Science of the Total Environment 615 (2018) 1028–1047

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

Uncertainty of modelled flow regime for flow-ecological assessment in Southern Europe

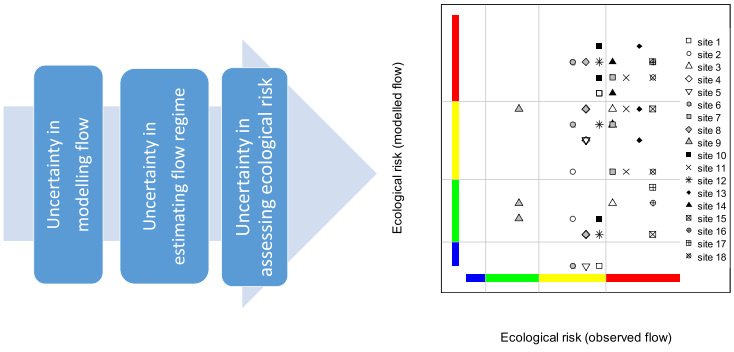
Olga Vigiak [a,b,*], Stefanie Lutz [c], Angeliki Mentzafou [d], Gabriele Chiogna [e,f], Ye Tuo [e], Bruno Majone [g], Hylke Beck [a], Ad de Roo [a], Anna Malagó [a], Fayçal Bouraoui [a], Rohini Kumar [h], Luis Samaniego [h], Ralf Merz [c], Christos Gamvroudis [i], Nikolaos Skoulikidis [d], Nikolaos P. Nikolaidis [i], Alberto Bellin [g], Vicenç Acuña [j], Nataša Mori [k], Ralf Ludwig [b], Alberto Pistocchi [a]

[a] European Commission, Joint Research Centre (JRC), Directorate D – Sustainable Resources, Ispra, Italy
[b] Ludwig-Maximilians-Universitaet Muenchen, Department of Geography, Munich, Germany
[c] UFZ-Helmholtz Centre for Environmental Research, Department Catchment Hydrology, Theodor-Lieser-Str. 4, 06120 Halle (Saale), Germany
[d] Hellenic Centre for Marine Research, Institute of Marine Biological Resources and Inland Waters, Anavyssos Attica, Greece
[e] Technical University of Munich, Chair of Hydrology and River Basin Management, Munich, Germany
[f] University of Innsbruck, Institute of Geography, Innsbruck, Austria
[g] University of Trento, Department of Civil, Environmental and Mechanical Engineering, Trento, Italy
[h] UFZ-Helmholtz Centre for Environmental Research, Department of Computational Hydrosystems, Permoserstraße 15, 04318 Leipzig, Germany
[i] School of Environmental Engineering, Technical University of Crete, Chania, Greece
[j] Catalan Institute for Water Research (ICRA), Girona, Spain
[k] National Institute of Biology, Department of Organisms and Ecosystems Research, Ljubljana, Slovenia

HIGHLIGHTS

- We assessed uncertainty in modelling flow regime in four basins with six models.
- Except for medium to high flow magnitude, indicator uncertainty was high.
- The lack of a reliable reference flow regime hampers flow-ecological assessments.
- Use of indicators in ecological assessments should account for their uncertainty.
- Simulation of anthropogenic alterations in hydrological models should be improved.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 24 August 2017
Received in revised form 26 September 2017
Accepted 27 September 2017
Available online 10 October 2017

Keywords:
Flow regime
Ecohydrological assessment

ABSTRACT

Sustainable water basin management requires characterization of flow regime in river networks impacted by anthropogenic pressures. Flow regime in ungauged catchments under current, future, or natural conditions can be assessed with hydrological models. Developing hydrological models is, however, resource demanding such that decision makers might revert to models that have been developed for other purposes and are made available to them ('off-the-shelf' models). In this study, the impact of epistemic uncertainty of flow regime indicators on flow-ecological assessment was assessed at selected stations with drainage areas ranging from about 400 to almost 90,000 km^2 in four South European basins (Adige, Ebro, Evrotas and Sava). For each basin, at least two models were employed. Models differed in structure, data input, spatio-temporal resolution, and calibration strategy, reflecting the variety of conditions and purposes for which they were initially developed. The

* Corresponding author at: European Commission, Joint Research Centre (JRC), Directorate D – Sustainable Resources, Ispra, Italy.
E-mail addresses: olga.vigiak@gmail.com, olga.vigiak@ext.ec.europa.eu (O. Vigiak).

https://doi.org/10.1016/j.scitotenv.2017.09.295

SWAT
mHM
Lisflood
HYPERstream

uncertainty of modelled flow regime was assessed by comparing the modelled hydrologic indicators of magnitude, timing, duration, frequency and rate of change to those obtained from observed flow. The results showed that modelled flow magnitude indicators at medium and high flows were generally reliable, whereas indicators for flow timing, duration, and rate of change were affected by large uncertainties, with correlation coefficients mostly below 0.50. These findings mirror uncertainty in flow regime indicators assessed with other methods, including from measured streamflow. The large indicator uncertainty may significantly affect assessment of ecological status in freshwater systems, particularly in ungauged catchments. Finally, flow-ecological assessments proved very sensitive to reference flow regime (i.e., without anthropogenic pressures). Model simulations could not adequately capture flow regime in the reference sites comprised in this study. The lack of reliable reference conditions may seriously hamper flow-ecological assessments. This study shows the pressing need for improving assessment of natural flow regime at pan-European scale.

1. Introduction

The ecological status of water bodies can be assessed on the basis of flow regime, morphology, water quality, and biological elements (EC, 2000; Pistocchi et al., 2016). Flow-ecological frameworks developed to provide guidelines for sustainable water basin management are typically based on characterization of flow regime and its alterations (Poff et al., 2010; Laizé et al., 2014). Under these frameworks, the assessment of ecological status is inferred by the deviation of flow regime from what are indicated as 'natural flow' conditions. The rationale is that an aquatic ecosystem that has developed in a specific flow regime would be subject to increasing stress when the flow regime is altered by human or other pressures (e.g. Poff et al., 1997; Poff et al., 2010). Five main characteristics of flow regime are to be considered: magnitude, timing/seasonality, frequency, duration, and rate of change (Richter et al., 1997; Poff et al., 2010; Kumar et al., 2010; Archfield et al., 2014; Laizé et al., 2014). Almost two hundred Indicators of Hydrologic Alteration (IHAs), each assessing one or more flow regime characteristics, have been used in flow-ecological assessments. While any IHAs can be important locally depending on the environmental pressure and the biota responses, some IHAs have been considered consistently in several studies (Murphy et al., 2013; Archfield et al., 2014; Laizé et al., 2014).

Given their pivotal role in the eco-hydrological frameworks, research started questioning the impact of the estimation methods on IHAs. For example, observational data uncertainty, stemming from measurement methods, representativeness and data management, may propagate to hydrological indicators (McMillan et al., 2012). Ultimately, indicators uncertainty varies largely, depending on the reach hydrological regime (e.g. flow variability), gauging site conditions (e.g. affecting the stage-discharge rating curve), and the indicator definition (Westerberg and McMillan, 2015; Westerberg et al., 2016). Uncertainty of mean annual flow has been estimated at about ± 10–15%, but it may exceed ± 20% at low or high flow percentiles. Uncertainty for indicators of frequency and duration of high and low flow is even higher, in the range of ± 30–40%, especially if the indicators are defined in relation to a threshold (e.g. as a multiple of mean or median flow or flows crossing specified quantile levels; Kumar et al., 2010; Westerberg and McMillan, 2015; Westerberg et al., 2016). Kennard et al. (2010) analyzed the impact of the period length on the accuracy of 120 IHAs and found that indicator accuracy quickly improved when the period of analysis increases from one to 15 years, but after that indicators tended to stabilize, and did not change substantially for periods longer than 30 years. IHAs that were most sensitive to the length of the analysis period were those linked to rare events, e.g. skewness in annual and maximum flow, low flow magnitude, duration, timing, and rate of change. They also found that with sufficiently long analysis periods (of 15 years or more), an overlap in measurement days of 50% of records or more would suffice to enable comparisons of IHAs between monitoring sites. Estimating indicators at ungauged sites adds further uncertainty depending on the regionalization method; the most uncertain indicators, like those measuring flow dynamics, are also the most difficult to transfer to ungauged sites (Carlisle et al., 2010; Murphy et al., 2013; Zhang et al., 2014; Westerberg et al., 2016; Yang et al., 2016; Peňas et al., 2016; Eddy et al., 2017).

In flow-hydrological frameworks, the added value of using hydrologic models resides in the potential capacity of assessing ex-ante changes in flow regime under alternative, foreseeable conditions. Hydrological models can be used to (i) assess flow regime conditions (i.e., IHA sets) in ungauged reaches, (ii) classify river reaches based on flow regime similarity, and (iii) assess the 'natural flow regime conditions' of reference, which allow quantifying the deviation of current conditions (Poff et al., 2010; Casper et al., 2012; Schneider et al., 2013). However, hydrological model outputs are also affected by uncertainty. Shrestha et al. (2014) found that a process-based model's ability of estimating flow regime indicators was reasonable in terms of magnitudes at both low and high flows, but poorer in characterizing seasonality, duration and rate of change. In comparing different models at five gauging stations in the U.S., Caldwell et al. (2015) found that regional-scale models had comparable performance in estimating flow regime indicators than more complex, fine-scale models. Vis et al. (2015) showed that estimation of 12 IHAs in 27 catchments with a process-based model was sensitive to calibration objective functions, and demonstrated how models that were calibrated for correct simulation of high flows were unfit to assess low flow characteristics. Similarly, Zhang et al. (2016) found that a multi-objective calibration that included several flow regime characteristics in its objective function improved model simulations for ecological purposes; however, they noted trade-offs between metrics, so that multi-purpose calibration degraded to some extent the simulation of high flows. Thus, the ability of models to characterize flow regime depends not only on the model structure and input data, but as well on the calibration method and, ultimately, on the purpose for which the model was developed.

Despite being generally accepted from a conceptual point of view, application of eco-hydrological frameworks is still limited in the practice of catchment planning and water management. One factor limiting larger adoption is that assessing flow regime indicators is resource demanding (e.g. Swirepik et al., 2016): limited time and budgets may persuade resource managers to drop the analysis. On the other hand, global and continental hydrological models are increasingly made available and could provide information that can be accessed with limited effort. In other cases, local hydrological models that have been developed for other purposes may be available to decision makers. Using these models that were pre-existent to the ecohydrological analysis and that were developed for other purposes, termed herein 'off-the-shelf' as in contrast to tailor-made models, could possibly reduce the burden on local basin management resources. However, decision makers should be aware of potentially large epistemic uncertainties in IHAs estimated with off-the-shelf models, while continental hydrological modelling are affected by epistemic uncertainty as well, chiefly due to scale mismatch with respect to the ecological analysis.

The general aim of this study was to assess to which extent off-the-shelf hydrological models and continental scale models can be used in

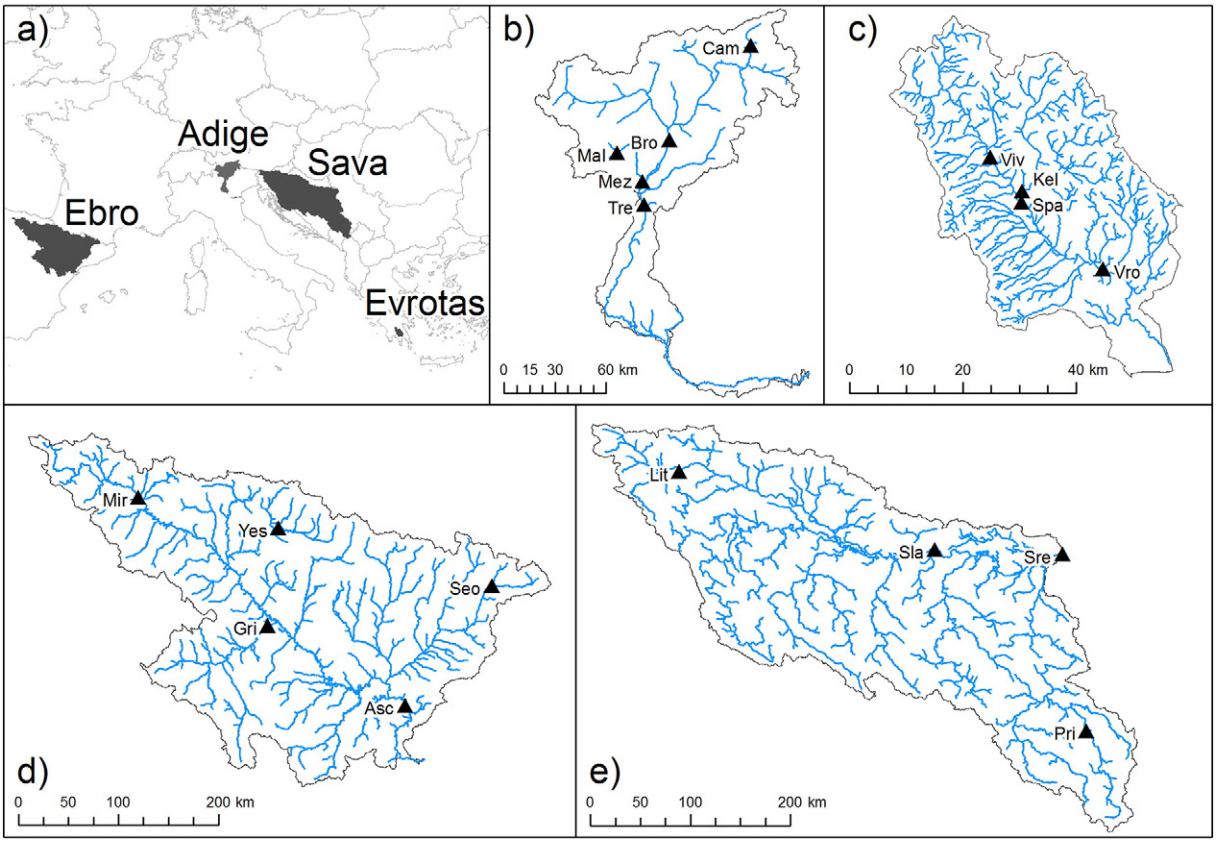


Fig. 1. (a) Location of the four studied basins in Mediterranean Europe, in clockwise order (b) Adige, (c) Evrotas, (d) Ebro, and (e) Sava. For each basin, the monitoring stations used for the analysis are indicated with triangles.

flow-ecological framework analysis. Specific objectives were to assess (i) the uncertainty of modelled flow regime indicators; (ii) the influence of the period of analysis on the uncertainty of modelled indicators; and (iii) the impact of modelling natural reference conditions on assessing ecological risk.

2. Materials and methods

The evaluation of model uncertainty was conducted for selected monitoring stations in four South European basins. After evaluation of hydrological model outputs in terms of monthly flow simulations, uncertainty was assessed for commonly used daily and monthly flow regime indicators. The sensitivity to the length of analysis period was assessed by comparing indicators accuracy when indicators were estimated from ten or 20 years of flow records. Finally, the impact of using modelled reference regime was assessed by applying a flow-ecological framework to the sites.

2.1. Study areas and selection of stations

The analysis was conducted in four river basins of South Europe: the Adige, the Ebro, the Evrotas and the Sava (Fig. 1). The basins are part of

Table 1
Overview of gauged stations selected for the study. Monitoring period refers to continuous daily records used as reference periods.

