water fluxes, *Submitted to*
5 *Geoscientific Model Development*.
- 6 BMI - Bundesministerium des Innern (German Federal Ministry of the Interior) 2002:
7 Bundesregierung zieht vorläufige Schadensbilanz der Hochwasserkatastrophe (Federal
8 Government draws preliminary balance of burdens during the flood catastrophe).
9 Pressemitteilung (Press release; 6.11.2002).
- 10 Burke, E. J., Brown, S. J., and Christidis, N.. 2006. Modelling the recent evolution of
11 drought and projections for the twenty-first century with the Hadley Centre climate
12 model, *J. Hydrometeorol.*, **7**, 1113–1125
- 13 Clark, D. B., Mercado, L. M., Sitch, S., Jones, C. D., Gedney, N., Best, M. J., Pryor, M.,
14 Rooney, G., Essery, R. L. H., Blyth, E., Boucher, O., Harding, R. J., and Cox, P. M.,
15 submitted: The Joint UK Land Environment Simulator (JULES), Model description,
16 Part 2: Carbon fluxes and vegetation, *Submitted to Geoscientific Model Development*
- 17 Cox, P.M., Betts, R.A., Bunton, C.B., Essery, R.L.H., Rowntree, P.R. and Smith, J.,
18 1999. The impact of new land surface physics on the GCM simulation of climate and
19 climate sensitivity. *Climate Dynamics*, **15**: 183-203
- 20 Dai, A., Qian, T., Trenberth, K. E., and Milliman, J. D., 2009: Changes in continental
21 freshwater discharge from 1948 to 2004, *J. Climate*, **22**, 2773–2791

- 1 Dankers, R., Feyen, L., 2008. Climate change impact on flood hazard in Europe: an
2 assessment based on high resolution climate simulations. *Journal of Geophysical*
3 *Research*, **113**, D19105.
- 4 Dankers, R., Feyen, L., 2009. Flood hazard in Europe in an ensemble of regional climate
5 scenarios. *Journal of Geophysical Research*, **114**, D16108
- 6 Döll P. and Siebert S., 2002. Global Modeling of Irrigation Water Requirements, *Water*
7 *Resources Research* **38** (4): 8.1-8.10.
- 8 Döll P., Kaspar F., Lehner B., 2003. A global hydrological model for deriving water
9 availability indicators: model tuning and validation. *J of Hydrology*, **270**, 105-134
- 10 Douglas, E.M., Vogel, R.M., Kroll, C.N., 2000. Trends in floods and low flows in the
11 United States: impact of spatial correlation. *J. Hydrology* **240**: 90-105
- 12 Environment Agency, 2010. *The costs of the summer 2007 floods in England and Wales*.
13 Project Report SC 070039/R1, Environment Agency, Bristol. 51pp.
- 14 Feyen, L. and Dankers, R., 2009. Impact of global warming on streamflow drought in
15 Europe, *J. of Geophysical Research*, **114**, D17116, doi: 10.1029/2008JD011438
- 16 Fink AH, Brücher T, Krüger A, Leckebush GC, Pinto JG, Ulbrich U., 2004. The 2003
17 European summer heatwaves and drought – synoptic diagnosis and impacts. *Weather*
18 **59**, 209 – 216.
- 19 Flörke M. and Alcamo J., 2004. *European Outlook on Water Use*, Final Report
20 EEA/RNC/03/007, Center for Environmental Systems Research - University of
21 Kassel.

- 1 Fox-Maule, C. Thejll, P., Christensen, J.H., Svendsen, S.H. and Hannaford, J, submitted.
2 Improved confidence in RCM simulations of precipitation from the ENSEMBLES
3 models using drought statistics. Submitted: *Journal of Climate*.
- 4 Gedney, N., Cox, P.M., Betts, R.A., Boucher, O., Huntingford, C. and Stott, P.A., 2006.
5 Detection of a direct carbon dioxide effect in continental river runoff records. *Nature*,
6 **439**, 835 - 837
- 7 Gudmundsson L, Tallaksen L. M. and Stahl K. (2010) Spatial cross-correlation patterns
8 of European low, mean and high flows. *Hydrological Processes*, doi:
9 10.1002/hyp.7807
- 10 Hamlet A. F., Mote P. W., Clark M. P., and Lettenmaier D. P., 2007: 20th Century
11 Trends in Runoff, Evapotranspiration, and Soil Moisture in the western US, *J.*
12 *Climate*, **20**, 1468–1486.
- 13 Haddeland, I., Clark, D. B., Franssen, W., Ludwig, F., Voß, F., Arnell, N. W., Bertrand,
14 N., Best, M., Folwell, S., Gerten, D., Gomes, S., Gosling, S. N., Hagemann, S.,
15 Hanasaki, N., Harding, R., Heinke, J., Kabat, P., Koirala, S., Oki, T., Polcher, J.,
16 Stacke, T., Viterbo, P., Weedon, G., and Yeh, P., accepted: Multi-model estimate of
17 the global water balance: setup and first results, *accepted by J. Hydrometeorol.*
- 18 Hagemann S. and Dümenil L., 1998. A parametrization of the lateral waterflow for the
19 global scale. *Climate dynamics*, **14**, 17-31
- 20 Hagemann S. and Dümenil Gates L.D., 2003. Improving a subgrid runoff
21 parameterization scheme for climate models by the use of high resolution data derived
22 from satellite observations. *Climate dynamics*, **21**, 349-359

- 1 Hannaford, J. and Marsh, T.J., 2008. high flow and flood trends in a network of
2 undisturbed catchments in the UK. *International J. Climatology*, **28**, 1325 – 1338.
- 3 Hannaford, J., Lloyd-Hughes, B., Keef, C., Parry, S., Prudhomme, C, 2010. Examining
4 the large-scale spatial coherence of European drought using regional indicators of
5 rainfall and streamflow deficit. *Hydrological Processes*.DOI:10.1002/hyp.7725
- 6 Huntington, T. G., 2006. Evidence for intensification of the global water cycle: Review
7 and synthesis. *Journal of Hydrology* **319**: 83 - 95.
- 8 Khaliq, M. N., Oaurda, T. B. M. J., Gachon, P., and Sushama, L., 2008: Temporal
9 evolution of low flow regimes in Canadian rivers, *Water Resour. Res.*, **44**, W08436,
10 doi:10.1029/2007WR006132,
- 11 Lehner, B., Doll P., Alcamo J., Henrichis T., Kaspar F., 2006. Estimating the impact of
12 global change on flood and drought risks in Europe: a continental, integrated analysis.
13 *J. of Climate*, **75**, 273-299
- 14 Lloyd-Hughes B, Prudhomme C, Hannaford J, Parry S, Keef C, Rees HG., 2010. *Drought*
15 *Catalogues for UK and Europe*. Environment Agency Science Report SC070079/SR
16 Environment Agency, Bristol.
- 17 Lloyd-Hughes B, 2011. A spatio-temporal structure-based approach to drought
18 characterisation. *International Journal of Climatology*: doi: 10.1002/joc.2280
- 19 Lohmann, D., Mitchell, K. E., Houser, P. R., Wood, E. F., Schaake, J. C., Robock, A.,
20 Cosgrove, B. A., Sheffield, J., Duan, Q., Luo, L., Higgins, R. W., Pinker, R. T., and
21 Tarpley, J. D., 2004. Streamflow and water balance intercomparisons of four land
22 surface models in the North American Land Data Assimilation System project, *J.*
23 *Geophys. Res.*, **109**, D07S91,30 doi:10.1029/2003JD003517.

