Doctoral Dissertation
Doctoral Program in Civil and Environmental Engineering (35^{th}cycle)

River flood characterization in North-Western Italy: statistical analysis and regional rainfall-runoff modelling

By

Matteo Pesce

******

Supervisor(s):
Prof. Alberto Viglione, Supervisor
Prof. Jost von Hardenberg, Co-Supervisor

Doctoral Examination Committee:
Prof. Elena Toth, Referee, Università di Bologna
Prof. Mojca Šraj, Referee, University of Ljubljana
Prof. Daniele Ganora, Politecnico di Torino
Prof. Marta Tuninetti, Politecnico di Torino
Prof. Marco Borga, Università di Padova

Politecnico di Torino
2024

Declaration

I hereby declare that, the contents and organization of this dissertation constitute my own original work and does not compromise in any way the rights of third parties, including those relating to the security of personal data.

Matteo Pesce
2024

* This dissertation is presented in partial fulfillment of the requirements for **Ph.D. degree** in the Graduate School of Politecnico di Torino (ScuDo).

Acknowledgements

I would like to thank my Supervisor Prof. Alberto Viglione for contributing to the realization of this Thesis and giving me the opportunity to work with him in the past years. Despite some difficulties due to the pandemic period, he encouraged me with useful advice along the journey.

Thanks to my Co-Supervisor, Prof. Jost von Hardenberg, for the suggestions he gave me especially in the first part of this work.

I would like to express my gratitude to Prof. Elena Toth and Prof. Mojca Šraj for their reviews, which allowed to improve this manuscript.

I would like also to thank Prof. Ralf Merz and the Department Catchment Hydrology of UFZ for the warm welcome and hospitality and in particular Larisa Tarasova for the supervision during my stay in Germany.

Last but not least, I really thank my family and my friends for the support, during this period of my life.

Abstract

This research work deals with the identification of flood event drivers in North-Western Italy, both through a statistical analysis of streamflow annual maxima and the application of a conceptual semi-distributed hydrological model. The work provides different perspectives on how to study extreme events at the regional scale, focusing on a set of non-regulated catchments located in North-Western Italy, within the Alpine region.

After a preliminary analysis of the study area and the main data used, in Chapter 2 a correlation analysis between climate and flood indices at different temporal scales is widely discussed. In particular, annual maximum flows, extracted from the daily streamflow series, are correlated with the standard Climate Change Indices (ETCCDI) of precipitation and temperature, which are commonly used in climate research. A temporal correlation analysis, performed in order to identify which climate drivers better explain the interannual variability of floods, is followed by a spatial correlation analysis of temporal trends of the variables, with the aim of capturing the influence of climate (decadal) variability on the tendency of annual maximum discharges. The results show that, while at the annual timescale floods are highly correlated with indices of precipitation extremes, the tendencies of discharge maxima seem to be better explained by the mean precipitation over the catchment. A following step of the work involves the characterization and classification of different runoff event types over the region, which potentially allows to study how different flood event types regionally change over time.

To this aim, in the second part (Chapter 3-4) a conceptual semi-distributed hydrologic model is calibrated over the study area, first with locally observed discharge data and then regionally, by using the PArameter Set Shuffling (PASS) procedure (Merz et al., 2020), a robust and well documented regionalization procedure that allows to transfer the information contained into locally calibrated parameters and catchment descriptors to the entire domain. In this work, PASS is implemented with a

decision tree machine learning algorithm for the regionalization of model parameters. In particular, the advantage of using snow information in the calibration procedure is further investigated. In addition, a newly developed R package, useful to make the application of the procedure more flexible, is presented, together with examples of application based on a well known comprehensive U.S. hydrologic database that is publicly accessible. It appears from the results that PASS can be efficiently used for the regionalization of model parameters in the study area, by providing consistent relationships among climatic or geomorphological characteristics and model parameters while confirming the effect of reduction of parameter equifinality. The inclusion of snow in the model efficiency function doesn't significantly improve model simulations but provides more consistent results for snow parameters and, overall, less uncertain model simulations.

In Chapter 5, the regionally calibrated model is used for identifying, characterizing and classifying runoff events in the same study region. The aim is to extend the observed dataset in space and time in order to get a timeseries of spatially distributed simulated events spanning 60 years from 1961 to 2020. First, the ability of the model in reproducing observed runoff event characteristics (i.e. runoff coefficient, event duration, event peak time, event peak, event volume) is evaluated by comparing model simulations with observations in gauged sites. Then, regionally distributed runoff event characteristics for the period 1961-2020 are obtained by considering a wider catchment dataset, i.e. the European Catchments and Rivers Network System (ECRINS, 2012), over which the model with regionally calibrated parameters is applied. The results for the gauged catchments show that the model is able to properly capture the spatial pattern of observed runoff characteristics, in particular runoff event peak and volume, with a median Nash-Sutcliffe Efficiency (NSE) greater than 0.5, while the performance for runoff coefficient, duration and peak time is lower. It is worth noting that the value of the runoff coefficient, event peak and event volume is maximum in the southern and northeastern part of the region, in catchments located at medium elevation in the proximity of Alps and Apennines. Consistent results are also obtained by running the model in a distributed mode in ungauged sites. By using several climatic indicators describing different event features of the observed data that are not limited to discharge (i.e., type of inducing event, space-time organization, wetness state of the catchment and spatial interaction of precipitation and soil moisture), the first-order controls of event runoff response are identified in the gauged catchments using the framework presented in Tarasova et al. (2020), and this reveals four distinct clusters (sub-regions) with homogeneous

event type frequency. In particular, cluster 1 mainly consists of lowland catchments where intensity-dominated and local volume-dominated events under dry conditions constitute a relevant quota of total events, suggesting that the main runoff generation mechanism is a local one with possible infiltration excess or event-fed saturation. These mechanisms indicate that convective activity is a very likely phenomenon leading to floods in these catchments. Cluster 2 includes catchments that are located both in lowland and at medium elevation. The fraction of events characterized by the presence of snow is higher compared to the previous cluster and extensive and steady rainfall events, both intensity and volume-dominated, are dominant types for this cluster. This indicates the potential occurrence of orographic slow-moving storms. The third cluster covers a large portion of the Alpine range, from South-West to North, and includes all high elevation catchments strongly impacted by snow processes and large valley catchments characterized by high-elevation zones. Indeed, the majority of events in this cluster is represented by mixture of rainfall and snowmelt and a moderate fraction is given by pure snowmelt events. Finally, cluster 4 includes catchments located all over the region at quite high elevation, both along the Alps and the Apennines. This cluster has some similarities with cluster 2 but the fraction of events impacted by snow processes is much higher and extensive volume-dominated events during dry conditions prevail, suggesting extensive event fed-saturation as a major runoff generation mechanism, with possible event-induced connectivity. By applying the same framework regionally in ungaged sites using simulated events that span the period 1961-2020, we find that the spatial pattern of event type occurrence as obtained by the model is coherent with the event typology from observed discharges and reflects the hydroclimatic conditions of the area. The main differences concern cluster 1, which shows a higher quota of unsteady events (mainly volume-dominated), and cluster 3, in which the fraction of snowmelt events overcomes that of rain-on-snow events. The event types classification allows to better explain the spatial distribution of event characteristics. The highest values of runoff coefficient, event peak and volume are found for catchments of cluster 2 or 4, where rain-on-snow events and orographic slow-moving storms, with an extensive and steady structure, play a role in the runoff generation. Instead, catchments pertaining to cluster 1 and 3, showing lower values of these characteristics, are strongly impacted by either local runoff events that massively depend on the intensity of inter-event evapotranspiration and on soil moisture state, unsteady volume-dominated events and mixture of rainfall and snowmelt events.
Finally, Chapter 6 is dedicated to final Discussion and Conclusions.

This Thesis represents a contribution to the hydrological community by providing insights on the added value of using regionally calibrated distributed hydrological models to describe flood events in a snow-dominated area, compared to a standard statistical analysis of extremes. It is also provided a coded version of an established regionalization procedure, to allow a flexible use of such models for a variety of hydrological regimes. Future research can build upon the time-series of simulated events and the results of event classification to study the possible spatial and temporal correlations among climate variables and specific flood event types.

Contents

List of Figures

List of Tables

Introduction

The hydrological cycle is strongly impacted by the pattern variation of climate variables such as temperature and precipitation. Higher temperatures induce an increase of potential evaporation and, at high elevation, changes in snow dynamics, which affect runoff production.

In the last decades, high elevation areas are given particular attention as their hydrological regimes, being strongly influenced by snow dynamics and changes in glaciated areas, are especially vulnerable to climate change (IPCC, 2019). Over the last century, the Alps have experienced an increase in temperature by a factor of 1.6 higher than the average worldwide temperature increase (Brunetti et al., 2009). In the Alpine region, monthly and extreme runoff is characterized by a strong seasonality, with maximum runoff typically occurring in spring and summer, driven by the snowmelt, and minimum runoff in winter. The interest of the hydrologic community is to understand how this pattern has changed in the past and is going to change in the future. For this reason, the impact of precipitation and temperature patterns on river flows is a well discussed topic and the attention is particularly devoted to floods.

River floods are, indeed, one of the most impacting natural hazards, leading to huge annual average damages in different sectors of the society, which are expected to rapidly increase (IPCC, 2022). Large floods have occurred in Europe in the last decades; among these we distinguish events in Central Europe in 2002, 2013, and 2021 (e.g., Blöschl et al., 2013; Kreienkamp et al., 2021; Ulbrich et al., 2003), winter floods in North-West England in 2009 and 2015/2016 (e.g., Barker et al., 2016; Miller et al., 2012), autumn floods in North-Western Italy such as 1994, 2000, 2016 and 2020 in Piemonte (e.g., Cassardo et al., 2001; Grazzini et al., 2020), 2011 in Liguria (Silvestro et al., 2012; Silvestro et al., 2016) and finally the spring flood of 2023 in Emilia-Romagna, Italy (Arrighi and Domeneghetti, 2024). Based on this

evidence, many studies focus on the detection of past changes in flood hazard. In these studies, Mann-Kendall test is generally adopted to detect changes in the mean annual flood magnitude and frequency (Mediero et al., 2014; Petrow and Merz, 2009; Prosdocimi et al., 2014; Villarini et al., 2011). Among them, Blöschl et al. (2019) analyse the most comprehensive dataset of observations in Europe (Hall et al., 2015) and extract spatial patterns of trends in the annual maximum streamflow for the period 1960-2010. An attribution to possible drivers of floods is also performed by considering as candidates the annual maximum seven-day precipitation, the highest monthly soil moisture and the spring temperature as a proxy for snowmelt and snow-to-rain transition. Many studies consider non-stationary flood frequency analysis for flood change attribution, by modelling distribution parameters with time-varying climatic covariates (e.g., Prosdocimi et al., 2014; Prosdocimi et al., 2015; Šraj et al., 2016; Viglione et al., 2016). Bertola et al. (2020) analyse the differences between small and large flood changes in Europe and Bertola et al. (2021) attribute them to corresponding drivers. Other studies focus on future flood projections, by considering changes in the magnitude of annual extreme flows (Hanus et al., 2021) or flood quantiles (e.g., Alfieri et al., 2015; Rojas et al., 2012). Alfieri et al. (2015) compare an ensemble of European flood projections for different future time periods with flood simulations for an historical period. They analyse the possible interconnections among two possible drivers of change, such as the annual precipitation and the annual maximum daily precipitation, and flood change at the regional scale. Both the annual precipitation amount and the annual maximum daily precipitation are examples of standard climate indices, as defined by the Commission for Climatology/World Climate Research Programme/Technical Commission for Oceanography and Marine Meteorology (CCI/WCRP/JCOMM) Expert Team on Climate Change Detection and Indices (ETCCDI, see e.g., Zhang et al., 2005). ETCCDI indices are widely used in the climate literature to represent temperature and precipitation extremes and they can be applied to study a variety of extreme events such as heavy rain, floods, droughts, heat waves, etc. It is of interest to evaluate whether and which ETCCDI indices are relevant for characterizing and, therefore, predicting flood changes, particularly in a snow-dominated region.

So, the first part of this Thesis is devoted to explore the possible correlations between the annual maximum daily discharges and ETCCDI indices time series at the catchment scale in North-Western Italy (Pesce et al., 2022). This provides an indication of which extreme precipitation and temperature indices could be used as

covariates for estimating annual flood probabilities and their temporal change over this region.

It is well established in literature that the use of hydrological models can be of particular importance for a more comprehensive quantification and characterization of the changes occurring in the hydrological regime, whose drivers are phenomena tipically occurring at large scales, up to the global scale (e.g. climate change). This can be performed taking advantage of distributed hydrological models through the estimation of spatially consistent model parameters that represent different physical processes. The consistency of the parameter values is obtained by finding a functional relationship between the parameters and climatic and geomorphological characteristics of the area over which the model is calibrated, also defined as catchment descriptors. The idea is to obtain parameter sets for each gridded element (more generally hydrologic unit) having the same functional relationship with catchment descriptors, following a process-based approach (Archfield et al., 2015; Clark et al., 2016; Gupta et al., 2014; Mizukami et al., 2017; Paniconi and Putti, 2015). Parameters regionalization techniques also find wide application in the context of runoff prediction in ungauged basins (Blöschl et al., 2013; Merz and Blöschl, 2004; Parajka et al., 2013; Seibert, 1999; Troch et al., 2003). Among the different techniques, an innovative one is the PArameter Set Shuffling (PASS) procedure, proposed by Merz et al. (2020), based on the concept of using machine learning algorithms to derive relationships linking locally calibrated parameters and catchment descriptors, which can be used to predict spatially distributed parameters. When dealing with the calibration of hydrological models, one main topic is the choice of the objective function and many studies demonstrate the adding value, in terms of improving the prediction and reducing the uncertainty, of multiple objective calibration i.e. constraining the model to some additional hydrological variables other than runoff, such as snow cover or soil moisture (Parajka and Blöschl, 2008; Tong et al., 2021).

On this premise, in the second part, a conceptual semi-distributed hydrologic model is regionally calibrated on the same study area, by taking advantage of the PASS method. The method is implemented with a decision-tree algorithm that transfers parameter values to the entire domain, by using the information contained in local parameters and catchment descriptors. A multi-objective calibration of the model is used by considering MODIS satellites products as snow cover information. The main assumption is that including snow information in the calibration procedure could improve the representation of the hydrological processes in an Alpine setting,

leading to improved streamflow simulations. A newly developed R package fostering the flexibility of the procedure is also presented and examples are provided with openly accessible, well known U.S. hydrologic database.

In the final part of this Dissertation, the regionally calibrated model is used for identifying, characterizing and classifying runoff events in North-Western Italy. The identification of runoff events is performed by using the algorithm proposed by Giani et al. (2022) that allows to identify and extract the main runoff characteristics (i.e. runoff coefficient, event duration, peak time, event peak and volume). Given the reasonably acceptable performance of the model in reproducing the above mentioned observed characteristics, the model is run all over the region in order to get distributed regional statistics and it is then used for the characterization and classification of the identified runoff events. Many studies recently cover the topic of process-based classification of large runoff events, at different spatial scales (Merz and Blöschl, 2003; Berghuijs et al., 2016; Berghuijs et al., 2019; Stein et al., 2019). However, they lack in taking into account pre-event wetness and its connection with precipitation events, whose importance is highlighted in Viglione et al. (2010). This explains the work of Tarasova et al. (2019), discussing the main drivers of runoff generation by also including catchment wetness, leading up to the definition of a process-based framework for event characterization and classification (Tarasova et al., 2020), which is used in this Thesis. In particular, using several climatic indicators, the main mechanisms responsible for event runoff response are identified, revealing four zones of North-Western Italy with homogeneous event type frequency. During my PhD, I had the opportunity to spend three months at the Department of Catchment Hydrology of the Helmholtz Centre for Environmental Research (UFZ), led by Prof. Ralf Merz, where I worked on event identification, classification and characterization, taking advantage of the expertise of the research team, in particular of Dr. Phd Larisa Tarasova.

The Thesis is structured as follows. Chapter 1 gives an overview of the study region and the main data used. Chapter 2 presents the correlation analysis between annual maximum discharges and ETCCDI indices. Chapter 3 describes the regional calibration of the semi-distributed model over North-Western Italy, using the PASS procedure. Chapter 4 presents the R package 'hydroPASS', used to implement the PASS procedure, with examples from U.S. database. Chapter 5 is dedicated to runoff event identification, characterization and classification. Finally, Chapter 6 provides final Discussion and Conclusions.

Chapter 1

Study Region and Data

1.1 Climate and flow data

The study region is broadly coincident with the upper part of the drainage basin of the Po River and drains the semicircle of Alps and Apennines surrounding the region on three sides. The climate is temperate, and of type continental, becoming progressively temperate-cold and cold as altitude rises. Rainfall falls mainly in spring and autumn on most of the territory, and in summer in the higher inland Alpine areas. The heterogeneity of this area in terms of elevation and dominance of snow related processes leads to peculiar effects of precipitation and temperature change on floods. Figure 1.1 shows the study area in terms of elevation and river network. For the elevation, the digital elevation model provided by EarthEnv at 90m resolution is used (https://www.earthenv.org/DEM.html), while the river network is extracted from the EEA Catchments and Rivers Network System ECRINS v1.1 (ECRINS, 2012).

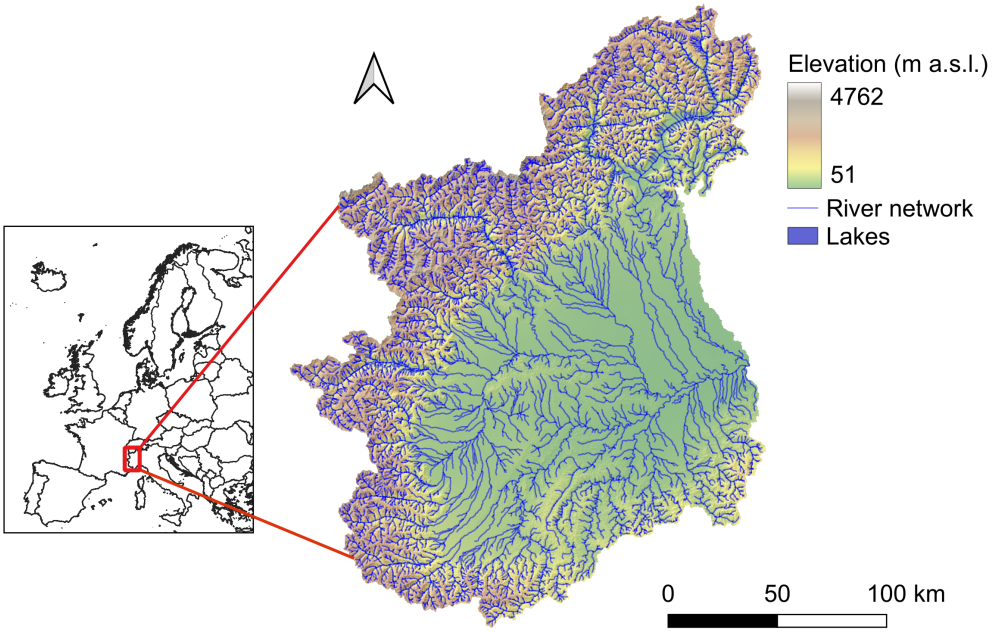


Fig. 1.1 Map of North-Western Italy region, with elevation and river network.

In this Thesis, daily precipitation [mm d^{-1}] and minimum and maximum daily temperature [°C] data are provided by a gridded dataset, covering the period from 01-12-1957 up to 31-12-2019, with cell resolution of 0.125° x 0.125°. This is derived by spatial regridding through Optimal Interpolation (OI) of daily observations from meteorological stations, collected by the Hydrographic Office network and the network of the Regional Environmental Protection Agengy (ARPA) telemetry stations (NWOI dataset, https://www.arpa.piemonte.it/scheda-informativa/dataset-griglia-nwioi). The technique allows to obtain a regular grid by homogeneization of observational data from different networks and sources. The potential evapotranspiration is calculated with the modified Blaney-Criddle equation (Doorenbos and Pruitt, 1977), by considering the mean daily temperature and the mean daily percentage of annual daytime hours for a latitude of 45° N.

Figure 1.2 reports some mean climatic characteristics of the study area, referred to the period 2000-2019, calculated using the OI dataset. The aridity index, defined

as the ratio between mean annual potential evapotranspiration and mean annual precipitation, is provided as it is a widely used climate indicator in hydrology (Blöschl et al., 2013). It is worth noting that in the Alps, particularly in the northern area, there is a tendency towards more humid conditions, characterized by high precipitation amounts, while the central lowland area is experiencing more arid conditions, as shown by potential evapotranspiration and aridity index maps. Both the mean annual precipitation and the precipitation extremes show the highest rainfall values in the northern part of the region and in the south-east, along the Apennines. Appendix A provides some geographical and topographical characteristics of 197 sites located over the study area, extracted from the Atlante dei Bacini Imbriferi Piemontesi (AtlanteBaciniImbriferi.pdf), together with information about mean climatic characteristics at the catchment scale, calculated over the period 2000-2019 (mean annual precipitation, mean annual potential evapotranspiration and aridity index).

In this Thesis, data from the regional stream gauge network managed by the regional environmental protection agency (Arpa Piemonte) are used. Data can be downloaded from Arpa Piemonte website (https://www.arpa.piemonte.it/temi/acqua). In particular, mean daily discharges in m^3/s are considered, which are obtained by averaging the 48 half-hourly values recorded each day.

1.2 Catchment Water Balance

In this section the water balance of the catchments located over the region is analysed. The water balance is governed by the following mass balance equation:

$$\frac{dS}{dt} = P - Q - G - E \tag{1.1}$$

where $\frac{dS}{dt}$ represents the variation of storage in time, P represents precipitation, Q is the outlet discharge, G the recharge of acquifers and E the actual evapotranspiration. At long time scale (e.g., 20 years) it is assumed that the storage variation is neglibile, as all the water entering the control volume, here assumed as delimited by the catchment boundaries, is equal to the water exiting it. For evaluating the water balance of the catchments it is also assumed that the groundwater flux is negligible, so that the water balance equation reduces to:

$$P = Q + E \tag{1.2}$$

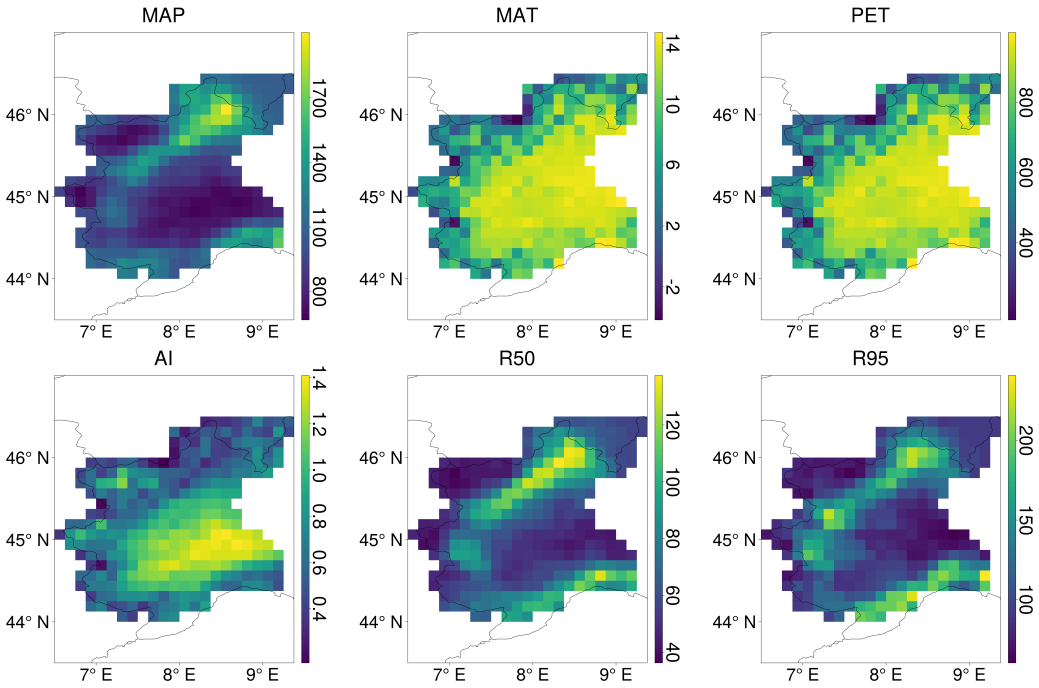
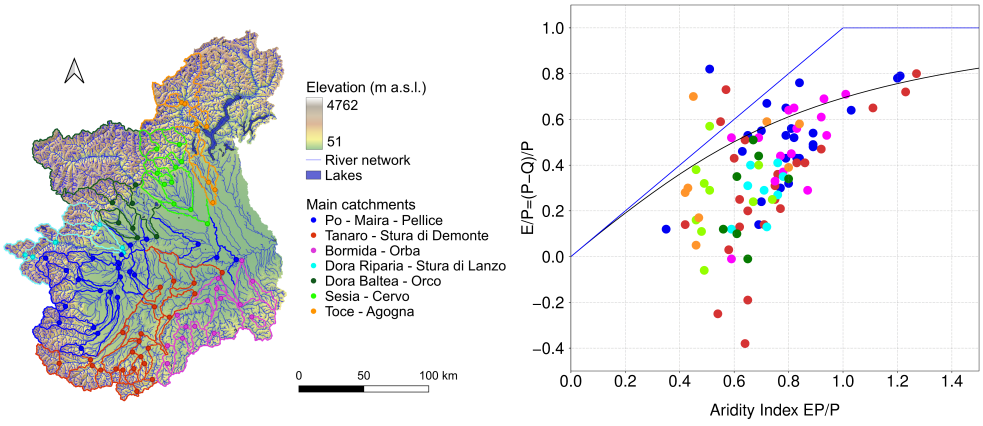


Fig. 1.2 Maps of the mean annual precipitation in mm/yr (MAP), temperature in °C (MAT), potential evapotranspiration in mm/yr (PET), aridity index (AI), median of the annual maximum daily precipitation in mm/d (R50) and 95th quantile of the annual maximum daily precipitation in mm/d (R95) for the study area, derived from the Optimal Interpolation (OI) database. Reference period: 2000-2019 (hydrologic year: 1st October 1999-30th September 2019)

.

where the terms are long-term rate of precipitation (P), long-term rate of evapotranspiration (E) and long-term rate of runoff (Q). According to the prevailing climate conditions, a catchment can be energy limited or water limited. It is energy limited when the ratio between the long-term potential evapotranspiration and rainfall (i.e. the Aridity Index) is lower than 1, so the long-term evapotranspiration corresponds to potential evapotranspiration (E = EP). On the contrary, it is water limited when the Aridity Index is greater than 1 and the actual long-term evapotranspiration corresponds to precipitation (E = P). These two cases are well represented in the Budyko diagram (Figure 1.3b).



(a) (b)

Fig. 1.3 a) Map of North-Western Italy region, with elevation, river network, catchment boundaries and outlets colored by main rivers. b) Budyko curve of the catchments located over the study area (black solid line). The blue line represents the energy and water limits to the evaporative index. The statistics are calculated over the period 2000-2019 (hydrologic year: 1st October 1999 - 30th September 2019).

Budyko was a Russian climatologist, one of the founders of physical climatology. The Budyko diagram (Budyko, 1974) defines the space where points should be located if the water balance is met. Catchments that are energy limited, so experiencing humid conditions, should be located along the bisecting line or just below. Instead, catchments that are water limited, so placed in arid conditions, should be located along the straight line corresponding to E/P = 1, or just below. Budyko determined the analytical expression of the curve representing points in the diagram, which is given below (black solid line in Figure 1.3b):

$$\frac{E}{P} = \left\{ \frac{EP}{P} \left[1 - exp\left(-\frac{EP}{P} \right) \right] tanh \left[\left(\frac{EP}{P} \right)^{-1} \right] \right\}^{1/2} \tag{1.3}$$

The water balance is evaluated for a selection of 95 stations over the study region (Figure 1.3a), having at least 9 years of data in the period 1990-2019. The records start not earlier than 1996. Table 1.1 provides some topographical and climatic characteristics of the 95 sites. For uniformity reasons, the long-term fluxes are calculated over the period 2000-2019, by considering the hydrologic year. The results are shown in Figure 1.3b. Most catchments are confined in the energy limited region of the Budyko diagram, so they experience humid conditions. It can be observed

that the majority of catchments are placed close to the Budyko curve, but a bunch of catchments shows very low values of the ratio among long-term evapotranspiration and precipitation. This is probably due to issues of precipitation undercatch, which appears particularly in the mountains when precipitation is snowfall, leading to an underestimation of the precipitation flux which in turn decreases the numerator of the ratio and so the ratio. For six catchments, the long term discharge appears to be higher than precipitation, leading to negative values of E/P. The most negative values refer to two stations along Corsaglia river and Pesio river catchment at San Bartolomeo. The average mean elevation of catchments for which the balance is not met is 1590 m.a.s.l. with two stations located over 2000 m.a.s.l. Overall, catchments showing low values of long-term actual evapotranspiration (E/P $\leq$ 0.2) are located at medium to high elevation with an average of 1625 m a.s.l. Finally, few catchments are outside the allowed region with anomalies in the E/P value, meaning that the actual evapotranspiration is higher than the potential evapotranspiration. This can be the result of a water withdrawal, for example for irrigation needs in these catchments or the presence of other fluxes (e.g., acquifer recharge), which decreases the contribution to runoff discharge.

Table 1.1 Statistics of topographical and climatic characteristics for the 95 catchments considered in the study. Reference period: 2000–2019 (hydrologic year: 1st October 1999–30th September 2019).

	mean	**CV**	**min**	**25%**	**median**	**75%**	**max**
Area (km^2)	1596	2.32	38	146	336	951	25640
Mean elevation (m a.s.l.)	1186	0.488	244	678.5	1125	1666	2339
Mean Annual precipitation (mm yr^{-1})	1095	0.210	722	932	1051	1212	1827
Mean annual runoff (mm yr^{-1})	698	0.495	148	460	644	897	1583
Aridity index (-)	0.722	0.254	0.350	0.606	0.718	0.816	1.266

1.3 Land use and Curve Number data

In this section, the land cover characteristics of the study area are discussed. Data are provided by the Corine Land Cover dataset 2018 (https://land.copernicus.eu/pan-european/corine-land-cover), which gives land cover information at 100 m spatial resolution over Europe. Starting from the entire dataset, the North-Western Italy region is selected and land use classes are reclassified into five classes (Urban, Small vegetation, Agriculture, Forest, Wetland), according to Table 1.2.

Table 1.2 Corine Land Use 2018 Classes vs Reclassified Classes.

Corine Land Use Class	**Reclassified Class**
Continuos urban fabric	
Discontinuos urban fabric	
Industrial or commercial units	Urban
Road and rail networks and associated land	
Port Areas	
Airports	
Mineral extraction sites	
Dump sites	
Construction sites	Small veg
Green urban areas	
Sport and leisure facilities	
Non-irrigated arable land	
Permanently irrigated land	
Rice fields	
Vineyards	
Fruit trees and berry Plantation	
Olive groves	Agriculture
Pastures	
Annual crops associated with permanent crops	
Complex cultivation patterns	
Land principally occupied by agriculture	
with significant areas of natural vegetation	
Agro-forestry areas	
Broad-leaved forest	
Coniferous forest	
Mixed forest	
Natural grasslands	Forest
Moors and heathland	
Sclerophyllous vegetation	
Transitional woodland-shrub	
Beaches, dunes, sands	
Bare rocks	
Sparsely vegetated areas	Small veg
Burnt areas	
Glaciers and perpetual snow	
Inland marshes	
Peat bogs	
Salt marshes	
Salines	
Intertidal flats	Wetland
Water courses	
Water bodies	
Coastal lagoons	
Estuaries	
Sea and ocean	

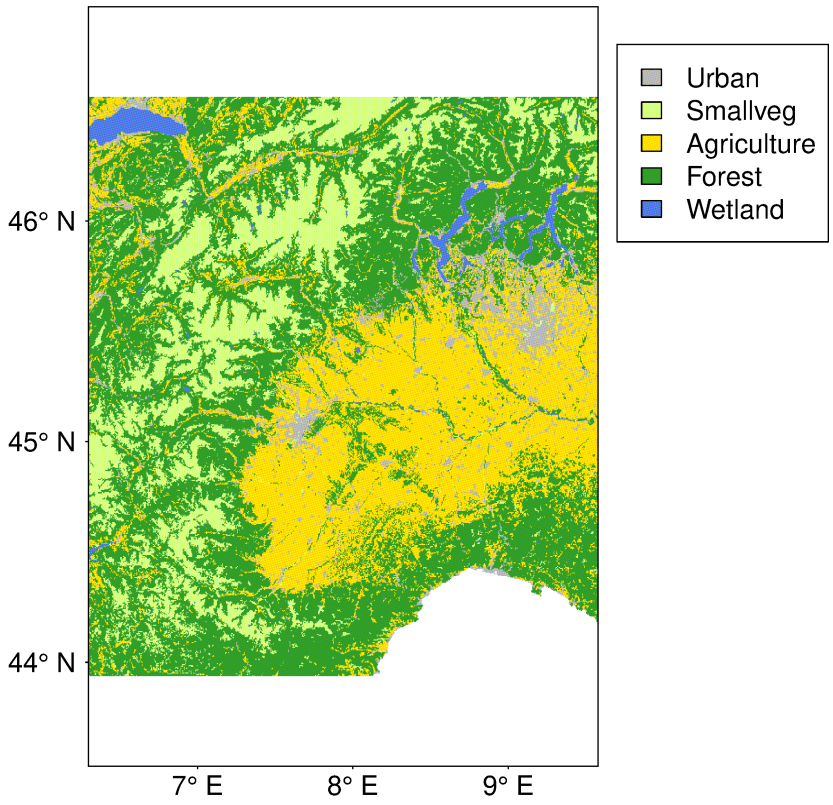


Fig. 1.4 Reclassified Corine Land Cover Dataset over North-Western Italy.

Figure 1.4 shows the spatial distribution of land cover type over North-Western Italy. The urban settlements, in particular Turin in Piemonte and Milan in Lombardia, can be spotted. The Po Valley is characterized by agricultural fields, while the mountain ranges, (i.e. Alps and Appennines) are characterized by forests and, at very high elevation, sparse vegetation and perennial snow and glaciers. Also water bodies, such as lakes, are quite visible in the map. The Optimal Interpolation grid is considered to calculate the percentage of pixels pertaining to a specific class inside each cell and the result is provided in Figure 1.5. The new rasters are used for the application of the parameter regionalization method which is discussed in Chapter 3.

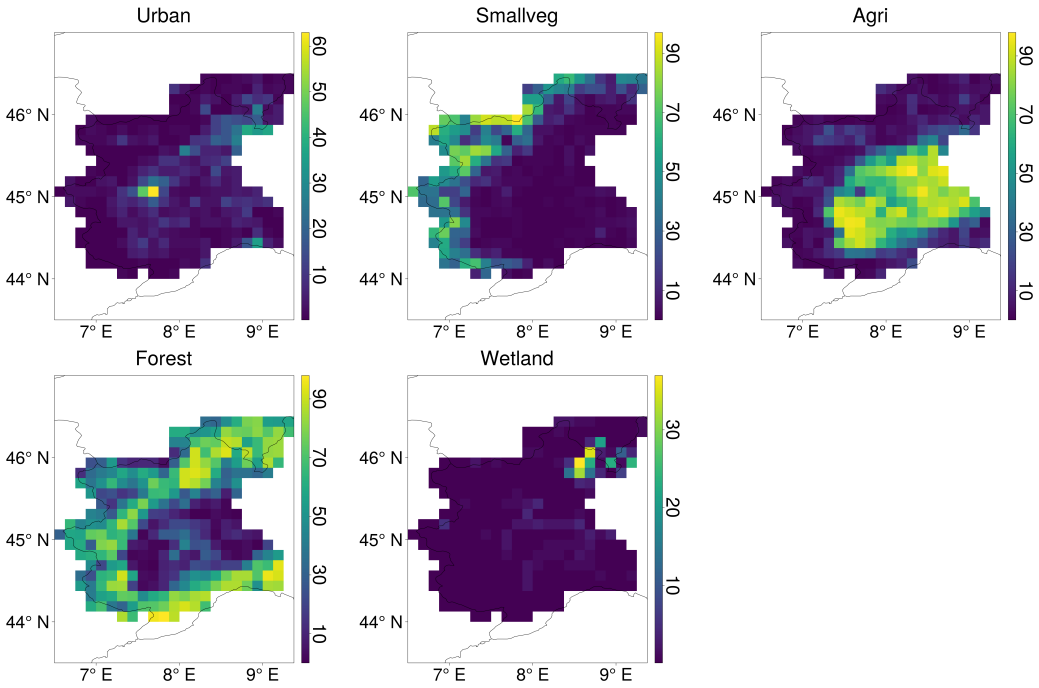


Fig. 1.5 Maps of the fraction (%) of land use classes (urban, small vegetation, agriculture, forest, wetland) for the study area.

Land use or land cover information is essential to determine the Curve Number (CN). CN is a parameter whose value ranges from 0 to 100, summarizing three characteristics related to soil: the soil typology, land use and soil moisture level before the rainfall event. The theoretical limits of CN = 0 and CN = 100 represent the case when all the precipitation infiltrates in the soil and when all the precipitation is transformed into surface runoff, respectively. The Soil Conservation Service-Curve Number (SCS-CN) method classifies the soil in four hydrologic groups (USDA-NRCS, 1986):

- A - Soils with low potential for surface runoff. This group includes deep sands with very low silt and clay and also deep permeable gravel.
- B - Soils with quite low potential for surface runoff. This group includes the majority of sandy soils, less deep compared to group A. High infiltration capacity even at saturation conditions.
- C - Soils with quite high potential for surface runoff. It includes thin soils with substantial amount of clay and colloids (but less than in group D). Low infiltration capacity at saturation.

- D - Soils with high potential for surface runoff. This group includes clay soils with high swelling capacity but also thin soils with an impermeable layer close to the surface.

At first, a CN reflecting averaged soil moisture conditions (CN_{II}) is considered. Tables provided by the United States Department of Agriculture (USDA-NRCS, 1986; USDA-NRCS, 2004) reporting CN_{II} values for rural and urbanized areas are used to determine values of CN_{II} associated to the different hydrologic soil types and the land use classes of the Corine Land Cover 2018 (Table 1.3). The CN_{II} values are attributed to the new classes by direct comparison with the classes provided by USDA or by physical consistency among the two sources. Starting from the rasters of soil characteristics identified by using the Harmonized World Soil Database (HWSD; http://webarchive.iiasa.ac.at/Research/LUC/External-World-soil-database/), it is defined a raster of soil typology for our study area, by adopting the following simplified criteria, based on the mass fraction [% wt.] of sand, clay and silt in the topsoil (0-30 cm) and subsoil (30-100 cm):

- A - Soils with S_SAND ≥ 70 & (S_CLAY + S_SILT) ≤ 30, T_SAND ≥ 70 & (T_CLAY + T_SILT) ≤ 30
- B - Soils with $50 \leq$ S_SAND < 70 & $30 <$ (S_CLAY + S_SILT) ≤ 50, $50 \leq$ T_SAND < 70 & $30 <$ (T_CLAY + T_SILT) ≤ 50
- C - Soils with $20 \leq$ S_SAND < 50 & $50 <$ (S_CLAY + S_SILT) ≤ 80 & S_CLAY ≥ 10, $20 \leq$ T_SAND < 50 & $50 <$ (T_CLAY + T_SILT) ≤ 80 & T_CLAY ≥ 10
- D - Soils with $10 \leq$ S_SAND < 20 & $80 <$ (S_CLAY + S_SILT) ≤ 90 & S_CLAY ≥ 20, $10 \leq$ T_SAND < 20 & $80 <$ (T_CLAY + T_SILT) ≤ 90 & T_CLAY ≥ 20

The CN_{II} is used to determine the corrisponding CN_I for drier soil moisture conditions and CN_{III} for wetter conditions. The formula used are the following (Mishra et al., 2008):

$$CN_I = \frac{CN_{II}}{2.3 - 0.013 CN_{II}} \tag{1.4}$$

$$CN_{III} = \frac{CN_{II}}{0.43 + 0.0057 CN_{II}} \tag{1.5}$$

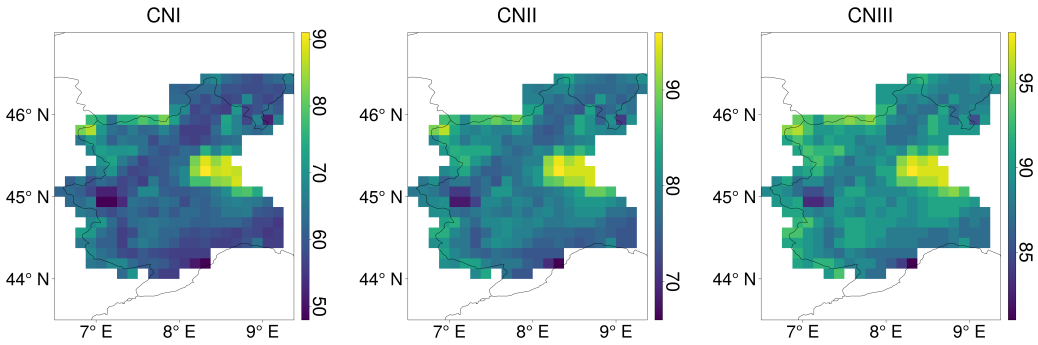


Fig. 1.6 Maps of the Curve Number (CN_I, CN_{II}, CN_{III}) for the study area.

Figure 1.6 represents the spatial distribution of the mean CN values over North-Western Italy, by considering the Optimal Interpolation grid. It is worth noting that the highest values are found around Vercelli province, in the East, where rice fields represent a dominant cultivation, and in the Alps (northwest area), characterized by glaciers and perpetual snow. The lowest CNs are instead found in forestry areas in the West and in the South, in Liguria region, where vineyards and olive groves are typically present. As for the land cover, the gridded CNs are used as descriptors in the PASS regionalization method described in Chapter 3.

Table 1.3 CN_{II} values associated to the Corine Land Use classes and the four hydrologic soil types A-B-C-D, derived by USDA-NRCS (1986) and USDA-NRCS (2004).

Corine Land Use Class	**A**	**B**	**C**	**D**
Continuos urban fabric	77	85	90	92
Discontinuos urban fabric	61	75	83	87
Industrial or commercial units	89	92	94	95
Road and rail networks and associated land	98	98	98	98
Port Areas	98	98	98	98
Airports	98	98	98	98
Mineral extraction sites	76	85	89	91
Dump sites	76	85	89	91
Construction sites	76	85	89	91
Green urban areas	39	61	74	80
Sport and leisure facilities	49	69	79	84
Non-irrigated arable land	61	73	81	84
Permanently irrigated land	63	73	80	83
Rice fields	96	96	96	96
Vineyards	66	74	80	82
Fruit trees and berry plantation	62	71	78	81
Olive groves	62	71	78	81
Pastures	30	58	71	78
Annual crops associated with permanent crops	64	73	79	82
Complex cultivation patterns	64	73	79	82
Land principally occupied by agriculture with significant areas of natural vegetation	64	73	79	82
Agro-forestry areas	64	73	79	82
Broad-leaved forest	36	60	73	79
Coniferous forest	45	66	77	83
Mixed forest	36	60	73	79
Natural grasslands	49	69	79	84
Moors and heathland	35	56	70	77
Sclerophyllous vegetation	35	56	70	77
Transitional woodland-shrub	43	65	76	82
Beaches, dunes, sands	49	68	79	84
Bare rocks	76	85	89	91
Sparsely vegetated areas	63	77	85	88
Burnt areas	63	77	85	88
Glaciers and perpetual snow	98	98	98	98
Inland marshes	98	98	98	98
Peat bogs	98	98	98	98
Salt marshes	98	98	98	98
Salines	98	98	98	98
Intertidal flats	98	98	98	98
Water courses	98	98	98	98
Water bodies	98	98	98	98
Coastal lagoons	98	98	98	98
Estuaries	98	98	98	98
Sea and ocean	98	98	98	98

Chapter 2

Correlation of Climate and Flood Indices

In this chapter, a correlation analysis is performed between annual maximum mean daily discharges and standard climate indices of precipitation and temperature extremes, as defined by the Commission for Climatology/World Climate Research Programme/Technical Commission for Oceanography and Marine Meteorology (CCI/WCRP/JCOMM) Expert Team on Climate Change Detection and Indices (ETCCDI, see e.g., Peterson, 2005; Zhang et al., 2005). ETCCDI indices are tipically used in the climate literature to study a variety of extreme events such as heavy rain, floods, droughts, heat waves, etc. In particular, they find wide application in climate change studies, not limited to past climate change, but also future changes by using projections of climate models (e.g., Sardella et al., 2020). Nonetheless, few studies have focused on the relationship between ETCCDI indices and peak discharges and it is of interest in hydrology to evaluate which ETCCDI indices are relevant for characterizing and predicting flood changes in the Alpine region.

2.1 Data and Methods

A description of the 27 ETCCDI indices considered in this study is provided in Table 2.1. ETCCDI indices are calculated at the annual timescale using the NWOI dataset. For this purpose the `climdex.pcic.ncdf` R library is used (https://github.com/pacificclimate/climdex.pcic.ncdf), which performs an automatic calculation that saves the gridded

outputs as netCDF files. Indices are referred to thresholds calculated over the base period 1961-1990. The annual maximum mean daily discharge is selected as typical flood index, as reported in literature on flood change (e.g., Blöschl et al., 2017; Blöschl et al., 2019). The choice of mean daily flow rather than peak flow is explained by the better consistency among the former and the space-time scale of climate indices. Years with missing daily data covering a period greater or equal to 3 months are discarded from the analysis. Both the annual maximum discharge and the ETCCDI indices are calculated considering the hydrologic year (1^{st} October - 30^{th} September). Average annual indices at the catchment scale are obtained by clipping the gridded dataset based on catchment boundaries, by making use of a weighted average, considering the proportion of each cell inside the catchment. The indices are coupled with the flow annual maxima, so only years with available discharge data are considered. The choice of a quite coarse data resolution is made based on two reasons: first, this is consistent with the outputs of regional climate models (EURO-CORDEX); secondly, the aim of this research work is to describe regional floods, not local flash floods.

Figure 2.1 shows several information about flood data used. In particular, Figure 2.1a and Figure 2.1b show the dependence of the mean annual specific flood (MAF) and the coefficient of variation (CV) of annual specific floods on catchment area. MAF and CV decrease with catchment area, as expected. Fitting a linear model to data, which is equivalent to assuming a power law relationship between the variables and catchment area, unveils a pattern which is already found in other studies (e.g., Lun et al., 2021; Merz and Blöschl, 2003; Merz and Blöschl, 2005). The coefficients (β) of MAF and CV found for North-Western Italy area (-0.136 and -0.049, respectively) are consistent with the ones found in Lun et al. (2021) for the Alpine area (-0.208 and -0.020, respectively), but closer to values pertaining to the Atlantic region (-0.184 and -0.042, respectively). However, the number of sites considered here and their record length, which are smaller than in Lun et al. (2021), determine a remarkable scatter.

Two types of correlation analysis are considered: on the one hand a temporal correlation is performed at the annual time scale among maximum discharges and ETCCDI indices for each catchment, in order to capture the best covariates explaining the annual variability of floods (Section 2.1.1). On the other hand, in the spirit of comparative hydrology (Falkenmark and Chapman, 1989), trends of maximum discharges and trends of ETCCDI indices are spatially correlated, in order to find

Table 2.1 ETCCDI climate indices

Index	Description
FD - Number of frost days	Annual count of days when TN (daily minimum temperature) < 0°C [days]
SU - Number of summer days	Annual count of days when TX (daily maximum temperature) > 25°C [days]
ID - Number of icing days	Annual count of days when TX (daily maximum temperature) < 0°C [days]
TR - Number of tropical nights	Annual count of days when TN (daily minimum temperature) > 20°C [days]
GSL - Growing season length	Annual (1^{st} Jan to 31^{st} Dec in Northern Hemisphere (NH), 1^{st} July to 30^{th} June in Southern Hemisphere (SH)) count between first span of at least 6 days with daily mean temperature TG > 5°C and first span after July 1^{st} (Jan 1^{st} in SH) of 6 days with TG < 5°C.
TXx	Annual maximum value of daily maximum temperature [°C]
TNx	Annual maximum value of daily minimum temperature [°C]
TXn	Annual minimum value of daily maximum temperature [°C]
TNn	Annual minimum value of daily minimum temperature [°C]
TN10p	Percentage of days when daily minimum temperature $< 10^{th}$ percentile
TX10p	Percentage of days when daily maximum temperature $< 10^{th}$ percentile
TN90p	Percentage of days when daily minimum temperature $> 90^{th}$ percentile
TX90p	Percentage of days when daily maximum temperature $> 90^{th}$ percentile
WSDI - Warm speel duration index	Annual count of days with at least 6 consecutive days when daily maximum temperature $> 90^{th}$ percentile [days]
CSDI - Cold speel duration index	Annual count of days with at least 6 consecutive days when daily minimum temperature $< 10^{th}$ percentile [days]
DTR - Daily temperature range	Annual mean difference between daily maximum temperature and daily minimum temperature [°C]
Rx1day	Annual maximum 1-day precipitation [mm]
Rx5day	Annual maximum consecutive 5-day precipitation [mm]
SDII - Simple precipitation intensity index	Average precipitation rate on wet days (R $\geq$ 1mm) [mm/day]
R10mm	Annual count of days when precipitation $\geq$ 10mm [days]
R20mm	Annual count of days when precipitation $\geq$ 20mm [days]
R1mm	Annual number of wet days [days]
CDD	Maximum length of dry spell i.e. number of consecutive days with precipitation $<$ 1mm [days]
CWD	Maximum length of wet spell i.e. number of consecutive days with precipitation $\geq$ 1mm [days]
R95pTOT	Annual total precipitation when daily precipitation is greater than the 95^{th} percentile [mm]
R99pTOT	Annual total precipitation when daily precipitation is greater than the 99^{th} percentile [mm]
PRCPTOT	Annual total precipitation in wet days [mm]

which covariates best explain the regional variability of the decadal tendency of floods (Section 2.1.2). By focusing on multi-year tendency rather than on annual variability, the latest analysis can be useful to select specific ETCCDI indices as possible climate covariates of flood discharges, for regional non-stationary flood

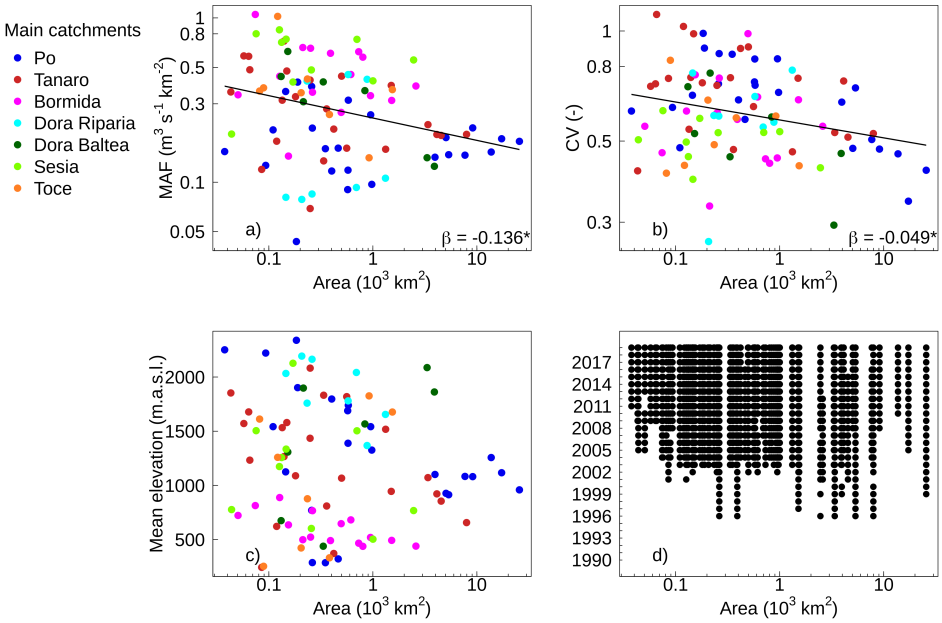


Fig. 2.1 a) Mean annual specific flood (MAF) vs. catchment area. b) Coefficient of variation (CV) of annual specific floods vs. catchment area, colored as in Figure 1.3a. Lines are ordinary least squares regression lines. The values of the slope (β) for a double logarithmic relationship are also reported. *Indicates statistical significance for a one-sided t-test at the 5% significance level. c) Area vs. mean catchment elevation. d) Data consistency vs. catchment area.

frequency analysis. Finally, circular statistics is used to describe the average timing and concentration of observed maximum discharges (Section 2.1.3).

2.1.1 Correlation measures

In the first analysis, Spearman's rank correlation is applied to annual data in order to determine which indices show the highest temporal correlation with annual maximum daily discharges. The choice of the Spearman correlation instead of other measures of variable association (e.g., Pearson) is justified by the non-linear relationship between precipitation, temperature and discharge. The Pearson correlation coefficient between two variables x and y can be expressed as follows (Helsel et al., 2020, Chapter 8.2):

$$r_{x,y} = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{\sigma_x}\right)\left(\frac{y_i - \bar{y}}{\sigma_y}\right) = \frac{cov(x,y)}{\sigma_x \sigma_y} \tag{2.1}$$

where n is the length of the sample, $\bar{x}$ and $\bar{y}$ are the mean values of x and y, σ_{x} and σ_{y} are the standard deviations of x and y. The Spearman correlation coefficient is defined as the Pearson correlation coefficient among the ranks (Rx and Ry) of the variables:

$$\rho_{Rx,Ry} = \frac{cov(Rx,Ry)}{\sigma_{\mathrm{Rx}}\sigma_{\mathrm{Ry}}} \tag{2.2}$$

where $cov(Rx,Ry)$ is the covariance of rank variables, σ_{Rx} and σ_{Ry} are the standard deviations of the rank variables. By substituting the ranks into Equation (2.1), the coefficient can be computed with the following formulation, which holds in case of no ties (Helsel et al., 2020, Chapter 8.3):

$$\rho = \frac{\sum_{i=1}^{n}(Rx_{\mathrm{i}}Ry_{\mathrm{i}}) - n\left(\frac{n+1}{2}\right)^2}{n(n^2-1)/12} \tag{2.3}$$

The significance of ρ under the null hypothesis of no-correlation between the variables is tested with the test statistic S (Helsel et al., 2020, Chapter 8.3):

$$S = \sum_{i=1}^{n}(Rx_{\mathrm{i}} - Ry_{\mathrm{i}})^2 = (1-\rho)(n^3-n)/6 \tag{2.4}$$

with the right hand formulation that holds in case of no ties. For small sample sizes ($n < 20$), the algorithm AS 89 (Best and Roberts, 1975) allows to compute exact p-values, by calculating the discrete probability distribution of S. For large sample sizes ($n > 20$), the test is computed on the transformed variable:

$$t = \frac{\rho\sqrt{n-2}}{1-\rho^2} \tag{2.5}$$

where n is the length of the two tested samples and t follows a Student's t-distribution with n-2 degrees of freedom, under the null hypothesis of no-correlation between the variables (Helsel et al., 2020, Chapter 8.2). Accordingly, p-values are calculated with the following formulation:

$$p = 1 - F_{\mathrm{t}}(abs(t), n-2) \tag{2.6}$$

where $F_{\mathrm{t}}(abs(t), n-2)$ is the non-exceedance probability associated with $abs(t)$ for a Student's t-distribution with $n-2$ degrees of freedom. If ties are present, this is the approach used for the calculation of p-values. One-sided tests at 5% significance

level are considered, both for positive and negative correlation. Confidence intervals of the correlation are given by the Fisher z transform of the correlation (Fisher, 1915):

$$z = \frac{1}{2}\ln\left(\frac{1+\rho}{1-\rho}\right) = \operatorname{arctanh}(\rho) \tag{2.7}$$

assuming z as normally distributed with standard deviation:

$$\sigma_z = \sqrt{\frac{1}{n-3}} \tag{2.8}$$

Finally, a transformation allows to obtain confidence intervals in correlation units.

The same procedure is also applied to detrended data, to check for the potential impact of trends in the data on the results of annual correlation among maximum discharges and ETCCDI indices. To this aim, the Theil-Sen linear regression model with time is adopted (Section 2.1.2). The detrending is performed for each catchment, by subtracting the predicted values to the observed values of the variable and then the correlation is calculated on the residuals. For each index, a regional mean Spearman correlation coefficient is calculated, weighted for the uncertainty associated with the single correlation values. In particular, the weights are function of the confidence intervals of the correlation coefficients:

$$w_i = (up_i - low_i)^{-1} \tag{2.9}$$

where w_i is the weight associated to catchment i, up_i and low_i are, respectively, the upper and lower bounds of the confidence interval. In this Thesis, the Spearman's rank correlation is also applied to evaluate the correlation among dedacal tendencies of annual maximum discharges and tendencies of climate indices. The choice of this measure is justified by its application in other studies that analyse the interconnection among discharge maxima and climate variables (e.g., Blöschl et al., 2019).

2.1.2 Tendency measures

The potential presence of decadal tendencies in the data is checked for both the annual maximum discharge and climate indices. To this aim, the Theil-Sen model is adopted, as defined by Theil (1950), with further investigations by Sen (1968). This

is a robust nonparametric linear regression model:

$$\hat{y} = \alpha + \beta \cdot x \tag{2.10}$$

where the slope estimator (β) represents the median of the slopes calculated for all possible pair of values assumed by the variable over the years:

$$\beta = median\left(\frac{y_j - y_i}{j - i}\right), i < j \tag{2.11}$$

where y refers to the annual values of the variable and i, j refer to distinct years. The decadal tendencies will be plotted as the percentage of the mean value of the variable per year (i.e., $100 \cdot \beta / mean(y)$). The intercept ($\alpha$) of the regression line is obtained following the approach used by Conover (1999):

$$\alpha = y_{med} - \beta \cdot x_{med} \tag{2.12}$$

where x_{med} and y_{med} are the medians of x and y, which represent time and the selected variable, respectively.

The trend significance is evaluated with one-sided Mann-Kendall tests (Mann, 1945) at the 5% significance level. The test statistic is calculated by computing the sum of the sign of differences for all $\frac{n(n-1)}{2}$ pairs extracted from the n observations:

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} sgn(x_j - x_i) \tag{2.13}$$

$$sgn(x_j - x_i) = \begin{cases} +1, \ if \ (x_j - x_i) > 0 \\ 0, \ if \ (x_j - x_i) = 0 \\ -1, \ if \ (x_j - x_i) < 0 \end{cases} \tag{2.14}$$

For $n \geq 10$ (Kendall, 1948), the normal approximation test can be used. The test statistic Z is defined as:

$$Z = \begin{cases} \frac{S-1}{\sqrt{var(S)}}, \ if \ S > 0 \\ 0, \ if \ S = 0 \\ \frac{S+1}{\sqrt{var(S)}}, \ if \ S < 0 \end{cases} \tag{2.15}$$

with $var(S)$ being the variance of S:

$$var(S) = \frac{n(n-1)(2n+5) - \sum_{j=1}^{g} t_{\mathrm{j}}(t_{\mathrm{j}}-1)(2t_{\mathrm{j}}+5)}{18} \tag{2.16}$$

where g represents the number of groups of tied values and t_{j} the number of ties in group j. In case of no ties the formulation becomes:

$$var(S) = \frac{n(n-1)(2n+5)}{18} \tag{2.17}$$

Z is distributed as a standard normal distribution under the null hypothesis of no trend of the variable (Mann, 1945). If the test provides $Z > 0$, this is an indication of increasing trend (and viceversa) and p-values are computed as the exceedance probability associated with Z (one-sided tests). Besides the test, confidence intervals are computed for β, as a measure of uncertainty in the trends estimation, by selecting the upper and lower limits within the sample of slopes. Following Hollander et al. (1999), the critical value is given by the quantile of the standard normal distribution $Z_{\frac{\alpha}{2}}$, where α is the confidence level, and the upper and lower ranks of the slopes are found by:

$$R_{\mathrm{u}} = \frac{N + Z_{\frac{\alpha}{2}}\sqrt{var(S)}}{2} + 1 \tag{2.18}$$

$$R_{\mathrm{l}} = \frac{N - Z_{\frac{\alpha}{2}}\sqrt{var(S)}}{2} \tag{2.19}$$

where $N = \frac{n(n-1)}{2}$ is the number of computed slopes. The confidence level used in this analysis is 0.10, which is coherent with the Mann-Kendall test at the 5% significance level, applied for positive and negative trends separately. Also for the analysis of tendency, a regional mean tendency is provided, weighting for the uncertainty associated with the single tendencies. The weights are given by:

$$w_{\mathrm{i}} = e^{(-k(up_{\mathrm{i}} - low_{\mathrm{i}}))} \tag{2.20}$$

with k = 0.25, to constrain the weights range in case of confidence range equal to 0.

2.1.3 Circular statistics

Circular statistics (Bayliss and Jones, 1993; Mardia, 1972) is adopted to provide information on the seasonality of flood indices. The choice of this approach is justified by its application in previous large-scale studies on river flood timing in Europe (Blöschl et al., 2017). For each site, the average day of the year on which floods have occurred is calculated. The date of occurrence D_i of a flood in year i is converted to an angle θ_i:

$$\theta_i = D_i \cdot \frac{2\pi}{m_i} \quad 0 \leq \theta_i \leq 2\pi \tag{2.21}$$

where $D_i = 1$ corresponds to January 1 and $D_i = \mathrm{m}_i$ to December 31 , and m_i represents the number of days in that year. The average date of occurrence $\bar{D}$ of a flood is calculated as follows:

$$\bar{D} = \begin{cases} \tan^{-1}\left(\frac{\bar{y}}{\bar{x}}\right) \cdot \frac{\bar{m}}{2\pi} & \bar{x} > 0, \bar{y} \geq 0 \\ \tan^{-1}\left(\frac{\bar{y}}{\bar{x}}\right) \cdot \frac{\bar{m}}{2\pi} + \pi & \bar{x} \leq 0 \\ \tan^{-1}\left(\frac{\bar{y}}{\bar{x}}\right) \cdot \frac{\bar{m}}{2\pi} + 2\pi & \bar{x} > 0, \bar{y} < 0, \end{cases} \tag{2.22}$$

with

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} \cos(\theta_i) \tag{2.23}$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^{n} \sin(\theta_i) \tag{2.24}$$

$$\bar{m} = \frac{1}{n} \sum_{i=1}^{n} m_i \tag{2.25}$$

where $\bar{x}$ and $\bar{y}$ are the cosine and sine components of the average date, respectively, $\bar{m}$ is the average number of days per year (365.25), and n is the total number of flood peaks observed at the site. The concentration R of the date of occurrence around the average date is:

$$R = \sqrt{\bar{x}^2 + \bar{y}^2} \quad 0 \leq R \leq 1 \tag{2.26}$$

R spans from $R = 0$ (no concentration, i.e. floods are widely dispersed over the year) to $R = 1$ (all floods at a site occur on the same day of the year).

2.2 Results

2.2.1 Temporal correlation of annual climate and flood indices

In this section, the results of the correlation among the annual time series of ETCCDI indices and the annual maximum discharges are shown. The analysis gives similar outcomes when using original and detrended data, indicating that the analysis is robust and is not affected by the presence of tendencies in the data. The Spearman's rank correlation coefficients for the ETCCDI indices showing the highest mean regional correlation are reported in Figure 2.2. As one could expect, annual maximum flows show the highest correlation with indices of precipitation extremes: the annual maximum 1-day precipitation (Rx1day), the annual maximum consecutive 5-day precipitation (Rx5day), the annual total precipitation when daily precipitation is above the 95^{th} daily percentile (R95pTOT) and the annual total precipitation when daily precipitation is above the 99^{th} daily percentile (R99pTOT). Indices reflecting average rainfall conditions, such as the simple precipitation intensity index (SDII) and the annual total precipitation (PRCPTOT) show a lower mean correlation. Figure 2.2 reveals that the strength of the correlation depends on catchment area. This is particularly clear for Rx5day, which shows higher correlation values compared to Rx1day for large catchments ($A > 2000$ km^2), while the opposite happen for the small ones ($A < 100$ km^2). Moreover, R99pTOT shows more significant correlation with maximum discharges than R95pTOT, especially for medium-to-large sized catchments ($A > 500$ km^2). Results for PRCPTOT generally indicate a weaker correlation compared to other indices and no significant results are found for very large catchments located over the Po River valley. Temperature indices don't reveal high and significant correlation with discharge extremes.

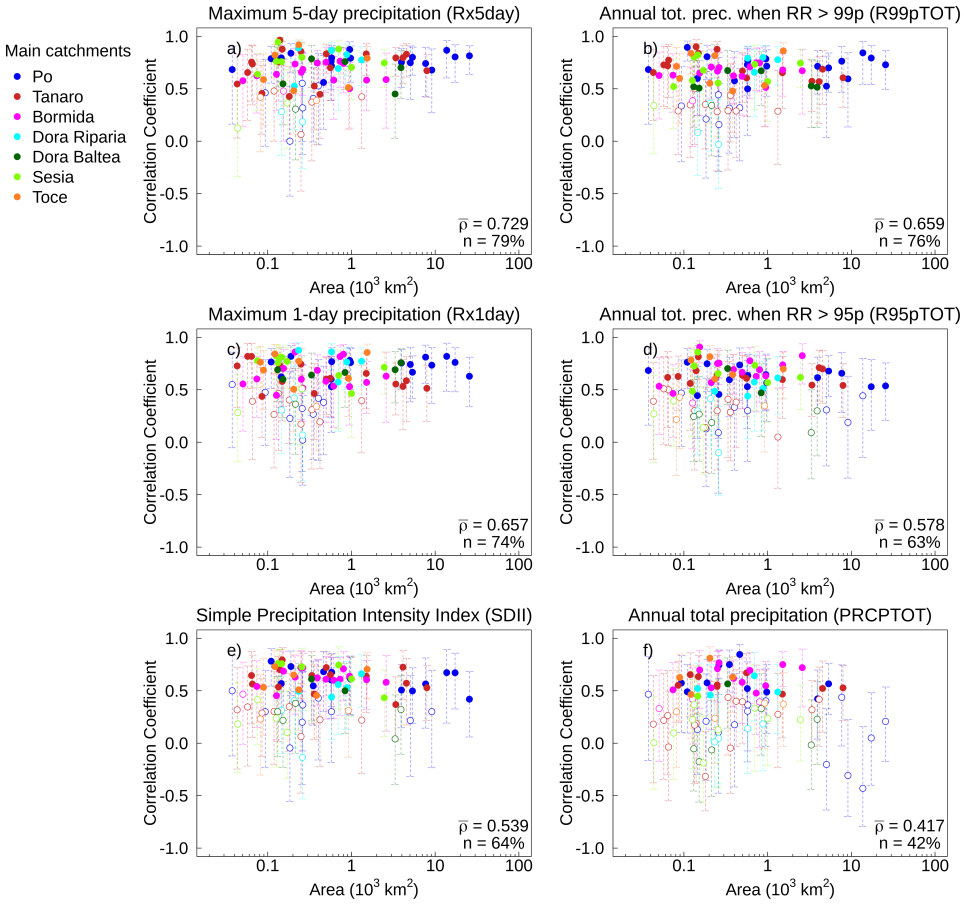


Fig. 2.2 Spearman's rank correlation coefficients among annual maximum mean daily discharges and a) maximum 5-day precipitation (Rx5day), b) annual total precipitation above the 99th percentile (R99pTOT), c) maximum 1-day precipitation (Rx1day), d) annual total precipitation above the 95th percentile (R95pTOT), e) simple precipitation intensity index (SDII), f) annual total precipitation (PRCPTOT) for all catchments vs. catchment area, colored as in Figure 1.3a. For each index, the regional mean correlation coefficient ($\bar{\rho}$) and the percentage of significant cases (n) (one-sided tests at 5% level), are reported. Full dots represent catchments with significant positive correlation, while empty dots represent not significant positive correlation.

2.2.2 Decadal tendency of climate and flood indices

It is of interest to investigate whether the decadal tendency of flood magnitudes is function of the same climate indices which are relevant to explain the annual floods. In this section the tendencies of both annual maximum discharges and ETCCDI indices over the period 2000-2019 are evaluated. Figure 2.3 shows the estimated trends in the annual maximum flows. The results show there is not a dominant

tendency at the regional scale and some noise appears in the data, mainly due to the limited length of the time series (7% of the sites have a significant trend according to one-sided Mann-Kendall tests at the 5% significance level). Many stations are placed on the same river from upstream to downstream and some spatial coherence in terms of tendency can be observed. Bormida and Tanaro catchments (pink and red points in Figure 2.3, respectively), which are located in the southern part of the region, show a positive tendency, while the sign is less homogeneous in other macro-catchments. Looking at the Po River (blue points), for example, the tendency is negative for small catchments, which are tipically the Alpine ones, while for medium-sized hilly catchments it is positive and for the largest valley catchments no distint sign can be recognized. Also in the smallest Sesia catchments (light green points) the tendency is negative.

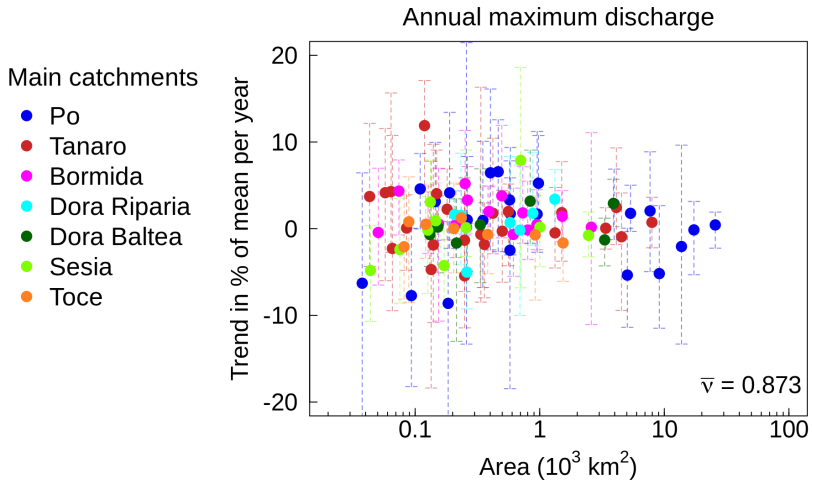


Fig. 2.3 Trends of annual maximum mean daily discharge for each catchment vs. catchment area, colored as in Figure 1.3a. The regional mean trend ($\bar{v}$) is also reported.

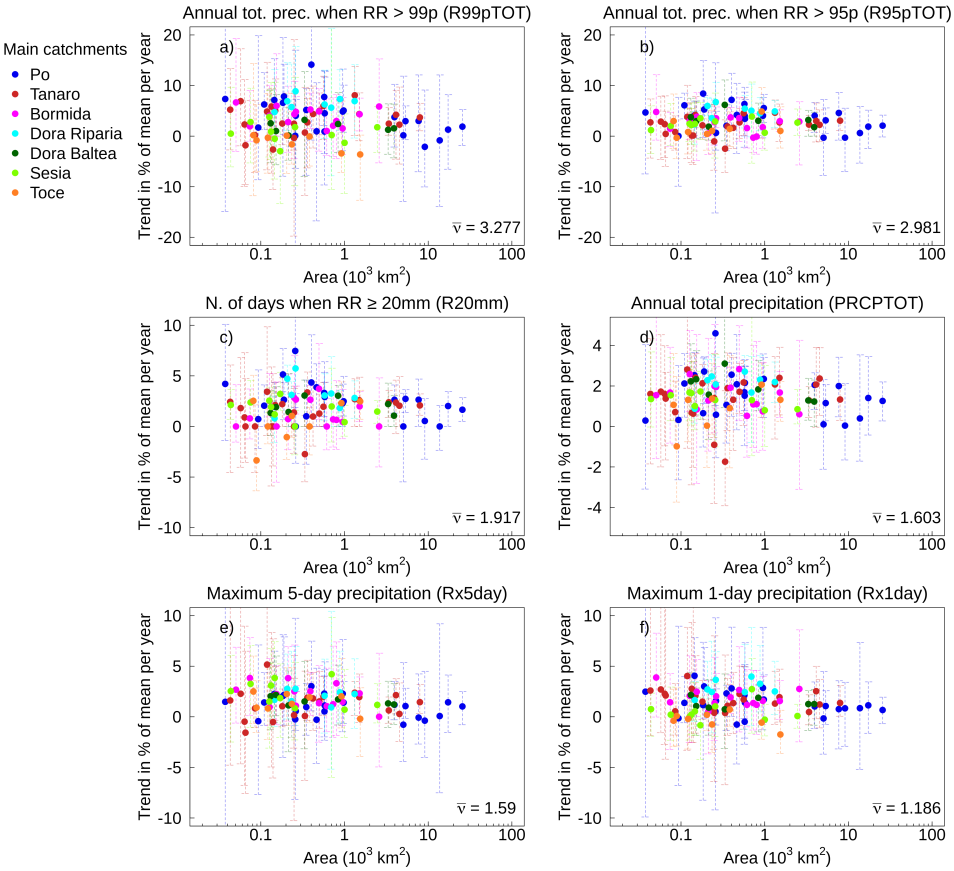


Fig. 2.4 Trends of a) annual total precipitation above the 99th percentile (R99pTOT), b) annual total precipitation above the 95th percentile (R95pTOT), c) annual number of days when precipitation is above 20mm (R20mm), d) annual total precipitation (PRCPTOT), e) maximum 5-day precipitation (Rx5day), f) maximum 1-day precipitation (Rx1day), for each catchment vs. catchment area, colored as in Figure 1.3a. For each index, the regional mean trend ($\bar{v}$) is reported.

Contrary to river flows, ETCCDI indices reveal significant tendencies both for precipitation and temperature. Figure 2.4 shows the main outcomes for precipitation indices. The trends are mainly positive and extreme indices (R99pTOT, R95pTOT) reveal a decreasing tendency for increasing catchment area, especially for the Po River catchments (blue points). Total precipitation (PRCPTOT) shows a clear pattern as it experiences a positive significant tendency in around 50% of the catchments, but the spatial variability does not seem to be a function of catchment area. R20mm, Rx5day and Rx1day show lack of a strong spatial heterogeneity, indicating that they may not be ideal in explaining the spatial variability of flood decadal tendencies.