Basin	River	Station	ID	Monitoring period	Drainage area (km^2)	Flow regime pressures
Adige	Aurino	Caminata	Cam	1994–2013	420	Limited pressure; reference site
	Noce	Malé	Mal	1994–2013	466	Limited hydropeaking
	Noce	Mezzolombardo	Mez	1994–2013	1353	Hydropeaking; irrigation
	Adige	Bronzolo	Bro	1994–2013	6913	Hydropeaking; irrigation
	Adige	Trento Ponte S. Lorenzo	Tre	1994–2013	9768	Hydropeaking; irrigation
Ebro	Valira	Seo d'Urgell	Seo	1995–2014	559	Limited pressure; reference site
	Aragon	Yesa	Yes	1995–2014	2191	Dams
	Ebro	Miranda	Mir	1980–1998	5481	Dams
	Jalon	Grisen	Gri	1995–2014	9694	Irrigation
	Ebro	Asco Coca	Asc	1995–2014	82,245	Dams
Evrotas[a]	Evrotas	Vivari Sellasia	Viv	1994–2013 (Jan–Nov 2007)	459	Limited pressure; reference site
	Oinountas	Kelefina Kladas	Kel	2006–2015 (Nov 2007–Jan 2008)	309	Limited pressure; reference site
	Evrotas	Sparti Bridge	Spa	2006–2015 (Nov 2006–Jan 2008)	800	Irrigation; water abstraction
	Evrotas	Vrontamas Bridge	Vro	1990–1995; 2004–2015 (Mar 2007–Mar 2008)	1589	Irrigation; water abstraction
Sava	Lim	Prijepolje	Pri	1991–2010	3160	Limited pressure; reference site
	Sava	Litija	Lit	1993–2012	4768	Hydropower dams
	Sava	Slavonski Brod	Sla	1982; 2004–2012	54,134	Irrigation, water abstraction
	Sava	Sremska Mitrovica	Sre	1992–2010	87,966	Irrigation, water abstraction

[a] In the Evrotas basin, the monitoring period refers to periodic (once or twice per month) sampling; periods of continuous daily monitoring are indicated in brackets.

Table 2
Overview of hydrological models in their applications to the case study basins. More information on input data and hydrological processes implemented in the models can be found in the Supplementary Information (Tables SI.1 and SI.2).

Model	LISFLOOD	mHM	ADSWAT	EUSWAT	EVSWAT	HYPERstream
Basins of application	All basins	Ebro, Sava	Adige	Ebro, Sava	Evrotas	Adige
Spatial resolution	Pixel 5 × 5 km^2	Pixel 24 × 24 km^2	Subbasins of 260 km^2	Subbasins of 140 km^2	Subbasins of 10 km^2	Pixel 5 × 5 km^2
Simulation period	1996–2015	1995–2014	2001–2010	Ebro: 1995–2004 Sava: 1995–2009	1990–2015	1990–2013

the Globaqua project (Navarro-Ortega et al., 2015) study sites, and were chosen to encompass a range of geographic and socio-ecological conditions. All basins are affected by water scarcity either due to climatic or societal reasons, but anthropogenic stressors acting on the freshwater network are different. The Adige is an alpine basin located in the north-eastern part of Italy. With a size of about 12,000 km^2 and a length of about 409 km, it is the third largest Italian river basin. Its source is in the southeastern Alps near the border with Austria and Switzerland, and it enters the Adriatic Sea south of the Venice lagoon. The climate is characterized by dry winters, snowmelt in spring, and humid summer and falls. The long term annual mean temperature (1961–1990) is 3 °C. The annual mean precipitation is 1456 mm and ranges between 400 and

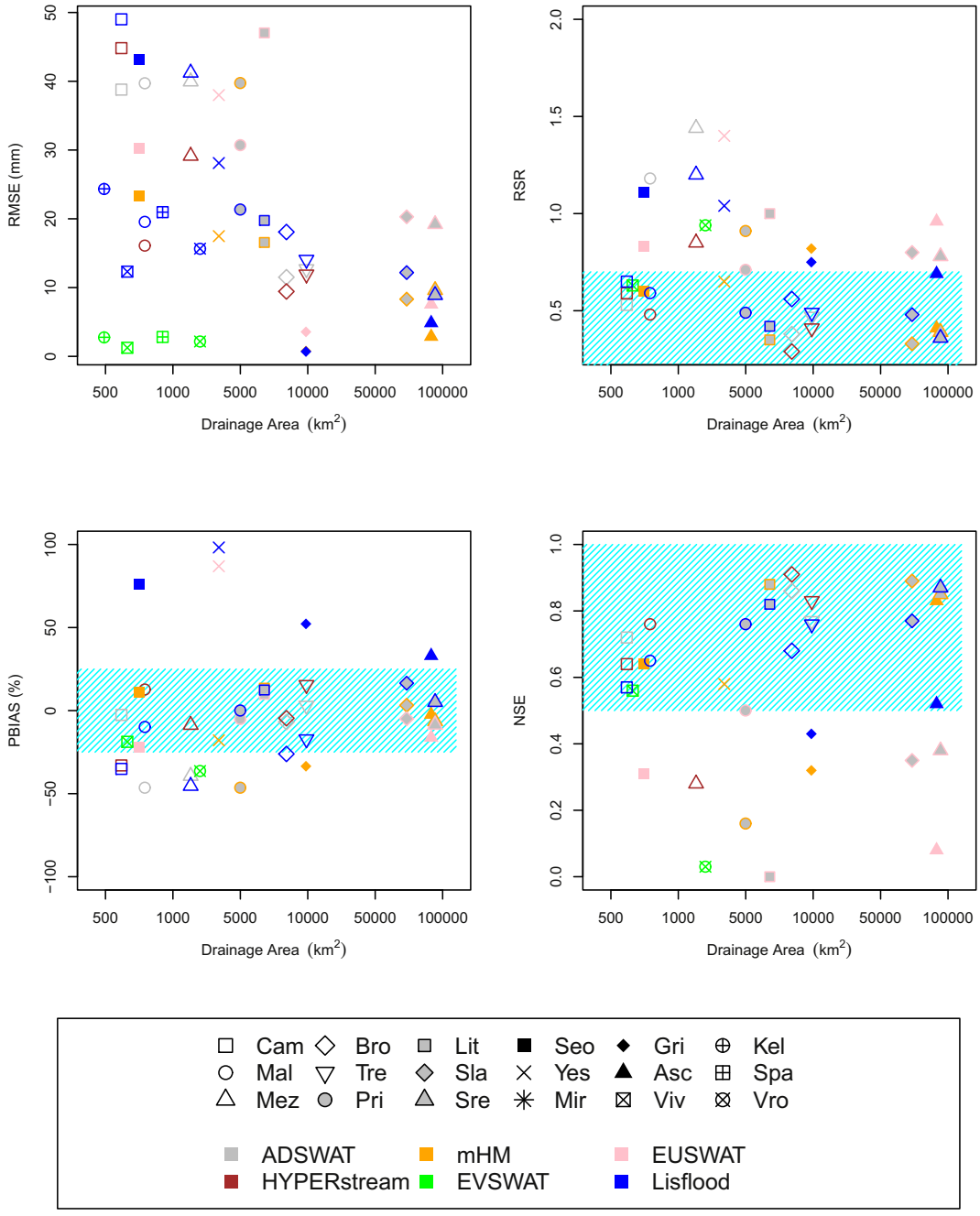


Fig. 2. Models performance in simulating monthly flow in terms of RMSE (a), RSR (b), PBIAS (c) and NSE (d). The x-axis indicates the drainage area of the station (in log10 scale). Hatched areas indicate acceptable performance according to Moriasi et al. (2007). Not all station-model combinations may be shown as the y-axis was cut to central values.

500 mm in the north-west part of the catchment and 1600 mm in the central-east part (period from 1961 to 1990; Chiogna et al., 2016). Land use consists of forest (56%), agriculture (12%), grassland and sparse vegetation (both around 17%). Starting from the beginning of the last century, with acceleration in the 50s, 34 large reservoirs have been built mainly for hydropower generation. Hydropower exploitation has induced significant streamflow alterations in the main stem and tributaries (Majone et al., 2016), particularly at intermediate and low flow regimes (Zolezzi et al., 2009; Bellin et al., 2016).

The Ebro River is a major river in Spain; it drains a basin of 85,550 km^2 over a river length of 928 km. Its Delta forms one of the largest wetland areas (320 km^2) in the western Mediterranean region. Climate ranges from meso- to supra-Mediterranean. Mean annual precipitation is 650 mm, varying from 300 mm in the central area to 1700 mm in the Pyrenees mountain range. Precipitation in form of snow is abundant in winter and early spring (Bejarano et al., 2010). The basin hosts >2.7 million inhabitants and approximately 45% of the population concentrate in five cities located next to the Ebro River or its tributaries. Rainfed agriculture covers 37% of the basin, whereas irrigated agriculture represents 15%. Forests cover 24%, and shrublands and grasslands cover 23% of the total basin area. Water abstractions for agricultural and industrial activities and the impact of waste water treatment plants have deteriorated soil and water quality (Lutz et al., 2016). The Ebro is largely regulated by dams and channels, which have altered its hydrological regime, especially in its middle and lower reaches (Bejarano et al., 2010; Majone et al., 2012).

The Evrotas river basin is located at the south-eastern part of the Peloponnese (Prefecture of Laconia, Greece). It covers an area of 1739 km^2, and has a mean altitude of 627 m. The Evrotas River is about 90 km in length and is fed by numerous intermittent and ephemeral tributaries. The basin has a typical Mediterranean climate, with an average annual temperature of 16 °C (range between 11 and 21 °C) and mean annual precipitation of 858 mm (1970–2016). Most of the river basin is covered by forest and seminatural vegetation (65%), followed by agricultural

Table 3
Flow regime indicators as derived from selected literature and considered in this study. IQR = interquartile range, i.e. the 75th percentile–25th percentile of the distribution.

Flow regime characteristics	Water Resource Indicators (WRIs)	Monthly Flow Risk Indicators (MFRI)	7 parsimonious daily flow indicators (Mag7)	Daily indicators of Hydrologic Alteration (IHAs)
	Shrestha et al. (2014)	Laizé et al. (2014)	Archfield et al. (2014)	Murphy et al. (2013)
Magnitude	Annual flow = Mean total annual flow [mm/year]	med_h = number of months above the upper threshold[a] IQR_h = IQR of flows above the upper threshold	MDF = mean daily flow [mm/day] CV = coefficient of variation of daily flow [mm/day]	E85 = 85% exceedance of daily streamflow [mm/day] MA26 = variability of March (standard deviation/mean) [%] MH10 = maximum October streamflow [mm/day]
		med_l = number of months below the lower threshold[a] IQR_l = IQR of number of months below the lower threshold	Skew = skewness of daily flow Kurt = kurtosis of daily flow	ML18 = variability in baseflow (standard deviation of the ratios of 7-days moving average flows to mean annual flow [%] ML20 = baseflow (ratio of total flow to total base flow) SEP_med = median September daily streamflow [mm/day]
Timing	Winter = median winter flow (Dec–Feb) [mm/day]	med_Jan = median flow in January [mm/day] IQR_Jan = IQR range of flow in January	Ampl = amplitude	TA1 = constancy, stability of flow regime
	Spring = median spring flow (Mar–May), [mm/day]	med_Apr = median flow in April [mm/day] IQR_Apr = IQR range of flow in April	Phase	TH1 = Julian day of occurrence of annual maximum flow [day]
	Summer = median summer flow (Jun-Aug), [mm/day]	med_Jul = median flow in July [mm/day] IQR_Jul = IQR range of flow in July		TL1 = Julian day of occurrence of annual minimum flow [day]
	Autumn = median autumn flow (Sep–Nov), [mm/day] Day50 = median Julian day of occurrence of 50% of annual flow [day]	med_Oct = median flow in October [mm/day] IQR_Oct = IQR of flow in October		
Frequency		Mon_h = modal month of maximum flow		FH6 = frequency of moderate flooding (three times median annual flow), average number of moderate flow events [number of events/year]
		Mon_l = modal month of minimum flow		FH7 = frequency of high flows (seven times median annual flow), average number of high flow events per year FL2 = variability in low pulse count; coefficient of variation for the number of annual occurrences of daily flows less than 25th percentile
Duration		med_seq = median number of sequences at least 2-month long below the lower threshold IQR_seq = IQR of number of sequences at least 2-month long below the lower threshold		DH13 = average of the annual maximum of 30 day moving average divided by median for the entire period DH16 = variability in high pulse duration, standard deviation of annual average high pulse duration (daily flows greater than 75th percentile) DL6 = variability of annual minimum daily average flow, standard deviation for the minimum daily streamflow divided by the mean streamflow pf the period
Rate of change			AR1 = autoregressive of 1 day	RA5 = number of day rises (number of days in which streamflow is greater than previous day divided by the total number of days in the period) RA7 = rate of streamflow recession, median change of log of flow for days in which the change is negative [mm/day] RA8 = flow direction reversals, average number of days per year when flow changes from rising of falling or from falling to raising [n/year]

[a] The upper threshold is the 95th percentile of the reference (naturalized) flow; the lower threshold is the 5th percentile of the reference (naturalized) flow.

areas (34%; EEA, 2012). The population density of the basin is about 26 inhabitants/km^2, i.e., about 45,000 inhabitants (Hellenic Statistical Authority, 2011). The main anthropogenic pressures in the Evrotas basin comprise overexploitation of water resources for irrigation, disposal of agro-industrial wastes, and agrochemical pollution (Skoulikidis et al., 2011). There are many, mostly illegal, surface water abstraction points along the river and about 3500 bore-hole drillings for irrigation (Skoulikidis et al., 2011), which influence surface and groundwater interactions and often result in the dryness of Evrotas River network during summer (Gamvroudis et al., 2017).

The Sava River (945 km) is the largest tributary of the Danube River. The basin extends over 97,713.20 km^2 across Slovenia, Croatia, Bosnia and Herzegovina, Serbia, Montenegro, and Albania (ISRBC, 2009). The climate varies across the basin from alpine to moderate continental depending on orography and influence of the sea. Mean annual precipitation ranges from about 1100 mm in the alpine area of Slovenia to about 650 mm in the Serbian plains. Most precipitation occurs in summer season or in autumn; a substantial portion of precipitation in the basin falls as snow in winter, feeding spring to early summer flow (ISRBC, 2009). Average annual air temperature for the whole Sava Basin is about 9.5 °C and ranges from 5 °C in the headwaters to 12 °C at its mouth. Cultivated land covers 23.2% of the basin, pasture 6.7%, boreal forest 1.5%, mixed forest 31.7%, and deciduous forest 36.1% (Levi et al., 2015). The population in the Sava basin is around 8.2 million (ISBRC, 2009). Thermal and nuclear power plant cooling, 19 large dams for hydropower production, and flood protection structures exert major anthropogenic pressures on flow regime (Levi et al., 2015).

Several flow gauging stations representative of different spatial scales, flow regime conditions, and flow pressures were selected in each basin (Table 1). The minimum drainage area was set to about 400 km^2. All stations were gauged; the availability of monitored daily flow allowed assessing model performances of monthly flow simulations and estimating flow regime indicators from observations. Five stations were considered as reference sites, i.e. stations whose flow regime is not impacted by anthropogenic pressures and can be considered representative of natural conditions. The selection was based on expert knowledge from basin ecologists and hydrologists.