- 1 Marsh, T.J. and Dale, M., 2002. The UK floods of 2000/01 – a hydrometeorological
2 appraisal. *CIWEM J.* **16(3)**, 180-188
- 3 Marsh, T.J. and Hannaford, J., 2008. *The 2007 Summer floods in England and Wales – a*
4 *hydrological appraisal*. Centre for Ecology and Hydrology, 32pp
- 5 Milly, P. C. D., Wetherald, R.T., Dunne, K.A. and Delworth, T.I., 2002. Increasing risk of
6 great floods in a changing climate. *Nature*, **415**, 514 - 517
- 7 Milly, P. C. D., Dunne, K. A., and Vecchia, A. V., 2005. Global pattern of trends in
8 streamflow and water availability in a changing climate, *Nature*, **438**, 347–350.
- 9 Parry, S., Prudhomme, C., Hannaford, J., and Lloyd-Hughes, B., 2010. High and low
10 flow catalogues for Europe: Regional indicators as tools to characterise spatially-
11 coherent hydrological extremes. *Global change: facing risks and threats to water*
12 *resources. Proceedings of the Sixth FRIEND World Conference, Fez (Morocco)*,
13 *IAHS*, **340**, 653-660.
- 14 Prudhomme C, Sauquet E., 2006. *Modelling a regional drought index in France*. CEH –
15 Cemagref Joint Report. Centre for Ecology and Hydrology, Wallingford.
- 16 Sheffield, J. and Wood, E.F., 2007. Characteristics of global and regional drought, 1950-
17 2000: analysis of soil moisture data from off-line simulation of the terrestrial
18 hydrologic cycle, *J. of Geophysical Research*, **112**, D17115, doi:
19 10.1029/2006JD008288
- 20 Sheffield, J. and Wood, E. F., 2008a: Projected changes in drought occurrence under
21 future global warming from multi-model, multi-scenario, IPCC AR4 simulations,
22 *Clim. Dynam.*, **13**, 79–105.

- 1 Sheffield, J. and Wood, E.F., 2008b: Global trends and variability in soil moisture and
2 drought characteristics, 1950–2000, from observation-driven simulations of the
3 terrestrial hydrologic cycle, *J. Clim.*, **21**, 432–458.
- 4 Sheffield, J., Andreadis, K.M., Wood, E.F., Lettenmaier, D.P., 2009. Global and
5 continental drought in the second half of the twentieth century: severity-area-duration
6 analysis and temporal variability of large-scale events. *J. of Climate*, **22**, 1962-1981
- 7 Spearman, C. E., 1944. "General Intelligence" objectively determined and measured.
8 *American Journal of Psychology*, **15**, 201-293.
- 9 Stahl K., 2001. *Hydrological drought - a study across Europe*. Geowissenschaftlichen
10 Fakultät. Freiburg: Albert-Ludwigs Universität Freiburg.
- 11 Stahl K, Demuth S., 1999. Linking streamflow drought to the occurrence of atmospheric
12 circulation patterns. *Hydrological Sciences Journal* **44**, 467 – 482
- 13 Stahl, K., Hisdal, H., Hannaford, J., Tallaksen, L. M., van Lanen, H. A. J., Sauquet, E.,
14 Demuth, S., Fendekova, M., and Jódar, J., 2010: Streamflow trends in Europe:
15 evidence from a dataset of near-natural catchments, *Hydrol. Earth Syst. Sci. Discuss.*,
16 **7**, 5769-5804, doi:10.5194/hessd-7-5769-2010
- 17 Stott, P.A., Gillet, N.P., Hegerl, G.C., Karoly, D.J., Stone, D.A., Zhang. X. and Zwiers,
18 F., 2010. Detection and attribution of climate change: a regional perspective. *Wiley*
19 *Interdisciplinary Review, Climate Change*. **1**, 192 – 211.
- 20 Weedon, G.P., Gomes, S., Viterbo, P., Osterle, H., Adam, J.C., Bellouin, N., Boucher, O.,
21 Best, M., 2010. *The WATCH Forcing Data 1958-2001: a meteorological forcing*
22 *dataset for land-surface and hydrological models*. WATCH technical report 22, 41p.

- 1 Wilby, R.L., Beven, K.J. and Reynard, N.S., 2008. Climate change and fluvial flood risk
- 2 in the UK: more of the same? *Hydrological Processes*, **22**, 2511 – 2523
- 3 Zaidman MD, Rees HG, Young AR., 2001. Spatio-temporal development of streamflow
- 4 droughts in north-west Europe. *Hydrology and Earth System Sciences* **5**, 733 - 751
- 5
- 6
- 7

1 **Table captions**

2

3 Table 1: Main characteristics of models considered in the study. After
4 Haddeland et al. (accepted)

5

6 Table 2: Goodness-of-fit assessments for the RFI and RDI series and two
7 periods of low flow (1976) and high flow (2000) for six European regions

8

9

1

2 Table 1

Model name ¹	Model time step	Meteorological forcing variables ²	Energy balance	Evapotranspiration scheme ³	Runoff scheme ⁴	Snow scheme	Reference(s)
JULES	1 h	R, S, T, W, Q, LW, SW, SP	Yes	Penman-Monteith	Infiltration excess/ Darcy	Energy balance	Cox et al. (1999)
MPI-HM	Daily	P, T	No	Thorntwaite	Saturation excess/ Beta function	Degree day	Hagemann and Gates (2003), Hagemann and Dümenil (1998)
WaterGAP	Daily	P, T, LWn, SW	No	Priestley-Taylor	Beta function	Degree day	Alcamo et al. (2003)