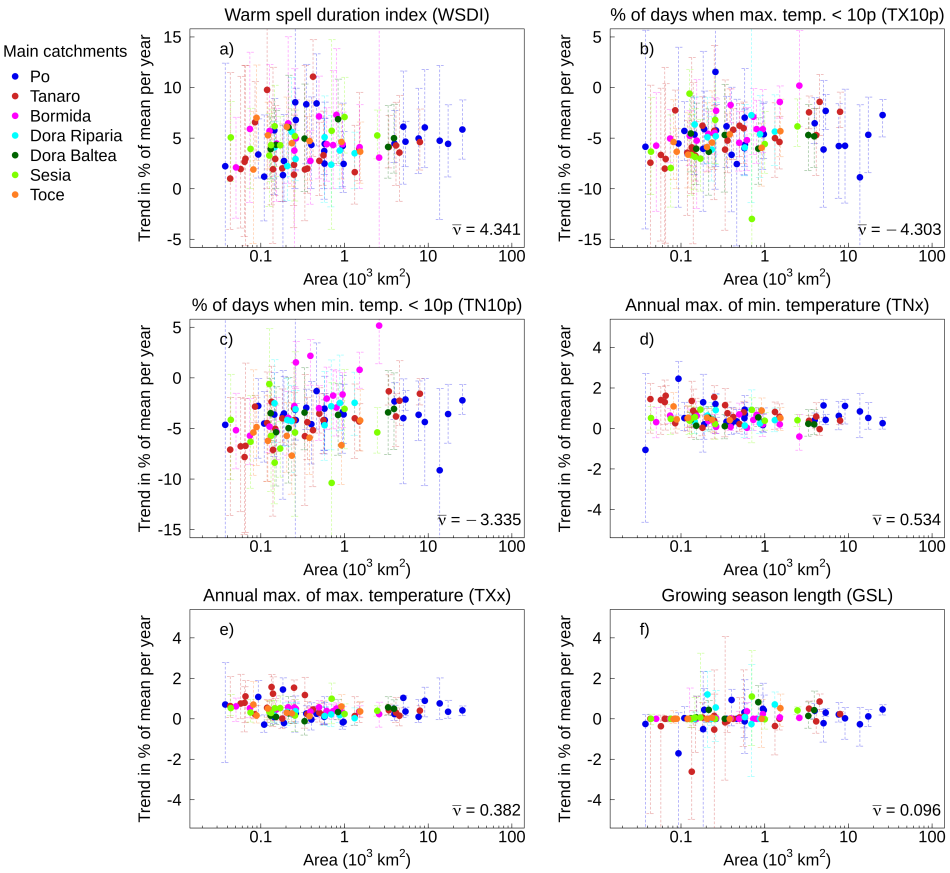


Fig. 2.5 Trends of a) warm spell duration index (WSDI), b) % of days when the maximum temperature is lower than the 10th percentile (TX10p), c) % of days when the minimum temperature is lower than the 10th percentile (TN10p), d) maximum value of daily minimum temperature (TNx), e) maximum value of daily maximum temperature (TXx), f) growing season length (GSL) for each catchment vs. catchment area, colored as in Figure 1.3a. For each index, the regional mean trend ($\bar{v}$) is reported.

The most important results for temperature indices are presented in Figure 2.5. Almost all catchments experience a marked negative tendency of % of cold days (TX10p) and % of cold nights (TN10p). Moreover, for these indices, the tendency increases for increasing catchment area. This means that the intensity of warming is actually decreasing for increasing area. Trends of the annual maximum of minimum temperature (TNx), annual maximum of maximum temperature (TXx) and the growing season length (GSL) are also reported. The relative trend magnitude for these indices is lower and the spatial variability seems also limited, but they show a significant relationship with flood trends (see Section 2.2.3).

2.2.3 Correlation of decadal tendencies of climate and flood indices

The most relevant results of the spatial correlation analysis among decadal tendencies of ETCCDI indices and decadal tendencies of annual maximum discharge are shown in Figure 2.6, where results are discretized according to mean catchment elevation (circle sizes).

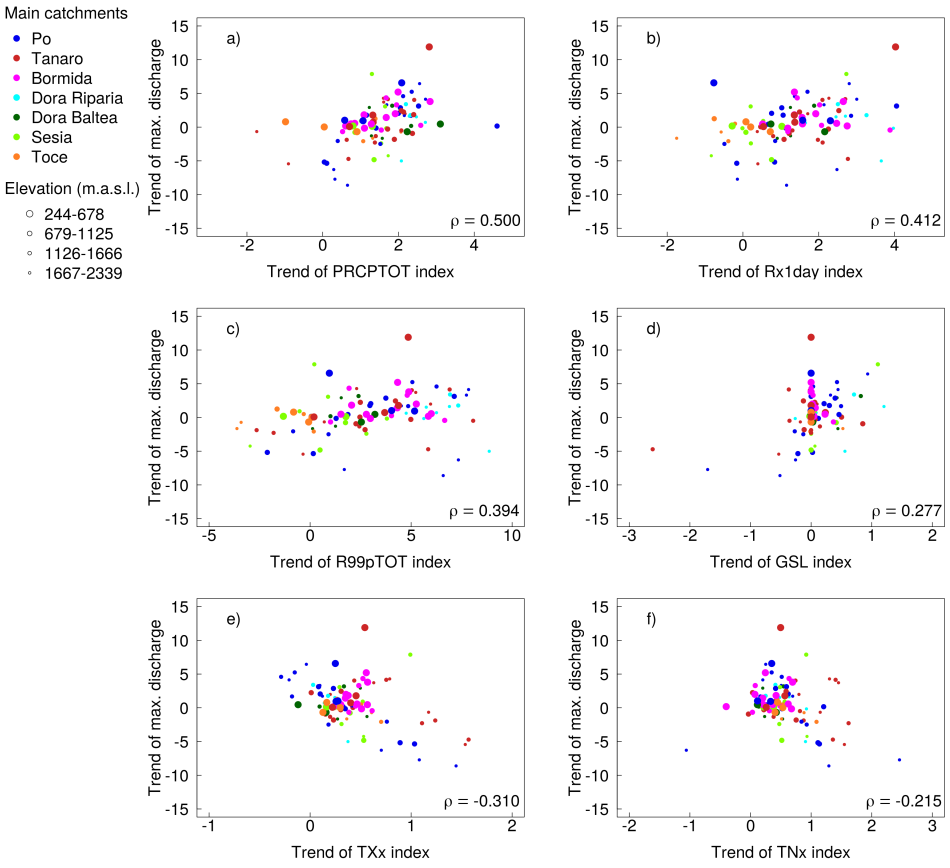


Fig. 2.6 Trends of a) annual total precipitation (PRCPTOT), b) maximum 1-day precipitation (Rx1day), c) annual total precipitation above the 99th percentile (R99pTOT), d) growing season length (GSL), e) maximum value of daily maximum temperature (TXx), f) maximum value of daily minimum temperature (TNx) vs. trends of annual maximum mean daily discharges, discretized by mean catchment elevation, colored as in Figure 1.3a. Spearman's rank correlation coefficients (ρ) are reported.

Looking at precipitation indices, the total annual precipitation (PRCPTOT) tendency has the highest significant correlation with annual flood tendency, as unveiled by a one-sided test at 5% significance level ($\rho = 0.5$), with lower values for the

annual maximum consecutive 1-day precipitation (Rx1day) and the annual total precipitation exceeding the 99th daily percentile (R99pTOT). This points out that the long-term variability of floods seems to have a stronger relation with the tendency of mean precipitation, rather than the one of extreme precipitation.

Temperature indices show weaker correlation with the tendency of floods. In particular, there is a significant positive correlation (ρ = 0.277) with the tendency of the growing season length (GSL) and a negative correlation with the tendency of the maximum value of the daily maximum temperature (TXx) and maximum value of the daily minimum temperature (TNx) (ρ = -0.310 and -0.215, respectively).

2.2.4 Seasonality of flood indices

The timing of observed annual discharge maxima reveals some seasonality over the study area. Figure 2.7 shows that floods are not evenly distributed over the year, in particular in the western and southeastern part of the region, characterized by the presence of Alps and Apennines, where floods typically occur in late spring and in the autumn season, respectively. On the contrary, in the central area and in the northern part of the region, floods are quite evenly distributed and a strong seasonality is lacking.

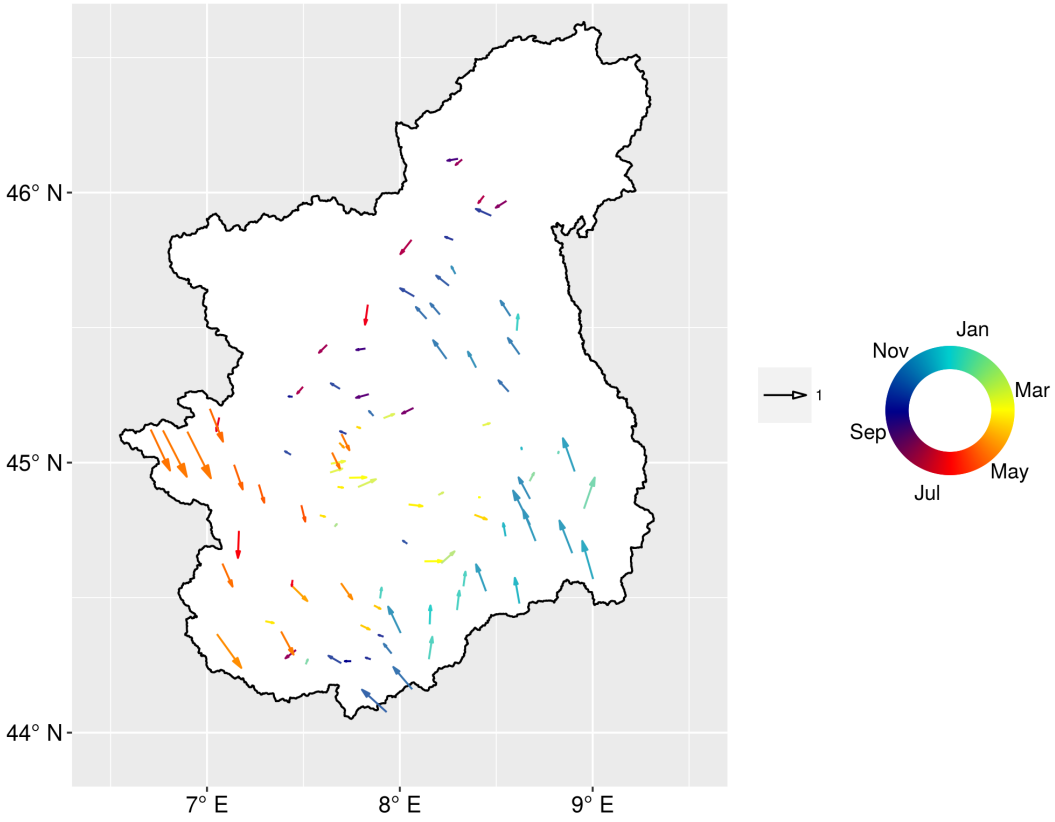


Fig. 2.7 Observed average timing of river floods in North-Western Italy, 2000-2019. Each arrow represents one site. Color and arrow direction indicate the average timing of floods, the length of the arrow indicate the concentration of the date of occurrence, where 1 indicates the flood occurs on the same date.

2.3 Conclusions

The work discussed in this chapter aims at understanding the possible interrelations among the annual maximum daily discharges and standard climate indices (ETCCDI) at the catchment scale in North-Western Italy, over the period 2000-2019. The first analysis is a temporal correlation performed at the annual scale, to assess which indices better explain the interannual variability of discharge maxima. The second analysis is a spatial correlation of the tendencies of annual discharge maxima and climate indices time series aimed at explaining the multi-annual tendency of floods with different potential climate drivers.

It is shown that indices of extreme precipitation such as R99pTOT, R95pTOT, Rx5day, Rx1day are highly positively correlated to annual discharge maxima at the

annual timescale, more than indices reflecting mean precipitation conditions, and for Rx5day and Rx1day the correlation value depends on catchment area. In contrast with this result, the decadal changes of extreme flows may be better explained by the decadal changes of the average precipitation. Temperature indices, instead, are not major controls of annual discharge maxima.

The observed maxima are timely concentrated during late spring over the western part of the region and during fall in the southeastern part, suggesting the diversity of the main runoff generation mechanisms that spatially drive floods in the study area.

Chapter 3

Implementation of TUWmodel and regionalization with PASS

In this chapter, an application of PASS in North-Western Italy, in the context of the Alpine region, is presented, based on the work of Merz et al. (2020). They test the PASS approach in 263 German catchments, to infer parameters of the distributed conceptual hydrological model SALTO. The median model efficiency obtained for training and test catchments is similar to the one obtained by other studies using similar approaches, but Merz et al. (2020) highlight that a unique combination of catchment descriptors controlling model parameters is not found and many regional functional relationships between descriptors and parameters give similar model performances. The aim of this analysis is to evaluate the applicability of the procedure in a diverse region compared to Germany, and the impact of using snow cover information on the simulation results.

3.1 Data

The different types of data used in the analysis are listed as follows:

- **Discharge** data [m^3/s] from the regional stream gauge network managed by the regional environmental protection agency (Arpa Piemonte);
- **Daily precipitation** and minimum and maximum **daily temperature**: from 1961 to 2020, provided by a gridded dataset with cell resolution 0.125° x

0.125°, derived by spatial interpolation of daily observations taken from a dense network of meteorological stations (Optimal Interpolation (OI) Dataset, Arpa Piemonte);

- A **digital elevation model** (DEM) at around 90 m resolution derived from https://www.earthenv.org;
- **Catchment boundaries** for 197 catchments in Piemonte and Valle d'Aosta;
- **Catchment descriptors** associated to the pixels, which are consistent with the Supporting Information S2 in Merz et al. (2020). In particular, 79 descritors are used, divided in 5 categories: Climate (OI dataset); ETCCDI indices (OI); Morphology (DEM); Land Use (Corine Land Cover 2018); Soil (Harmonized World Soil Database).
- **Snow cover** data provided by daily maps of the Normalized Difference Snow Index (NDSI), at 500 m spatial resolution, obtained combining the MODIS products from Terra (MOD10A1) and Aqua (MYD10A1) satellites (Parajka and Blöschl, 2008; Tong et al., 2021). The NDSI values range between 0 and 100 and snow cover is typically considered to be present if NDSI is larger than a threshold, which in this study is assumed to be 40. Values of NDSI larger than 100 represent other cases, like the presence of clouds (NDSI = 250, nsidc.org). The two datasets are combined to reduce the effect of clouds, which prevents getting information about the presence of snow. This means that pixels classified as clouds or missing in Terra are replaced by pixels from Aqua, if these are classified as snow covered or snow free. For each day of observation, the number of snow pixels, no-snow pixels and cloud pixels is counted for each cell of the domain (Figure 3.1) and the percentage of snow pixels is calculated, retaining the value if the percentage of clouds is lower than 60%.

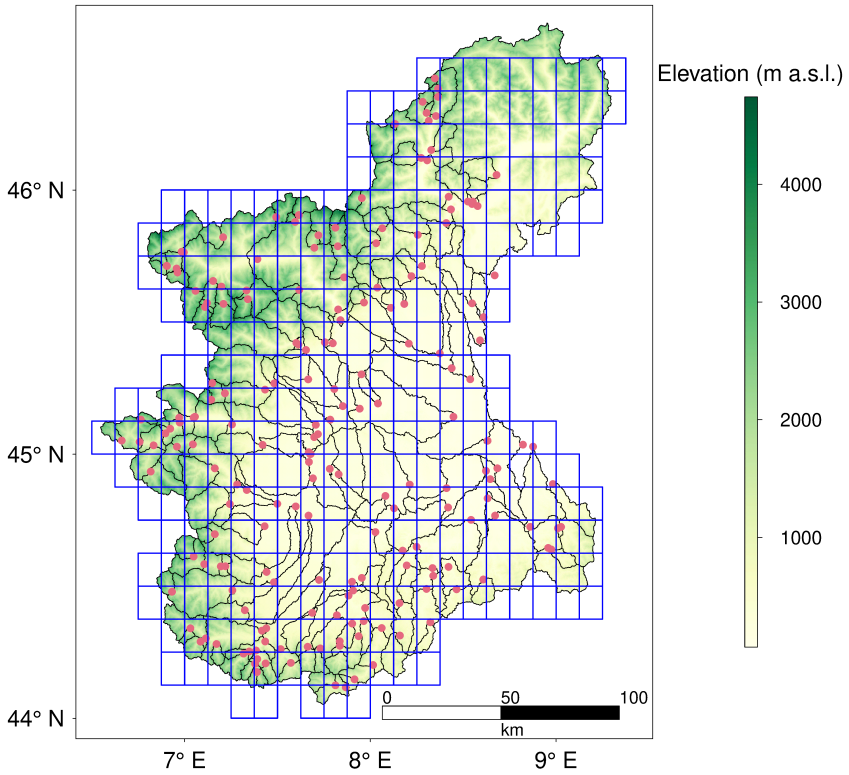


Fig. 3.1 Stream gauges, catchment boundaries and pixels with climate inputs of Piemonte and Valle d'Aosta. The background layer indicates elevation (m a.s.l.).

The catchment descriptors are described in details in Table 3.1

3.2 The TUWmodel

The rainfall-runoff model used for the analysis is the TUWmodel. TUWmodel is a semi-distributed version of the HBV (Hydrologiska Byråns Vattenbalansavdelning) model (e.g., Bergström, 1995; Parajka et al., 2007b), developed by Viglione and Parajka (2018) and available through the R-package *TUWmodel* (https://cran.r-project.org/web/packages/TUWmodel/). Recently, the model has found wide application in Austria (see Tong et al., 2021). The model consists of three routines: snow accumulation and melt, soil moisture routine and flow routing (hillslope and stream), which are described in detail in the following section. It is designed to be run in a semi-distributed way; this structure potentially allows the user to divide the catchment in

Table 3.1 Catchment descriptors used for the training of PASS in North-Western Italy

Category	Variable	Description
Climate Descriptors	MAP	Long-term mean annual precipitation [mm]
	MAT	Long-term mean annual temperature [°C]
	PET	Long-term mean potential evapotranspiration [mm]
	$\frac{PET}{MAP}$	Aridity index (ratio of mean annual potential evapotranspiration and mean annual precipitation) [-]
	Psum2win	Ratio of long-term summer precipitation (May-Oct) and winter precipitation (Nov-Apr) [-]
	R50	Long-term median maximum daily precipitation [mm/d]
	R95	Long-term 95th percentile of maximum daily precipitation [mm/d]
	dRD2D	Long-term mean absolute difference of rainfall amount between two consecutives days [mm]
Standard Climate Indices	FD	Frost days: Annual count of days when TN (daily minimum temperature) < 0°C [days]
(**ETCCDI**)	SU	Summer days: Annual count of days when TX (daily maximum temperature) > 25°C [days]
	ID	Icing days: Annual count of days when TX (daily maximum temperature) < 0°C [days]
	TR	Tropical nights: Annual count of days when TN (daily minimum temperature) > 20°C [days]
	GSL	Growing season length:Annual (1^{st} Jan to 31^{st} Dec in Northern Hemisphere (NH), 1^{st} July to 30^{th} June in Southern Hemisphere (SH)) count between first span of at least 6 days with daily mean temperature TG > 5°C and first span after July 1^{st} (Jan 1^{st} in SH) of 6 days with TG < 5°C.
	TXx	Annual maximum value of daily maximum temperature [°C]
	TNx	Annual maximum value of daily minimum temperature [°C]
	TXn	Annual minimum value of daily maximum temperature [°C]
	TNn	Annual minimum value of daily minimum temperature [°C]
	TN10p	Percentage of days when daily minimum temperature $< 10^{th}$ percentile
	TX10p	Percentage of days when daily maximum temperature $< 10^{th}$ percentile
	TN90p	Percentage of days when daily minimum temperature $> 90^{th}$ percentile
	TX90p	Percentage of days when daily maximum temperature $> 90^{th}$ percentile
	WSDI	Warm speel duration index: Annual count of days with at least 6 consecutive days when daily maximum temperature $> 90^{th}$ percentile [days]
	CSDI	Cold speel duration index: Annual count of days with at least 6 consecutive days when daily minimum temperature $< 10^{th}$ percentile [days]
	DTR	Daily temperature range: Annual mean difference between daily maximum temperature and daily minimum temperature [°C]
	Rx1day	Annual maximum 1-day precipitation [mm]
	Rx5day	Annual maximum consecutive 5-day precipitation [mm]
	SDII	Simple precipitation intensity index [mm/day]
	R10mm	Annual count of days when precipitation $\geq$ 10mm [days]
	R20mm	Annual count of days when precipitation $\geq$ 20mm [days]
	R1mm	Annual number of wet days [days]
	CDD	Maximum length of dry spell i.e. number of consecutive days with precipitation $<$ 1mm [days]
	CWD	Maximum length of wet spell i.e. number of consecutive days with precipitation $\geq$ 1mm [days]
	R95pTOT	Annual total precipitation when daily precipitation is greater than the 95^{th} percentile [mm]
	R99pTOT	Annual total precipitation when daily precipitation is greater than the 99^{th} percentile [mm]
	PRCPTOT	Annual total precipitation in wet days [mm]
Morphology	mean_dem	Mean elevation [m a.s.l.]
	cv_dem	Coefficient of variation of elevation [-]
	mean_slope	Mean slope [%]
	mean_aspect	Mean aspect [°]

Category	Variable	Description
Land use	urban	Percent of grid cell area covered with artificial surfaces [%]
	smallveg	Percent of grid cell area covered with herbaceous, little or no vegetation and open spaces [%]
	agri	Percent of grid cell area covered with agricultural areas [%]
	forest	Percent of grid cell area covered with various types of forests [%]
	wetland	Percent of grid cell area covered with wetlands and lakes [%]
Soil	S_GRAVEL	Mean Content of Gravel in Subsoil (30-100 cm) [%vol]
	S_SAND	Mean Fraction of Sand in Subsoil (30-100 cm) [%wt]
	S_SILT	Mean Fraction of Silt in Subsoil (30-100 cm) [%wt]
	S_CLAY	Mean Fraction of Clay in Subsoil (30-100 cm) [%wt]
	S_REF_BULK_DENSITY	Reference Bulk Density of Subsoil (30-100 cm) [kg/dm^3]
	T _GRAVEL	Mean Content of Gravel in Topsoil (0-30 cm) [%vol]
	T_SAND	Mean Fraction of Sand in Topsoil (0-30 cm) [%wt]
	T_SILT	Mean Fraction of Silt in Topsoil (0-30 cm) [%wt]
	T_CLAY	Mean Fraction of Clay in Topsoil (0-30 cm) [%wt]
	T_REF_BULK_DENSITY	Reference Bulk Density of Topsoil (0-30 cm) [kg/dm^3]
	AWC_LARGE	Percent of grid cell area with large (125-150 mm/m) available water content [%]
	AWC_MED	Percent of grid cell area with medium (75-125 mm/m) available water content [%]
	AWC_SMALL	Percent of grid cell area with small (15-75 mm/m) available water content [%]
	IL_TOP	Percent of grid cell area with impermeable layer located within 80 cm of soil profile [%]
	IL_MED	Percent of grid cell area with impermeable layer located within 80-150 cm of soil profile [%]
	IL_DEEP	Percent of grid cell area with no impermeable layer located within 150 cm of soil profile [%]
	SWR_NOTWET	Percent of grid cell area with dominant annual average soil water regime class: not wet within 80 cm for over 3 months and not wet within 40 cm for over 1 month [%]
	SWR_MEDIUMWET	Percent of grid cell area with dominant annual average soil water regime class: wet within 80 cm for 3 -6 months, but not wet within 40 cm for over 1 month [%]
	SWR_WET	Percent of grid cell area with dominant annual average soil water regime class: wet within 80 cm for over 6 months, but not wet within 40 cm for over 11 months [%]
	SWR_TOTALWET	Percent of grid cell area with dominant annual average soil water regime class: wet within 40 cm for over 11 months [%]
	T_TEXTURE_COARSE	Percent of grid cell area with coarse topsoil texture [%]
	T_TEXTURE_MEDIUM	Percent of grid cell area with medium topsoil texture [%]
	T_TEXTURE_FINE	Percent of grid cell area with fine topsoil texture [%]
	S_USDA_CLAY	Percent of grid cell area with clay subsoil according to USDA classification [%]
	S_USDA_SILTLOAM	Percent of grid cell area with silt and loam subsoil according to USDA classification [%]
	S_USDA_SAND	Percent of grid cell area with sand subsoil according to USDA classification [%]
	T_USDA_CLAY	Percent of grid cell area with clay topsoil according to USDA classification [%]
	T_USDA_SILTLOAM	Percent of grid cell area with silt and loam topsoil according to USDA classification [%]
	T_USDA_SAND	Percent of grid cell area with sand topsoil according to USDA classification [%]
	DRAINAGE_LARGE	Percent of grid cell area belonging to “excessive” and “well” drainage class, based on FAO (2006) [%]
	DRAINAGE_MED	Percent of grid cell area belonging to “moderate” and “imperfect” drainage class, based on FAO (2006) [%]
	DRAINAGE_SMALL	Percent of grid cell area belonging to “poor” and “very poor” drainage class, based on FAO (2006) [%]
	CN1	Mean Curve Number of Type 1
	CN2	Mean Curve Number of Type 2
	CN3	Mean Curve Number of Type 3

sub-basins (e.g., based on elevation zones or other classifications), which separately contribute to the outlet streamflow. The model runs at the daily or shorter time scale. The inputs are:

- P: vector/matrix of precipitation [mm/d]
- T: vector/matrix of daily temperature [°C]
- PET: vector/matrix of potential evapotranspiration [mm/d]
- area: vector of percentage of catchment area for each zone [-]
- param: vector/matrix of parameters

Each climate variable is a matrix n x m where n is the length of the time series and m the number of zones in which the catchment is divided. The catchment outflow is obtained by averaging the contributions from each zone taking into account the area. While model inputs and states are defined over each zone, model parameters can be either constant for the entire catchment or differentiated across the zones. In this analysis, zones correspond to grid cells over the domain, so that model inputs and outputs are estimated for each grid cell and model parameters can be either lumped or differentiated across the cells for each catchment.

3.2.1 Description of model routines

The snow routine is based on the simple degree-day approach and is governed by five parameters. The partition between rain and snow is given by two threshold temperature values, $\mathbf{T_R}$ and $\mathbf{T_S}$, which are used to separate the total precipitation input at time t (P_t) into rainfall ($P_{R,t}$) and snowfall ($P_{S,t}$):

$$P_{R,t} = \begin{cases} P_t & if\ T_t \geq T_R \\ P_t \frac{T_t - T_S}{T_R - T_S} & if\ T_S < T_t < T_R \\ 0 & if\ T_t \leq T_S \end{cases} \tag{3.1}$$

$$P_{S,t} = P_t - P_{R,t} \tag{3.2}$$

Melt starts when air temperature is higher than a threshold $\mathbf{T_M}$:

$$M_t = (T_t - \mathbf{T_M})\mathbf{DDF} \qquad if\ T_t > \mathbf{T_M}\ \&\ SWE > 0 \tag{3.3}$$

where M is the amount of snowmelt per time step, **DDF** is the degree-day factor and SWE is the snow water equivalent. To overcome the problem of catch deficit of precipitation gauges during snowfall, the snowfall is corrected by a snow correction factor **SCF**, so that the snow accumulation process is described by the following equation:

$$SWE_t = SWE_{t\text{-}1} + (\mathbf{SCF} * P_{S,t} - M_t)\Delta t \tag{3.4}$$

where Δt = 1 day.

The soil moisture routine describes runoff generation and soil moisture state changes in the catchment. It involves three parameters: the maximum soil moisture storage **FC**, which represents a saturation condition; **LP**, a parameter representing the soil moisture state above which evapotranspiration is at its potential rate, and a parameter β ruling the non-linear runoff generation process. The portion of rainfall and snowmelt generating runoff (ΔS_{UZ}) is a non linear function of the soil moisture level SM, the parameters FC and β:

$$\Delta S_{UZ,t} = \left(\frac{SM_{t\text{-}1}}{\mathbf{FC}}\right)^{\beta} (P_{R,t} + M_t) \tag{3.5}$$

The level of soil moisture storage at each time step is computed with the water balance equation:

$$SM_t = SM_{t\text{-}1} + P_{R,t} + M_t - \Delta S_{UZ,t} \tag{3.6}$$

If SM exceeds **FC**, surplus water is summed to $\Delta S_{UZ,t}$. The actual evapotranspiration is another flux exiting the system and is function of the potential evapotranspiration (PET) input, **FC** and **LP**:

$$EA_t = \min(PET_t \frac{SM_t}{\mathbf{LP*FC}},\ PET_t) \tag{3.7}$$

Adding this flux to the water balance one gets:

$$SM_t = SM_t - EA_t \tag{3.8}$$

The runoff routine includes the response function, representing runoff routing on hillslopes, and the routing routine, which represents the runoff routing in streams. Seven additional parameters are involved: $\mathbf{L_{UZ}}$, **k0**, **k1**, **k2**, $\mathbf{C_{PERC}}$, $\mathbf{B_{MAX}}$ and $\mathbf{C_{ROUTE}}$. The response function is given by an upper and lower reservoir. Excess

rainfall and snowmelt reach the upper zone reservoir of level S_{UZ}, as highlighted by Eq. (3.9).

$$S_{UZ,t} = S_{UZ,t-1} + \Delta S_{UZ,t} - q_{0,t} - q_{1,t} - C_{PERC} \tag{3.9}$$

The outflow is splitted in three contributions: an outlet from the reservoir ruled by a fast storage coefficient **k1** (Eq. 3.11); a constant percolation flux entering the lower zone reservoir of level S_{LZ} (Eq. 3.13) and, if $S_{UZ} > L_{UZ}$, an outlet ruled by a very fast storage coefficient **k0** (Eq. 3.10). The three contributions subtract water from the upper reservoir state (Eq. 3.12). The water outflow from the lower reservoir is ruled by a slow storage coefficient **k2**. (Eq. 3.14 and Eq. 3.15).

$$q_{0,t} = \frac{S_{UZ,t} - \mathbf{L_{UZ}}}{\mathbf{k0}} e^{-\frac{1}{\mathbf{k0}}} \tag{3.10}$$

$$q_{1,t} = \frac{S_{UZ,t}}{\mathbf{k1}} e^{-\frac{1}{\mathbf{k1}}} \tag{3.11}$$

$$S_{UZ,t} = S_{UZ,t} - q_{0,t} - q_{1,t} - C_{PERC} \tag{3.12}$$

$$S_{LZ,t} = S_{LZ,t-1} + C_{PERC} \tag{3.13}$$

$$q_{2,t} = \frac{S_{LZ,t}}{\mathbf{k2}} e^{-\frac{1}{\mathbf{k2}}} \tag{3.14}$$

$$S_{LZ,t} = S_{LZ,t} - q_{2,t} \tag{3.15}$$

The outflow from both reservoirs is then routed by a triangular transfer function representing the stream runoff routing:

$$B_{Q,t} = \max(\mathbf{B_{MAX}} - \mathbf{C_{ROUTE}}(q_{0,t} + q_{1,t} + q_{2,t}), 1) \tag{3.16}$$

where B_Q is the base of the triangular function, $\mathbf{B_{MAX}}$ is the maximum base at low flows and $\mathbf{C_{ROUTE}}$ is a free scaling parameter. All the parameters involved in the model are summarized in Table 3.2, while the model scheme is reported in Figure 3.2.

Table 3.2 TUWmodel parameters.

Parameter	Description	Units	Typical range
SCF	Snow Correction Factor	-	0.9-1.5
DDF	Degree Day Factor	mm/°C/day	0.0-5.0
T_R	Temperature threshold for rainfall	°C	1.01-3.0
T_S	Temperature threshold for snowfall	°C	-3.0-1.0
T_M	Temperature threshold for snow melt	°C	-2.0-2.0
LP	Parameter related to the evaporation threshold	-	0.0-1.0
FC	Field capacity (i.e. maximum soil moisture storage)	mm	0.0-600.0
β	Non linear parameter for runoff production	-	0.0-20.0
k0	Storage coefficient for very fast response	days	0.0-2.0
k1	Storage coefficient for fast response	days	2.0-30.0
k2	Storage coefficient for slow response	days	30.0-250.0
L_{UZ}	Threshold of storage state triggering very fast response	mm	1.0-100.0
C_{PERC}	Constant percolation rate	mm	0.0-8.0
B_{MAX}	Maximum base at low flows	days	0.0-30.0
C_{ROUTE}	Scaling parameter	$days^2/mm$	0.0-50.0

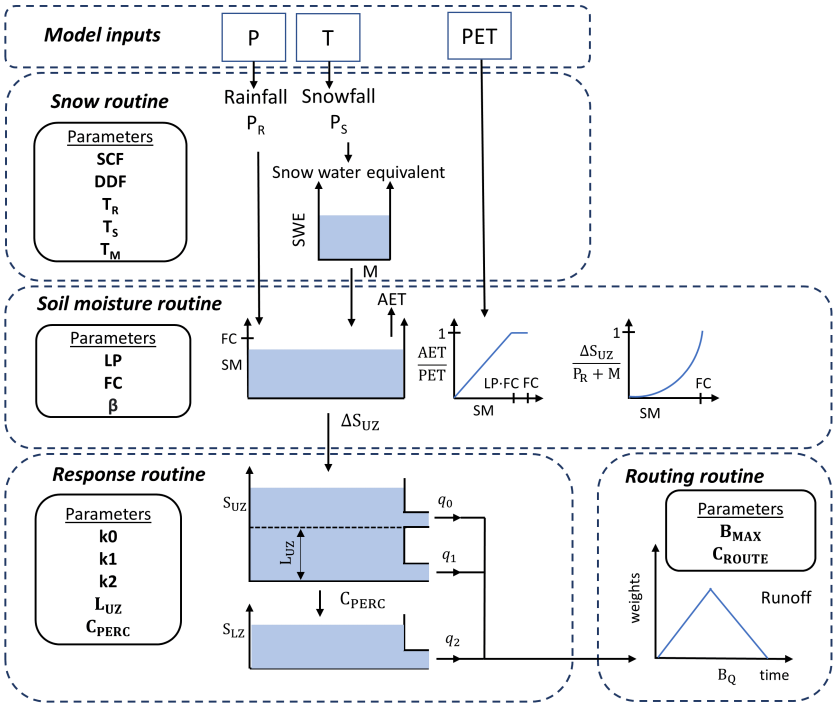


Fig. 3.2 Scheme of TUWmodel.

3.3 Local calibration of TUWmodel

Rainfall-runoff models can be calibrated at gauged sites by minimizing the difference between simulated and observed discharges. Although ideally the procedure can be performed manually by tuning the single parameter values, it is more efficiently

done by machines with high computational performance by adopting automatic optimization algorithms that make use of an objective function.

3.3.1 The objective function

The objective function is the real-valued mathematical function whose value is to be maximized or minimized during the calibration procedure. The choice of the objective function for the calibration of rainfall-runoff models must fit the purpose of the application. The most used objective function in hydrological modelling is the Nash-Sutcliffe Efficiency (NSE), which is a standardized version of the Root Mean Square Error (RMSE). MSE-based metrics are considered a useful tool to reduce simulation errors associated with high flows, as these kind of metrics typically magnify errors in the high flows rather than in low flows (Mizukami et al., 2019). As described in Mizukami et al. (2019), the optimization of many different aspects of the hydrograph is a difficult task; as an example, improving the flow variability representation affects the mean and correlation properties of the simulation. Nevertheless, Gupta et al. (2009) point out that using NSE in calibration leads to an underestimation of the runoff variability, which ultimately means an underestimation and overestimation of high flows and low flows, respectively. For this reason, they propose the Kling-Gupta Efficiency metric, an improvement of NSE which considers three aspects: the daily correlation, the mean flow and the flow variability:

$$\mathrm{KGE} = 1 - \sqrt{(r-1)^2 + (\alpha-1)^2 + (\beta-1)^2} \tag{3.17}$$

where r is the Pearson product-moment correlation coefficient, α is the ratio between the standard deviations of the simulated and observed runoff values and β is the ratio between the means of the simulated and observed runoff. Given the possibility of capturing different aspects of the runoff hydrograph, KGE is widely used in literature of rainfall-runoff modelling (e.g., Kling et al., 2012; Hirpa et al., 2018; Pool et al., 2018; Becker et al., 2019; Liu, 2020; Quintero et al., 2020; Merz et al., 2020) and is selected for the calibration of the rainfall-runoff model used in this Dissertation (i.e. TUWmodel). For the application of PASS in North-Western Italy, two alternative objective functions are considered. On the one hand, the objective function is given by the KGE of discharge values, in which the variability term is given by the ratio of the coefficients of variation (CV) of simulated and observed runoff values, as

specified in Kling et al. (2012):

$$\mathrm{KGE} = 1 - \sqrt{(\mathrm{r}-1)^2 + (\mathrm{CV_{sim}}/\mathrm{CV_{obs}} - 1)^2 + (\mu_{\mathrm{sim}}/\mu_{\mathrm{obs}} - 1)^2} \tag{3.18}$$

where r represents the linear correlation between observations and simulations, $\mathrm{CV_{sim}}/\mathrm{CV_{obs}}$ is the ratio of the coefficients of variation of simulated and observed runoff, $\mu_{\mathrm{sim}}/\mu_{\mathrm{obs}}$ is the ratio of the means of simulated and observed runoff. On the other hand, in order to test the impact of using snow information on the simulation results, the objective function is defined as the arithmetic average of two quantities:

1. the Kling-Gupta Efficiency:

$$\mathrm{KGE} = 1 - \sqrt{(\mathrm{r}-1)^2 + (\mathrm{CV_{sim}}/\mathrm{CV_{obs}} - 1)^2 + (\mu_{\mathrm{sim}}/\mu_{\mathrm{obs}} - 1)^2} \tag{3.19}$$

2. a Snow Cover Efficiency defined as:

$$\mathrm{SC} = 1 - \frac{1}{n \cdot N} \sum_{1}^{n} \left(\sum_{1}^{N} (|\mathrm{SC_{sim}} - \mathrm{SC_{obs}}| \right) \tag{3.20}$$

   where N is the number of days, n the number of cells, $\mathrm{SC_{sim}}$ is 1 if the simulated snow water equivalent (SWE) exceeds 1 mm and 0 otherwise, and $\mathrm{SC_{obs}}$ is 1 if the percentage of snow pixels within the cell is greater than 10% and 0 otherwise.

The model efficiency is so defined:

$$\mathrm{ME} = 0.5 \cdot \mathrm{KGE} + 0.5 \cdot \mathrm{SC} \tag{3.21}$$

3.3.2 Optimization Algorithm: Differential Evolution

An optimization algorithm is a mathematical implementation used for solving an optimization problem. It can be considered as a decision problem where the final goal is to maximize (or minimize) an objective function, which depends on one or more variables (parameters). Almost all optimization algorithms work using an iterative process in order to look for the solution, but they are based on different approaches. A first distinction is to be made among local and global methods. The former look for the optimum in a limited portion of the parameter space and the solution is

affected by the starting point, while the latter are able to extend the research to the entire parameter space. For this reason, global optimization methods are preferred for most optimization problems. Global optimization techniques can be divided in deterministic and stochastic and the latter have found wider applicability for the calibration of hydrological models. A big distinction among optimization methods must be made among direct or exact methods, which consider the gradient vector as a result of the derivation of the objective function, and heuristic methods. The most famous examples of heuristic methods are evolutionary algorithms. These techniques are optimization methods inspired by the process of natural selection and have been used since the 1950s (Mitchell, 1998). Evolutionary algorithms are particularly useful when the optimization function is non-derivable and/or non-continuos. In fact, the research process towards parameter spaces that likey contain the solution is guided by probabilistic rules, but being a non-deterministic and non-exact method, there is no criteria to evaluate the quality of the obtained solutions and the degree of exploration of the parameter space. Heuristic methods are able to provide good quality near-optimal solutions for solving optimization problems but they don't guarantee the convergence to local or global optimal solutions (Rardin and Uzsoy, 2001). A type of evolutionary algorithm is the genetic algorithm, which was invented by John Holland in the 1960s (Holland, 1975). Genetic algorithms have proven to be useful in particular for combinational optimization problems: they apply logical operations on bit strings in order to perform crossover, mutation and selection on a population so that, over successive generations, the members of the population have a higher probability of representing the minimum of an objective function. Another type of evolutionary algorithm is represented by evolution strategies, in which members of a population are identified by floating point numbers and the population is subjected to transformations over successive generations, by using arithmetic operations. Evolution strategies have gained importance over time and in the 1990s an innovative evolution strategy called *differential evolution* (DE) was developed (Storn and Price, 1997). DE is suited for problems requiring the global optimum of a real-valued function of real-valued parameters and does not require a continuos or differentiable function. Since its invention, DE has been succesfully applied in a variety of fields, such as physics and operations research, as detailed in Price et al. (2006). In the context of this Dissertation, DE is the chosen methodology for the local calibration of TUWmodel in North-Western Italy. In particular the R package DEoptim (https://cran.r-project.org/web/packages/DEoptim/, Mullen et al., 2011) is adopted, for consistency with the R function used for the regional calibration of the

rainfall-runoff model (i.e. the PASS function). The structure of PASS itself mimics the iterative nature of DEoptim for the maximization of the objective function.

3.4 Regionalization of rainfall-runoff models

In the recent years, large scale hydrological modelling studies have followed a process-based approach, whereby the spatially distributed parameters are estimated so that they are consistent with the climatic and geomorphological characteristics of the examined study area (Archfield et al., 2015; Clark et al., 2016; Gupta et al., 2014; Mizukami et al., 2017; Paniconi and Putti, 2015). Regionalization of model parameters has become of great importance in hydrology, with the ambition of efficiently predicting runoff in ungauged catchments. In a broad sense, ungauged catchments are defined as the ones without past flow observations or those expecting to experience significant changes in the future (He et al., 2011). Regionalization techniques can be used to estimate either parameters of streamflow statistics (e.g., flood quantiles) or parameters of a continuos rainfall-runoff model for streamflow simulation. All regionalization methods can be divided in two main categories: (1) distance-based and (2) regression-based. While in the first category the entire parameter set of a gauged basin is assumed valid in the ungauged basins, regression methods find a relation linking individual model parameters and catchment characteristics (Parajka et al., 2013). A more detailed look at each group of methods is provided in the following sections.

3.4.1 Distance-based methods

Distance-based regionalization techniques are based on the concept of catchment similarity. The idea behind this type of methods is to identify similar or proxy catchments, i.e. catchments that share common features from a geographical and hydrological point of view.

Geographical similarity is based on the assumption that the hydrological response is likely to vary smoothly in space, indicating geographical proximity as a good measure of catchment similarity (Blöschl, 2005). Geographical distance between a pair of catchments is tipically the distance between the catchment outlets or centroids with the two catchments being defined as the target and the donor catchments (Li

et al., 2009; Randrianasolo et al., 2011; Zvolenskỳ et al., 2008). Target catchments are ungauged catchments that need information to be transferred from donor catchments. Donor catchments are gauged catchments that are considered similar to the target ones. Randrianasolo et al. (2011) use model parameters from neighbouring catchments for ensemble forecast and show good forecasts at the target catchments, with better performances for increasing number of neighbours. Other studies use different procedures that make use of parameters from gauged catchments and apply geo-statistical methodologies (e.g., kriging and inverse distance weighting) for the spatial interpolation, with mixed performance (e.g., Merz and Blöschl, 2004; Parajka et al., 2005; Vandewiele and Elias, 1995; Viviroli et al., 2009). For example, Merz and Blöschl (2004) find a very small decrease of performance using kriging compared to neighbouring Austrian catchments, while Parajka et al. (2005) find a slightly better performance with kriging in the same settings. However, spatial neighbouring catchments do not necessarily have similar hydrological functioning and responses.

A reasonable alternative is to choose the donor catchments based on the climate and catchment characteristics, so that similarity is conceived in the space of catchment descriptors having links with hydrological functioning, the so called hydrological similarity. A typical measure of this similarity, also defined physiographic-climatic similarity, is the root mean square difference of all the catchment descriptors in a pair of catchments, standardised by their standard deviation. Wagener et al. (2007) present metrics that combine similarities based on the catchment structure and hydroclimatic features to represent catchment response behavior. Studies adopting this approach use a wide range of climate and catchment characteristics. McIntyre et al. (2004) use catchment area, stardardised annual average precipitation and base-flow index to define similarity. Other studies use a larger number of characteristics, including geomorphological characteristics (e.g., Parajka et al., 2005; Zhang and Chiew, 2009). Another possibility is to use a model averaging technique, i.e. a weighted combination of the parameter sets from more than one donor catchments (e.g., Goswami et al., 2007; Kim and Kaluarachchi, 2008; Seibert and Beven, 2009).

The distance-based approach for regionalization allows to transfer the entire set of model parameters from the donor to the target catchments, with the advantage of not interfering with the integrity of model parameters as a set (Oudin et al., 2010). The drawback of this approach is related to the fact that also possible correlations and actual interactions between the parameters are entirely transfered to the target catchment.

3.4.2 Regression-based methods

The other macro category of regionalization methods includes regression-based methods. The idea is to relate individual model parameters for continuos streamflow simulation to catchment characteristics. The first attempts of regionalization using regression are described in Nash (1960), who correlate unit hydrographs with catchment characteristics. Three different versions of regression-based methods are discussed in literature: two-step regression, sequential regression and one-step regression, also defined as regional calibration (Parajka et al., 2013).

The two-step regression is the most commonly used method. Regionalization is performed in two steps: (1) calibrate model parameters in each catchment; (2) determine a relationship between model parameters and catchment descriptors. Many studies follow this approach for runoff prediction in ungauged catchments (e.g., Kokkonen et al., 2003; Merz and Blöschl, 2004; Seibert, 1999). The calibration of rainfall-runoff hydrologic models suffers from the problem of existence of multiple parameter combinations resulting in high performance, the so called equifinality issue (e.g., Beven, 1996, 2001).

A slightly different approach is represented by sequential regression. Instead of performing a simultaneous calibration of all model parameters, using this variant allows to calibrate the parameters in a sequential order, from the most identifiable to the least one. In each calibration round, the most identifiable model parameter is put in relation with selected catchment descriptors using a regression function. The parameter value obtained from the regression function is fixed in the next calibration round. This procedure continues until the last parameter is considered (Lamb et al., 2000; Lamb and Kay, 2004; Wagener and Wheater, 2006). Wagener and Wheater (2006) demonstrate the improvement in the parameter identifiability using sequential regression. However, this doesn't guarantee a stronger relationship between parameters and descriptors and the performance can be sometimes worse than two-step regression (Wagener and Wheater, 2006).

A final possibility is the one-step regression or regional calibration. The calibration of model parameters is not performed independently from catchment descriptors and the two steps are implemented concurrently using all objective functions at the same time. The aim is to find more reliable parameters taking advantage of the spatial variability of catchment characteristics. Fernandez et al. (2000) present the method and test it with a water balance model at the monthly scale in the Southern

United States. The results show no improvement compared to the traditional two-step method and the decrease of model performance when moving from calibration to validation is very similar to the one obtained with the two-step approach. Hundecha and Bárdossy (2004) use the same method of Fernandez et al. (2000) but selecting a different study region, model, resolution and descriptors. Other studies appply regional calibration considering homogenous groups (Szolgay et al., 2003) and geostatistical methods (Hundecha et al., 2008; Parajka et al., 2007a).

3.5 The PASS procedure

In the traditional prediction in ungauged basins (PUB) studies reported in the previos section and described in Parajka et al. (2013), usually only the best locally calibrated parameter set in each catchment is used to derive a functional relationship among model parameters and catchment descriptors, neglecting the existence of multiple optimal parameter sets with similar performance, the so called parameter equifinality issue (e.g., Beven and Freer, 2001). Parameter sets with slightly lower local model efficiency can show a better relationship with the descriptors. Based on this evidence, an innovative approach has been recently introduced under the name of PArameter Set Shuffling (PASS) approach (Merz et al., 2020). The objective is to find parameter sets having a good model performance in simulating runoff while preserving a consistent relationship with catchment descriptors at the model unit scale (e.g., grid scale). The underlying assumption is that the regional functional relationship among model parameters and descriptors can be inferred from spatial patterns of lumped parameters calibrated in small catchments (catchment area < 1000 km^2). Indeed, it is expected for spatial patterns of lumped parameters in small catchments to be similar to the spatial pattern of parameters at the model unit scale. The main advantage of using this methodology lies on the fact that it is a data driven approach that does not require any a-priori assumption on the type of relationship among model parameters and descriptors. The search for good functional relationships is performed over the entire parameter space by shuffling the combinations of parameters through an iterative process. Finally, PASS can be theoretically implemented with a variety of machine learning algorithms (e.g., Random Forest, Decision Trees, Artificial Neural Networks (ANNs)).

In this section, the steps of the methodology as formalized in this Thesis are presented. First, the tool used for determining a functional relationship among

local parameters and catchment descriptors is the Decision Tree, through the use of the R function Recursive Partitioning and Regression Trees (`rpart`), described in Breiman et al. (1984). The main idea of Decision Tree is to split the data space (catchments and catchment descriptors) in subgroups, based on the behavior of the response variable (parameters). The algorithm is based on several nodes; at each node catchments are divided according to some thresholds, applied to the descriptors, which minimize the variability of parameters in each subgroup (Figure 3.4).

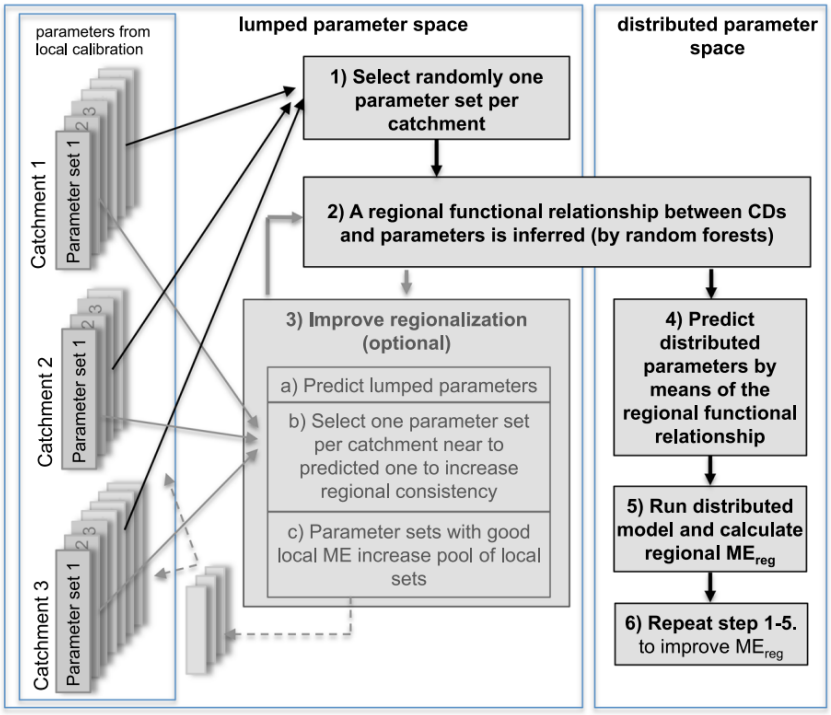


Fig. 3.3 Flow chart of PASS, adapted from (Merz et al., 2020).

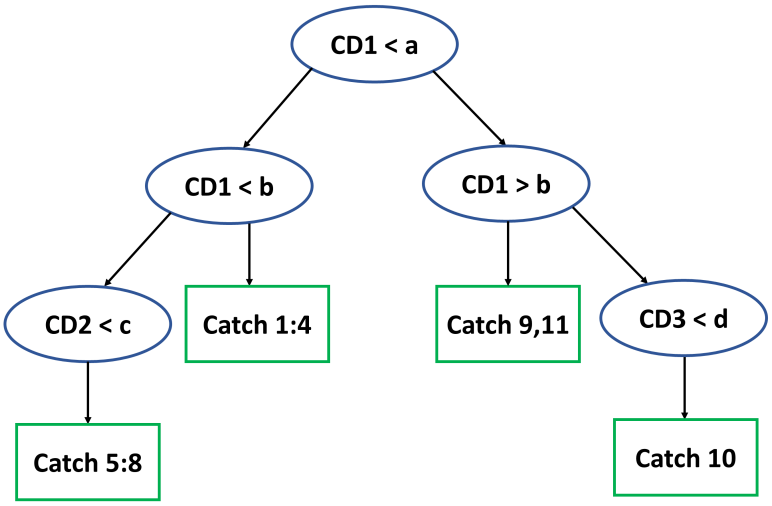


Fig. 3.4 Schematic representation of decision trees.

The method includes different stages (Figure 3.3) that can be overall summarized in three main phases:

1. **Regionalization**: in this phase, the chosen local parameter sets are used to infer the regionalization through the training of the decision tree algorithm. Combinations of parameter sets are performed in order to obtain the best regional model efficiency (ME). To improve regionalization, for each catchment the local parameter set which is more "similar" to the one predicted by the functional relationship at the previous step, is selected as training parameter set, following the "regional consistency" rule. The similarity is quantified by the sum of the differences among the locally calibrated parameters and parameters obtained by the functional relationship. The procedure is repeated until the variance explained by the decision tree is increasing.

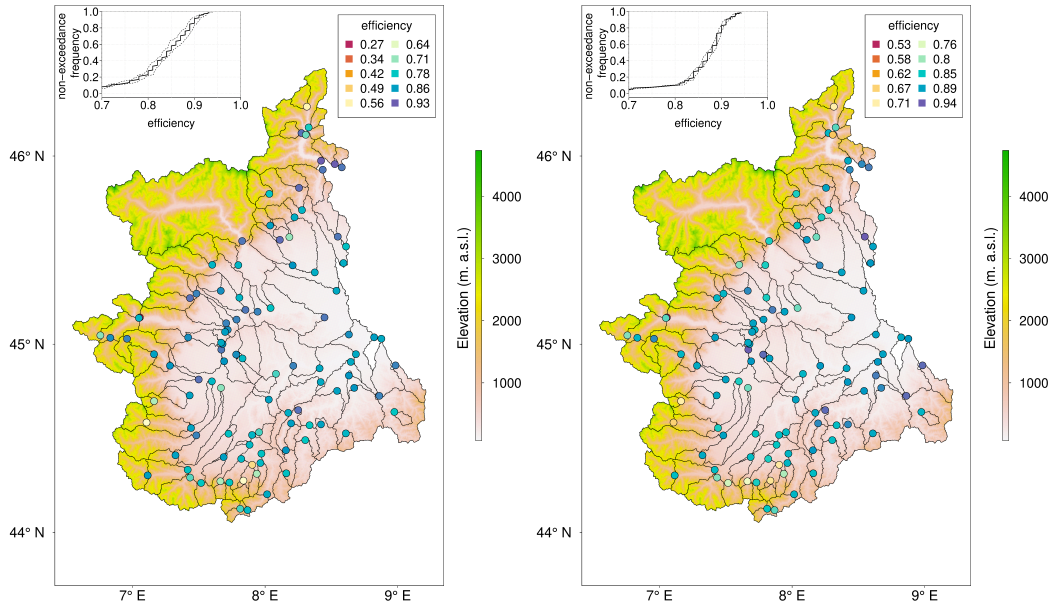
2. **Model run**: in this phase, the model is run with the found regional parameters, both in lumped and distributed mode. Regional parameter sets (lumped) with ME $>$ 0.9MEmax, where MEmax is the highest efficiency value found from local calibration, are added to the pool of local parameter sets and used in the next iterations.

3. **Calculation of mean model efficiency**: in this phase, the mean of the model efficiencies obtained in each catchment by running the model with regional

parameters in distributed mode is calculated. This represents a measure of the performance of the functional relationship. Steps 1-3 are repeated up to a maximum value.

3.6 Regional calibration

The time period considered for testing the PASS procedure with TUWmodel in North-Western Italy is 20 years, from 2000-10-01 to 2020-09-30. Different periods are chosen for the calibration and validation of the method as well as different approaches based on the possible use of snow information in the objective function, in order to provide a robust spectrum of possible calibration results, considering different scenarios (Table 3.3). From a set of 117 sites having daily discharge measurements in the period from 1961 to 2020, only sites with measures in both decades from 2000-10-01 to 2010-09-30 and from 2010-10-01 to 2020-09-30 are selected for local calibration, for a total of 104 stations widespread over the region. As an example, the mean model efficiencies obtained by local calibration in the period 2000-2010, for both cases when snow information is not used and when it is used, are reported in Figure 3.5a and 3.5b.



(a) (b)

Fig. 3.5 a) Local model efficiency (ME = KGE) for 104 sites, obtained without using snow information. b) Local model efficiency (ME = 0.5 · KGE + 0.5 · SC) for 104 sites, obtained using snow information. Note that the two quantities cannot be directly comparable.

Table 3.3 Scenarios for PASS regional calibration.

Calibration period	Validation period	Type
2000-2010	2010-2020	NO SNOW
2000-2010	2010-2020	SNOW
2010-2020	2000-2010	NO SNOW
2010-2020	2000-2010	SNOW
2005-2015	2000-2005, 2015-2020	NO SNOW
2005-2015	2000-2005, 2015-2020	SNOW

Out of the 104 catchments, the ones to be used for the regional calibration with PASS (called *training*) are selected according to the following criteria: local model efficiency (ME) > 0.75 given by at least five independent parameter sets; catchment area < 1000 km^2. This choice is based on two important assumptions that are also reported Merz et al. (2020): 1) the PASS method is based on shuffling parameter sets; 2) the PASS method is based on the idea of deriving regionalization rules for distributed models from lumped parameters obtained from local calibration in small catchments. The remaining catchments, called *test*, are used for validating the procedure. Table 3.4 and 3.5 illustrate the results in terms of regional model

efficiencies for all the scenarios, which consider different calibration - validation periods and type of information used in the objective function.

Table 3.4 Regional model efficiencies obtained with PASS for different calibration periods.

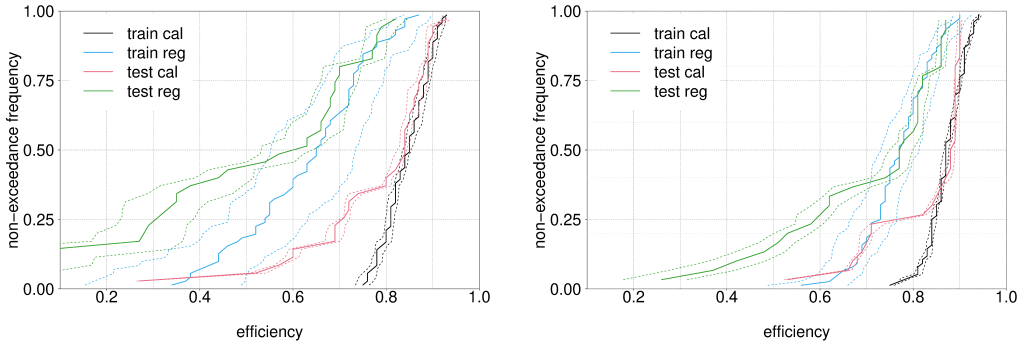
Calibration period	Type	n. training catchments	mean efficiency training catchments	median efficiency training catchments	mean efficiency test catchments	median efficiency test catchments
2000-2010	NO SNOW	70	0.6309	0.6500	0.4785	0.6000
2000-2010	SNOW	75	0.7680	0.7700	0.7017	0.7700
2010-2020	NO SNOW	72	0.6546	0.6600	0.3434	0.5000
2010-2020	SNOW	76	0.7716	0.7850	0.5954	0.7100
2005-2015	NO SNOW	72	0.6597	0.7000	0.4525	0.5750
2005-2015	SNOW	75	0.7771	0.8000	0.6790	0.7600

Table 3.5 Regional model efficiencies obtained with PASS for different validation periods.

Validation period	Type	n. training catchments	mean efficiency training catchments	median efficiency training catchments	mean efficiency test catchments	median efficiency test catchments
2010-2020	NO SNOW	70	0.5819	0.6350	0.3709	0.5500
2010-2020	SNOW	75	0.7472	0.7700	0.6255	0.7200
2000-2010	NO SNOW	72	0.5679	0.5700	0.4975	0.5650
2000-2010	SNOW	76	0.7289	0.7400	0.6804	0.7600
2000-2005, 2015-2020	NO SNOW	72	0.6182	0.7000	0.3803	0.5500
2000-2005, 2015-2020	SNOW	75	0.7478	0.7800	0.6689	0.7350

The results of the regional calibration with PASS in North-Western Italy by including snow information in the model efficiency function are presented in Pesce et al. (2024). In this Dissertation, the value of using snow data to obtain regionally consistent model parameters is further discussed in comparison with the case in which no snow data is used and the objective function coincides with the Kling-Gupta efficiency. For the sake of simplicity, the results for a single decade, i.e. the calibration period 2000-2010, are here further analyzed. The non-exceedance cumulative distributions of model efficiencies during the calibration period (Figure 3.6) show that, as expected, the statistics of the local model efficiencies are much higher compared to the regional model efficiencies, obtained by running the model with predicted distributed parameters. Moreover, it is worth noting that the median regional model efficiency for training catchments when no snow information is used (0.650) is higher compared to the median regional model efficiency for test catchments (0.600). Instead, when snow information is considered, the median performance is the same for training and test catchments (0.770). However, in both cases, when moving to an independent validation period (Figure 3.7), the decrease of regional model efficiencies is lower compared to the decrease of local model efficiencies (both for training and test catchments). This confirms the robustness of the procedure in the prediction of hydrologic variables for different spatial and

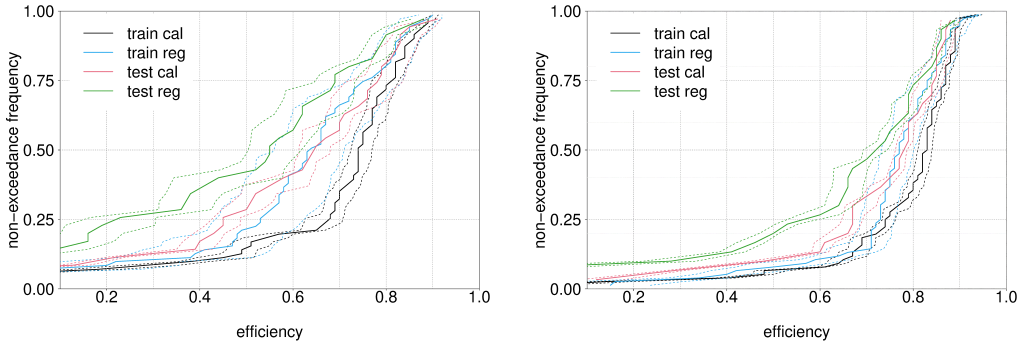
temporal domains. In this context, it is also evaluated the impact of using distributed model parameters, calibrated with snow data, just for streamflow simulation at the catchment outlet (ME = KGE), within the independent validation period. The results show that, compared to the case when snow information is not considered, there is a certain increase in model efficiency, but this is not a significant change (Figure 3.8). The median model efficiency for training catchments increases from 0.635 to 0.645, while the median model efficiency for test catchments increases from 0.550 to 0.575.



(a) No snow information (b) Snow information

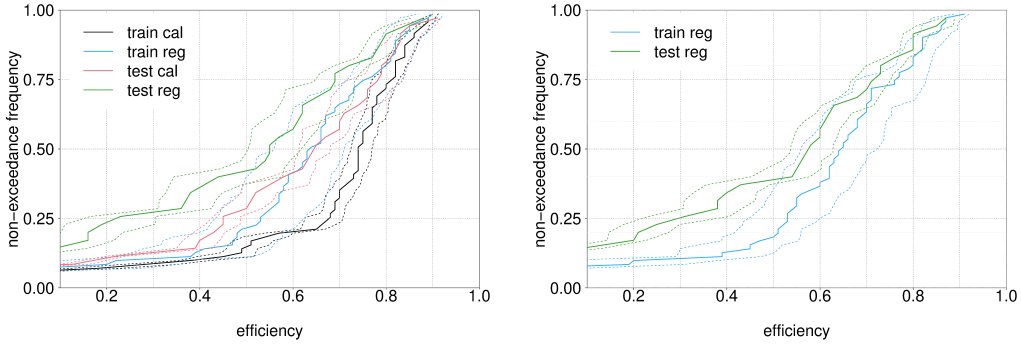
Fig. 3.6 Non-exceedance cumulative distribution of model efficiencies during the calibration period (2000-2010), a) without including snow information in the objective function and b) including snow information in the objective function. The continuos line represents the mean model efficiency, while the dashed lines represent the 10th and 90th percentiles of model efficiencies for 30 parameter sets.

It is of interest to compare the regionally calibrated model parameters, obtained by using PASS, with the locally calibrated parameters (Figure 3.9). On the left panel, the best 30 parameter sets found by local calibration are shown, while the right panel shows the best 30 regionally predicted parameter sets. The parameter values are normalized (the range is between 0 and 1) and the associated model efficiency is reported. Despite the fact that model efficiency given by regionally calibrated parameters is lower compared to the model efficiency associated with locally calibrated parameters, it is worth noting that most of the locally calibrated parameters span the whole possible range of values, while parameters found by regional calibration show a much smaller spread. This is expected due to the issue of equifinality, as discussed in the previous section. Appendix D reports this comparison for all the 75 training catchments regionally calibrated in the period 2000-2010 using snow information.



(a) No Snow information (b) Snow information

Fig. 3.7 Non-exceedance cumulative distribution of model efficiencies during the validation period (2010-2020), a) without including snow information in the objective function and b) including snow information in the objective function. The continuos line represents the mean model efficiency, while the dashed lines represent the 10th and 90th percentiles of model efficiencies for 30 parameter sets.



(a) No Snow information (b) Snow information

Fig. 3.8 a) Non-exceedance cumulative distribution of model efficiencies during the validation period (2010-2020) without including snow information in the objective function. b) Non-exceedance cumulative distribution of regional model efficiencies during the validation period (2010-2020), obtained using parameters calibrated with snow information for discharge simulation only (ME = KGE).

The spatial distribution of parameters estimated using PASS is also investigated. Figure 3.10a shows regional maps of TUWmodel parameters obtained when snow cover information is not used. Similarly, Figure 3.10b shows the same result obtained by using snow cover information. It is worth noting that the spatial patterns of regionally distributed parameters seem to be in line with hydro-meteorological and landscape features of the area (see e.g. the Alpine range or the Po Valley). By

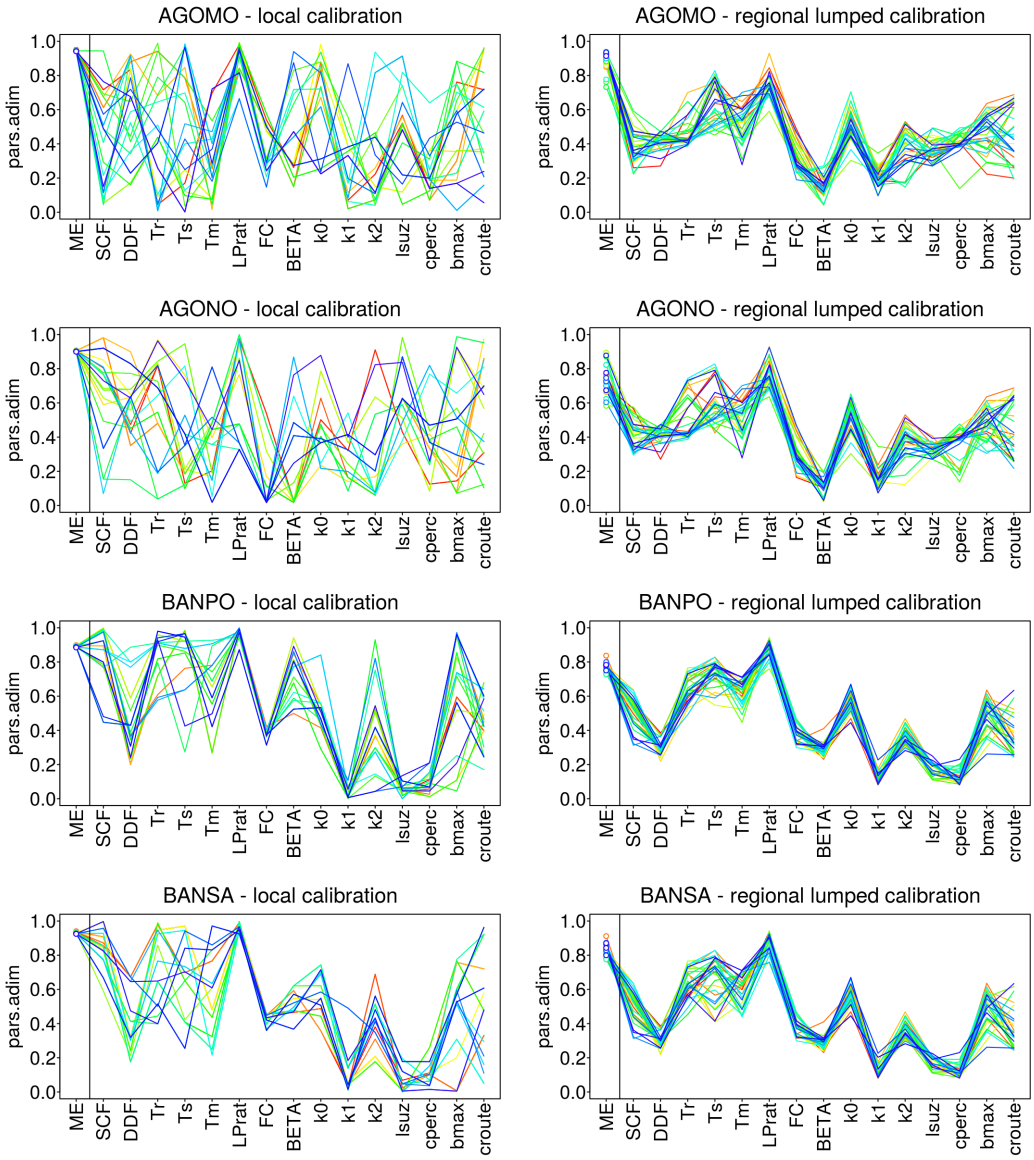
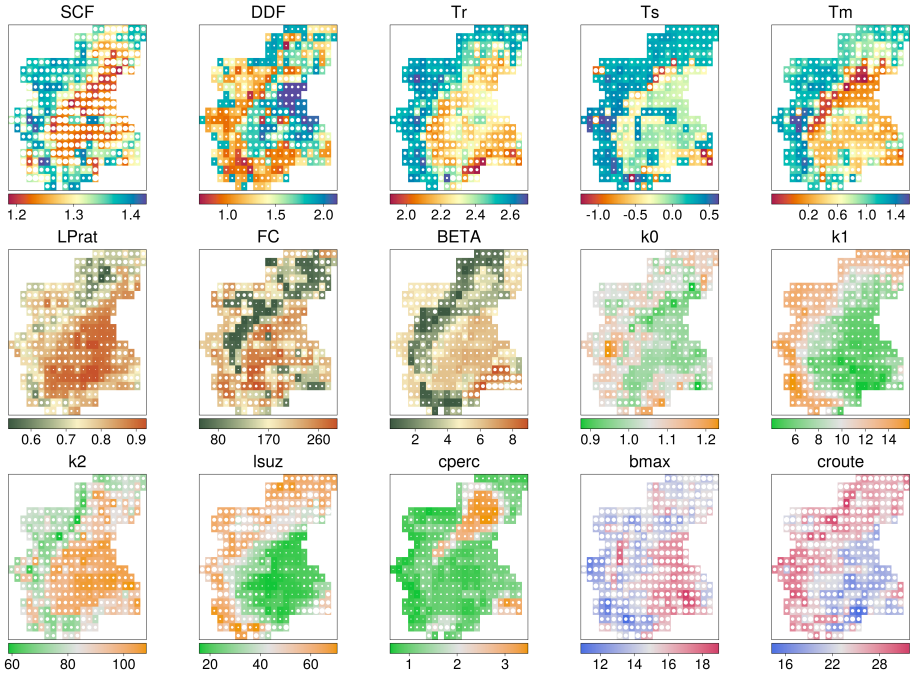
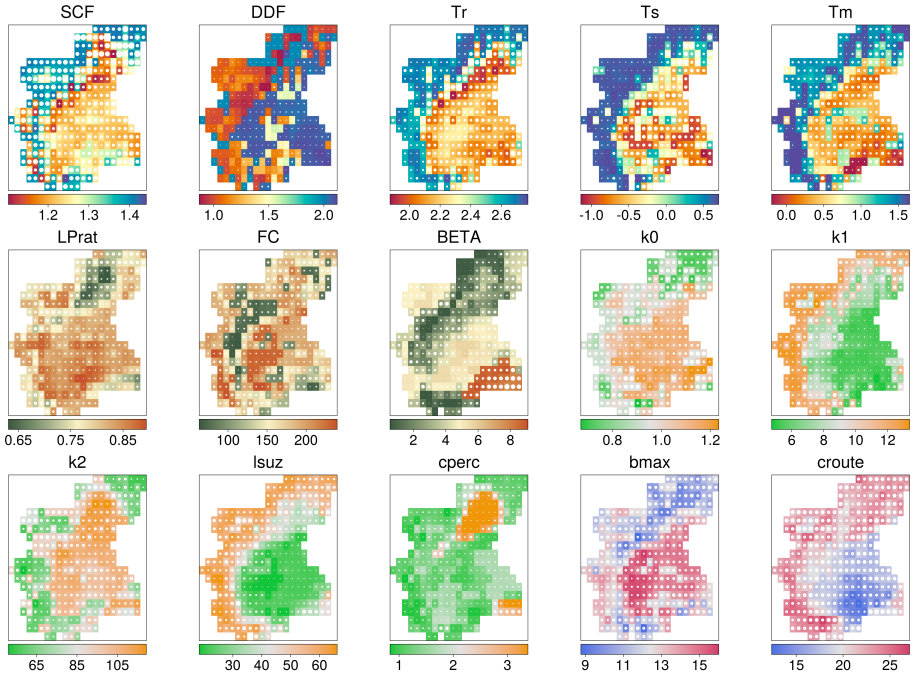


Fig. 3.9 Locally calibrated parameter sets (30 sets, left) vs regionally calibrated (lumped) parameter sets (30 sets, right) obtained with PASS for 4 sites.

comparing Figure 3.10a with Figure 3.10b, it is quite evident that snow parameters (SCF, DDF, Tr, Ts, Tm) show a better coherence in terms of spatial pattern when snow information is considered for the calibration, compared to the case when it is not used. In fact, the gradient of grid values when moving from the central Po Valley to the Alpine range (basically going north and west) is sharper and values are more consistent with what can be expected by physical processes (Figure 3.10b).



(a) No snow cover



(b) Snow cover

Fig. 3.10 a) Regional maps of distributed TUWmodel parameters obtained with the PASS procedure, without considering snow cover information. The color indicates the mean value while the circle indicates its variability among 30 regionalizations. b) Regional maps of distributed TUWmodel parameters obtained with the PASS procedure, by considering snow cover information. The color indicates the mean value while the circle indicates its variability among 30 regionalizations.