2.2. Overview of hydrological models

Several pre-existing spatially distributed hydrological models were available for the study. All models provided daily outputs, but differed in terms of input data, simulation period, spatial resolution, and calibration method, reflecting the variety in 'off-the-shelf' hydrological models. Most frequently, calibration was based on one or several commonly used performance metrics: percent bias (PBIAS), Nash-Sutcliffe

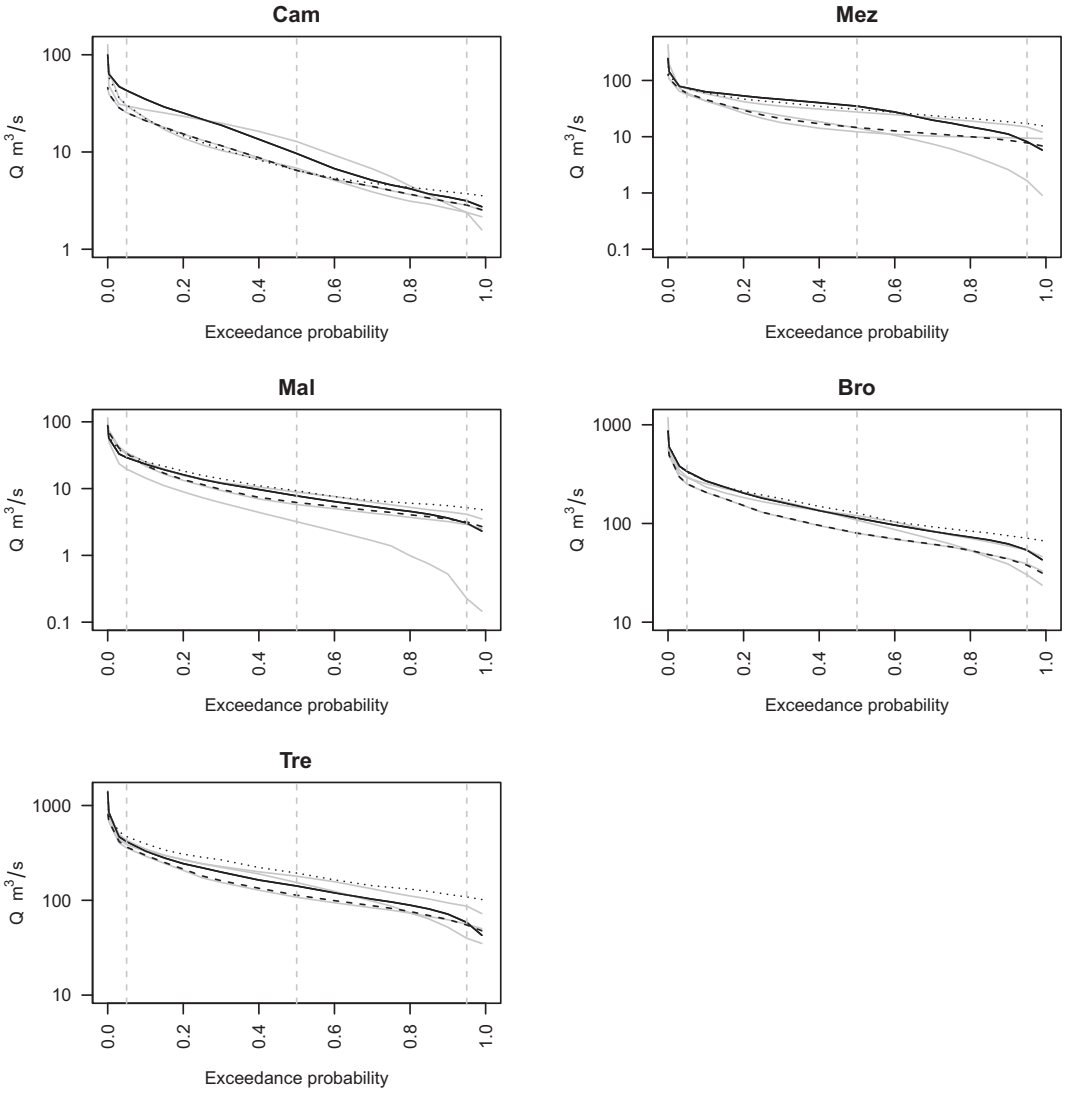


Fig. 3. Flow duration curve of daily flow at the Adige basin stations. The black line indicates observed flow; grey lines indicate modelled flow. The grey dashed vertical lines indicate the 5th, 50th and 95th percentiles. Zero flow is set at 0.001 m^3/s. Modelled naturalized flow is reported as black dashed (LisQNat) or dotted (HyperNat) lines.

Efficiency (NSE), Pearson's correlation coefficient (R2), Root Mean Square error (RMSE), and the ratio of RMSE divided by the standard deviation of measured data (RSR; Moriasi et al., 2007). Only a short description of each model and its application to the study areas are provided here; the interested reader is referred to the Supplementary Information (SI) and cited literature for more details. Table 2 summarizes the spatial and temporal resolution of the models in the application to the study basins.

2.2.1. Lisflood

Lisflood is a spatially distributed model designed to simulate hydrological processes in large European river basins for flood forecasting and assessing the effects of river regulation, climate and land use changes (Van der Knijff et al., 2008). The model accounts for snow melt, infiltration, interception, leaf drainage, evapotranspiration, preferential flow in the soil, deep drainage, and generation of quick flow (surface runoff plus lateral flow) and slow flow. The spatial resolution of the model for this application was at 5 × 5 km pixel size. Calibration is done for eight parameters related to the main water fluxes (Van der Knijff et al., 2008); it starts from most upstream gauged stations and moves downstream, adapting calibration parameter sets to the enlarged basin, on the basis of RMSE and NSE

2.2.2. mHM

The mesoscale Hydrologic Model (mHM; www.ufz.de/mhm) is a grid-based distributed hydrological model that simulates canopy interception, snow accumulation and melting, soil moisture dynamics, infiltration, deep percolation, surface runoff, evapotranspiration, storage in the subsurface and groundwater, discharge generation, baseflow, fast and slow interflow (Samaniego et al., 2010; Kumar et al., 2013). The spatial distribution of model parameters is derived from catchment characteristics such as soil types, geological classes and land cover types using a multiscale parameter regionalization technique (Samaniego et al., 2010). Rakovec et al. (2016a, 2016b) report on simulation and validation of mHM in several European basins, including the Ebro and the Sava. For this study, mHM was applied at 24 × 24 km pixel size in the Ebro and the Sava. Daily streamflow for 1995–2014 was used to calibrate the model at all selected stations with the Dynamically Dimensioned Search (DDS; Tolson and Shoemaker, 2007) calibration strategy. The objective function consisted of a combination of the NSE between observed and modelled flows and their logarithms to capture high, average, and low flows.

2.2.3. SWAT

The Soil and Water Assessment Tool (SWAT, Neitsch et al., 2011; Arnold et al., 2012) simulates daily water cycle, crop development,

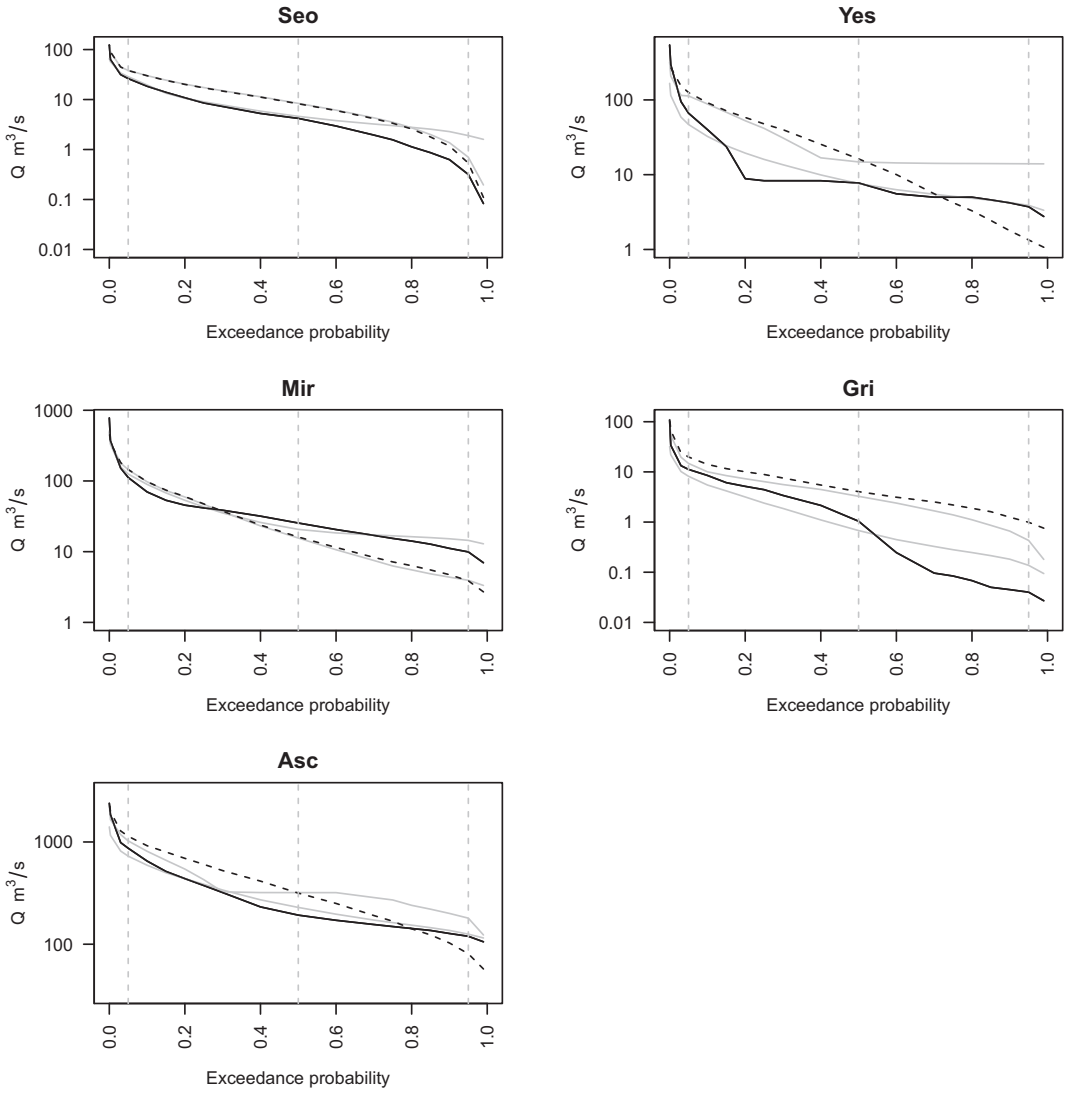


Fig. 4. Flow duration curve of daily flow at the Ebro basin stations. The black line indicates observed flow; grey lines indicate modelled flow. The grey dashed vertical lines indicate the 5th, 50th and 95th percentiles. Zero flow is set at 0.001 m^3/s. Modelled naturalized flow is reported as black dashed line (LisQNat).

sediment, nutrient and pesticide transport in a basin. The daily water balance considers precipitation, irrigation, evapotranspiration, surface runoff, lateral flow, and percolation to shallow and/or deep aquifers. Daily water yield is routed through the river network in the cascading sequence of streams along the network. Application of SWAT in the four basins differed for purposes and spatial resolution; therefore, distinct SWAT applications are further defined.

2.2.3.1. EUSWAT. SWAT applications of the Ebro and Sava basins were part of large scale modelling with subbasins of about 180 km^2 of size. Simulations in the Ebro were extracted from a model developed for the entire Iberian Peninsula for the decade 1995–2004 (Malagó et al., 2015). The Sava Basin was modelled as part of the Danube Basin and the simulation period covered 1995–2009 (Malagó et al., 2017). A regionalized calibration and validation procedure was applied to ensure good simulation of monthly streamflow and its components. Calibration was conducted in headwaters and transferred to ungauged subbasins based on hydrological similarity. Criteria and time scale of model outputs considered in the procedure depended on the processes under consideration (Malagó et al., 2015, 2017), but the main performance criteria were PBIAS, NSE, and the coefficient of determination multiplied by the coefficient of the regression line bR2 (Moriasi et al., 2007). None of the stations in the Sava selected for this study were part of the model calibration dataset. The version used in this study was SWAT2009 for Ebro and SWAT2012 for Sava (Neitsch et al., 2011; Arnold et al., 2012). As this was the only model that was calibrated at monthly resolution, its simulations were only used for the assessment of monthly indicators.

2.2.3.2. ADSWAT. The SWAT model of the Adige basin comprised 43 subbasins with the size of 260 km^2 on average. The period 2001–2005 was used for model calibration of daily flow at Trento Ponte S. Lorenzo and Bronzolo stations (Fig. 1, Table 1), whereas the model was validated in the period 2006–2010 (Tuo et al., 2016). Sensitive parameters were identified with a preliminary one-at-a-time sensitivity analysis; snow parameters were fixed on the basis of information available at neighboring regions (Tuo et al., 2016). Model calibration was performed using the Sequential Uncertainty Fitting algorithm version 2 (SUFI-2; Abbaspour, 2015) in four iterations. Initial parameter ranges were bounded to physically reasonable intervals according to literature (Arnold et al., 2012; Grusson et al., 2015; Vu et al., 2012). After each iteration, parameter ranges were modified (normally narrowed down) according to calibration results within their reasonable physical limitations. NSE was used to assess goodness of fit of simulations, following Moriasi et al. (2007) guidelines for acceptable performance.

2.2.3.3. EVSWAT. The SWAT model of the Evrotas river basin was developed by Gamvroudis (2016). The basin was delineated into 150 subbasins which are further subdivided into homogeneous hydrologic response units HRUs. The amount of irrigation was estimated based on the agricultural usage of electricity and the direct withdrawal from the river for irrigation purposes (Tzoraki et al., 2011). The model was calibrated against daily flow for the years 2009–2011 at seven sites and

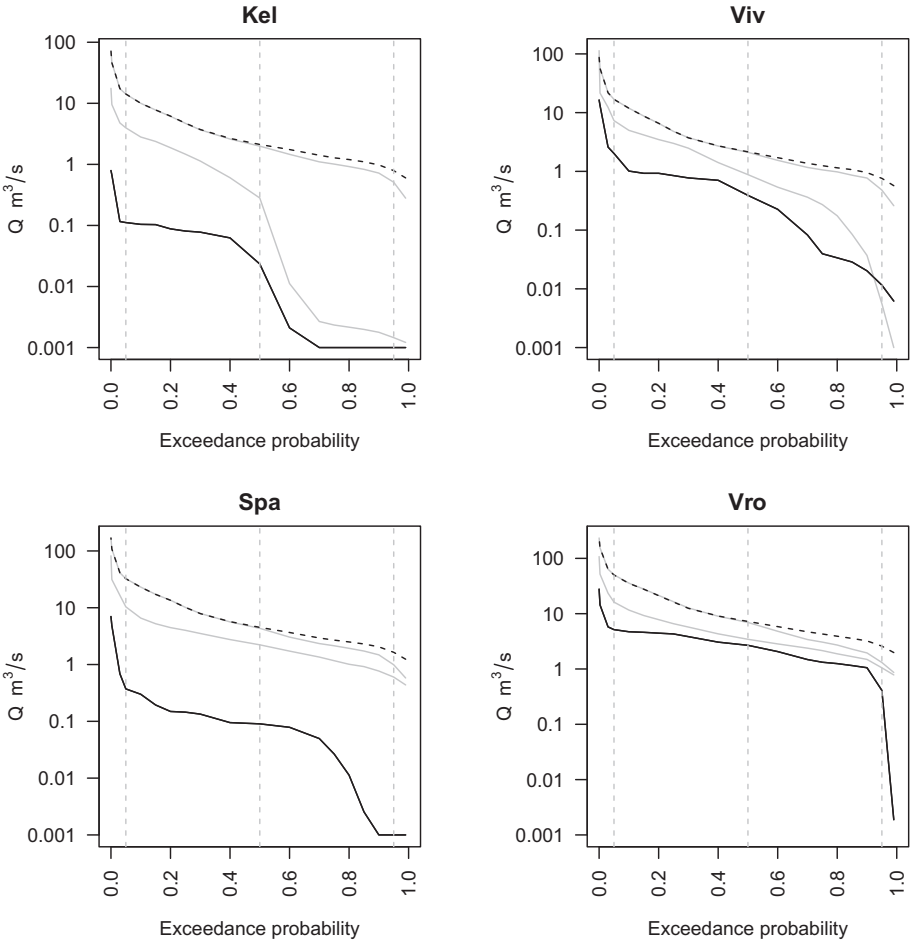


Fig. 5. Flow duration curve of daily flow at the Evrotas basin stations. The black line indicates observed flow; grey lines indicate modelled flow. The grey dashed vertical lines indicate the 5th, 50th and 95th percentiles. Zero flow is set at 0.001 m^3/s. Modelled naturalized flow is reported as black dashed line (LisQNat).

validated for the years 2004–2009 at two sites using PBIAS, NSE and RSR to assess model performance (Gamvroudis et al., 2015, 2017). Model simulation of streamflow was in good agreement with field observations in all gauging stations along Evrotas River and its tributaries. Input data (precipitation and temperature) were further complemented in order to run the model for the period 1973–2015.