3

4

5

6

7

8

9

10

11

Table 2

		RFI						RDI					
Region		SE Great Britain	NW Spain	W & C France	High Alps	Germany & Czech	Scandinavia via	SE Great Britain	NW Spain	W & C France	High Alps	Germany & Czech	Scandinavia via
	NS	37	14	48	35	8	37	37	14	48	35	8	37
	H	8.30	10.70	7.40	8.80	8.20	10.70	8.30	10.70	7.40	8.80	8.20	10.70
WaterGAP	RME	-4.30	-8.16	-11.67	-1.65	-4.69	-1.48	8.47	14.78	13.50	22.06	16.73	4.64
	Spread	1.25	1.11	0.91	1.03	0.92	0.94	1.23	1.28	1.07	1.17	1.09	1.04
	Rho	0.61	0.54	0.52	0.39	0.47	0.62	0.77	0.50	0.61	0.55	0.59	0.62
	Rho 76	0.69	0.52	0.49	0.39	0.30	0.62	0.91	0.46	0.82	0.55	0.68	0.78
	Rho 00	0.73	0.67	0.50	0.39	0.41	0.55	0.12	0.45	0.01	0.27	0.10	0.79
JULES	RME	-7.69	-9.28	-12.50	-2.59	-3.72	-2.71	8.18	13.81	10.99	21.13	16.63	4.90
	Spread	1.16	1.02	0.89	1.07	0.91	0.93	1.24	1.18	1.04	1.38	1.09	1.00
	Rho	0.59	0.49	0.55	0.47	0.48	0.65	0.77	0.48	0.59	0.54	0.53	0.61
	Rho 76	0.65	0.23	0.41	0.34	0.30	0.67	0.91	0.43	0.77	0.50	0.64	0.69
	Rho 00	0.83	0.60	0.45	0.25	0.67	0.67	NA*	0.24	0.12	0.45	0.44	0.84
MPI-HM	RME	-1.99	-13.28	-9.29	-1.07	-4.89	0.01	6.65	18.57	3.73	10.08	13.94	1.90
	Spread	1.33	1.09	0.95	1.15	0.96	1.06	1.22	1.42	1.03	1.33	1.21	1.02
	Rho	0.56	0.48	0.50	0.41	0.40	0.51	0.63	0.60	0.67	0.51	0.53	0.59
	Rho 76	0.59	0.50	0.45	0.34	0.30	0.52	0.74	0.56	0.78	0.50	0.62	0.70
	Rho 00	0.74	0.60	0.54	0.35	0.46	0.48	0.67	0.65	0.35	0.48	0.57	0.78

12

13

14

15

16

17

* RDI is always equal to 0 in 2000 in SE Great Britain as modelled by JULES. NS: Number of Stations. H: Homogeneity index. RME: Relative Mean Error; Spread: ratio of Standard Deviation of simulated series over Standard Deviation of Observed series; Rho: Spearman correlation. Bold values are significant at 0.95% level (only for Rho). 76: estimate for 75 to 77; 00: estimate for 1 Jan 2000 to 31 Dec 2000

1 **Figure captions**

2 Figure 1. Distribution of gauging stations over Europe and hydrological regions used in
3 this study.

4 Figure 2. Correlation between observed RDI and RDI simulated by JULES, MPI-HM and
5 WaterGAP. Regions station density (shades) and region homogeneity measure H
6 (numbers) are shown on the top-left graph.

7 Figure3. Same as Figure 2 but for RFI

8 Figure 4. Scatter plots of region homogeneity (black) or station density (grey) against
9 RME (top), spread reproduction (middle) and spearman correlation (bottom) for RDI
10 (left) and RFI (right) for WaterGAP (circle) JULES (square) and MPI-HM (diamond)

11 Figure 5. Observed RDI catalogues (OBS) and RDI simulated by JULES, MPI_HM and
12 WaterGAP (WGAP) for the regions of South East Great Britain, North West Spain and
13 western and central France

14 Figure 6. Same as Figure 5 but for the regions of High Alps, East Germany and Czech
15 Republic and North West Scandinavia

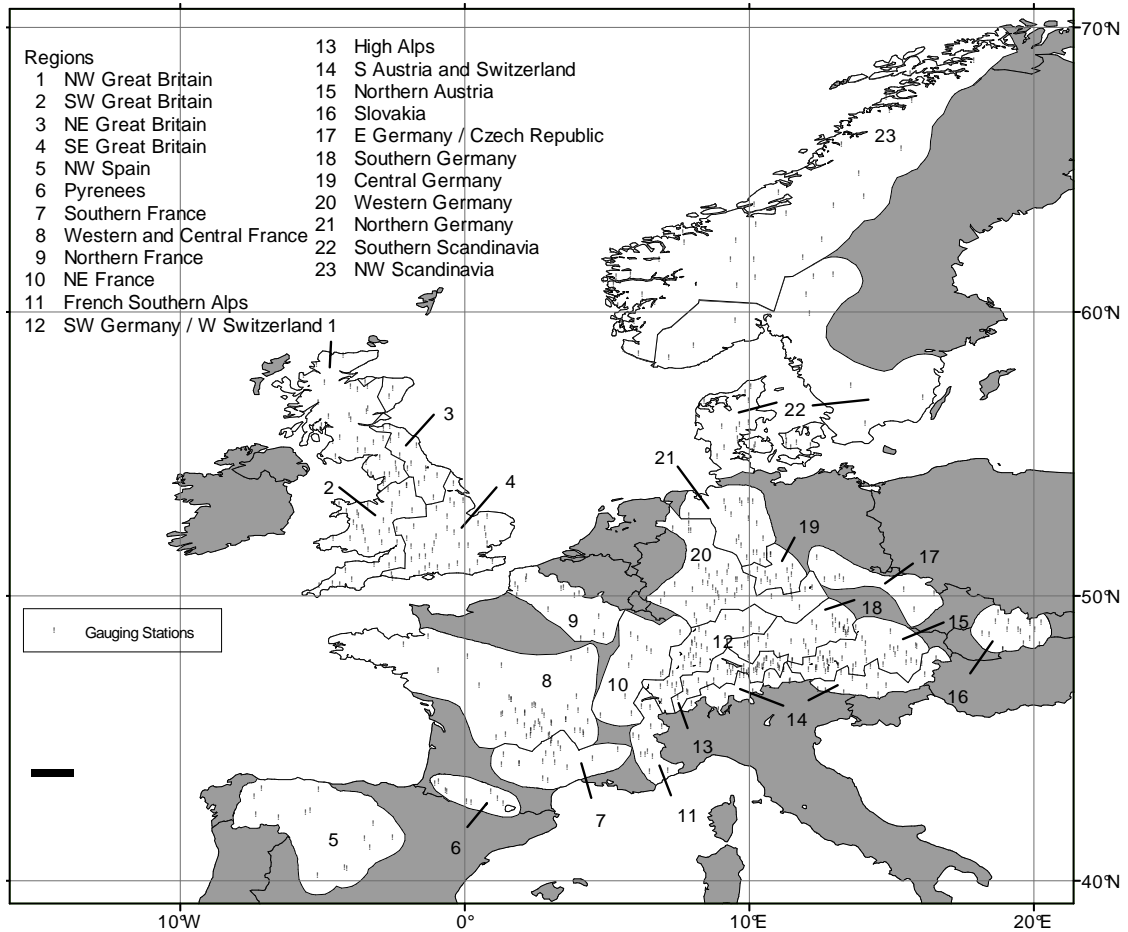
16 Figure 7: As Figure 5 but for RFI

17 Figure 8: As Figure 6 but for RFI

18 Figure 9: RDI for the 1975-76 drought for six contrasting regions, from top to bottom:
19 South East Great Britain, North West Spain, western and central France, High Alps, East
20 Germany and Czech Republic, North West Scandinavia. For each region, 10-daily RDI
21 are showed in grey (observed) and black (simulated) for WaterGAP (top row), JULES
22 (middle row) and MPI-HM (bottom). Corresponding Spearman correlation on the right
23 (significant at $\alpha = 0.05$)