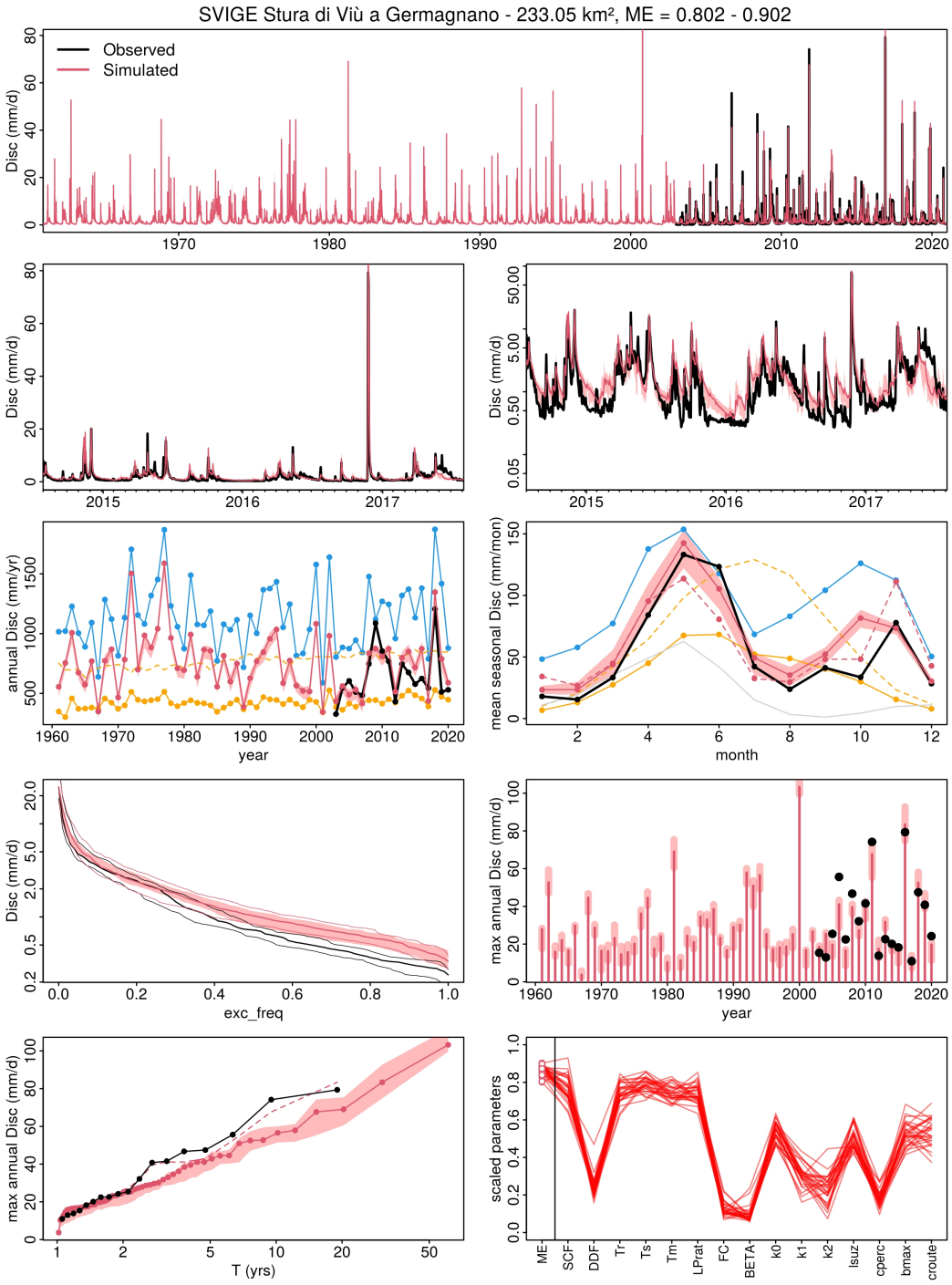


Fig. 3.11 Observed discharges vs. simulated discharges with regional PASS parameters of TUWmodel, obtained by calibration over the period 2000-2010 without using snow information ($ME = KGE$), for Stura di Viù River catchment at Germagnano.

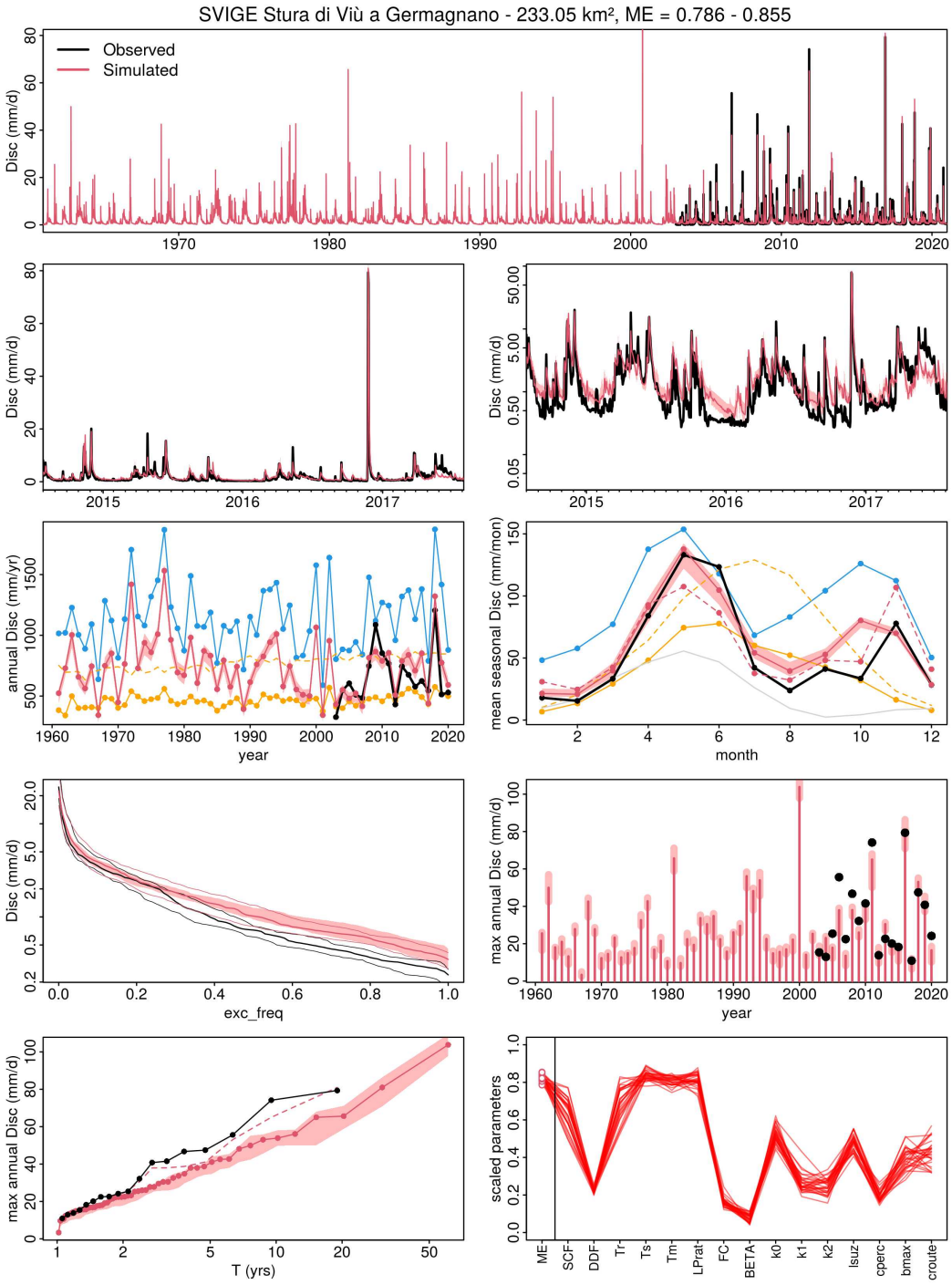


Fig. 3.12 Observed discharges vs. simulated discharges with regional PASS parameters of TUWmodel, obtained by calibration over the period 2000-2010 using snow information ($ME = 0.5 \cdot KGE + 0.5 \cdot SC$), for Stura di Viù River catchment at Germagnano.

Figure 3.11 and 3.12 report the comparison between simulated and observed runoff signatures for a single site, i.e. Stura di Viù River catchment at Germagnano, located in the central-western part of the region with a mean elevation of 1759 m a.s.l. The simulated runoff is obtained by regionally calibrating the model over the period 2000-2010 without using snow cover information (Figure 3.11) and by using snow cover information (Figure 3.12). For a detailed description of all the shown signatures please refer to Figure Legend of Appendix E, where all the results from the regional calibration of 75 training catchments using snow information are reported. It can be noticed that using snow information in the calibration procedure doesn't significantly improve the overall ability of the model in reproducing streamflow signatures (e.g. annual flows, seasonality). Nevertheless, when snow information is used, the range of simulated streamflow is narrower, indicating a more consistent model behavior.

3.7 Conclusions

This chapter deals with the application of the PASS procedure in North-Western Italy, exploiting a decision-tree based algorithm. The results show that the PASS procedure can be efficiently applied in the North-Western Italy territory, with similarly good model performances as for the German case study. Despite differences in the results based on the choice of the model efficiency function and the calibration/validation period, it is observed that the median model efficiency for training and test catchments in calibration mode is around 0.6-0.7 and the decrease of efficiency when moving to an independent validation period is minimal. Including snow information in the model efficiency function increases model performance in reproducing streamflow but not significantly and streamflow signatures (e.g., annual flows or seasonality) don't seem to be better captured. Nevertheless, the range of values assumed by simulated streamflow when snow information is considered is narrower, indicating a more consistent model behavior and less uncertain results. This is also confirmed by the higher spatial coherence of the snow parameter values obtained by regional calibration using snow, compared to the case when snow information is not included.

Chapter 4

The hydroPASS R package

In this chapter, a newly developed R package is presented, with the aim of making the PASS procedure open and reproducible and its application with different hydrologic models more flexible. An example of application of the package with U.S. database is also shown.

4.1 The package structure

The package aims at providing a consistent and reproducible version of the PASS methodology, in order to be able to apply it with different distributed or semi-distributed hydrological models and forcing datasets. For this purpose, the Git version controlling system is adopted, together with its graphical interface GitHub (https://github.com/alviglio/hydroPASS), which allows all contributors to work on the package, make updates and release versions of it (Figure 4.1). In the current version (v0.1-beta.2), the source package has the following structure:

- **DESCRIPTION** file, which contains general information about the package (e.g., version, authors, etc.);
- Directory **src/** which contains the Fortran code for the model SALTO;
- Directory **R/** which contains R code for the functions SALTO, PASS, catchDescrip, topology;
- Directory **data/** which contains some data for the examples provided as support of the functions;

- Directory **man/** with the help documentation for the code and data (i.e., that builds the manual written for the package, available in Appendix B);
- Directory **vignettes/** which contains a vignette about obtaining and processing U.S. input data for running PASS.

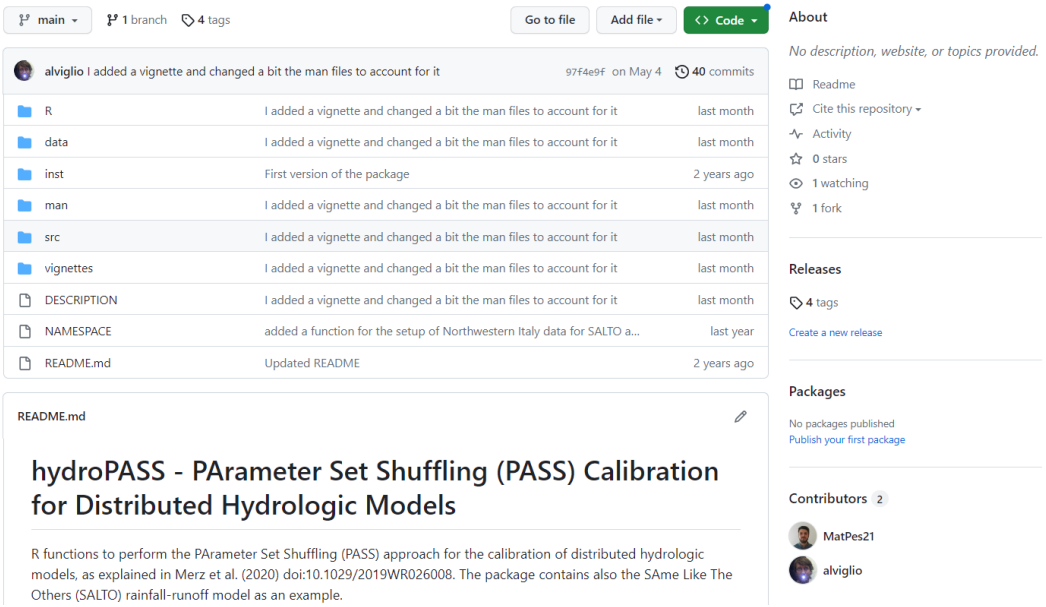


Fig. 4.1 Screenshot of the GitHub web page of the hydroPASS package.

From the source package, installation packages are built for Windows, Linux and Mac operating systems and are also available through GitHub (e.g., download hydroPASS 0.1-2.zip on Windows, or hydroPASS 0.1-2.tar.gz on Linux or Mac). The current version of the package allows to run PASS with different hydrologic models, through the definition of a model efficiency function (objective function). The objective function can be of low or high complexity, for example including snow cover, but it returns a single model efficiency value that must be optimized by PASS. For example, the procedure can be applied with the model TUWmodel, which is available in R at: https://cran.r-project.org/web/packages/TUWmodel. The following code provides an example on how to couple it with PASS by using a model efficiency function:

```
library(TUWmodel)
?TUWmodel

ME.TUWmodel <-
  function(param,      # vector or matrix parameters for catchment i
           cat.number, # number of catchment i
```

```
            grdname,    # list of grid names associated to all catchments
            prec,       # matrix of precipitation data for all grids
            airt,       # matrix of air temperature data for all grids
            ep,         # matrix of potential evapotranspiration...
            area,       # list of proportion of catchment area contained
                        #  in each grid for all catchments
            disc,       # matrix of runoff discharge for all rivers
            snow,       # matrix of snow cover for all grids
            iwarmup=303) {
  GRDNAME <- grdname[[cat.number]]
  AREA <- area[[cat.number]]
  PREC <- prec[, as.character(GRDNAME)]
  AIRT <- airt[, as.character(GRDNAME)]
  EP <- ep[, as.character(GRDNAME)]
  DISC <- disc[, cat.number]
  SNOW <- snow[, as.character(GRDNAME)]
  if (!is.null(dim(param))) param <- t(param)
  # global variables are in the function
  simu <- TUWmodel(prec=PREC, airt=AIRT, ep=EP,
                   area=AREA, param=param)
  swesim <- (simu$swe > 1)[-c(1:iwarmup),]
  sweobs <- (SNOW > 10)[-c(1:iwarmup),]
  # Snow Cover Efficiency
  SC <- mean(1 - apply(abs(swesim - sweobs), 1, mean, na.rm=T), na.rm=T)
  # Runoff Kling-Gupta Efficiency
  simu <- simu$q[-c(1:iwarmup)]  # remove the warming period
  simu[is.na(simu)] <- -999
  obse <- DISC[-c(1:iwarmup)]  # remove the warming period
  r <- cor(simu, obse, method='pearson', use='pairwise.complete.obs')
  beta <- mean(simu)/mean(obse, na.rm=TRUE)
  gamma <- (sd(simu)/mean(simu))/(sd(obse, na.rm=TRUE)/mean(obse, na.rm=TRUE))
  kgeQ <- 1 - sqrt((r - 1)^2 + (beta - 1)^2 + (gamma - 1)^2)
  me <- 0.5*kgeQ + 0.5*SC  # mean of Runoff Kling-Gupta Efficiency
                           #  and Snow Cover Efficiency
  if (is.na(me)) me <- -999
  return(me)
}

# Running PASS
run01 <- PASS(Y=train.parameters,
              X.cat=cat.CD,
              X.grd=grd.CD,
              grd2cat=sapply(topology, function(x) x$grd.name),
              model.eff.fn=ME.TUWmodel,
              lower=c(SCF=0.9, DDF=0.0, Tr=1.01, Ts=-3.0, Tm=-2.0,
                      LPrat=0.0, FC=0.0, BETA=0.0, k0=0.0, k1=2.0, k2=30.0,
                      lsuz=1.0, cperc=0.0, bmax=0.0, croute=0.0),
              upper=c(SCF=1.5, DDF=5.0, Tr=3.0, Ts=1.0, Tm=2.0,
                      LPrat=1.0, FC=600.0, BETA=20.0, k0=2.0, k1=30.0, k2=250.0,
                      lsuz=100.0, cperc=8.0, bmax=30.0, croute=50.0),
              options=PASS.options(maxLoops=100, nGroups=10,
                                   REGloops=5, sampling='random'),
              prec=data118cat305pxl$prec,
              airt=data118cat305pxl$tmean,
              ep=data118cat305pxl$pet,
```

```
            area=sapply(topology, function(x) x$grd.weightCat),
            grdname=sapply(topology, function(x) x$grd.name),
            disc=qobs, snow=data118cat305pxl$snow)
```

4.1.1 SALTO

SAme Like The Others (SALTO) is a distributed conceptual rainfall-runoff model. The model represents a soil moisture accounting scheme and is similar to other well known models, such as the HBV model. The version implemented in the package consists of 21 parameters and 1 soil moisture layer. However, the structure of the model allows to consider many soil layers, increasing the number of parameters and so the complexity of the model. A further description of the model and its implementation in the package is provided in the manual (see Appendix B).

4.1.2 PASS

As already described, the PArameter Set Shuffling (PASS) algorithm derives regional relationships between model parameters and catchment descriptors from observed patterns of locally calibrated parameters and descriptors through a machine learning procedure, and provides regionally consistent parameter sets. In this section, the PASS function implemented in hydroPASS using R code is described in details. An example of application with the model SALTO is provided in the manual (see Appendix B).

The PASS function inputs are shown here:

```
# PASS
PASS <- function(Y,
                 X.cat,
                 X.grd,
                 grd2cat,
                 model.eff.fn,
                 lower,
                 upper,
                 options=PASS.options(),
                 ...)

PASS.options <- function(maxLoops=100, nGroups=10, REGloops=5,
                         generalized.mean.power=-1, proportion.max.eff.update=0.95,
                         sampling='random', optim.subset.cat=0.7)
```

`Y` is a list of dataframes with locally lumped calibrated model parameters or previously obtained PASS output, which is used as starting point for further search of regionally consistent model parameters. `X.cat` is a matrix or data.frame of catchment descriptors, `X.grd` is a matrix or data.frame of model unit/pixel descriptors, `grd2cat` is a list of model unit/pixel names belonging to each catchment, `model.eff.fn` is the objective function to be optimized (maximixed), `upper` and `lower` are two vectors identifying scalar real lower and upper bounds for each parameter to be optimized, `PASS.options` is a vector containing a list of options for the PASS approach and ... refers to other arguments to be passed to `model.eff.fn`, such as climate input variables. The elements read by PASS.options function are here clarified. `maxLoops` is the maximum number of iterations for each PASS run; `nGroups` is the number of groups over which regional optimisation is run, in order to avoid falling in local optima; `REGloops` is the number of loops for the regional consistency algorithm, as specified in Section 3.5 (i.e. for each catchment the local parameter set which is the closest (most similar) to the one predicted by the regional relationship at the previous step is selected as training set); `generalized.mean.power` is the value of the power (p) exponent of the generalized mean of the efficiencies obtained in all catchments with the fully distributed regional model; `proportion.max.eff.update` sets a condition for adding regionalized lumped parameters to the pool of training sets; `sampling` allows to choose among two possible running modes: "random", so that the selection of model parameters for the regionalization is random or "optim", so that model parameters are selected from the ones obtained in the previous run of PASS, except for the proportion indicated in optim.subset.cat; `optim.subsect.cat` indicates the proportion of catchments from which the parameters for the regionalization are selected randomly.

At the beginning of the algorithm run, a list of n different groups is set and, at the first run, a null starting model efficiency is associated to each group. If the "random" sampling is selected, at each iteration the parameter sets are randomly selected from a constantly updated pool of possible parameter sets, a regional functional relationship is determined and a regional mean model efficiency ($\overline{ME}_{\text{reg}}$) is calculated. This $\overline{ME}_{\text{reg}}$ is saved in the output of the group with the currently lowest ME, prioritizing the null ones. This implies that in the first n iterations of a run, the null ME are substituted with a real scalar number; from the n+1 iteration, the new ME substitutes the old one in the group with the lowest ME value. In case the new ME is lower than the previous MEs, it is simply discarded.

The option "optim" requires at least a first run at random before being executed. If "optim", a random group is first selected. Then, the parameter sets belonging to that group found in the previous PASS run are partially substituted by performing a random parameter selection on a proportion of catchments equal to optim.subset.cat. The new set of selected parameters is used to infer the regionalization. The calculated $\overline{ME}_{\text{reg}}$ is compared with the overall efficiency of the same group. This is done in order to optimize single groups.

The output of the regionalization procedure, which is saved for each group, includes the following elements:

- `overall.eff`: the mean of the MEs obtained running the model with regionalized parameters in a distributed mode in each catchment;
- `selected.parameters`: a matrix containing the locally calibrated parameter sets used to infer regionalization.
- `regionalized.parameters`, divided in `cat.par.pred`, which identifies the predicted lumped parameters for each catchment and `grd.par.pred`, which identifies the predicted distributed parameters over the reference grid.
- `cat.eff.lump`: vector containing the MEs obtained with the predicted lumped parameters in each catchment.
- `cat.eff.dist`: vector containing the MEs obtained with the predicted distributed parameters in each catchment.

An additional list is also stored called `train.parameters.updated`, which contains, for each catchment, matrices with locally lumped model efficiency and calibrated model parameters used for the training of the procedure, possibly updated by PASS.

4.1.3 catchDescrip

The function catchDescrip reads in input a list of matrices of catchment descriptors, a spatial polygon data frame of catchments and an array of pixel/model unit coverage of the catchment area for each catchment (weights) to create data of descriptors both at grid and catchment scale. Basically, descriptors are averaged over catchments by applying a weighted mean over the grid values, using the above mentioned weights.

4.1.4 topology

The function topology reads in input a spatial polygon data frame of catchments and a raster of digital elevation model (dem) to derive the topology of catchments. In particular, a sink filling of the dem and the calculation of flow directions is performed within the function. The output includes four pieces of information for each catchment:

- grd - index of the pixels covering the catchment;
- effarea – pixels coverage inside the catchment;
- flowto – index of the downstream pixel;
- level – level of the pixel (1 = no pixel upstream, 2 = pixels of level 1 upstream, etc.);

A description of both catchDescrip and topology functions is provided in the manual (Appendix B).

4.2 hydroPASS with U.S. data

The data provided in the hydroPASS package are U.S. hydrologic database pertaining the Contiguos United States of America (CONUS), i.e. the area corresponding to the continental portion of the USA with the exception of Alaska. The main reason for this choice is the recent publication of the CAMELS dataset (Newman et al., 2015), a very large catchment scale hydrometeorological dataset of 671 catchments covering the country and with limited impact of human activities. The dataset is supplemented with data of catchment attributes (Addor et al., 2017). Together with CAMELS, other datasets of climate and geomorphological characteristics are used. In particular the following data are downloaded:

- Discharge data and catchment shapefiles: CAMELS data set;
- Rainfall, temperature and potential evapotranspiration: NCA-LDAS Noah-3.3 Land Surface Model L4 Daily 0.125 x 0.125 degree V2.0;
- DEM: https://www.hydrosheds.org/;

- Land Use: National Land Cover Database;
- Soil: NACP MsTMIP - Unified North American Soil Map.

A Vignette describing the source of the data and how to manipulate them in order to use the PASS procedure is included in the package and reported in Appendix C.

Chapter 5

Identification, characterization and classification of runoff events

In this chapter, the regionally calibrated TUWmodel is used for the identification, characterization and classification of runoff events in North-Western Italy, with the aim of obtaining an event typology from simulated runoff for this area, which is coherent with the correspondent event typology that is provided by observed discharges at gauged sites.

First, the event separation method proposed by Giani et al. (2022) is adopted to extract events for the 117 sites with available discharge measurements over the period 1961-2020, and observed and simulated runoff characteristics (i.e. runoff coefficient, event duration, event peak time, event peak and event volume) are spatially compared. The same procedure is then performed for a wider catchment dataset, i.e. the European Catchments and River Network System (ECRINS), to get distributed statistics of runoff event characteristics by considering events in around 700 catchments.

Secondly, the characterization and classification of runoff events following the framework described in Tarasova et al. (2020) is performed. By using several climatic indicators, the first-order controls of event runoff response are identified in a wide variety of catchments. This reveals distinct regions with homogeneous event type frequency and the spatio-temporal pattern of event type occurrence as obtained by the simulated streamflow is coherent with the event typology from observed discharges, reflecting the hydroclimatic conditions of the area.

5.1 Runoff Event Identification

The runoff event extraction is performed using the methodology for rainfall-runoff event identification proposed by Giani et al. (2022). The method is based on a time series analysis that simultaneously considers rainfall and discharge time series without making any a priori assumption about the baseflow separation. In fact, this is produced a posteriori, by connecting the delimiters of the identified streamflow events. The method can be efficiently applied at different time resolutions (i.e. hourly or daily), as long as the resolution is sufficient to capture the time delay between precipitation and runoff response. By looking at the simultaneous evolution in time of rainfall and streamflow time series, the method is able to identify events as "system" realizations. This technique is based on the Detrending Moving-average Cross-correlation Analysis (DMCA)-based method for the estimation of the catchment response time (Giani et al., 2021) and for this reason is named DMCA-Event Separation Routine (DMCA-ESR).

5.1.1 DMCA-Event Separation Routine (DMCA-ESR)

The DMCA-ESR is based on the DMCA-based method for the estimation of the catchment response time (Tr) (Giani et al., 2021). Tr is used for a hydrologically meaningful selection of rainfall-runoff events that implies the grouping of all the rainfall contributions building a streamflow event. According to the mentioned method, Tr is defined as the average lag between the center of mass of rainfall and the center of mass of streamflow across all the events. The method provides an estimate of Tr by identifying the time scale for which the correlation among the two time series is the highest. DMCA-ESR looks at the evolution in time of rainfall and streamflow, allowing to define the event as a system. An important feature of this method is its flexibility, as it can be applied to different time resolutions by adjusting only one parameter, i.e. the minimum rainfall intensity which is considered significant at a given time resolution.

In this section, the steps required by the method to perform the extraction of rainfall-runoff events from a continuos time series are described in detail.

1: Search of typical catchment response time

The first step is to find the characteristic Tr for the analysed catchment using the DMCA-based method. The estimate depends on the temporal data resolution but it is essential that the resolution is high enough to capture the delay between rainfall and runoff response. Tr is particularly important in order to evaluate the rainfall-streamflow interactions when contributions in the two series are grouped at the window scale associated to Tr, i.e. L_{min}. Once an estimate of Tr is produced ($Tr = \frac{Lmin-1}{2}$), the steps of DMCA-based method are re-applied using the window L_{min} to obtain timeseries of rainfall and streamflow fluctuations, which are used to identify rainfall and streamflow events (Figure 5.1).

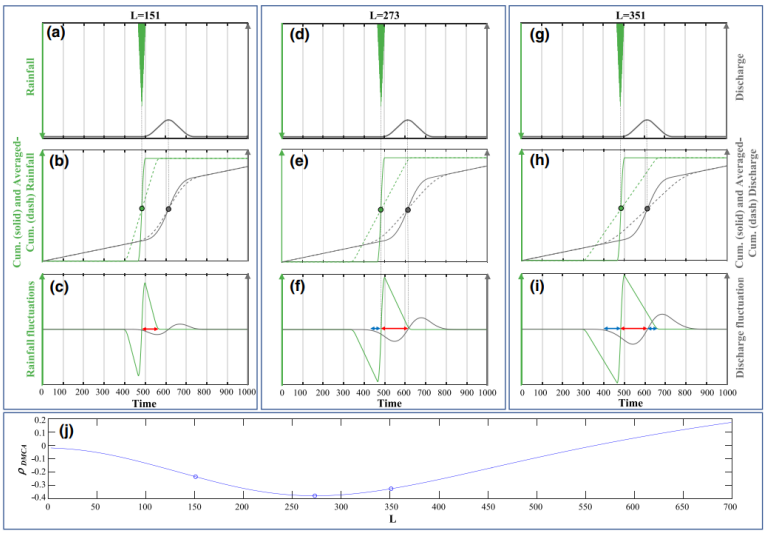


Fig. 5.1 (a–i) Graphic representation of steps (I, II, III) of DMCA-based methodology for moving average window lengths L = 151, L = 273, L = 351. Green lines relate to rainfall, gray lines to streamflow. The red (blue) arrows underline periods of negative (positive) bivariate fluctuations. (j) DMCA-based correlation coefficient variability with L, with circles showing correlation for the three window lengths above. Adapted from Giani et al. (2021).

Step 2: Setting a rainfall fluctuation tolerance

This step is used to adjust the time series analysis technique for the long dry or steady period required to break down the different contributions into different events. When the fluctuation between cumulative timeseries and averaged-cumulative timeseries

is zero, the original time series is steady, i.e. no event is occurring. These steady state periods can be used as break points for event identification if the period has a minimum duration of L_{min}. Due to the fact that any rainfall contribution within a window equal to L_{min}, even small, would prevent the discretization of the timeseries in different events, a tolerance for rainfall fluctuation ($R_{fluct.tol}$) is introduced. Any rainfall fluctuation smaller than the tolerance is set to zero. To define the rainfall fluctuation tolerance, it is considered the maximum absolute rainfall fluctuation between cumulative rainfall and averaged-cumulative rainfall generated by a given rainfall intensity, R_{min} (Giani et al., 2022):

$$R_{\text{fluct.tol}} = \frac{R_{\text{min}}}{L_{\text{min}}} \cdot \frac{L_{\text{min}} - 1}{2} \tag{5.1}$$

where $\frac{R_{\text{min}}}{L_{\text{min}}}$ is the absolute increment in fluctuation per time step and $\frac{L_{\text{min}}-1}{2}$ is the maximum number of timesteps for which there is an increase in fluctuation. In Eq. (5.1) the only parameter is R_{min}, which must correspond to the smallest rainfall intensity considered significant at the specific time resolution. Following the indication provided in Giani et al. (2021), in the range of values between 0.1 and 1 mm/hour the event selection is not very sensitive to the value of R_{min}. For the conversion of R_{min} at daily resolution, the Intensity-Duration-Frequency approach (IDF) is suggested (i.e. finding the intensity associated with the 24h duration, considering the same return period). Following this approach, the range 0.1-1 mm/h converts to 0.02-0.2 mm/h. For the analysis conducted in the context of this Dissertation, the value of 0.02 mm/h is selected.

Step 3: Selection of cores of rainfall and streamflow events

The core of an event is defined as the time interval within each event when both rainfall and discharge are unsteady. This implies that rainfall and streamflow events are associated to each other if they share the same core. The core of the rainfall-streamflow event is identified by looking at the bivariate fluctuation (Figure 5.2b), that is the product of rainfall fluctuations, corrected with the tolerance, and streamflow fluctuations (Figure 5.2a). When bivariate fluctuations are zero, this reveals that either rainfall or streamflow fluctuations (or both) are zero. This way, all the small rainfall contributions which do not generate runoff are excluded a priori. Zero bivariate fluctuations are break points for core identification only if two consecutive

time steps have zero values, to prevent that a change of fluctuations sign can be misinterpreted. The identified delimiters of the core represent the starting point to set the delimiters of rainfall and streamflow events.

Step 4: Identification of the end of rainfall events

The first guess is that the end of the rainfall event coincides with the end of the core. The position of the end of the rainfall event is then adjusted considering three cases:

1. The core has ended because rainfall fluctuations are zero and the rain at that point is zero. This is the most common case, explained by the fact that fluctuations are different from zero for a duration equal to L_{min} after the rainfall ends. In this situation, the delimiter is put backward in time, at the first time step of non-zero rainfall (Figure 5.2c).
2. The core has ended because rainfall fluctuations are zero and the rain at that point is different from zero. This case is generated by the introduction of the rainfall fluctuation tolerance; the delimiter is put forward in time, at the first time step when rainfall is lower than R_{min} (Figure 5.2d).
3. The core has ended because streamflow fluctuations are zero; in this situation the delimiter is put backward in time until the rainfall becomes larger than R_{min} (Figure 5.2e).

In the case when both rainfall and streamflow fluctuations are zero at the end of the core, the delimiter for the end of rainfall event is put following the approach for cores which end because of zero rainfall fluctuations.

Step 5: Identification of the beginning of rainfall events

The first guess is that the beginning of the rainfall event coincides with the beginning of the core. The position of the beginning of the rainfall event is then adjusted considering three cases:

1. Just before the core started, rainfall fluctuations are zero and the rain at that point is also zero. This is the most common case, explained by the fact that

fluctuations are different from zero for a duration equal to L_{min} before the rainfall starts. In this situation, the delimiter is put forward in time, at the first time step of non-zero rainfall (Figure 5.2f).

2. Just before the core started, rainfall fluctuations are zero and the rain at that point in time is different from zero. This case is generated by the introduction of the rainfall fluctuation tolerance. The delimiter is put backward in time, at the first time step when rainfall is lower than R_{min} (Figure 5.2g).

3. Just before the core started, the streamflow fluctuations are zero; in this situation the delimiter is put backward in time until the rainfall becomes lower than R_{min} (Figure 5.2h).

Also in this case, when both rainfall and streamflow fluctuations are zero, the delimiter for the beginning of rainfall event is put following the case of zero rainfall fluctuations.

Step 6: Check on rainfall events

During the definition of rainfall events, there is the possibility of finding unrealistic rainfall events which end even before starting. For this reason, at this stage all the rainfall events which show their beginning after their end, together with the events which are not delimited by dry periods or periods of rainfall lower than R_{min}, are discarded.

Step 7: Identification of the end of streamflow events

The position of the end of the streamflow event is adjusted considering two cases:

1. The core has ended because the rainfall fluctuations are zero. In this situation, as first guess it is assumed that the end of the streamflow event coincides with the end of the rainfall event. As it is expected to find the end of the streamflow event after the rainfall event, the delimiter is put forward in time at the end of positive streamflow fluctuations (Figure 5.2i).

2. The core has ended because the streamflow fluctuations are zero. In this situation, as first guess it is assumed that the end of the streamflow event coincides with the end of the core. As it is expected to have positive streamflow fluctuations at the end of the event, the delimiter is put backward in time when the streamflow fluctuations become positive (Figure 5.2j).

Again, if both rainfall and streamflow fluctuations are zero, it is followed the procedure indicated for zero rainfall fluctuations.

Step 8: Identification of the beginning of streamflow events

The position of the beginning of the streamflow event is adjusted considering two cases:

1. Just before the core started the rainfall fluctuations are zero. In this situation, as first guess it is assumed that the beginning of the streamflow event coincides with the beginning of the rainfall event. As it is expected to find the start of the streamflow event after the start of the rainfall event, the delimiter is put forward in time when the streamflow fluctuations become negative (Figure 5.2k).

2. Just before the core started, the streamflow fluctuations are zero. In this situation, as first guess it is assumed that the beginning of the streamflow event coincides with the beginning of the core. The delimiter is put forward in time when the streamflow fluctuations become negative (Figure 5.2l).

Again, if both rainfall and streamflow fluctuations are zero, it is followed the procedure indicated for zero rainfall fluctuations.

Step 9: Check on streamflow events

During the definition of streamflow events, there is the possibility of finding unrealistic streamflow events which end even before starting. For this reason, at this stage all the streamflow events which show their beginning after their end, together with their

corresponding rainfall events, are discarded. Other categories of events are discarded here:

- events for which the beginning (or end) delimiter of the streamflow event occurs earlier than the beginning (or end) delimiter of the rainfall event.
- streamflow events, and the corresponding rainfall events, which do not start with negative fluctuations and they do not end with positive ones.

Step 10: Check on rainfall-streamflow events

Given the way rainfall-streamflow events are defined, there might be some overlaps between contiguous rainfall events or between contiguous streamflow events. At this stage, overlapping events are lumped together into one single event.

5.1.2 Observed and simulated runoff event characteristics for gauged sites

In this section, the runoff event extraction is performed over the 117 stations, distributed over the region, for which streamflow data are available within the period from 1961-01-01 to 2020-12-31. The runoff event extraction algorithm (see section 5.1.1) is run both with observed and simulated discharge data. Parameters that are regionally calibrated with the PASS procedure during the period 2000-2010, by using snow cover information, are used to run the simulations. In particular, all the 30 groups of parameters are used to run the simulations in the 117 catchments and the results in terms of model efficiency are saved. In the next step, only the group showing the highest model efficiency is used to run the simulation in each catchment and the simulated series from 1961 to 2020 is saved. Finally, the event separation is performed using the simulated streamflow and precipitation for all the catchments. The precipitation input for event separation is the sum of liquid precipitation (i.e. rainfall) and snow melt, according to the snow routine embedded in TUWmodel. An example of event separation on a continuos streamflow series for a single site in North-Western Italy is shown in Figure 5.3.

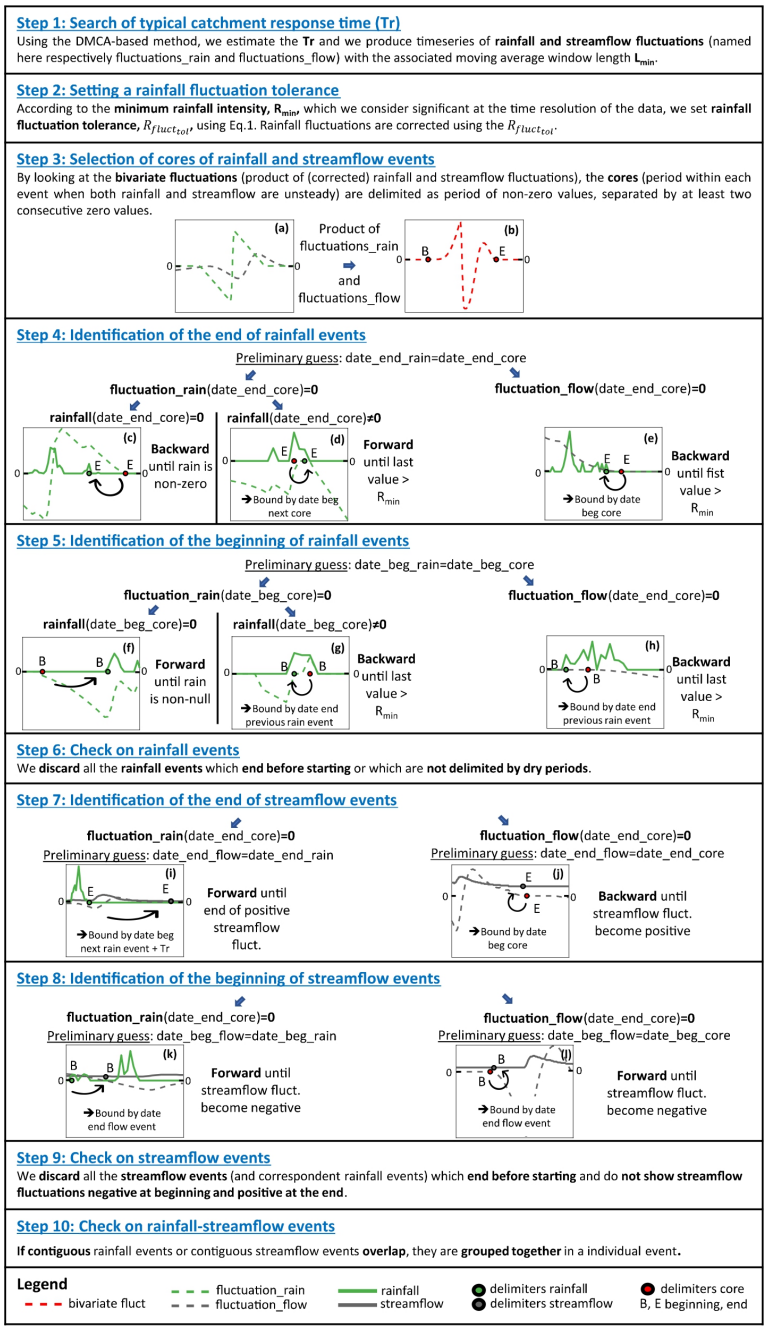


Fig. 5.2 Summary of DMCA-ESR. The legend applies to all the individual subfigures (5.2a-5.2l). Adapted from Giani et al. (2022).

Among the entire set of 117 sites, 108 catchments for which both the observed and simulated runoff event extraction works properly are selected. For each event, the following runoff event characteristics are calculated:

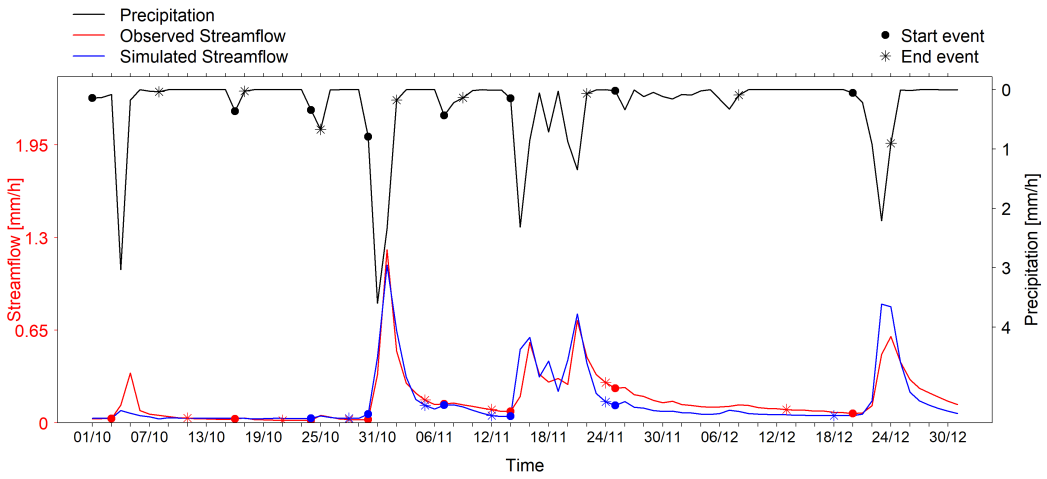


Fig. 5.3 Example of event separation performed on Agogna river catchment at Momo (204 km^2) in North-Western Italy, using the method described in Giani et al. (2022).

1. Runoff coefficient [-]: it is defined as the ratio between the direct runoff and the precipitation volume;
2. Event duration [d]: it is defined as the length in time of the streamflow event;
3. Event peak time [d]: it is defined as the time step that corresponds to the peak flow;
4. Event peak flow [mm/h]: it is defined as the maximum discharge level reached during the event;
5. Event volume [mm]: it is defined as the direct runoff of the event.

First, among the set of observed and simulated events, the events showing a runoff coefficient $>$ 1 and/or event duration $>$ 120 days are discarded. Runoff coefficients $>$ 1 can rise for two main reasons: either a quota of precipitation is given by convective storms that are not well captured by spatially distributed rain gauges or the event is driven by intense, not captured, snowmelt. Overall, the number of events with runoff coefficient greater than 1 is quite small, as reported in Table 5.1. The statistics for the discarded events are reported in Table 5.2. It is worth noting that, compared to

Table 5.1 Statistics of the number of observed and simulated events having rc > 1.

Type	Min	1st Quantile	Median	Mean	3rd Quantile	Max
Observed	0	0	1	1.66	2	25
Simulated	0	0	0	0.49	0	12

Table 5.2 Statistics of the number of observed and simulated events which are discarded.

Type	Min	1st Quantile	Median	Mean	3rd Quantile	Max
Observed	0	0	1	1.71	2	25
Simulated	0	0	1	3.09	3	65

the observed events, the number of simulated events having runoff coefficient greater than one is lower, while the model identifies more events having extra long duration.

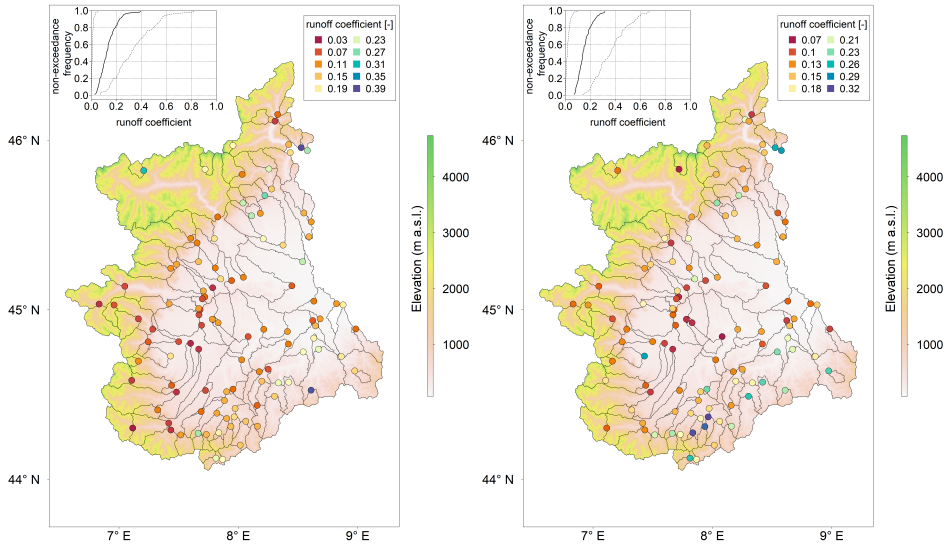
Secondly, the set of observed events is analyzed to get the number of events captured by the event separation on the simulated time series. The criterion used for the identification deals with the potential intersection among the observed and simulated event. If this is greater or equal than 40% of the maximum length among the length of observed and simulated event, the observed event is saved in the list of the captured events. At the end, the ratio among captured and total observed events in each catchment is calculated. The statistics for this variable is provided in Table 5.3.

Table 5.3 Statistics of the fraction of captured observed events over total observed events [%].

Min	1st Quantile	Median	Mean	3rd Quantile	Max
11	37	44.5	43.7	50	76

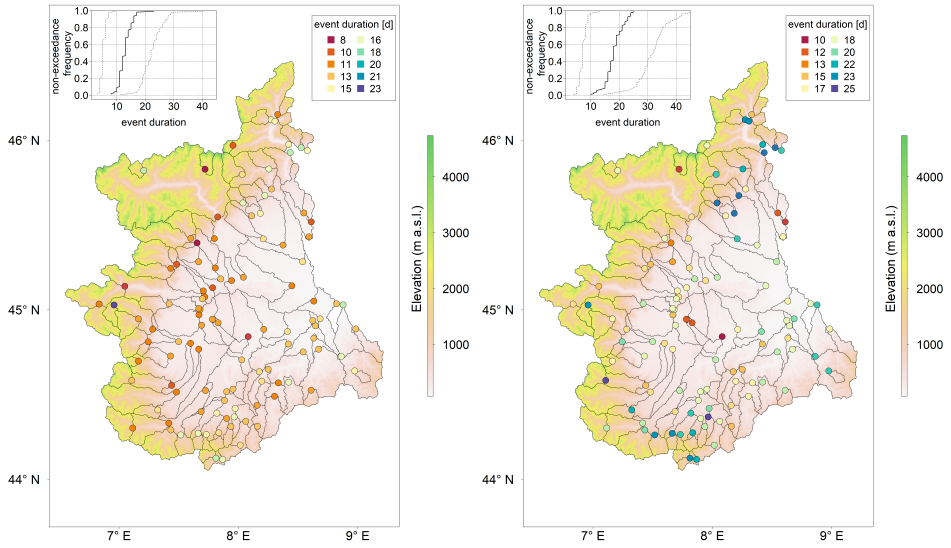
The matching observed and simulated events are finally compared for all the catchments in terms of event characteristics, i.e. runoff coefficient, event duration, event peak time, event peak, event volume (Figure 5.4 - 5.8).

The results show some clear spatial patterns of runoff characteristics over the region. The highest values of runoff coefficient (Figure 5.4) are found in the north-eastern part of the region (e.g., San Bernardino river catchment), in Valle d'Aosta and in the sourthern part of the region (Bormida, Orba and Tanaro river catchments), at medium-high elevation (Figure 5.9a). The values are lower in the central Po Valley and in the western part. Overall, the simulations quite overestimate the mean runoff coefficient (Figure 5.9a), especially in the South (Figure 5.4), but the spatial pattern is captured and the range of the mean is maintained between 0 and 0.4



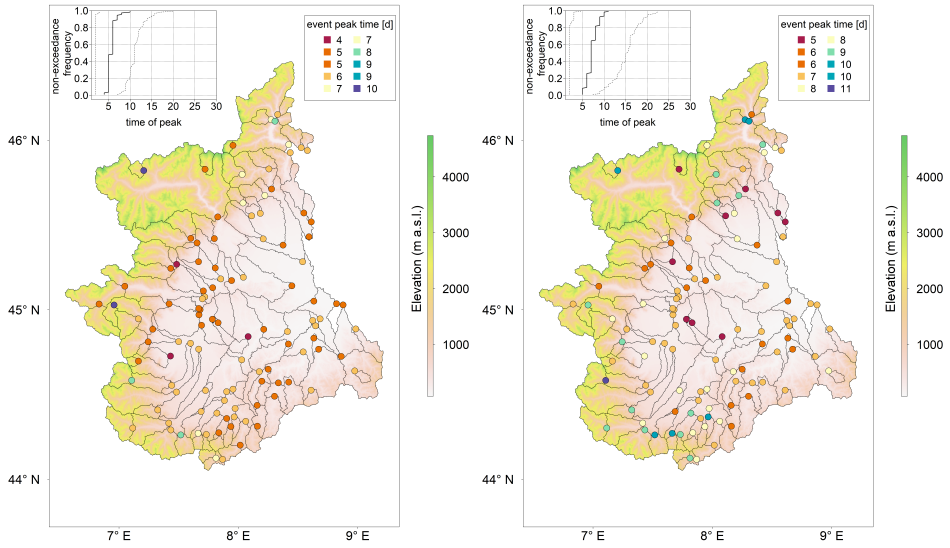
(a) Observed (b) Simulated

Fig. 5.4 Mean runoff coefficient for a) observed runoff events and b) corresponding simulated runoff events by using the regionally calibrated TUWmodel, in 108 catchments. In the inset, the continuos line represents the non-exceedance probability of mean runoff coefficient, while the dashed lines represent the non-exceedance probability of the 10th and 90th percentiles of runoff coefficient.



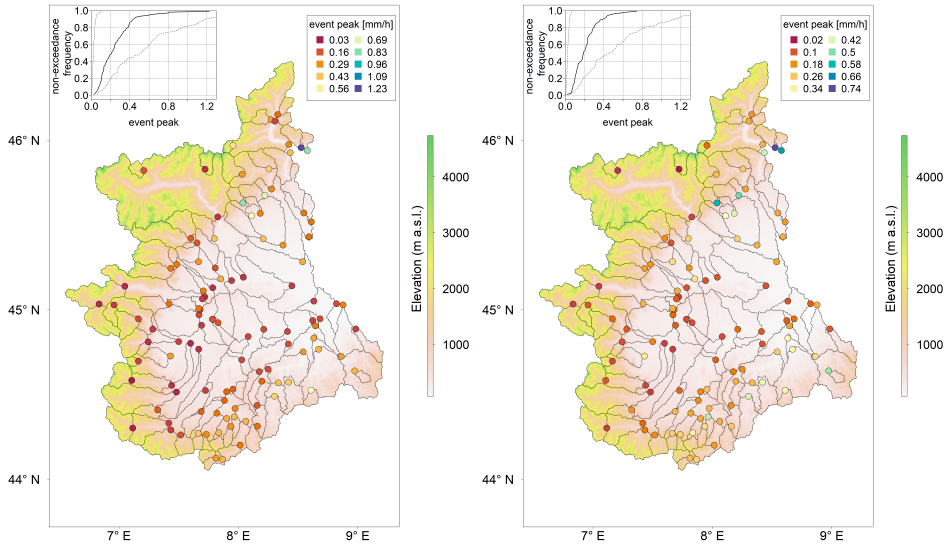
(a) Observed (b) Simulated

Fig. 5.5 Mean event duration for a) observed runoff events and b) corresponding simulated runoff events by using the regionally calibrated TUWmodel, in 108 catchments. In the inset, the continuos line represents the non-exceedance probability of mean event duration, while the dashed lines represent the non-exceedance probability of the 10th and 90th percentiles of event duration.



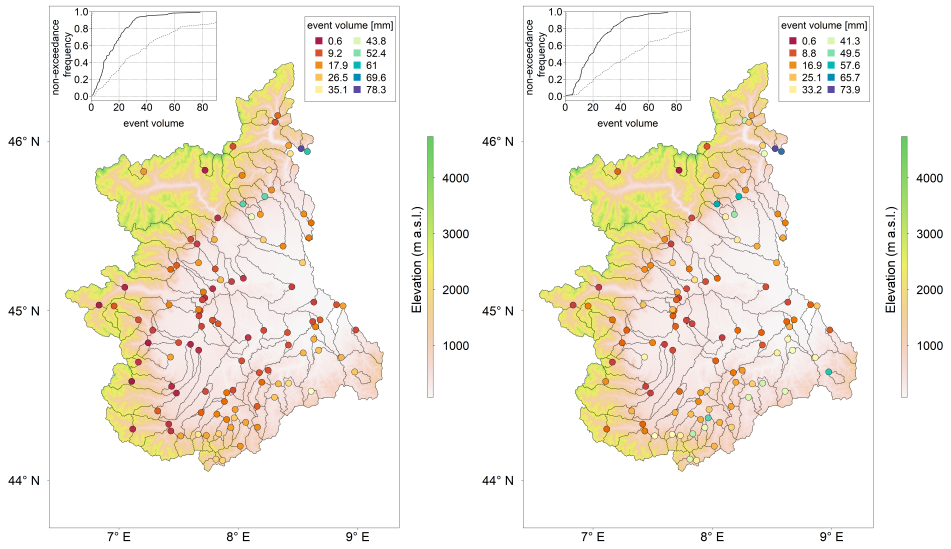
(a) Observed (b) Simulated

Fig. 5.6 Mean event peak time for a) observed runoff events and b) corresponding simulated runoff events by using the regionally calibrated TUWmodel, in 108 catchments. In the inset, the continuos line represents the non-exceedance probability of mean event peak time, while the dashed lines represent the non-exceedance probability of the 10th and 90th percentiles of peak time.



(a) Observed (b) Simulated

Fig. 5.7 Mean event peak for a) observed runoff events and b) corresponding simulated runoff events by using the regionally calibrated TUWmodel, in 108 catchments. In the inset, the continuos line represents the non-exceedance probability of mean event peak, while the dashed lines represent the non-exceedance probability of the 10th and 90th percentiles of event peak.



(a) Observed (b) Simulated

Fig. 5.8 Mean event volume for a) observed runoff events and b) corresponding simulated runoff events by using the regionally calibrated TUWmodel, in 108 catchments. In the inset, the continuos line represents the non-exceedance probability of mean event volume, while the dashed lines represent the non-exceedance probability of the 10th and 90th percentiles of event volume.

both for observed and simulated events. Event duration (Figure 5.5) shows a quite limited spatial variability, but it is worth noting that there are some similarities with the spatial pattern of runoff coefficient. In this case the simulated events, despite capturing the observed behavior, show higher mean and variability compared to observed events. The event peak time mimics the pattern of event duration (Figure 5.6). Finally, event peak and event volume (Figure 5.7 and 5.8), representing the intensity and magnitude of the events, reflect the spatial pattern of runoff coefficient, with the highest values in the North-East and in the southern catchments, and lower values in the central and western part of the region. Figure 5.9b and 5.9c highlight that the highest values of event peak and event volume are found for catchments located at medium elevation (between 1000 and 1500 m a.s.l.), while lower values are observed for both lowland and high-elevation Alpine catchments. The observed pattern of event peak seems to be well reproduced by the model, which, instead, tends to slightly overestimate the event volume.

Figure 5.10 and 5.11 show, respectively, the distribution of R^2 and Nash Sutcliffe Efficiency (NSE) computed on each catchment by considering the corresponding extracted observed and simulated runoff events. The results show that while the

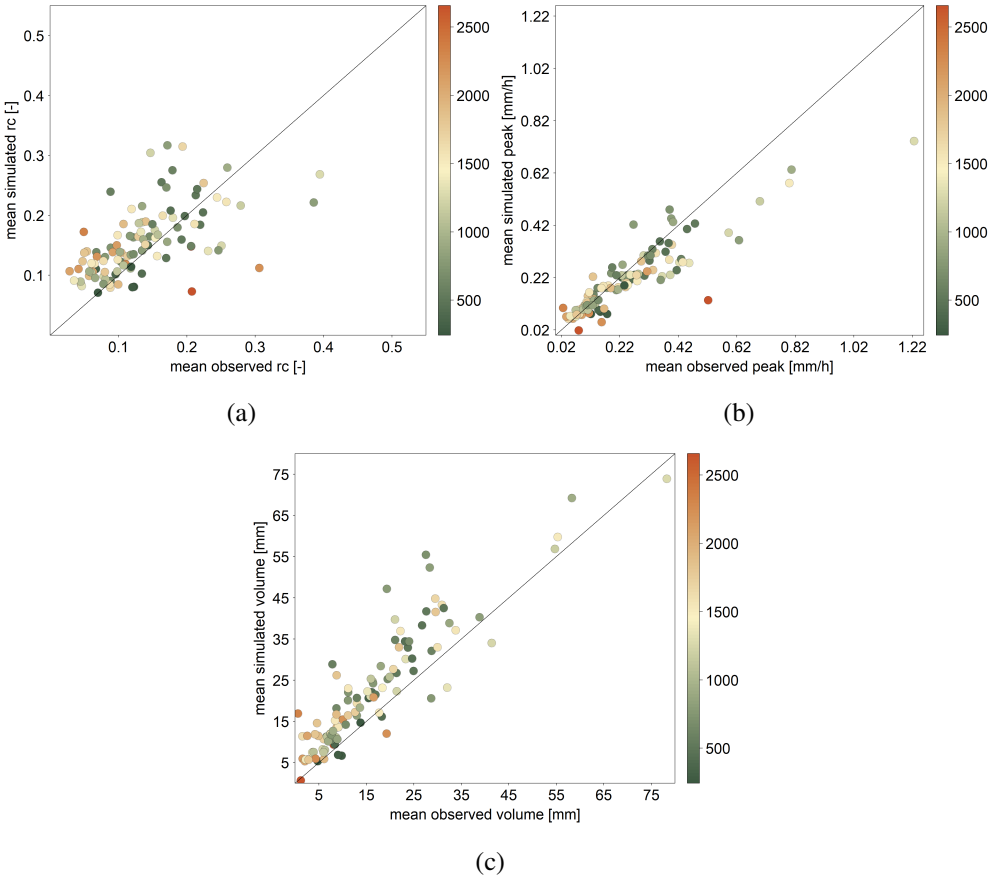


Fig. 5.9 a) Mean observed vs. simulated runoff coefficient. b) Mean observed vs simulated runoff event peak. c) Mean observed vs simulated runoff event volume. Points are discretized by mean catchment elevation [m a.s.l.].

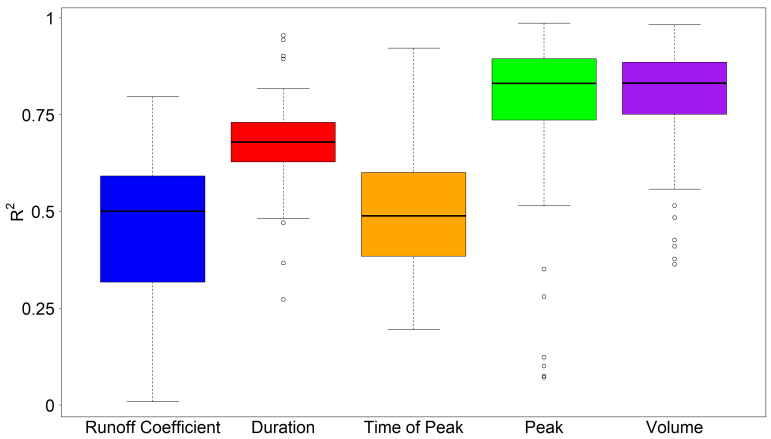


Fig. 5.10 Distribution of R^2 for observed vs. simulated runoff event characteristics, calculated for each catchment.

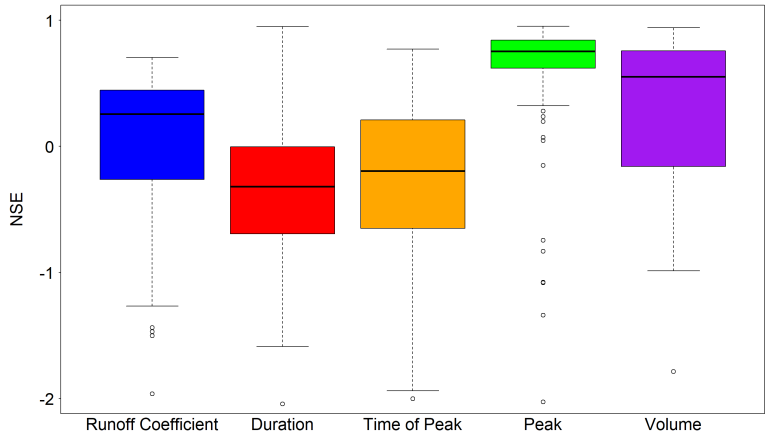


Fig. 5.11 Distribution of Nash Sutcliffe Efficiency (NSE) for observed and simulated runoff event characteristics, calculated for each catchment.

observed event duration and peak time are poorly represented by the model (and peak time also shows a low correlation value), the event peak and the event volume are well represented (median NSE ≥ 0.5). The median R^2 of the runoff coefficient is set to 0.5 but the median NSE proves to be positive, meaning that the model predictions are quite accurate.

5.1.3 Simulated runoff event characteristics for ungauged sites

In this section, the TUWmodel is used to extract simulated distributed runoff event characteristics over North-Western Italy. This is performed by taking advantage of a wide catchment dataset, i.e. the European Environment Agency (EEA) Catchments and River Network System (ECRINS, 2012).

European Catchments and Rivers Network System (ECRINS) dataset

The European Catchments and River Network System (ECRINS) is the geographical information system of European watersheds, river and lakes adopted by the European Environment Agency (EEA) and serving as reference system for the Water Information System for Europe (WISE). In particular, ECRINS is the final product obtained by the Joint Research Centre (JRC) Catchment Characterisation and Modelling (CCM) activity, which first developed a Catchment Characterisation Model (CCM) in 2003 with further geographical extension and error correction (JRC, 2008).

The dataset is generated from a digital elevation model of 100m resolution and results in digitised objects (mainly rivers and watersheds) with a unique code based on the location of the feature in the hydrological system. The covered area is the geographical Europe plus catchments with springs in Turkey and the western Caspian catchments, for a total area of around 10 million km^2 (Figure 5.12).

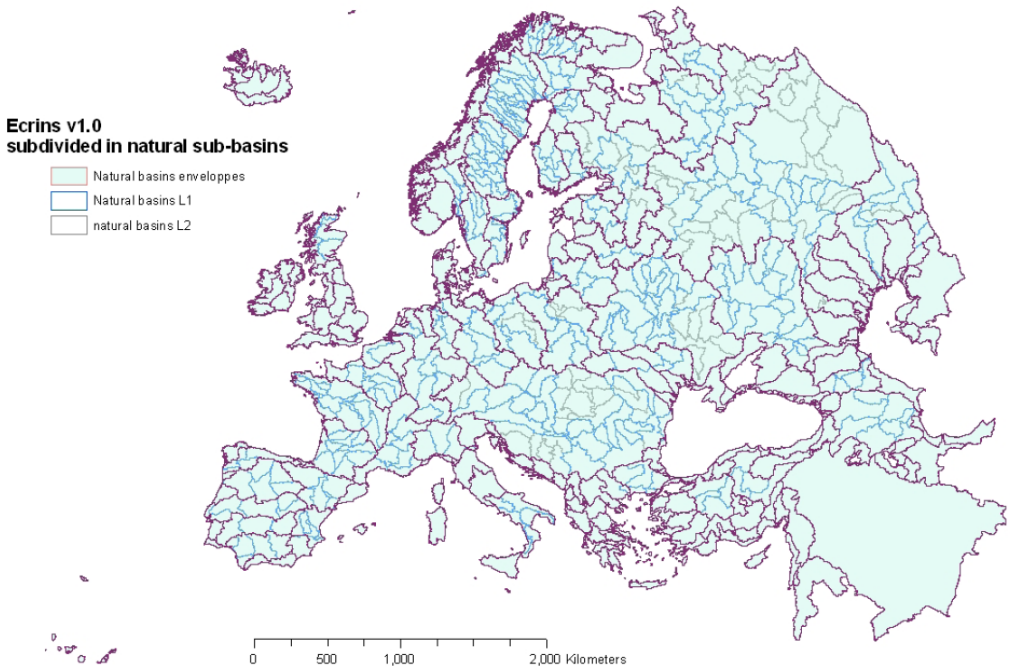


Fig. 5.12 Geographical extension of ECRINS, displayed with natural catchments.

The main component of this database is the Functional Elementary Catchment (FEC), which represents the smallest drainage area delineated under the imposed modelling constraints. These elementary catchments are defined by the accumulation flow which identifies a likely drainage area. FEC can be either a 'continental FEC' when built by aggregating elementary CCM catchments from a non-coastal basin, or a 'coastal FEC' when elementary CCM catchments belong to a coastal basin. The rationale followed to build the FECs is based on the Strahler level of catchments (ECRINS, 2012, Section 5.3). The FEC data structure is detailed in the feature class C_Zhyd of ECRINS dataset (ECRINS, 2012, Table A1.1) and the main information is reported in Table 5.4.

Table 5.4 Main FECs information in feature class C_Zhyd.

Variable	Description
OBJECTID	n/a
Bas0_ID	ID of the FEC envelope
BASINNAME	Name of the Bas0_ID
ZHYD	FEC ID
NextDown	ID of the downstream FEC
Outlet	n/a
Surf	FEC area [km^2]
Surfc	Cumulated area upstream (FEC not included) [km^2]
Surffinal	Cumulated area, FEC included [km^2]
MeanElev	Mean FEC altitude [m]

Simulated runoff event characteristics using TUWmodel and ECRINS

In this section, the results of events separation obtained by running TUWmodel over North-Western Italy considering the ECRINS distributed dataset are presented. The parameter sets used for discharge simulation are the outputs of the regional calibration with PASS, during the period 2000-2010 by using snow cover information. In order to run the model, a proper selection of the catchments falling inside the grid of input data for North-Western Italy (OI dataset grid) is performed. In particular, starting from few downstream catchments that are falling inside the grid, identified by the ZHYD code, the catchments having as 'NextDown' these catchments are chosen. This operation is iteratively repeated up to the headwaters. A subset of 695 catchments is finally identified and represented in Figure 5.13.

For each catchment, all the 30 parameter sets (corresponding to the 30 groups obtained as output from PASS procedure) are used, thus producing an ensemble of 30 runoff time series, from 1961 to 2020. The runoff event separation code is run for all the 30 time series, for each catchment. Later, the results are cleaned and, in particular, potential inconsistent cases, identified events having runoff coefficient $>$ 1 or event duration $>$ 120 days are removed.

Figure 5.14 - 5.16 report the mean value of runoff event characteristics calculated over the 695 catchments. The mean value is obtained by taking the mean of the variable over each of the 30 groups (corresponding to the 30 parametrizations) and then averaging these values. It is worth noting that the spatial pattern of event characteristics is in line with the results found for gauged sites (Figure 5.4-5.8). Moreover, event duration and peak time (Figure 5.14b and 5.15a) show a similar pattern as well as event peak and event volume (Figure 5.15b and 5.16), as can be

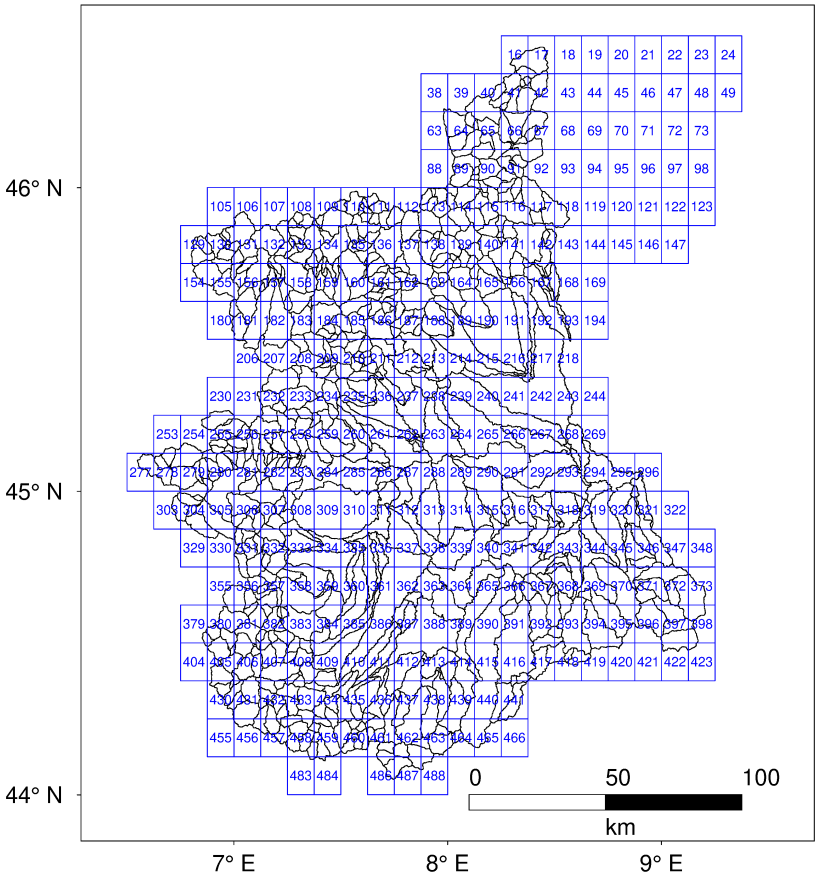
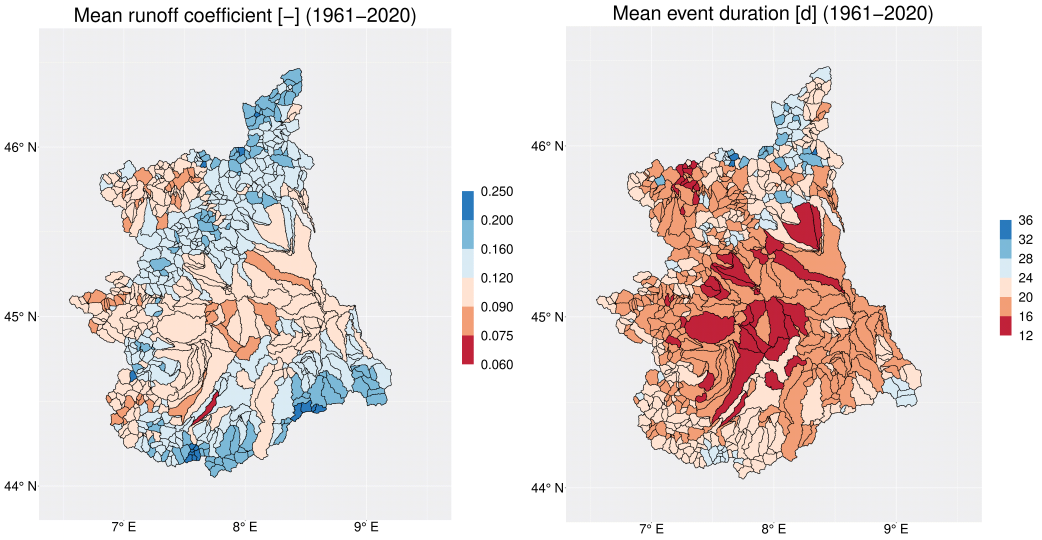


Fig. 5.13 ECRINS catchments with superimposed input data grid over North-Western Italy.

expected. Table 5.5 summarizes the statistics of the five event characteristics over the region.

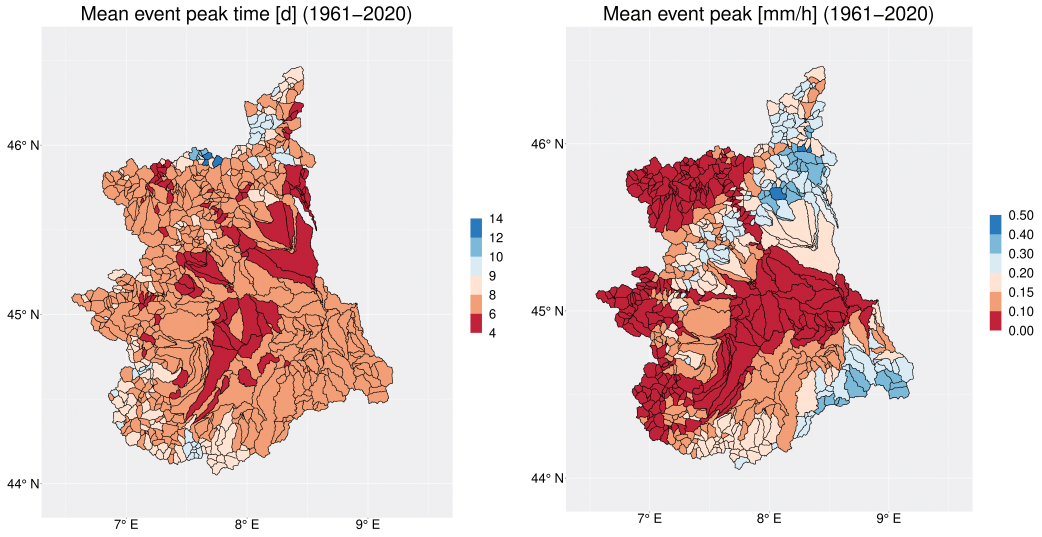
5.2 Runoff Event Characterization and Classification

In the context of climate change, which impacts natural and physical processes through a variety of mechanisms, understanding the generation processes of runoff events is of particular importance to highlight drivers behind possible shifts in river flows seasonality, long-term trends of floods and flood hazard changes. To this aim, the use of a classification framework is pivotal to link runoff generation processes to individual catchment responses at the event time scale. The importance of such a framework lies in the reduction of a great amount of information into a fixed number



(a) (b)

Fig. 5.14 a) Mean runoff coefficient calculated for simulated runoff events in the period 1961-2020 b) Mean event duration calculated for simulated runoff events in the period 1961-2020.



(a) (b)

Fig. 5.15 a) Mean event peak time calculated for simulated runoff events in the period 1961-2020 b) Mean event peak calculated for simulated runoff events in the period 1961-2020.

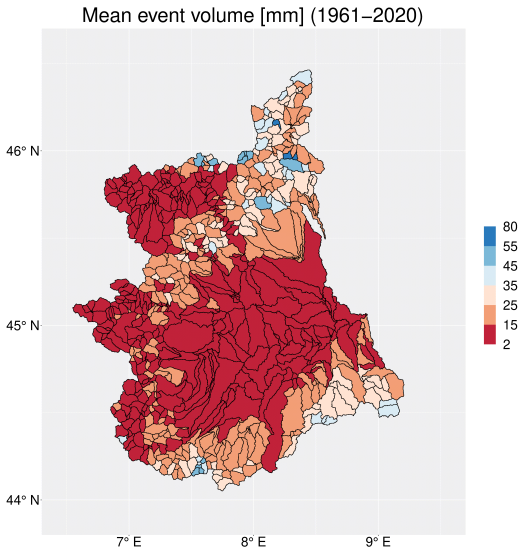


Fig. 5.16 Mean event volume calculated for simulated runoff events in the period 1961-2020.

Table 5.5 Statistics of simulated mean runoff event characteristics distributed over North-Western Italy.

Runoff event characteristic	**Min**	**1st quartile**	**Median**	**Mean**	**3rd quartile**	**Max**
Runoff coefficient [-]	0.073	0.104	0.130	0.131	0.152	0.247
Event duration [d]	12.54	18.11	19.98	20.46	22.31	34.98
Event peak time [d]	4.80	6.39	7.01	7.16	7.88	12.80
Event peak [mm/h]	0.037	0.07	0.127	0.138	0.182	0.482
Event volume [mm]	2.86	9.20	15.36	17.35	23.12	77.11

of classes that can be managed. Many studies focus on process-based classification of large runoff events (e.g., maximum annual floods) at different spatial scales, from regional to continental (e.g., Hirschboeck, 1987; Weingartner and Diezig, 2007; Nied et al., 2014; Sikorska et al., 2015; Merz and Blöschl, 2003; Berghuijs et al., 2016; Berghuijs et al., 2019; Stein et al., 2019). However, these studies don't consider the large range of variability of runoff magnitudes and tend to neglect pre-event wetness states and the interaction among soil moisture and precipitation, which instead are key drivers of runoff generation (Tarasova et al., 2019). Analyzing and quantifying the spatial and temporal variability of drivers of runoff, such as soil moisture, or moments of flood response at the catchment scale is crucial for correctly interpreting

runoff triggers (Viglione et al., 2010) and new available hydrometeorological datasets (e.g. from satellites) are useful tools to accomplish this goal.