2.2.4. HYPER (HYPERstream)

Hydrological simulations in the Adige were performed with the HYPERstream routing scheme (Piccolroaz et al., 2015), coupled with a continuous SCS-CN module for surface flow generation (Michel et al., 2005). The HYPERstream routing scheme is based on the width function instantaneous unit hydrograph (WFIUH) theory and is designed for coupling with climate models and, in general, with gridded climate datasets. In this study, the SCS-CN methodology is coupled with a non-linear bucket model for soil moisture depletion (Majone et al., 2010), a degree-day approach for snow dynamics, and a linear reservoir model for simulating groundwater dynamics and the associated base flow (Laiti et al., 2017). Evapotranspiration is computed by multiplying Hargreaves-Samani potential evapotranspiration by a linear limiting function depending on soil moisture (Bellin et al., 2016). For the purpose of this work, HYPERstream was applied over a 5-km computational grid during the period 1990–2013. Daily flow was calibrated by means of a Particle Swarming Optimization tool (PSO, in analogy with Piccolroaz et al., 2015) in the period 1992–2013 at Bronzolo and Trento Ponte S. Lorenzo stations (Table 1) to maximize flow NSE. A second calibration was performed with the same procedure in period 1926–1949, preceding the period of intense damming of the river basin, to obtain a calibration for almost pristine conditions.

To provide an initial evaluation of model simulation accuracy at the selected stations, model performance metrics for monthly flow simulations in the decade 2000–2009 are shown in Fig. 2 (monthly model simulations at all stations are shown in SI). Performances were comparable across models: none of the models succeeded in simulating monthly flow acceptably at all stations according to Moriasi et al. (2007) criteria (indicated as hatched areas). Conversely, all models performed well in some stations. Local calibration generally resulted in acceptable performance, for example at Bro and Tre stations in the Adige for ADSWAT and HYPERstream. Moreover, models that were not calibrated locally performed well in several cases, e.g. EUSWAT in the Sava. Modelling monthly flow was particularly difficult in the Ebro basin, where only local independent calibration of mHM at Seo, Yes, and Asc resulted in acceptable performances. Modelling flows in the Evrotas basin was challenging: local calibration allowed acceptable performance of EVSWAT in some but not all stations, whereas the pan-European Lisflood model overestimated flow (see SI). Conversely, flow simulations in the Sava were generally good for all models. Performance generally improved with increasing drainage areas, but did not differ significantly between stations of small and medium size drainage area.

The low performances in monthly simulations indicate that model outputs were at times uncertain and inaccurate, but this situation reflects what may likely occur in ungauged reaches, or when using 'off-the-shelf' models. Given that performances were generally comparable across models and drainage areas, uncertainty of flow regime indicators

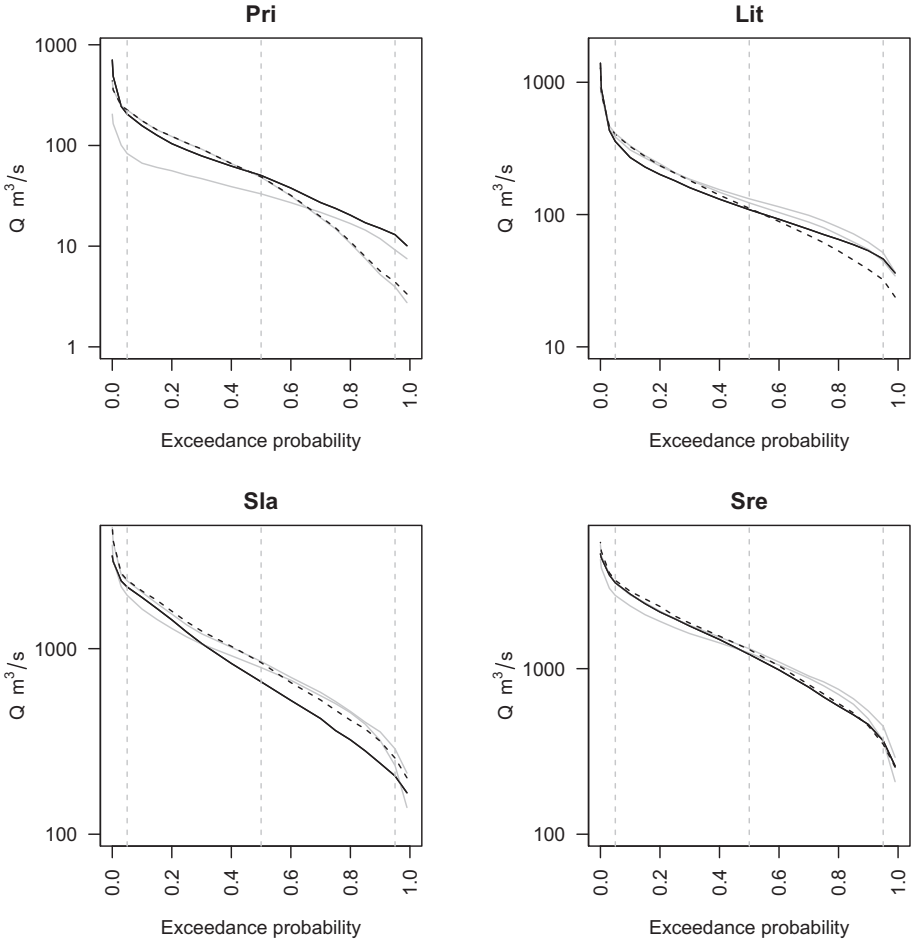


Fig. 6. Flow duration curve of daily flow at the Sava basin stations. The black line indicates observed flow; grey lines indicate modelled flow. The grey dashed vertical lines indicate the 5th, 50th and 95th percentiles. Zero flow is set at 0.001 m^3/s. Modelled naturalized flow is reported as black dashed line (LisQNat).

was conducted considering all stations and all models together (specific results per model and per station are shown in the SI).

2.3. Flow regime indicators

Flow regime characteristics comprise magnitude, duration, timing, frequency, and rate of change (Poff et al., 1997, 2010; Laizé et al., 2014). Magnitude can be visualized with daily flow duration curves, i.e. cumulative frequency distribution of daily flows, which were therefore included in the analysis. While useful, flow duration curves do not provide information about duration, timing, frequency, and rate of change. All applied flow regime indicators are summarized and described in Table 3.

In principle, relevant indicators should be selected based on the ecological response of importance for water management (Poff et al., 2010). In the absence of locally relevant information, however, and in the light of the high redundancy between indicators, parsimonious flow regime indicator sets that have been proposed in literature are explored herein (Table 3). These indicators include (i) six Water Resources Indicators (WRIs) that consider water availability and seasonality, i.e.

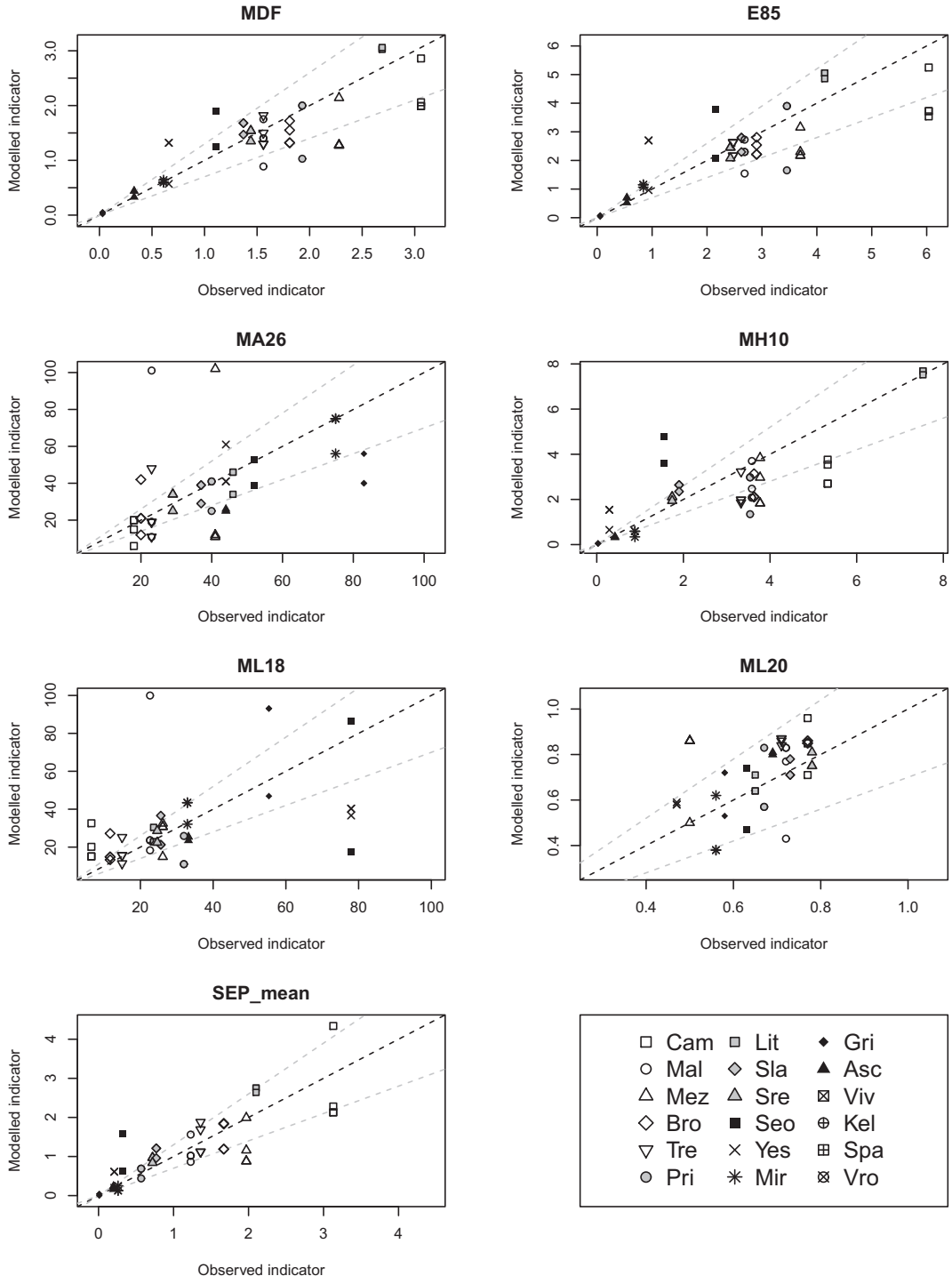


Fig. 7. Magnitude IHAs (Mean Daily Flow MDF plus magnitude IHAs selected from Murphy et al., 2013) in the reference decade 2000–2009 calculated with different hydrological models (y-axis) and from observed flow (x-axis) for all study sites. Black dashed line indicates 1:1 relationship; grey dashed lines indicate deviations of ±30% from observed flow. Symbols represent the selected gauging stations and are repeated for each model realization.

focusing on magnitude and timing of flow regime (Shrestha et al., 2014); (ii) seven parameters (Mag7) that represent flow regime parsimoniously (Archfield et al., 2014); (iii) 19 IHAs that are most frequently used in flow-ecological frameworks (Murphy et al., 2013); and (iv) 16 Monthly Flow Risk Indicators (MFRI) derived from monthly flow time series that mostly focus on magnitude, frequency, timing and duration (Laizé et al., 2014; Table 3).

The uncertainty of each indicator was assessed by evaluating RMSE, RSR and R2 between the modelled indicators against those obtained from observed data at all stations. In order to standardize indicator results across stations, streamflow and derived indicators were expressed in mm, although other units had been sometimes used in the original formulations (e.g. Shrestha et al., 2014). The analysis was performed in R environment. Mag7 and daily IHAs were calculated with the EFlowStats R package (Thompson and Archfield, 2015).

2.4. Impact of analysis period length

Reliable estimates of flow regime indicators should be derived from a sufficiently long period, ideally 20 years (Richter et al., 1997; Kennard et al., 2010; Eddy et al., 2017). However, available daily flow records and model simulations were in many cases shorter or incomplete (Table 1).