24 Figure 10: As for Figure 9 but for RFI for the year 2000

1 Figure 1. Distribution of gauging stations over Europe and hydrological regions used in
 2 this study.

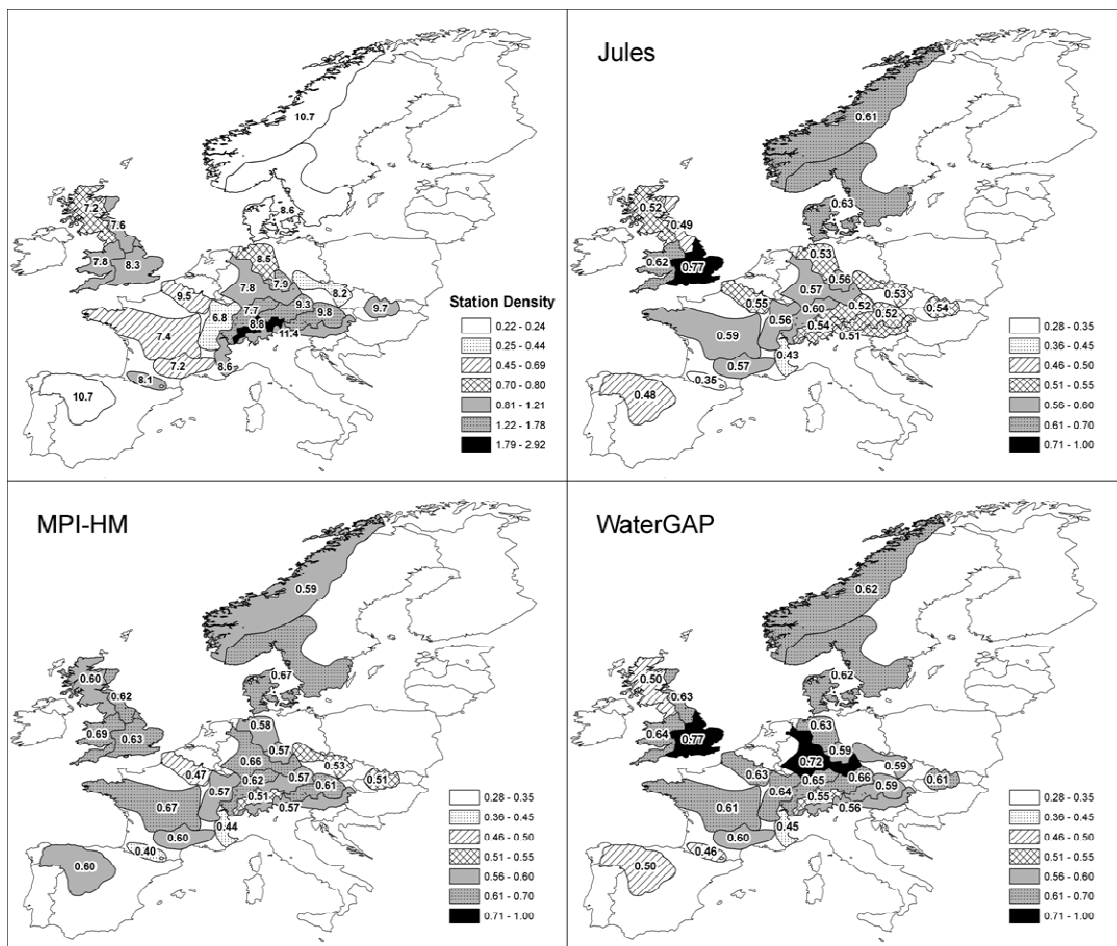


3

4

5

- 1 Figure 2. Correlation between observed RDI and RDI simulated by JULES, MPI-HM and
- 2 WaterGAP. Regions station density (shades) and region homogeneity measure H
- 3 (numbers) are shown on the top-left graph.