In this section, we take a further step in the event analysis by applying the consistent framework proposed in Tarasova et al. (2020) in North-Western Italy, with the aim of providing a picture of event types based on runoff-generation processes over the area. Observed event types for the gauged sites are compared with the simulated runoff event types, obtained by running TUWmodel over the study area in a distributed mode. The framework allows to use non dimensional indicators (mainly covariance and ratio based indicators), which are inherently less uncertain compared to indicators of absolute value, in order to characterize space-time dynamics of precipitation events and their interaction with antecedent states in terms of snow cover, frozen soils and soil moisture. The derived event typology captures the first-order controls of runoff response in a wide variety of catchments and allows to identify sub-regions of homogeneous event type frequency in North-Western Italy. The identified sub-regions prove to be spatially consistent when using observed and simulated runoff, confirming the ability of the model in reproducing hydrological processes at the event scale. The results show regional differences and similarities of runoff generation processes over the region that are consistent with the prevailing hydroclimatic conditions.

5.2.1 A process-based framework for event characterization

The characterization framework used in this Dissertation is a multi-layer process-based framework for runoff event characterization, which is based on a set of indicators describing different aspects of the transformation of rainfall into runoff (e.g., space-time dynamics of rainfall and snowmelt, antecedent catchment wetness states, snow cover and soil freezing conditions). The indicators are divided into two layers, as shown in Figure 5.17. The set of runoff events is split at each layer according to predefined thresholds, allowing to label each event with a specific category. All the indicators used, together with the applied thresholds, are shown in Table 5.6.

Table 5.6 Indicators and corresponding thresholds of each characterization layer.

Layer	Selected indicator (dimensionless)	Expression	Selected thresholds and possibile intervals	Performed split
a) Inducing event	Ratio of event snowmelt volume and total event precipitation volume	$\frac{M_{x,y,t}}{P_{x,y,t}}$	$Th\_ratio.Rain.Melt$ = 0.95 [0,1]	Snowmelt vs. any other type
	Ratio of event rainfall volume and total event precipitation volume	$\frac{R_{x,y,t}}{P_{x,y,t}}$	$Th\_ratio.Rain.Melt$ = 0.95 [0,1]	Rainfall or Rain-on-ice vs. Mixture or Rain-on-snow
	Normalized spatial covariance of event-averaged snow cover and rainfall	$\frac{cov_{x,y}(SWE_t, R_t)}{SWE_{x,y,t} * R_{x,y,t}}$	$Th_$cov = 0 (-∞, +∞)	Rain-on-snow vs. Mixture
	Normalized spatial covariance of pre-event level of soil freezing and event rainfall volume	$\frac{cov_{x,y}(SF(t_0), R_t)}{SF_{x,y,t} * R_{x,y,t}}$	$Th_$cov = 0 (-∞, +∞)	Rain-on-ice vs. Rainfall
Temporal organization	Temporal coefficient of variation of precipitation rate	$\frac{\sqrt{var_t(P_{x,y})}}{P_{x,y,t}}$	$Th_$cv.temp = 1 (0, +∞)	Volume-dominated vs. Intensity-dominated
	Ratio of maximum precipitation rate and total precipitation volume	$\frac{max(P_{x,y}(t))}{P_{x,y,t}}$	$Th_$ratio.Vol.Int. = 0.5 (0,1]	Volume-dominated vs. Intensity dominated
Space-time organization	Spatial coefficient of variation of precipitation event volume	$\frac{\sqrt{var_{x,y}(P_t)}}{P_{x,y,t}}$	$Th_$cv.space = Q2 (0, +∞)	Local vs. Extensive
	Mean normalized spatial covariance of precipitation rates between consecutive time steps	$\frac{1}{t+1}\sum_t \frac{cov_{x,y}(P(t), P(t+1))}{P_{x,y,t} * P_{x,y,t}}$	$Th_$cov = 0 (-∞, +∞)	Steady vs. Unsteady
b) Wetness state	Catchment-averaged antecedent soil moisture	$SM_{x,y}(t_0)$	$Th_$sm = max(K) [0,1]	Wet vs. Dry
Spatial interaction of precipitation and soil moisture	Spatial coefficient of variation of antecedent soil moisture	$\frac{\sqrt{var_{x,y}(SM_{t_0})}}{SM_{x,y,t}}$	$Th_$cv.space = Q2 (0, +∞)	Uniform vs. Patchy
	Normalized spatial covariance of precipitation event volume and antecedent soil moisture	$\frac{cov_{x,y}(SM(t_0), P_t)}{SM_{x,y,t} * P_{x,y,t}}$	$Th_$cov = 0 (-∞, +∞)	Patchy No Overlap vs. Patchy Overlap

It is worth mentioning that the proposed framework doesn't consider streamflow data to define categories. In the first layer of the framework the type of inducing meteorological event is analyzed. An event is a pure snowmelt event if the volumetric ratio of catchment-averaged event snowmelt and total precipitation is larger than 0.95. Similarly, for a pure rainfall event, the ratio of catchment-averaged rainfall and total precipitation is larger than 0.95. A rainfall event can be in particular a Rain-on-ice event (RoI), meaning that rain falls on frozen soils, based on the spatial covariance of event-averaged rainfall volume and pre-event degree of soil freezing ($SF(t_0)$). $SF(t_0)$ is defined as :

$$SF(t_0) = \sum_{i=0}^{n} |T_i| \tag{5.2}$$

where T is the air temperature of a snow-free pixel and n is the number of days with air temperature below -2 °C during the 5 days before the start of rainfall event. Within the sample of events that are mixture of rainfall and snowmelt, Rain-on-snow (RoS) events are determined if the normalized spatial covariance of event averaged snow water equivalent and event rainfall volume is positive.

Another aspect investigated in this layer is the temporal organization of the inducing event. In this respect, the coefficient of variation in time of the catchment-averaged precipitation rate and the ratio of the maximum intensity to total volume of the inducing event are considered as indicators. For the latter, the threshold used is 0.5, so that if the ratio is higher than this threshold, the event is defined as intensity dominated, viceversa it is volume dominated. This corresponds to runoff generation mechanisms of infiltration-excess and saturation-excess (Horton, 1933; Dunne, 1978).

Going beyond the temporal dimension of the runoff events, the interaction between space and time dimensions is investigated, by using the spatial coefficient of variation of precipitation volume and the mean spatial covariance of precipitation rates between consecutive time steps (Table 5.6). This allows to categorize an event as *Local Steady*, *Local Unsteady*, *Extensive Steady* and *Extensive Unsteady* (Figure 5.17). A *Local* event is characterized by an uneven distribution of the precipitation volume over the catchment, suggesting a local runoff generation. On the opposite, the event is defined *Extensive* if the precipitation volume is evenly distributed, suggesting extensive runoff generation. A *Steady* event occurs mainly in the same part of the catchment area during consecutive days, leading to higher likelihood of saturation. An *Unsteady* event is characterized, instead, by the variation of the precipitation centroid in time, with the event involving different portions of the catchment.

The second characterization layer divides events according to the catchment wetness state (Figure 5.17). First, events are subdivided in *Wet* or *Dry* based on catchment-averaged antecedent soil moisture. In theory, since the non-linear storage-discharge relation can be different for varying catchments, defining a unique threshold of soil moisture to classify wet and dry states can be misleading. For this reason, in Tarasova et al. (2020) the value of soil moisture corresponding to the maximum curvature κ of a fitted exponential function describing the non-linear relation between the runoff coefficient (*rc*) and antecedent soil moisture (*sm*) is used as threshold. Nevertheless, for practical reasons, in this study a single representative value equal to 0.7 is selected as wetness state threshold.

Once defined the catchment wetness state, the spatial interaction between the inducing precipitation event and the catchment wetness state is investigated. To this aim, the spatial coefficient of variation of antedecent soil moisture gives the indication of a *Uniform* or *Patchy* spatial organization of soi moisture. In case of a *Patchy* (not homogeneous) organization, the spatial covariance of soil moisture state

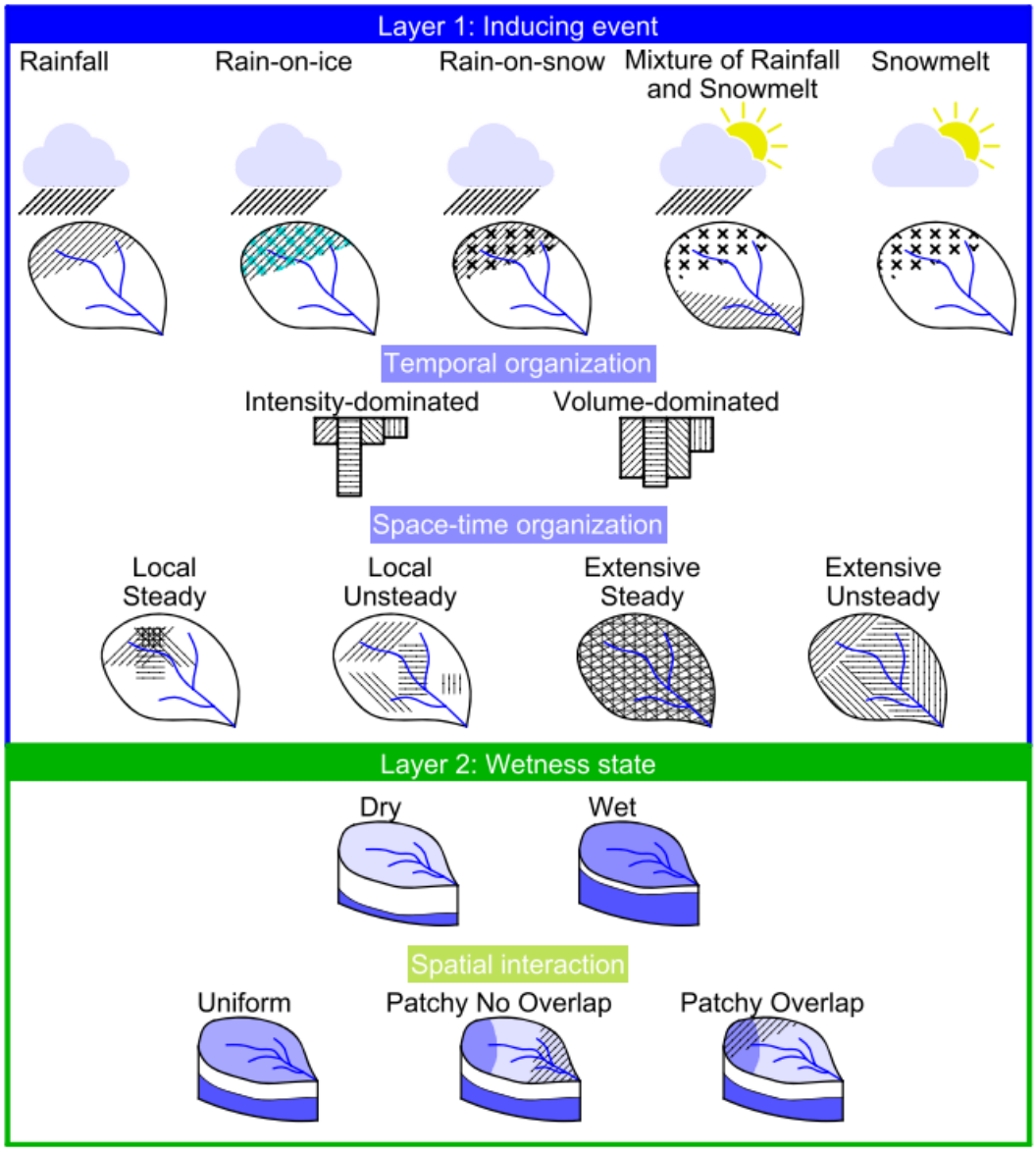


Fig. 5.17 The multi-layer framework for process-based characterization and categorization of runoff events (adapted from Tarasova et al. (2020)). Indicators and categorization thresholds used for each layer are indicated in Table 5.6.

and precipitation volume is used to detect overlaps of saturated areas and wetted areas due to the precipitation event. This leads to the following event categories: *Uniform*, *Patchy Overlap* or *Patchy No Overlap* (Figure 5.17). To summarize, the threshold used for covariance-based indicators is 0, as positive covariance corresponds to a spatial overlap of hydrometeorological variables. For indicators based on the temporal coefficient of variation, the threshold is 1. Finally, spatial coefficient of

variation is dependent on catchment size (larger catchments show higher spatial CV) and thus, for each catchment, the median of the event values is taken as threshold.

5.2.2 Hierarchical event classification

The combination of the event categories found by applying the characterization framework fully describes the runoff generation process of a single event. The combinations are chosen according to some a priori knowledge on runoff generation at catchment scale (Table 5.7), leading to a hierarchical classification (Figure 5.18).

Table 5.7 Event types and corresponding hypothesized runoff generation processes at catchment scale. Adapted from Tarasova et al. (2020).

Event type	Hypothesized runoff generation processes
Snowmelt	Radiation-induced snowmelt (usually *Wet.Volume.Steady*)
Rain-on-ice	Frozen soils prevent infiltration of rainfall (usually *Wet.Intensity.Uniform*)
Rain-on-snow	Several possible ways of runoff generation ranging from situations when snowpack prevents infiltration of rainfall or instead either stores substantial portion of rainfall water or is degraded due to rainfall-induced snowmelt (usually *Wet.Volume.Extensive.Steady.Uniform*)
Mixture of Rainfall and Snowmelt	Radiation-induced snowmelt and simultaneous in time but not in space rainfall (usually *Wet.Volume.Local.Unsteady*)
Rain.Wet.Intensity.Local	Local runoff generation; pre-event saturation or infiltration excess; possible connectivity[a]
Rain.Wet.Intensity.Extensive	Extensive runoff generation; pre-event saturation or infiltration excess; possible connectivity
Rain.Wet.Volume.Local.No.Overlap	Local runoff generation, event-fed saturation, possible connectivity
Rain.Wet.Volume.Local.Overlap	Local runoff generation, pre-event saturation, established connectivity
Rain.Wet.Volume.Extensive.No.Overlap	Extensive runoff generation, event-fed saturation, possible connectivity
Rain.Wet.Volume.Extensive.Overlap	Extensive runoff generation, pre-event saturation, established connectivity
Rain.Dry.Intensity.Local.Steady	Local runoff generation; possible infiltration excess or event-fed saturation; no connectivity
Rain.Dry.Intensity.Unsteady	Local runoff generation; possible infiltration excess; no connectivity
Rain.Dry.Intensity.Exrtensive.Steady	Possible extensive runoff generation; possible event-fed saturation and infiltration excess; possible event-induced connectivity
Rain.Dry.Volume.Local	Local event-fed saturation; no connectivity
Rain.Dry.Volume.Extensive.Steady	Extensive event fed-saturation; possible event-induced connectivity
Rain.Dry.Volume.Extensive.Unsteady	Extensive runoff generation; possible pre-event saturation; possible event-induced connectivity

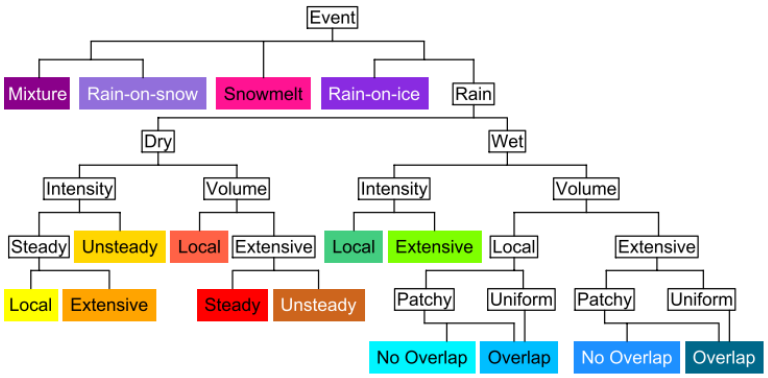


Fig. 5.18 A decision tree for the hierarchical classification of runoff events. Hypothesized runoff generation processes corresponding to each type of runoff events are described in Table 5.7. Adapted from Tarasova et al. (2020).

5.2.3 Observed and simulated event typology in North-Western Italy

In this section, the final classification of events in different types (Figure 5.18) is performed in North-Western Italy, by considering both the 108 catchments, for which observed discharge data is available, and the ECRINS distributed catchment dataset, over which TUWmodel is applied. The aim is to obtain a typology both for observed and simulated events and verify the ability of the model in reproducing the spatial distribution of observed event type frequency over the region. This is an exemplary application, similar to the one presented in Tarasova et al. (2020).

Observed event types

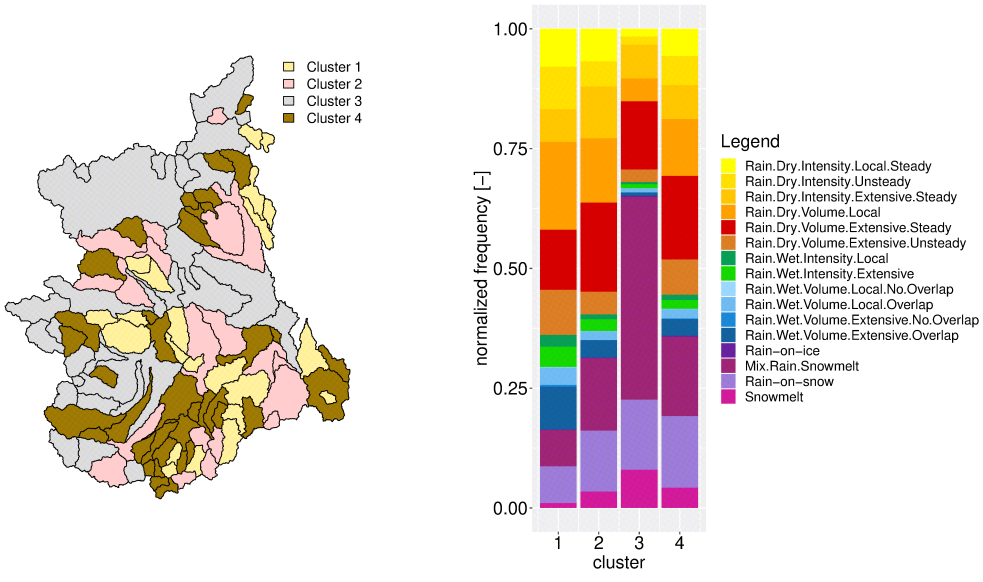
The catchments for which discharge data are available are clustered according to the frequency of occurrence of their event types. Four clusters are identified with specific event type distributions, based on thresholds of observed event type frequency in each catchment, deemed suitable to describe the data variability over the catchments. In particular, Cluster 2 gathers catchments for which the frequency of *Rain.Dry.Intensity.Extensive.Steady* events is $> 10\%$ and the sum of *Rain-on-snow* and *Mix.Rain.Snowmelt* frequencies is also $> 10\%$. For catchments of Cluster 3, the *Mix.Rain.Snowmelt*, *Rain-on-snow* and *Snowmelt* events are dominant with a total frequency $> 50\%$. Cluster 4 contains catchments having both a frequency of *Rain-on-snow* $> 10\%$ and a *Rain-on-snow* and *Mix.Rain.Snowmelt* total frequency $> 20\%$. The remaining catchments, typically having a dominant *Rain.Dry.Volume* component, are assigned to Cluster 1.

Looking at the results (Figure 5.19a), Cluster 1 consists of lowland catchments that are located in the central Po river Valley, partially in the North-East (e.g, Agogna river catchment) and in the southern part in the proximity of the Apennines (Bormida and Tanaro river). For this cluster, intensity-dominated events under dry conditions represent an important fraction of all events (24%) and the majority of these are unsteady (9%). Volume-dominated events with local characteristics under dry conditions (*Rain.Dry.Volume.Local*) represent the most dominant class (18%). Events occurring under wet conditions and events characterized by the presence of snow represent a small fraction of the total.

The second cluster consists of catchments located both in lowland and at medium elevation. With regard to Figure 1.3a, the results shows that Cervo and Sesia catchments are included as well as Orco and, in the South, catchments pertaining to Bormida, Tanaro and Orba river. The fraction of events impacted by snow (*Mix.Rain.Snowmelt*, *Rain-on-Snow*, *Snowmelt*) is higher compared to cluster 1 and extensive and steady rainfall events in dry conditions are dominant types for these catchments, both intensity (11%) and volume (19%) types.

Cluster 3 includes all the Alpine catchments from South-West to North (mainly Doria Riparia, Dora Baltea and Toce), where *Snowmelt* events are relevant (8% of the total), and also large catchments extending towards the Po Valley, characterized by high elevation zones and the presence, among others, of events being mixture of rainfall and snowmelt (*Mix.Rain.Snowmelt*, 42% of the total). The portion of rainfall events during wet conditions is very small for this cluster.

Finally, cluster 4 includes catchments that are widespread over the region but mainly located in mountainous area, at quite high elevation. It appears to be similar to Cluster 2 but the portion of *Mix.Rain.Snowmelt*, *Rain-on-snow* and *Snowmelt* is higher (36% in total). Extensive volume-dominated events during dry conditions prevail (the steady ones represent 17% of the total).



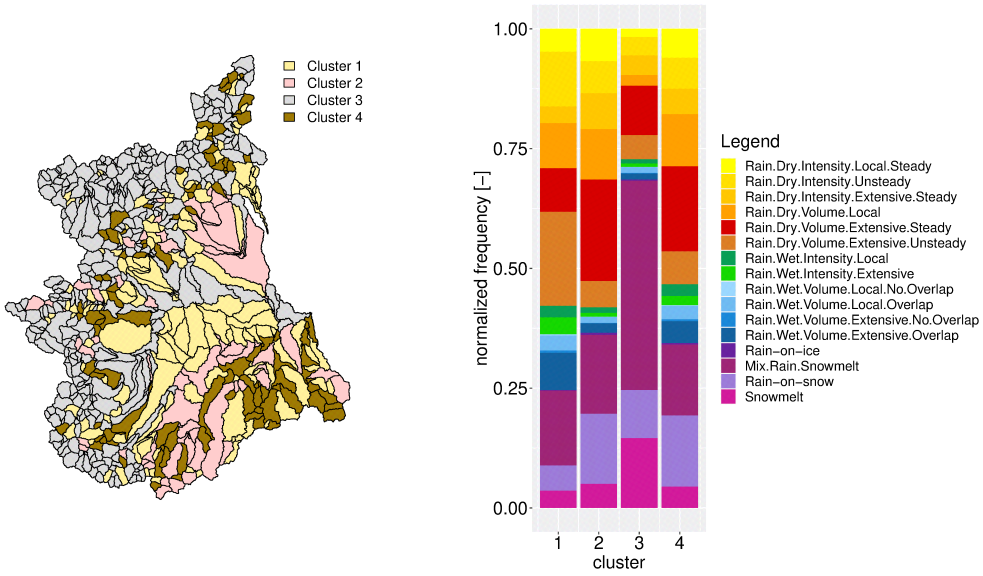
(a) (b)

Fig. 5.19 Regional pattern of observed event type frequency in North-Western Italy: a) spatial distribution of four event clusters with homogeneous frequency of event types. b) frequency of event types for each cluster.

Simulated event types

Similarly to the observed runoff events, the same characterization and classification procedure is performed with the simulated events, spanning the period from 1961 to 2020, by considering the ECRINS dataset. Since results in terms of event type frequency are not dependent on the specific parameterization used, one is chosen among the possible 30 groups derived from the regionalization procedure. The same criteria adopted for the observed events are here used to cluster the 695 catchments according to event frequency. Figure 5.20 reports the event type frequency associated to each cluster and the spatial distribution of the clusters. It emerges that both the frequency of events and the spatial pattern of clusters are consistent with the corresponding observed ones, despite some specific differences.

In cluster 1, the volume-dominated events in dry conditions are stil dominant, but the unsteady types (both intensity and volume) constitute an important portion of all events. In particular, *Rain.Dry.Volume.Extensive.Unsteady* events represent the 19% of the total, while *Rain.Dry.Intensity.Unsteady* the 11%. The portion of events that are not pure rainfall (*Rain-on-ice*, *Mix.Rain.Snowmelt*, *Rain-on-snow*, *Snowmelt*) is higher compared to the corresponding fraction in the observed events cluster. Simulated clusters 2, 3 and 4 (Figure 5.20b) are very similar to the corresponding observed clusters (Figure 5.19b) but the quota of the above mentioned event types is higher when considering simulations. It is worth noting that in cluster 3 the fraction of *Snowmelt* events (14%) is higher than the fraction of *Rain-on-snow* events (10%), as opposed to the results for observed events.



(a) (b)

Fig. 5.20 Regional pattern of simulated event type frequency in North-Western Italy: a) spatial distribution of four event clusters with homogeneous frequency of event types. b) frequency of event types for each cluster.

5.3 Conclusions

In this chapter, the regionally calibrated TUWmodel is used for identifying runoff events and determining the spatial distribution of event characteristics (i.e. runoff coefficient, event duration, event peak time, event peak and event volume) over the region. First, the event characteristics found by using the simulated runoff series are compared to observed event characteristics in 108 sites, where the event extraction procedure works properly both for observed and simulated discharges. Then, the model is applied with a wider catchment dataset to get distributed statistics of runoff event characteristics. In a second phase, the events are characterized and classified by using some dimensionless climatic indicators that allow to identify the runoff generation mechanisms and produce classes of events (i.e. event types).

Concerning runoff event characteristics, the results show that the mean observed runoff coefficient, event peak and event volume are maxima over catchments located in the northeastern and southern part of the study area. The model is able to properly capture the spatial pattern of observed runoff event characteristics, even if a quite

important discrepancy is found among the mean and variability of observed and simulated event duration. Runoff event peak and volume show quite good performances with median R^2 and Nash-Sutcliffe Efficiency (NSE) greater than 0.75 and 0.5, respectively, while the performance is lower for runoff coefficient, duration and peak time. A consistent spatial pattern of event characteristics is obtained by running the model in a distributed mode in ungauged sites. The highest values of the characteristics, in particular runoff coefficient, event peak and event volume, are found in catchments that are located close to the mountainous chains of North-Western Italy (i.e. Alps and Apennines), at an altitude range of 1000-1500 metres. Instead, catchments located at higher elevations over the Alpine range and in the Po River Valley generally display lower values.

By analysing the results obtained with event characterization and classification, catchments displaying the highest mean characteristics mainly belong to either cluster 2 or cluster 4, for which the portion of *Rain-on-Snow* and *Rain.Dry.Volume.Extensive. Steady* events represent an important fraction of the total number of events. Instead, the lowland catchments and the high-elevation catchments in the Alps belong to either cluster 1 and 3, where volume-dominated local/unsteady events and snow-impacted events are dominant, respectively.

Chapter 6

Discussion and Conclusions

This Dissertation deals with river flood characterization in North-Western Italy. After a brief description of the study region and the main data used in Chapter 1, a statistical analysis of annual maximum discharges and standard ETCCDI indices is performed, to get an indication on which climate variables better explain the interannual variability and the long-term tendency of floods. In the second part, the available runoff dataset is expanded in time and space by using the conceptual semi-distributed rainfall-runoff model TUWmodel, with the aim of characterizing runoff events over the same region and get a picture of the classified event types, which reflect different hydroclimatic conditions leading to floods. In this context, the model is regionally calibrated using a decision tree machine-learning algorithm and a documented regionalization procedure. The modelling results can be used to detect potential temporal changes of runoff characteristics of different flood event types and attribute these changes to climate drivers.

In Chapter 2, the temporal correlation analysis among ETCCDI indices and runoff extremes provides interesting results in terms of extreme precipitation: R99pTOT, R95pTOT, Rx5day, Rx1day are highly positively correlated to annual discharge maxima, more than indices reflecting mean precipitation conditions. In particular, R99pTOT shows higher correlation to annual floods compared to R95pTOT. This is probably due to the fact that the former incorporates a very high precipitation threshold, so it is a good proxy for the annual maximum flow events, while the latter includes less intense rainfall events, not necessarily leading to the largest floods. For this reason, R95pTOT could be used to describe other discharge indicators (e.g., flow volume, not considered in this work). By comparing the results of maximum 1-day

precipitation (Rx1day) and maximum consecutive 5-day precipitation (Rx5day), it is worth noting that Rx1day shows higher correlation to annual maxima for small catchments, while Rx5day performs better for large catchments. This can be explained by the interplay between the duration of rainfall events and the catchment response time. In fact, longer (shorter) rainfall events are relevant in causing floods in larger (smaller) catchments. The majority of intermediate-sized catchments (i.e., $100 < A < 600$ km^2) don't show high correlation for extreme indices. A possible improvement could be to adopt some ad-hoc indices to capture rainfall events with a time-scale that is more adequate for these catchment sizes. An example is the maximum consecutive 3-day precipitation (Rx3day), which is not considered in this Thesis, since it is not included in the standard ETCCDI indices.

The spatial correlation analysis of tendencies is based on the assumption that an indication of which indices are correlated to annual floods can be obtained by searching for indices whose tendency is correlated in space (considering the ensemble of catchments) with the tendency of floods. It is important clarifying that the identified ETCCDI indices explain flood tendencies at a regional level, not at a sub-regional or local level. The trend analysis highlights some limitations of the results in terms of the statistical significance of the differences within the region, i.e. taking into account uncertainty, the estimated trends are often overlapping. This can be mainly attributed to the limited length of the series (15 years, on average). Nonetheless, it is worth noting that finding trends which are significantly different from zero is not so relevant in this analysis, since the main purpose is to check for a coherent tendency correlation of climate indices and floods in all sites. To this regard, the spatial correlation analysis allows to attribute the flood tendency variability that characterizing the area.

From the results, it appears that the highest Spearman's rank correlation is found for the annual total precipitation, representing mean precipitation conditions. This suggests that while extreme precipitation is highly correlated with extreme discharges at the annual timescale, the decadal changes of extreme flows may be better explained by the decadal changes of the average precipitation. Indeed, there is an intimate connection between average precipitation and catchment saturation conditions that would explain flood tendency, as demonstrated by other studies (e.g., Šraj et al., 2016). Considering temperature indices, a quite weak negative correlation among the tendency of annual maximum daily temperatures and flood tendency is observed. Nonetheless, a closer look to the results reveals that by just

considering catchments at high elevation ($>$ 1600 m a.s.l.), the negative correlation is stronger. This suggests that the results for temperature indices are mainly driven by high elevation catchments, where snow/ice-related processes dominate in the runoff response. Indeed, there is a physical connection between temperatures and snowmelt and ice melt dynamics; increasing summer temperatures are causing the retreat of glacierized areas and can be themselves correlated to the occurrence of longer periods with no precipitation in the mountains, which penalize snow accumulation, as observed in the Alps in the last 20 years (Beniston et al., 2018). The result of these processes is a negative tendency of runoff peaks, typically in spring or summer season. However, this conclusion is not trivial and analyzing the long-term temporal tendency of specific flood event types in the same area (e.g., snowmelt or rain-on-snow floods as in Sezen et al. (2020)) could be useful to reinforce the attribution of the tendency of the observed discharge maxima.

The analysis of flood timing reveals a high seasonal concentration of floods in spring season in the western mountainous area and in fall in the eastern and southeastern part of the region, as expected. Spring floods are likely due to snowmelt events occurring thanks to the quite abundant snow accumulation in winter in the Alps, while autumn floods are tipically associated with the Atlantic circulation, which occasionally favors intense precipitation events characterized by huge precipitation amounts falling over Liguria and Piemonte, as reported in the Introduction of this Dissertation.

The uncertainties found for the tendency of the indices, mainly related to the record length and the size of the spatial domain, indicate that the correlations found in the spatial correlation analysis may be spurious. Since, from a statistical point of view, the regional differences are not strong, more robust approaches could be used that also account for the expected sensitivity of the flood behavior to the covariates (e.g., Bertola et al., 2019). Regional differences in precipitation and temperature tendencies have already been discussed in other studies. Libertino et al. (2019) find trends of annual maxima of precipitation over Italy, for different durations and spatial scales. Focusing on North-Western Italy, some variability is shown for different areas of the region, with significant positive or negative trends in hilly or mountainous stations, opposed to an undistinct pattern over the Po Valley for 12-h and 1-day precipitation. This is consistent with the results obtained in this analysis for indices of precipitation intensity (Rx1day, R99pTOT). In terms of frequency of extremes, it is here shown a general increase of the number of days with precipitation above

20 mm (R20mm) for small and medium-sized catchments and this result is in line with the increasing annual record breaking anomaly described in Libertino et al. (2019) for the upper Po Region, during the late '80s - early '90s. Temperature trends reveal a marked decrease of the % of cold days and cold nights (TX10p and TN10p), which is more exacerbated in small catchments located at high elevation, in accordance with previous studies (see e.g., Acquaotta et al., 2014). Moreover, also warm days and warm nights (TX90p and TN90p) regionally increase over time, in line with results of Fioravanti et al. (2015). The analysis is not multivariate and therefore it doesn't quantify the relative contribution of temperature and precipitation indices in explaining flood tendencies. Other statistical tools, such as machine learning techniques or conceptual models can be useful to measure the relative importance of different explanatory variables (see e.g., Bertola et al., 2021; Zeng et al., 2021). However, the results obtained here provide an indication of which ETCCDI indices are more connected to floods. Zeng et al. (2021) also identify total precipitation as the most important driver of streamflow variation in the U.S., shaping the hydrologic regime. The main implication of this study is that future projections of mean precipitation for the Alpine area, obtained from state-of-the-art regional climate models, may be used as covariates in non-stationary flood frequency analysis to produce flood change projections. The reliable prediction of flood change is undermined by the high uncertainty associated with climate models, in particular for scenarios including the end-of-century. In this respect, it is worth noting that the inherent uncertainty of projections of annual total precipitation is lower compared to the uncertainty of extreme precipitation, so the use of total precipitation as covariate can increase the reliability of future flood estimates.

In Chapter 3, the conceptual semi-distributed rainfall-runoff model TUWmodel is implemented in the same study region. In particular, starting from locally calibrated parameters, a regional calibration of the model using a decision-tree version of the PASS procedure (Merz et al., 2020) is performed. The aim is to investigate the applicability of the method in a diverse region compared to Germany and the impact of using snow information on the calibration results, as other studies have already discussed (Parajka and Blöschl, 2008; Tong et al., 2021). It is shown that PASS can be efficiently applied in the study area and the resulting median model efficiencies in calibration, for training and test catchments, are similarly good to what obtained for Germany (0.6-0.7). Moreover, the methodology is robust in predicting runoff when considering an independent validation period. Some important conclusions drawn from Merz et al. (2020) are confirmed in this study.

The parameter sets found by regional calibration show a smaller spread compared to local parameter sets, meaning that the regional PASS calibration considerably reduces the effect of parameter equifinality. Also, regionally distributed parameters seem to be in line with hydro-meteorological and landscape features of the area and this indicates that geomorphological and climatic descriptors control model parameter values. Including snow information in the model efficiency function only produces a little non-significant increase of model performance in simulating streamflow and it doesn't guarantee higher performance in reproducing streamflow signatures (e.g., annual flows, seasonality). Nevertheless, when using snow information, the simulations with the regionally calibrated parameters prove to be less uncertain and the spatial pattern of snow parameters is more coherent with the climatic and geomorphological characteristics of the study area. Chapter 4 describes the newly developed R package called hydroPASS, designed for a flexible implementation of PASS. hydroPASS is supported with examples of application by using large U.S. hydrologic database (e.g., CAMELS).

The calibrated model is finally used to extract runoff events and identify runoff event characteristics such as runoff coefficient, event duration, event peak time, event peak and event volume (Chapter 5). The results obtained by applying the event separation algorithm of Giani et al. (2022) highlight that TUWmodel allows to capture the spatial pattern of observed runoff event characteristics and the performance is particularly good for runoff event peak and volume. For this reason, it can be used in a distributed mode in ungauged locations, providing consistent results. Nevertheless, the event separation method shows some limitations in the identification of event duration and peak time of events. When applied at the daily time-scale (a coarser resolution compared to hourly scale), this method tends to identify longer and bigger events (Figure 4, Giani et al., 2022), without distinguishing multiple peaks as potentially separated events. This feature appears to be exacerbated when the snowmelt component is an important fraction of the liquid water input, thus producing very long events in the northeastern and southern part of the region (Figure 5.5 and 5.6). Different event separation methods could be implemented in the same study region to overcome this issue such us the one proposed by Tarasova et al. (2018b), which allows to refine multiple-peaks events. The highest mean values of runoff coefficient, event peak and event volume are found in the northeastern and southern part of the region, at medium elevation. Tarasova et al. (2018a) obtain clusters of runoff event characteristics for Germany, identifying regions with homogenous runoff response. Catchments displaying the highest event characteristics in North-Western Italy share

some common features with the ones included in cluster 1 of Tarasova et al. (2018a), i.e. western mountainous catchments (the Rhenish Massif, Black Forest and Swabian Jura) and the Bavarian Forest catchments, which display very high runoff coefficients, a moderate melt component and a quite fast response. Catchments located in the western mountainous area of North-Western Italy and, especially, in the central Po Valley display lower values of event characteristics such as runoff coefficient, peak and volume. They result comparable to catchments of cluster 2 and cluster 4 described in Tarasova et al. (2018a), mainly representing the Alpine Foreland and the North German Plain, respectively.

The runoff event classification, performed first on observed runoff based on Tarasova et al. (2020), shows that in cluster 1 intensity-dominated (mainly unsteady) events and volume-dominated local events are relevant types, suggesting that the main runoff mechanism is local and both infiltration excess and event-fed saturation are possible. This in turns indicates convective activity, typical in summer, as very likely in these catchments. This cluster shows strong similarities with cluster 5 of Tarasova et al. (2020), mainly grouping catchments located at low elevation, in Germany. Cluster 2 identified here, instead, is characterized by extensive and steady rainfall events in dry conditions, both intensity and volume dominated. Considering the proximity with mountainous chains, this suggests the occurrence of slow moving orographic storms. This cluster shares common features with cluster 6 found in Tarasova et al. (2020), which occupies a large portion of Southern and Central Germany. Similarly, in cluster 4 of this analysis extensive volume-dominated events are prevalent, with a lower fraction of intensity-dominated events and a higher presence of snow-impacted events, meaning that extensive event fed saturation, with possible event-induced connectivity, is the most likely runoff trigger. Finally, in cluster 3 the sum of all snow-impacted events represents more than 50% of total events. It is interesting to note that in this study the number of dry events is way higher than the number of wet events in all the identified clusters, and this matches with the results found in Tarasova et al. (2020) for clusters sharing similar features such as cluster 5 and 6, despite the thresholds used for separating wet and dry events are different in the two studies. Other classification schemes have been recently adopted in literature, such as the one proposed by Brazda et al. (2022) for floods over catchments located in Europe and North-America. The method is based on the one proposed by Sikorska et al. (2015), with an additional distinction between wet and dry events. The results highlight that snowmelt floods are a predominant flood type in the selected catchments, especially for low and medium-elevation catchments where

snow-related floods constitute a much relevant portion of all events. This seems not to be the case for the catchments analyzed in this Dissertation, even if snow-related events constitute the dominant fraction in cluster 3. Moreover, differently from our results, Brazda et al. (2022) show a net predominance of events with wet initial conditions over dry events. However, these differences can be expected due to the higher average latitude and the different climatic conditions of the catchments considered in their study, possibly impacting both the antecedent wetness state before floods and the flood types themselves.

The event typology obtained from simulated events in ungauged catchments is consistent with the one obtained from observed data, both in terms of event frequency and spatial pattern of clusters. The event type frequency of simulated events for cluster 2, 3 and 4 proves to be very similar to that of the corresponding observed clusters, while in cluster 1 the unsteady component, both intensity and volume dominated, is more relevant compared to observations. Moreover, in cluster 3 the fraction of *Snowmelt* events is higher than the fraction of *Rain-on-snow* events, as opposed to the results for observed events. This can be explained by the fact that simulations span the period 1961-2020 so they also reflect past climatic conditions which favored colder temperatures, even during spring, leading to likely pure snowmelt events. At the contrary, *Rain-on-snow* events become more important in the context of climate change, as they are a signal of the transition between snow and rain induced by the increase of temperature. Due to the fact that available discharge series mainly cover the period 2000-2020, global warming is likely to have an impact on the classification results.

The spatial distribution of the event characteristics reflects the results obtained by event types classification. In fact, the highest values of runoff coefficient, event peak and volume are found in the northeastern and southern part of the region, at altitude range of 1000-1500 metres, where both rain-on-snow events and orographic slow-moving storms play an important role (cluster 2 and 4). These event types are typically associated with high values of runoff coefficient and effective runoff volume, which are also likely determined by the presence of shallow, low-permeable soils. Catchments located in the central valley and in the western Alpine area (cluster 1 and 3) generally display lower values of mean runoff event characteristics. On the one hand, this can be explained by local runoff generation patterns, typically occurring when the inter-event evapotranspiration is a consistent contribution to the water balance, controlling the soil moisture of lowland catchments with high

water storage. On the other hand, the enhanced effect of global warming in high elevation Alpine catchments is also responsible for decreased runoff peaks and volumes. Moreover, cluster 1 of simulated events shows the dominance of the unsteady component, which is likely associated to lower runoff coefficient, peak and volume.

To conclude, the analysis performed in the second part of this Dissertation provides mean regional maps of runoff event characteristics and event type frequency that give an indication of the processes and mechanisms that spatially control runoff (and flood) generation in North-Western Italy. Despite the already mentioned limitations related to the application of the methodology in a snow-impacted area, with the current availability of precipitation and temperature data from NWOI dataset, covering a 60-year period from 1961 to 2020, and of classified simulated runoff events in the same period, a possible following step of the work can be performing a statistical analysis, similarly to what done in Chapter 2, with the aim of describing the temporal variability of flood event type characteristics. This will allow to detect possible long-term trends in the event characteristics and potentially attribute the tendencies to variations in the local climate. Studying trend detection and attribution of different river flood event types can further stimulate the discussion of the research community on the topic of climate change impact on water extremes, which is still very debated.

References

Acquaotta, F., Fratianni, S., and Garzena, D. (2014). Temperature changes in the North-Western Italian Alps from 1961 to 2010. *Theoretical and Applied Climatology*, 122(3-4):619–634.

Addor, N., Newman, A. J., Mizukami, N., and Clark, M. P. (2017). The CAMELS data set: catchment attributes and meteorology for large-sample studies. *Hydrology and Earth System Sciences*, 21(10):5293–5313.

Alfieri, L., Burek, P., Feyen, L., and Forzieri, G. (2015). Global warming increases the frequency of river floods in Europe. *Hydrology and Earth System Sciences*, 19(5):2247–2260.

Archfield, S. A., Clark, M., Arheimer, B., Hay, L. E., McMillan, H., Kiang, J. E., Seibert, J., Hakala, K., Bock, A., Wagener, T., Farmer, W. H., Andréassian, V., Attinger, S., Viglione, A., Knight, R., Markstrom, S., and Over, T. (2015). Accelerating advances in continental domain hydrologic modeling. *Water Resources Research*, 51(12):10078–10091.

Arrighi, C. and Domeneghetti, A. (2024). Brief communication: On the environmental impacts of the 2023 floods in emilia-romagna (italy). *Natural Hazards and Earth System Sciences*, 24(2):673–679.

Barker, L., Hannaford, J., Muchan, K., Turner, S., and Parry, S. (2016). The winter 2015/2016 floods in the UK: a hydrological appraisal. *Weather*, 71(12):324–333.

Bayliss, A. C. and Jones, R. C. (1993). *Peaks-over-threshold flood database: Summary statistics and seasonality*. Institute of Hydrology, Wallingford, UK, 1993.

Becker, R., Koppa, A., Schulz, S., Usman, M., aus der Beek, T., and Schüth, C. (2019). Spatially distributed model calibration of a highly managed hydrological system using remote sensing-derived ET data. *Journal of Hydrology*, 577:123944.

Beniston, M., Farinotti, D., Stoffel, M., Andreassen, L. M., Coppola, E., Eckert, N., Fantini, A., Giacona, F., Hauck, C., Huss, M., Huwald, H., Lehning, M., López-Moreno, J.-I., Magnusson, J., Marty, C., Morán-Tejéda, E., Morin, S., Naaim, M., Provenzale, A., Rabatel, A., Six, D., Stötter, J., Strasser, U., Terzago, S., and Vincent, C. (2018). The European mountain cryosphere: a review of its current state, trends, and future challenges. *The Cryosphere*, 12(2):759–794.

Berghuijs, W. R., Harrigan, S., Molnar, P., Slater, L. J., and Kirchner, J. W. (2019). The relative importance of different flood-generating mechanisms across Europe. *Water Resources Research*, 55(6):4582–4593.

Berghuijs, W. R., Woods, R. A., Hutton, C. J., and Sivapalan, M. (2016). Dominant flood generating mechanisms across the United States. *Geophysical Research Letters*, 43(9):4382–4390.

Bergström, S. (1995). The HBV model. *Computer models of watershed hydrology*, pages 443–476.

Bertola, M., Viglione, A., and Blöschl, G. (2019). Informed attribution of flood changes to decadal variation of atmospheric, catchment and river drivers in Upper Austria. *Journal of Hydrology*, 577:123919.

Bertola, M., Viglione, A., Lun, D., Hall, J., and Blöschl, G. (2020). Flood trends in Europe: are changes in small and big floods different? *Hydrology and Earth System Sciences*, 24(4):1805–1822.

Bertola, M., Viglione, A., Vorogushyn, S., Lun, D., Merz, B., and Blöschl, G. (2021). Do small and large floods have the same drivers of change? a regional attribution analysis in Europe. *Hydrology and Earth System Sciences*, 25(3):1347–1364.

Best, D. and Roberts, D. (1975). Algorithm AS 89: the upper tail probabilities of Spearman's rho. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 24(3):377–379.

Beven, K. (1996). Equifinality and uncertainty in geomorphological modelling. In Thorn, C. and Rhoads, B., editors, *The Scientific Nature of Geomorphology: Proceedings of the 27th Binghamton Symposium in Geomorphology, held 27–29 September 1996*, volume 27. John Wiley & Sons, NJ, USA.

Beven, K. (2001). How far can we go in distributed hydrological modelling? *Hydrology and Earth System Sciences*, 5(1):1–12.

Beven, K. and Freer, J. (2001). Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the glue methodology. *Journal of Hydrology*, 249(1–4):11–29.

Blöschl, G. (2005). Rainfall-runoff modeling of ungauged catchments.

Blöschl, G., Hall, J., Parajka, J., Perdigão, R. A. P., Merz, B., Arheimer, B., Aronica, G. T., Bilibashi, A., Bonacci, O., Borga, M., Čanjevac, I., Castellarin, A., Chirico, G. B., Claps, P., Fiala, K., Frolova, N., Gorbachova, L., Gül, A., Hannaford, J., Harrigan, S., Kireeva, M., Kiss, A., Kjeldsen, T. R., Kohnová, S., Koskela, J. J., Ledvinka, O., Macdonald, N., Mavrova-Guirguinova, M., Mediero, L., Merz, R., Molnar, P., Montanari, A., Murphy, C., Osuch, M., Ovcharuk, V., Radevski, I., Rogger, M., Salinas, J. L., Sauquet, E., Šraj, M., Szolgay, J., Viglione, A., Volpi, E., Wilson, D., Zaimi, K., and Živković, N. (2017). Changing climate shifts timing of European floods. *Science*, 357(6351):588–590.

Blöschl, G., Hall, J., Viglione, A., Perdigão, R. A. P., Parajka, J., Merz, B., Lun, D., Arheimer, B., Aronica, G. T., Bilibashi, A., Boháč, M., Bonacci, O., Borga, M., Čanjevac, I., Castellarin, A., Chirico, G. B., Claps, P., Frolova, N., Ganora, D., Gorbachova, L., Gül, A., Hannaford, J., Harrigan, S., Kireeva, M., Kiss, A., Kjeldsen, T. R., Kohnová, S., Koskela, J. J., Ledvinka, O., Macdonald, N., Mavrova-Guirguinova, M., Mediero, L., Merz, R., Molnar, P., Montanari, A., Murphy, C., Osuch, M., Ovcharuk, V., Radevski, I., Salinas, J. L., Sauquet, E., Šraj, M., Szolgay, J., Volpi, E., Wilson, D., Zaimi, K., and Živković, N. (2019). Changing climate both increases and decreases European river floods. *Nature*, 573(7772):108–111.

Blöschl, G., Nester, T., Komma, J., Parajka, J., and Perdigão, R. A. P. (2013). The June 2013 flood in the Upper Danube Basin, and comparisons with the 2002, 1954 and 1899 floods. *Hydrology and Earth System Sciences*, 17(12):5197–5212.

Blöschl, G., Sivapalan, M., Wagener, T., Savenije, H., and Viglione, A. (2013). *Runoff prediction in ungauged basins: synthesis across processes, places and scales*. Cambridge University Press.

Brazda, S., Šraj, M., and Bezak, N. (2022). Classification of floods in Europe and North America with focus on compound events. *ISPRS International Journal of Geo-Information*, 11(12):580.

Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J. (1984). *Classification And Regression Trees*. Wadsworth.

Brunetti, M., Lentini, G., Maugeri, M., Nanni, T., Auer, I., Boehm, R., and Schoener, W. (2009). Climate variability and change in the Greater Alpine Region over the last two centuries based on multi-variable analysis. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, 29(15):2197–2225.

Budyko, M. I. (1974). *Climate and life*. D. H. Miller, Academic Press.

Cassardo, C., Cremonini, R., Gandini, D., Paesano, G., Pelosini, R., and Qian, M. (2001). Analysis of the severe flood of 13t-16th October 2000 in Piedmont (Italy). *Cuadernos de Investigación Geográfica*, 27:147–162.

Clark, M. P., Schaefli, B., Schymanski, S. J., Samaniego, L., Luce, C. H., Jackson, B. M., Freer, J. E., Arnold, J. R., Moore, R. D., Istanbulluoglu, E., and Ceola, S. (2016). Improving the theoretical underpinnings of process-based hydrologic models. *Water Resources Research*, 52(3):2350–2365.

Conover, W. J. (1999). *Practical nonparametric statistics*, volume 350. john wiley & sons.

Doorenbos, J. and Pruitt, W. (1977). Crop water requirements. FAO irrigation and drainage paper 24. *Land and Water Development Division, FAO, Rome*, 144(1).

Dunne, T. (1978). Field studies of hillslope flow processes. *Hillslope Hydrology*, pages 227–293.

ECRINS (2012). EEA Catchments and Rivers Network System–ECRINS v1.1. Technical Report 7, European Environment Agency.

Falkenmark, M. and Chapman, T. (1989). *Comparative hydrology: An ecological approach to land and water resources*. UNESCO.

Fernandez, W., Vogel, R. M., and Sankarasubramanian, A. (2000). Regional calibration of a watershed model. *Hydrological Sciences Journal*, 45(5):689–707.

Fioravanti, G., Piervitali, E., and Desiato, F. (2015). Recent changes of temperature extremes over Italy: an index-based analysis. *Theoretical and Applied Climatology*, 123(3-4):473–486.

Fisher, R. A. (1915). Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika*, 10(4):507–521.

Giani, G., Rico-Ramirez, M. A., and Woods, R. A. (2021). A practical, objective, and robust technique to directly estimate catchment response time. *Water Resources Research*, 57(2).

Giani, G., Tarasova, L., Woods, R. A., and Rico-Ramirez, M. A. (2022). An objective time-series-analysis method for rainfall-runoff event identification. *Water Resources Research*, 58(2).

Goswami, M., O'Connor, K., and Bhattarai, K. (2007). Development of regionalisation procedures using a multi-model approach for flow simulation in an ungauged catchment. *Journal of Hydrology*, 333(2–4):517–531.

Grazzini, F., Fragkoulidis, G., Pavan, V., and Antolini, G. (2020). The 1994 Piedmont flood: an archetype of extreme precipitation events in Northern Italy. *Bulletin of Atmospheric Science and Technology*, 1(3-4):283–295.

Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, 377(1-2):80–91.

Gupta, H. V., Perrin, C., Blöschl, G., Montanari, A., Kumar, R., Clark, M., and Andréassian, V. (2014). Large-sample hydrology: a need to balance depth with breadth. *Hydrology and Earth System Sciences*, 18(2):463–477.

Hall, J., Arheimer, B., Aronica, G. T., Bilibashi, A., Boháč, M., Bonacci, O., Borga, M., Burlando, P., Castellarin, A., Chirico, G. B., Claps, P., Fiala, K., Gaál, L., Gorbachova, L., Gül, A., Hannaford, J., Kiss, A., Kjeldsen, T., Kohnová, S., Koskela, J. J., Macdonald, N., Mavrova-Guirguinova, M., Ledvinka, O., Mediero, L., Merz, B., Merz, R., Molnar, P., Montanari, A., Osuch, M., Parajka, J., Perdigão, R. A. P., Radevski, I., Renard, B., Rogger, M., Salinas, J. L., Sauquet, E., Šraj, M., Szolgay, J., Viglione, A., Volpi, E., Wilson, D., Zaimi, K., and Blöschl, G. (2015). A European Flood Database: facilitating comprehensive flood research beyond administrative boundaries. *Proceedings of the International Association of Hydrological Sciences*, 370:89–95.

Hanus, S., Hrachowitz, M., Zekollari, H., Schoups, G., Vizcaino, M., and Kaitna, R. (2021). Future changes in annual, seasonal and monthly runoff signatures in contrasting Alpine catchments in Austria. *Hydrology and Earth System Sciences*, 25(6):3429–3453.

He, Y., Bárdossy, A., and Zehe, E. (2011). A review of regionalisation for continuous streamflow simulation. *Hydrology and Earth System Sciences*, 15(11):3539–3553.

Helsel, D. R., Hirsch, R. M., Ryberg, K. R., Archfield, S. A., and Gilroy, E. J. (2020). *Statistical methods in water resources*.

Hirpa, F. A., Salamon, P., Beck, H. E., Lorini, V., Alfieri, L., Zsoter, E., and Dadson, S. J. (2018). Calibration of the Global Flood Awareness System (GloFAS) using daily streamflow data. *Journal of Hydrology*, 566:595–606.

Hirschboeck, K. K. (1987). Hydroclimatically-defined mixed distributions in partial duration flood series. In *Hydrologic Frequency Modeling: Proceedings of the International Symposium on Flood Frequency and Risk Analyses, 14–17 May 1986, Louisiana State University, Baton Rouge, USA*, pages 199–212. Springer.

Holland, J. (1975). Adaptation in natural and artificial systems. *University of Michigan Press*.

Hollander, M., Wolfe, D. A., and Chicken, E. (1999). Nonparametric statistical methods john wiley & sons. *New York*, 57:58–59.

Horton, R. E. (1933). The rôle of infiltration in the hydrologic cycle. *Transactions, American Geophysical Union*, 14(1):446.

Hundecha, Y. and Bárdossy, A. (2004). Modeling of the effect of land use changes on the runoff generation of a river basin through parameter regionalization of a watershed model. *Journal of Hydrology*, 292(1–4):281–295.

Hundecha, Y., Ouarda, T. B. M. J., and Bárdossy, A. (2008). Regional estimation of parameters of a rainfall-runoff model at ungauged watersheds using the "spatial" structures of the parameters within a canonical physiographic-climatic space. *Water Resources Research*, 44(1).

IPCC (2019). *IPCC special report on the ocean and cryosphere in a changing climate*. Cambridge University Press, Cambridge, UK.

IPCC (2022). *Climate change 2022: impacts, adaptation and vulnerability. Contribution of working group II to the sixth assessment report of the intergovernmental panel on climate change*. Cambridge University Press, Cambridge, UK.

JRC (2008). *CCM2 River and Catchment Database for Europe, Version 2.1 Release Notes*. European Commission-Joint Research Centre–Institute for Environment and Sustainability, Ispra, Italy.

Kendall, M. G. (1948). *Rank correlation methods.* Griffin.

Kim, U. and Kaluarachchi, J. J. (2008). Application of parameter estimation and regionalization methodologies to ungauged basins of the Upper Blue Nile River Basin, Ethiopia. *Journal of Hydrology*, 362(1-2):39–56.

Kling, H., Fuchs, M., and Paulin, M. (2012). Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. *Journal of Hydrology*, 424-425:264–277.

Kokkonen, T. S., Jakeman, A. J., Young, P. C., and Koivusalo, H. J. (2003). Predicting daily flows in ungauged catchments: model regionalization from catchment descriptors at the coweeta hydrologic laboratory, north carolina. *Hydrological Processes*, 17(11):2219–2238.

Kreienkamp, F., Philip, S. Y., Tradowsky, J. S., Kew, S. F., Lorenz, P., Arrighi, J., Belleflamme, A., Bettmann, T., Caluwaerts, S., Chan, S. C., et al. (2021). Rapid attribution of heavy rainfall events leading to the severe flooding in Western Europe during July 2021. *World Weather Atribution*.

Lamb, R., Crewett, J., and Kay, A. (2000). Progress in the spatial generalisation of "continuos simulation" flood frequency modeling. In Toensmann, F. and Koch, M., editors, *River flood defence. Vol. 1. Proceedings, Kassel Reports of Hydraulic Engineering*, volume 9/2000, pages D117–D125.

Lamb, R. and Kay, A. L. (2004). Confidence intervals for a spatially generalized, continuous simulation flood frequency model for Great Britain. *Water Resources Research*, 40(7).

Li, H., Zhang, Y., Chiew, F. H., and Xu, S. (2009). Predicting runoff in ungauged catchments by using xinanjiang model with modis leaf area index. *Journal of Hydrology*, 370(1–4):155–162.

Libertino, A., Ganora, D., and Claps, P. (2019). Evidence for increasing rainfall extremes remains elusive at large spatial scales: The case of Italy. *Geophysical Research Letters*, 46(13):7437–7446.

Liu, D. (2020). A rational performance criterion for hydrological model. *Journal of Hydrology*, 590:125488.

Lun, D., Viglione, A., Bertola, M., Komma, J., Parajka, J., Valent, P., and Blöschl, G. (2021). Characteristics and process controls of statistical flood moments in Europe – a data-based analysis. *Hydrology and Earth System Sciences*, 25(10):5535–5560.

Mann, H. B. (1945). Nonparametric tests against trend. *Econometrica: Journal of the econometric society*, pages 245–259.

Mardia, K. V. (1972). *Statistics of Directional Data*. Academic Press.

McIntyre, N., Lee, H., Wheater, H., and Young, A. (2004). Tools and approaches for evaluating uncertainty in streamflow predictions in ungauged uk catchments. In Pahl-Wostl, C., Schmidt, S., and Jakeman, T., editors, *Proceedings of the International Congress on Environmental Modelling and Software*.

Mediero, L., Santillán, D., Garrote, L., and Granados, A. (2014). Detection and attribution of trends in magnitude, frequency and timing of floods in Spain. *Journal of Hydrology*, 517:1072–1088.

Merz, R. and Blöschl, G. (2003). A process typology of regional floods. *Water Resources Research*, 39(12).

Merz, R. and Blöschl, G. (2004). Regionalisation of catchment model parameters. *Journal of Hydrology*, 287(1-4):95–123.

Merz, R. and Blöschl, G. (2005). Flood frequency regionalisation—spatial proximity vs. catchment attributes. *Journal of Hydrology*, 302(1-4):283–306.

Merz, R., Tarasova, L., and Basso, S. (2020). Parameter's controls of distributed catchment models—How much information is in conventional catchment descriptors? *Water Resources Research*, 56(2).

Miller, J. D., Kjeldsen, T. R., Hannaford, J., and Morris, D. G. (2012). A hydrological assessment of the November 2009 floods in Cumbria, UK. *Hydrology Research*, 44(1):180–197.

Mishra, S. K., Jain, M. K., Suresh Babu, P., Venugopal, K., and Kaliappan, S. (2008). Comparison of AMC-dependent CN-conversion formulae. *Water Resources Management*, 22(10):1409–1420.

Mitchell, M. (1998). *An introduction to genetic algorithms*. MIT press.

Mizukami, N., Clark, M. P., Newman, A. J., Wood, A. W., Gutmann, E. D., Nijssen, B., Rakovec, O., and Samaniego, L. (2017). Towards seamless large-domain parameter estimation for hydrologic models. *Water Resources Research*, 53(9):8020–8040.

Mizukami, N., Rakovec, O., Newman, A. J., Clark, M. P., Wood, A. W., Gupta, H. V., and Kumar, R. (2019). On the choice of calibration metrics for "high-flow" estimation using hydrologic models. *Hydrology and Earth System Sciences*, 23(6):2601–2614.

Mullen, K., Ardia, D., Gil, D., Windover, D., and Cline, J. (2011). Deoptim: An R package for global optimization by differential evolution. *Journal of Statistical Software*, 40(6):1–26.

Nash, J. E. (1960). A unit hydrograph study, with particular reference to British catchments. *Proceedings of the Institution of Civil Engineers*, 17(3):249–282.

Newman, A. J., Clark, M. P., Sampson, K., Wood, A., Hay, L. E., Bock, A., Viger, R. J., Blodgett, D., Brekke, L., Arnold, J. R., Hopson, T., and Duan, Q. (2015). Development of a large-sample watershed-scale hydrometeorological data set for the contiguous USA: data set characteristics and assessment of regional variability in hydrologic model performance. *Hydrology and Earth System Sciences*, 19(1):209–223.

Nied, M., Pardowitz, T., Nissen, K., Ulbrich, U., Hundecha, Y., and Merz, B. (2014). On the relationship between hydro-meteorological patterns and flood types. *Journal of Hydrology*, 519:3249–3262.

Oudin, L., Kay, A., Andréassian, V., and Perrin, C. (2010). Are seemingly physically similar catchments truly hydrologically similar? *Water Resources Research*, 46(11).

Paniconi, C. and Putti, M. (2015). Physically based modeling in catchment hydrology at 50: Survey and outlook. *Water Resources Research*, 51(9):7090–7129.

Parajka, J. and Blöschl, G. (2008). The value of MODIS snow cover data in validating and calibrating conceptual hydrologic models. *Journal of Hydrology*, 358(3-4):240–258.

Parajka, J., Blöschl, G., and Merz, R. (2007a). Regional calibration of catchment models: Potential for ungauged catchments. *Water Resources Research*, 43(6).

Parajka, J., Merz, R., and Blöschl, G. (2005). A comparison of regionalisation methods for catchment model parameters. *Hydrology and Earth System Sciences*, 9(3):157–171.

Parajka, J., Merz, R., and Blöschl, G. (2007b). Uncertainty and multiple objective calibration in regional water balance modelling: case study in 320 Austrian catchments. *Hydrological Processes*, 21(4):435–446.

Parajka, J., Viglione, A., Rogger, M., Salinas, J. L., Sivapalan, M., and Blöschl, G. (2013). Comparative assessment of predictions in ungauged basins – Part 1: Runoff-hydrograph studies. *Hydrology and Earth System Sciences*, 17(5):1783–1795.

Pesce, M., Viglione, A., von Hardenberg, J., Tarasova, L., Basso, S., Merz, R., Parajka, J., and Tong, R. (2024). Regional multi-objective calibration for distributed hydrological modelling: a decision tree based approach. *Proceedings of IAHS*, 385:65–69.

Pesce, M., von Hardenberg, J., Claps, P., and Viglione, A. (2022). Correlation between climate and flood indices in Northwestern Italy at different temporal scales. *Journal of Hydrology and Hydromechanics*, 70(2):178–194.

Peterson, T. (2005). Climate change indices. *World meteorological organization bulletin*, 54(2):83.

Petrow, T. and Merz, B. (2009). Trends in flood magnitude, frequency and seasonality in Germany in the period 1951–2002. *Journal of Hydrology*, 371(1-4):129–141.

Pool, S., Vis, M., and Seibert, J. (2018). Evaluating model performance: towards a non-parametric variant of the Kling-Gupta efficiency. *Hydrological Sciences Journal*, 63(13-14):1941–1953.

Price, K., Storn, R. M., and Lampinen, J. A. (2006). *Differential evolution: a practical approach to global optimization*. Springer Science & Business Media.

Prosdocimi, I., Kjeldsen, T. R., and Miller, J. D. (2015). Detection and attribution of urbanization effect on flood extremes using nonstationary flood-frequency models. *Water Resources Research*, 51(6):4244–4262.

Prosdocimi, I., Kjeldsen, T. R., and Svensson, C. (2014). Non-stationarity in annual and seasonal series of peak flow and precipitation in the UK. *Natural Hazards and Earth System Sciences*, 14(5):1125–1144.

Quintero, F., Krajewski, W. F., Seo, B.-C., and Mantilla, R. (2020). Improvement and evaluation of the Iowa Flood Center Hillslope Link Model (HLM) by calibration-free approach. *Journal of Hydrology*, 584:124686.

Randrianasolo, A., Ramos, M. H., and Andréassian, V. (2011). Hydrological ensemble forecasting at ungauged basins: using neighbour catchments for model setup and updating. *Advances in Geosciences*, 29:1–11.

Rardin, R. L. and Uzsoy, R. (2001). Experimental evaluation of heuristic optimization algorithms: A tutorial. *Journal of Heuristics*, 7(3):261–304.

Rojas, R., Feyen, L., Bianchi, A., and Dosio, A. (2012). Assessment of future flood hazard in Europe using a large ensemble of bias-corrected regional climate simulations. *Journal of Geophysical Research: Atmospheres*, 117(D17):n/a–n/a.

Sardella, A., Palazzi, E., von Hardenberg, J., Grande, C. D., Nuntiis, P. D., Sabbioni, C., and Bonazza, A. (2020). Risk mapping for the sustainable protection of cultural heritage in extreme changing environments. *Atmosphere*, 11(7):700.

Seibert, J. (1999). Regionalisation of parameters for a conceptual rainfall-runoff model. *Agricultural and Forest Meteorology*, 98-99:279–293.

Seibert, J. and Beven, K. J. (2009). Gauging the ungauged basin: how many discharge measurements are needed? *Hydrology and Earth System Sciences*, 13(6):883–892.

Sen, P. K. (1968). Estimates of the regression coefficient based on Kendall's tau. *Journal of the American statistical association*, 63(324):1379–1389.

Sezen, C., Šraj, M., Medved, A., and Bezak, N. (2020). Investigation of rain-on-snow floods under climate change. *Applied Sciences*, 10(4):1242.

Sikorska, A. E., Viviroli, D., and Seibert, J. (2015). Flood-type classification in mountainous catchments using crisp and fuzzy decision trees. *Water Resources Research*, 51(10):7959–7976.

Silvestro, F., Gabellani, S., Giannoni, F., Parodi, A., Rebora, N., Rudari, R., and Siccardi, F. (2012). A hydrological analysis of the 4 November 2011 event in Genoa. *Natural Hazards and Earth System Sciences*, 12(9):2743–2752.

Silvestro, F., Rebora, N., Rossi, L., Dolia, D., Gabellani, S., Pignone, F., Trasforini, E., Rudari, R., Angeli, S. D., and Masciulli, C. (2016). What if the 25 October 2011 event that struck Cinque Terre (Liguria) had happened in Genoa, Italy? Flooding scenarios, hazard mapping and damage estimation. *Natural Hazards and Earth System Sciences*, 16(8):1737–1753.

Šraj, M., Viglione, A., Parajka, J., and Blöschl, G. (2016). The influence of non-stationarity in extreme hydrological events on flood frequency estimation. *Journal of Hydrology and Hydromechanics*, 64(4):426–437.

Stein, L., Pianosi, F., and Woods, R. (2019). Event-based classification for global study of river flood generating processes. *Hydrological Processes*, 34(7):1514–1529.

Storn, R. and Price, K. (1997). Differential evolution - a simple and efficient heuristic for global optimization over continuos spaces. *Journal of Global Optimization*, 11(4):341–359.

Szolgay, J., Hlavcová, K., Kohnová, S., and Danihlik, R. (2003). Regional estimation of parameters of a monthly water balance model. *Journal of Hydrology and Hydromechanics*, 51(4).

Tarasova, L., Basso, S., Poncelet, C., and Merz, R. (2018a). Exploring controls on rainfall-runoff events: 2. Regional patterns and spatial controls of event characteristics in Germany. *Water Resources Research*, 54(10):7688–7710.

Tarasova, L., Basso, S., Wendi, D., Viglione, A., Kumar, R., and Merz, R. (2020). A process-based framework to characterize and classify runoff events: The event typology of Germany. *Water Resources Research*, 56(5).

Tarasova, L., Basso, S., Zink, M., and Merz, R. (2018b). Exploring controls on rainfall-runoff events: 1. Time series-based event separation and temporal dynamics of event runoff response in Germany. *Water Resources Research*, 54(10):7711–7732.

Tarasova, L., Merz, R., Kiss, A., Basso, S., Blöschl, G., Merz, B., Viglione, A., Plötner, S., Guse, B., Schumann, A., Fischer, S., Ahrens, B., Anwar, F., Bárdossy, A., Bühler, P., Haberlandt, U., Kreibich, H., Krug, A., Lun, D., Müller-Thomy, H., Pidoto, R., Primo, C., Seidel, J., Vorogushyn, S., and Wietzke, L. (2019). Causative classification of river flood events. *WIREs Water*, 6(4).

Theil, H. (1950). A rank-invariant method of linear and polynominal regression analysis (Parts 1-3). In *Ned. Akad. Wetensch. Proc. Ser. A*, volume 53, pages 1397–1412.

Tong, R., Parajka, J., Salentinig, A., Pfeil, I., Komma, J., Széles, B., Kubáň, M., Valent, P., Vreugdenhil, M., Wagner, W., and Blöschl, G. (2021). The value of ASCAT soil moisture and MODIS snow cover data for calibrating a conceptual hydrologic model. *Hydrology and Earth System Sciences*, 25(3):1389–1410.

Troch, P. A., Paniconi, C., and van Loon, E. E. (2003). Hillslope-storage Boussinesq model for subsurface flow and variable source areas along complex hillslopes: 1. formulation and characteristic response. *Water Resources Research*, 39(11).

Ulbrich, U., Brücher, T., Fink, A. H., Leckebusch, G. C., Krüger, A., and Pinto, J. G. (2003). The central European floods of August 2002: Part 1 – Rainfall periods and flood development. *Weather*, 58(10):371–377.

USDA-NRCS (1986). *Urban Hydrology for Small Watersheds*, volume 55. US Department of Agriculture, Soil Conservation Service.

USDA-NRCS (2004). *Chapter 9: Hydrologic Soil-Cover Complexes, NRCS National Engineering Handbook, Part 630: Hydrology*. US Department of Agriculture, Natural Resources Conservation Service.

Vandewiele, G. and Elias, A. (1995). Monthly water balance of ungauged catchments obtained by geographical regionalization. *Journal of Hydrology*, 170(1–4):277–291.

Viglione, A., Chirico, G. B., Komma, J., Woods, R., Borga, M., and Blöschl, G. (2010). Quantifying space-time dynamics of flood event types. *Journal of Hydrology*, 394(1-2):213–229.

Viglione, A., Merz, B., Dung, N. V., Parajka, J., Nester, T., and Blöschl, G. (2016). Attribution of regional flood changes based on scaling fingerprints. *Water Resources Research*, 52(7):5322–5340.

Viglione, A. and Parajka, J. (2018). Tuwmodel: Lumped Hydrological Model for Education Purposes. *R package version*, pages 1–0.

Villarini, G., Smith, J. A., Serinaldi, F., and Ntelekos, A. A. (2011). Analyses of seasonal and annual maximum daily discharge records for central Europe. *Journal of Hydrology*, 399(3-4):299–312.

Viviroli, D., Mittelbach, H., Gurtz, J., and Weingartner, R. (2009). Continuous simulation for flood estimation in ungauged mesoscale catchments of Switzerland – Part II: Parameter regionalisation and flood estimation results. *Journal of Hydrology*, 377(1–2):208–225.

Wagener, T., Sivapalan, M., Troch, P., and Woods, R. (2007). Catchment classification and hydrologic similarity. *Geography Compass*, 1(4):901–931.

Wagener, T. and Wheater, H. S. (2006). Parameter estimation and regionalization for continuous rainfall-runoff models including uncertainty. *Journal of Hydrology*, 320(1–2):132–154.

Weingartner, R. and Diezig, R. (2007). Hochwasserprozesstypen-Schlüssel zur Hochwasserabschätzung. *Wasser und Abfall*, 9(4):18–26.

Zeng, X., Schnier, S., and Cai, X. (2021). A data-driven analysis of frequent patterns and variable importance for streamflow trend attribution. *Advances in Water Resources*, 147:103799.

Zhang, X., Hegerl, G., Zwiers, F. W., and Kenyon, J. (2005). Avoiding inhomogeneity in percentile-based indices of temperature extremes. *Journal of Climate*, 18(11):1641–1651.

Zhang, Y. and Chiew, F. H. S. (2009). Relative merits of different methods for runoff predictions in ungauged catchments. *Water Resources Research*, 45(7).

Zvolenskỳ, M., Kohnova, S., Hlavčová, K., Szolgay, J., and Parajka, J. D. (2008). Regionalisation of rainfall-runoff model parameters based on geographical location of gauged catchments. *Journal of hydrology and hydromechanics*, 56(3):176–189.