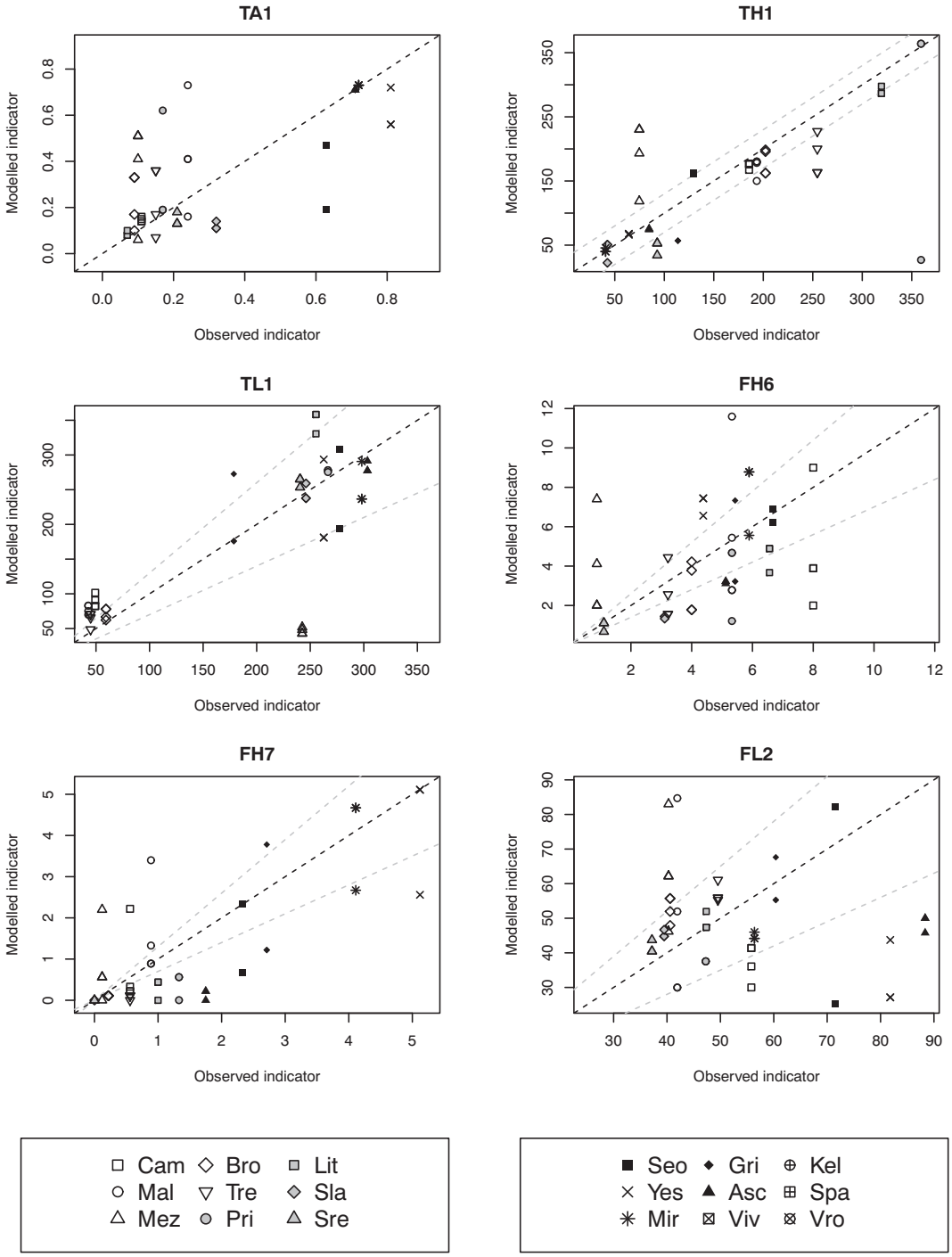


Fig. 8. Timing and frequency IHAs (selected from Murphy et al., 2013) in the reference decade calculated with hydrological models (y-axis) and from observed flow (x-axis). Black dashed line indicates 1:1 relationship; grey dashed lines indicate deviations of ±30% from observed flow; in case of Julian day indicators (TH1 and TL1) dashed lines indicate ±30 days. Symbols represent the selected gauging stations and are repeated for each model realization.

Some flexibility to define the period of analysis had thus to be accepted, while maximizing the overlap between simulated or observed periods among the gauging stations. The simulation period focused on the 2000–2009 decade, for which daily observed flow records were available in most stations (Table 1), and overlap with modelled flow was at least 50% (Kennard et al., 2010). Measurements at Miranda (Ebro) became sporadic after June 1999, and then started regularly after 2009. However, it is likely that the station location and rating curve have changed in the meantime, so only the decade prior to the 1999 measurement interruption was considered. Observed data in the Evrotas were particularly limited; continuous daily records lasted for less than two years, while for the rest of the period only monthly data were available. As the observation period was too short for reliable assessment (Kennard et al., 2010), the impact of analysis period length on modelled indicators uncertainty could not be assessed in Evrotas basin.

The sensitivity of indicator accuracy to the length of the analysis period was assessed by enlarging the analysis from ten to 20 years (1995–2014) or the 20-year period that mostly overlapped with this time window (Table 1). The analysis was limited to stations with sufficient observed data (thus excluding Mir in Ebro and all Evrotas stations), and to models with 20 year simulations, i.e. HYPERstream, mHM, and Lisflood. The impact of analysis period length on each

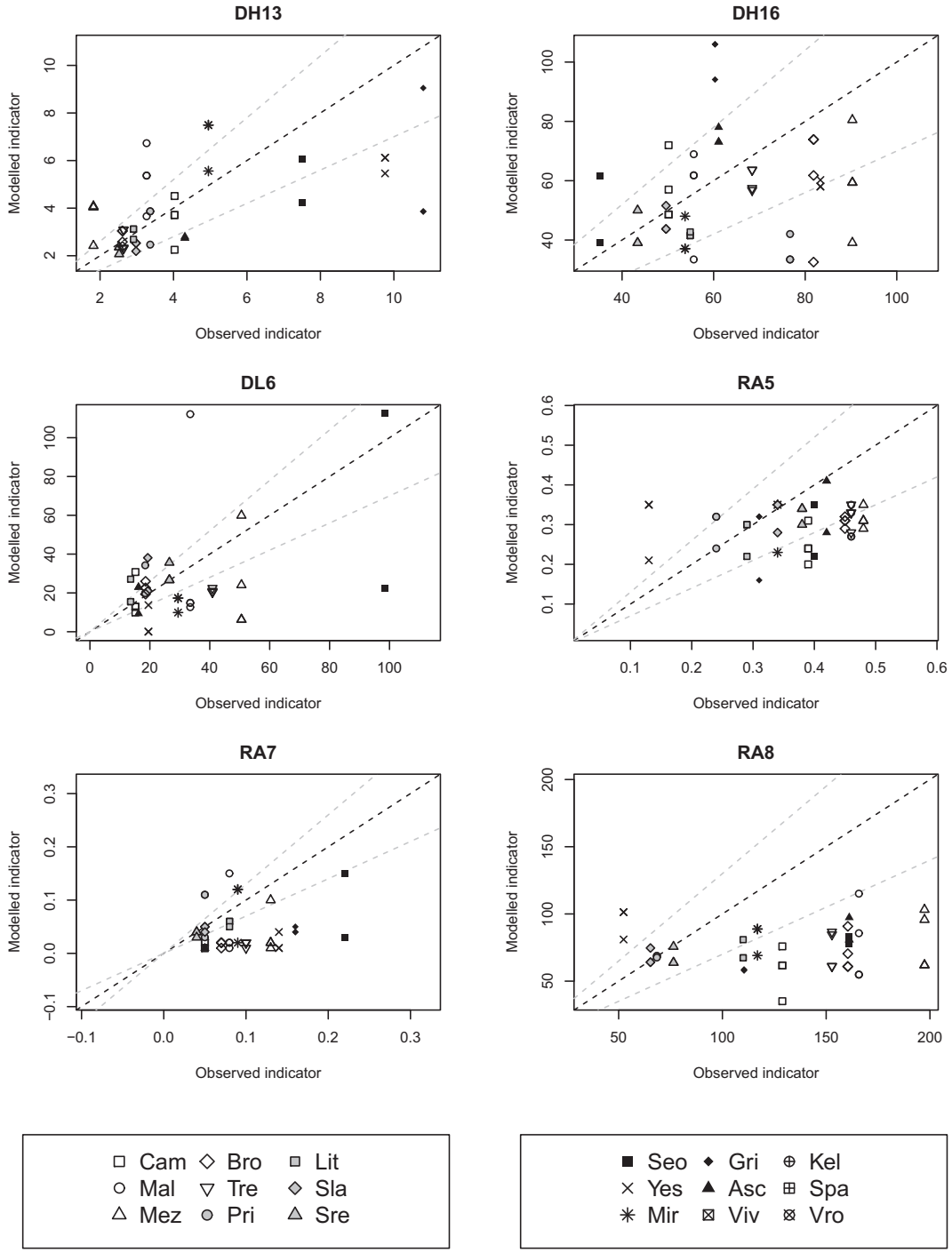


Fig. 9. Duration and rate of change IHAs (selected from Murphy et al., 2013) in the reference decade calculated with hydrological models (y-axis) and from observed flow (x-axis). Black dashed line indicates 1:1 relationship; grey dashed lines indicate deviations of ±30% from observed flow. Symbols represent the selected gauging stations and are repeated for each model realization.

indicator was assessed by evaluating changes in RMSE and R2 between modelled indicators and those obtained from observed data for the two periods.

2.5. Impact of natural flow regime simulation

Flow-ecological assessment frameworks are based on the deviation of the current conditions from natural flow regime, which can be inferred but not directly observed in altered reaches. Several methods exist for inferring natural flow regime indicators (e.g. Murphy et al., 2013; Peňas et al., 2016; Yang et al., 2016), among which the use of hydrological model simulations after removing human impacts (Poff et al., 2010; Laizé et al., 2014; Eddy et al., 2017; Bellin et al., 2016). Of all the models available for the study, only Lisflood and HYPERstream provided "naturalized flow" simulations. Lisflood naturalized flow (named from now on LisQnat) was obtained after removing dams and omitting water abstractions. HYPERstream naturalized flow (HyperQnat) was instead obtained by applying to the 1990–2015 period the pre-dam calibration set of 1926–1949.

The impact of using a modelled naturalized flow in flow-ecological assessment was determined by applying the flow-ecological ERFA framework (Laizé et al., 2014) as an example. ERFA was chosen because (i) its rules are clearly defined; (ii) it is based on monthly data, which was the minimum temporal resolution for which all models were calibrated; and (iii) it has been applied at pan-European scale (Schneider et al., 2013; Laizé et al., 2014), which is coherent with the scope of the work. As in other frameworks, ERFA assesses ecological risk as a function of the number of indicators that deviate significantly from a natural status (Poff et al., 2010; Eddy et al., 2017). The results of the analysis could thus be extended to other flow-ecological frameworks. In ERFA framework, ecological risk is assessed with the ERFA score, which is the number of MFRIs that deviates by more than ±30% (±1 month for modal month of high flows Mon_h and low flows Mon_l, Table 3) from the reference regime; the higher the score, the higher the

Table 4
Uncertainty of modelled flow regime indicators quantified as the RMSE, RSR, and Pearson's correlation coefficient R2 calculated for all models and stations compared to indicators obtained from observed flow for reference decade. Type = flow regime characteristic measured by indicator: M = magnitude, T = Timing, F = frequency, D = duration, R = rate of change. Fr dev = fraction of significant deviations, i.e. fraction of data entries falling outside boundaries of ±30% of indicator estimated from observed flow (±30 days for Julian day; ±1 month for Mon_h and Mon_l). Numbers in italic indicate high modelled uncertainty (fr dev > 0.50; RSR > 0.70, and R2 < 0.50). ρ = Spearman's rank correlation coefficient between indicator error and modelled mean daily streamflow error; the symbol * indicates if the coefficient was significant at probability level of 0.05. Daily indicator sample comprised 33 data entries (station x model combinations). Monthly indicator sample comprised 42 data entries. MFRIs threshold of low and high flow were set as LisQnat 5th and 95th percentiles.

Indicator set	Indicator	Type	Fr dev	RMSE	RSR	R2	ρ
WRIS (Shrestha et al., 2014)	Annual flow (mm/y)	M	0.36	191	0.60	0.65	0.87*
	Day50 (day)	T	0.15	22	*0.78*	0.57	0.45*
	Winter (mm/d)	T	0.39	0.47	*0.74*	0.50	0.51*
	Spring (mm/d)	T	0.36	0.54	0.68	0.59	0.61*
	Summer (mm/d)	T	*0.52*	0.77	0.49	0.81	0.83*
	Autumn (mm/d)	T	*0.58*	0.54	*0.73*	0.54	0.82*
Mag7 (Archfield et al., 2014)	MDF (mm/d)	M	0.33	0.48	0.56	0.70	0.87*
	CV (mm/d)	M	0.06	0.09	*0.81*	*0.42*	−0.45*
	Skew	M	0.24	0.13	*0.93*	*0.26*	−0.46*
	Kurt	M	0.55	0.11	*0.90*	*0.21*	−0.23
	Ampl	T	0.39	0.21	*0.87*	*0.37*	−0.32
	Phase	T	*0.94*	0.24	0.46	0.83	−0.36*
	AR1	R	0.15	0.08	*1.56*	*0.08*	−0.06
IHAs (Murphy et al., 2013)	E85 (mm/d)	M	0.33	0.95	0.61	*0.08*	0.67*
	MA26 (%)	M	0.48	23	*1.24*	*0.17*	0.004
	MH10 (mm/d)	M	*0.58*	1.26	0.63	0.64	0.71*
	ML18 (%)	M	0.42	23	*1.07*	*0.19*	−0.29
	ML20 (ratio)	M	0.12	0.14	*1.45*	*0.23*	−0.08
	SEP_mean (mm/d)	M	0.45	0.52	0.57	0.71	0.86*
	TA1	T	*0.58*	0.19	*0.72*	0.53	−0.68
	TH1 (day)	T	0.45	74	*0.78*	*0.46*	−0.07
	TL1 (day)	T	0.48	72	0.69	0.58	0.09
	FH6	F	*0.67*	136	*1.30*	*0.13*	−0.13
	FH7	F	*0.76*	71	*0.73*	0.54	−0.26
	FL2	F	0.33	144	*1.50*	*0.03*	−0.10
	DH13	D	0.39	2.06	*0.79*	*0.39*	−0.55*
	DH16	D	0.36	23	*1.43*	*0.01*	0.24
	DL6	D	*0.68*	25	*1.16*	*0.19*	0.08
	RA5	R	0.45	0.12	*1.31*	*0.08*	0.34
	RA7 (mm/d)	R	*0.81*	0.07	*1.54*	*0.06*	−0.13
	RA8	R	*0.76*	68	*1.56*	*0.05*	0.23
MFRIs (Laizé et al., 2014)	Med_h	M	0.40	0.90	*1.00*	*0.10*	0.58*
	IQR_h	M	*0.52*	0.96	*0.94*	*0.19*	0.47*
	Med_low	M	0.48	1.87	*1.35*	*0.18*	−0.42*
	IQR_low	M	0.45	1.21	*1.07*	*0.13*	−0.47*
	Med_Jan (mm/d)	T	0.36	0.51	*0.82*	*0.40*	0.55*
	IQR_Jan	T	*0.76*	0.44	*1.42*	*0.05*	0.02
	Med_Apr (mm/d)	T	0.40	0.63	0.61	0.72	0.58*
	IQR_Apr	T	*0.79*	0.65	*1.01*	*0.24*	0.41*
	Med_Jul (mm/d)	T	0.55	0.76	0.50	0.80	0.81*
	IQR_Jul	T	*0.76*	0.30	*0.72*	0.68	0.32
	Med_Oct (mm/d)	T	0.52	0.51	0.69	0.65	0.84*
	IQR_Oct	T	*0.74*	0.48	*1.09*	*0.30*	0.59
	Mon_h	F	0.31	3.87	*1.49*	*0.01*	0.41*
	Mon_l	F	0.38	3.21	*0.91*	*0.26*	0.00
	Med_seq	D	0.43	0.87	*1.74*	*0.09*	−0.43*
	IQR_seq	D	0.40	0.65	*1.56*	*0.02*	−0.30

ecological risk. The ecological risk class is defined according to ERFA score as: 0 (no risk), 1–5 (low risk), 6–10 (medium risk) and 11–16 (high risk). The analysis was applied for both periods of ten and 20 years (i) to all stations using LisQnat as reference conditions; (ii) to the Adige stations using HyperQnat as reference, and (iii) to the reference sites using observed flow as reference.

3. Results

3.1. Uncertainty of flow regime indicators

Daily flow duration curves (Figs. 3–6) provide visual information about observed and simulated flow magnitude for the 2000–2009 decade. Mismatches between models and observed flow occurred particularly at medium and low flows with high exceeding probability. Low flows proved to be difficult to simulate accurately and large scatter across modelled flow duration curves occurred at probability exceedance of 95% or more.