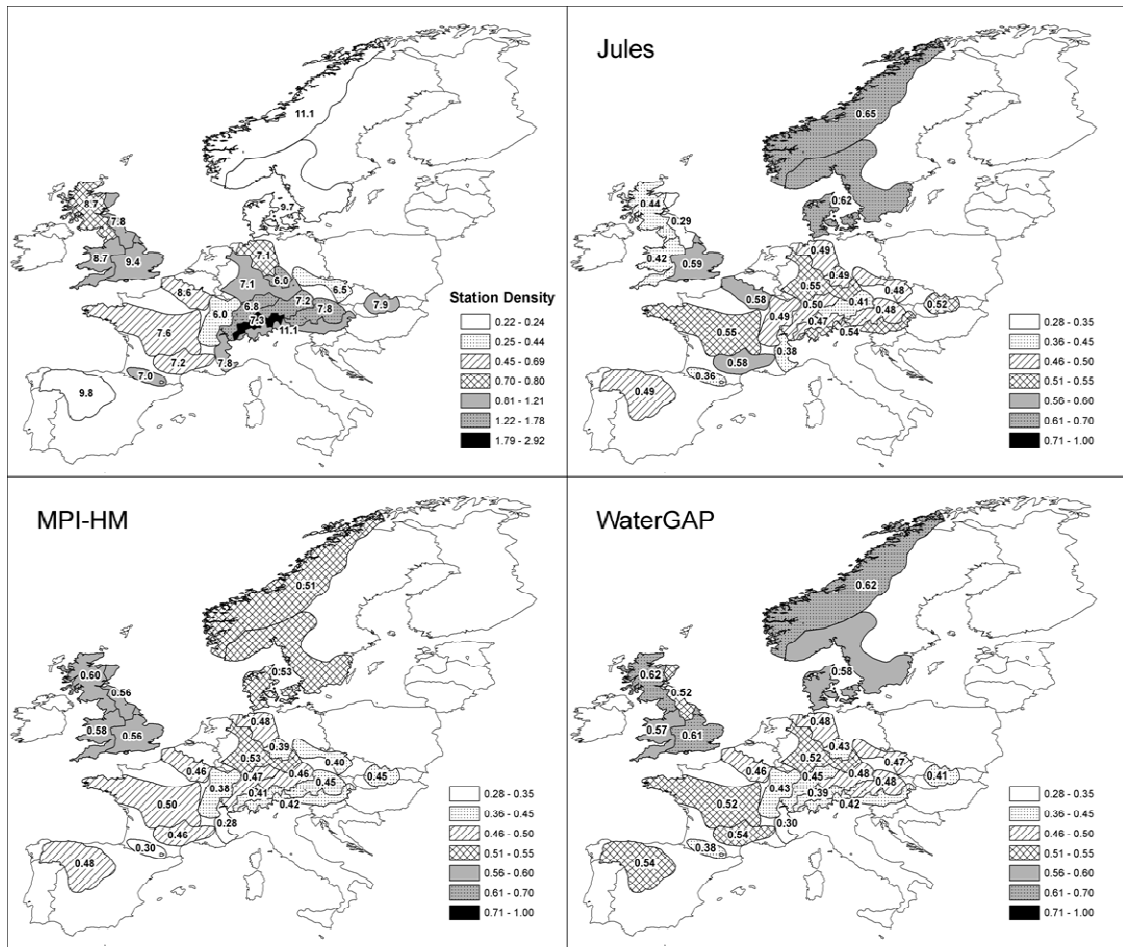


4

5

6

1 Figure3. Same as Figure 2 but for RFI

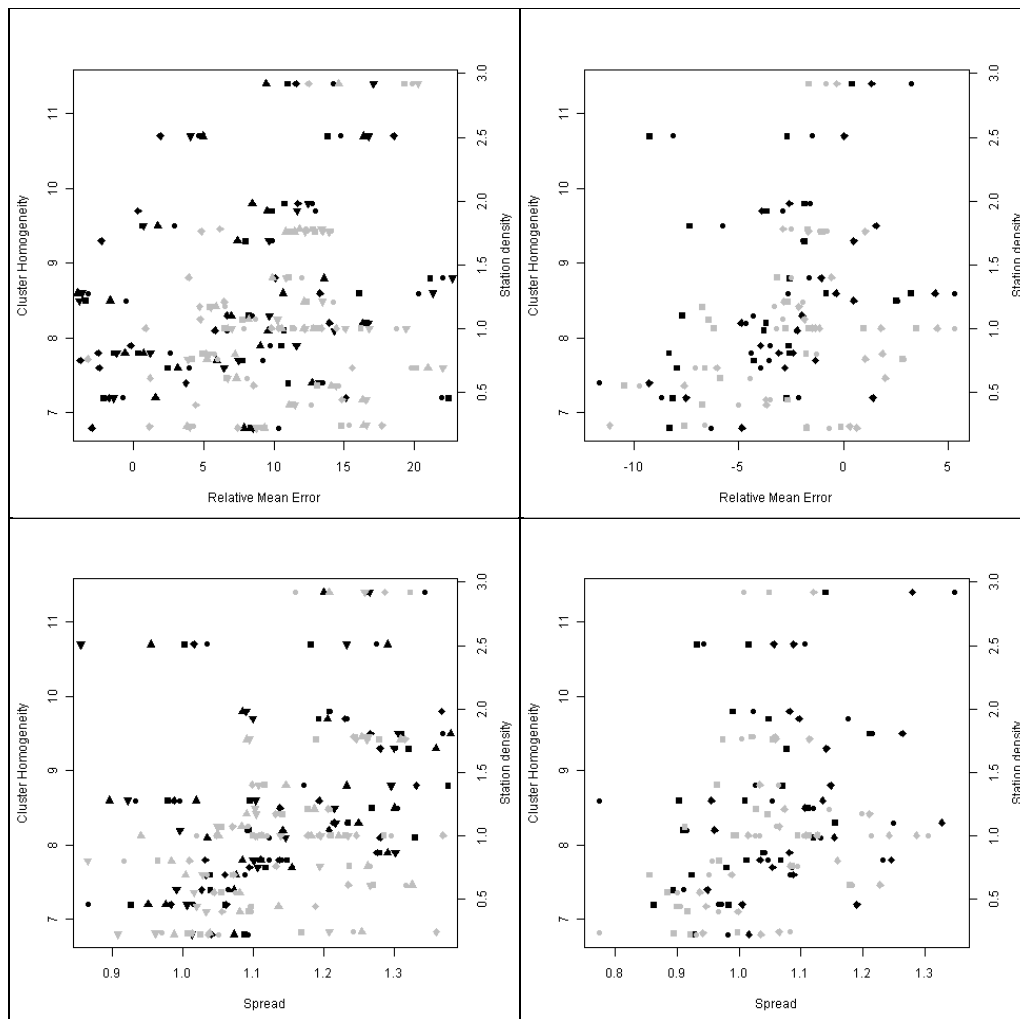


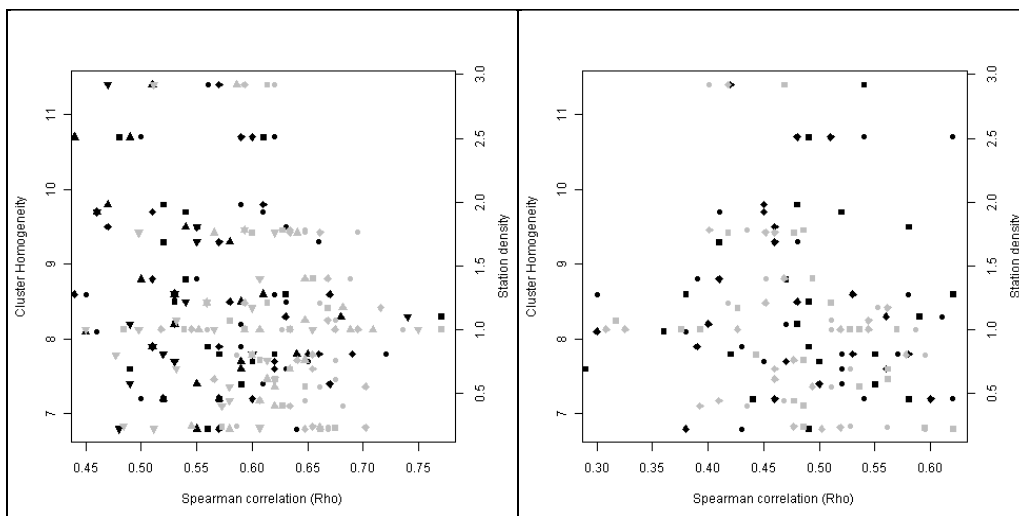
2

3

4

- 1 Figure 4. Scatter plots of region homogeneity (black) or station density (grey) against
- 2 RME (top), spread reproduction (middle) and spearman correlation (bottom) for RDI
- 3 (left) and RFI (right) for WaterGAP (circle) JULES (square) and MPI-HM (diamond)