Appendix A

Summary information of catchments

Catchment	Coord.E [UTM32N]	Coord.N [UTM32N]	Catchment Area [km^2]	Mean Elevation [m a.s.l.]	Mean annual precipitation [mm yr^{-1}]	Mean annual PET [mm yr^{-1}]	Aridity Index [-]
Agogna a Momo	464709	5046499	204	423	1300	934	0.718
Agogna a Novara	467981	5030915	382	332	1175	939	0.799
Anza a Macugnaga	418994	5091217	41	2657	1237	330	0.267
Artanavaz a Saint Oyen	360765	5075765	69	2239	866	443	0.512
Ayasse a Champorcher	392220	5052946	43	2368	1114	582	0.522
Banna a Poirino	407635	4975171	262	288	764	914	1.196
Banna a Santena	403910	4977515	350	287	758	915	1.207
Belbo a Castelnuovo Belbo	454039	4960718	421	372	742	909	1.225
Belbo a Rocchetta Belbo	434525	4942715	119	622	787	870	1.105
Bogna a Domodossola	443911	5107834	81	1612	1467	624	0.425
Borbera a Baracche	500788	4951811	202	862	1129	812	0.719
Borbera a Pertuso	502162	4952371	193	875	1134	809	0.713
Borbore a San Damiano d'Asti	427419	4965805	85	244	722	914	1.266
Bormida a Alessandria	472015	4972514	2594	440	961	885	0.921
Bormida a Cassine	463786	4955433	1516	493	925	865	0.935
Bormida di Mallare a Ferrania	446089	4912508	50	604	1162	871	0.750
Bormida di Millesimo a Camerana	432905	4920644	263	767	1014	816	0.805
Bormida di Millesimo a Cessole	440472	4944272	496	647	903	838	0.928
Bormida di Millesimo a Murialdo	432976	4907028	127	888	1041	778	0.747
Bormida di Spigno a Mombaldone	447206	4935280	391	491	1005	871	0.867
Bormida di Spigno a Piana Crixia	444555	4926519	252	523	1058	869	0.821
Bormida di Spigno a Valla	447475	4932026	68	468	1004	870	0.867
Bousset a Tetti Porcera	375046	4896152	37	1991	1263	720	0.570
Breuil a Alpette	336956	5064247	28	2459	928	478	0.515
Brobbio a Margarita	395470	4917082	127	729	1111	875	0.788
Bucera a Ponte Rovine	370555	4896060	28	2152	1306	613	0.469
Buthier a Place Moulin	383131	5083798	74	2800	900	390	0.433
Cannobino a Traffiume	475222	5100470	107	1107	1926	808	0.420
Casotto a Monasterolo Casotto	415233	4906954	66	1233	1048	749	0.715
Cenischia a Susa	347163	5000542	146	2033	881	672	0.763
Cervo a Passobreve	425027	5053441	75	1505	1569	726	0.463
Cervo a Quinto Vercellese	451013	5025503	1002	504	1180	879	0.745
Cervo a Vigliano Biellese	430621	5044910	133	1257	1500	764	0.509
Chiavanne a Alpette	336855	5064354	22	2485	923	489	0.530
Chisola a La Loggia	395104	4980550	464	322	912	943	1.034
Chisola a Fenestrelle	346006	4988947	154	2154	794	579	0.729
Chisone a Pinerolo	368464	4969243	586	1727	973	680	0.699
Chisone a San Martino	364406	4971647	581	1739	972	677	0.697
Chisone a Soucheres Basses	339221	4988138	93	2221	763	528	0.692
Chiusella a Gurzia	402638	5030792	142	1374	1318	793	0.602
Chiusella a Parella	405965	5030307	152	1308	1308	799	0.611
Clarea a presa centrale Chiomonte	340471	5000302	28	2298	858	565	0.659
Corsaglia a Frabosa Soprana	407136	4903042	64	1678	1156	752	0.651
Corsaglia a Presa Centrale Moline	407096	4904936	89	1525	1142	721	0.631
Corsaglia a Torre Mondovì	4125545	4912293	140	1266	1119	714	0.638
Curone a Volpedo	498654	4970400	154	636	849	861	1.014

Catchment	Coord.E [UTM32N]	Coord.N [UTM32N]	Catchment Area [km^2]	Mean Elevation [m a.s.l.]	Mean annual precipitation [mm yr^{-1}]	Mean annual PET [mm yr^{-1}]	Aridity Index [-]
Devero a Baceno	447186	5123403	106	2057	1362	650	0.477
Dora Baltea a Aosta	374919	5066248	1846	2265	835	565	0.677
Dora Baltea a Beauregard	348670	5053499	92	2607	831	546	0.657
Dora Baltea a Cignana	390990	5081622	13	2725	899	367	0.408
Dora Baltea a Mazzè	417900	5017085	3854	1883	923	617	0.668
Dora Baltea a Ponte Baio	409329	5039992	3359	2071	895	578	0.646
Dora Baltea a Ponte di Mombardone	344193	5069807	372	2412	933	505	0.541
Dora Baltea a Tavagnasco	408421	5044504	3321	2087	890	575	0.646
Dora Baltea a Verolengo	424689	5004726	3903	1862	923	621	0.673
Dora di Bardonecchia a Beaulard	323449	4990582	207	2193	795	572	0.719
Dora di Courmayeur a Prè Saint Didier	343253	5070108	222	2472	978	497	0.508
Dora di Rhemes a Notre Dame	353389	5048222	69	2664	867	460	0.531
Dora di Rhemes a Pèlaud	352495	5046587	57	2710	880	449	0.510
Dora di Rhemes a Saint Georges	356155	5057480	119	2485	829	494	0.596
Dora Riparia a Chiomonte	340659	4998256	623	2083	766	582	0.760
Dora Riparia a Oulx	329337	4988965	260	2164	767	547	0.713
Dora Riparia a Salbertrand	334481	4993854	548	2120	765	574	0.750
Dora Riparia a Sant'Antonino di Susa	362771	4996936	1041	1899	827	633	0.765
Dora Riparia a Susa	346587	5000202	694	2042	776	588	0.758
Dora Riparia a Torino Ponte Washington	399094	4992175	1322	1655	868	678	0.781
Ellero a Mondovì	406021	4915888	180	1089	1139	688	0.604
Ellero a Rastello	398721	4902052	43	1853	1238	713	0.576
Ellero a Carisio	437996	5029759	256	603	1201	831	0.692
Ellero a Sordevolo	419328	5047278	34	1533	1371	683	0.498
Ellero a Cartosio	454040	4935727	212	499	1088	840	0.772
Ellero a Sassello	457432	4926268	91	595	1267	840	0.663
Evancon a Brusson	398845	5070594	134	2491	896	413	0.461
Evancon a Champoluc	400645	5075837	102	2634	930	376	0.404
Galamba a presa centrale Chiomonte	336552	4995777	15	2410	775	623	0.804
Germanasca a Perrero	355062	4978653	189	1902	1013	642	0.634
Gesso a Andonno	375079	4905369	336	1832	1229	677	0.551
Gesso della Barra a San Giacomo	371050	4892659	20	2103	1286	648	0.504
Gesso della Valletta a San Lorenzo	371071	4901660	111	2095	1238	594	0.480
Gesso di Entracque a Diga di Piastra	371532	4898058	87	2047	1283	658	0.513
Gesso di Entracque a Entracque	368094	4901309	160	1880	1272	686	0.539
Gesso di Monte Colombo a San Giacomo	371309	4892550	25	2160	1258	699	0.556
Ghiandone a Staffarda	375770	4953868	129	605	1061	826	0.779
Grana a Monterosso	366585	4918708	110	1542	1000	768	0.768
Grand'Eyvia a Cretaz	370093	5053194	180	2582	878	595	0.678
Isorno a Pontetto	448127	5111126	70	1617	1566	742	0.474
Lys a d'Ejola	407796	5078893	30	3136	1062	240	0.226
Lys a Gressoney Saint Jean	408689	5071087	91	2645	1004	419	0.417
Lys a Guillemore	411124	5058000	201	2260	1102	535	0.485
Maira a Busca	379305	4930199	573	1689	937	675	0.720
Maira a Racconigi	394538	4957991	976	1326	922	751	0.815
Maria a San Damiano Macra	361439	4927199	452	1899	935	627	0.671
Maira a Saretto	335576	4927181	56	2411	953	593	0.622
Malone a Brandizzo	409801	5003811	333	439	1147	917	0.799
Malone a Front	395360	5015295	131	675	1301	894	0.687
Marmore a Perreres	392605	5084497	54	2739	931	279	0.300
Mastallone a Varallo	442190	5075482	147	1335	1614	775	0.480
Melezet a Melezet	315867	4991349	43	2389	896	402	0.449
Meris a Sant'Anna Valdieri	365656	4900453	23	2110	1174	640	0.545
Mongia a Mombasiglio	417497	4913268	54	818	1059	790	0.746
Negrone a Ormea	404967	4886432	68	1798	1223	703	0.575
Niguglia a Omegna	454127	5080362	125	655	1635	897	0.549
Nontey a Valnontey	370540	5049494	53	2781	828	570	0.688
Orba a Basaluzzo	474020	4957173	731	467	1151	896	0.778
Orba a Casal Cermelli	470871	4964452	801	438	1112	900	0.809
Orba a Tiglieto	468888	4930332	74	814	1372	813	0.593
Orco a Cuorgnè	394628	5027688	629	1909	1198	687	0.573
Orco a Pont Canavese	391506	5030005	614	1938	1193	684	0.573
Orco a San Benigno Canavese	406313	5011059	838	1567	1196	726	0.607
Pellice a Luserna San Giovanni	361138	4963467	216	1634	1095	673	0.615
Pellice a Villafranca Piemonte	381346	4963166	953	1542	1019	712	0.699
Pesio a Carrù	411141	4924107	360	810	1075	830	0.772
Pesio a San Bartolomeo	393299	4902836	57	1571	1264	685	0.542
Po a Carignano	396682	4973652	3956	1101	931	778	0.836
Po a Casale Monferrato	4565575	4998852	13732	1257	947	750	0.792
Po a Castiglione Torinese	404428	4998161	7685	1083	935	788	0.843
Po a Crissolo	354451	4950954	38	2251	1012	354	0.350
Po at Isola Sant'Antonio	485975	4986981	25640	959	962	802	0.834
Po a Meirano Moncalieri	395654	4983807	5310	919	912	814	0.893

Catchment	Coord.E [UTM32N]	Coord.N [UTM32N]	Catchment Area [km^2]	Mean Elevation [m a.s.l.]	Mean annual precipitation [mm yr^{-1}]	Mean annual PET [mm yr^{-1}]	Aridity Index [-]
Po a Moncalieri	395895	4983881	5048	927	909	811	0.892
Po a San Mauro Torinese	404376	4998217	7682	1083	935	788	0.843
Po a San Sebastiano	416953	5002644	9114	1081	967	791	0.818
Po a Torino Murazzi	397499	4990961	5355	914	912	815	0.894
Po a Valenza	470988	4988677	17291	1117	983	776	0.789
Rea a Dogliani	416921	4931461	69	558	798	882	1.105
Rio Bagni a Bagni di Vinadio	347299	4905908	61	2142	1071	662	0.618
Rio del Piz a Pietraporzio	343090	4911605	22	2182	1082	582	0.538
Rio Freddo a Rio Freddo	354141	4904744	37	2123	1122	540	0.481
Ripa a Bousson	327797	4977949	147	2330	792	514	0.649
Rochemolles a Rochemolles	324207	4999927	24	2641	860	559	0.650
Rutor a Promise	340966	5063082	46	2525	808	540	0.668
Rutor a La Joux	341334	5061405	42	2570	806	544	0.675
San Bernardino a San Bernardino Santino	463224	5089336	122	1259	1827	772	0.423
San Bernardino a Trobaso	464983	5088294	129	1211	1828	783	0.428
San Giovanni a Possaccio	465266	5089297	54	1002	1940	859	0.443
San Giovanni a Verbania	467272	5087416	60	933	1929	869	0.450
Sangone a Moncalieri	395076	4984680	258	770	985	876	0.889
Sangone a Trana	375550	4988107	145	1125	1027	824	0.802
Savara a Eau Rousse	360294	5047742	82	2682	848	470	0.554
Savara a Fenille	359591	5055032	131	2594	828	473	0.571
Scrivia a Guazzora	490258	4986143	949	520	1061	882	0.831
Scrivia a Isola del Cantone	496640	4943533	216	668	1429	844	0.591
Scrivia a Serravalle Scrivia	488989	4943533	216	668	1229	849	0.691
Sermenza a Rimasco	427374	5078429	82	1855	1391	628	0.451
Sesia a Borgosesia Ponte Aranco	443803	5062352	703	1504	1480	721	0.487
Sesia a Campertogno	424762	5072125	171	2127	1264	619	0.490
Sesia a Palestro	463769	5014565	2464	768	1260	843	0.669
Sesia a Vercelli	455957	5019293	2214	837	1297	830	0.640
Sessera a Pray	439377	5058122	126	1174	1666	774	0.465
Soana a Pont Canavese	390596	5030736	215	1898	1222	680	0.556
Strona di Omegna a Gravellona Toce	456135	5086079	235	877	1693	788	0.465
Strona di Vallemosso a Cossato	436248	5046479	44	777	1556	790	0.508
Stura di Demonte a Fossano	398628	4930850	1327	1518	1100	700	0.636
Stura di Demonte a Gaiola	373755	4909963	560	1819	1068	658	0.616
Stura di Demonte a Pianche	349597	4907221	180	2077	1047	574	0.548
Stura di Demonte a Roccasparvera	375650	4910905	584	1785	1067	665	0.623
Stura di Demonte a Vinadio	349628	4907193	250	2082	1054	598	0.567
Stura di Lanzo a Lanzo	380982	5013879	578	1780	1183	700	0.592
Stura di Lanzo a Torino	398253	4996117	880	1368	1177	766	0.651
Stura di Viù a Germagnano	377063	5011307	233	1759	1104	729	0.660
Stura di Viù a Lago della Rossa	354795	5014526	4	2962	1192	536	0.450
Stura di Viù a Malciaussia	354204	5007535	26	2622	920	832	0.904
Stura di Viù a Usseglio	360063	5010116	76	2393	1003	688	0.686
Tanaro a Alba	422966	4950663	3385	1072	1024	776	0.758
Tanaro a Alessandria	470235	4976004	5323	767	914	832	0.910
Tanaro a Asti	437791	4970463	4123	922	971	803	0.827
Tanaro a Clavesana	413139	4926203	1485	951	1053	788	0.748
Tanaro a Farigliano	412709	4929894	1502	945	1051	790	0.752
Tanaro a Garessio	421310	4894772	249	1435	1132	736	0.650
Tanaro a Masio	453443	4968849	4535	855	947	815	0.861
Tanaro a Montecastello	475104	4977074	7956	657	928	850	0.916
Tanaro a Nucetto	425191	4910271	375	1232	1093	758	0.694
Tanaro a Ormea	413231	4888845	176	1524	1171	740	0.632
Tanaro a Piantorre	418169	4918794	500	1067	1045	782	0.748
Tanaro a Ponte di Nava	409392	4885563	149	1580	1189	739	0.622
Tanaro a San Martino Alfieri	431007	4960571	3542	1036	1011	783	0.774
Terdoppio a Caltignaga	469533	5040651	89	254	1092	922	0.844
Ticino a Miorina	474278	5058225	6655	1290	1358	709	0.522
Toce a Agaro	446314	5126747	11	2110	1334	678	0.508
Toce a Cadarese	450177	5125409	187	2150	1259	520	0.413
Toce a Candoglia	455210	5091406	1539	1677	1437	656	0.457
Toce a Codelago	444732	5131430	26	2322	1327	666	0.502
Toce a Domodossola	446394	5106791	919	1826	1414	637	0.450
Toce a Lago Busin	451099	5133475	3	2524	1342	677	0.504
Toce a Lago d'Avino	433177	5122226	6	2550	1464	552	0.377
Toce a Lago del Sabbione	449883	5141238	14	2765	1142	344	0.301
Toce a Lago Vannino	451161	5136937	12	2527	1199	444	0.370
Uzzone a Cortemilia	436164	4936487	87	559	787	862	1.095
Varaita a Castello	345174	4941766	68	2406	956	356	0.372
Varaita a Polonghera	389172	4961944	577	1389	889	698	0.785
Varaita a Rore	358666	4937361	278	2118	925	492	0.532
Varaita a Rossana	376220	4934616	402	1797	917	597	0.651
Varaita a Sampeyre	356706	4937355	247	2197	926	490	0.529
Varaita a Torrette	349657	4938438	184	2339	933	473	0.507
Vermenagna a Limone	385934	4896191	57	1675	1276	473	0.371
Vermenagna a Robilante	381656	4902050	134	1533	1266	528	0.417
Vobbia a Isola del Cantone	497957	4942898	51	723	1351	793	0.587

Appendix B

Manual of the R package 'hydroPASS'

Package 'hydroPASS'

July 1, 2024

Type Package

Title PArameter Set Shuffling (PASS) Calibration for Distributed Hydrologic Models

Version 0.1-2

Date 2023-04-30

Depends R (>= 4.0.0), stats, rpart, graphics, raster, sp

Description R functions to perform the PArameter Set Shuffling (PASS) approach for the calibration of distributed hydrologic models, as explained in Merz et al. (2020) \{ }doi{ 10.1029/2019WR026008 }. The package contains also the SAme Like The Others (SALTO) rainfall-runoff model as an example.

NeedsCompilation yes

Encoding UTF-8

License GPL (>= 3)

R topics documented:

`hydroPASS-package` *PArameter Set Shuffling (PASS) Calibration for Distributed Hydrologic Models*

Description

R functions to perform the PArameter Set Shuffling (PASS) approach for the calibration of distributed hydrologic models, as explained in Merz et al. (2020) <doi:10.1029/2019WR026008>. The package contains also the SAme Like The Others (SALTO) rainfall-runoff model as an example.

Details

Package: hydroPASS
Version: 0.0

Main changes in version 0.0
0.0-1: very first version;
0.0-4: version with both PASS and SALTO functions + data;
0.0-7: revisited version that allows using models from other R packages (exemple with TUW-model);
0.0-10: version with U.S. data;
0.1-1: revision of SALTO model, check on U.S. analysis;
0.1-2: inclusion of vignette (see it with `vignette('vignCamelDataPrep')`);

Author(s)

Alberto Viglione, Ralf Merz, Matteo Pesce

Maintainer: Alberto Viglione <viglione@hydro.tuwien.ac.at>

References

All the manual references are listed here:

Merz, R., L. Tarasova and S. Basso (2020). Parameter's controls of distributed catchment models - How much information is in conventional catchment descriptors?, Water Resources Research, 56, e2019WR026008, <https://doi.org/10.1029/2019WR026008>.

`catchDescrip` *Function for the setup of catchment descriptors*

Description

The catchDescrip function reads gridded catchment descriptors and a shapefile of catchments to create data of descriptors at grid and catchment scale, in order to apply the PASS procedure.

Usage

```
catchDescrip(cds, catchments, weights, lon.mast, lat.mast)
```

Arguments

`cds`	arrays of catchment descriptors
`catchments`	SpatialPolygonsDataFrame of catchments
`weights`	array of pixel/model unit coverage of the catchment area for each catchment
`lon.mast, lat.mast`	numeric vectors of input master grid coordinates

Value

`train.CD_4cat104pxl`
list of two matrices of catchment/model unit descriptors:
`cat` – matrix with N rows corresponding to each catchment whose columns are catchment descriptors
`grd` – matrix with M rows corresponding to each pixel/model unit whose columns are pixel/model unit descriptors

See Also

`outDataPrep`

Examples

```
data(outDataPrep)

ls()

names(inputCatchDescr)
CDS <- inputCatchDescr$cds
CATCHMENTS <- inputCatchDescr$catchments
WEIGHTS <- inputCatchDescr$weights
LON.M <- inputCatchDescr$lon.m
LAT.M <- inputCatchDescr$lat.m

train.CD_4cat104pxl <- catchDescrip(cds=CDS, catchment=CATCHMENTS, weights=WEIGHTS,
                                    lon.mast=LON.M, lat.mast=LAT.M)
ls()
str(train.CD_4cat104pxl)
```

`outDataPrep` *Data-sample*

Description

Example output data from the vignette 'vignCamelDataPrep'

Usage

```
data(outDataPrep)
```

Format

The data are following:

- `train.data`: list of input timeseries for each of 215 pixels in the study area and of observed runoff discharges for each of 25 catchments. The elements of the list are: `prec` – matrix of precipitation input [mm/day] whose 215 columns contain the precipitation for each pixel of the region; `temp` – matrix of temperature input [degC] whose 215 columns contain the temperature for each pixel of the region; `pet` – matrix of potential evapotranspiration input [mm/day] whose 215 columns contain the potential evapotranspiration for each pixel of the region; `qobs` – matrix of observed stream runoff [mm/day] whose 215 columns contain the stream runoff at each catchment outlet in the region.
- `inputTopology`: list of objects which are used as input for the function calculating the topology of catchments. The elements of the list are: `catchments` – SpatialPolygonsDataFrame of catchments; `dem` – RasterLayer of digital elevation model at 30 arc-second resolution; `lon.m,lat.m` – numeric vectors of input master grid coordinates; `lon.dem,lat.dem` – numeric vectors of digital elevation model coordinates.
- `inputCatchDescr`: list of objects which are used as input for the function calculating catchment descriptors. The elements of the list are: `cds` – list of numeric arrays of catchment descriptors of dimensions lat, lon; `catchments` – SpatialPolygonsDataFrame of catchments; `weights` – array of pixel/model unit coverage of the catchment area for each catchment; `lat.m,lon.m` – numeric vectors of input master grid coordinates.

References

Gridded rainfall and temperature data are available from NCAR:

Jasinski, M.F., S.V. Kumar, J.S. Borak, D.M. Mocko, C.D. Peters-Lidard, M. Rodell, H. Rui, H. Kato Beaudoing, B.E. Vollmer, K.R. Arsenault, B. Li, and J.D. Bolten, 2018: NCA-LDAS Noah-3.3 Land Surface Model L4 Daily 0.125 x 0.125 degree V2.0, Greenbelt, Maryland, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC), https://doi.org/10.5067/7V3N5DO04MAS

Daily runoff observations are available from CAMELS data set:

Addor, N., A. J. Newman, N. Mizukami, and M. P. Clark, 2017: The CAMELS data set: Catchment attributes and meteorology for large-sample studies. Hydrol. Earth Syst. Sci. , 21, 5293–5313, https://doi.org/10.5194/hess-21-5293-2017

Catchment descriptors are calculated from sources listed in Appendix D of:

Merz, R., A. Miniussi, S. Basso, K. J. Petersen, and L. Tarasova, 2022: More Complex is Not Necessarily Better in Large-Scale Hydrological Modeling: A Model Complexity Experiment across the Contiguous United States. Bulletin of the American Meteorological Society, 103(8), E1947-E1967, https://doi.org/10.1175/BAMS-D-21-0284.1

Examples

```
data(outDataPrep)

ls()

str(train.data)
names(inputTopology)
```

```
str(inputTopology)
names(inputCatchDescr)
str(inputCatchDescr)
```

`outTrainParameters` *Data-sample*

Description

Example input SALTO parameters for PASS from U.S. datasets (see vignette 'vignCamelDataPrep')

Usage

```
data(outTrainParameters)
```

Format

The data are following:

- `train.parameters_25cat`: list of 25 matrices with locally lumped model efficiency (1st column) and calibrated model parameters used for the training of the machine-learning regionalisation (one column per parameter, one row per set).

See Also

SALTO, PASS, outDataPrep

Examples

```
data(outTrainParameters)

ls()

str(train.parameters_25cat)


data(outDataPrep)

str(train.data)
names(inputTopology)
str(inputTopology)
names(inputCatchDescr)
str(inputCatchDescr)
```

PASS *PArameter Set Shuffling algorithm*

Description

The PArameter Set Shuffling (PASS) algorithm derives relationship between model parameters and catchment descriptors from observed patterns of calibrated parameters and available catchment descriptors through machine-learning, and provides regionally consistent parameter sets.

Usage

```
PASS (Y, X.cat, X.grd, grd2cat,
      model.eff.fn, lower, upper,
      options=PASS.options(),
      ...)

PASS.options (maxLoops=100, nGroups=10, REGloops=5,
              generalized.mean.power=-1, proportion.max.eff.update=0.95,
              sampling='random', optim.subset.cat=0.7)

## S3 method for class 'PASS'
 print(x, ...)
```

Arguments

Y	list (N.cat) of data.frames (XXX x N.par) with locally lumped calibrated model parameters OR output of a previous run of PASS on the same data, which serves as a starting point for further search of regionally consistent model parameters
X.cat	matrix or data.frame (N.cat x N.dsc) of catchment descriptors
X.grd	matrix or data.frame (N.grd x N.dsc) of model unit/pixel descriptors
grd2cat	list (N.cat) of model unit/pixel names belonging to each catchment
model.eff.fn	user defined function to be optimized (maximized). The function should have as its first argument the vector/matrix of real-valued parameters to optimize, as second argument the catchment index, and return a scalar real result with maximum equal to 1 ('NA' and 'NaN' values are not allowed). The second argument of the function should be the catchment number (the model will be applied to one catchment only). The other arguments of the function should use has input all the data available for the region (all catchments, all model units/pixels). Examples are provided below
lower, upper	two vectors specifying scalar real lower and upper bounds on each parameter to be optimized, so that the i-th element of 'lower' and 'upper' applies to the i-th parameter
options	list of options for the PASS approach, see PASS.options
...	other arguments (for PASS further arguments to be passed to 'model.eff.fn')
maxLoops	number of loops for the serach

`nGroups` number of groups over which optimisation is run (used to avoid being stuck in local optima)

`REGloops` number of loops for the regional consistency procedure (i.e., the parameter selected for machine-learning regionalisation are the closest ones to the identified parameters in a previous machine-learning regionalisation within REGloops)

`generalized.mean.power`
value of the power exponent of the generalized mean of the efficiencies obtained in all catchments with the fully distributed regional model. The generalized mean is used to decide whether a regionalization outperforms another. If the power p=1 => arithmetic mean, p->0 => geometric mean, p=-1 => harmonic mean, p=-Inf => minimum, p=+Inf => maximum

`proportion.max.eff.update`
if regionalized lumped parameters are found so that the model efficiency exceeds the 'proportion.max.eff.update' multiplied by the maximum value in the training parameter set, then the regionalized lumped parameters are added in the training set and used in fllowing loops

`sampling` if `random` the selection of model parameters of each catchment for the machine-learning regionalisation is random, if `optim` and `previousPASSout` exists the model parameters of each catchment for the machine-learning regionalisation are selected from the ones obtained in `previousPASSout` except for the proportion indicated in `optim.subset.cat`, which is selected randomly

`optim.subset.cat`
proportion of catchments from which the parameters for the machine-learning regionalisation are selected randomly (the others being kept from `previousPASSout`)

`x` output of `PASS`

Details

Here explanation of what the function does...

Value

`groups` list of outputs of the PASS approach for each of the `nGroups` containing:
`overall.eff` – overall regional efficiency of the model on all catchments with regionally consistent parameters identified by PASS (at the moment calculated as the median efficiency for the N catchments)
`selected.parameters` – matrix of the model efficiency (1st column) and selected parameters for training the machine-learning regionalisation that give the best overall regional efficiency
`regionalized.parameters` – list of two matrices, one with N rows for the lumped models (`cat.par.pred`), and the other with M rows for each pixel/model unit (`grd.par.pred`)
`cat.eff.lump` – vector of the N efficiencies of the lumped models with parameters in `regionalized.parameters$cat.par.pred`
`cat.eff.dist` – vector of the N efficiencies of the distributed models with parameters in `regionalized.parameters$grd.par.pred`

`train.parameters.updated`
: list of N matrices with locally lumped model efficiency (1st column) and calibrated model parameters used for the training of the machine-learning regionalisation (one column per parameter, one row per set) which may have been updated by PASS if the regionalisation procedure has identified parameter sets whose efficiency is comparable to the local calibration

`PASS.options`
: list of options used in the PASS approach (see above)

See Also

SALTO

Examples

```
## Not run:
# ------------------------------------------------------------------------------------------ #
#             1) PASS with model SALTO and KGE efficiency                                    #
# ------------------------------------------------------------------------------------------ #

# data:
data(outDataPrep)
data(outTrainParameters)

CDS <- inputCatchDescr$cds
CATCHMENTS <- inputCatchDescr$catchments
WEIGHTS <- inputCatchDescr$weights
LON.M <- inputCatchDescr$lon.m
LAT.M <- inputCatchDescr$lat.m
train.CD_25cat215pxl <- catchDescrip(cds=CDS, catchment=CATCHMENTS, weights=WEIGHTS,
                                     lon.mast=LON.M, lat.mast=LAT.M)

CATCHMENTS <- inputTopology$catchments
DEM <- inputTopology$dem
LON.M <- inputTopology$lon.m
LAT.M <- inputTopology$lat.m
LON.DEM <- inputTopology$lon.dem
LAT.DEM <- inputTopology$lat.dem
# the following takes some time
train.topology_25cat215pxl <- topology(catchments=CATCHMENTS, dem=DEM, lon.mast=LON.M,
                                       lat.mast=LAT.M, lon.dem=LON.DEM, lat.dem=LAT.DEM)

# Define an efficiency function:
# The function should have as its first argument the vector/matrix of real-valued parameters
#  to optimize, as second argument the catchment index, and return a scalar real result with
#  maximum equal to 1 ('NA' and 'NaN' values are not allowed). The second argument of the
#  function should be the catchment number (the model will be applied to one catchment only).
#  The other arguments of the function should use has input all the data available for the
#  region (all catchments, all model units/pixels).
KGE.SALTO <- function(param, cat.number, prec, temp, pet,
                      effarea, grdname, flowto, level, disc, iwarmup=100) {
  GRDNAME <- grdname[[cat.number]]
  EFFAREA <- effarea[[cat.number]]
```

```
  FLOWTO <- flowto[[cat.number]]
  LEVEL <- level[[cat.number]]
  PREC <- prec[, as.character(GRDNAME)]
  TEMP <- temp[, as.character(GRDNAME)]
  PET <- pet[, as.character(GRDNAME)]
  DISC <- disc[, cat.number]
  # global variables are in the function
  simu <- SALTO(prec=PREC, temp=TEMP, pet=PET,
                effarea=EFFAREA, grdname=GRDNAME, flowto=FLOWTO, level=LEVEL,
                param=param)$q
  if (all(!is.na(simu))) {
    simu[is.na(simu)] <- -999
    simu <- simu[-c(1:iwarmup)]  # remove the warming period
    obse <- DISC[-c(1:iwarmup)]  # remove the warming period
    r <- cor(simu, obse, method='pearson', use='pairwise.complete.obs')
    beta <- mean(simu)/mean(obse)
    gamma <- (sd(simu)/mean(simu))/(sd(obse)/mean(obse))
    kgeQ <- 1 - sqrt((r - 1)^2 + (beta - 1)^2 + (gamma - 1)^2)
  } else kgeQ <- -999
  return(kgeQ)  # Kling-Gupta Efficiency
}


# this takes long
run01 <- PASS(Y=train.parameters_25cat,
              X.cat=train.CD_25cat215pxl$cat,
              X.grd=train.CD_25cat215pxl$grd,
              grd2cat=sapply(train.topology_25cat215pxl, function(x) x$grd),
              model.eff.fn=KGE.SALTO,
              lower=c(-3,0,1,0.1,0.1,-99,-3,  0,  10,0.1,0.1,
                      0.1,0.1, 0,0.1, 30,0.1,  100,0.1,0.1,0.1),
              upper=c( 3,0,1, 10, 10,-99, 3,100,3000, 20,  2,
                       50,  5,10,  5,400,  5,15000,  5,  5, 30),
         options=PASS.options(maxLoops=2000, nGroups=10, REGloops=5, sampling='random'),
              prec=train.data$prec,
              temp=train.data$temp,
              pet=train.data$pet,
              effarea=sapply(train.topology_25cat215pxl, function(x) x$effarea),
              grdname=sapply(train.topology_25cat215pxl, function(x) x$grd),
              flowto=sapply(train.topology_25cat215pxl, function(x) x$flowto),
              level=sapply(train.topology_25cat215pxl, function(x) x$level),
              disc=train.data$qobs)
#
#     Loop 200 out of 2000 loops
#     Loop 400 out of 2000 loops
#     Loop 600 out of 2000 loops
#     Loop 800 out of 2000 loops
#     Loop 1000 out of 2000 loops
#     Loop 1200 out of 2000 loops
#     Loop 1400 out of 2000 loops
#     Loop 1600 out of 2000 loops
#     Loop 1800 out of 2000 loops
#     Loop 2000 out of 2000 loops
```

```
#
# Time difference of 31.28067 mins
 run01
# Output of the PArameter Set Shuffling algorithm:
#
#  number of catchments: 4
#  number of model units (e.g. pixels): 104
# sampling strategy: random
#  number of loops: 2000
#  number of groups: 10
#  number of loops for regional consistency: 5
# distributed regionalization efficiencies:
#         Group1 Group2 Group3 Group4 Group5 Group6 Group7 Group8 Group9 Group10
# Min.     0.413  0.395  0.373  0.393  0.398  0.396  0.394  0.373  0.363   0.348
# 1st Qu.  0.427  0.409  0.388  0.429  0.426  0.422  0.418  0.388  0.388   0.376
# Median   0.479  0.471  0.436  0.462  0.438  0.463  0.475  0.436  0.447   0.419
# Mean     0.480  0.475  0.478  0.493  0.487  0.492  0.473  0.478  0.474   0.475
# 3rd Qu.  0.532  0.536  0.526  0.525  0.499  0.534  0.531  0.526  0.532   0.518
# Max.     0.548  0.564  0.666  0.654  0.674  0.646  0.548  0.667  0.639   0.715
#  updated number of train parameter sets: 1029

run02 <- PASS(Y=run01,
              X.cat=train.CD_25cat215pxl$cat,
              X.grd=train.CD_25cat215pxl$grd,
             grd2cat=sapply(train.topology_25cat215pxl_125cat215pxl, function(x) x$grd),
              model.eff.fn=KGE.SALTO,
              lower=c(-3,0,1,0.1,0.1,-99,-3,  0,  10,0.1,0.1,
                      0.1,0.1, 0,0.1, 30,0.1,  100,0.1,0.1,0.1),
              upper=c( 3,0,1, 10, 10,-99, 3,100,3000, 20,  2,
                       50,  5,10,  5,400,  5,15000,  5,  5, 30),
              options=PASS.options(maxLoops=2000, nGroups=10, REGloops=4,
                                   sampling='optim', optim.subset.cat=0.7),
              prec=train.data$prec,
              temp=train.data$temp,
              pet=train.data$pet,
              effarea=sapply(train.topology_25cat215pxl, function(x) x$effarea),
              grdname=sapply(train.topology_25cat215pxl, function(x) x$grd),
              flowto=sapply(train.topology_25cat215pxl, function(x) x$flowto),
              level=sapply(train.topology_25cat215pxl, function(x) x$level),
              disc=train.data$qobs)
#
#     Loop 200 out of 2000 loops
#     Loop 400 out of 2000 loops
#     Loop 600 out of 2000 loops
#     Loop 800 out of 2000 loops
#     Loop 1000 out of 2000 loops
#     Loop 1200 out of 2000 loops
#     Loop 1400 out of 2000 loops
#     Loop 1600 out of 2000 loops
#     Loop 1800 out of 2000 loops
#     Loop 2000 out of 2000 loops
#
# Time difference of 26.98754 mins
```

```
 run02
# Output of the PArameter Set Shuffling algorithm:
#
#  number of catchments: 4
#  number of model units (e.g. pixels): 104
# sampling strategy: optim
#  number of loops: 2000
#  number of groups: 10
#  number of loops for regional consistency: 4
#  proportion of randomized parameters (when optim): 0.7
# distributed regionalization efficiencies:
#        Group1 Group2 Group3 Group4 Group5 Group6 Group7 Group8 Group9 Group10
# Min.     0.413  0.395  0.389  0.393  0.398  0.396  0.389  0.373  0.363   0.348
# 1st Qu.  0.427  0.409  0.398  0.429  0.426  0.422  0.398  0.388  0.388   0.376
# Median   0.479  0.471  0.459  0.462  0.438  0.463  0.460  0.436  0.447   0.419
# Mean     0.480  0.475  0.495  0.493  0.487  0.492  0.493  0.478  0.474   0.475
# 3rd Qu.  0.532  0.536  0.556  0.525  0.499  0.534  0.555  0.526  0.532   0.518
# Max.     0.548  0.564  0.672  0.654  0.674  0.646  0.665  0.667  0.639   0.715
#  updated number of train parameter sets: 3831

run03 <- PASS(Y=run02,
              X.cat=train.CD_25cat215pxl$cat,
              X.grd=train.CD_25cat215pxl$grd,
              grd2cat=sapply(train.topology_25cat215pxl, function(x) x$grd),
              model.eff.fn=KGE.SALTO,
              lower=c(-3,0,1,0.1,0.1,-99,-3,  0,  10,0.1,0.1,
                      0.1,0.1, 0,0.1, 30,0.1,  100,0.1,0.1,0.1),
              upper=c( 3,0,1, 10, 10,-99, 3,100,3000, 20,  2,
                       50,  5,10,  5,400,  5,15000,  5,  5, 30),
              options=PASS.options(maxLoops=2000, nGroups=10, REGloops=3,
                                   sampling='optim', optim.subset.cat=0.5),
              prec=train.data$prec,
              temp=train.data$temp,
              pet=train.data$pet,
              effarea=sapply(train.topology_25cat215pxl, function(x) x$effarea),
              grdname=sapply(train.topology_25cat215pxl, function(x) x$grd),
              flowto=sapply(train.topology_25cat215pxl, function(x) x$flowto),
              level=sapply(train.topology_25cat215pxl, function(x) x$level),
              disc=train.data$qobs)
#
#      Loop 200 out of 2000 loops
#      Loop 400 out of 2000 loops
#      Loop 600 out of 2000 loops
#      Loop 800 out of 2000 loops
#      Loop 1000 out of 2000 loops
#      Loop 1200 out of 2000 loops
#      Loop 1400 out of 2000 loops
#      Loop 1600 out of 2000 loops
#      Loop 1800 out of 2000 loops
#      Loop 2000 out of 2000 loops
#
# Time difference of 23.23943 mins
 run03
```

```
# Output of the PArameter Set Shuffling algorithm:
#
#  number of catchments: 4
#  number of model units (e.g. pixels): 104
# sampling strategy: optim
#  number of loops: 2000
#  number of groups: 10
#  number of loops for regional consistency: 3
#  proportion of randomized parameters (when optim): 0.5
# distributed regionalization efficiencies:
#         Group1 Group2 Group3 Group4 Group5 Group6 Group7 Group8 Group9 Group10
# Min.     0.413  0.395  0.389  0.393  0.398  0.396  0.389  0.406  0.363   0.400
# 1st Qu.  0.427  0.409  0.398  0.429  0.426  0.422  0.398  0.441  0.388   0.439
# Median   0.479  0.471  0.459  0.462  0.438  0.463  0.460  0.488  0.447   0.483
# Mean     0.480  0.475  0.495  0.493  0.487  0.492  0.493  0.488  0.474   0.483
# 3rd Qu.  0.532  0.536  0.556  0.525  0.499  0.534  0.555  0.535  0.532   0.528
# Max.     0.548  0.564  0.672  0.654  0.674  0.646  0.665  0.570  0.639   0.566
#  updated number of train parameter sets: 4945


run04 <- PASS(Y=run03,
              X.cat=train.CD_25cat215pxl$cat,
              X.grd=train.CD_25cat215pxl$grd,
              grd2cat=sapply(train.topology_25cat215pxl, function(x) x$grd),
              model.eff.fn=KGE.SALTO,
              lower=c(-3,0,1,0.1,0.1,-99,-3,  0,  10,0.1,0.1,
                      0.1,0.1, 0,0.1, 30,0.1,  100,0.1,0.1,0.1),
              upper=c( 3,0,1, 10, 10,-99, 3,100,3000, 20,  2,
                       50,  5,10,  5,400,  5,15000,  5,  5, 30),
              options=PASS.options(maxLoops=2000, nGroups=10, REGloops=2,
                                   sampling='optim', optim.subset.cat=0.3),
              prec=train.data$prec,
              temp=train.data$temp,
              pet=train.data$pet,
              effarea=sapply(train.topology_25cat215pxl, function(x) x$effarea),
              grdname=sapply(train.topology_25cat215pxl, function(x) x$grd),
              flowto=sapply(train.topology_25cat215pxl, function(x) x$flowto),
              level=sapply(train.topology_25cat215pxl, function(x) x$level),
              disc=train.data$qobs)
#
#      Loop 200 out of 2000 loops
#      Loop 400 out of 2000 loops
#      Loop 600 out of 2000 loops
#      Loop 800 out of 2000 loops
#      Loop 1000 out of 2000 loops
#      Loop 1200 out of 2000 loops
#      Loop 1400 out of 2000 loops
#      Loop 1600 out of 2000 loops
#      Loop 1800 out of 2000 loops
#      Loop 2000 out of 2000 loops
#
# Time difference of 19.6393 mins
 run04
```

```
# Output of the PArameter Set Shuffling algorithm:
#
#  number of catchments: 4
#  number of model units (e.g. pixels): 104
# sampling strategy: optim
#  number of loops: 2000
#  number of groups: 10
#  number of loops for regional consistency: 2
#  proportion of randomized parameters (when optim): 0.3
# distributed regionalization efficiencies:
#        Group1 Group2 Group3 Group4 Group5 Group6 Group7 Group8 Group9 Group10
# Min.     0.413  0.395  0.389  0.393  0.398  0.396  0.389  0.406  0.363   0.400
# 1st Qu.  0.427  0.409  0.398  0.429  0.426  0.422  0.398  0.441  0.388   0.439
# Median   0.479  0.471  0.459  0.462  0.438  0.463  0.460  0.488  0.447   0.483
# Mean     0.480  0.475  0.495  0.493  0.487  0.492  0.493  0.488  0.474   0.483
# 3rd Qu.  0.532  0.536  0.556  0.525  0.499  0.534  0.555  0.535  0.532   0.528
# Max.     0.548  0.564  0.672  0.654  0.674  0.646  0.665  0.570  0.639   0.566
#  updated number of train parameter sets: 5202


## End(Not run)
```

SALTO *SAme Like The Others rainfall-runoff model*

Description

The SAme Like The Others (SALTO) model is a distributed conceptual rainfall-runoff model. The model is a representative of a soil moisture accounting scheme and similar to many other well-known models in hydrology, such as the HBV model or elements of the SUPERFLEX modeling approach.

Usage

```
SALTO (prec, temp, pet, effarea=1, grdname=1,
       flowto=0, level=1, sm_layer=1,
       param=c(TS=0, TS_R=0, BETA_SNOW=1,
               DDF_NR=2, DDF_R=2, DDF_INC=-99,
               TM=0, SM_MIN_1=100, SM_MAX_1=1000,
               BETA_RC_1=1, BETA_AET_1=0.5,
               K_RS_1=10, BETA_RS_1=1,
               PERCMAX_1=5, BETA_PERC_1=1,
               K_GW=100, BETA_GW=1, GW_S_MAX=5000, BETA_DQ_GW=1,
               K_RIVER=1, BETA_RIVER=1), itimesteps=NULL)
## S3 method for class 'SALTO'
 print(x, ...)
## S3 method for class 'SALTO'
 plot(x, time, which=1, ...)
```

Arguments

prec	matrix of precipitation input [mm/timestep] whose columns contain the precipitation for each pixel/part of the catchment
temp	matrix of temperature input [degC] whose columns contain the temperature for each pixel/part of the catchment
pet	matrix of potential evapotranspiration input [mm/timestep] whose columns contain the potential evapotranspiration for each pixel/part of the catchment
effarea	vector of what proportion of catchment area is contained in each pixel/part of the catchment, the sum should be equal to 1
grdname	pixel/model unit number covering the catchment
flowto	downstream pixel/model unit (use -99 for the outlet pixel)
level	pixel/model unit level (1=highest headwater, with no upstream neighbour, 2=only level 1 upstream neighbours, etc.)
sm_layer	number of soil moisture layers, which affects the number of parameters of the model
param	parameters of the SALTO model (one row per pixel/part of the catchment, or a single vector for the same parameters for each pixel/part of the catchment) which are:

TS	threshold temperature above which precipitation is rain [degC]
TS_R	???
BETA_SNOW	the non linear parameter for snowmelt [-]
DDF_NR	degree day factor [mm/degC/timestep] for ???
DDF_R	degree day factor [mm/degC/timestep] for ???
DDF_INC	degree day factor [mm/degC/timestep] for ???
TM	threshold temperature above which melting starts [degC]
SM_MIN_1	min. soil moisture storage [mm] for the first soil moisture layer
SM_MAX_1	max. soil moisture storage [mm] for the first soil moisture layer
BETA_RC_1	the non linear parameter for runoff production [-] for the first soil moisture layer
BETA_AET_1	the non linear parameter for evapotranspiration [-] for the first soil moisture layer
K_RS_1	storage coefficient for fast response [timestep] for the first soil moisture layer
BETA_RS_1	power coefficient for fast response [-] for the first soil moisture layer
PERCMAX_1	max percolation rate [mm/timestep] for the first soil moisture layer
BETA_PERC_1	the non linear parameter for percolation [-] for the first soil moisture layer
SM_MIN_2	min. soil moisture storage [mm] for the second soil moisture layer
SM_MAX_2	max. soil moisture storage [mm] for the second soil moisture layer
BETA_RC_2	the non linear parameter for runoff production [-] for the second soil moisture layer
BETA_AET_2	the non linear parameter for evapotranspiration [-] for the second soil moisture layer
K_RS_2	storage coefficient for slow response [timestep] for the second soil moisture layer
BETA_RS_2	power coefficient for slow response [-] for the second soil moisture layer
PERCMAX_2	max percolation rate [mm/timestep] for the second soil moisture layer
BETA_PERC_2	the non linear parameter for percolation [-] for the second soil moisture layer
SM_MIN_3	min. soil moisture storage [mm] for the third soil moisture layer
SM_MAX_3	max. soil moisture storage [mm] for the third soil moisture layer
BETA_RC_3	the non linear parameter for runoff production [-] for the third soil moisture layer
BETA_AET_3	the non linear parameter for evapotranspiration [-] for the third soil moisture layer

K_RS_3	storage coefficient for slow response [timestep] for the third soil moisture layer
BETA_RS_3	power coefficient for slow response [-] for the third soil moisture layer
PERCMAX_3	max percolation rate [mm/timestep] for the third soil moisture layer
BETA_PERC_3	the non linear parameter for percolation [-] for the third soil moisture layer
K_GW	storage coefficient for groundwater response [timestep]
BETA_GW	the non linear parameter for groundwater response [-]
GW_S_MAX	max. groundwater storage [mm]
BETA_DQ_GW	power coefficient for groundwater diverted to river [-]
K_RIVER	storage coefficient for river routing [timestep]
BETA_RIVER	power coefficient for river routing [-]

itimesteps	number of timesteps ???
x	output of SALTO
time	time variable, for plots
which	which plot: 1 = full timeseries; 2 = annual timeseries
...	other arguments

Details

Here explanation of what the function does...

Value

itimesteps	number of timesteps
igrdicat	number of pixels/parts of the catchment
upsneighb	square matrix of 0s and 1s with dimention equal to the number of pixel/part of the catchment, upsneighb[34,58] = 1 if pixel 58 is upstream of pixel 34, otherwhise 0
effarea	vector of what proportion of catchment area is contained in each pixel/part of the catchment, the sum should be equal to 1
prec	matrix of precipitation input [mm/timestep] whose rows contain the precipitation for each pixel/part of the catchment
temp	matrix of temperature input [degC] whose rows contain the temperature for each pixel/part of the catchment
pet	matrix of potential evapotranspiration input [mm/timestep] whose rows contain the potential evapotranspiration for each pixel/part of the catchment
sm_layer	number of soil moisture layers, which affects the number of parameters of the model
TS	threshold temperature above which precipitation is rain [degC]
TS_R	???
BETA_SNOW	the non linear parameter for snowmelt ???
TM	threshold temperature above which melting starts [degC]
DDF_NR	degree day factor [mm/degC/timestep] for ???

DDF_R	degree day factor [mm/degC/timestep] for ???
DDF_INC	degree day factor [mm/degC/timestep] for ???
SM_MIN	min. soil moisture storage [mm]
SM_MAX	max. soil moisture storage [mm]
BETA_RC	the non linear parameter for runoff production [-]
BETA_AET	the non linear parameter for evapotranspiration [-]
K_RS	storage coefficient for fast/slow response [timestep]
BETA_RS	power coefficient for fast/slow response [-]
PERCMAX	max percolation rate [mm/timestep]
BETA_PERC	the non linear parameter for percolation [-]
K_GW	storage coefficient for groundwater response [timestep]
BETA_GW	the non linear parameter for groundwater response [-]
GW_S_MAX	max. groundwater storage [mm]
BETA_DQ_GW	power coefficient for groundwater diverted to river [-]
K_RIVER	storage coefficient for river routing [timestep]
BETA_RIVER	the non linear parameter for river routing [-]
SWE_START	initial snow water equivalent in each pixel/part of the catchment [mm]
SM_START	initial soil moisture in each pixel/part of the catchment [mm]
RS_S_START	initial lateral runoff storage in each pixel/part of the catchment [mm]
GW_S_START	initial groundwater storage in each pixel/part of the catchment [mm]
RIVER_S_START	initial storage in the river in each pixel/part of the catchment [mm]
swe	output timeseries of snow water equivalent in each pixel/part of the catchment [mm]
sm	output timeseries of soil moisture in each pixel/part of the catchment [mm]
aet	output timeseries of actual evapotranspiration in each pixel/part of the catchment [mm/timestep]
sm_perc	output timeseries of soil moisture percolation in the next soil layer in each pixel/part of the catchment [mm/timestep]
rs_s	output timeseries of fast runoff storage in each pixel/part of the catchment [mm]
rs_q	output timeseries of fast runoff in each pixel/part of the catchment [mm/timestep]
gw_s	output timeseries of groundwater storage in each pixel/part of the catchment [mm]
gw_q	output timeseries of groundwater discharge in each pixel/part of the catchment [mm/timestep]
river_s	output timeseries of river storage in each pixel/part of the catchment [mm]
gw_out	output timeseries of groundwater outflow in each pixel/part of the catchment [mm/timestep]
qsim	output timeseries of produced runoff in each pixel/part of the catchment [mm/timestep]
q	output timeseries of produced runoff at the catchment outlet [mm/timestep]

See Also

PASS

Examples

```
## Not run:
data(outDataPrep)
data(outTrainParameters)

CATCHMENTS <- inputTopology$catchments
DEM <- inputTopology$dem
LON.M <- inputTopology$lon.m
LAT.M <- inputTopology$lat.m
LON.DEM <- inputTopology$lon.dem
LAT.DEM <- inputTopology$lat.dem
# the following takes some time
train.topology_25cat215pxl <- topology(catchments=CATCHMENTS, dem=DEM, lon.mast=LON.M,
                                       lat.mast=LAT.M, lon.dem=LON.DEM, lat.dem=LAT.DEM)

# select data for one catchment
whichcatchment = 2
whichpixels <- which(match(colnames(train.data_4cat104pxl$prec),
                           as.character(train.topology_4cat104pxl[[whichcatchment]]$grd),
                            nomatch=0) > 0)
PREC <- train.data$prec[, whichpixels]
TEMP <- train.data$temp[, whichpixels]
PET <- train.data$pet[, whichpixels]
EFFAREA <- train.topology_25cat215pxl[[whichcatchment]]$effarea
GRDNAME <- train.topology_25cat215pxl[[whichcatchment]]$grd
FLOWTO <- train.topology_25cat215pxl[[whichcatchment]]$flowto
LEVEL <- train.topology_25cat215pxl[[whichcatchment]]$level
PARAM <- train.parameters_4cat[[whichcatchment]][3, -1]
QOBS <- train.data$qobs[, whichcatchment]
TIME <- as.Date(rownames(train.data$qobs))

# run the model
output <- SALTO(prec=PREC, temp=TEMP, pet=PET,
                effarea=EFFAREA, grdname=GRDNAME, flowto=FLOWTO, level=LEVEL,
                param=PARAM)
output
# Output of the model SALTO:
#
# simulations of 3652 days
# with inputs on 12 pixels/model units
#
# mean annual temperature of 7.2 degC
# mean annual potential evapotranspiration of 728 mm/a
# mean annual precipitation of 1003 mm/a
# mean annual actual evapotranspiration of 465 mm/a
# mean annual runoff of 868 mm/a
# quantiles of the simulated daily stream runoff (mm/d):
#    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
```

```
#   0.000   0.758   1.408   2.377   2.768  46.703
str(output)

# plot daily observed vs. simulated streamflow runoff (mm/d)
plot(output, TIME, which=1)
 lines(TIME, QOBS)

# plot annual values (mm/a)
QOBSa <- tapply(QOBS, substr(TIME, 1, 4), mean, na.rm=TRUE)*365
plot(output, TIME, which=2)
 lines(as.numeric(names(QOBSa)), QOBSa, col=1)


# Calibration example
require(DEoptim)

negME.SALTO <- function(param, prec, temp, pet, effarea, grdname, flowto, level, disc) {
  # global variables are in the function
  simu <- SALTO(prec=prec, temp=temp, pet=pet,
            effarea=effarea, grdname=grdname, flowto=flowto, level=level, param=param)$q
  if (!all(is.na(simu))) {
    simu[is.na(simu)] <- -999
    simu <- simu[-c(1:303)]  # remove the warming period
    obse <- disc[-c(1:303)]  # remove the warming period
    r <- cor(simu, obse, method='pearson', use='pairwise.complete.obs')
    beta <- mean(simu)/mean(obse)
    gamma <- (sd(simu)/mean(simu))/(sd(obse)/mean(obse))
    kgeQ <- 1 - sqrt((r - 1)^2 + (beta - 1)^2 + (gamma - 1)^2)
  } else kgeQ <- -999
  return(-kgeQ)  # negative Kling-Gupta Efficiency
}
calibr_out <- DEoptim(fn=negME.SALTO,
                      lower=c(TS=-3.0, TS_R=0, BETA_SNOW=1,
                              DDF_NR=0.1, DDF_R=0.1, DDF_INC=-99,
                              TM=-3.0, SM_MIN_1=0, SM_MAX_1=10,
                              BETA_RC_1=0.1, BETA_AET_1=0.1,
                              K_RS_1=0.1, BETA_RS_1=0.1,
                              PERCMAX_1=0.0, BETA_PERC_1=0.1,
                              K_GW=30, BETA_GW=0.1, GW_S_MAX=100, BETA_DQ_GW=0.1,
                              K_RIVER=0.1, BETA_RIVER=0.1),
                       upper=c(TS=3.0, TS_R=0, BETA_SNOW=1,
                               DDF_NR=10, DDF_R=10, DDF_INC=-99,
                               TM=3.0, SM_MIN_1=100, SM_MAX_1=3000,
                               BETA_RC_1=20, BETA_AET_1=2.0,
                               K_RS_1=50, BETA_RS_1=5.0,
                               PERCMAX_1=10, BETA_PERC_1=5.0,
                               K_GW=400, BETA_GW=5.0, GW_S_MAX=15000, BETA_DQ_GW=5.0,
                               K_RIVER=5.0, BETA_RIVER=30),
               control=DEoptim.control(NP=NA, itermax=200, reltol=1e-3, steptol=10, trace=5),
                      prec=PREC, temp=TEMP, pet=PET,
                      effarea=EFFAREA, grdname=GRDNAME, flowto=FLOWTO, level=LEVEL,
      disc=QOBS)
# Iteration: 5 bestvalit: -0.522257 bestmemit:    1.358058    ...
```

```
# Iteration: 10 bestvalit: -0.522257 bestmemit:    1.358058    ...
# Iteration: 15 bestvalit: -0.627414 bestmemit:    0.387819    ...
# Iteration: 20 bestvalit: -0.631475 bestmemit:   -0.729270    ...
# Iteration: 25 bestvalit: -0.646940 bestmemit:   -0.729270    ...
# Iteration: 30 bestvalit: -0.650348 bestmemit:    0.273490    ...
# Iteration: 35 bestvalit: -0.659448 bestmemit:   -2.720806    ...
# Iteration: 40 bestvalit: -0.669184 bestmemit:   -0.663057    ...
# Iteration: 45 bestvalit: -0.691207 bestmemit:    1.554097    ...
# Iteration: 50 bestvalit: -0.692687 bestmemit:    1.554097    ...
# Iteration: 55 bestvalit: -0.695200 bestmemit:    1.554097    ...
# Iteration: 60 bestvalit: -0.702177 bestmemit:    0.340355    ...
# Iteration: 65 bestvalit: -0.702213 bestmemit:    0.340355    ...
# Iteration: 70 bestvalit: -0.705200 bestmemit:   -1.303549    ...
# Iteration: 75 bestvalit: -0.705612 bestmemit:   -1.303549    ...
# Iteration: 80 bestvalit: -0.705612 bestmemit:   -1.303549    ...

calibr_sim <- SALTO(prec=PREC, temp=TEMP, pet=PET,
                    effarea=EFFAREA, grdname=GRDNAME, flowto=FLOWTO, level=LEVEL,
                    param=calibr_out$optim$bestmem)
 calibr_sim
# Output of the model SALTO:
#
# simulations of 3652 days
# with inputs on 12 pixels/model units
#
# mean annual temperature of 7.2 degC
# mean annual potential evapotranspiration of 728 mm/a
# mean annual precipitation of 1003 mm/a
# mean annual actual evapotranspiration of 583 mm/a
# mean annual runoff of 835 mm/a
# quantiles of the simulated daily stream runoff (mm/d):
#     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
#  0.07153  0.74396  1.42160  2.28617  2.67042 36.99902


plot(calibr_sim, TIME, which=1)
 lines(TIME, QOBS)

## End(Not run)
```

`topology` *Function for the setup of catchment topology*

Description

The topology function reads a shapefile of catchments, a digital elevation model and master grid and dem coordinates to derive the topology of catchments.

Usage

```
topology(catchments, dem, lon.mast, lat.mast, lon.dem, lat.dem)
```

Arguments

`catchments`	SpatialPolygonsDataFrame of catchments
`dem`	RasterLayer of digital elevation model
`lon.mast, lat.mast`	numeric vectors of input master grid coordinates
`lon.dem, lat.dem`	numeric vectors of DEM coordinates

Value

`train.topology_4cat104pxl`

list of N data.frames (N = number of catchments) each of them having the following columns:
`grd` – pixel/model unit number covering the catchment
`effarea` – pixel/model unit coverage of the catchment area
`flowto` – downstream pixel/model unit (use -99 for the outlet pixel)
`level` – pixel/model unit level (1 = highest headwater, with no upstream neighbour, 2 = only level 1 upstream neighbours, etc.)

See Also

`outDataPrep`

Examples

```
data(outDataPrep)

ls()

names(inputTopology)
CATCHMENTS <- inputTopology$catchments
DEM <- inputTopology$dem
LON.M <- inputTopology$lon.m
LAT.M <- inputTopology$lat.m
LON.DEM <- inputTopology$lon.dem
LAT.DEM <- inputTopology$lat.dem

## Not run:
train.topology_4cat104pxl <- topology(catchments=CATCHMENTS, dem=DEM, lon.mast=LON.M,
                                      lat.mast=LAT.M, lon.dem=LON.DEM, lat.dem=LAT.DEM)
ls()
str(train.topology_4cat104pxl)

## End(Not run)
```

Index

Appendix C

Vignette for U.S. data preparation for the R package 'hydroPASS'

U.S. Data Preparation for hydroPASS

Matteo Pesce, Alberto Viglione

Politecnico di Torino — May 4, 2023

Introduction

With this document, the preparation of US-CAMEL data for the application of the model SALTO and the PASS regional calibration routine is demonstrated. The following data have been downloaded:

- Discharge data and catchment shapefiles: CAMELS data set, https://ral.ucar.edu/solutions/products/camels;
- Rainfall, temperature and potential evapotranspiration: NCA-LDAS Noah-3.3 Land Surface Model L4 Daily 0.125 x 0.125 degree V2.0, https://disc.gsfc.nasa.gov/datasets/NCALDAS_NOAH0125_D_2.0/summary;
- DEM: https://www.hydrosheds.org/;
- Land Use: National Land Cover Database, https://www.usgs.gov/centers/eros/science/national-land-cover-database;
- Soil: NACP MsTMIP - Unified North American Soil Map, https://daac.ornl.gov/cgi-bin/dsviewer.pl?ds_id=1242.

For the implementation of the scripts, the following R packages are needed:

```
require(raster)
require(rgdal)
require(maptools)
require(rgrass)
```

1 Catchments and DEM

Read shapefiles of Hydrologic Unit Code (HUC) = 1:

```
catshape.file = paste('/work/users/matteo/Models/PASS/testModels/UFZHalleWork/',
                      'AmericanData/CAMELS/basin_timeseries_v1p2_metForcing_obsFlow/',
                      'basin_dataset_public_v1p2/shapefiles/Region_01_nhru_simplify_100.shp', sep='')
shape <- readOGR(catshape.file)
```

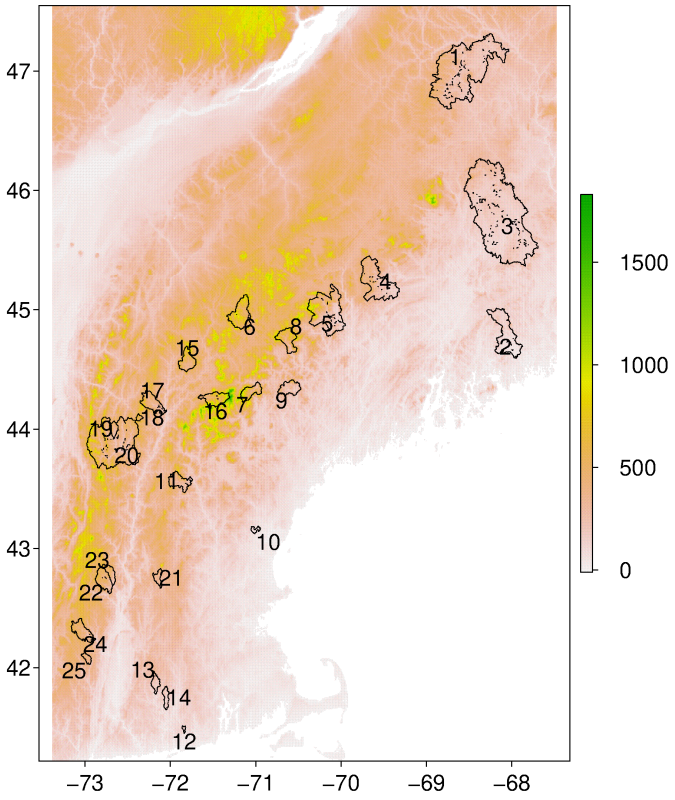
Merge polygons by gageid:

```
shape <- spTransform(shape, CRS("+proj=longlat +ellps=WGS84 +datum=WGS84 +no_defs"))
shape <- shape[!(shape$GAGEID == '01150900'),] # remove this catchment as obs. discharge not present
shape <- shape[!(shape$GAGEID == '01195100'),] # incomplete series (no 2000-2010)
shape <- shape[!(shape$GAGEID == '04296000'),] # incomplete series (no 2000-2010)
GAGE.ID <- shape$GAGEID
shape.union <- unionSpatialPolygons(shape, GAGE.ID)
shape.agg <- data.frame(aggregate(shape[,c(10,52,53)], list(GAGE.ID), FUN = sum))
 rownames(shape.agg) <- shape.agg$Group.1
catchments <- SpatialPolygonsDataFrame(shape.union, shape.agg)
 rownames(catchments@data) <- 1:length(catchments)
 catchments@data
extent_Bacini <- extent(catchments)*1.08
```

Read the DEM file:

```
dem.file <- '/work/users/matteo/Models/PASS/testModels/UFZHalleWork/AmericanData/hydroSHEDS/hyd_na_dem_30s.tif'
dem <- raster(dem.file)
dem <- crop(dem, extent_Bacini)

# add dem coordinates
points <- data.frame(rasterToPoints(dem, spatial=TRUE))
lat.dem <- rev(unique(points$y))
lon.dem <- sort(unique(points$x))
```

```
plot(dem)
 plot(catchments, add=TRUE)
 pointLabel(coordinates(catchments), labels=rownames(catchments@data))
```

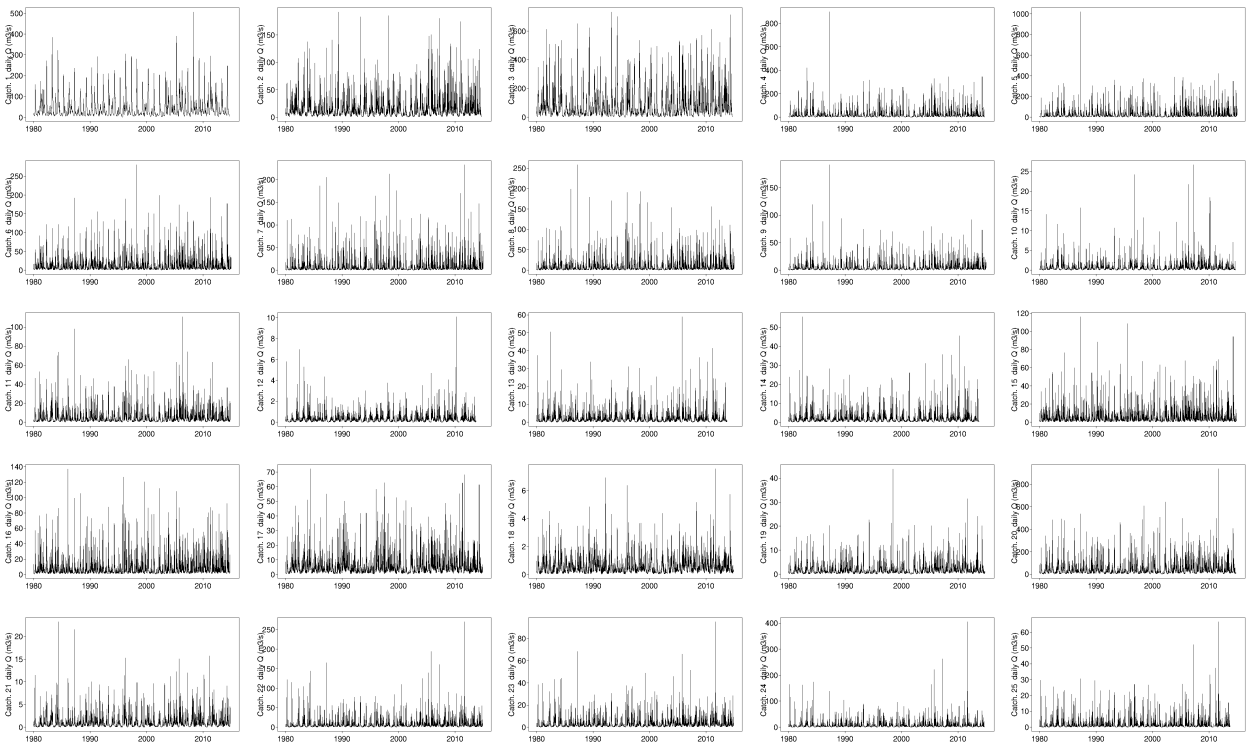


2 Discharge

Read txt files of streamflow of Hydrologic Unit Code (HUC) = 1:

```
path_disch <- paste('/work/users/matteo/Models/PASS/testModels/UFZHalleWork/',
                    'AmericanData/CAMELS/basin_timeseries_v1p2_metForcing_obsFlow/',
                    'basin_dataset_public_v1p2/usgs_streamflow/01', sep='')
disch.files <- list.files(path = path_disch, pattern = "_streamflow_qc.txt$", full.names=TRUE)
gauges.codes <-  gsub('/', "", gsub(path_disch, "",(gsub("_streamflow_qc.txt", "", disch.files))))

obs_Q <- lapply(disch.files, function(x) read.table(x, header=FALSE, sep=''))
obs_Q <- lapply(obs_Q, setNames, c('ID_GAUGE','YEAR','MONTH','DAY','DISCHARGE','FLAG'))
 names(obs_Q) <- gauges.codes
obs_Q <- obs_Q[-which(names(obs_Q) %in% c('01195100','04296000'))]
 #str(obs_Q)
discharge <- matrix(nrow = nrow(obs_Q[[1]]), ncol = length(obs_Q))
 rownames(discharge) <- as.character(seq(as.Date('1980-01-01'), as.Date('2014-12-31'), 'days'))
for (i in 1:ncol(discharge)){
  discharge[,i] <- obs_Q[[i]]$DISCHARGE
}
 colnames(discharge) <- names(obs_Q)
 str(discharge)


##  num [1:12784, 1:25] 655 640 625 620 605 585 570 555 540 535 ...
##  - attr(*, "dimnames")=List of 2
##   ..$ : chr [1:12784] "1980-01-01" "1980-01-02" "1980-01-03" "1980-01-04" ...
##   ..$ : chr [1:25] "01013500" "01022500" "01030500" "01031500" ...


discharge[discharge == -999] <- NA
discharge <- discharge/35.315 # cubic feet --> cubic meter
```

```
plot(as.Date(rownames(discharge)), discharge[,i], type='l', xlab='', ylab=paste('Catch.', i, ' daily Q (m3/s)'))
```



Select the period of interest:

```
discharge <- discharge[rownames(discharge)>='2000-10-01' & rownames(discharge)<='2010-09-30', ]
```

Divide by area (official information on area)

```
gauges.info <- paste('/work/users/matteo/Models/PASS/testModels/UFZHalleWork/',
                     'AmericanData/CAMELS/basin_timeseries_v1p2_metForcing_obsFlow/',
                     'basin_dataset_public_v1p2/basin_metadata/gauge_information.txt', sep='')
GAUGES_INFO <- read.delim(gauges.info, header=T, sep='\t')
GAUGES_INFO <- GAUGES_INFO[,c(1:6)]
 colnames(GAUGES_INFO) <- c('HUC', 'GAGE_ID', 'GAGE_NAME', 'LAT', 'LONG', 'DRAINAGE_AREA')
GAUGES_INFO$GAGE_ID <- paste0("0",GAUGES_INFO$GAGE_ID)
GAUGES_INFO <- GAUGES_INFO[,c(2,4,5,6)]
GAUGES_INFO <- GAUGES_INFO[GAUGES_INFO$GAGE_ID %in% names(obs_Q), ]
 GAUGES_INFO # 25 basins


##     GAGE_ID      LAT      LONG DRAINAGE_AREA
## 1  01013500 47.23739 -68.58264       2252.70
## 2  01022500 44.60797 -67.93524        573.60
## 3  01030500 45.50097 -68.30596       3676.17
## 4  01031500 45.17501 -69.31470        769.05
## 5  01047000 44.86920 -69.95510        909.10
## 6  01052500 44.87739 -71.05749        383.82
## 7  01054200 44.39044 -70.97964        180.98
## 8  01055000 44.64275 -70.58878        250.64
## 9  01057000 44.30399 -70.53968        190.92
## 10 01073000 43.14870 -70.96506         31.30
## 11 01078000 43.56646 -71.74786        222.46
## 12 01118300 41.47482 -71.83424         10.36
## 13 01121000 41.84371 -72.16897         70.25
## 14 01123000 41.67177 -72.05230         77.85
## 15 01134500 44.51172 -71.83731        195.13
## 16 01137500 44.26867 -71.63036        228.55
## 17 01139000 44.15034 -72.06509        246.33
## 18 01139800 44.09284 -72.33565         22.80
## 19 01142500 43.93451 -72.65788         82.17
## 20 01144000 43.71424 -72.41815       1790.24
## 21 01162500 42.68259 -72.11508         49.71
## 22 01169000 42.63842 -72.72509        230.64
## 23 01170100 42.70342 -72.67065        106.99
## 24 01181000 42.23731 -72.89565        243.50
## 25 01187300 42.03732 -72.93899         53.92


disch <- round(t(t(discharge)*86.4/GAUGES_INFO$DRAINAGE_AREA), 4)
```

Add information on gauges coordinates:

```
catchments$lat <- GAUGES_INFO[,c(2)]
catchments$lon <- GAUGES_INFO[,c(3)]

IDcatchments <- catchments$Group.1
 IDcatchments


##  [1] "01013500" "01022500" "01030500" "01031500" "01047000" "01052500"
##  [7] "01054200" "01055000" "01057000" "01073000" "01078000" "01118300"
## [13] "01121000" "01123000" "01134500" "01137500" "01139000" "01139800"
## [19] "01142500" "01144000" "01162500" "01169000" "01170100" "01181000"
## [25] "01187300"
```

3 Input climate variables

Read the data file names:

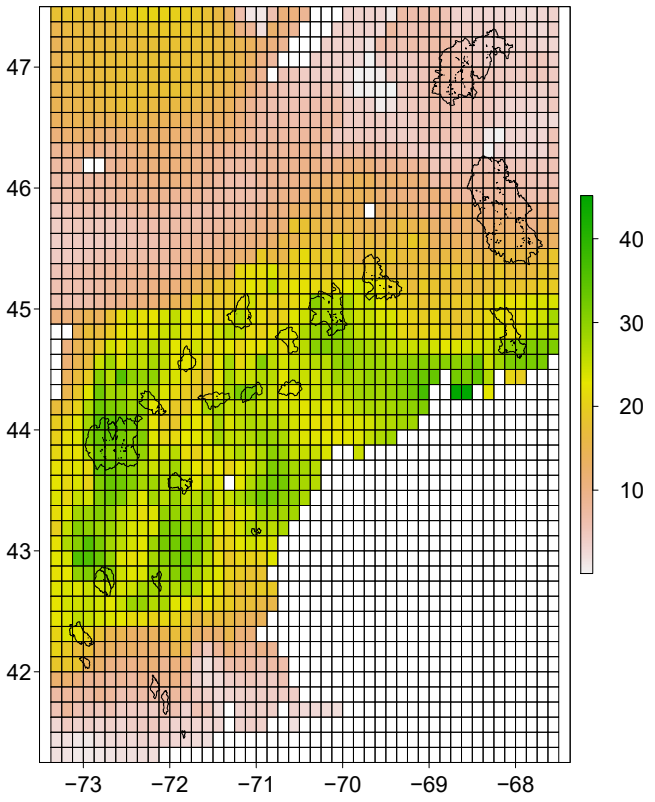
```
path_clim <- '/work/users/matteo/Models/PASS/testModels/UFZHalleWork/AmericanData/NCALDAS'
clim.files <- list.files(path = path_clim, pattern = '.nc4', recursive = TRUE, full.names=TRUE)
```

Select the period of interest: e.g. 2000-2010

```
files10yrs <- clim.files[clim.files>= grep('20001001', clim.files, value=TRUE) &
                         clim.files <= grep('20100930', clim.files, value=TRUE)]
```

Select one day, as an example, and plot the rain in mm/day:

```
file_test <- files10yrs[6]  # it rained a bit in the 6th day
Y <- raster(file_test, varname='Rainf')
# delimit the region of interest
#extent_Bacini <- extent(catchments)*1.08

#crop area
Ycrop <- crop(Y, extent_Bacini)*86400  # in mm/day
lon.m <- sort(unique(rasterToPoints(Ycrop, spatial=FALSE)[,1]))
lat.m <- rev(unique(rasterToPoints(Ycrop, spatial=FALSE)[,2]))

res.m <- res(Ycrop)[1]
irow.m <- nrow(Ycrop)
icol.m <- ncol(Ycrop)
grd.id <- matrix(1:(irow.m*icol.m), nrow=irow.m, byrow=TRUE)

plot(Ycrop)
 plot(catchments, add=T)
grid.master <-  Ycrop
m.grid <- rasterToPolygons(grid.master, na.rm=FALSE)
lines(m.grid)
# text(m.grid, labels=as.numeric(rownames(m.grid@data)), cex=0.50, col=2, font=2)
# ok
```



Build a raster as a grid for the inputs:

```
r <- raster(
         xmn=range(lon.m)[1] - (0.5*res.m), xmx=range(lon.m)[2] + (0.5*res.m),
         ymn=range(lat.m)[1] - (0.5*res.m), ymx=range(lat.m)[2] + (0.5*res.m),
         ncol=icol.m, nrow=irow.m,
         crs=CRS('+proj=longlat +datum=WGS84 +no_defs'),
         vals=1
       )
r100 <- raster(res=res(r)/100, extent(r), crs=CRS('+proj=longlat +datum=WGS84 +no_defs'), vals=1)
```

Create the raster of weights:

```
weights <- array(NA, dim=c(irow.m, icol.m, length(catchments)))
for (icat in 1:length(catchments)){
        dummy <- mask(r100, catchments[icat,])
        dummy[is.na(dummy)] <- 0
        weights[,,icat] <- as.matrix(aggregate(dummy, fact=100, fun=sum)/10000) #matrix of catchment area coverage
        weights[,,icat][weights[,,icat] <= 0.005] <- 0
}  # it takes 2 min
 dimnames(weights)[[3]] <- IDcatchments
```

Populate the input arrays:

```
indices <- list()
for (i in 1:length(catchments)){
 res <- which(!(weights[,,i] == 0), arr.ind=T)
 indices[[i]] <- sort(grd.id[res])
}

pixels_cov <- unique(unlist(indices))

prec <- array(numeric(0),c(length(lat.m), length(lon.m), length(files10yrs)))
temp <- array(numeric(0),c(length(lat.m), length(lon.m), length(files10yrs)))
tmax <- array(numeric(0),c(length(lat.m), length(lon.m), length(files10yrs)))
tmin <- array(numeric(0),c(length(lat.m), length(lon.m), length(files10yrs)))
pet <- array(numeric(0),c(length(lat.m), length(lon.m), length(files10yrs)))

for (i in 1:dim(prec)[3]){
  print(i)
  Pr <- raster(files10yrs[i], varname='Rainf') # load the variable as raster
  Pr_crop <- crop(Pr, extent_Bacini) # extent of the selected Basins (North-Eastern US)
  prec[,,i] <- (as.matrix(Pr_crop))*86400 # convert to matrix and to [mm/d]
  Temp <- raster(files10yrs[i], varname='Tair_f')
  Temp_crop <- crop(Temp, extent_Bacini)
  temp[,,i] <- (as.matrix(Temp_crop))-273.15
  Temp_max <- raster(files10yrs[i], varname='Tair_f_max')
  Temp_max_crop <- crop(Temp_max, extent_Bacini)
  tmax[,,i] <- (as.matrix(Temp_max_crop))-273.15
  Temp_min <- raster(files10yrs[i], varname='Tair_f_min')
  Temp_min_crop <- crop(Temp_min, extent_Bacini)
  tmin[,,i] <- (as.matrix(Temp_min_crop))-273.15
  PotEvap <- raster(files10yrs[i], varname='PotEvap')
  PotEvap_crop <- crop(PotEvap, extent_Bacini)
  pet[,,i] <- (as.matrix(PotEvap_crop))*86400
}  # i from 1 to 3652 ; it takes long: 18 min
```