In Figs. 7–9, a sample of modelled (i.e. derived from simulated streamflow) IHAs is compared with IHAs derived from observed streamflow data for all gauging stations. Following Murphy et al. (2013), IHAs set was used as an example as it provides the most comprehensive suite of indicators, considered representative of all flow regime characteristics (Table 3; results for all indicators of Table 3 are shown in the SI). In the interpretation of differences between indicators obtained from models and from observed flow, deviations from reference conditions are considered significant if larger than ±30% (Shrestha et al., 2014; Murphy et al., 2013; Laizé et al., 2014), or, in the case of Julian day indicators, larger than ±30 days (Schneider et al., 2013). Differences in modelled and observed flow magnitude are mirrored in the uncertainty of magnitude indicators (Fig. 7). While average flow conditions (e.g. Mean daily flow MDF, or mean flow in September, Sep_mean) were reasonably modelled, the results for maximum flow in October (MH10) were less robust. The highest uncertainty in magnitude indicators was observed in indicators describing flow variability, low flow and baseflow conditions (e.g. 85% exceedance E85, variability of March MA26, and variability of baseflow ML18 and baseflow ratio ML20; Fig. 7). In terms of timing and frequency, TH1 (Julian day of maximum flow) and frequency of high flows (FH7) were well correlated to observations, whereas timing and frequency of low

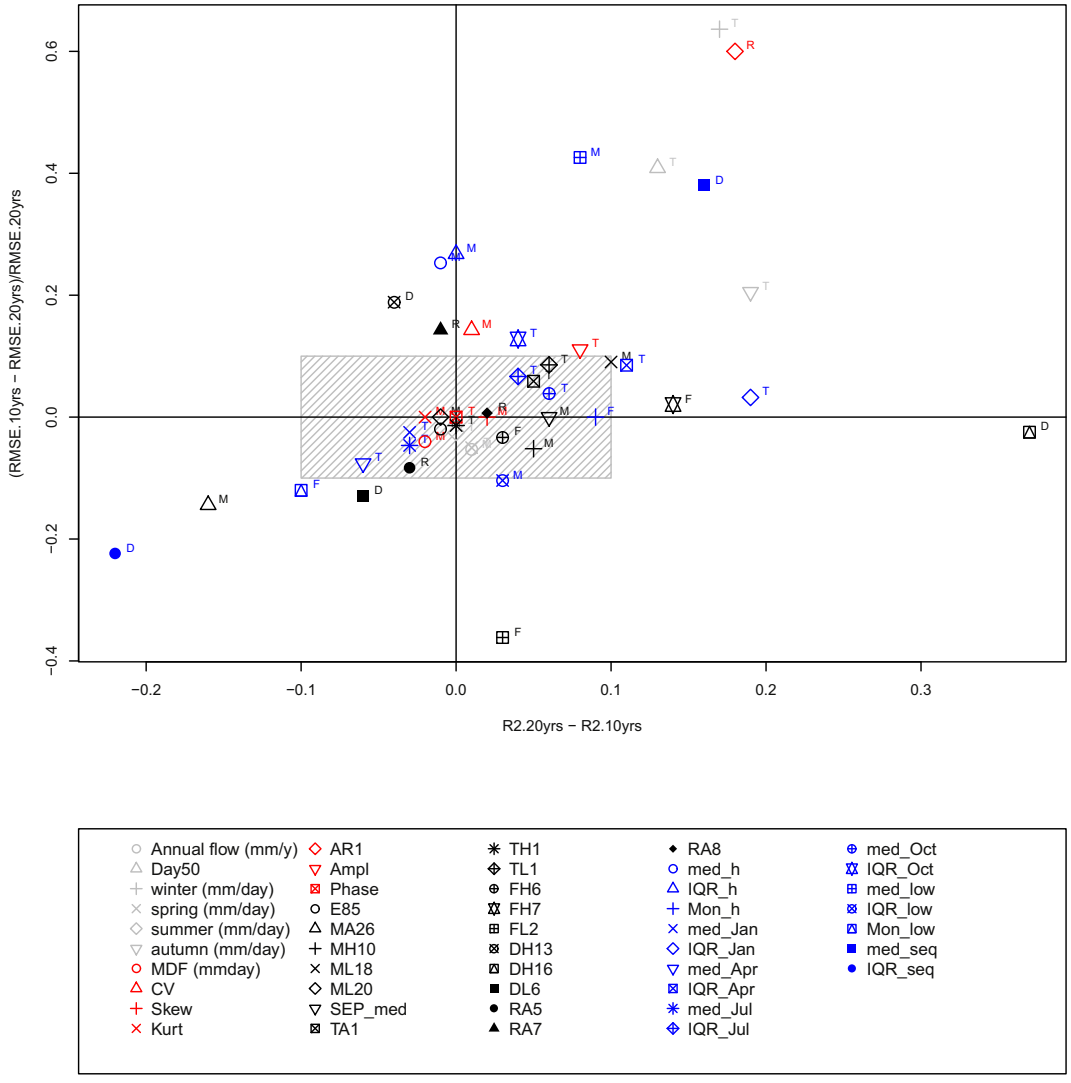


Fig. 10. Relative change in RMSE versus change of Pearson's correlation coefficients R2 of flow regime indicators evaluated with simulated and observed streamflows. The changes are evaluated according to the parameters obtained with 20 and 10 years of observations. Changes of ±10% (hatched grey area) can be considered non-significant given the small data ensemble sample (26 station-model combinations). Indicators falling outside the hatched area were sensitive to the length of the period considered for the estimation. Improvements in indicator estimation (reduction of RMSE and increase of R2) when passing from ten to 20 years fall in the positive (upper-right) panel. Letters indicate which type of flow characteristic the indicators refer to: M = magnitude; T = timing; F = frequency; D = duration, and R = rate of change.

flows were poorly captured (Fig. 8). The largest uncertainty affected duration and rate of change indicators (Fig. 9) as evidenced by the large scattering in modelled indicators.

Table 4 summarizes modelled indicator uncertainty for the data ensemble (all station-model combinations) in terms of (i) the fraction of data entries showing significant deviations from indicators derived from observed flow (i.e. falling outside the boundaries indicated in Figs. 7–9), and (ii) the deviation of model-derived indicators from observation-derived indicators expressed as RMSE, RSR and R2. RMSE indicates the magnitude of uncertainty of the single indicator, but it does not allow comparisons between different indicator types, given the different dimensions of the indicators. Instead, RSR can be used for assessing the uncertainty relative to observed variability as it provides a measure of the error in comparison with observed variability. Finally, R2 gives a measure of the predictive power of the model ensemble in relative terms. The presence of a monotonic relationship between indicator estimation error and modelled mean daily flow error was evaluated with Spearman's rank coefficient ρ.

3.2. Impact of analysis period length

Increasing the analysis period from ten to 20 years allowed enlarging the number of extremes in the observed flow, thereby capturing better flow variability. Observed flow daily mean and standard deviation did

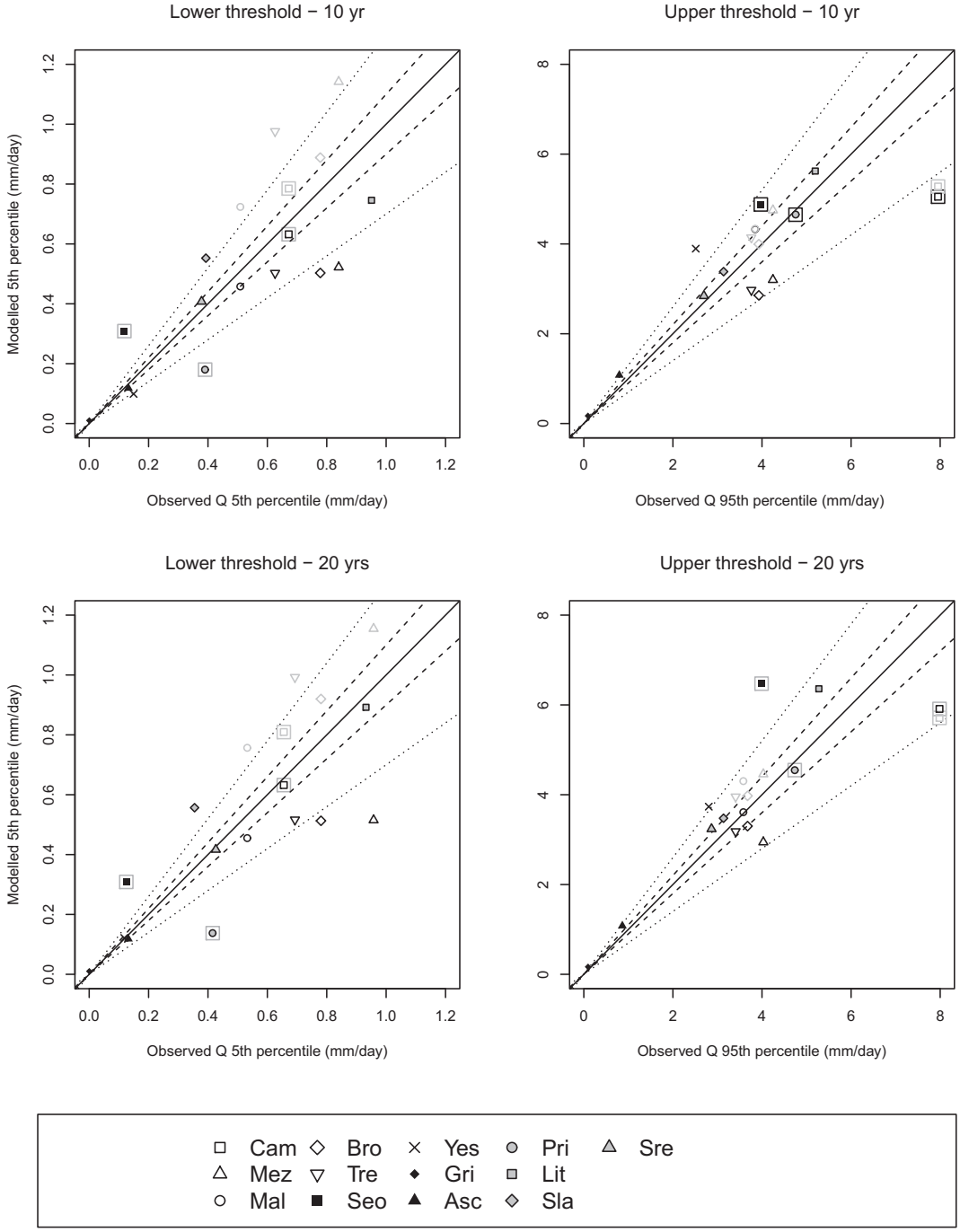


Fig. 11. Low and high flow ERFA thresholds defined from 5th and 95th percentile of observed flow (x-axis) and modelled naturalized flow (y-axis, LisQnat in black and HyperNat in grey). Large grey squares indicate reference sites. The continuous black line indicates the 1:1 relationship. Dashed lines indicate a deviation of ± 10% from the 1:1 relationship. Dotted lines indicate a deviation of ± 30% from the 1:1 relationship.

not change substantially, whereas skewness decreased and kurtosis increased (Figs. in SI). Observed flow duration curves (Figs in the SI) changed only slightly, generally showing an increase in the medium to low flow percentiles. However, in 83% of stations, percentile changes were <10%. Flow magnitude changes were more marked in the Ebro basin, especially at Seo, whereas in the Adige and the Sava basins observed flow magnitude did not change between the two periods. Similarly, model performances in simulating flow did not change significantly for the two periods. Among the different models, Lisflood was the most sensitive to the period length, i.e. estimated changes in flow were larger for Lisflood than for other models.

The sensitivity to period length of modelled indicators in comparison with those obtained from observed flow could be summarized by changes in indicator RMSE and Pearson's R2 for the data ensemble when passing from ten to 20 years. Fig. 10 shows the relative change of RMSE, as ($RMSE_{10}$ − $RMSE_{20}$) / $RMSE_{20}$, versus the change of R2 (x-axis; $R2_{20\ yrs}$ − $R2_{10\ yrs}$) with respect to the two periods. Indicators whose estimation improved (reduction of RMSE and increase in R2) using the 20 years period fell in the positive (upper-right) panel. Conversely, indicators whose estimation worsened fell in the negative (bottom-left) panel. Changes within a 10% (hatched area) could be considered non-significant, especially given the limited number of data entries (26 station/model combinations). Most indicators were sensitive to the length of estimation period (i.e. fell outside the hatched grey box area), and, with some exceptions, the longer analysis period improved indicator estimation both in terms of accuracy and correlation. The largest improvements occurred for WRI Day50, winter and autumn flow, Mag7 AR1, MFRI med_low and med_seq (i.e., low flow conditions and occurrence). All five duration indicators considered in the analysis fell outside the ±10% change, indicating that this flow regime characteristic is the most sensitive to the length of the analysis period, followed by frequency (three indicators out of five), and rate of change (two cases out of four). For timing and magnitude, about 60% of indicators fell within the ±10% area, i.e. these flow regime characteristics were the least sensitive to the length of the analysis period.

3.3. Impact of modelling naturalized flow

Results of Sections 3.1 and 3.2 showed that significant uncertainty affects modelled indicators, and that the majority of indicators are sensitive to the analysis period length. The following analysis aimed at identifying how the uncertainty in modelled flow regime may affect assessing flow hydrologic alteration and decision making.

The ERFA framework (Laizé et al., 2014) was used as an example for the assessment of hydrologic alteration and associated flow-ecological status. ERFA is particularly sensitive to reference flow conditions because six out the 16 MFRIs depend on the setting of high and low flow thresholds (5th and 95th percentile of natural flow). Therefore, we analyzed how these two thresholds were set when using modelled (LisQNat or HyperNat) or observed flow (Fig. 11) for a period of reference conditions of ten (upper panels) or 20 years (lower panels). The comparison indicates two different things. At the reference sites (highlighted in Fig. 11), we would expect that thresholds estimated with the different methods would be similar, as the observed flow at these sites is unaltered. Thus, deviations from the 1:1 line can be attributed to modelling errors. Modelling low and high flow thresholds at the three reference sites considered for this analysis (Evrotas stations were excluded due to lack of observed data) was particularly uncertain: only in one case for the lower threshold (Cam by LisQnat, Fig. 11a and c) and one case for upper threshold (Pri by LisQnat, Fig. 11b and d), modelled thresholds differed from the observed one by <10%. In most cases, thresholds in reference sites set by LisQNat or HyperNat differed by more than ±30% from those derived from observed flow. Conversely, in all other stations, deviations between naturalized flow and observed flow result from the combined effect of alterations of the flow regime and the unknown modelling error. For example, at Mez the large deviation in the lower threshold can reflect the high anthropogenic pressures; unfortunately, the two modelled naturalized thresholds available for this site indicate opposite deviations from observed flow. In general, the scatter in Fig. 11 for altered sites indicate high uncertainty in setting the ERFA thresholds. Enlarging the period of analysis to 20 years did not significantly improve the threshold estimation (Fig. 11).