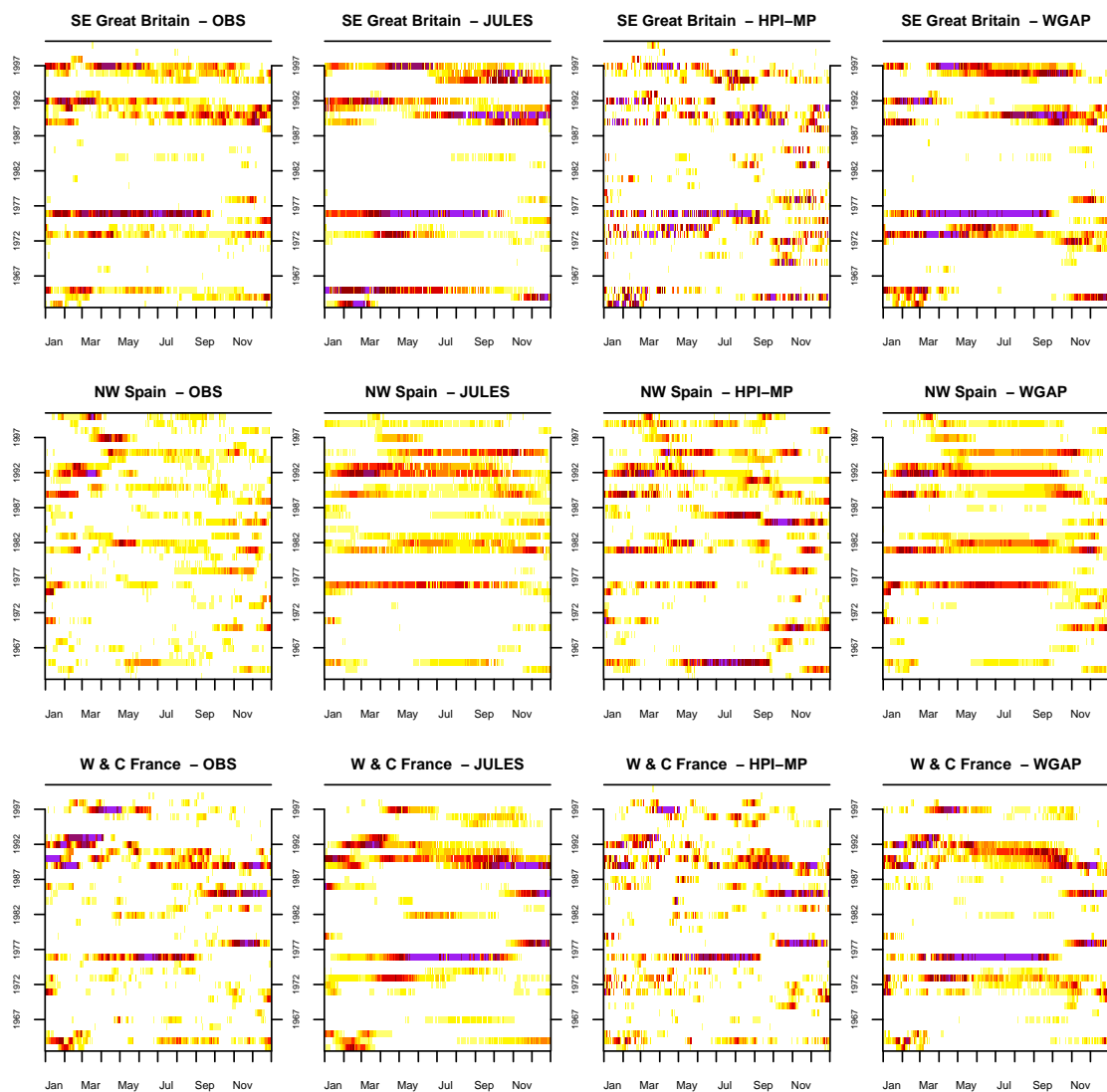


1

2

3

- 1 Figure 5. Observed RDI catalogues (OBS) and RDI simulated by JULES, MPI_HM and
- 2 WaterGAP (WGAP) for the regions of South East Great Britain, North West Spain and
- 3 western and central France