Convert the inputs into a 2D structure:

```
train.data <- vector('list', 4)
names(train.data) <- c('prec','temp','pet','qobs')
t.arr <- matrix(NA, nrow = dim(prec)[3], ncol = dim(prec)[1]*dim(prec)[2])
for (i in 1:3) train.data[[i]] <- t.arr

z <- 1
for (i in 1:dim(prec)[1]){
  for (j in 1:dim(prec)[2]){
    train.data$prec[,z] <- prec[i,j,]
```

```
    train.data$temp[,z] <- temp[i,j,]
    train.data$pet[,z] <- pet[i,j,]
    z <- z + 1
  }
}

for (i in 1:3){
  colnames(train.data[[i]]) <- 1:(dim(prec)[1]*dim(prec)[2])
  rownames(train.data[[i]]) <- as.character(seq(as.Date('2000-10-01'), as.Date('2010-09-30'), "days"))
}

train.data$qobs <- disch

train.data$prec <- train.data$prec[, c(pixels_cov)]
train.data$temp <- train.data$temp[, c(pixels_cov)]
train.data$pet <- train.data$pet[, c(pixels_cov)]
 dim(train.data$prec)
```

```
## [1] 3652  215
```

```
 colnames(train.data$prec)
```

```
##   [1] "87"   "88"   "89"   "131"  "132"  "133"  "134"  "135"  "136"  "137"
##  [11] "177"  "178"  "179"  "180"  "181"  "182"  "183"  "184"  "224"  "225"
##  [21] "226"  "227"  "228"  "229"  "271"  "272"  "273"  "274"  "275"  "319"
##  [31] "320"  "322"  "935"  "936"  "981"  "982"  "983"  "1029" "1030" "1031"
##  [41] "1076" "1077" "1078" "1125" "463"  "464"  "509"  "510"  "511"  "512"
##  [51] "513"  "556"  "557"  "558"  "559"  "560"  "561"  "603"  "604"  "605"
##  [61] "606"  "607"  "608"  "609"  "651"  "652"  "653"  "654"  "655"  "656"
##  [71] "698"  "699"  "700"  "701"  "702"  "703"  "746"  "747"  "748"  "749"
##  [81] "750"  "751"  "793"  "794"  "795"  "796"  "797"  "798"  "781"  "782"
##  [91] "783"  "828"  "829"  "830"  "831"  "875"  "876"  "877"  "878"  "879"
## [101] "923"  "924"  "925"  "926"  "871"  "872"  "873"  "917"  "918"  "919"
## [111] "920"  "965"  "966"  "967"  "968"  "1012" "1013" "1014" "1015" "911"
## [121] "912"  "957"  "958"  "959"  "1004" "1005" "1006" "1148" "1193" "1194"
## [131] "1195" "1240" "1241" "1008" "1009" "1010" "1056" "1057" "1150" "1151"
## [141] "1152" "1197" "1198" "1199" "1617" "1618" "1422" "1468" "1469" "1470"
## [151] "1471" "1517" "2222" "2269" "2078" "2125" "2126" "2173" "1047" "1093"
## [161] "1094" "1095" "1141" "1189" "1190" "1191" "1192" "1237" "1238" "1239"
## [171] "1184" "1185" "1231" "1232" "1233" "1280" "1277" "1278" "1275" "1276"
## [181] "1322" "1323" "1273" "1274" "1320" "1321" "1324" "1367" "1368" "1369"
## [191] "1370" "1371" "1372" "1414" "1415" "1416" "1417" "1418" "1749" "1750"
## [201] "1796" "1797" "1744" "1745" "1791" "1792" "1882" "1883" "1929" "1930"
## [211] "1931" "1977" "1978" "2024" "2025"
```

```
str(train.data)
```

```
## List of 4
##  $ prec: num [1:3652, 1:215] 0 0 0.2784 0 0.0504 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : chr [1:3652] "2000-10-01" "2000-10-02" "2000-10-03" "2000-10-04" ...
##   .. ..$ : chr [1:215] "87" "88" "89" "131" ...
##  $ temp: num [1:3652, 1:215] 12.41 14.14 13.03 12.29 8.11 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : chr [1:3652] "2000-10-01" "2000-10-02" "2000-10-03" "2000-10-04" ...
##   .. ..$ : chr [1:215] "87" "88" "89" "131" ...
##  $ pet : num [1:3652, 1:215] 1.43 2.24 1.7 2.04 2.05 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : chr [1:3652] "2000-10-01" "2000-10-02" "2000-10-03" "2000-10-04" ...
##   .. ..$ : chr [1:215] "87" "88" "89" "131" ...
##  $ qobs: num [1:3652, 1:25] 0.116 0.113 0.113 0.105 0.101 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : chr [1:3652] "2000-10-01" "2000-10-02" "2000-10-03" "2000-10-04" ...
##   .. ..$ : chr [1:25] "01013500" "01022500" "01030500" "01031500" ...
```

4 Catchment descriptors

4.1 Climate descriptors

Temperature:

```
tmean_series <- matrix(NA, nrow = dim(temp)[3], ncol = dim(temp)[1]*dim(temp)[2])
z <- 1
for (i in 1:dim(temp)[1]){
  for (j in 1:dim(temp)[2]){
    tmean_series[,z] <- temp[i,j,]
    z <- z + 1
  }
}
 colnames(tmean_series) <- 1:ncol(tmean_series)
library(zoo)
tmean_series <- zoo(tmean_series, seq(as.Date('2000-10-01'), as.Date('2010-09-30'), "days"))
```

Precipitation:

```
prec_series <- matrix(NA, nrow = dim(temp)[3], ncol = dim(temp)[1]*dim(temp)[2])
z <- 1
for (i in 1:dim(prec)[1]){
  for (j in 1:dim(prec)[2]){
    prec_series[,z] <- prec[i,j,]
    z <- z + 1
  }
}
 colnames(prec_series) <- 1:ncol(prec_series)
prec_series <- zoo(prec_series, seq(as.Date('2000-10-01'), as.Date('2010-09-30'), "days"))
```

Potential evapotranspiration:

```
pet_series <- matrix(NA, nrow = dim(temp)[3], ncol = dim(temp)[1]*dim(temp)[2])
z <- 1
for (i in 1:dim(pet)[1]){
  for (j in 1:dim(pet)[2]){
    pet_series[,z] <- pet[i,j,]
    z <- z + 1
  }
}
 colnames(pet_series) <- 1:ncol(pet_series)
pet_series <- zoo(pet_series, seq(as.Date('2000-10-01'), as.Date('2010-09-30'), "days"))
```

Climatic descriptors:

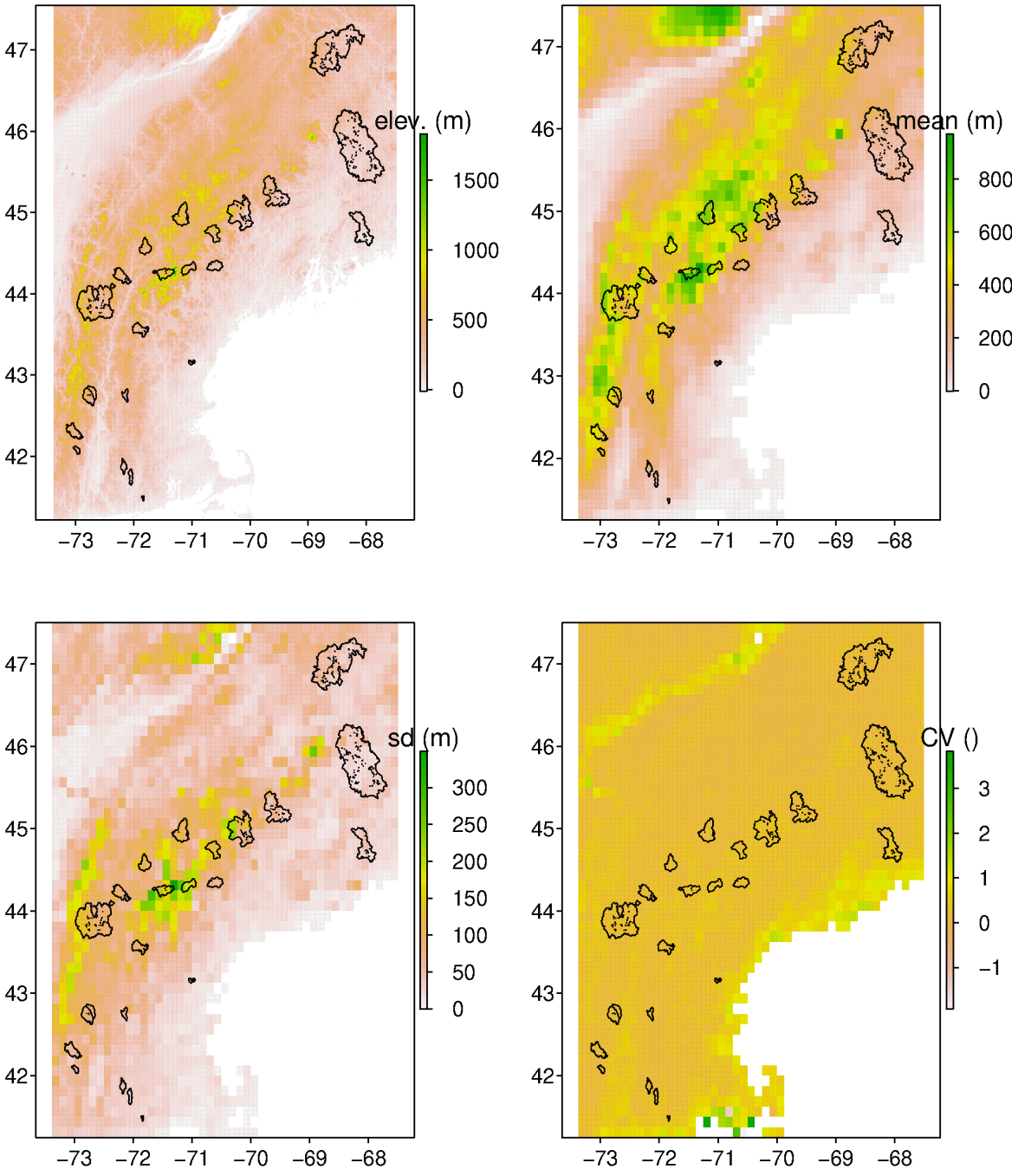
```
# Long-term mean annual precipitation
CL_MAP <- round(apply(prec_series, 2, mean)*365.25)  # [mm]
# Long-term mean annual temperature
CL_MAT <- round(apply(tmean_series, 2, mean), 1)  # [degC]
# Long-term mean potential evapotranspiration
CL_PET <- round(apply(pet_series, 2, mean)*365.25)  # [mm]
# Aridity index (Budyko, 1974) as ratio of mean annual potential evaporation and mean annual precipitation
CL_PETovP <- round(CL_PET/CL_MAP, 2) # [-]

monthlyP <- aggregate(prec_series, format(time(prec_series), '%m'), mean)*30  # [mm]
# ratio of long-term summer precipitation (May-Oct) and winter precipitation (Nov-Apr)
CL_Psum2win <- round(apply(monthlyP[5:10,], 2, sum)/apply(monthlyP[-c(5:10),], 2, sum), 2)
# Long-term median maximum daily precipitation
CL_R50 <- round(apply(aggregate(prec_series, format(time(prec_series), '%Y'), max), 2,
                median), 1)  # [mm/day]
# Long-term 95th percentile of maximum daily prec.
CL_R95 <- round(apply(aggregate(prec_series, format(time(prec_series), '%Y'), max), 2,
                quantile, probs=0.95, na.rm=T), 1) # [mm/day]
# Long-term mean absolute difference of rainfall amount between two consecutives days
CL_dRD2D <- round(apply(prec_series, 2, FUN=function(x) mean(abs(diff(x)))), 2) # [mm]
```

4.2 Morphology and topology descriptors

Calculate elevation statistics over the cells:

```
incells <- extract(dem, m.grid)  # it takes 1 min

elevstats <- list()
elevstats$ncells <- unlist(lapply(incells, function(x) sum(!is.na(x))))
elevstats$meanelev <- unlist(lapply(incells, function(x) if (sum(!is.na(x)) > 0) round(mean(x, na.rm=TRUE)) else NA))
elevstats$sd <- unlist(lapply(incells, function(x) if (sum(!is.na(x)) > 0) sd(x, na.rm=TRUE) else NA))
elevstats$CV <- round(elevstats$sd/elevstats$meanelev, 4)
elevstats$maxelev <- unlist(lapply(incells, function(x) if (sum(!is.na(x)) > 0) max(x, na.rm=TRUE) else NA))
elevstats$minelev <- unlist(lapply(incells, function(x) if (sum(!is.na(x)) > 0) min(x, na.rm=TRUE) else NA))
elevstats$q01 <- unlist(lapply(incells, function(x) if (sum(!is.na(x)) > 0) quantile(x, prob=.01, na.rm=TRUE) else NA))
elevstats$q05 <- unlist(lapply(incells, function(x) if (sum(!is.na(x)) > 0) quantile(x, prob=.05, na.rm=TRUE) else NA))
elevstats$q10 <- unlist(lapply(incells, function(x) if (sum(!is.na(x)) > 0) quantile(x, prob=.10, na.rm=TRUE) else NA))
elevstats$q25 <- unlist(lapply(incells, function(x) if (sum(!is.na(x)) > 0) quantile(x, prob=.25, na.rm=TRUE) else NA))
elevstats$q50 <- unlist(lapply(incells, function(x) if (sum(!is.na(x)) > 0) quantile(x, prob=.50, na.rm=TRUE) else NA))
elevstats$q75 <- unlist(lapply(incells, function(x) if (sum(!is.na(x)) > 0) quantile(x, prob=.75, na.rm=TRUE) else NA))
elevstats$q90 <- unlist(lapply(incells, function(x) if (sum(!is.na(x)) > 0) quantile(x, prob=.90, na.rm=TRUE) else NA))
elevstats$q95 <- unlist(lapply(incells, function(x) if (sum(!is.na(x)) > 0) quantile(x, prob=.95, na.rm=TRUE) else NA))
elevstats$q99 <- unlist(lapply(incells, function(x) if (sum(!is.na(x)) > 0) quantile(x, prob=.99, na.rm=TRUE) else NA))
elevstats$ncells2000 <- unlist(lapply(incells, function(x) if (sum(!is.na(x)) > 0) sum(x >= 2000, na.rm=TRUE) else NA))
elevstats <- as.data.frame(elevstats)
 str(elevstats)

elevstats$CV[which(elevstats$CV=='Inf')] <- NaN
MP_mean_dem <- elevstats$meanelev
MP_cv_dem <- elevstats$CV
```

```
layout(matrix(1:4, nrow=2, byrow=TRUE))
plot(dem, legend.args=list(text='elev. (m)'))
 plot(catchments, add=TRUE)

values(grid.master) <- elevstats$meanelev
plot(grid.master, legend.args=list(text='mean (m)'))
 plot(catchments, add=TRUE)

values(grid.master) <- elevstats$sd
plot(grid.master, legend.args=list(text='sd (m)'))
 plot(catchments, add=TRUE)

values(grid.master) <- elevstats$CV
plot(grid.master, legend.args=list(text='CV ()'))
 plot(catchments, add=TRUE)
```



Set GRASS environment and database location:

```
initGRASS(gisBase='/usr/lib/grass78', home=getwd(), gisDbase='GRASS_TEMP', mapset='PERMANENT', override=TRUE)
execGRASS('g.proj', flags = 'c', epsg = 4326)
execGRASS('g.mapset', flags = 'c', mapset = 'new_mapset')
```

Import raster to GRASS and set region:

```
writeRaster(dem, filename = 'DTM.tif', format = 'GTiff', overwrite = TRUE)
execGRASS('r.in.gdal', flags='o', parameters=list(input='DTM.tif', output='elev'))
execGRASS('g.region', flags='p', parameters=list(raster='elev'))

execGRASS('r.watershed', flags='overwrite',
          parameters=list(elevation='elev', threshold=1000, convergence=5, memory=300,
                          accumulation='ACCUMp', tci='TCIp', spi='SPIp', drainage='DRAINp', basin='BASINp', stream='STREAMp',
                          length_slope='LSLOPEp', slope_steepness='STEEPp'))
dir.create(file.path("./", "GRASSoutputs"), showWarnings = FALSE)
```

Run grass tools and export results:

```
execGRASS('r.out.gdal', flags='overwrite', parameters=list(input=c('ACCUMp'),
                                                           output=c('./GRASSoutputs/ACCUMp.tif'), nodata=-9999))
execGRASS('r.out.gdal', flags='overwrite', parameters=list(input=c('TCIp'),
                                                           output=c('./GRASSoutputs/TCIp.tif'), nodata=-9999))
execGRASS('r.out.gdal', flags='overwrite', parameters=list(input=c('SPIp'),
                                                           output=c('./GRASSoutputs/SPIp.tif'), nodata=-9999))
execGRASS('r.out.gdal', flags='overwrite', parameters=list(input=c('DRAINp'),
                                                           output=c('./GRASSoutputs/DRAINp.tif'), nodata=-9999))
execGRASS('r.out.gdal', flags='overwrite', parameters=list(input=c('BASINp'),
                                                           output=c('./GRASSoutputs/BASINp.tif')))
execGRASS('r.out.gdal', flags='overwrite', parameters=list(input=c('STREAMp'),
                                                           output=c('./GRASSoutputs/STREAMp.tif')))
execGRASS('r.out.gdal', flags='overwrite', parameters=list(input=c('LSLOPEp'),
                                                           output=c('./GRASSoutputs/LSLOPEp.tif'), nodata=-9999))
execGRASS('r.out.gdal', flags='overwrite', parameters=list(input=c('STEEPp'),
                                                           output=c('./GRASSoutputs/STEEPp.tif'), nodata=-9999))
```

Read exported files:

```
DRAINp <- raster('./GRASSoutputs/DRAINp.tif')
 #plot(DRAINp)
ACCUMp <- raster('./GRASSoutputs/ACCUMp.tif')
 #plot(ACCUMp)
 #plot(log(ACCUMp))
BASINp <- raster('./GRASSoutputs/BASINp.tif')
 #plot(BASINp)
STREAMp <- raster('./GRASSoutputs/STREAMp.tif')
 #plot(STREAMp)
TCIp <- raster('./GRASSoutputs/TCIp.tif')
 #plot(TCIp)
SPIp <- raster('./GRASSoutputs/SPIp.tif')
 #plot(SPIp)
LSLOPEp <- raster('./GRASSoutputs/LSLOPEp.tif')
 #plot(LSLOPEp)
STEEPp <- raster('./GRASSoutputs/STEEPp.tif')
 #plot(STEEPp)
```

From STREAMp (raster) calculate the shapefile of river network (STREAMvect.shp):

```
execGRASS('r.thin', flags='overwrite', parameters=list(input='STREAMp', output='thinned'))
execGRASS('r.to.vect', flags=c('overwrite', 'v'), parameters=list(input='thinned',
                                                                  output='STREAMvect', type='line'))
#export
execGRASS("v.out.ogr", flags='overwrite', parameters=list(input=c('STREAMvect'),
                                                          output=c('./GRASSoutputs/STREAMvect.shp')))
#read exported file
streams <- readOGR('./GRASSoutputs/STREAMvect.shp')
#plot(dem)
 #lines(streams)
 #ok
```

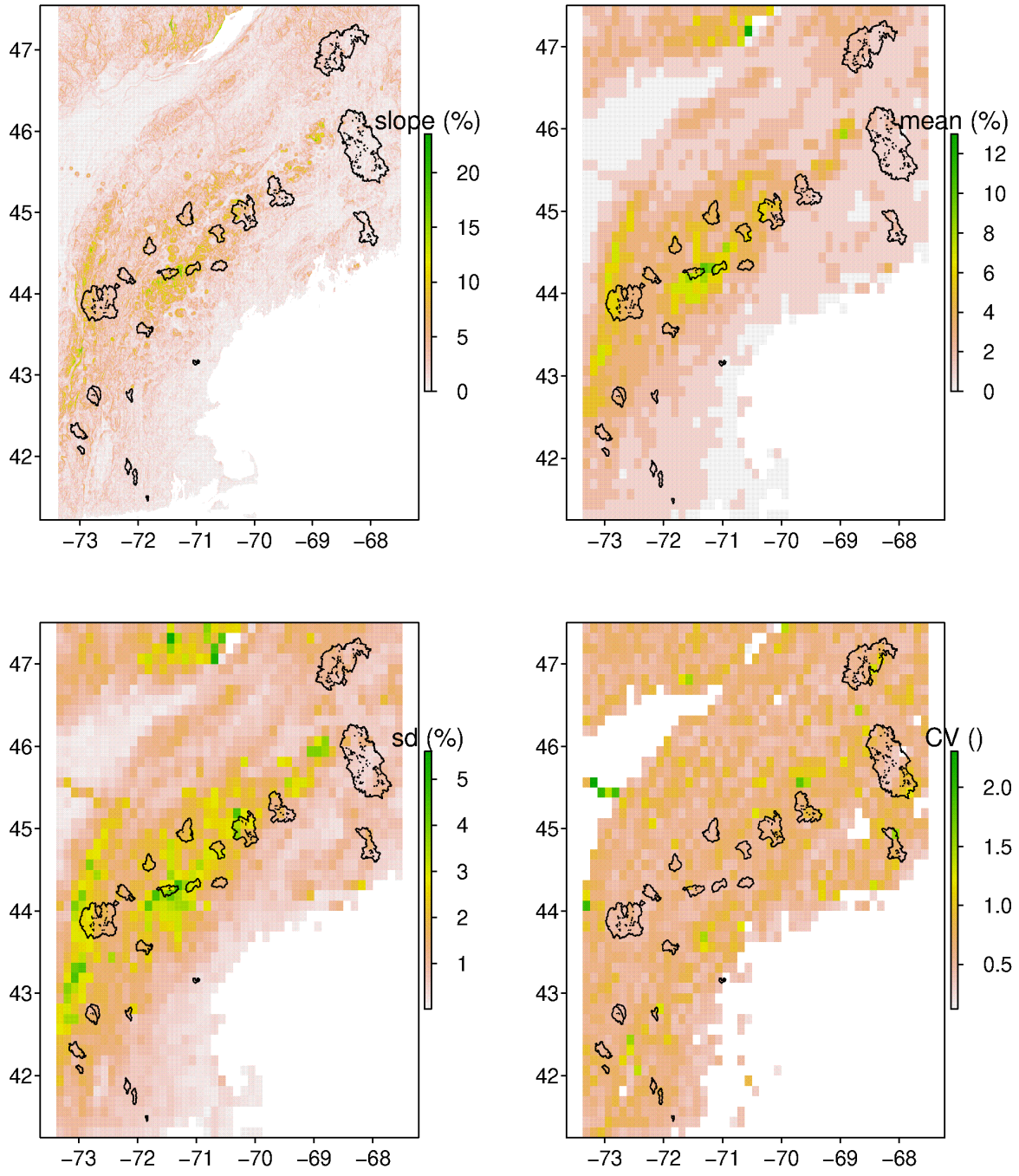
Calculate slope and aspect:

```
execGRASS('r.slope.aspect', flags='overwrite', parameters=list(elevation='elev', slope='SLOPEp', aspect='ASPECTp',
                                                               pcurvature='PCURVp', tcurvature='TCURVp'))
#export
execGRASS('r.out.gdal', flags='overwrite', parameters=list(input=c('SLOPEp'),
                                                           output=c('./GRASSoutputs/SLOPEp.tif'), nodata=-9999))
execGRASS('r.out.gdal', flags='overwrite', parameters=list(input=c('ASPECTp'),
                                                           output=c('./GRASSoutputs/ASPECTp.tif'), nodata=-9999))
execGRASS('r.out.gdal', flags='overwrite', parameters=list(input=c('PCURVp'),
                                                           output=c('./GRASSoutputs/PCURVp.tif'), nodata=-9999))
execGRASS('r.out.gdal', flags='overwrite', parameters=list(input=c('TCURVp'),
                                                           output=c('./GRASSoutputs/TCURVp.tif'), nodata=-9999))
#read exported file
SLOPEp <- raster('./GRASSoutputs/SLOPEp.tif')
 #plot(SLOPEp)
ASPECTp <- raster('./GRASSoutputs/ASPECTp.tif')
 #plot(ASPECTp)
PCURVp <- raster('./GRASSoutputs/PCURVp.tif')
 #plot(PCURVp)
TCURVp <- raster('./GRASSoutputs/TCURVp.tif')
 #plot(TCURVp)
```

Calculate slope statistics over the cells:

```
incells <- extract(SLOPEp, m.grid)  # it takes 4 min

slopestats <- list()
slopestats$ncells <- unlist(lapply(incells, function(x) sum(!is.na(x))))
slopestats$mean <- unlist(lapply(incells, function(x) if (sum(!is.na(x)) > 0) round(mean(x, na.rm=TRUE)) else NA))
slopestats$sd <- unlist(lapply(incells, function(x) if (sum(!is.na(x)) > 0) sd(x, na.rm=TRUE) else NA))
slopestats$CV <- slopestats$sd/slopestats$mean
slopestats$max <- unlist(lapply(incells, function(x) if (sum(!is.na(x)) > 0) max(x, na.rm=TRUE) else NA))
slopestats$min <- unlist(lapply(incells, function(x) if (sum(!is.na(x)) > 0) min(x, na.rm=TRUE) else NA))
slopestats$q01 <- unlist(lapply(incells, function(x) if (sum(!is.na(x)) > 0) quantile(x, prob=.01, na.rm=TRUE) else NA))
slopestats$q05 <- unlist(lapply(incells, function(x) if (sum(!is.na(x)) > 0) quantile(x, prob=.05, na.rm=TRUE) else NA))
slopestats$q10 <- unlist(lapply(incells, function(x) if (sum(!is.na(x)) > 0) quantile(x, prob=.10, na.rm=TRUE) else NA))
slopestats$q25 <- unlist(lapply(incells, function(x) if (sum(!is.na(x)) > 0) quantile(x, prob=.25, na.rm=TRUE) else NA))
slopestats$q50 <- unlist(lapply(incells, function(x) if (sum(!is.na(x)) > 0) quantile(x, prob=.50, na.rm=TRUE) else NA))
slopestats$q75 <- unlist(lapply(incells, function(x) if (sum(!is.na(x)) > 0) quantile(x, prob=.75, na.rm=TRUE) else NA))
slopestats$q90 <- unlist(lapply(incells, function(x) if (sum(!is.na(x)) > 0) quantile(x, prob=.90, na.rm=TRUE) else NA))
slopestats$q95 <- unlist(lapply(incells, function(x) if (sum(!is.na(x)) > 0) quantile(x, prob=.95, na.rm=TRUE) else NA))
slopestats$q99 <- unlist(lapply(incells, function(x) if (sum(!is.na(x)) > 0) quantile(x, prob=.99, na.rm=TRUE) else NA))
slopestats <- as.data.frame(slopestats)

MP_mean_slope <- slopestats$mean
```

```
layout(matrix(1:4, nrow=2, byrow=TRUE))
plot(SLOPEp, legend.args=list(text='slope (%)'))
 plot(catchments, add=TRUE)

values(grid.master) <- slopestats$mean
plot(grid.master, legend.args=list(text='mean (%)'))
 plot(catchments, add=TRUE)

values(grid.master) <- slopestats$sd
plot(grid.master, legend.args=list(text='sd (%)'))
 plot(catchments, add=TRUE)

values(grid.master) <- slopestats$CV
plot(grid.master, legend.args=list(text='CV ()'))
 plot(catchments, add=TRUE)
```

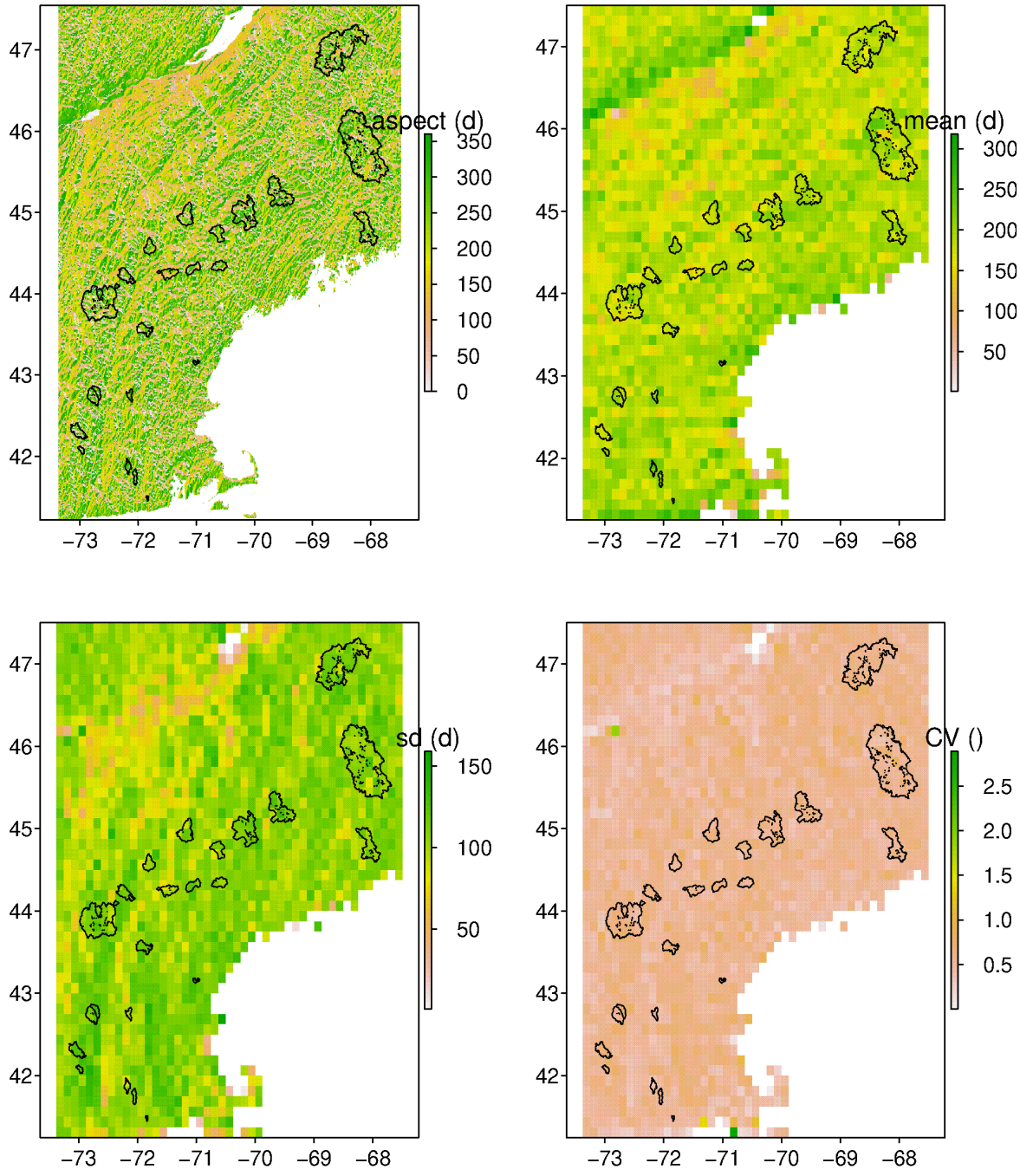


Calculate aspect statistics over the cells:

```
incells <- extract(ASPECTp, m.grid)  # it takes 4 min

aspectstats <- list()
aspectstats$ncells <- unlist(lapply(incells, function(x) sum(!is.na(x))))
aspectstats$mean <- unlist(lapply(incells, function(x) if (sum(!is.na(x)) > 0) round(mean(x, na.rm=TRUE)) else NA))
aspectstats$sd <- unlist(lapply(incells, function(x) if (sum(!is.na(x)) > 0) sd(x, na.rm=TRUE) else NA))
aspectstats$CV <- aspectstats$sd/aspectstats$mean
```

```
aspectstats$max <- unlist(lapply(incells, function(x) if (sum(!is.na(x)) > 0) max(x, na.rm=TRUE) else NA))
aspectstats$min <- unlist(lapply(incells, function(x) if (sum(!is.na(x)) > 0) min(x, na.rm=TRUE) else NA))
aspectstats$q01 <- unlist(lapply(incells, function(x) if (sum(!is.na(x)) > 0) quantile(x, prob=.01, na.rm=TRUE) else NA))
aspectstats$q05 <- unlist(lapply(incells, function(x) if (sum(!is.na(x)) > 0) quantile(x, prob=.05, na.rm=TRUE) else NA))
aspectstats$q10 <- unlist(lapply(incells, function(x) if (sum(!is.na(x)) > 0) quantile(x, prob=.10, na.rm=TRUE) else NA))
aspectstats$q25 <- unlist(lapply(incells, function(x) if (sum(!is.na(x)) > 0) quantile(x, prob=.25, na.rm=TRUE) else NA))
aspectstats$q50 <- unlist(lapply(incells, function(x) if (sum(!is.na(x)) > 0) quantile(x, prob=.50, na.rm=TRUE) else NA))
aspectstats$q75 <- unlist(lapply(incells, function(x) if (sum(!is.na(x)) > 0) quantile(x, prob=.75, na.rm=TRUE) else NA))
aspectstats$q90 <- unlist(lapply(incells, function(x) if (sum(!is.na(x)) > 0) quantile(x, prob=.90, na.rm=TRUE) else NA))
aspectstats$q95 <- unlist(lapply(incells, function(x) if (sum(!is.na(x)) > 0) quantile(x, prob=.95, na.rm=TRUE) else NA))
aspectstats$q99 <- unlist(lapply(incells, function(x) if (sum(!is.na(x)) > 0) quantile(x, prob=.99, na.rm=TRUE) else NA))
aspectstats <- as.data.frame(aspectstats)

MP_mean_aspect <- aspectstats$mean
```

```
layout(matrix(1:4, nrow=2, byrow=TRUE))
plot(ASPECTp, legend.args=list(text='aspect (d)'))
 plot(catchments, add=TRUE)

values(grid.master) <- aspectstats$mean
plot(grid.master, legend.args=list(text='mean (d)'))
 plot(catchments, add=TRUE)

values(grid.master) <- aspectstats$sd
plot(grid.master, legend.args=list(text='sd (d)'))
 plot(catchments, add=TRUE)

values(grid.master) <- aspectstats$CV
plot(grid.master, legend.args=list(text='CV ()'))
 plot(catchments, add=TRUE)
```



Here circular statistics shoud have been used.

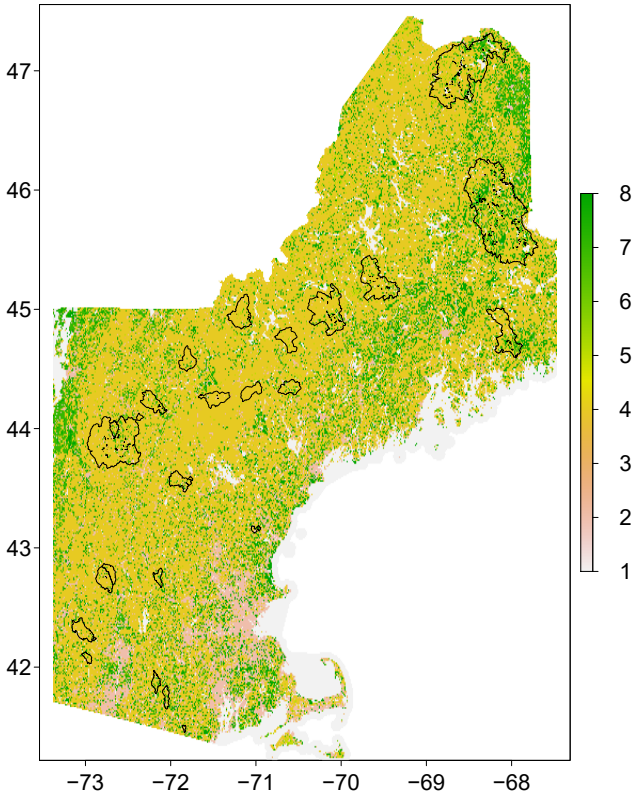
5 Land use

```
landuse.file <- paste('/work/users/matteo/Models/PASS/testModels/UFZHalleWork/',
                      'AmericanData/NLCD_landcover/nlcd_2019_land_cover_l48_20210604.img', sep='')
land <- raster(landuse.file)
land0 <- crop(land, extent(1000005, 2342655, 2299995, 3310005))
land <- projectRaster(land0, res=c(0.0125,0.0125), crs="+proj=longlat +ellps=WGS84 +datum=WGS84 +no_defs", method='ngb')
land_crop <- crop(land, extent_Bacini)
```

Land cover reclassification:

```
#unique(as.matrix(land_crop))
rclmat <- matrix(c(10.9,12,1,  # classes 11 and 12 Water
                   20.9,24,2, # classes 21 to 24 Developed
                   30.9,31,3,       # class 31 Barren
                   40.9,43,4,  # classes 41 to 43 Forest
                   50.9,52,5,  # classes 51 to 52 Shrubland
                   70.9,74,6,  # classes 71 to 74 Herbaceous
                   80.9,82,7,  # classes 81 to 82 Planted/cultivated
                   89.9,90,8,   # classes 90 and 95 to Wetlands
                   94.9,95,8,
                   -1,0,NA), ncol=3, byrow=TRUE)
landcrop_rcl <- reclassify(land_crop, rclmat)
```

```
plot(landcrop_rcl)
 #plot(m.grid, add=TRUE)
 plot(catchments, add=TRUE)
```



Landcover statistics over the cells:

```
incells <- extract(landcrop_rcl, m.grid)  # it takes less than 1 min

# 8 classes:
# Percent of the catchment covered with water
LD_water <- round(100*unlist(lapply(incells, function(x) sum(x == 1, na.rm=T)/sum(!is.na(x)))))  # [%]
# Percent of the catchment covered with artificial surfaces
LD_developed <- round(100*unlist(lapply(incells, function(x) sum(x == 2, na.rm=T)/sum(!is.na(x)))))  # [%]
# Percent of the catchment covered with barren land
LD_barren <- round(100*unlist(lapply(incells, function(x) sum(x == 3, na.rm=T)/sum(!is.na(x)))))  # [%]
# Percent of the catchment covered with forest
LD_forest <- round(100*unlist(lapply(incells, function(x) sum(x == 4, na.rm=T)/sum(!is.na(x)))))  # [%]
# Percent of the catchment covered with shrubland
LD_shrubland <- round(100*unlist(lapply(incells, function(x) sum(x == 5, na.rm=T)/sum(!is.na(x)))))  # [%]
# Percent of the catchment covered with herbaceous
LD_herbaceous <- round(100*unlist(lapply(incells, function(x) sum(x == 6, na.rm=T)/sum(!is.na(x)))))  # [%]
# Percent of the catchment covered with planted/cultivated
LD_planted <- round(100*unlist(lapply(incells, function(x) sum(x == 7, na.rm=T)/sum(!is.na(x)))))  # [%]
# Percent of the catchment covered with wetlands
LD_wetlands <- round(100*unlist(lapply(incells, function(x) sum(x == 8, na.rm=T)/sum(!is.na(x)))))  # [%]
```

6 Soil

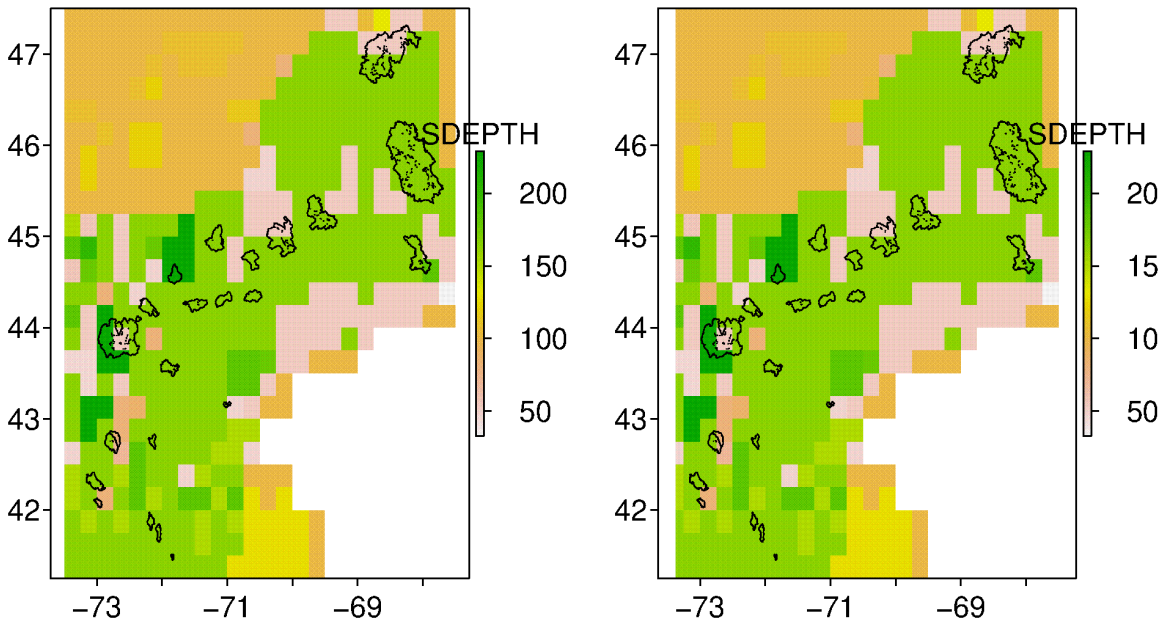
Read soil data:

```
path_soil <- '/work/users/matteo/Models/PASS/testModels/UFZHalleWork/AmericanData/NACP-SOIL/data'
soil.files <- list.files(path = path_soil, pattern = '.tif', recursive = TRUE, full.names=TRUE)
nameschar <- c("SCOMP_DOM", "SDEPTH", "S_CAEX", "S_CLAY", "S_GRAVEL", "S_OC", "S_PH", "S_REF_BULK_DENSITY",
               "S_SAND", "S_SILT", "T_CAEX", "T_CLAY", "T_GRAVEL", "T_OC", "T_PH", "T_REF_BULK_DENSITY",
               "T_SAND", "T_SILT")
SOILchar <- list()
for (i in 1:length(soil.files)){
  SOILchar[[i]] <- crop(raster(soil.files[i]), extent_Bacini)
}
 names(SOILchar) <- nameschar
```

Soil depth:

```
incells <- extract(SOILchar$SDEPTH, m.grid)  # it takes 10 sec
SDEPTH <- unlist(lapply(incells, function(x) if (sum(!is.na(x)) > 0) round(mean(x, na.rm=TRUE),2) else NA))
```

```
layout(matrix(1:2, nrow=1, byrow=TRUE))
plot(SOILchar$SDEPTH, legend.args=list(text='SDEPTH'))
 plot(catchments, add=TRUE)

values(grid.master) <- SDEPTH
plot(grid.master, legend.args=list(text='SDEPTH'))
 plot(catchments, add=TRUE)
```

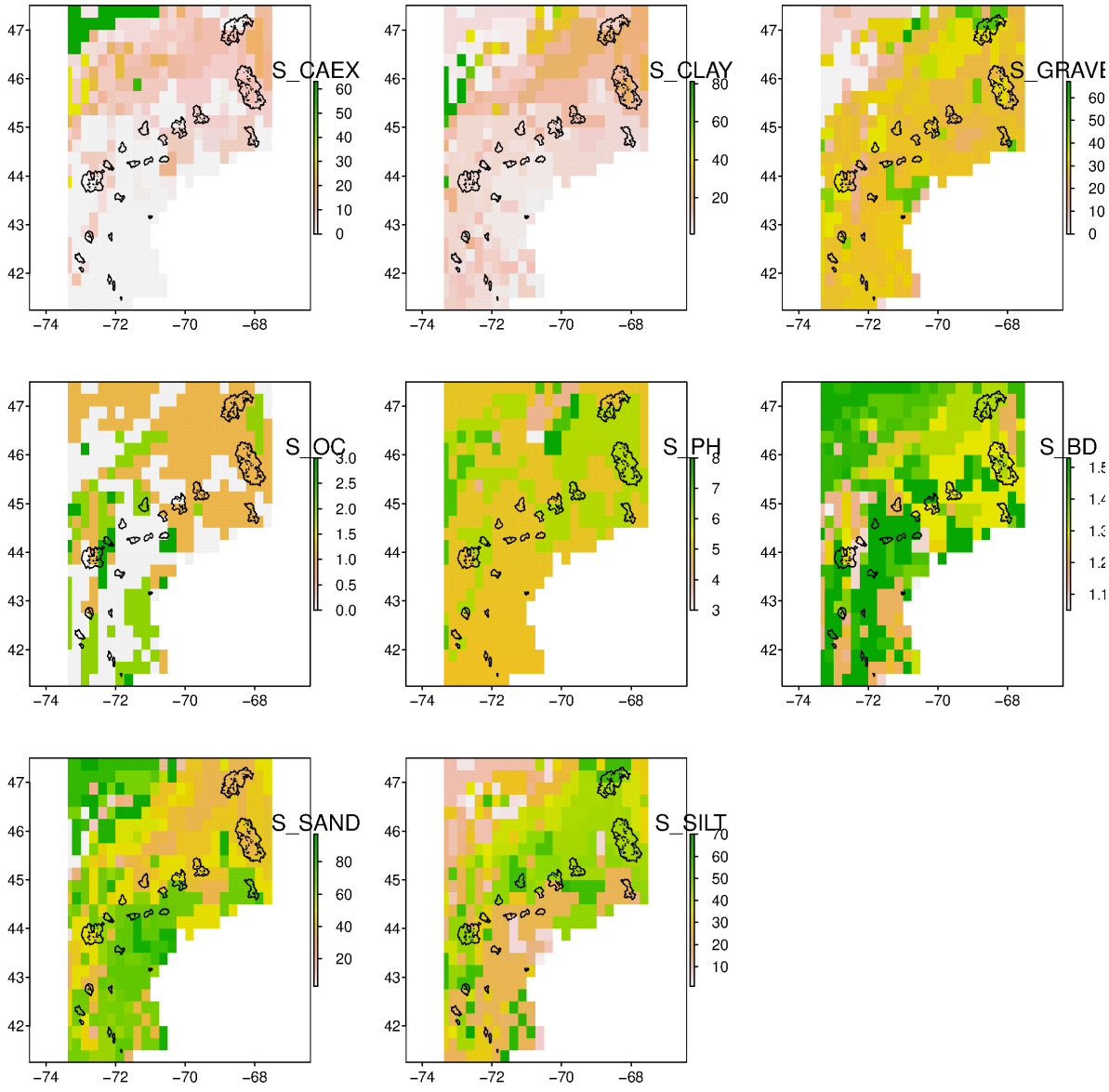


Sub-soil:

```
incells <- extract(SOILchar$S_CAEX, m.grid)  # it takes 16 sec
S_CAEX <- unlist(lapply(incells, function(x) if (sum(!is.na(x)) > 0) round(mean(x, na.rm=TRUE),2) else NA))
incells <- extract(SOILchar$S_CLAY, m.grid)  # it takes 10 sec
S_CLAY <- unlist(lapply(incells, function(x) if (sum(!is.na(x)) > 0) round(mean(x, na.rm=TRUE)) else NA))
incells <- extract(SOILchar$S_GRAVEL, m.grid)  # it takes 10 sec
S_GRAVEL <- unlist(lapply(incells, function(x) if (sum(!is.na(x)) > 0) round(mean(x, na.rm=TRUE)) else NA))
incells <- extract(SOILchar$S_OC, m.grid)
S_OC <- unlist(lapply(incells, function(x) if (sum(!is.na(x)) > 0) round(mean(x, na.rm=TRUE)) else NA))
incells <- extract(SOILchar$S_PH, m.grid)
S_PH <- unlist(lapply(incells, function(x) if (sum(!is.na(x)) > 0) round(mean(x, na.rm=TRUE)) else NA))
incells <- extract(SOILchar$S_REF_BULK_DENSITY, m.grid)
S_REF_BULK_DENSITY <- unlist(lapply(incells, function(x) if (sum(!is.na(x)) > 0) round(mean(x, na.rm=TRUE),2) else NA))
incells <- extract(SOILchar$S_SAND, m.grid)
S_SAND <- unlist(lapply(incells, function(x) if (sum(!is.na(x)) > 0) round(mean(x, na.rm=TRUE)) else NA))
incells <- extract(SOILchar$S_SILT, m.grid)
S_SILT <- unlist(lapply(incells, function(x) if (sum(!is.na(x)) > 0) round(mean(x, na.rm=TRUE)) else NA))
```

```
layout(matrix(1:9, nrow=3, byrow=TRUE))
values(grid.master) <- S_CAEX
plot(grid.master, legend.args=list(text='S_CAEX'))
 plot(catchments, add=TRUE)

values(grid.master) <- S_CLAY
plot(grid.master, legend.args=list(text='S_CLAY'))
 plot(catchments, add=TRUE)

values(grid.master) <- S_GRAVEL
plot(grid.master, legend.args=list(text='S_GRAVEL'))
 plot(catchments, add=TRUE)

values(grid.master) <- S_OC
plot(grid.master, legend.args=list(text='S_OC'))
 plot(catchments, add=TRUE)

values(grid.master) <- S_PH
plot(grid.master, legend.args=list(text='S_PH'))
 plot(catchments, add=TRUE)
```

```
values(grid.master) <- S_REF_BULK_DENSITY
plot(grid.master, legend.args=list(text='S_BD'))
 plot(catchments, add=TRUE)

values(grid.master) <- S_SAND
plot(grid.master, legend.args=list(text='S_SAND'))
 plot(catchments, add=TRUE)

values(grid.master) <- S_SILT
plot(grid.master, legend.args=list(text='S_SILT'))
 plot(catchments, add=TRUE)
```

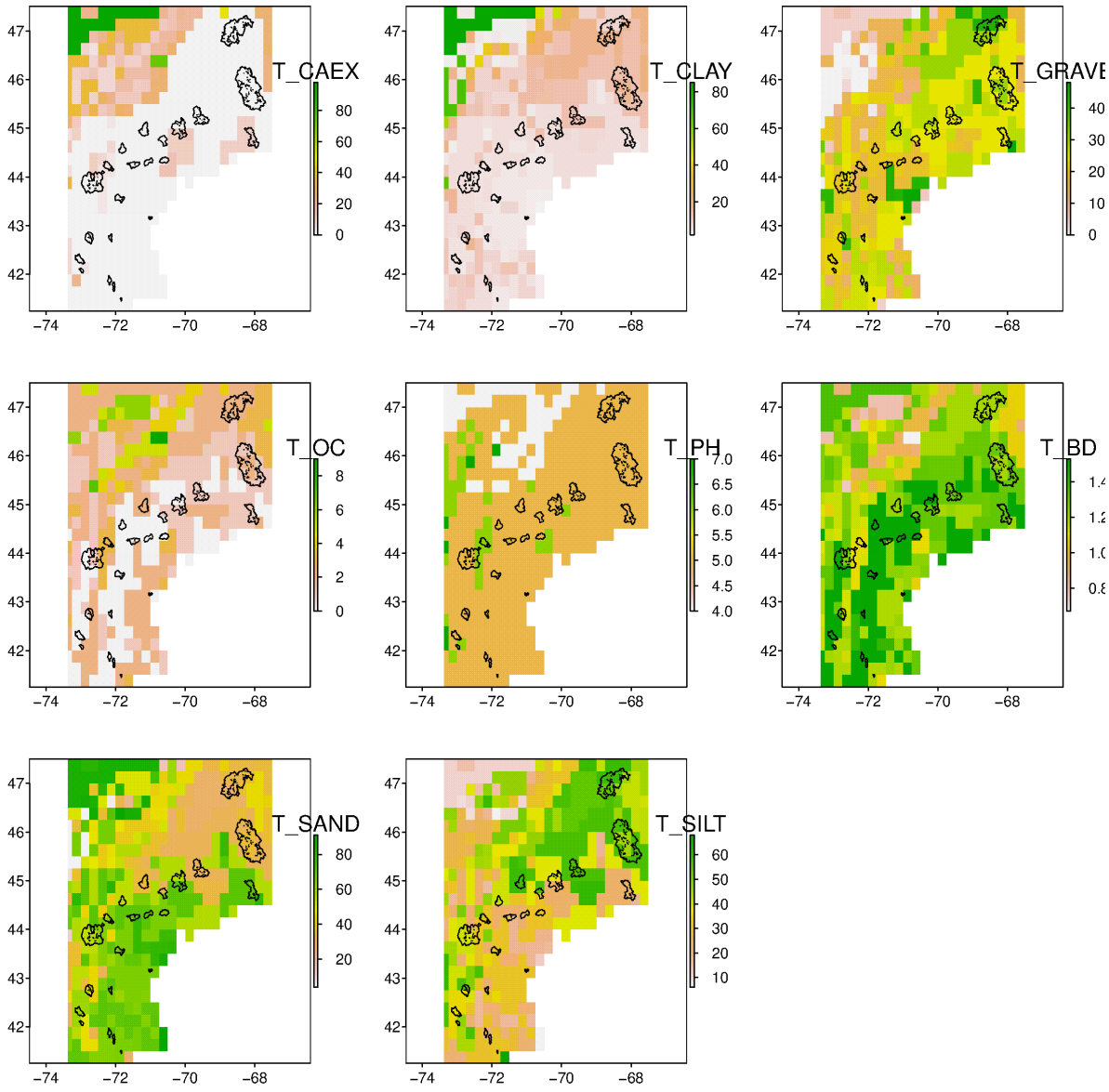


Top-soil:

```
incells <- extract(SOILchar$T_CAEX, m.grid)
T_CAEX <- unlist(lapply(incells, function(x) if (sum(!is.na(x)) > 0) round(mean(x, na.rm=TRUE),2) else NA))
```

```
incells <- extract(SOILchar$T_CLAY, m.grid)
T_CLAY <- unlist(lapply(incells, function(x) if (sum(!is.na(x)) > 0) round(mean(x, na.rm=TRUE)) else NA))
incells <- extract(SOILchar$T_GRAVEL, m.grid)
T_GRAVEL <- unlist(lapply(incells, function(x) if (sum(!is.na(x)) > 0) round(mean(x, na.rm=TRUE)) else NA))
incells <- extract(SOILchar$T_OC, m.grid)
T_OC <- unlist(lapply(incells, function(x) if (sum(!is.na(x)) > 0) round(mean(x, na.rm=TRUE)) else NA))
incells <- extract(SOILchar$T_PH, m.grid)
T_PH <- unlist(lapply(incells, function(x) if (sum(!is.na(x)) > 0) round(mean(x, na.rm=TRUE)) else NA))
incells <- extract(SOILchar$T_REF_BULK_DENSITY, m.grid)
T_REF_BULK_DENSITY <- unlist(lapply(incells, function(x) if (sum(!is.na(x)) > 0) round(mean(x, na.rm=TRUE),2) else NA))
incells <- extract(SOILchar$T_SAND, m.grid)
T_SAND <- unlist(lapply(incells, function(x) if (sum(!is.na(x)) > 0) round(mean(x, na.rm=TRUE)) else NA))
incells <- extract(SOILchar$T_SILT, m.grid)
T_SILT <- unlist(lapply(incells, function(x) if (sum(!is.na(x)) > 0) round(mean(x, na.rm=TRUE)) else NA))
```

```
layout(matrix(1:9, nrow=3, byrow=TRUE))
values(grid.master) <- T_CAEX
plot(grid.master, legend.args=list(text='T_CAEX'))
 plot(catchments, add=TRUE)

values(grid.master) <- T_CLAY
plot(grid.master, legend.args=list(text='T_CLAY'))
 plot(catchments, add=TRUE)

values(grid.master) <- T_GRAVEL
plot(grid.master, legend.args=list(text='T_GRAVEL'))
 plot(catchments, add=TRUE)

values(grid.master) <- T_OC
plot(grid.master, legend.args=list(text='T_OC'))
 plot(catchments, add=TRUE)

values(grid.master) <- T_PH
plot(grid.master, legend.args=list(text='T_PH'))
 plot(catchments, add=TRUE)

values(grid.master) <- T_REF_BULK_DENSITY
plot(grid.master, legend.args=list(text='T_BD'))
 plot(catchments, add=TRUE)

values(grid.master) <- T_SAND
plot(grid.master, legend.args=list(text='T_SAND'))
 plot(catchments, add=TRUE)

values(grid.master) <- T_SILT
plot(grid.master, legend.args=list(text='T_SILT'))
 plot(catchments, add=TRUE)
```




```
cds_list <- list(CL_MAP, CL_MAT, CL_PET, CL_PETovP, CL_R50, CL_R95, CL_dRD2D,
                 MP_mean_dem, MP_cv_dem, MP_mean_slope, MP_mean_aspect,
                 LD_water, LD_developed, LD_barren, LD_forest, LD_shrubland,
                 LD_herbaceous, LD_planted, LD_wetlands,
                 SDEPTH, S_CAEX, S_CLAY, S_GRAVEL, S_OC, S_PH, S_REF_BULK_DENSITY, S_SAND, S_SILT,
                 T_CAEX, T_CLAY, T_GRAVEL, T_OC, T_PH, T_REF_BULK_DENSITY, T_SAND, T_SILT)

cds_names <- c("CL_MAP","CL_MAT","CL_PET", "CL_PETovP", "CL_R50", "CL_R95", "CL_dRD2D",
               "MP_mean_dem", "MP_cv_dem", "MP_mean_slope", "MP_mean_aspect",
               "LD_land_use_water", "LD_land_use_developed", "LD_land_use_barren", "LD_land_use_forest",
               "LD_land_use_shrubland", "LD_land_use_herbaceous",
               "LD_land_use_planted", "LD_land_use_wetlands",
               "SOIL_maximum_soil_depth", "SOIL_subsoil_cation_exchange", "SOIL_mean_subsoil_clay",
               "SOIL_mean_subsoil_gravel", "SOIL_mean_subsoil_organic_carbon", "SOIL_mean_subsoil_ph",
               "SOIL_mean_subsoil_bulk_density", "SOIL_mean_subsoil_sand",
               "SOIL_mean_subsoil_silt", "SOIL_topsoil_cation_exchange",
               "SOIL_mean_topsoil_clay", "SOIL_mean_topsoil_gravel",
               "SOIL_mean_topsoil_organic_carbon", "SOIL_mean_topsoil_ph",
```

```
              "SOIL_mean_topsoil_bulk_density", "SOIL_mean_topsoil_sand", "SOIL_mean_topsoil_silt")

cds <- vector("list", length(cds_list))
for (i in 1:length(cds_list)){
   cds[[i]] <- matrix(cds_list[[i]], length(lat.m), length(lon.m), byrow=TRUE)
}
 names(cds) <- cds_names
```

7 Save variables

Save the variables to be used with SALTO and PASS:

```
str(train.data)
```

```
## List of 4
##  $ prec: num [1:3652, 1:215] 0 0 0.2784 0 0.0504 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : chr [1:3652] "2000-10-01" "2000-10-02" "2000-10-03" "2000-10-04" ...
##   .. ..$ : chr [1:215] "87" "88" "89" "131" ...
##  $ temp: num [1:3652, 1:215] 12.41 14.14 13.03 12.29 8.11 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : chr [1:3652] "2000-10-01" "2000-10-02" "2000-10-03" "2000-10-04" ...
##   .. ..$ : chr [1:215] "87" "88" "89" "131" ...
##  $ pet : num [1:3652, 1:215] 1.43 2.24 1.7 2.04 2.05 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : chr [1:3652] "2000-10-01" "2000-10-02" "2000-10-03" "2000-10-04" ...
##   .. ..$ : chr [1:215] "87" "88" "89" "131" ...
##  $ qobs: num [1:3652, 1:25] 0.116 0.113 0.113 0.105 0.101 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : chr [1:3652] "2000-10-01" "2000-10-02" "2000-10-03" "2000-10-04" ...
##   .. ..$ : chr [1:25] "01013500" "01022500" "01030500" "01031500" ...
```

```
inputTopology <- list(catchments, dem, lon.m, lat.m, lon.dem, lat.dem)
 names(inputTopology) <- c("catchments", "dem", "lon.m", "lat.m", "lon.dem", "lat.dem")
#str(inputTopology)

inputCatchDescr <- list(cds, catchments, weights, lat.m, lon.m)
 names(inputCatchDescr) <- c("cds", "catchments", "weights", "lat.m", "lon.m")
#str(inputCatchDescr)
```

```
save(list=c('train.data','inputTopology','inputCatchDescr'),
     #file='inputTopologyCatchDescr47x50_traindata25cat215pxl.RData',
     file='outDataPrep.RData',
     compress='xz')  # 7.9M
```

8 Prepare input to pass

Prepare the catchment descriptor and topology inputs for PASS:

```
library(hydroPASS)

load('outDataPrep.RData')
# analogous to: data('outDataPrep')

CDS <- inputCatchDescr$cds
CATCHMENTS <- inputCatchDescr$catchments
```

```
WEIGHTS <- inputCatchDescr$weights
LON.M <- inputCatchDescr$lon.m
LAT.M <- inputCatchDescr$lat.m
train.CD_25cat215pxl <- catchDescrip(cds=CDS, catchment=CATCHMENTS, weights=WEIGHTS,
                                     lon.mast=LON.M, lat.mast=LAT.M)

CATCHMENTS <- inputTopology$catchments
DEM <- inputTopology$dem
LON.M <- inputTopology$lon.m
LAT.M <- inputTopology$lat.m
LON.DEM <- inputTopology$lon.dem
LAT.DEM <- inputTopology$lat.dem
# the following takes some time
train.topology_25cat215pxl <- topology(catchments=CATCHMENTS, dem=DEM, lon.mast=LON.M,
                                       lat.mast=LAT.M, lon.dem=LON.DEM, lat.dem=LAT.DEM)
```

Local calibration of SALTO:

```
library(hydroPASS)
require(DEoptim)

negME.SALTO <- function(param, prec, temp, pet, effarea, grdname, flowto, level, disc) {
  # global variables are in the function
  simu0 <- SALTO(prec=prec, temp=temp, pet=pet,
                 effarea=effarea, grdname=grdname, flowto=flowto, level=level, param=param)$qsim
  simu <- simu0[which.max(apply(simu0, 1, sum)), ]    # <- this is a quick fix that should be taken care of
  if (!all(is.na(simu))) {
    simu[is.na(simu)] <- -999
    simu <- simu[-c(1:303)]  # remove the warming period
    obse <- disc[-c(1:303)]  # remove the warming period
    r <- cor(simu, obse, method='pearson', use='pairwise.complete.obs')
    beta <- mean(simu)/mean(obse)
    gamma <- (sd(simu)/mean(simu))/(sd(obse)/mean(obse))
    kgeQ <- 1 - sqrt((r - 1)^2 + (beta - 1)^2 + (gamma - 1)^2)
  } else kgeQ <- -999
  #if (is.na(kgeQ)) kgeQ <- -999
  return(-kgeQ)  # negative Kling-Gupta Efficiency
}

load('outDataPrep.RData')
# analogous to: data('outDataPrep')

qualicodici <- colnames(train.data$qobs)

parametriLOCcalibrati <- array(NA, dim=c(25, 22, 30),
                               dimnames=list(qualicodici,
                                             c('ME', 'TS','TS_R','BETA_SNOW',
                                               'DDF_NR','DDF_R','DDF_INC',
                                               'TM','SM_MIN','SM_MAX',
                                               'BETA_RC','BETA_AET','K_RS','BETA_RS',
                                               'PERCMAX','BETA_PERC',
                                               'K_GW','BETA_GW','GW_S_MAX',
                                               'BETA_DQ_GW','K_RIVER','BETA_RIVER'),
                                             paste('run', 1:30, sep=''))))
for (jjj in 1:30) {
  for (iii in 1:length(qualicodici)) {
    codice <- qualicodici[iii]
    print(paste(iii, '-', codice, 'run', jjj))

    whichcatchment <- which(colnames(train.data$qobs) == codice)
    whichpixels <- which(match(colnames(train.data$prec),
                               as.character(train.topology_25cat215pxl[[whichcatchment]]$grd),
                               nomatch=0) > 0)
    PREC <- train.data$prec[, whichpixels]
    TEMP <- train.data$temp[, whichpixels]
    PET <- train.data$pet[, whichpixels]
    EFFAREA <- train.topology_25cat215pxl[[whichcatchment]]$effarea
    GRDNAME <- train.topology_25cat215pxl[[whichcatchment]]$grd
    FLOWTO <- train.topology_25cat215pxl[[whichcatchment]]$flowto
    LEVEL <- train.topology_25cat215pxl[[whichcatchment]]$level
    QOBS <- train.data$qobs[, whichcatchment]
```

```
    TIME <- as.Date(rownames(train.data$qobs))

    # Calibration
    timestamp()
    calibr <- try(DEoptim(fn=negME.SALTO,
                          lower=c(TS=-2.0, TS_R=0, BETA_SNOW=1,
                                  DDF_NR=0.1, DDF_R=0.1, DDF_INC=-99,
                                  TM=-3.0, SM_MIN_1=0, SM_MAX_1=10,
                                  BETA_RC_1=0.1, BETA_AET_1=0.1,
                                  K_RS_1=0.1, BETA_RS_1=0.1,
                                  PERCMAX_1=0.0, BETA_PERC_1=0.1,
                                  K_GW=30, BETA_GW=0.1, GW_S_MAX=100, BETA_DQ_GW=0.1,
                                  K_RIVER=0.1, BETA_RIVER=0.1),
                          upper=c(TS=2.0, TS_R=0, BETA_SNOW=1,
                                  DDF_NR=10, DDF_R=10, DDF_INC=-99,
                                  TM=3.0, SM_MIN_1=100, SM_MAX_1=1000,
                                  BETA_RC_1=20, BETA_AET_1=5.0,
                                  K_RS_1=50, BETA_RS_1=5.0,
                                  PERCMAX_1=10, BETA_PERC_1=5.0,
                                  K_GW=400, BETA_GW=5.0, GW_S_MAX=15000, BETA_DQ_GW=5.0,
                                  K_RIVER=5.0, BETA_RIVER=30),
                          control=DEoptim.control(NP=NA, itermax=200, reltol=1e-3, steptol=10, trace=5,
                                                  parallelType=1, limitCores=32, packages=c('hydroPASS')),
                          prec=PREC, temp=TEMP, pet=PET,
                          effarea=EFFAREA, grdname=GRDNAME, flowto=FLOWTO, level=LEVEL,
                          disc=QOBS))
    if (class(calibr) != 'try-error') {
      parametriLOCcalibrati[as.character(codice), 1, jjj] <- -calibr$optim$bestval
      parametriLOCcalibrati[as.character(codice), -1, jjj] <- calibr$optim$bestmem
    }
    timestamp()
  }
}

train.parameters_25cat <- vector('list', dim(parametriLOCcalibrati)[1])
 names(train.parameters_25cat) <- dimnames(parametriLOCcalibrati)[[1]]
for (i in 1:length(train.parameters_25cat)) {
  train.parameters_25cat[[i]] <- t(parametriLOCcalibrati[i,,])
}

save(train.parameters_25cat, file='outTrainParameters.RData', compress='xz')
```

Appendix D

Locally vs. regionally calibrated lumped parameters

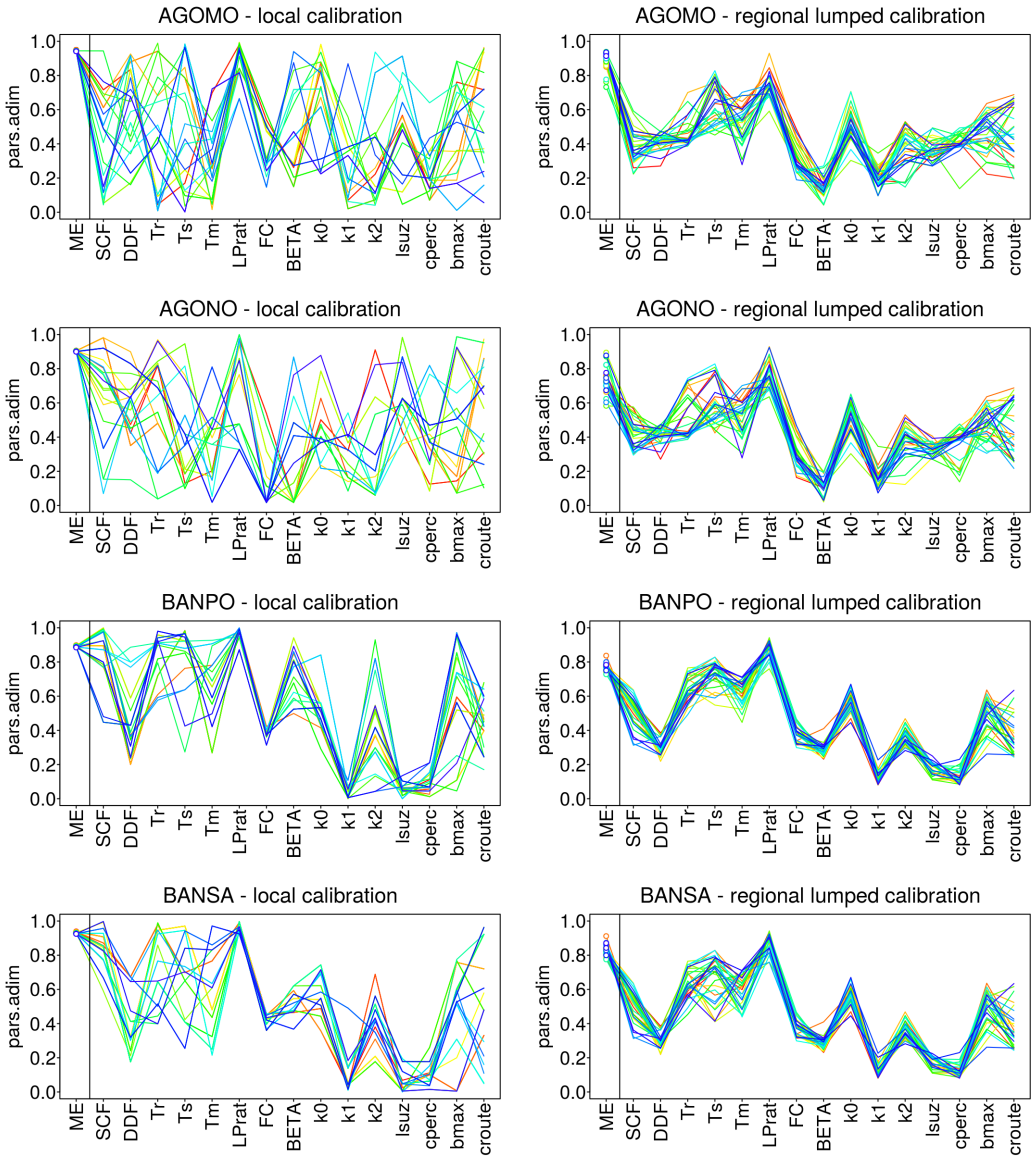


Fig. D.1 Locally calibrated parameters (30 sets, left) vs. regionally calibrated parameters (30 sets, right) with PASS for four sites.

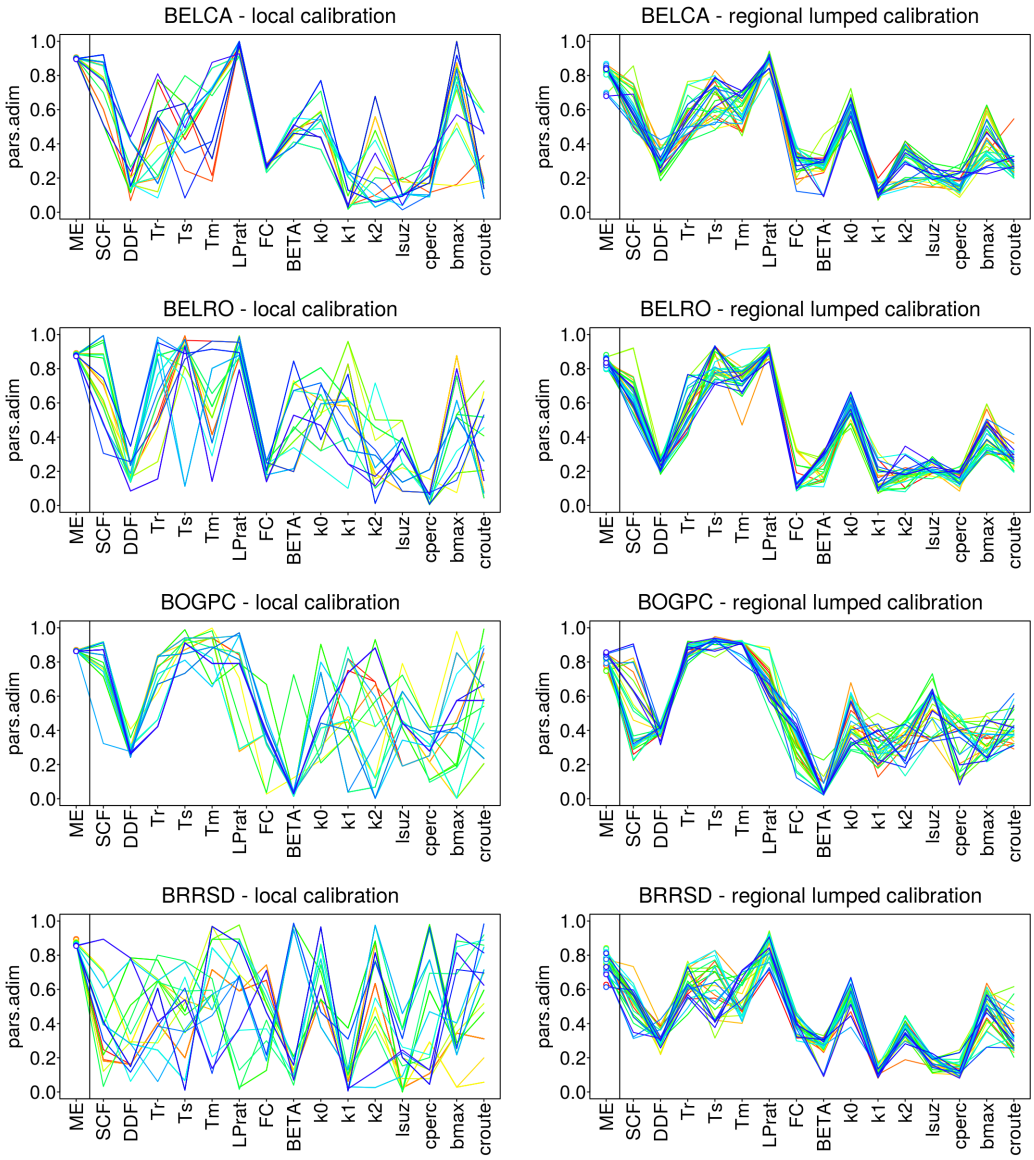


Fig. D.2 Locally calibrated parameters (30 sets, left) vs. regionally calibrated parameters (30 sets, right) with PASS for four sites.