The choice of thresholds and reference conditions (Fig. 11), together with high uncertainty in MFRIs (Table 4), has important repercussions on assessing ecological risk. Fig. 12 presents the ERFA scores obtained by comparing MFRIs calculated with combinations of (i) different methods to estimate current conditions (modelled or observed flow) with (ii) different reference regimes, i.e. either modelled naturalized flow (LisQNat or HyperNat) or observed flow in reference sites. The analysis was performed for ten (upper panel) and 20 years (lower panel). Considerable uncertainty existed in ecological risk assessment, with most stations being judged at either low, medium, or high ecological risk depending on which combination of methods was used to assess current and reference flow regime. In other words, ERFA scores were inconsistent between models and periods. Stations in the Ebro river basin show the largest ERFA scores among the three basins. This suggests severe flow alteration, which is consistent with the large degree of damming in this basin. In contrast, stations in the Sava river basin show the smallest degree of flow alteration especially for the 20-years period. In the Sava most dams are located in the upstream areas; this may result in lower ecological impact in the downstream stations. However, the inconsistencies highlighted in Fig. 12 imply that it would be difficult for a decision maker to correctly identify highly altered sites.

4. Discussion

Differences in modelled flow simulations at all temporal scales were expected given the heterogeneity in input data, particularly precipitation, temperature and evapotranspiration (Schneider et al., 2013), model structure, and calibration approaches. Despite these diversities, model performances were comparable, and locally calibrated models did not outperform large scale models consistently (Fig. 2). These results are comparable to Caldwell et al. (2015) and confirm that more complex, fine-scale models may not outperform regional-scale models. In addition, the expected spatial scale response, whereby better simulations would occur for local models in the smaller catchments, was not proven in these stations. Considerable simulation errors occurred at small and large spatial scales and no spatial pattern in model simulation pitfalls was identified.

Uncertainty in flow duration curve, particularly at low and medium flow, may on one hand reflect calibration procedures that generally aim at matching high flows, but may also indicate limited capacity to capture existing anthropogenic alterations of flow. For example, at Yes and Gri in the Ebro (Fig. 4), observed flows are clearly impacted by dam regulations and show an abrupt decrease at lower flows, which is not reflected in the model simulations. In the Evrotas stations (Fig. 5) models overestimated flow magnitude. Probably models did not include, or underestimated, water abstractions at these gauging stations, though observed flow was limited to a few continuous months, during a period of pronounced water scarcity (2007 drought event). However, in the Evrotas the observation period was short and characterized by higher than usual water abstractions for irrigation (Skoulikidis et al., 2011) thus observed duration curves are also uncertain. Similarly, models overestimated flow magnitude at medium and low flow at Sla in the Sava (Fig. 6). Irrigation, drainage and land reclamation for flood protection can be concurrent reasons explaining the lower observed flows (ISRBC, 2009, 2014). Hydropeaking impact was less apparent in the flow duration curves: for example, at Mez (Fig. 3), i.e. the Adige gauging station that is heavily impacted by hydropeaking, the higher medium flow in observation flow duration curve than modelled ones may be attributed to an impact of water releases. Conversely, the

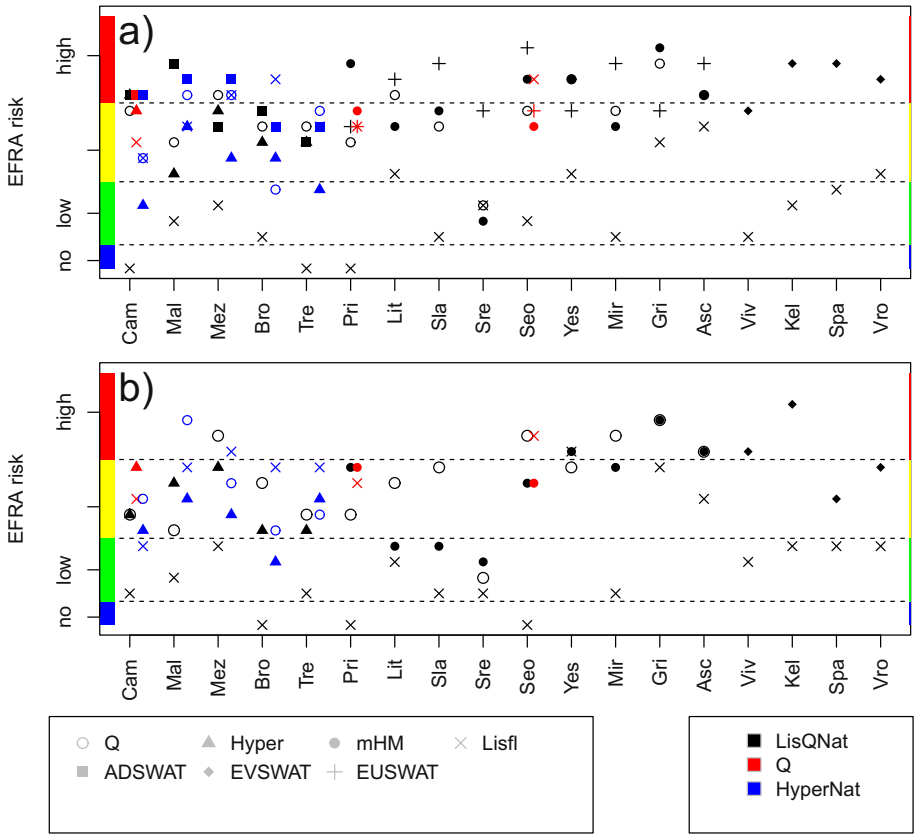


Fig. 12. ERFA risk scores (Laizé et al., 2014) as assessed for a period of (a) ten or (b) 20 years. Symbols (grey legend) indicate the method used to assess current flow regime. Symbol colors indicate the reference flow regime method (i.e. modelled naturalized flow or observed flow). Color code on the y-axis indicate the ERFA risk class (blue = no risk, green = low risk, yellow = high risk, and red = high risk). Q = observed flow.

underestimation of low and medium flows at reference sites Pri in the Sava (Fig. 6) and Kel in the Evrotas (Fig. 5) cannot be ascribed to anthropogenic impact but most probably to specific hydrogeological characteristics of the catchment that models did not fully considered (e.g., the conspicuous presence of karstic areas in the Sava and Evrotas basins; Kostić et al., 2016; Gamvroudis et al., 2015).

Uncertainties in modelled flow regime indicators (Table 4, Figs. 7–9) mirror those reported for observed flow (e.g. Westerberg and McMillan, 2015; Westerberg et al., 2016) in that while estimation of long-term average indicators is generally reliable, characterization of frequency, duration and rate of change is much more elusive. It is also important to note that in this study we considered observed flow as the reference against which to evaluate model output results; however, observed flow and observed flow regime indicators suffer from considerable uncertainty as well (Westerberg and McMillan, 2015; Westerberg et al., 2016). Indicators uncertainty varied not only depending on flow characteristic but also on how the indicators were mathematically defined. WRIs showed to be quite robust, although Day50, summer and autumn flow showed significant uncertainty. Magnitude indicators of Mag7 showed increasing uncertainty in going from mean flow to Kurtosis, confirming problems in modelling flow variability and peaks. Timing and rate of change Mag7 indicators were highly uncertain. Among the MFRIs, median flow indicators were more robust than monthly IQR (measuring variability of flow). Magnitude and duration MFRIs, which are calculated on the basis of set thresholds, were among the most uncertain indicators. In agreement with Westerberg and McMillan (2015), our results indicate that the mathematical formulation of any indicator has an impact on the propagation of uncertainty from the original data (observed flow or modelled output) to the indicator. In particular, uncertainty is large for indicators that are dependent on thresholds, such as, for example, the magnitude MFRIs in comparison to, for example, flow percentiles (Table 4).

The analysis of the period length showed that when using models to estimate indicators a period of at least 15 to 20 years should be used to better capture flow regime. These results are in agreement with the conclusions that Kennard et al. (2010) drew on indicators derived from observed flow: indicator accuracy generally improved when enlarging the analysis period from ten to 20 years. Hence, also from this point of view modelled indicator uncertainty reflects that of indicators obtained from observed flow.

The current analysis revealed serious shortcomings in 'off-the-shelf' models to capture flow regime, especially flow timing, duration, rate of change, and, to a lesser extent, magnitude and variability of low flows. We expected uncertainties in indicators derived from 'off-the-shelf' models to be larger than what could be achieved when using locally calibrated models. Local calibration of streamflow alone however did not guarantee better model performance in estimating flow regime indicators. For example, ADSWAT and HYPER were both locally calibrated at Bro and Tre stations. ADSWAT and HYPER flow regime indicators at these stations, however, were not consistently better than those derived from Lisflood (see figures in SI). Spearman's rank correlation coefficients ρ (Table 4) indicate that the mean error in modelling daily flow had an impact on the estimation of most magnitude and timing indicators, however only for a few indicators of frequency, duration, and rate of change the estimation error was monotonically related to streamflow modelling error. This confirms findings of Caldwell et al. (2015) that local calibration of streamflow was not sufficient to characterize the full spectrum of flow regime characteristics. Instead, improvements could be expected when calibration would consider flow characteristics of interest (e.g. Shrestha et al., 2014; Vis et al., 2015; Zhang et al., 2016).

However, 'off-the-shelf' models could still be useful, especially when magnitude of medium to high flow and long-term averages are of interest. The calibration procedure and purpose for which an existing model was developed can guide a decision maker regarding what indicator could be reliably estimated with an already available model. For example, if high flows were the main focus of modelling in the first place, we can expect that high flow estimation will be more reliable. Instead, if frequency, duration, and rate of change were the most important flow regime characteristics, then 'off-the-shelf' models will probably not be appropriate, and a tailor-made hydrological modelling would be needed to provide reliable information. This might simply mean to re-calibrate an existing model to the flow regime characteristic of interest by redefining appropriate calibration objective functions (e.g. Zhang et al., 2016). Future research should aim at verifying to which extent flow regime calibration could reduce indicators uncertainty in the stations considered in this study.

In agreement with other studies (Murphy et al., 2013; Westerberg and McMillan, 2015; Westerberg et al., 2016; Vis et al., 2015; Eddy et al., 2017), our analysis showed that indicator uncertainty varies greatly, which should be accounted for when natural resource management depends on flow regime indicators. Failure to estimate indicators can lead to errors in flow regime classification (Eddy et al., 2017), characterization of hydrological disturbance and projected flow regime in scenario analysis and can affect detection of ecological response if uncertainties are larger than signal change (Kennard et al., 2010). Uncertain or inaccurate model outputs may lead to wrong management decisions, e.g. in the identification of water availability and allocation, and assessment of environmental flow requirements. Differences in mean annual streamflow were in the order of 30% for most stations in this study. Uncertainty, however, increased when assessing seasonality and flow variability, which has important implications for determination of water availability throughout the year. Similarly, assessment of flow regime duration and rate of change were unreliable. This particularly may affect our capacity to detect the impact of dam regulation and hydropeaking on the freshwater ecosystems.

Flow-ecological frameworks currently used to assess freshwater system management (Poff et al., 2010; Laizé et al., 2014; Richter et al., 2012) require an accurate assessment of deviations of flow regime from reference conditions. The example of ERFA scoring (Fig. 12) provides evidence of how uncertainty in reference flow regime could affect management of hydrologic alterations. When the reference regime was set with LisQNat, Lisflood scores (black crosses in Fig. 12) were among the lowest, generally indicating no or low risk. Similarly, albeit to a lower degree, HYPERstream scores were generally low when reference regime was set with HyperNat (blue triangles in Fig. 12). This suggests that naturalized flow simulation does not differ sufficiently from modelled current conditions to capture hydrological alterations effectively. Furthermore, when using observed flow as reference regime (red symbols), reference sites were always classified as being at medium or high risk, i.e. models failed to identify unaltered hydrological conditions at these sites. Together with shortcomings shown in the analysis of flow duration curves, this highlights the need for improving simulation of anthropogenic alterations in hydrological models (Bellin et al., 2016).

The lack of a reliable reference flow regime can become a serious bottleneck in applying flow-ecological assessments. In this study, the naturalized flow simulations could not simulate flow regime at reference sites; given their importance for effective management, a larger collective effort should be made to provide reliable reference flow regime conditions. This may be achieved by calibrating models in natural sites for a comprehensive set of flow regime indicators, or with approaches that do not necessarily use hydrologic models (e.g., Peñas et al., 2016), for example using hydrologic similarity to expand results in ungauged reaches (Westerberg et al., 2016; Yang et al., 2016). Examples for the U.S. (e.g. Carlisle et al., 2010) show that such efforts are possible even at large scale and may indeed prove necessary to enable effective freshwater ecosystem management.

5. Conclusions

Hydrological models can be very useful in simulating flow in ungauged reaches, particularly for projecting potential changes to occur under future scenarios (Casper et al., 2012). Developing hydrologic models is, however, resource demanding. On the other hand, existing global, continental and local hydrological models may be available to decision makers. In this study, we assessed the potential of six 'off-the-shelf' models, i.e. models that were not specifically developed for assessing flow-ecological indicators, to conduct flow-ecological assessments. The models were generally able to simulate flow regime magnitude, especially at medium and high flows, but performed poorly in the prediction of other flow regime characteristics. In particular, flow timing, duration and rate of change were subject to large uncertainties. This has potentially far-reaching consequences in estimating ecological conditions in freshwater systems and consequential basin management, particularly in ungauged reaches, where model performance cannot be assessed against observed flow.

'Off-the-shelf' models should thus preferably be used for ecological risk assessment if flow magnitude is the most important aspect of flow regime alteration. If other characteristics of flow regime were of interest, then tailor-made hydrological models should be used instead, or existing models should be recalibrated with appropriate calibration functions (e.g. Zhang et al., 2016) to be made 'fit-for-purpose'. Furthermore, in agreement with indicator assessment based on observed data (Kennard et al., 2010), it is recommended that flow regime is assessed from model outputs of at least 15, or even better 20 years, especially when occurrence of rare events is of interest (e.g. extreme flows, duration of low flow, etc.).

In the light of the uncertainty of flow regime indicators assessed with 'off-the-shelf' models, flow-ecological assessments should consider the variable uncertainty that affects different indicators and flow regime characteristics. Hence, prevalent weight should be given to those indicators that are more robust to estimate, such as long-term mean annual indicators, compared to indicators that are highly uncertain, such as those defined in relation to set thresholds. Correspondingly, the ERFA system, which bases six out of 16 indicators on fixed thresholds, proved highly sensitive to inaccuracies in indicators estimation in this study.

Finally, flow-ecological assessments are very sensitive to reference conditions. In this study, naturalized flow simulations could not capture flow regime characteristics of reference sites satisfactorily. The lack of reliable reference flow regime might become the bottleneck in applying flow-ecological assessments in water basin management. Therefore, a collective effort is recommended to (i) provide reference conditions in the study areas and ideally at European scale, and (ii) advance the simulation of anthropogenic alterations in hydrological models in order to improve the characterization of current and natural flow conditions.

Acknowledgments

This work has received funding from the European Union's Seventh Programme for research, technological development and demonstration under grant agreement No. 603629 – project "Globaqua". G.C. acknowledges the support of the Stiftungsfonds für Umweltökonomie und Nachhaltigkeit GmbH (SUN).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2017.09.295.

References

Abbaspour, K., 2015. SWAT-CUP 2012: SWAT Calibration and Uncertainty Programs: A User Manual. Department of Systems Analysis. Integrated Assessment and Modelling (SIAM), Eawag, Swiss Federal Institute of Aquatic Science and Technology, Duebendorf, Switzerland.

Archfield, S.A., Kennen, J.G., Carlisle, D.M., Wolock, D.M., 2014. An objective and parsimonious approach for classifying natural flow regimes at a continental scale. River Res. Appl. 30, 1166–1183.