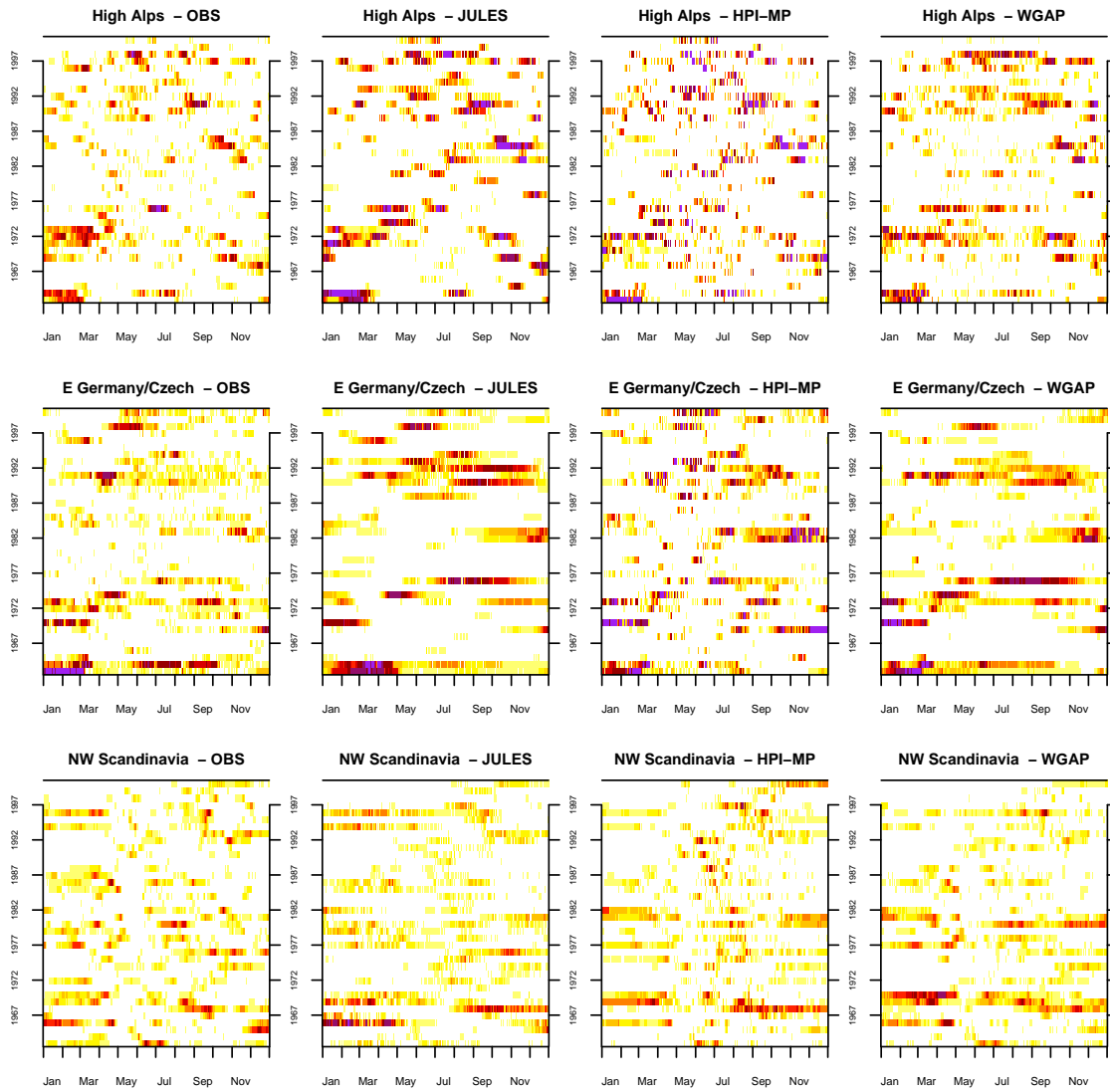


4

5

6

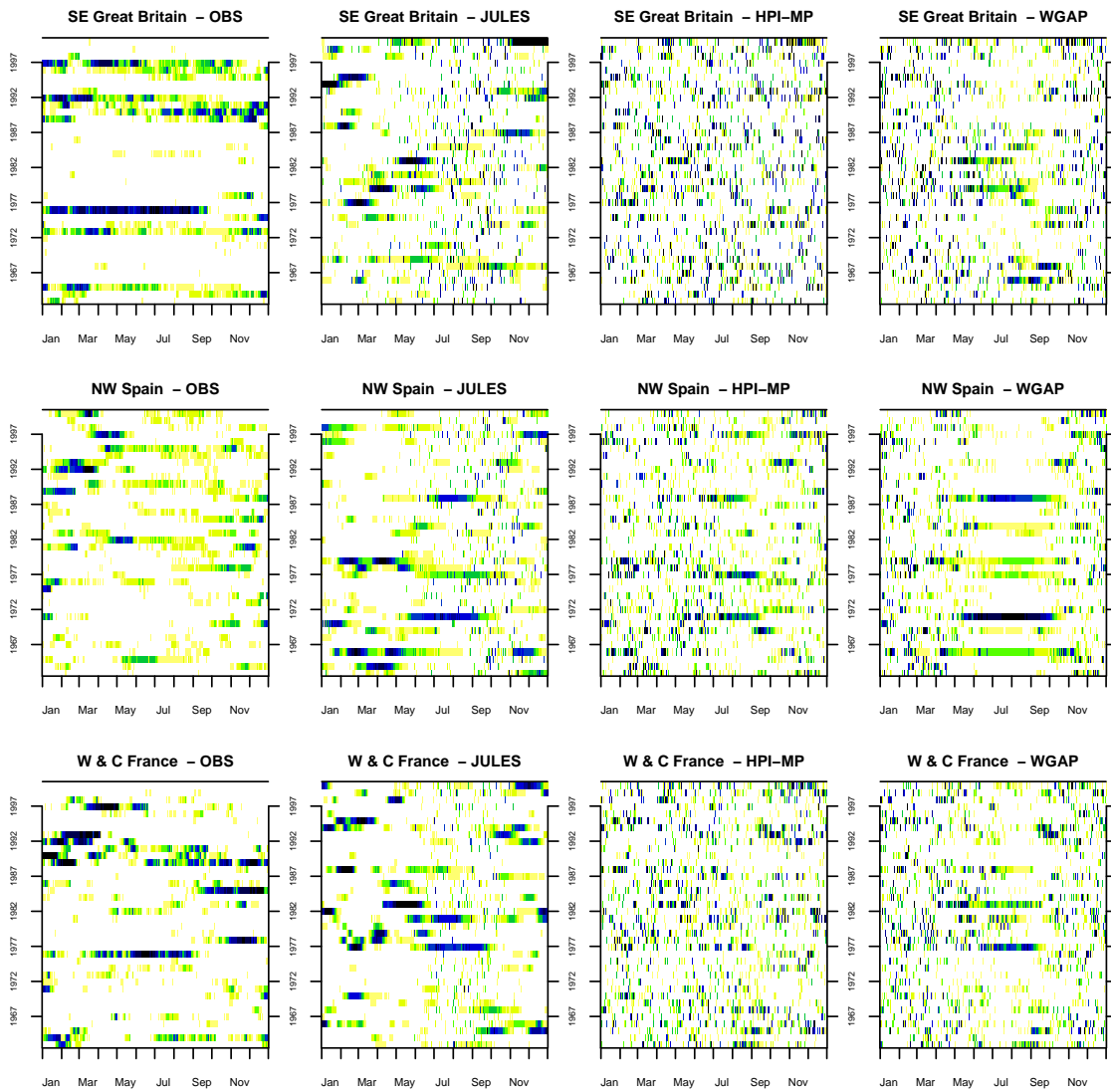
- 1 Figure 6. Same as Figure 5 but for the regions of High Alps, East Germany and Czech
- 2 Republic and North West Scandinavia