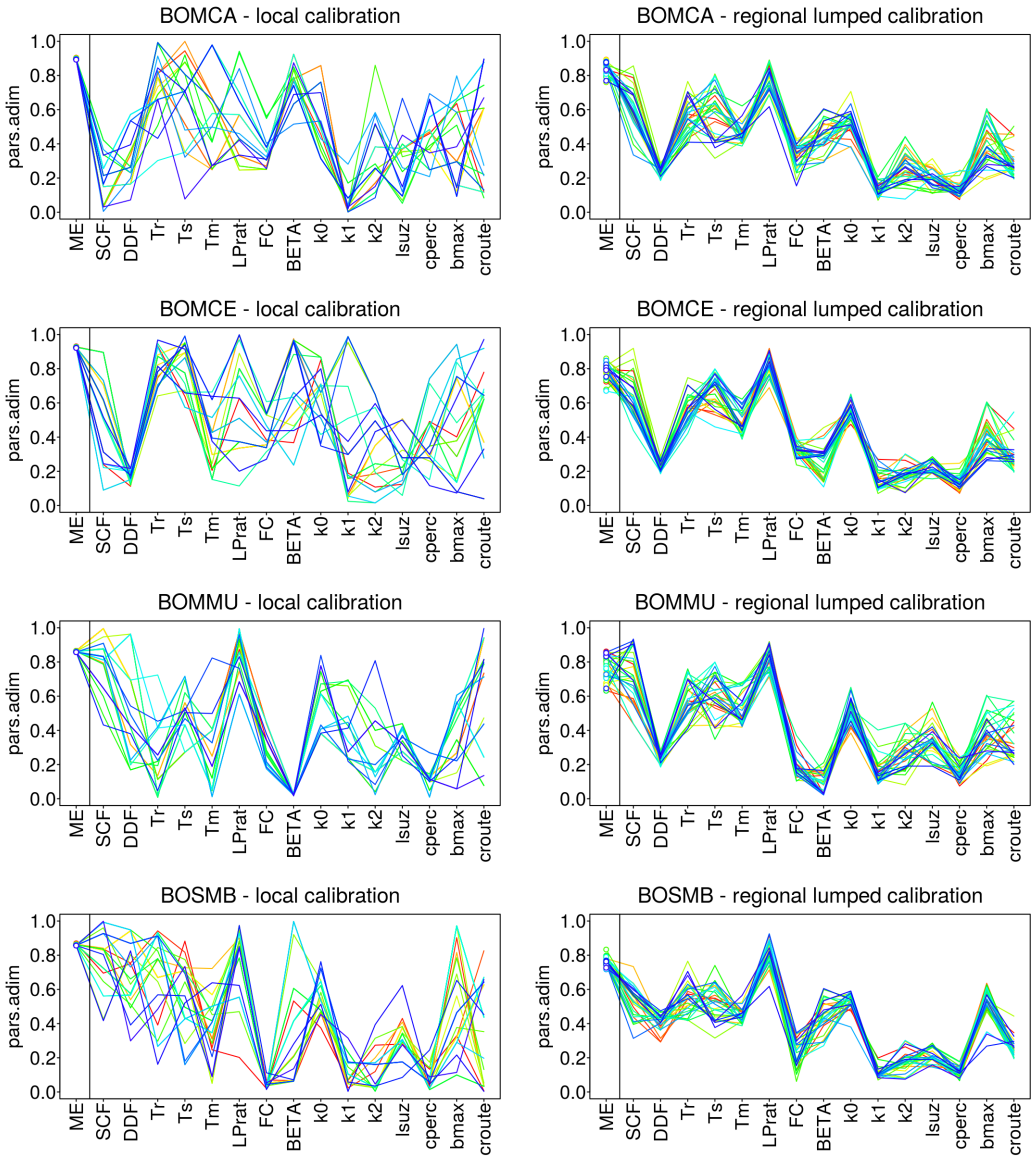


Fig. D.3 Locally calibrated parameters (30 sets, left) vs. regionally calibrated parameters (30 sets, right) with PASS for four sites.

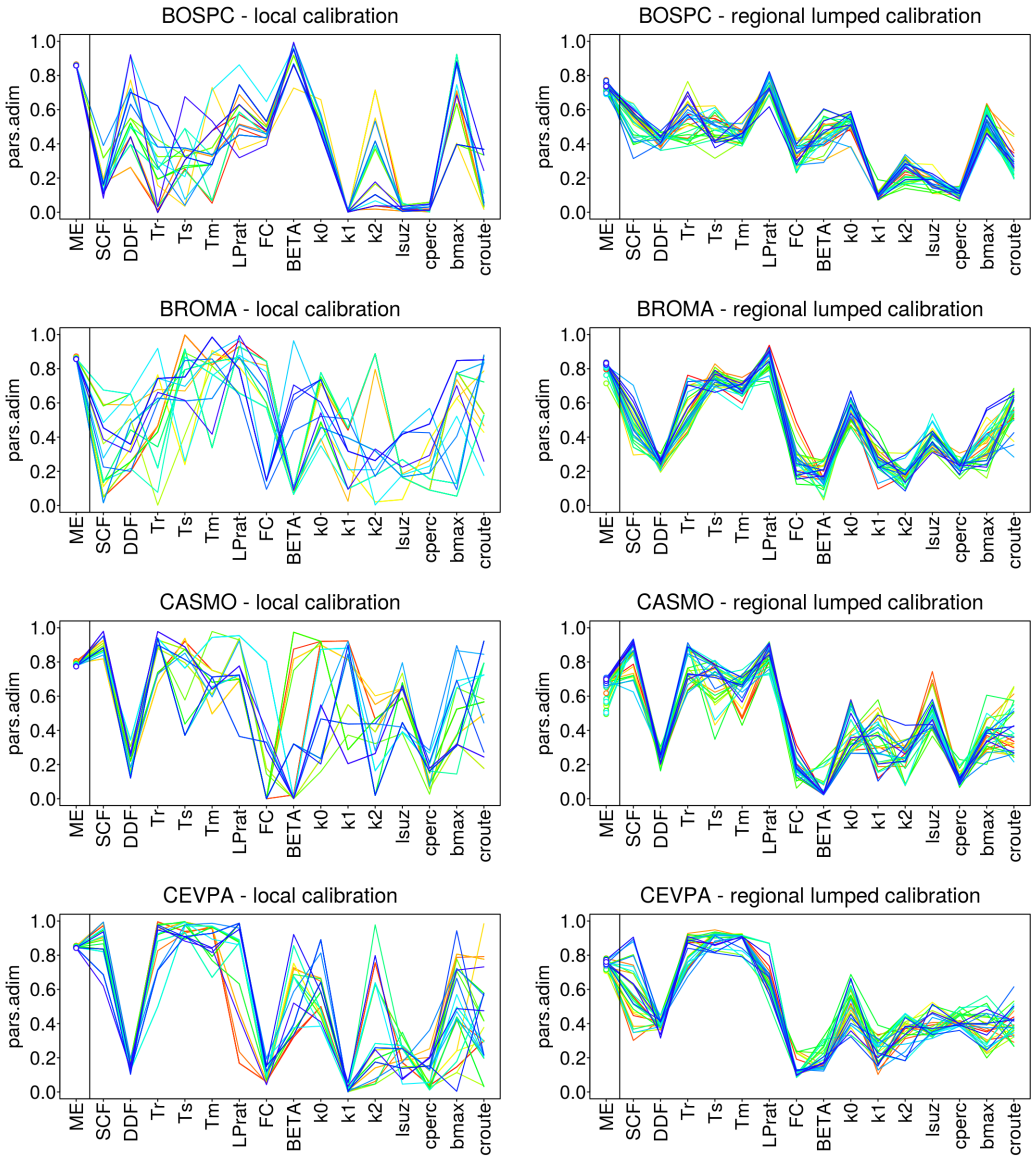


Fig. D.4 Locally calibrated parameters (30 sets, left) vs. regionally calibrated parameters (30 sets, right) with PASS for four sites.

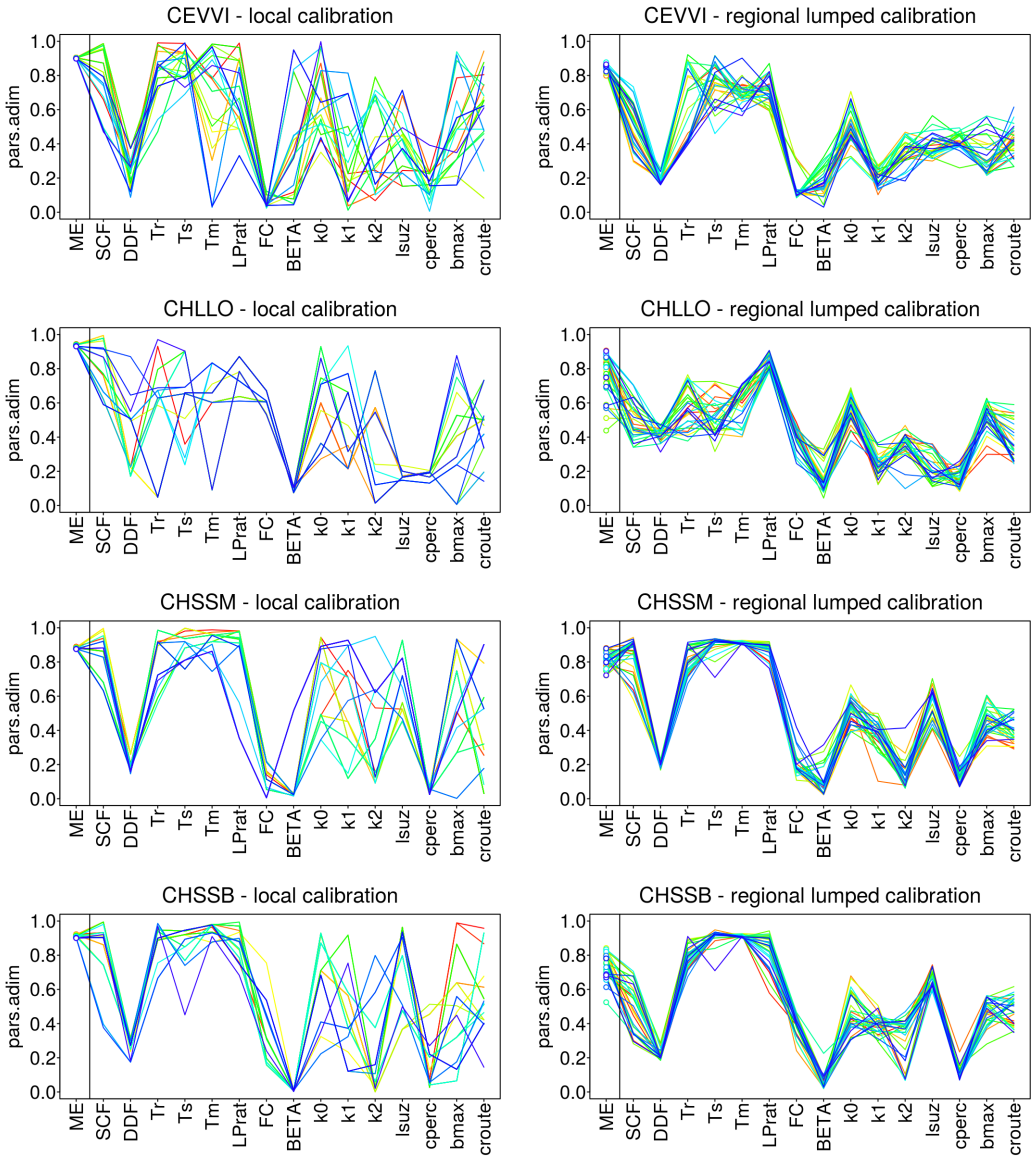


Fig. D.5 Locally calibrated parameters (30 sets, left) vs. regionally calibrated parameters (30 sets, right) with PASS for four sites.

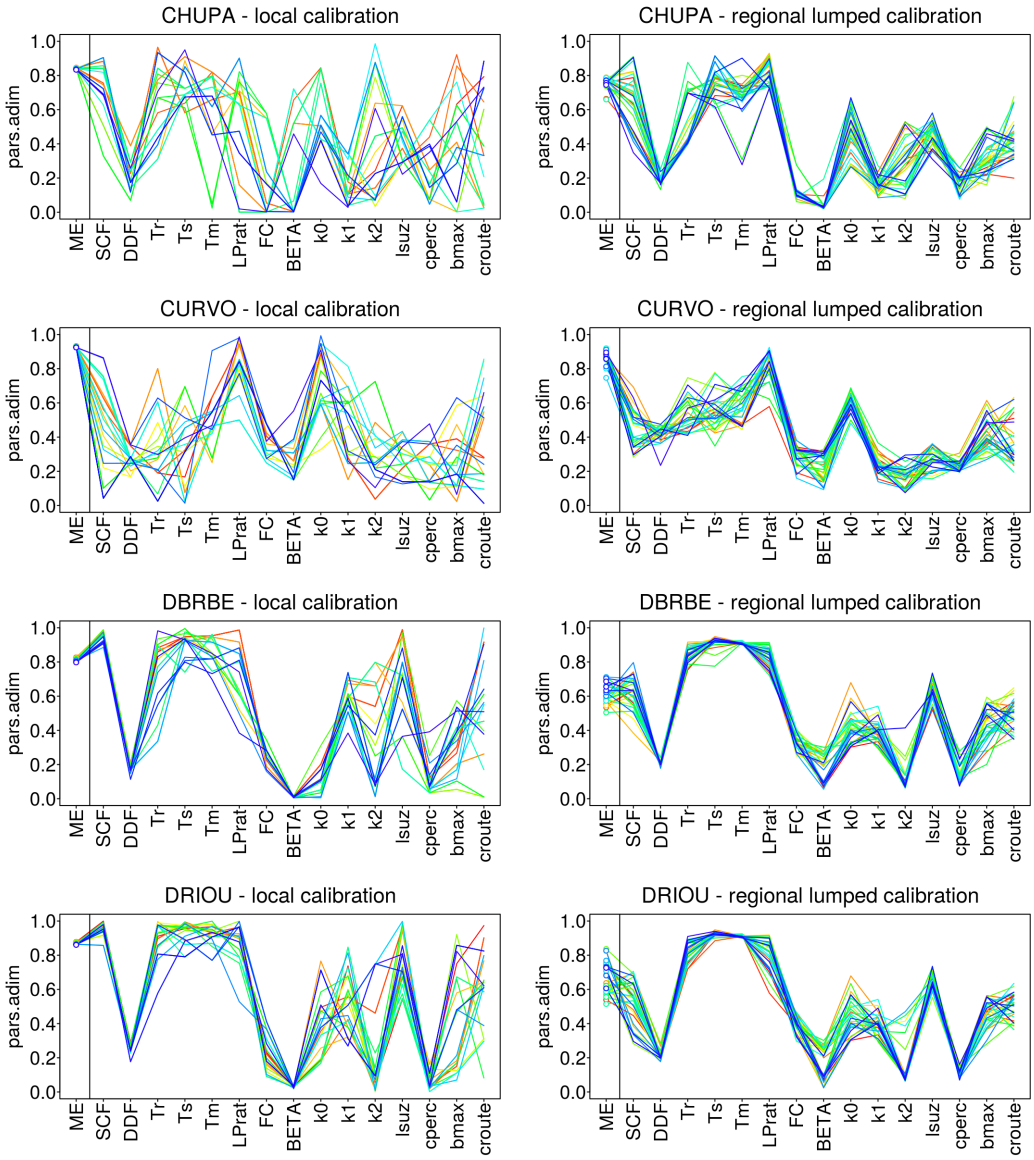


Fig. D.6 Locally calibrated parameters (30 sets, left) vs. regionally calibrated parameters (30 sets, right) with PASS for four sites.

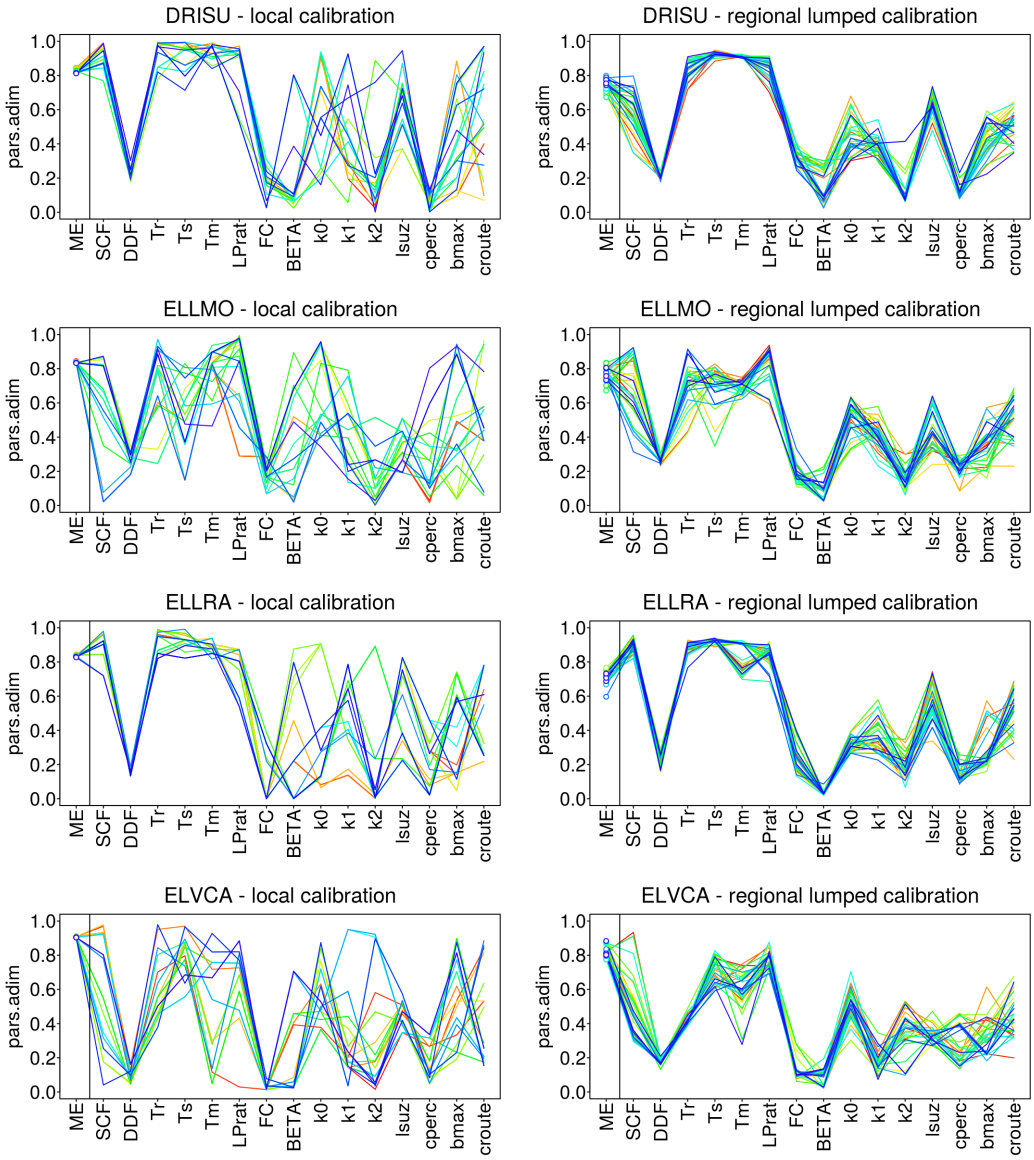


Fig. D.7 Locally calibrated parameters (30 sets, left) vs. regionally calibrated parameters (30 sets, right) with PASS for four sites.

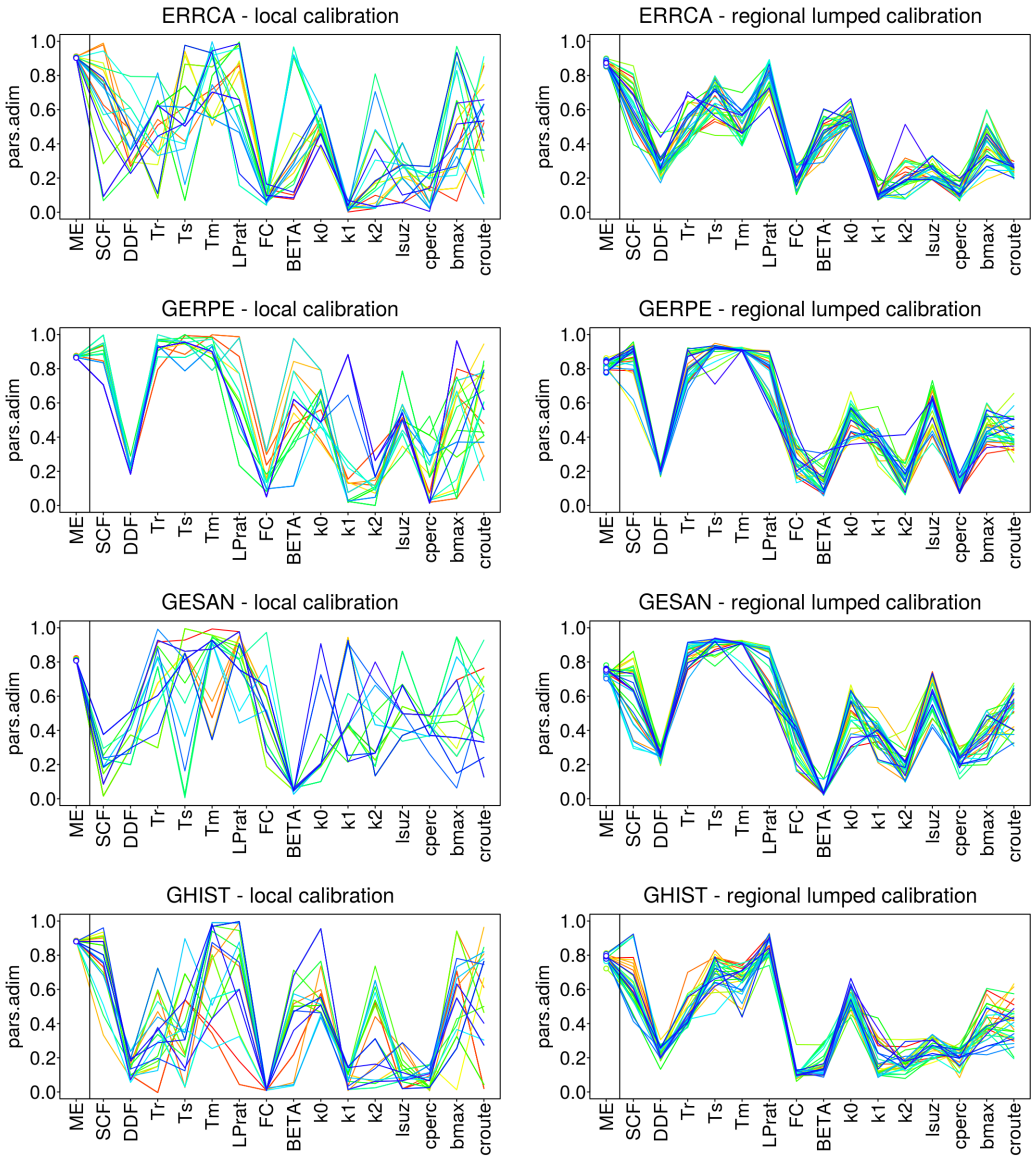


Fig. D.8 Locally calibrated parameters (30 sets, left) vs. regionally calibrated parameters (30 sets, right) with PASS for four sites.

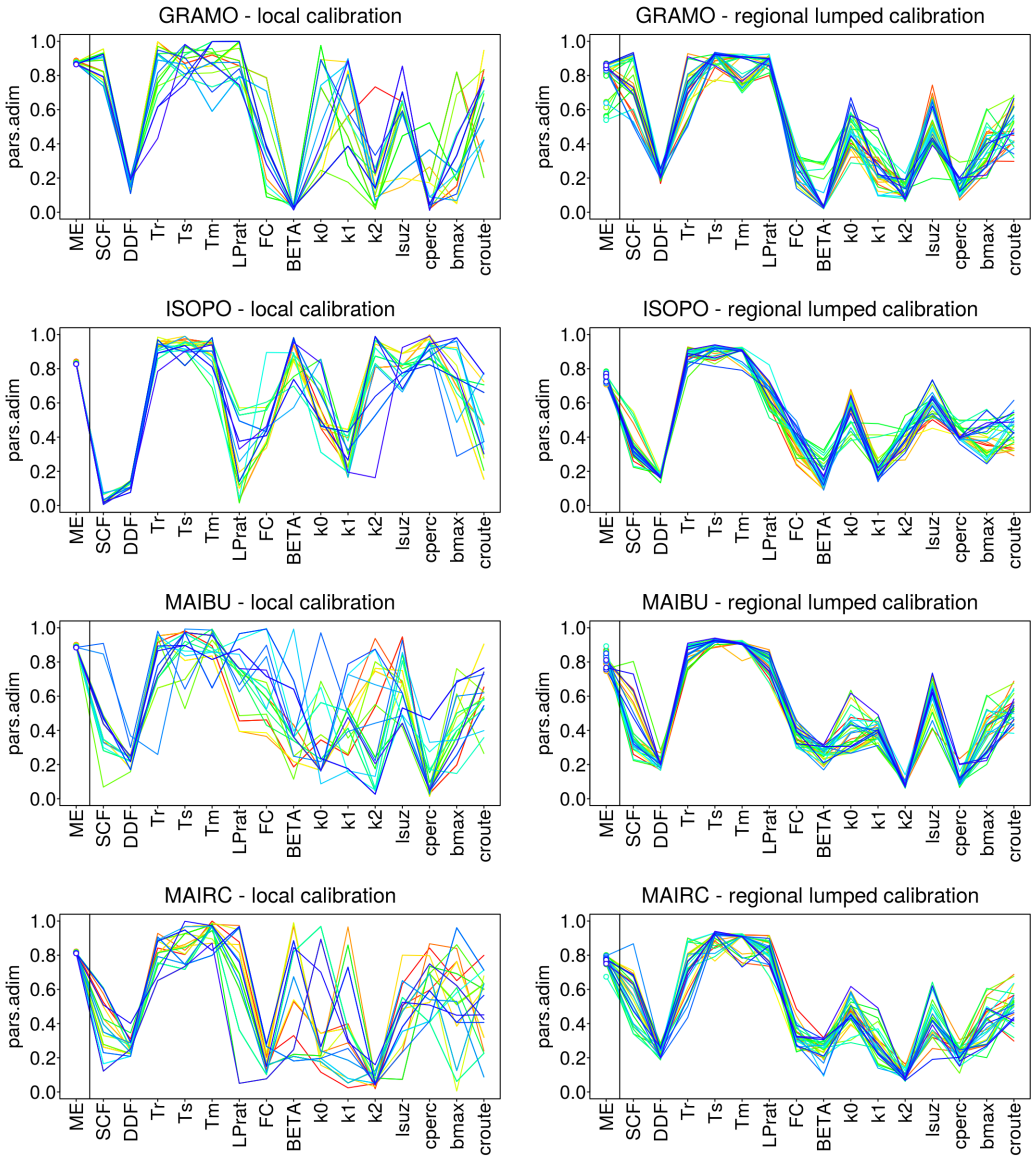


Fig. D.9 Locally calibrated parameters (30 sets, left) vs. regionally calibrated parameters (30 sets, right) with PASS for four sites.

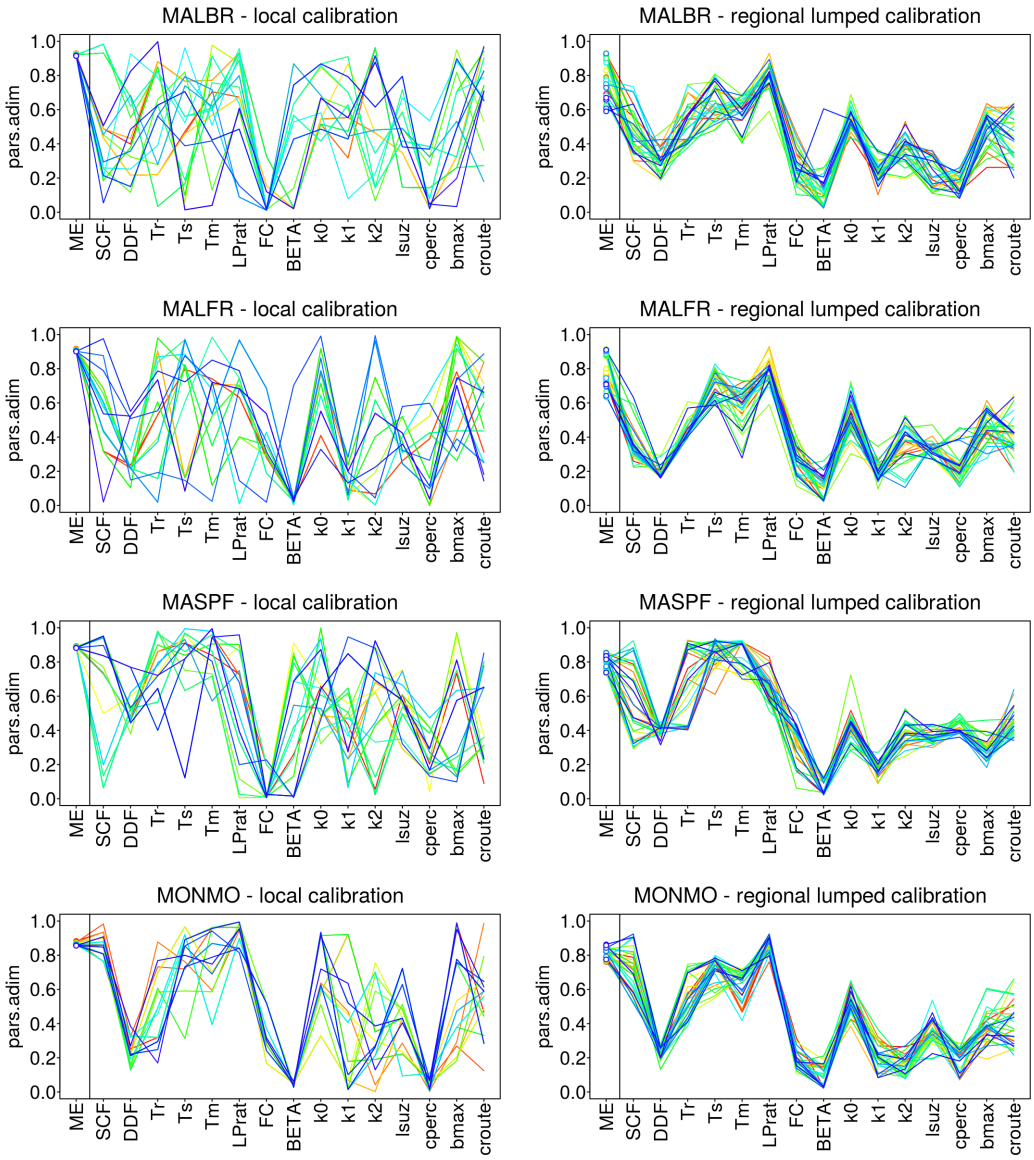


Fig. D.10 Locally calibrated parameters (30 sets, left) vs. regionally calibrated parameters (30 sets, right) with PASS for four sites.

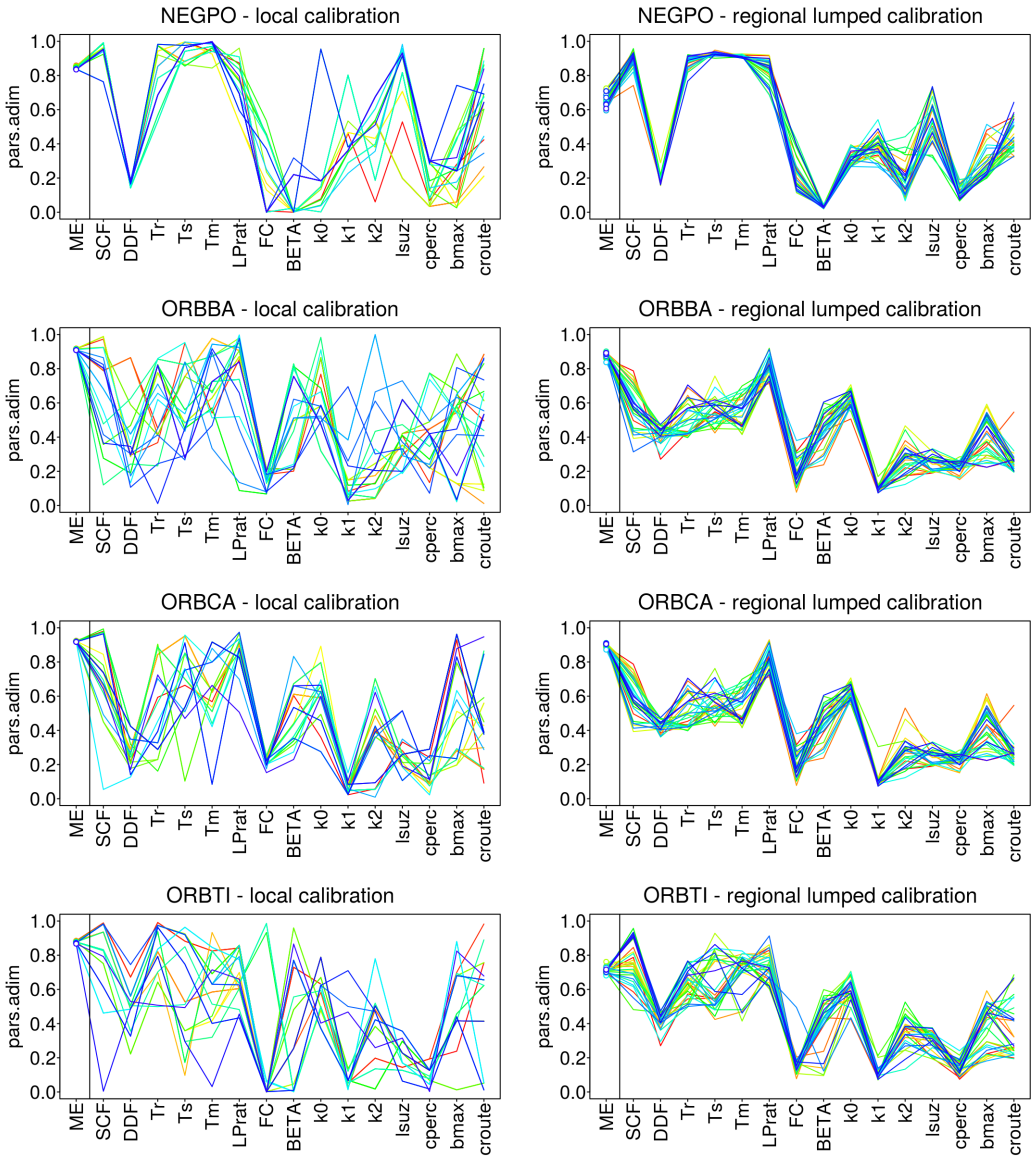


Fig. D.11 Locally calibrated parameters (30 sets, left) vs. regionally calibrated parameters (30 sets, right) with PASS for four sites.

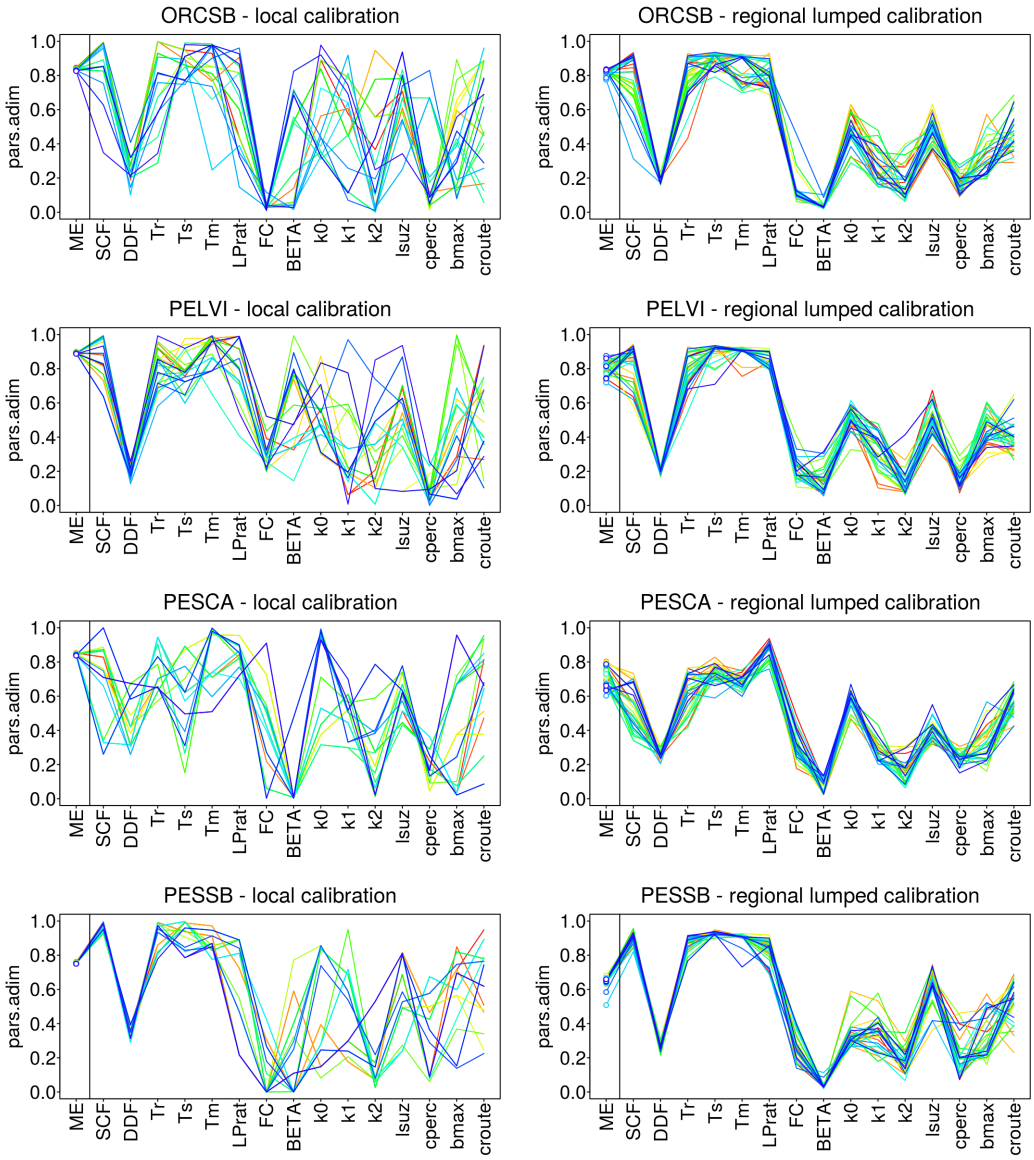


Fig. D.12 Locally calibrated parameters (30 sets, left) vs. regionally calibrated parameters (30 sets, right) with PASS for four sites.

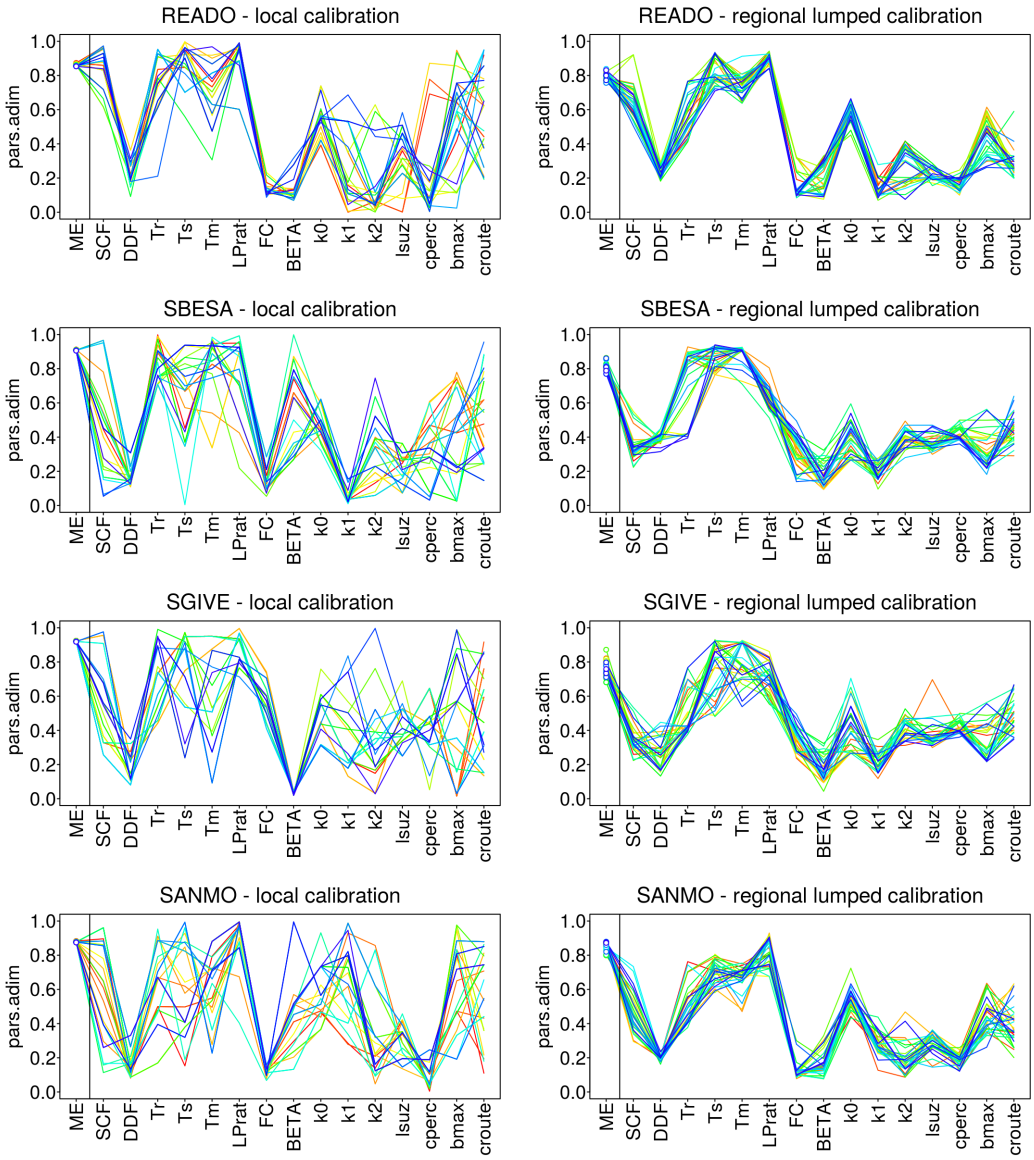


Fig. D.13 Locally calibrated parameters (30 sets, left) vs. regionally calibrated parameters (30 sets, right) with PASS for four sites.

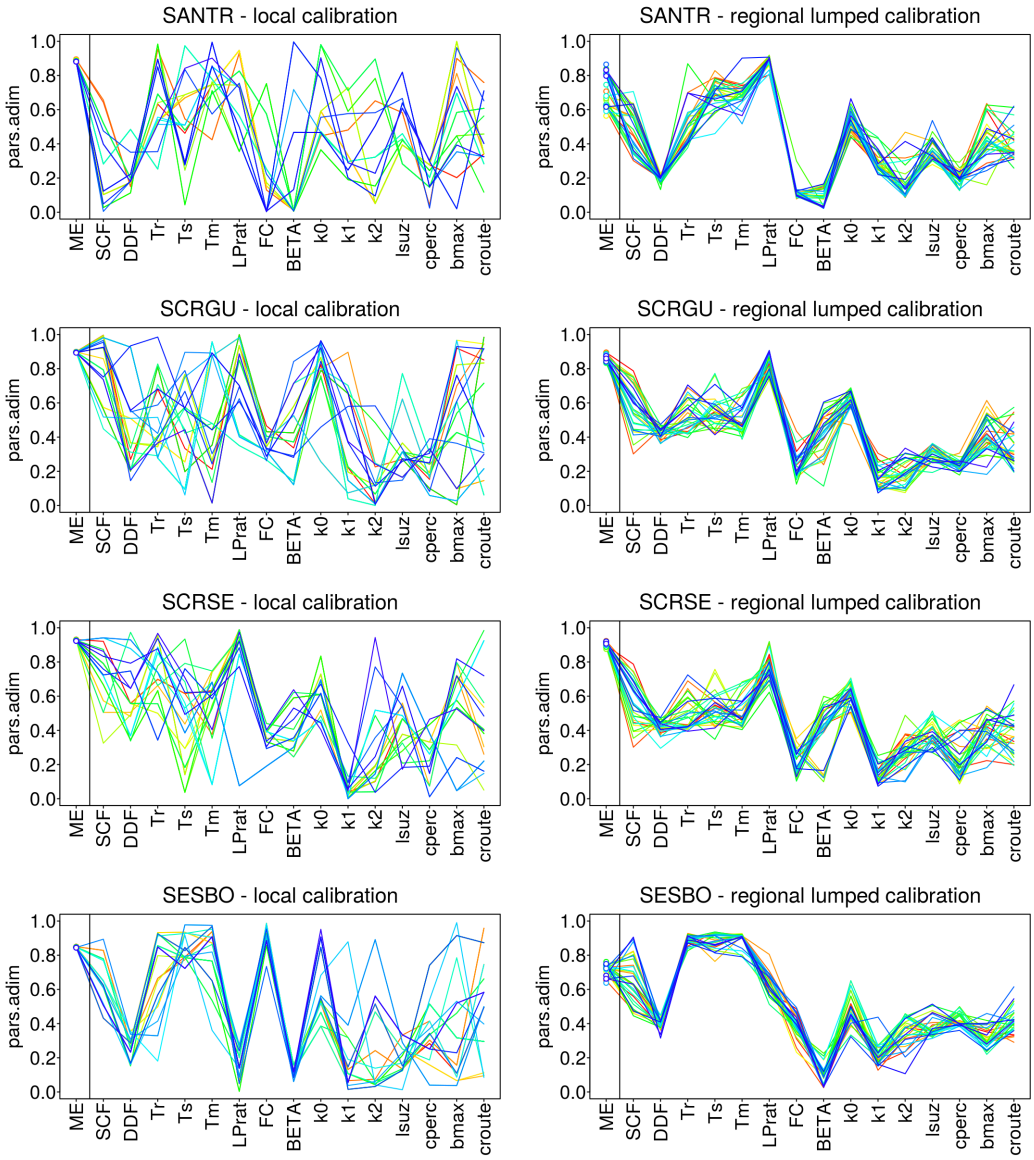


Fig. D.14 Locally calibrated parameters (30 sets, left) vs. regionally calibrated parameters (30 sets, right) with PASS for four sites.

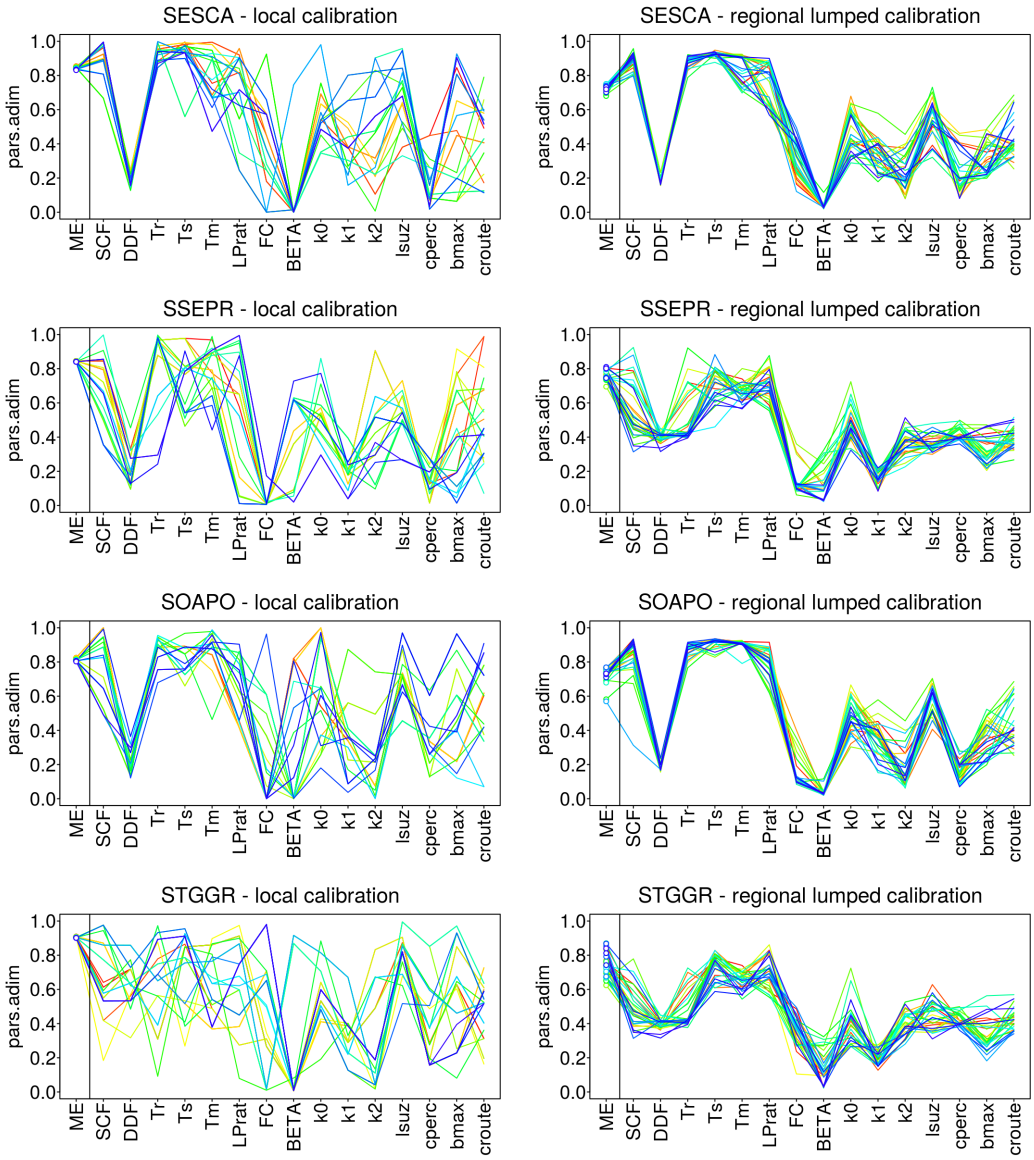


Fig. D.15 Locally calibrated parameters (30 sets, left) vs. regionally calibrated parameters (30 sets, right) with PASS for four sites.

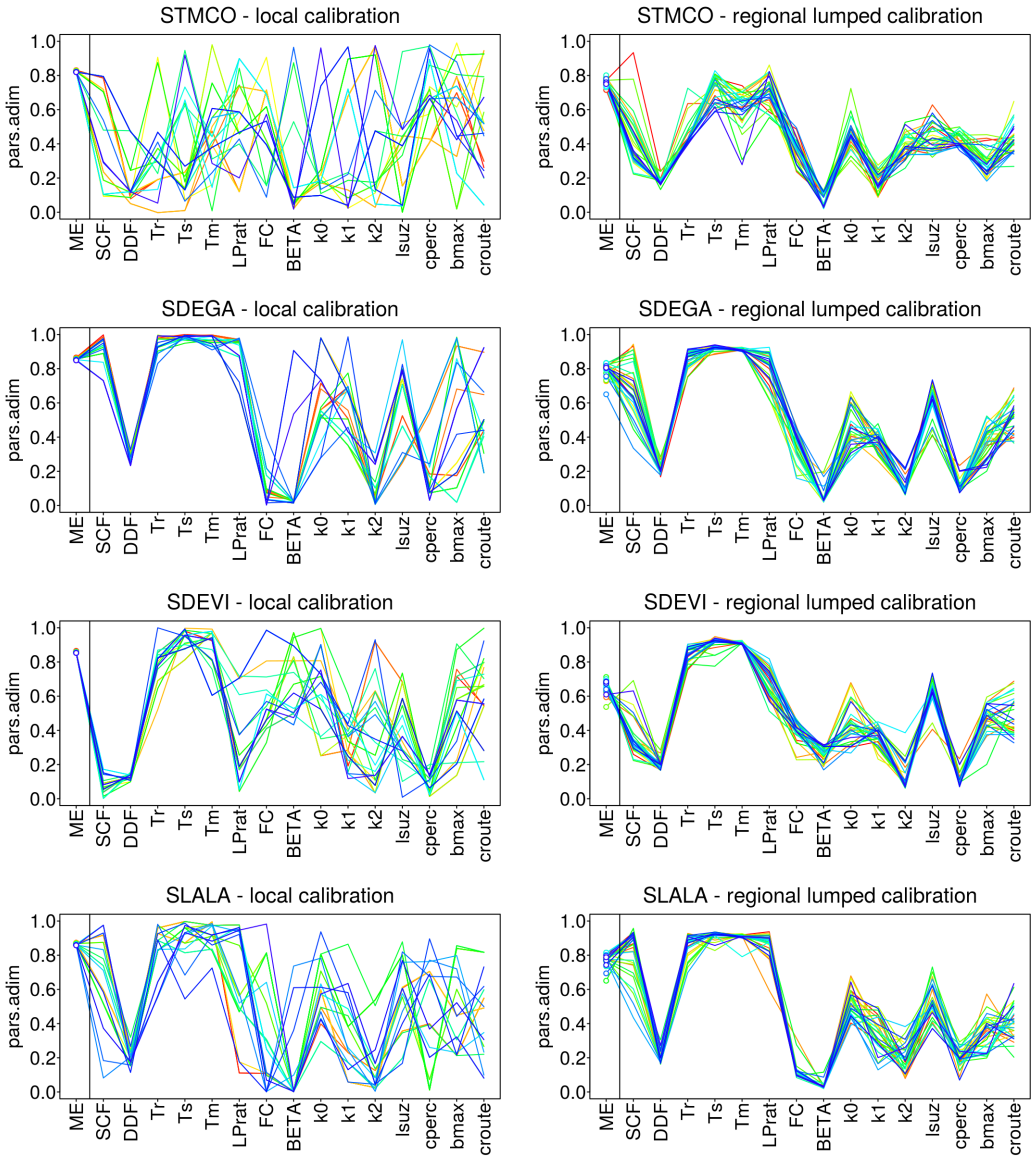


Fig. D.16 Locally calibrated parameters (30 sets, left) vs. regionally calibrated parameters (30 sets, right) with PASS for four sites.

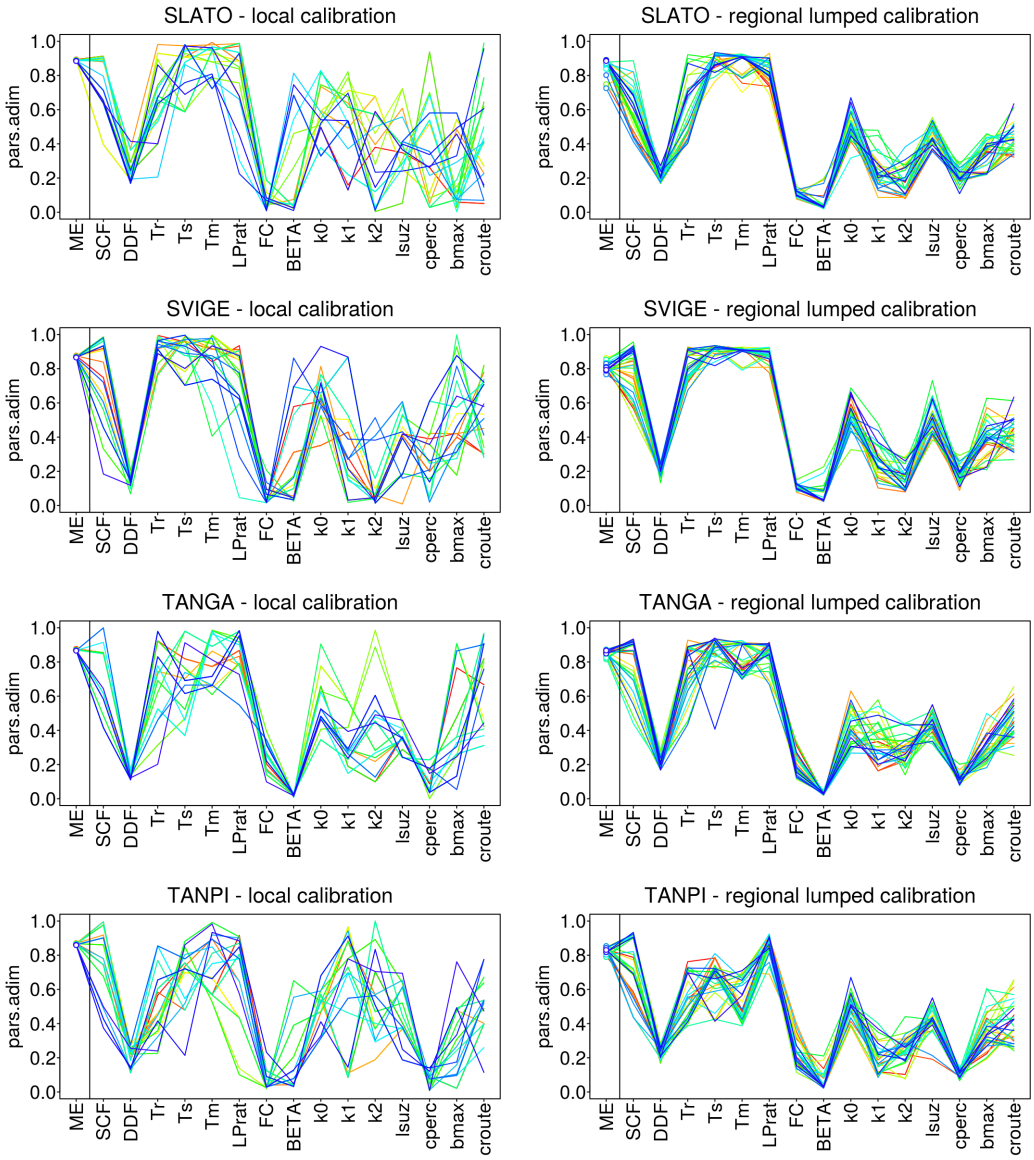


Fig. D.17 Locally calibrated parameters (30 sets, left) vs. regionally calibrated parameters (30 sets, right) with PASS for four sites.

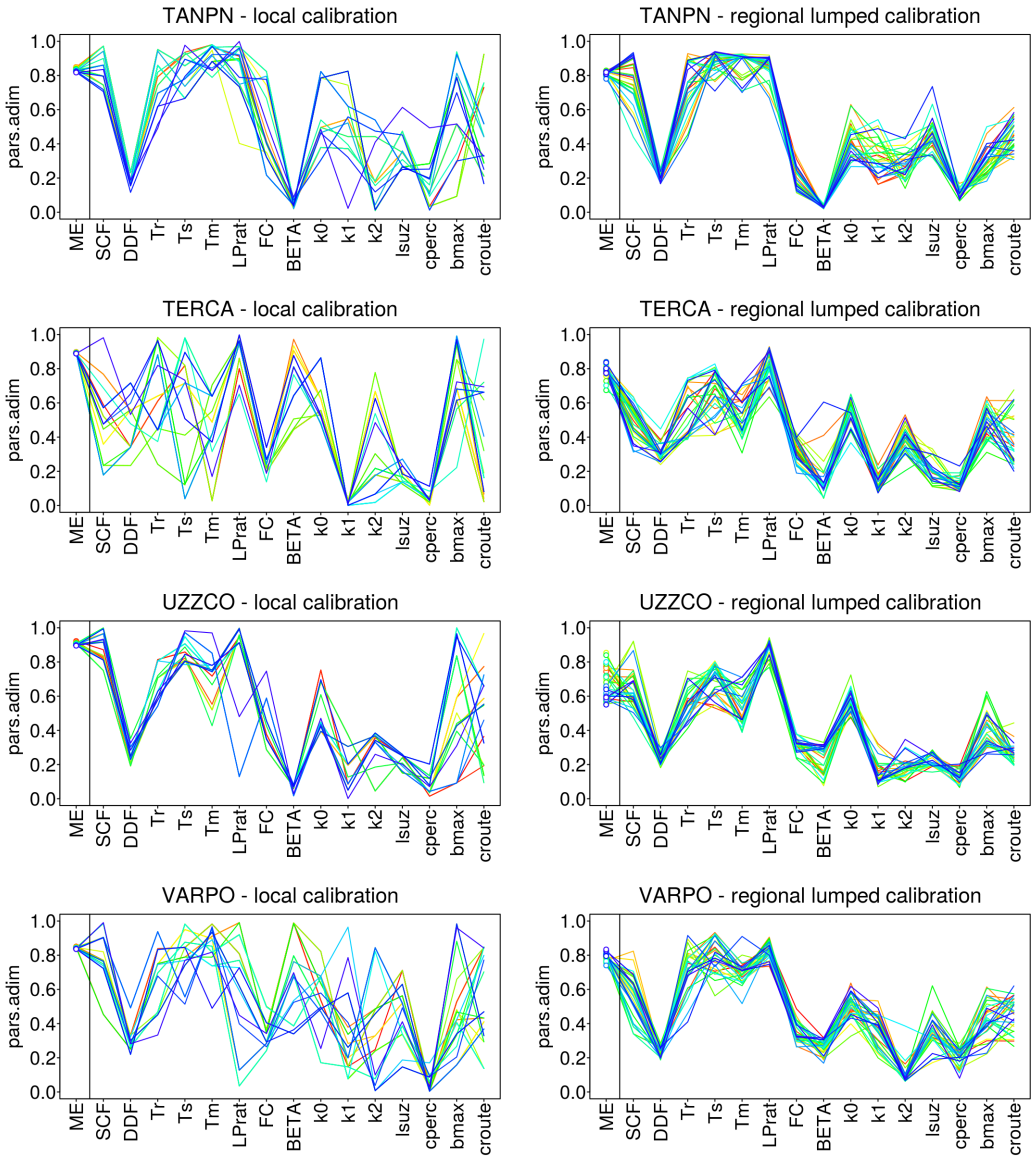


Fig. D.18 Locally calibrated parameters (30 sets, left) vs. regionally calibrated parameters (30 sets, right) with PASS for four sites.

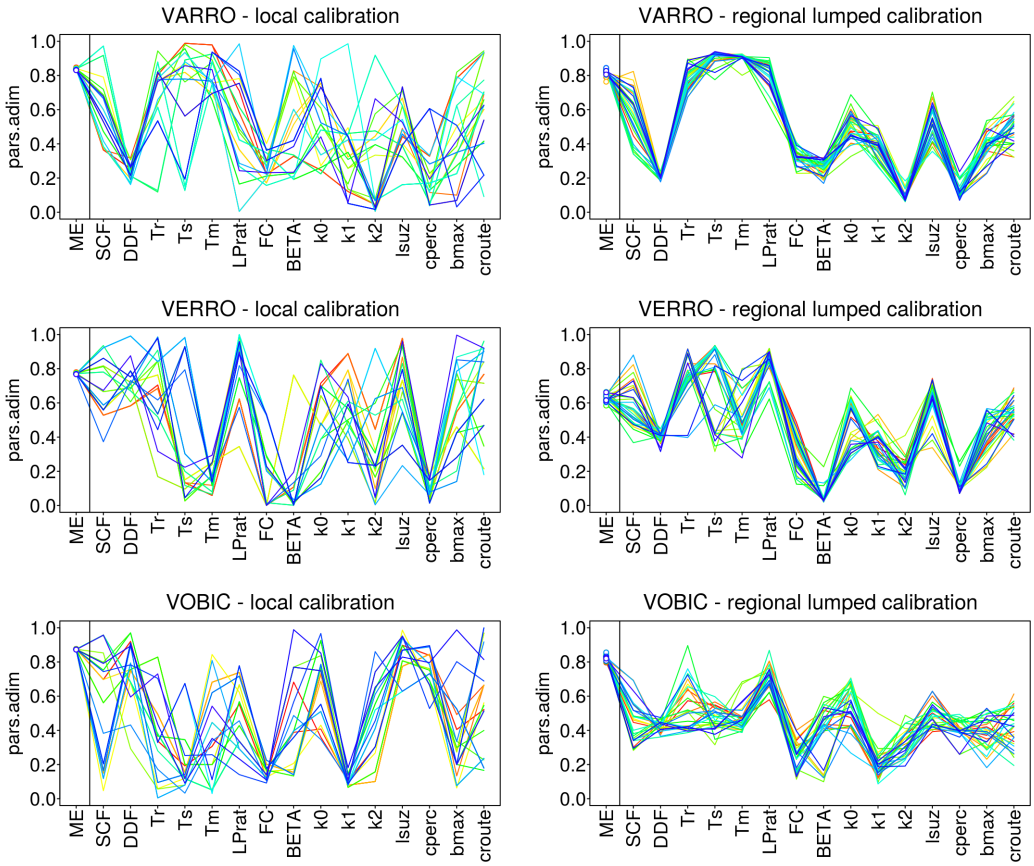


Fig. D.19 Locally calibrated parameters (30 sets, left) vs. regionally calibrated parameters (30 sets, right) with PASS for four sites.

Appendix E

Observed vs. Simulated discharges with regional PASS parameters

Legend for all the figures

- 1st row: simulated (red) vs. observed (black) daily discharges for the full simulation period (red shading indicates the range of the simulated daily discharges with the parameter sets obtained with PASS);
- 2nd row, 1st column: simulated (red) vs. observed (black) daily discharges for a couple of years (red shading indicates the range of the simulated daily discharges with the parameter sets obtained with PASS);
- 2nd row, 2nd column: simulated (red) vs. observed (black) daily discharges in logarithmic scale for a couple of years (red shading indicates the range of the simulated daily discharges with the parameter sets obtained with PASS);
- 3rd row, 1st column: annual components of the water balance: simulated runoff (red), observed runoff (black), observed precipitation (blue), observed potential evapotranspiration (dashed orange line), and simulated actual evapotranspiration (orange) (red shading indicates the range of the simulated annual runoff with the parameter sets obtained with PASS);
- 3rd row, 2nd column: seasonality of the water balance: simulated runoff (red), simulated runoff in the years when runoff was observed (dashed red line), observed runoff (black), observed precipitation (blue), observed potential evapotranspiration (dashed orange line), simulated actual evapotranspiration

(orange), simulated snow water equivalent (grey) (red shading indicates the range of the simulated seasonal runoff with the parameter sets obtained with PASS);

- 4th row, 1st column: simulated annual flow duration curve (mean = thick red, confidence bounds = thin red), observed annual flow duration curve (mean = thick black, confidence bounds = thin black) (red shading indicates the range of the simulated mean annual flow duration curve with the parameter sets obtained with PASS);
- 4th row, 2nd column: simulated maximum annual daily discharges (red bars), observed maximum annual daily discharges (black dots) (red shading indicates the range of the simulated maximum annual daily discharges with the parameter sets obtained with PASS);
- 5th row, 1st column: simulated empirical flood frequency curve (red), simulated empirical flood frequency curve in the years when runoff was observed (dashed red line), observed empirical flood frequency curve (black) (red shading indicates the range of the simulated empirical flood frequency curve with the parameter sets obtained with PASS);
- 5th row, 2nd column: fully distributed model efficiency $ME = 0.5 \cdot KGE + 0.5 \cdot SC$, and mean pixel parameter sets obtained with PASS (i.e., not to be confused with the lumped parameter sets of Appendix D).

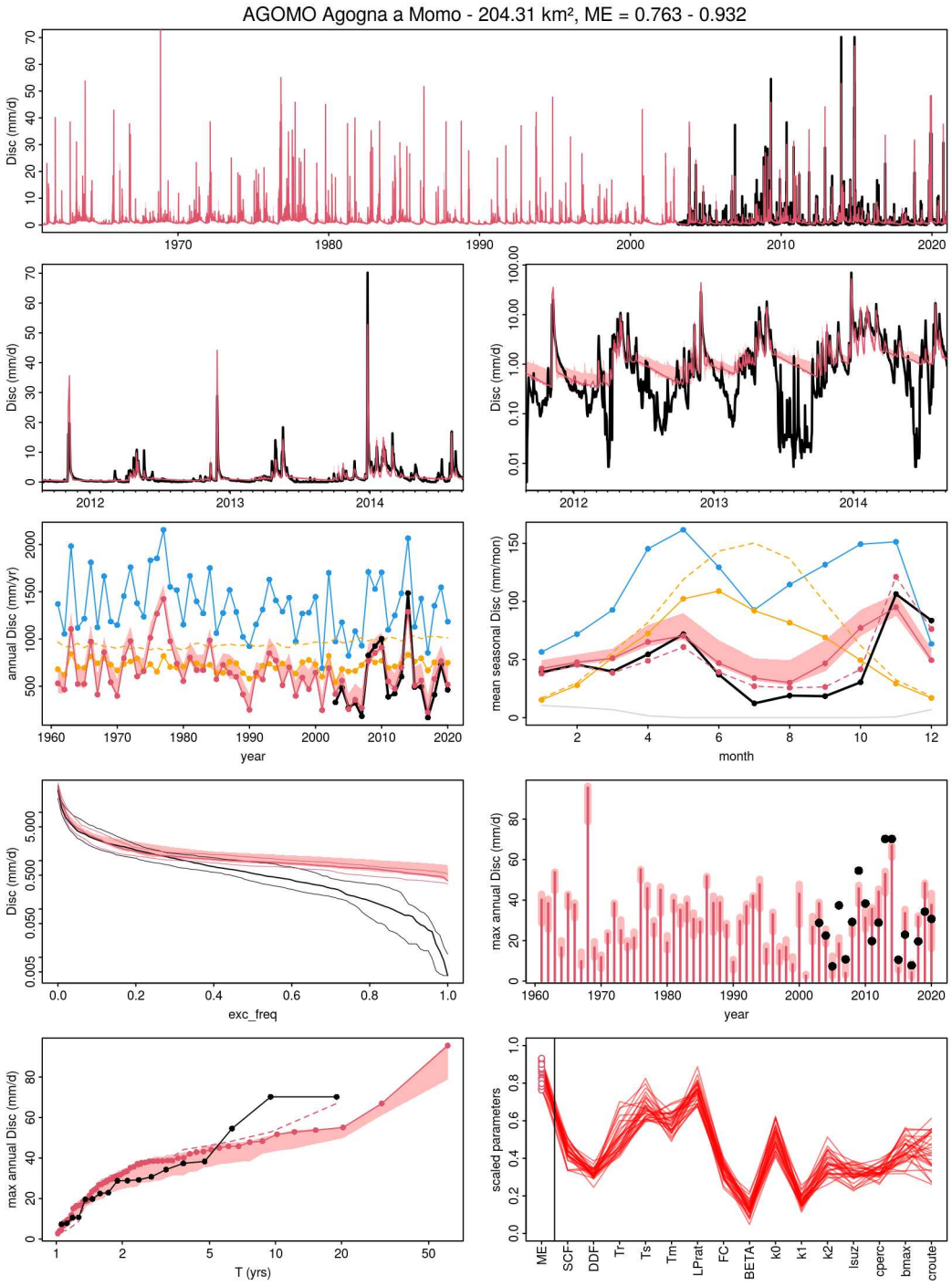


Fig. E.1 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 001.

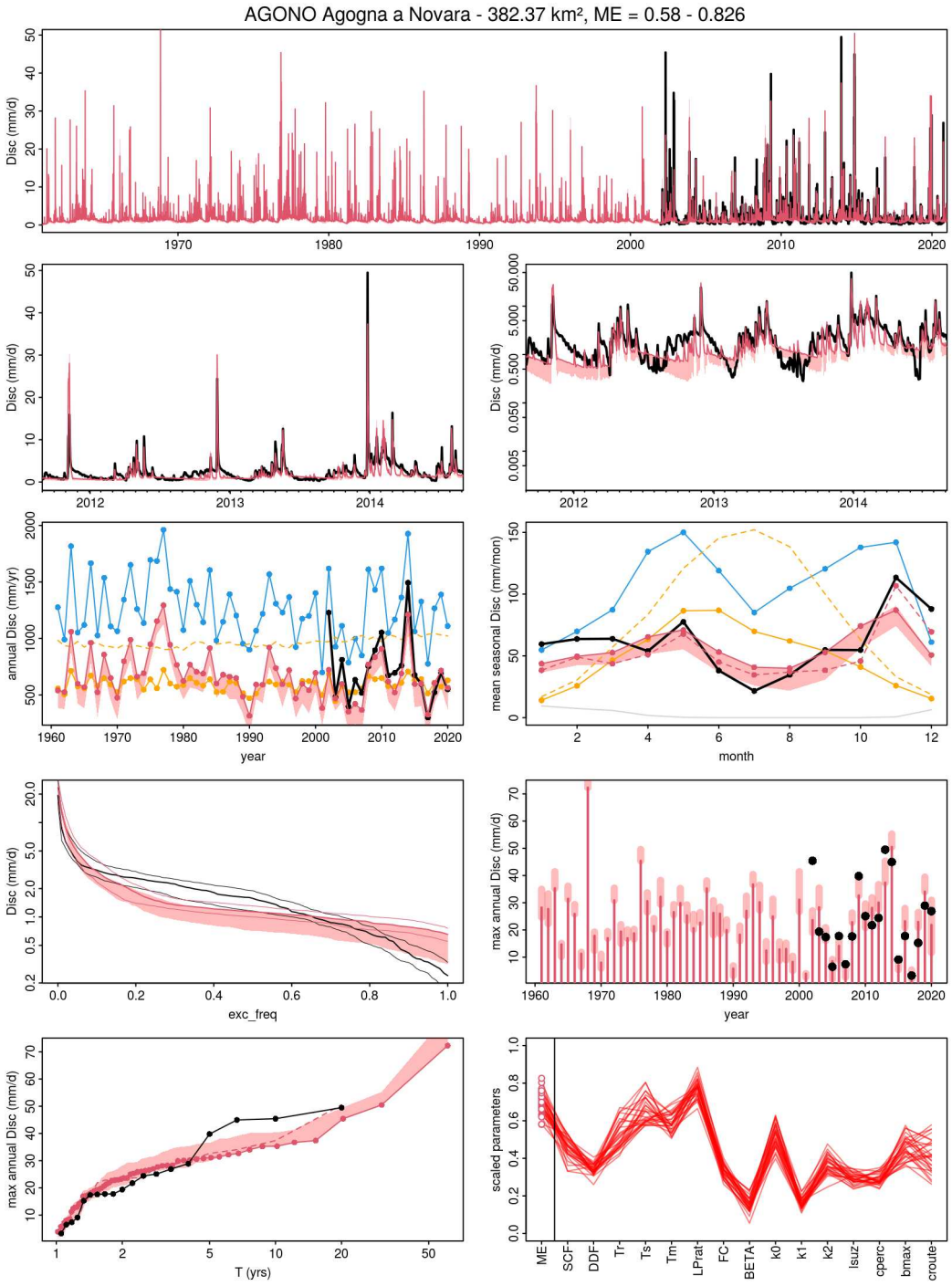


Fig. E.2 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 002.