Arnold, J.G., Kiniry, J.R., Srinivasan, R., William, J.R., Haney, E.B., Neitsch, S.L., 2012. Soil and Water Assessment Tool Input/Output Documentation: Version 2012. Texas Water Resources Institute.

Bejarano, M.D., Marchamalo, M., Garcia de Jalón, D., González del Tánago, M., 2010. Flow regime patterns and their controlling factors in the Ebro basin (Spain). J. Hydrol. 385, 323–335.

Bellin, A., Majone, B., Cainelli, O., Alberici, D., Villa, F., 2016. A continuous coupled hydrological and water resources management model. Environ. Model. Softw. 75, 176–192.

Caldwell, P.V., Kennen, J.G., Sun, G., Kiang, J.E., Butcher, J.B., Eddy, M.C., Hay, L.E., LaFontaine, J.H., Hain, E.F., Nelson, S.A.C., McNulty, S.G., 2015. A comparison of hydrologic models for ecological flows and water availability. Ecohydrology 8:1525–1546. https://doi.org/10.1002/eco.1602.

Carlisle, D.M., Falcone, J., Wolock, D.M., Meador, M.R., Norris, R.H., 2010. Predicting the natural flow regime: models for assessing hydrological alteration in streams. River Res. Appl. 26, 118–136.

Casper, M.C., Grigoryan, G., Gronz, O., Gutjahr, O., Heinemann, G., Ley, R., Rock, A., 2012. Analysis of projected hydrological behavior of catchments based on signature indices. Hydrol. Earth Syst. Sci. 16, 409–421.

Chiogna, G., Majone, B., Paoli, K.C., Diamantini, E., Stella, E., Mallucci, S., Lencioni, V., Zandonai, F., Bellin, A., 2016. A review of hydrological and chemical stressors in the Adige catchment and its ecological status. Sci. Total Environ. 540, 429–443.

Eddy, M.C., Phelan, J., Patterson, L., Allen, J., Pearsall, S., 2017. Evaluating flow metric-based stream classification systems to support the determination of ecological flows in North Carolina. J. Am. Water Resour. Assoc. 53, 30–41.

European Commission (EC), 2000. Establishing a framework for community action in the field of water policy. Directive 2000/60/EC of the European Parliament & of the Council of 23 October 2000, Official Journal of the European Communities, Brussels http://ec.europa.eu/environment/water/water-framework/index en.html.

European Environment Agency (EEA), 2012. CORINE Land Cover CLC2012. URL. http://land.copernicus.eu/, Accessed date: 20 November 2016.

Gamvroudis, C., 2016. Integrated Modeling Framework of Hydrologic, Water Quality and Sediment Transport in Temporary River Basins. Thesis. School of Environmental Engineering - Technical University of Crete in Greece.

Gamvroudis, C., Nikolaidis, N.P., Tzoraki, O., Papadoulakis, V., Karalemas, N., 2015. Water and sediment transport modeling of a large temporary river basin in Greece. Sci. Total Environ. 508, 354–365.

Gamvroudis, C., Dokou, Z., Nikolaidis, N.P., Karatzas, G.P., 2017. Impacts of surface and groundwater variability response to future climate change scenarios in a large Mediterranean watershed. Environ. Earth Sci. 76:385. https://doi.org/10.1007/s12665-017-6721-7.

Grusson, Y., Sun, X., Gascoin, S., Sauvage, S., Raghavan, S., Anctil, F., et al., 2015. Assessing the capability of the SWAT model to simulate snow, snow melt and streamflow dynamics over an alpine watershed. J. Hydrol. 531, 574–588.

Hellenic Statistical Authority (El.STAT.), 2011. URL. http://www.statistics.gr/portal/page/portal/ESYE/PAGE-census2011, Accessed date: 20 November 2014.

International Sava River Basin Commission (ISRBC), 2009. The Sava River Basin Analysis Report. ISBRC - International Sava River Basin Commission, Zagreb (289 pp.).

International Sava River Basin Commission (ISRBC), 2014. Sava River Basin Management Plan. ISBRC - International Sava River Basin Commission, Zagreb (240 pp.).

Kennard, M.J., Mackay, S.J., Pusey, B.J., Olden, J.D., Marsh, N., 2010. Quantifying uncertainty in estimation of hydrologic metrics for ecohydrological studies. River Res. Appl. 26, 137–156.

Kostić, S., Stojković, M., Prohaska, S., 2016. Hydrological flow rate estimation using artificial neural networks: model development and potential applications. Appl. Math. Comput. 291, 373–385.

Kumar, R., Samaniego, L., Attinger, S., 2010. The effects of spatial discretization and model parameterization on the prediction of extreme runoff characteristics. J. Hydrol. 392, 54–69.

Kumar, R., Samaniego, L., Attinger, S., 2013. Implications of distributed hydrologic model parameterization on water fluxes at multiple scales and locations. Water Resour. Res. 49. https://doi.org/10.1029/2012WR012195.

Laiti, L., Mallucci, S., Piccolroaz, S., Bellin, A., Zardi, D., Fiori, A., Nikulin, G., Majone, B., 2017. The use of hydrological modeling for testing the spatio-temporal coherence of high-resolution gridded precipitation and temperature datasets in the Alpine region. Water Resour. Res. (under review).

Laizé, C.L.R., Acreman, M.C., Schneider, C., Dunbar, M.J., Houghton-Carr, H.A., Flörke, M., Hannah, D.M., 2014. Projected flow alteration and ecological risk for pan-European rivers. River Res. Appl. 30:299–314. https://doi.org/10.1002/rra.2645.

Levi, L., Jaramillo, F., Andričević, R., Destouni, G., 2015. Hydroclimatic changes and drivers in the Sava River Catchment and comparison with Swedish catchments. Ambio 44: 624–634. https://doi.org/10.1007/s13280-015-0641-0.

Lutz, S.R., Mallucci, S., Diamantini, E., Majone, B., Bellin, A., Merz, R., 2016. Hydroclimatic and water quality trends across three Mediterranean river basins. Sci. Total Environ. 571, 1392–1406.

Majone, B., Bertagnoli, A., Bellin, A., 2010. A non-linear runoff generation model in small Alpine catchments. J. Hydrol. 385, 300–312.

Majone, B., Bovolo, C.I., Bellin, A., Blenkinsop, S., Fowler, H.J., 2012. Modeling the impacts of future climate change on water resources for the Gállego river basin (Spain). Water Resour. Res. 48, W01512. https://doi.org/10.1029/2011WR010985.

Majone, B., Villa, F., Deidda, R., Bellin, A., 2016. Impact of climate change and water use policies on hydropower potential in the south-eastern Alpine region. Sci. Total Environ. 543 (B), 965–980.

Malagó, A., Pagliero, L., Bouraoui, F., Franchini, M., 2015. Comparing calibrated parameter sets of the SWAT model for the Scandinavian and Iberian peninsulas. Hydrol. Sci. J. 60, 949–967.

Malagó, A., Bouraoui, F., Vigiak, O., Grizzetti, B., Pastori, M., 2017. Modelling water and nutrient fluxes in the Danube River Basin with SWAT. Sci. Total Environ. 603-604, 196–218.

McMillan, H., Krueger, T., Freer, J., 2012. Benchmarking observational uncertainties for hydrology: rainfall, river discharge and water quality. Hydrol. Process. 26, 4078–4111.

Michel, C., Andreassian, V., Perrin, C., 2005. Soil Conservation Service Curve Number method: how to mend a wrong soil moisture accounting procedure? Water Resour. Res. 41:W02,011. https://doi.org/10.1029/2004WR003191.

Moriasi, D.N., Arnold, J.G., van Liew, M.W., Bingner, R.L., Harmel, R.D., Veith, T.L., 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Trans. ASABE 50, 885–900.

Murphy, J.C., Knight, R.R., Wolfe, W.J., Gain, W.S., 2013. Predicting ecological flow regime at ungauged sites: a comparison of methods. River Res. Appl. 29, 660–669.

Navarro-Ortega, A., Acuña, V., Bellin, A., Burek, P., Cassiani, G., Choukr-Allah, R., Dolédec, S., Elosegi, A., Ferrari, F., Ginebreda, A., Grathwohl, P., Jones, C., Rault, P.K., Kok, K., Koundouri, P., Ludwig, R.P., Merz, R., Milacic, R., Muñoz, I., Nikulin, G., Paniconi, C., Paunović, M., Petrovic, M., Sabater, L., Sabater, S., Skoulikidis, N.T., Slob, A., Teutsch, G., Voulvoulis, N., Barceló, D., 2015. Managing the effects of multiple stressors on aquatic ecosystems under water scarcity. The GLOBAQUA project. Sci. Total Environ. 503 (504), 3–9.

Neitsch, S.L., Arnold, J.G., Kiniry, J.R., Williams, J.R., 2011. Soil and Water assessment tool – theoretical documentation. Texas Water Resources Institute Technical Report 406. Texas A&M University System, College Station, Texas, U.S. http://swat.tamu.edu/media/99192/swat2009-theory.pdf, Accessed date: 17 December 2014.

Peñas, F.J., Barquín, J., Álvarez, C., 2016. Assessing hydrologic alteration: evaluation of different alternatives according to data availability. Ecol. Indic. 60, 470–482.

Piccolroaz, S., Majone, B., Palmieri, F., Cassiani, G., Bellin, A., 2015. On the use of spatially distributed, time-lapse microgravity surveys to inform hydrological modeling. Water Resour. Res. 51, 7270–7288.

Pistocchi, A., Udias, A., Grizzetti, B., Gelati, E., Kondouri, P., Ludwig, R., Papandreou, A., Souliotis, I., 2016. An integrated assessment framework for the analysis of multiple pressures in aquatic ecosystems and the appraisal of management options. Sci. Total Environ. 575:1477–1488. https://doi.org/10.1016/j.scitotenv.2016.10.020.

Poff, N.L., Allan, J.D., Bain, M.B., Karr, J.R., Prestegaard, K.L., Richter, B.D., Sparks, R.E., Stromberg, J.C., 1997. The natural flow regime: a paradigm for river conservation and restoration. Bioscience 47:769–784. https://doi.org/10.2307/1313099.

Poff, N.L., Richter, B.D., Arthington, A.H., et al., 2010. The ecological limits of hydrologic alteration (ELOHA): a new framework for developing regional environmental flow standards. Freshw. Biol. 55:147–170. https://doi.org/10.111/j.1365-2427.2009.02204.x.

Rakovec, O., Kumar, R., Attinger, S., Samaniego, L., 2016a. Improving the realism of hydrologic model functioning through multivariate parameter estimation. Water Resour. Res. 52, 7779–7792.

Rakovec, O., Kumar, R., Mai, J., Cuntz, M., Thober, S., Zink, M., Attinger, S., Schäfer, D., Schrön, M., Samaniego, L., 2016b. Multiscale and multivariate evaluation of water fluxes and states over European river basins. J. Hydrometeorol. 17, 287–307.

Richter, B., Baumgartner, J., Robert, W., Braun, D., 1997. How much water does a river need? Freshw. Biol. 37, 231–249.

Richter, B.D., Davis, M.M., Apse, C., Konrad, C., 2012. A presumptive standard for environmental flow protection. River Res. Appl. 28:1312–1321. https://doi.org/10.1002/rra.1511.

Samaniego, L., Kumar, R., Attinger, S., 2010. Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale. Water Resour. Res. 46, W05523. https://doi.org/10.1029/2008WR007327.

Schneider, C., Laizé, C.L.R., Acreman, M.C., Flörke, M., 2013. How will climate change modify river flow regimes in Europe? Hydrol. Earth Syst. Sci. 17, 325–339.

Shrestha, R.R., Peters, D.L., Schnorbus, M.A., 2014. Evaluating the ability of a hydrologic model to replicate hydro-ecologically relevant indicators. Hydrol. Process. 28, 4294–4310.

Skoulikidis, N., Vardakas, L., Karaouzas, I., Economou, A., Dimitriou, E., Zogaris, S., 2011. Assessing water stress in Mediterranean lotic systems: insights from an artificially intermittent river in Greece. Aquatic Sciences 73, 581–597 Special Issue: Recent Perspectives on Temporary River Ecology.

Swirepik, J.L., Burns, I.C., Dyer, F.J., Neave, I.A., O'Brien, M.G., Pryde, G.M., Thompson, R.M., 2016. Establishing environmental water requirements for the Murray-Darling basin, Australia's largest developed river system. River Res. Appl. 32, 1153–1165.

Thompson, J., Archfield, S., 2015. The EflowStats R Package. USGS http://abouthydrology.blogspot.it/2013/12/usgs-r-packages-for-hydrology.html, Accessed date: 19 January 2017.

Tolson, B.A., Shoemaker, C.A., 2007. Dynamically dimensioned search algorithm for computationally efficient watershed model calibration. Water Resour. Res. 43, W01413. https://doi.org/10.1029/2005WR004723.

Tuo, Y., Duan, Z., Disse, M., Chiogna, G., 2016. Evaluation of precipitation input for SWAT modeling in Alpine catchment: a case study in the Adige river basin (Italy). Sci. Total Environ. 573, 66–82.

Tzoraki, O., Papadoulakis, V., Christodoulou, A., Vozinaki, E., Karalemas, N., Gamvroudis, C., Nikolaidis, N.P., 2011. Hydrologic modelling of a complex hydrogeologic basin: Evrotas River Basin. In: Lambrakis, N., Stournaras, G., Katsanou, K. (Eds.), Advances in the Research of Aquatic Environment. Springer, Berlin Heidelberg, pp. 179–186.

Van der Knijff, J.M., Younis, J., de Roo, A.P.J., 2008. Lisflood: a GIS-based distributed model for river basin scale water balance and flood simulation. Int. J. Geogr. Inf. Sci. 24. https://doi.org/10.1080/13658810802549154.

Vis, M., Knight, R., Pool, S., Wolfe, W., Seibert, J., 2015. Model calibration criteria for estimating ecological flow characteristics. Water 7:2358–2381. https://doi.org/10.3390/w7052358.

Vu, M.T., Raghavan, S.V., Liong, S.Y., 2012. SWAT use of gridded observations for simulating runoff – a Vietnam river basin study. Hydrol. Earth Syst. Sci. 16, 2801–2811.

Westerberg, I.K., McMillan, H.K., 2015. Uncertainty in hydrological signatures. Hydrol. Earth Syst. Sci. 19, 3951–3968.

Westerberg, I.K., Wagener, T., Coxon, G., McMillan, H.K., Castelklarin, A., Montanari, A., Freer, J., 2016. Uncertainty in hydrological signatures for gauged and ungauged catchments. |. Water Resour. Res. 52, 1847–1865.

Yang, H.-C., Suen, J.-P., Chou, S.-K., 2016. Estimating the ungauged natural flow regimes for environmental flow management. Water Resour. Manag. 30, 4571–4584.

Zhang, Y., Vaze, J., Chiew, F.H.S., Teng, J., Li, M., 2014. Predicting hydrological signatures in ungauged catchments using spatial interpolation, index model, and rainfall-runoff modelling. J. Hydrol. 517, 936–948.

Zhang, Y., Shao, Q., Zhang, S., Zhai, X., She, D., 2016. Multi-metric calibration of hydrological model to capture overall flow regimes. J. Hydrol. 539, 525–538.

Zolezzi, G., Bellin, A., Bruno, M.C., Maiolini, B., Siviglia, A., 2009. Assessing hydrological alterations at multiple temporal scales: Adige River, Italy. Water Resour. Res. 45, W12421. https://doi.org/10.1029/2008WR007266.