3

4

1 Figure 7: As Figure 5 but for RFI

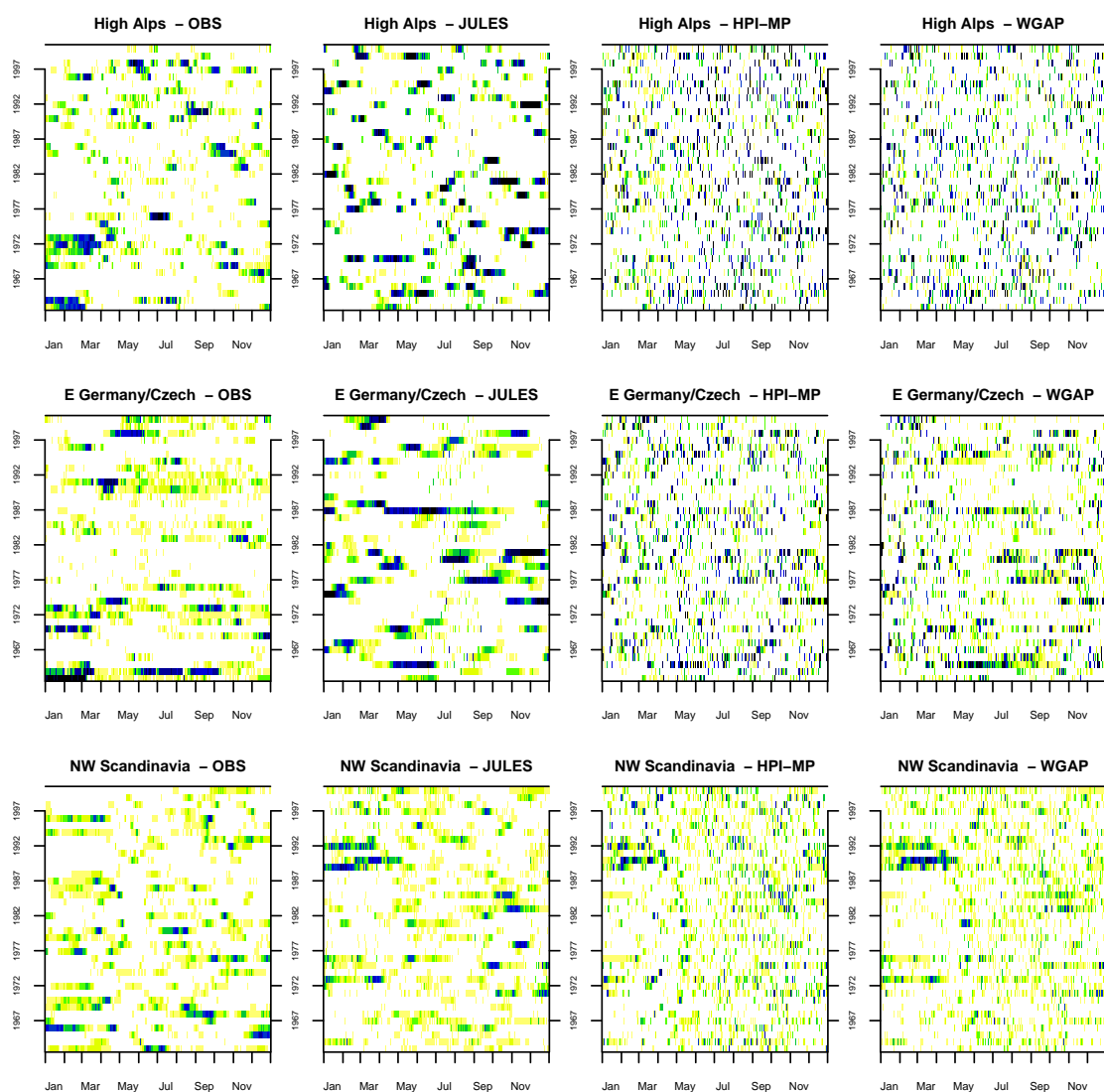


2

3

4

1 Figure 8: As Figure 6 but for RFI



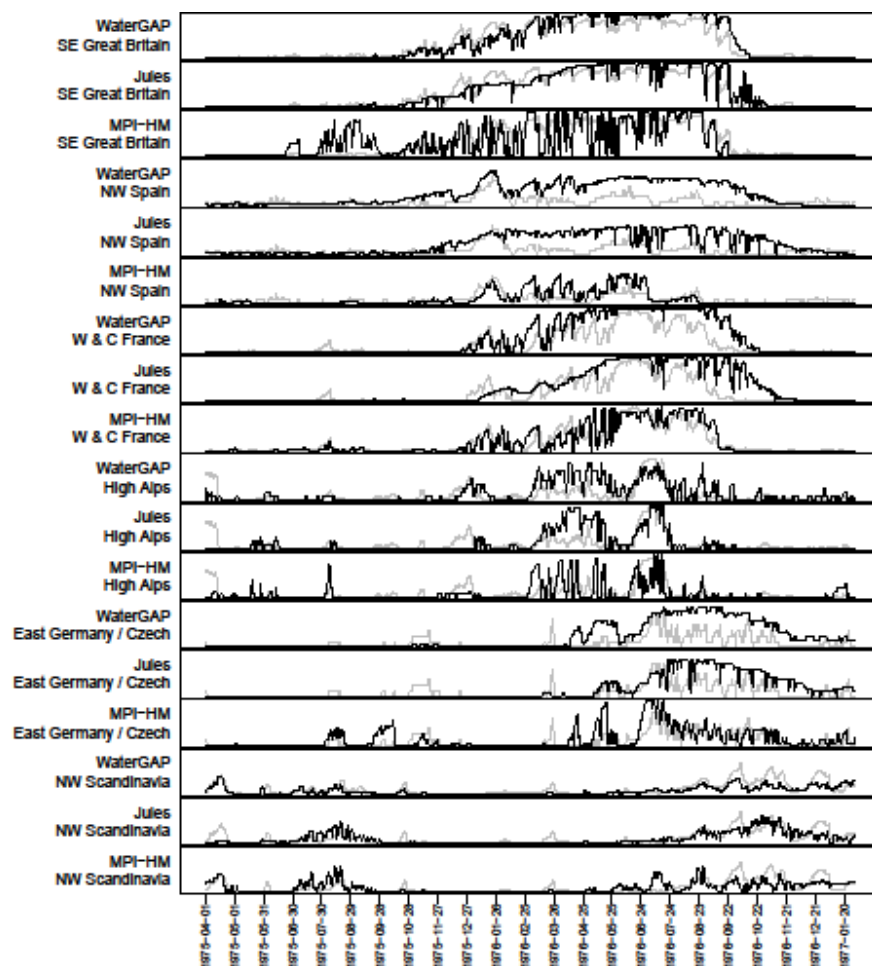
2

3

4

1 Figure 9: RDI for the 1975-76 drought for six contrasting regions, from top to bottom:
 2 South East Great Britain, North West Spain, western and central France, High Alps, East
 3 Germany and Czech Republic, North West Scandinavia. For each region, 10-daily RDI
 4 are showed in grey (observed) and black (simulated) for WaterGAP (top row), JULES
 5 (middle row) and MPI-HM (bottom). Corresponding Spearman correlation on the right
 6 (significant at $\alpha = 0.05$)

7

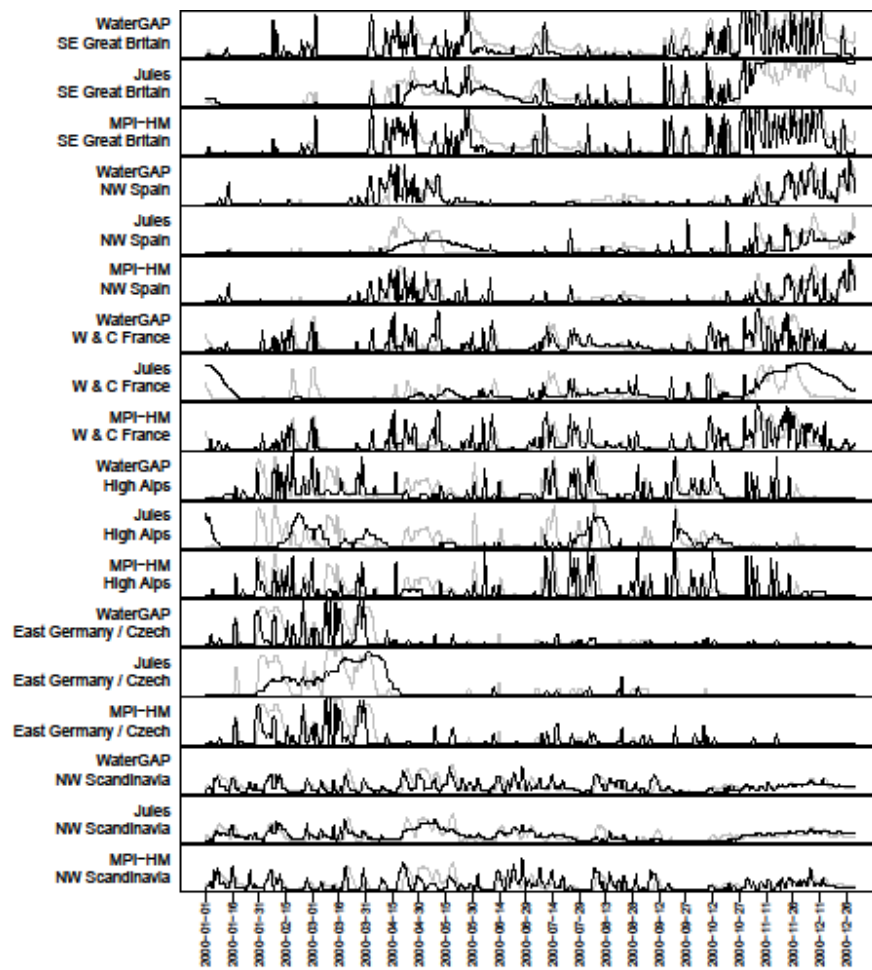


8

9

10

1 Figure 10: As for Figure 9 but for RFI for the year 2000

2
3