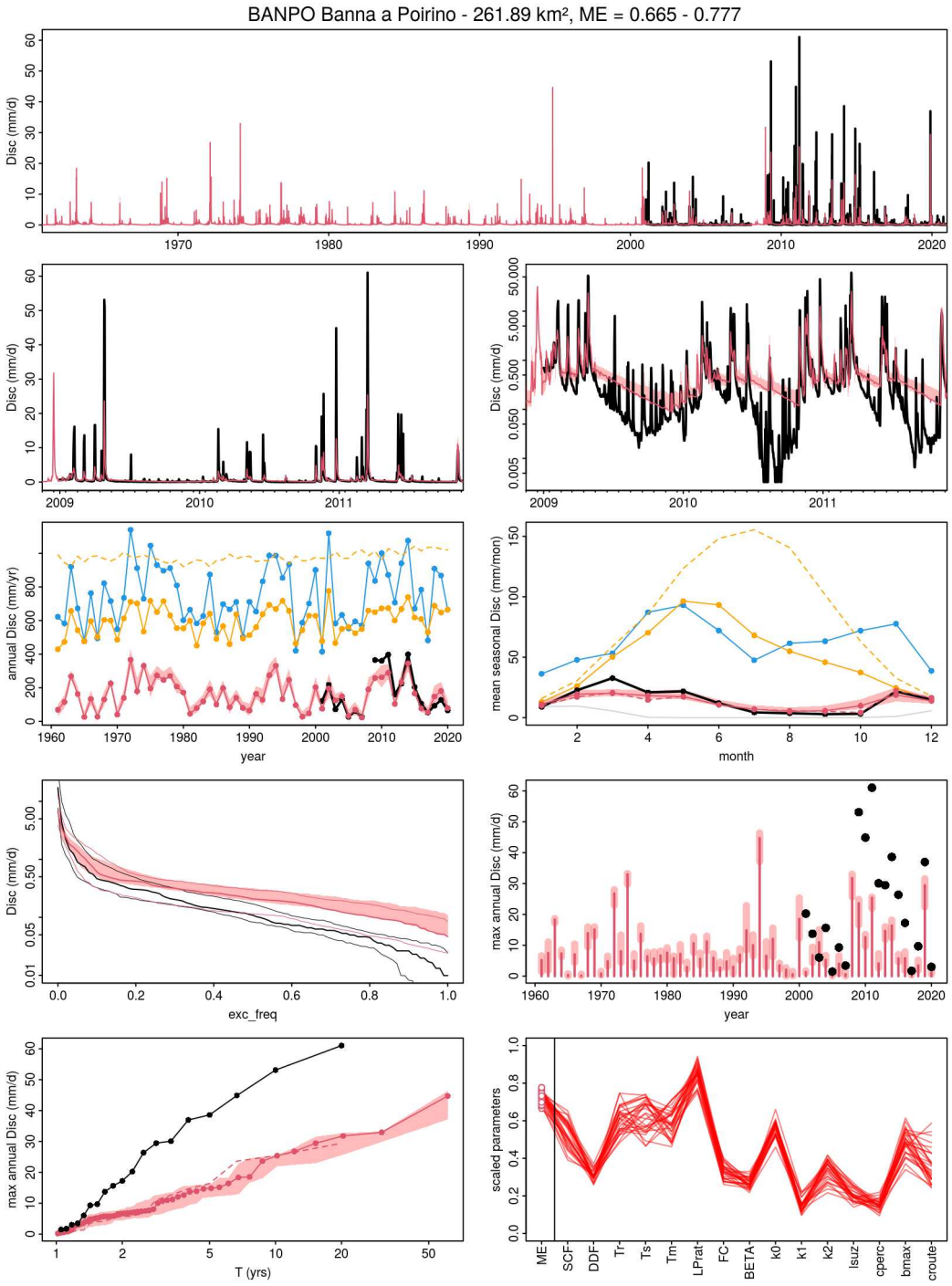


Fig. E.3 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 003.

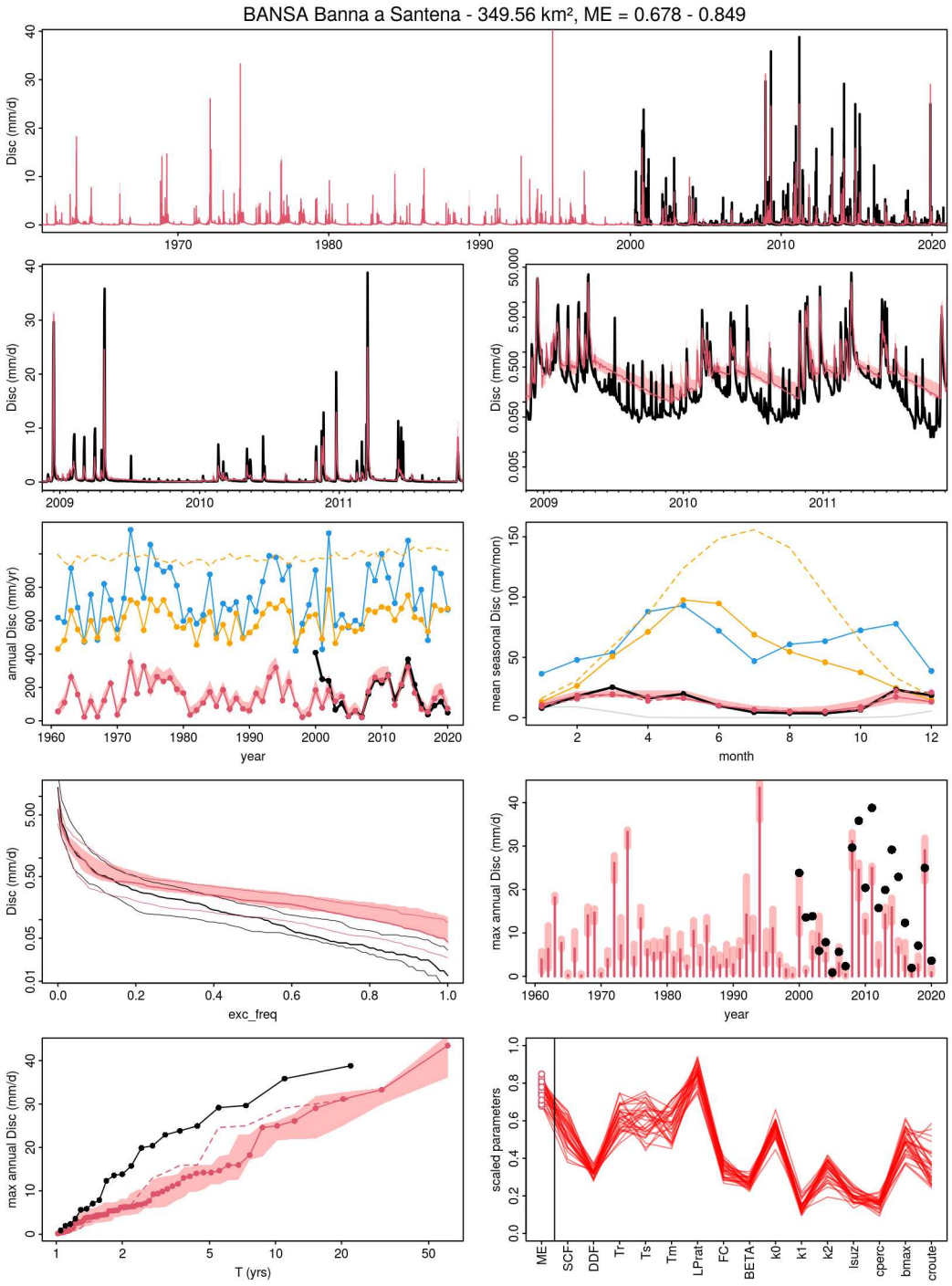


Fig. E.4 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 004.

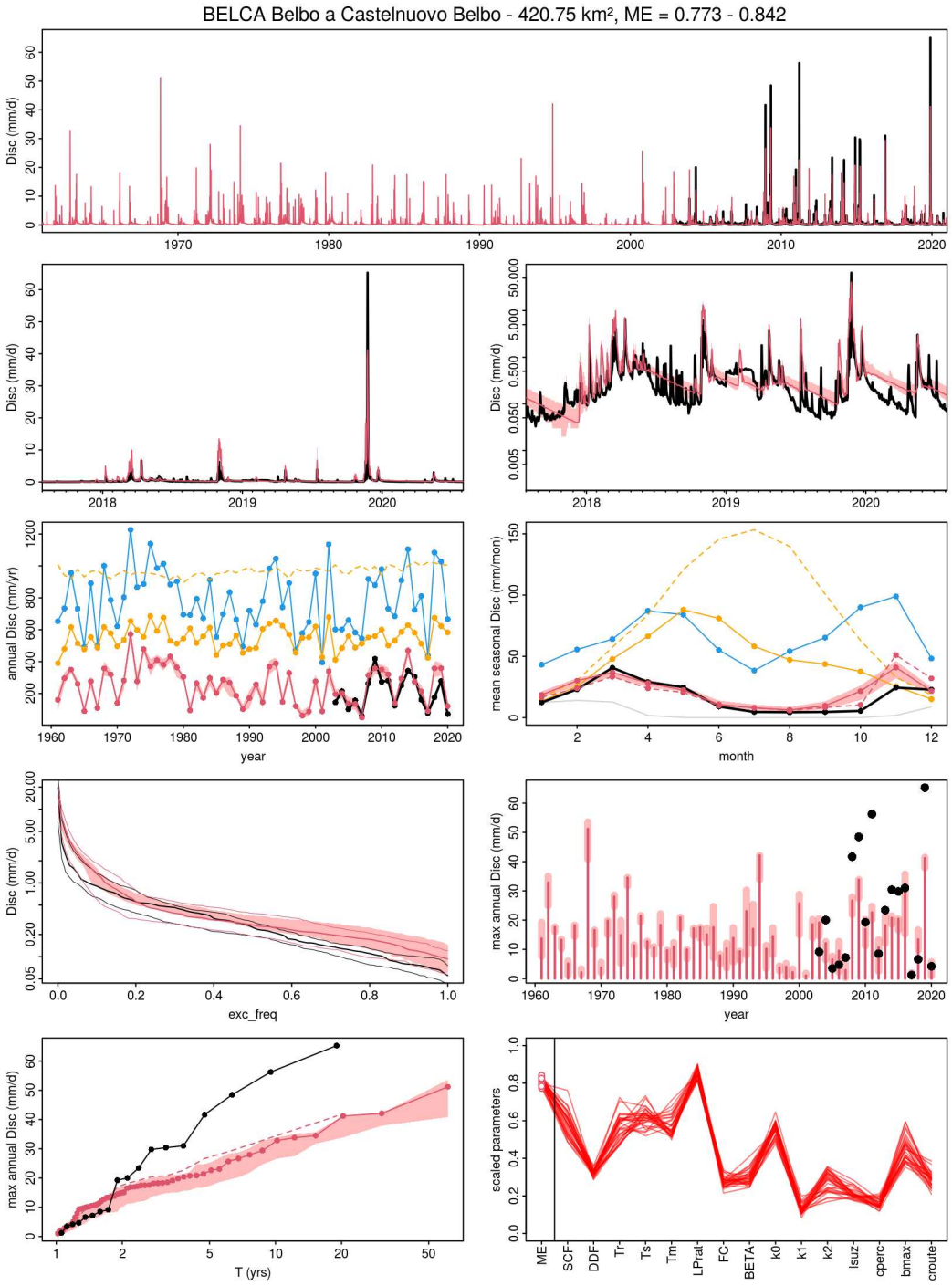


Fig. E.5 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 005.

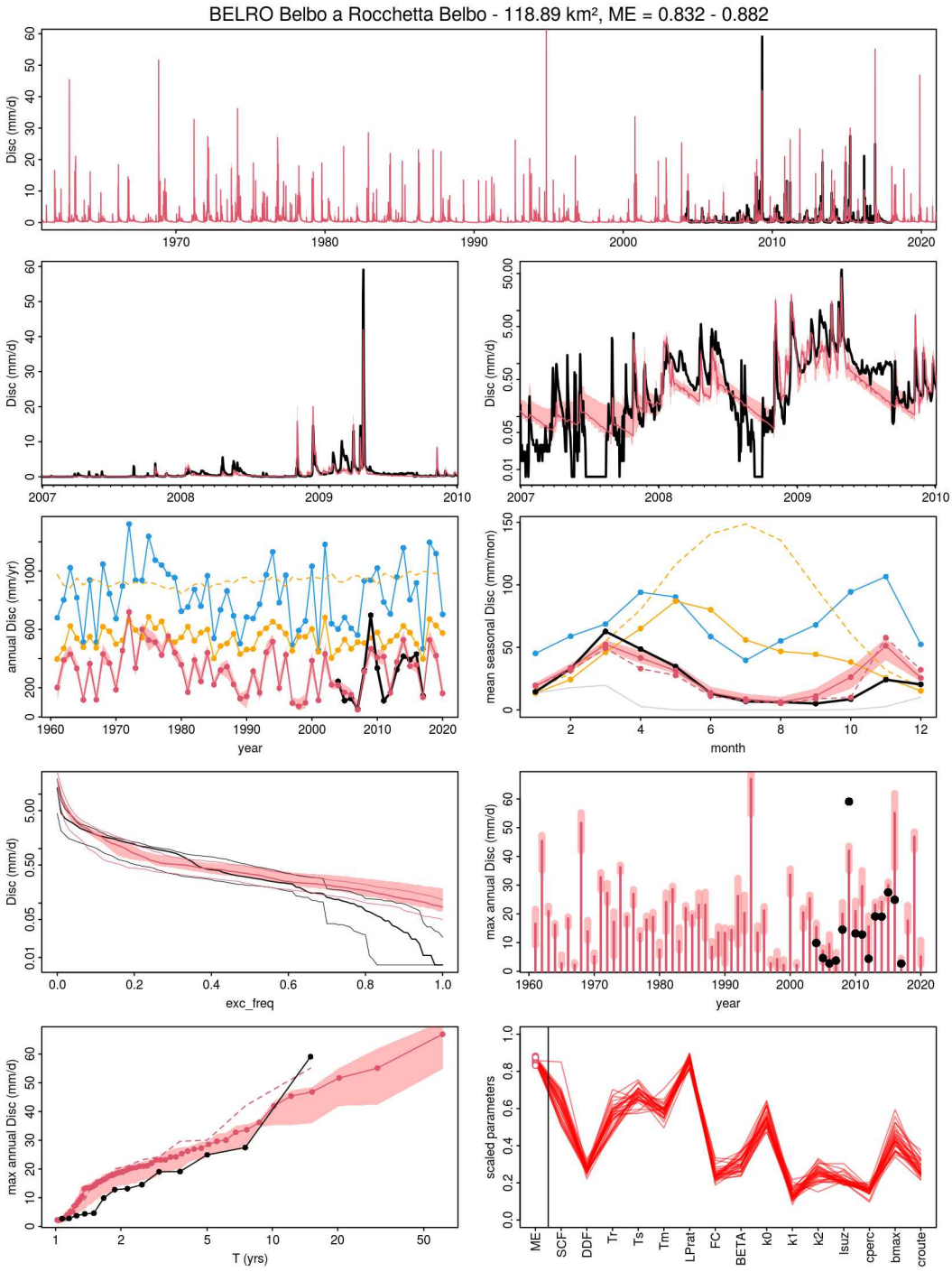


Fig. E.6 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 006.

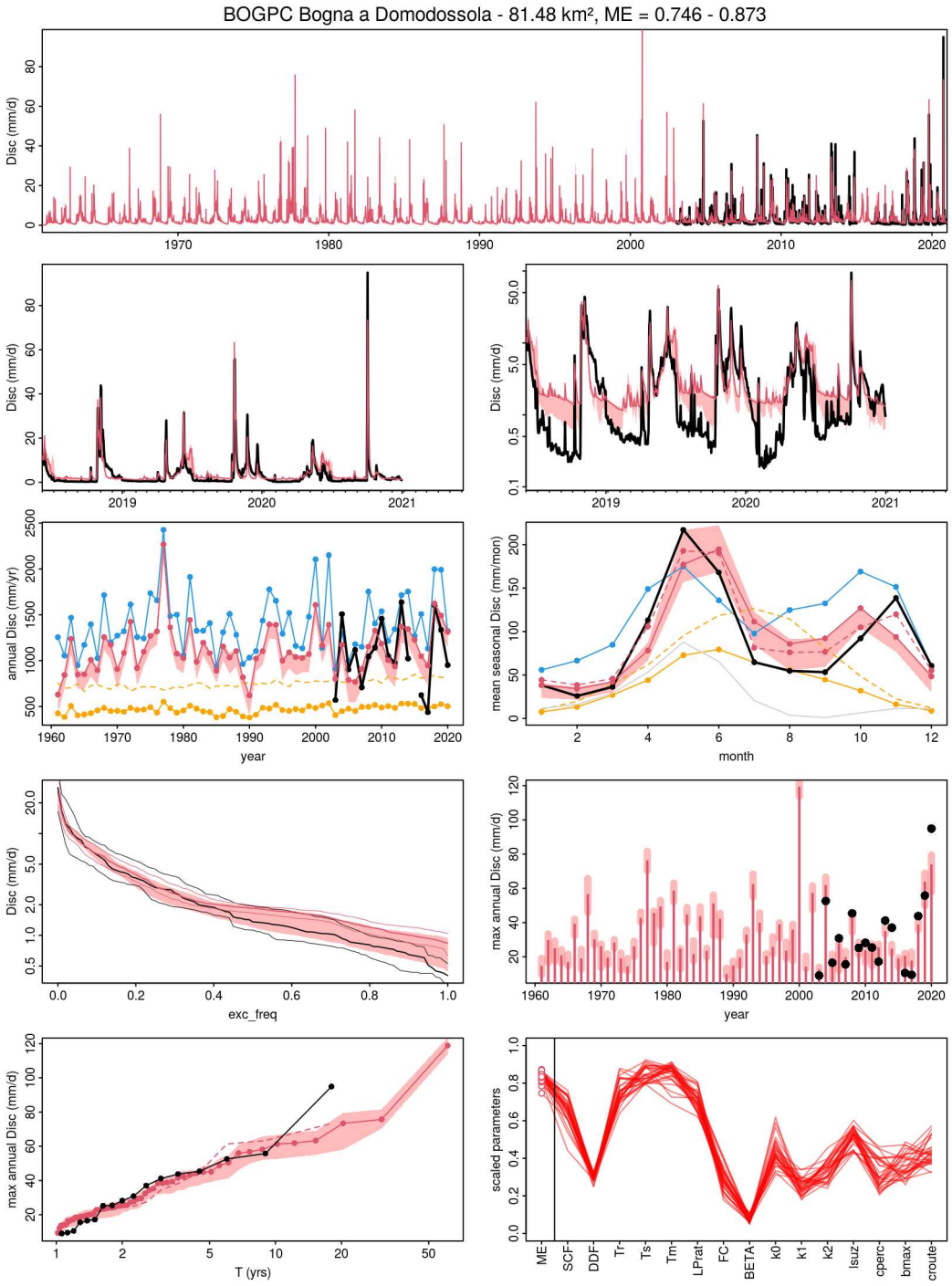


Fig. E.7 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 007.

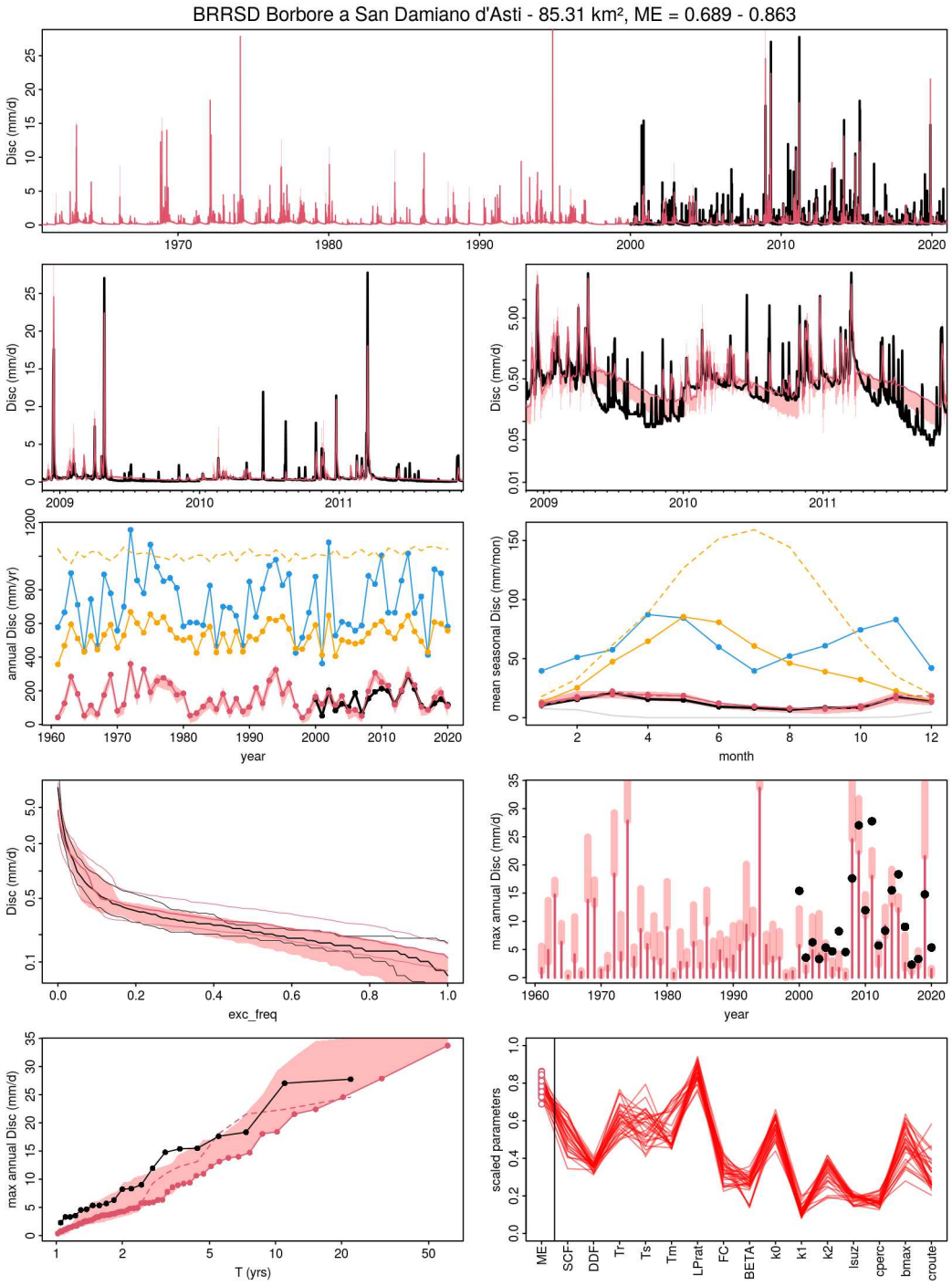


Fig. E.8 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 008.

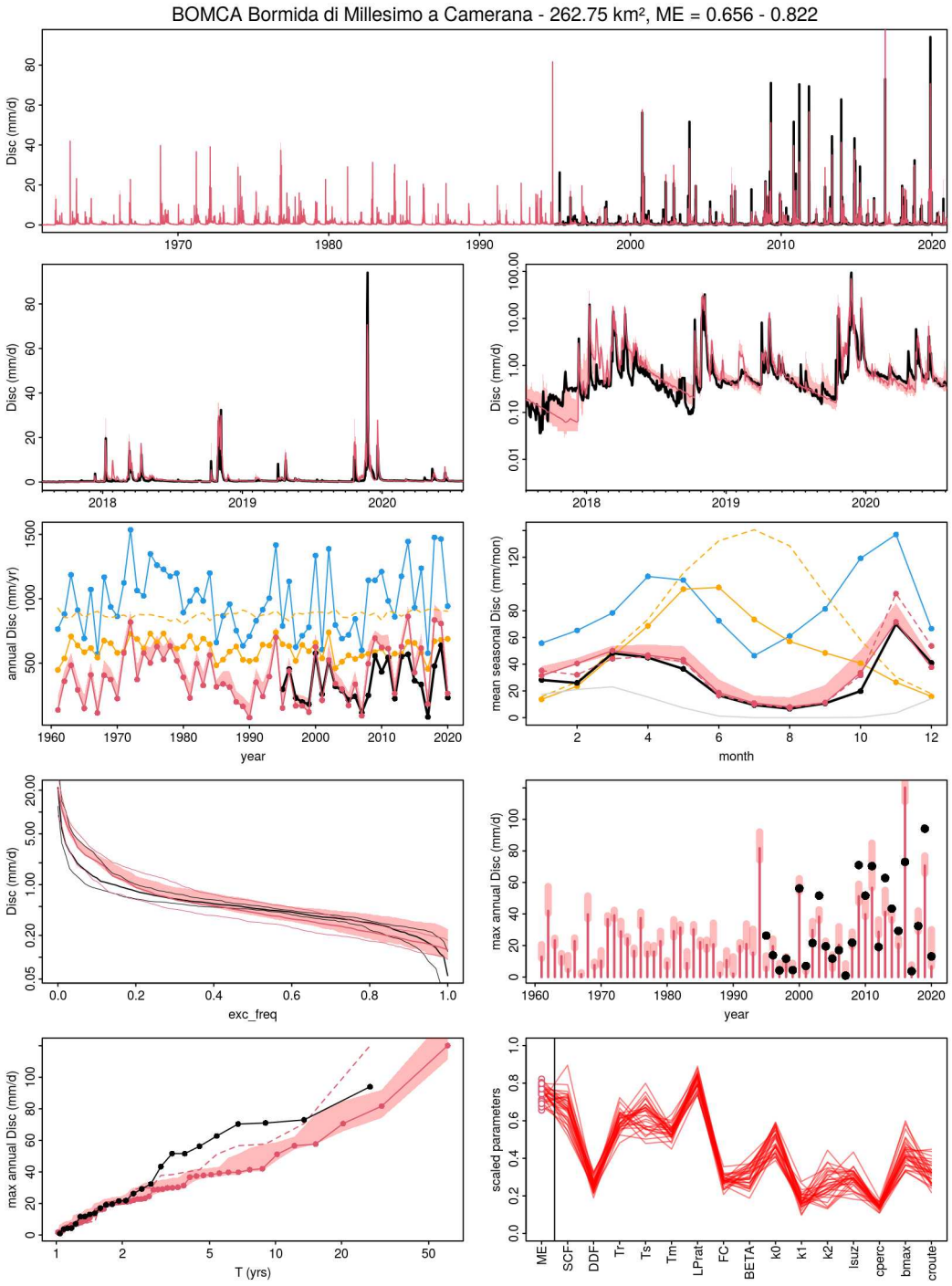


Fig. E.9 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 009.

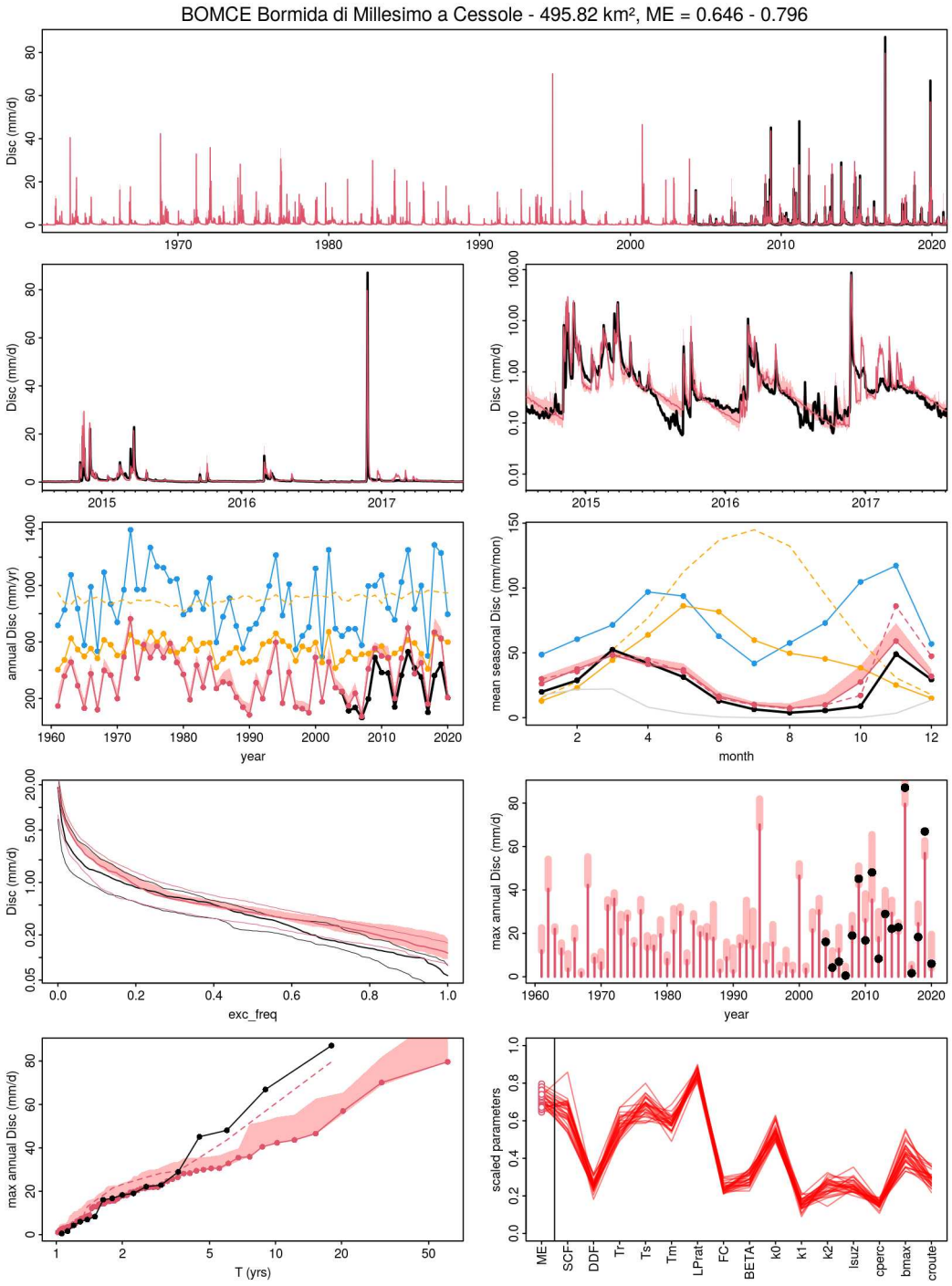


Fig. E.10 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 010.

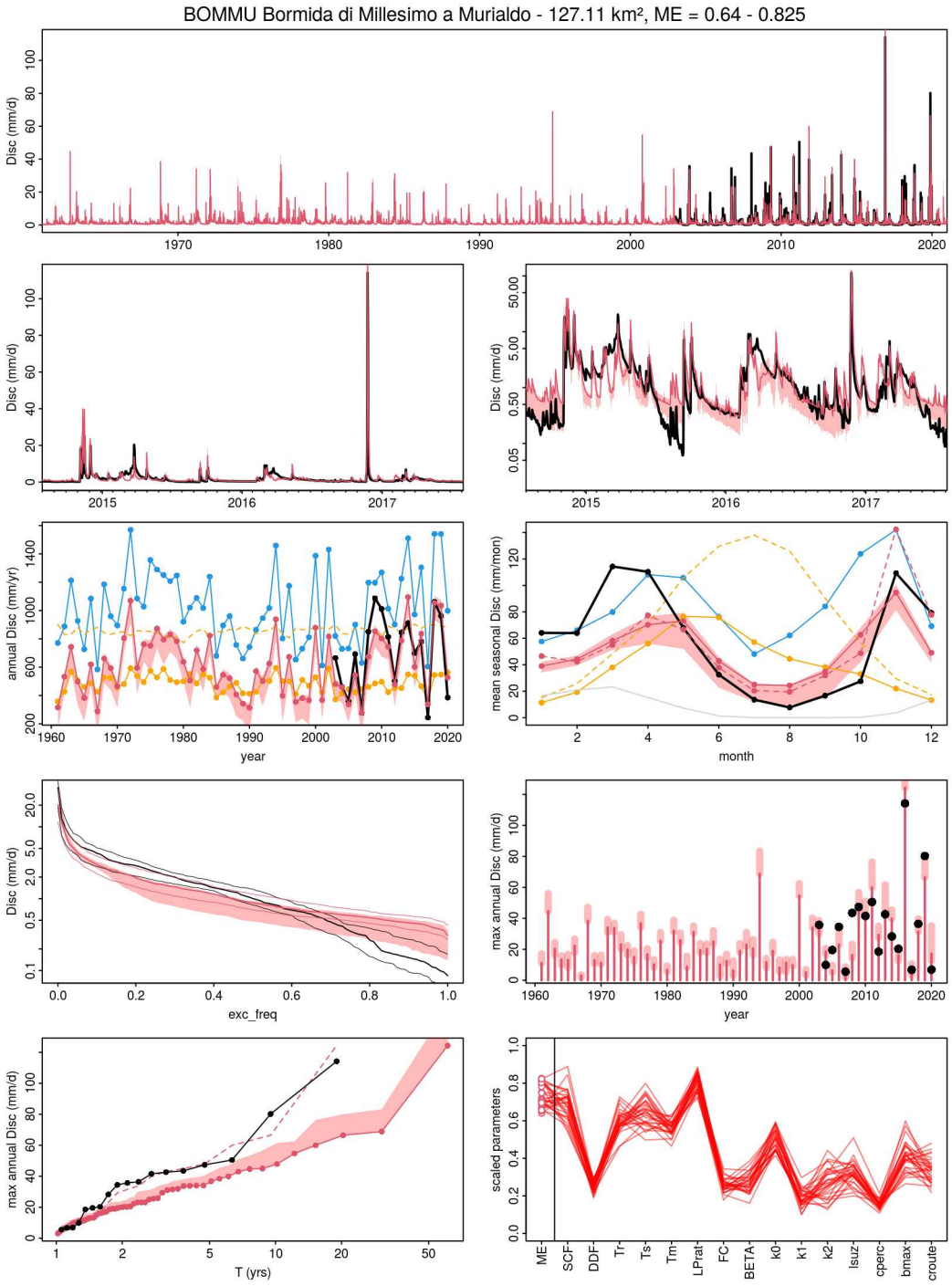


Fig. E.11 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 011.

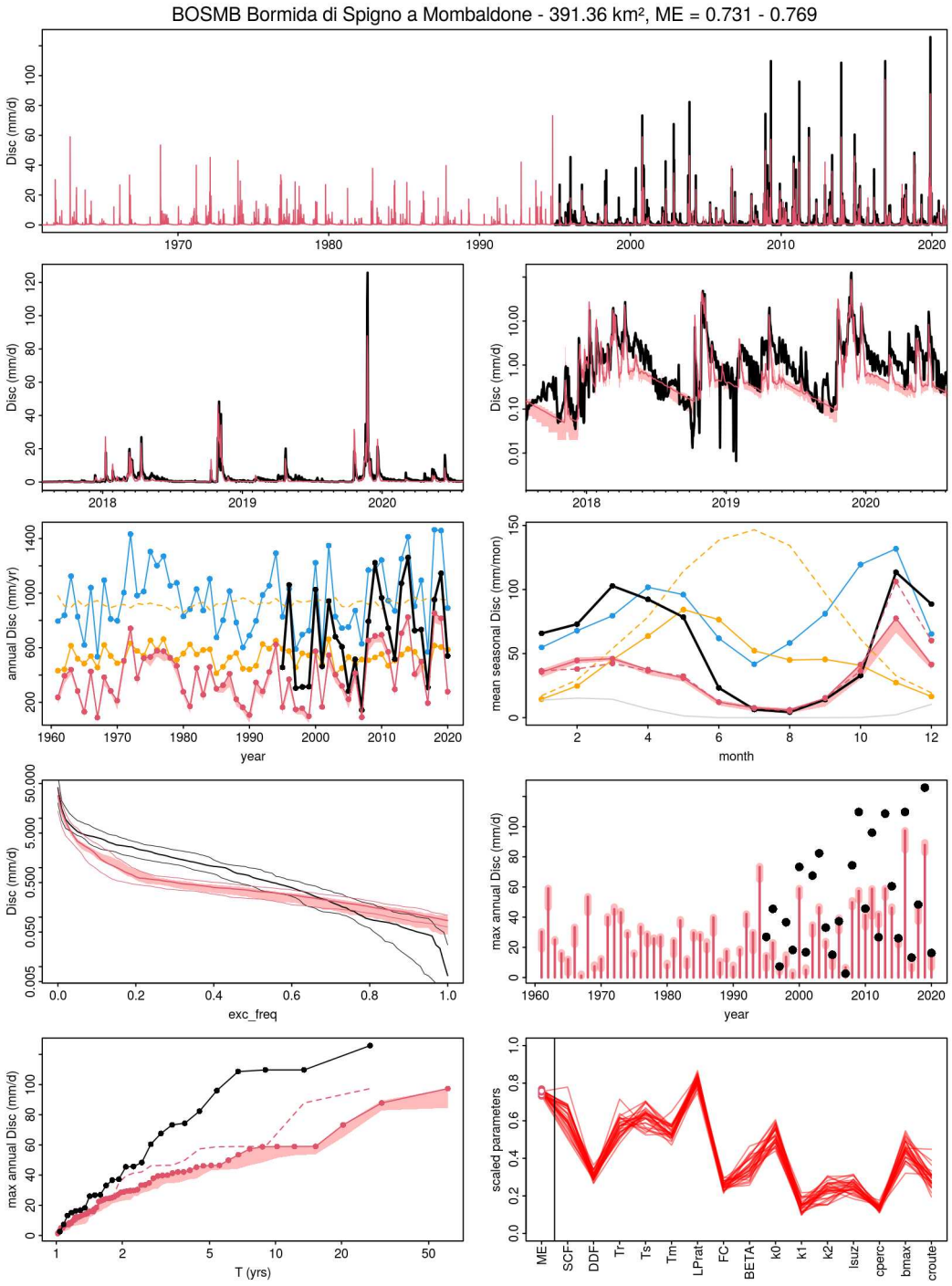


Fig. E.12 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 012.

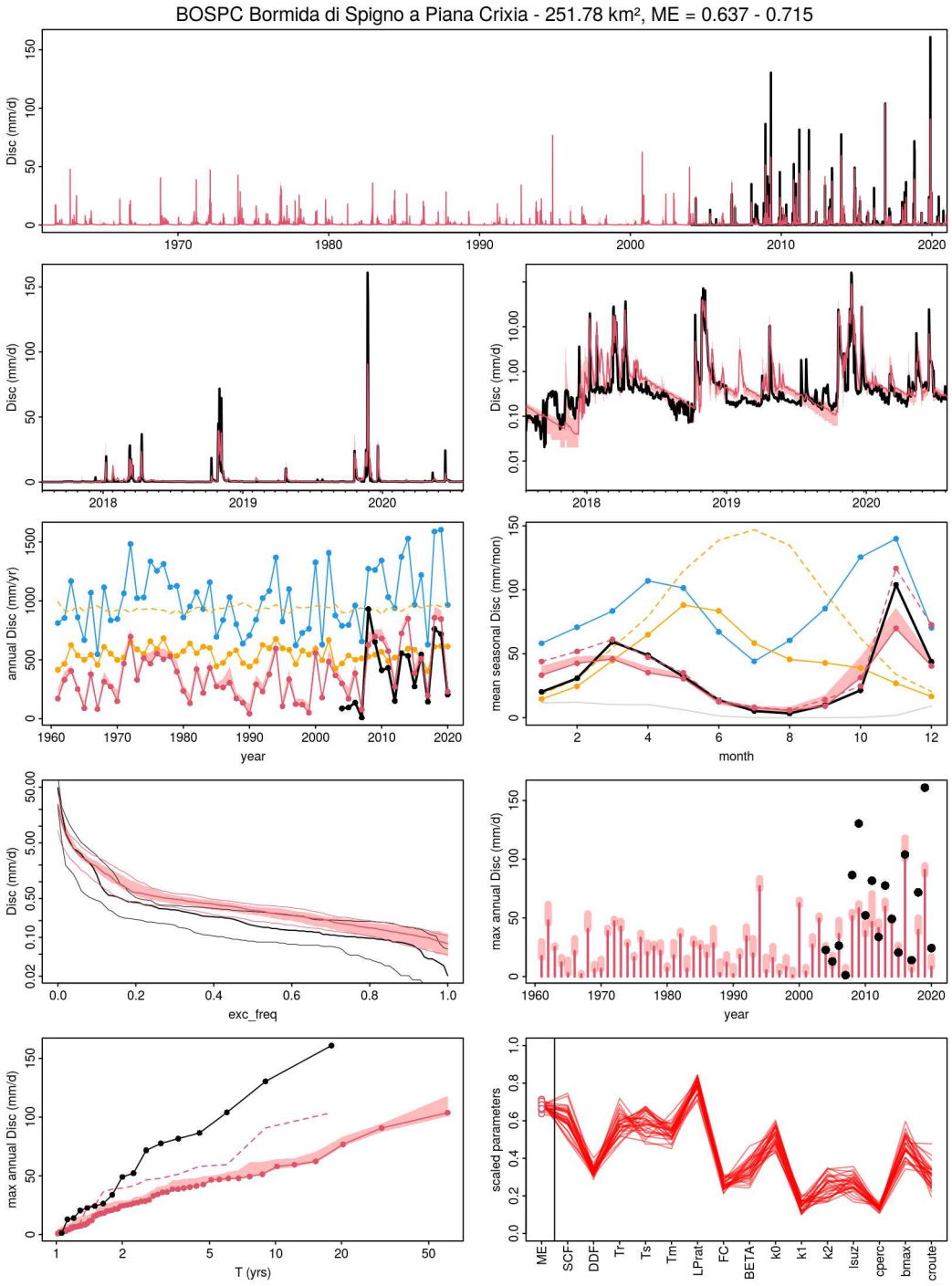


Fig. E.13 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 013.

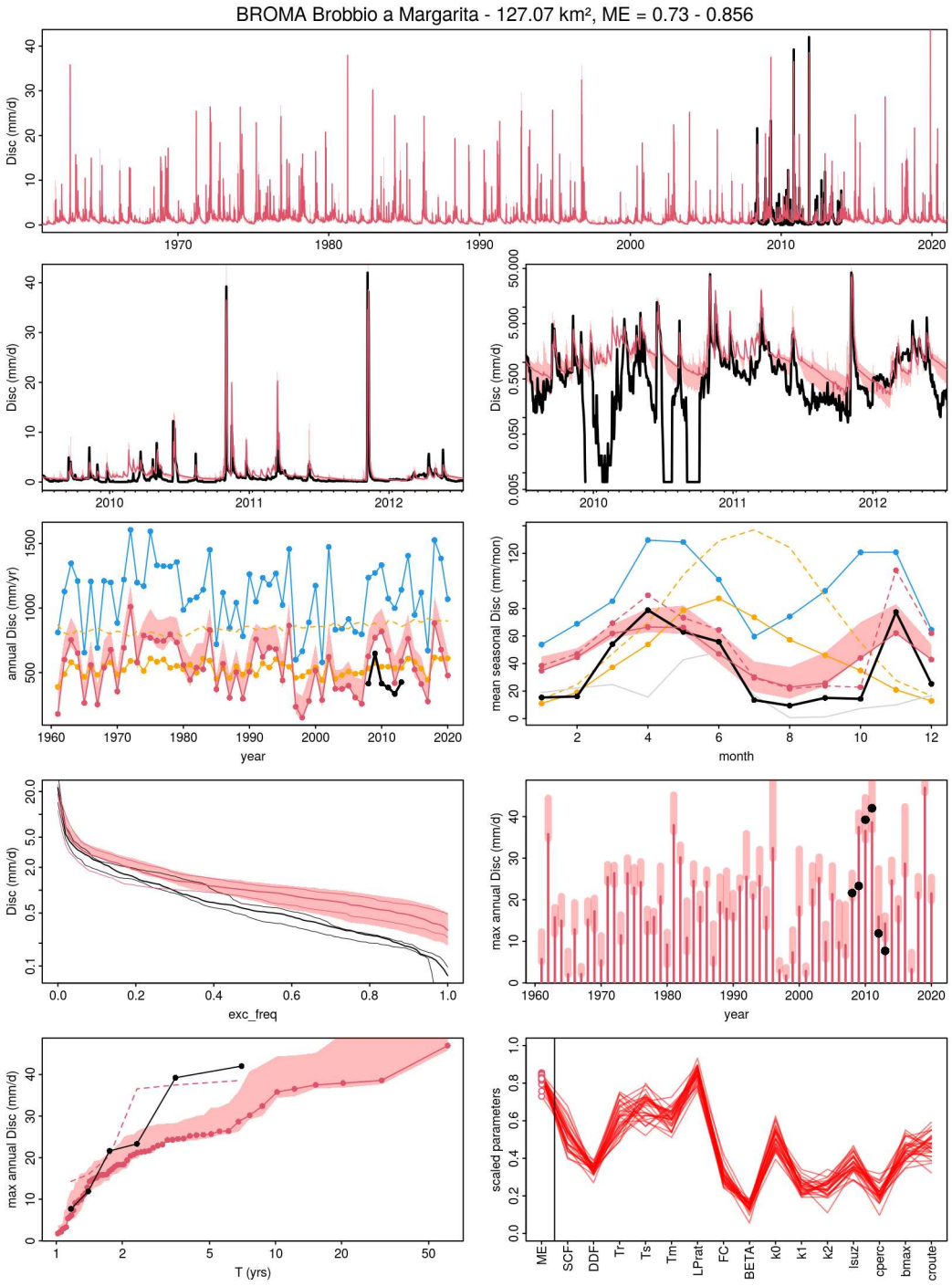


Fig. E.14 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 014.

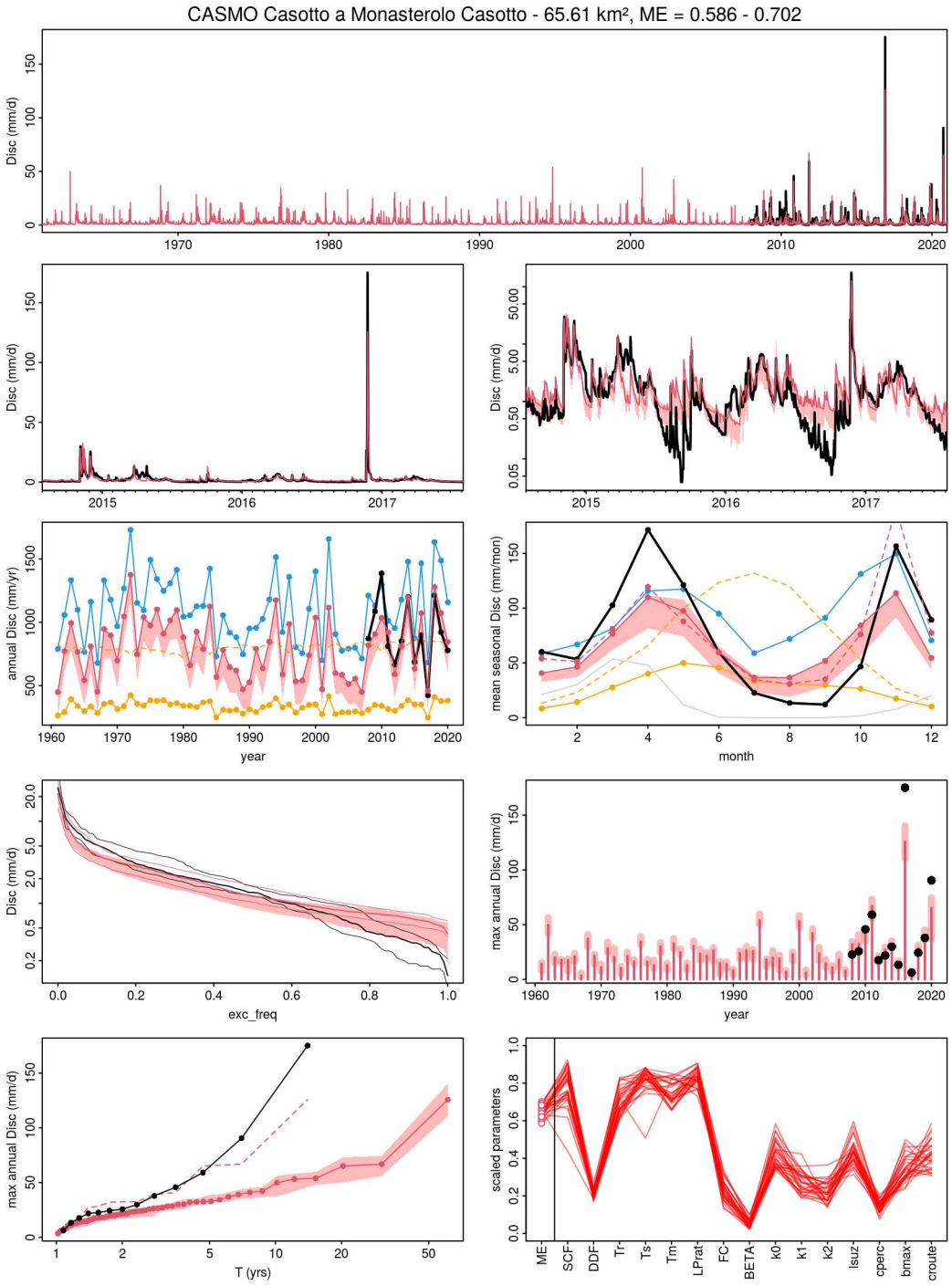


Fig. E.15 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 015.

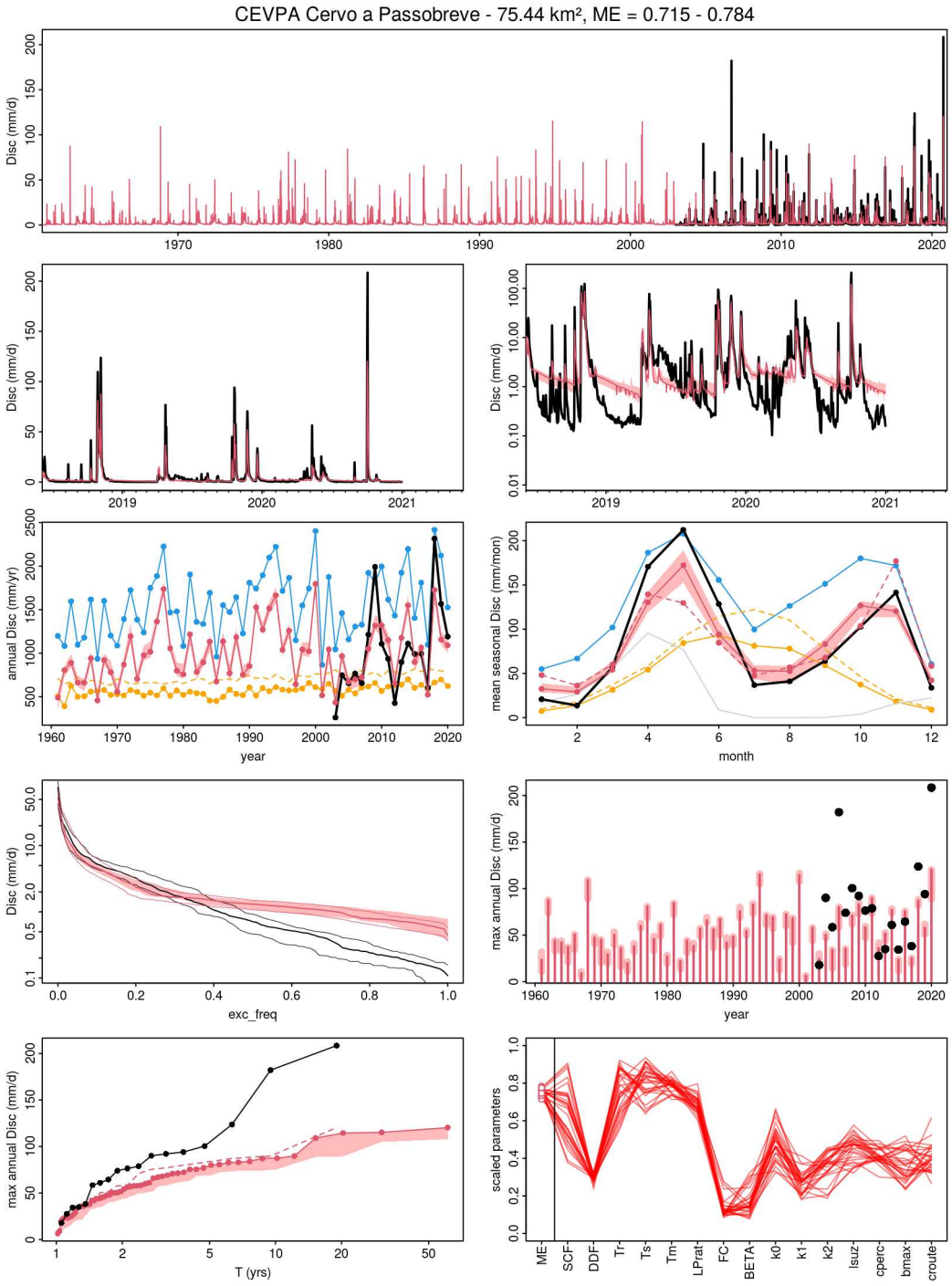


Fig. E.16 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 016.

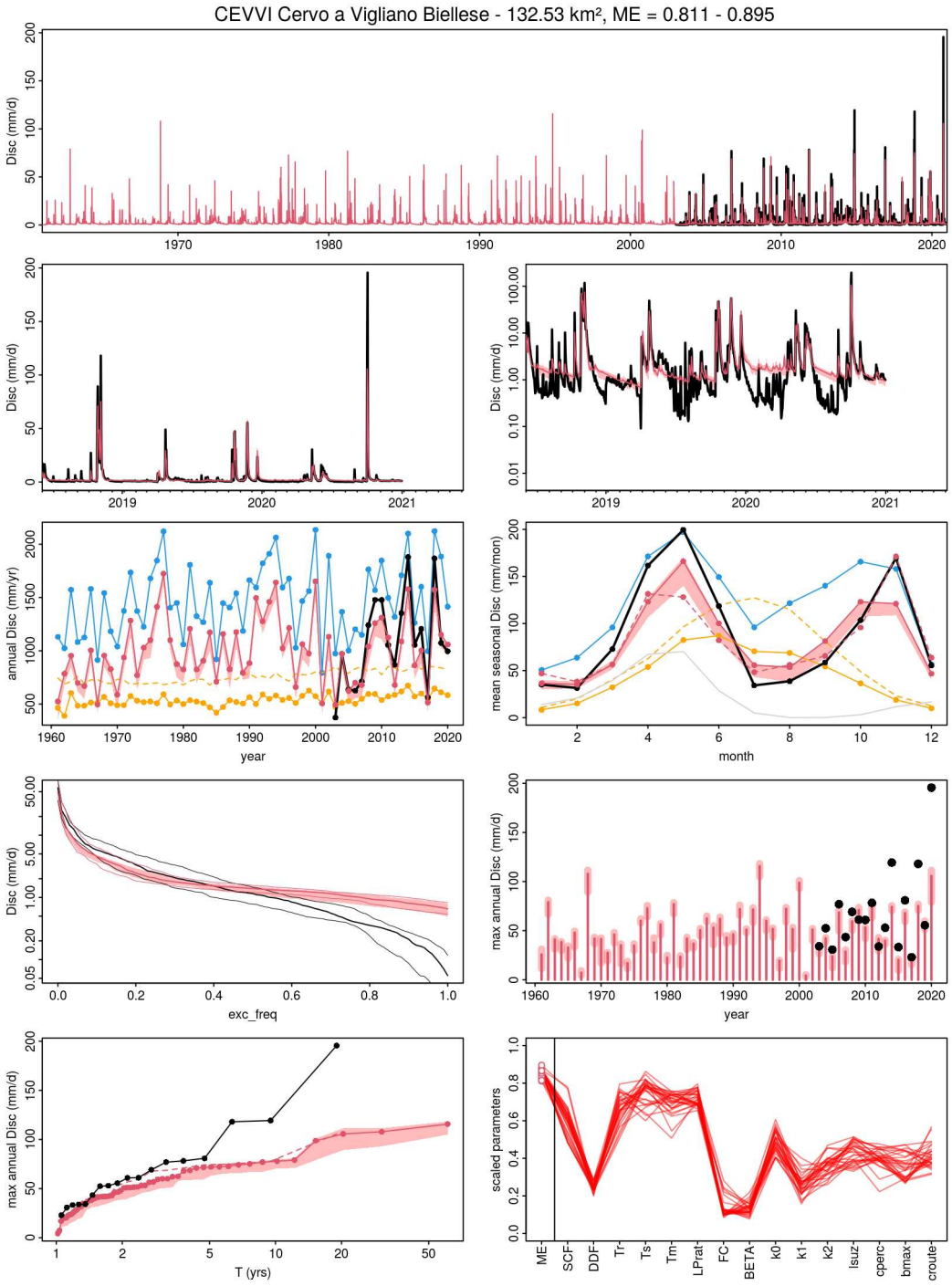


Fig. E.17 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 017.

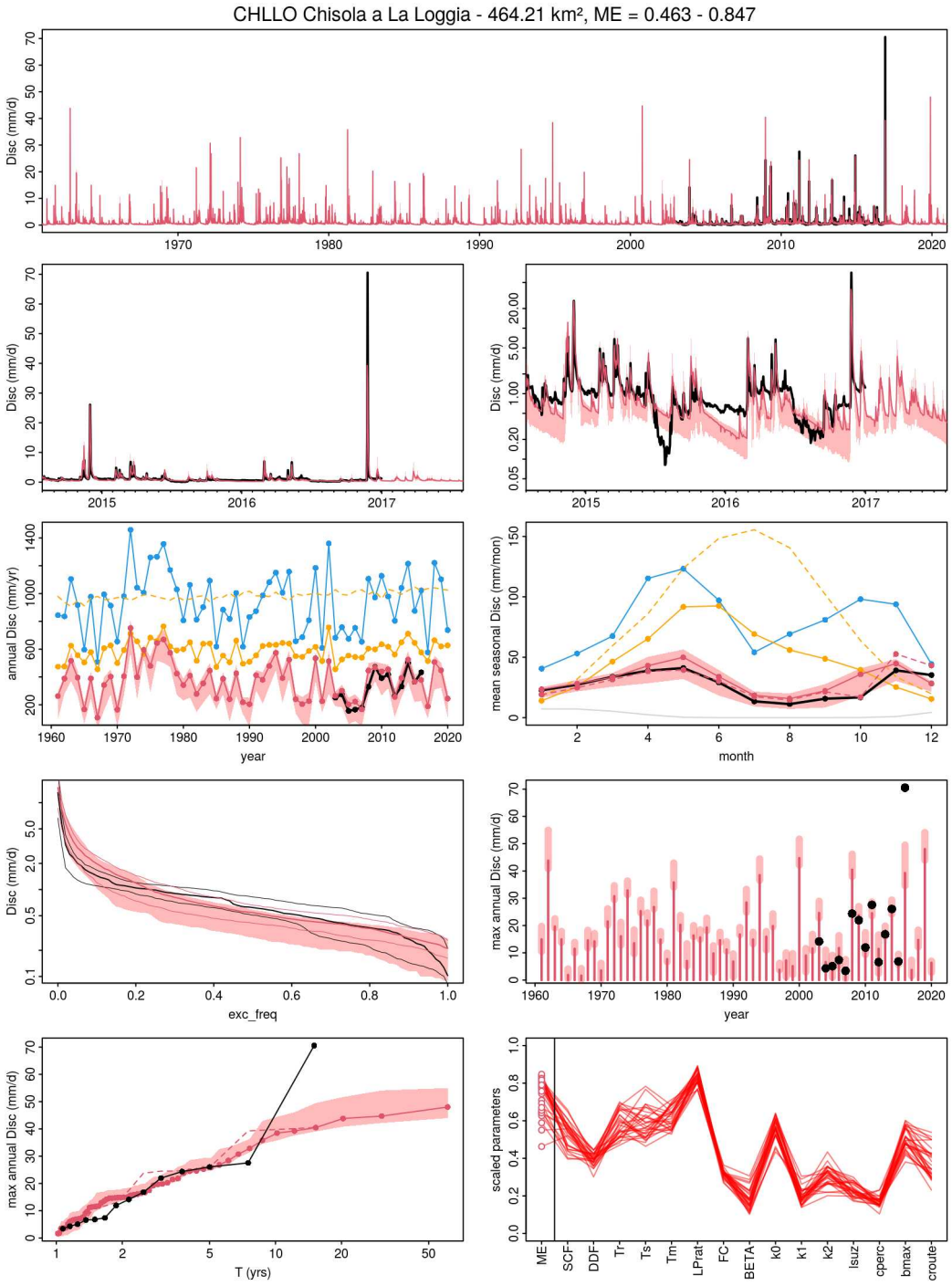


Fig. E.18 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 018.

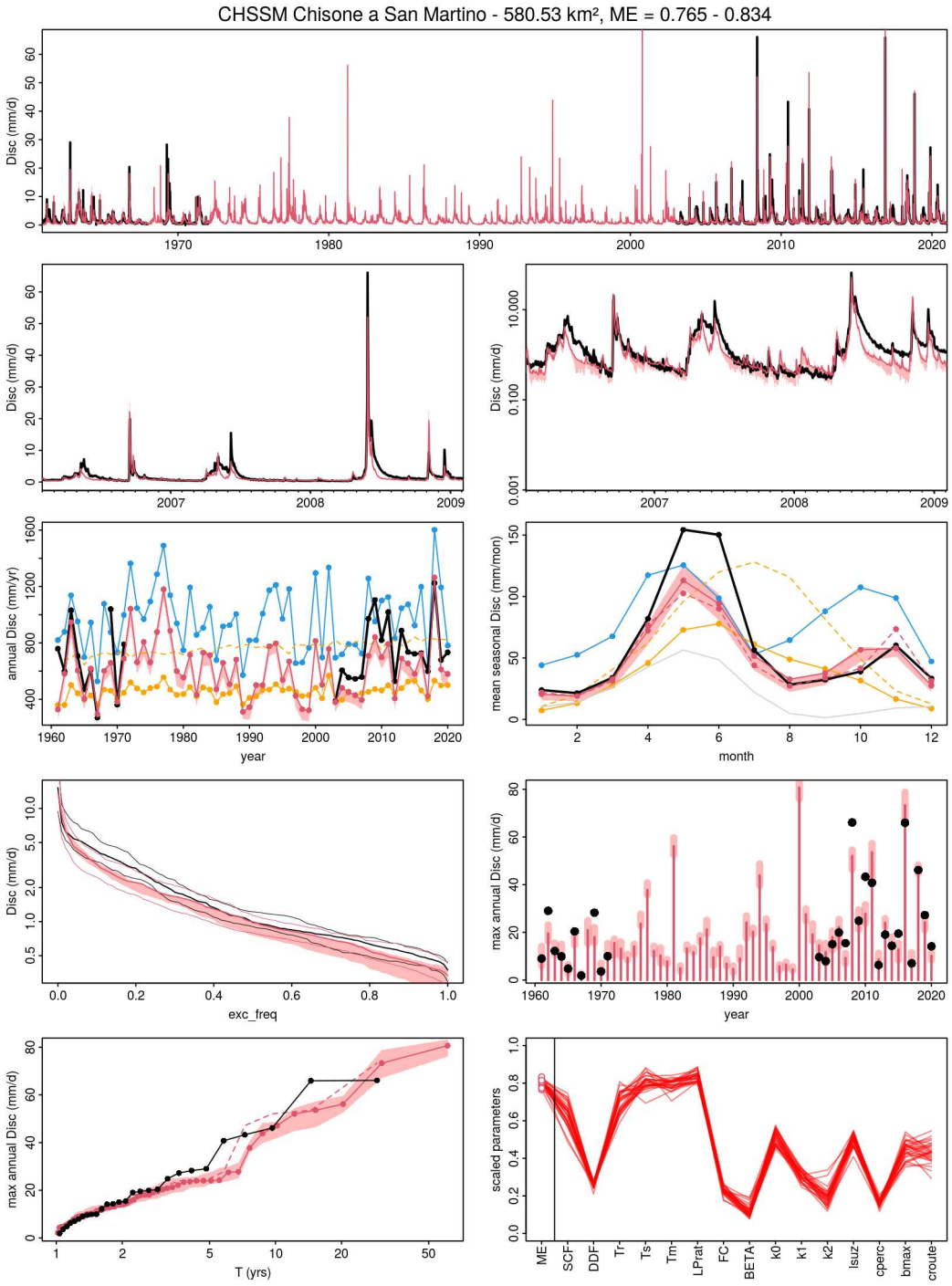


Fig. E.19 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 019.

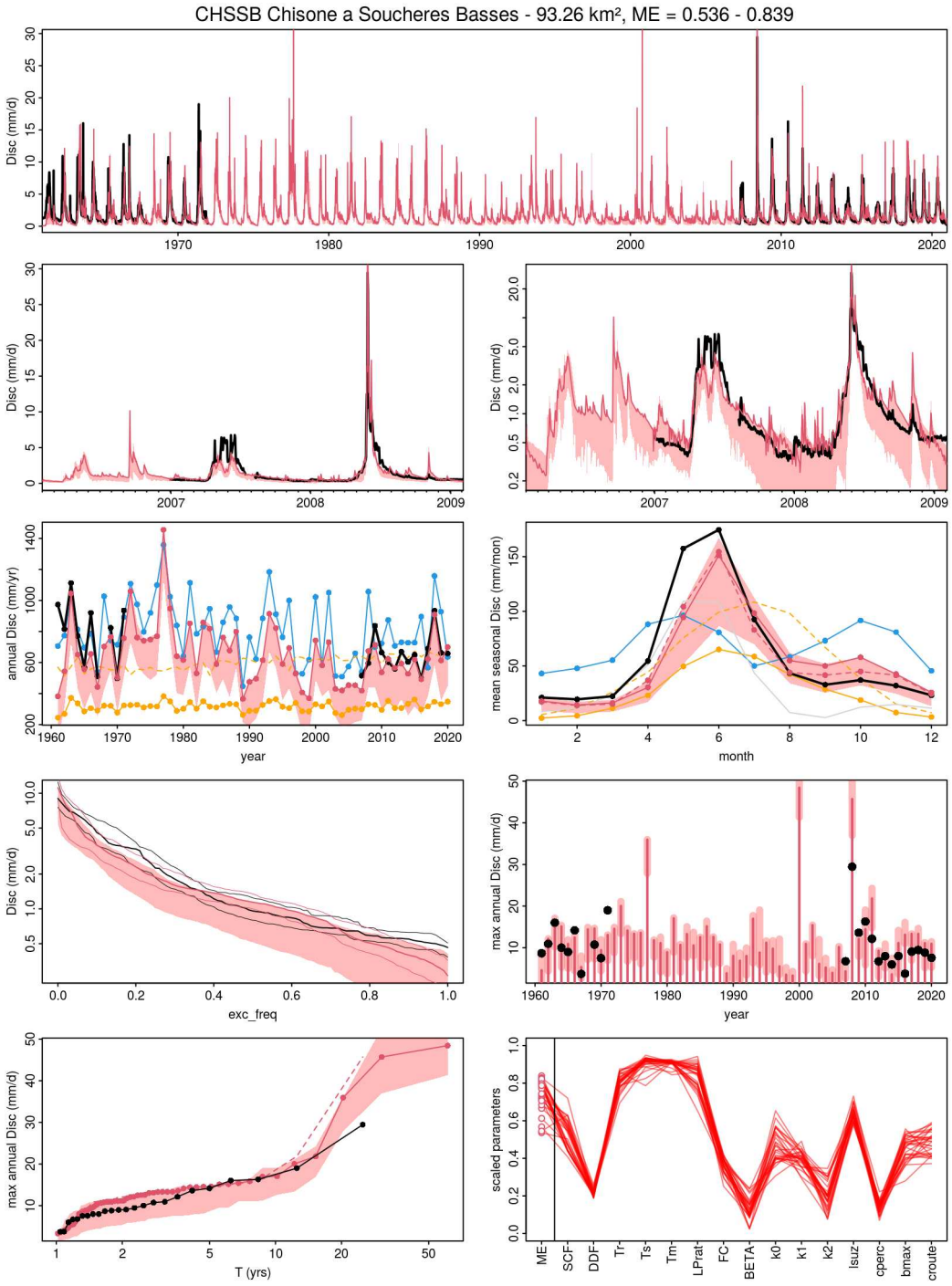


Fig. E.20 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 020.

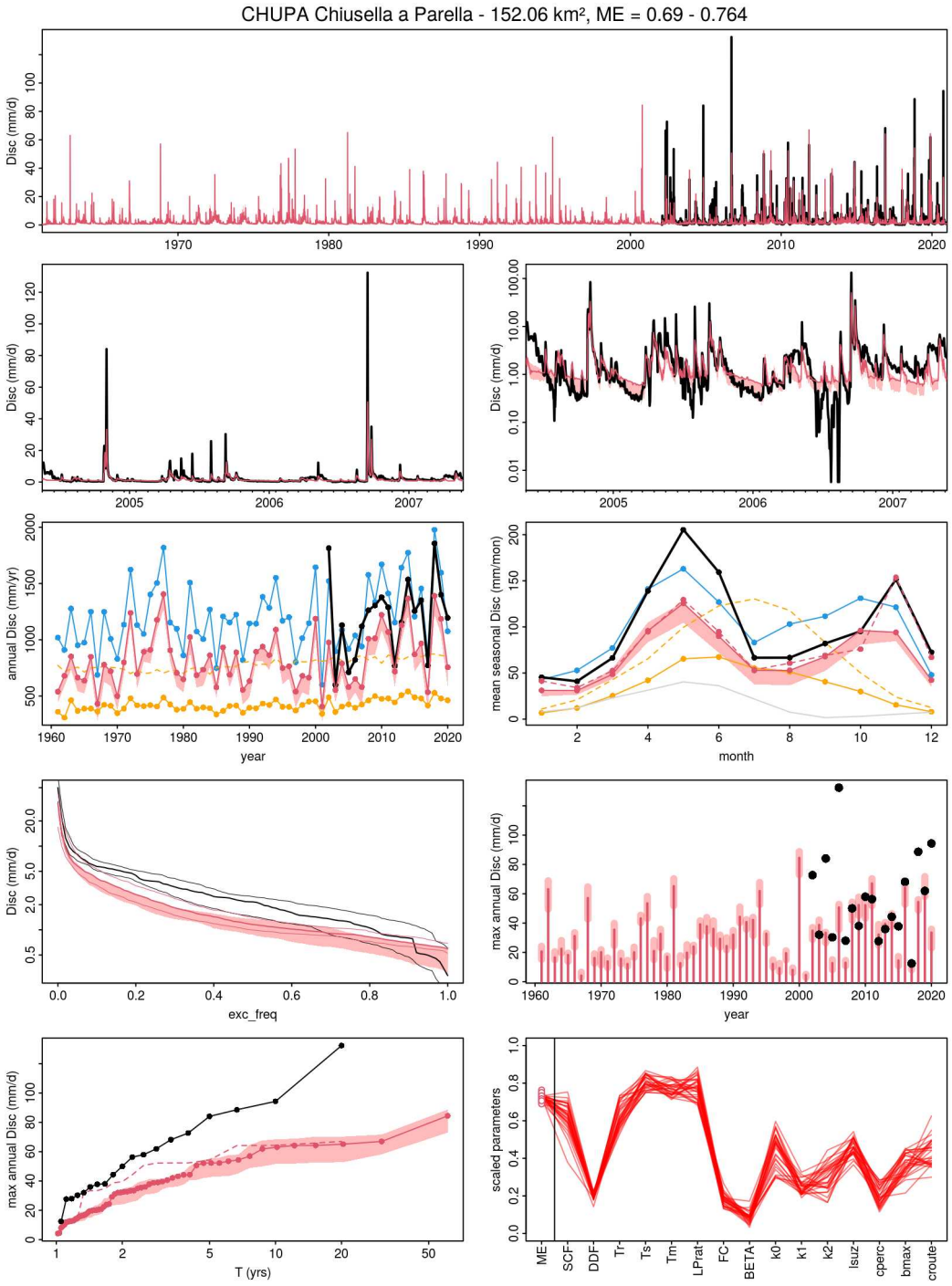


Fig. E.21 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 021.

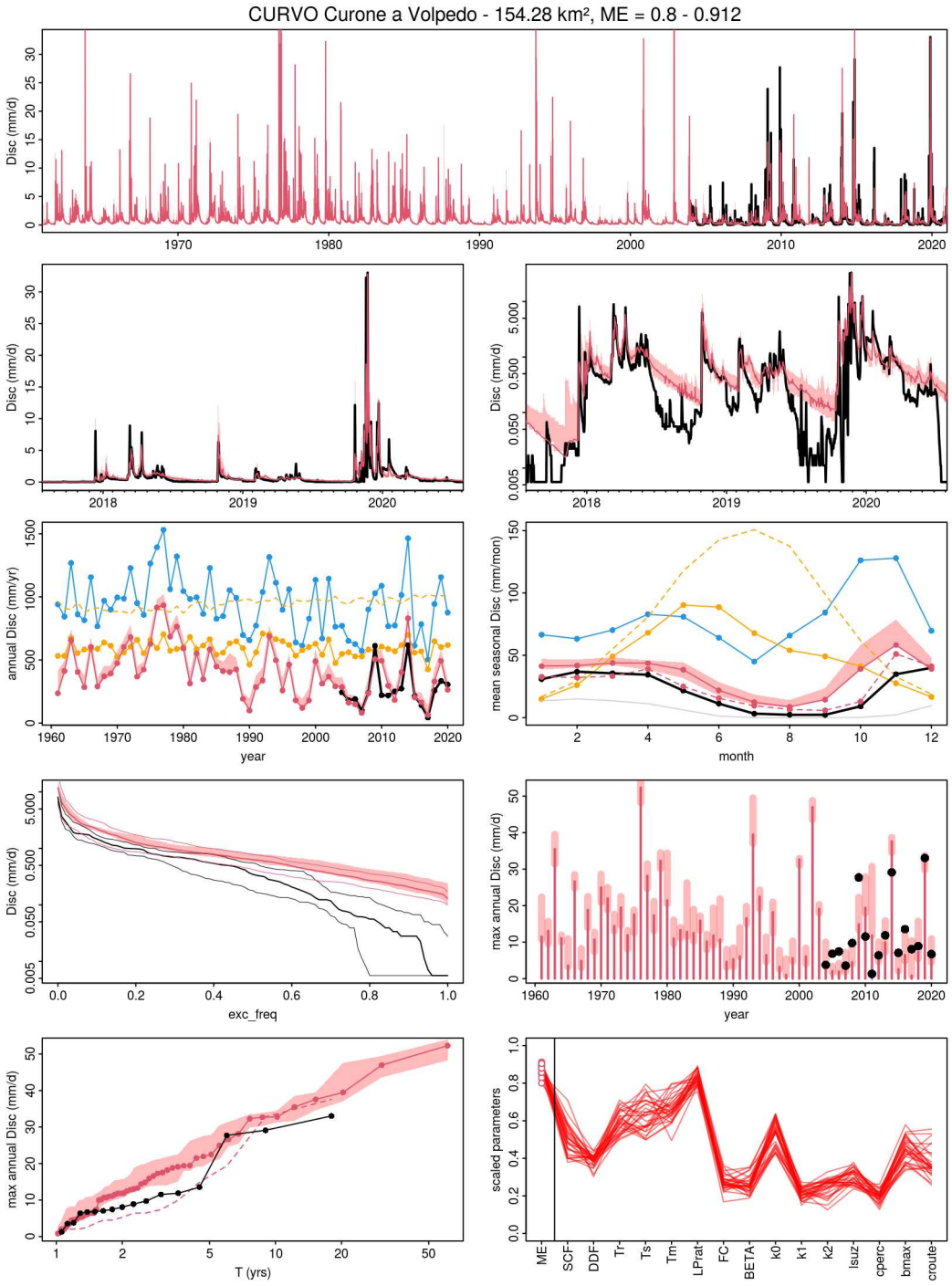


Fig. E.22 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 022.

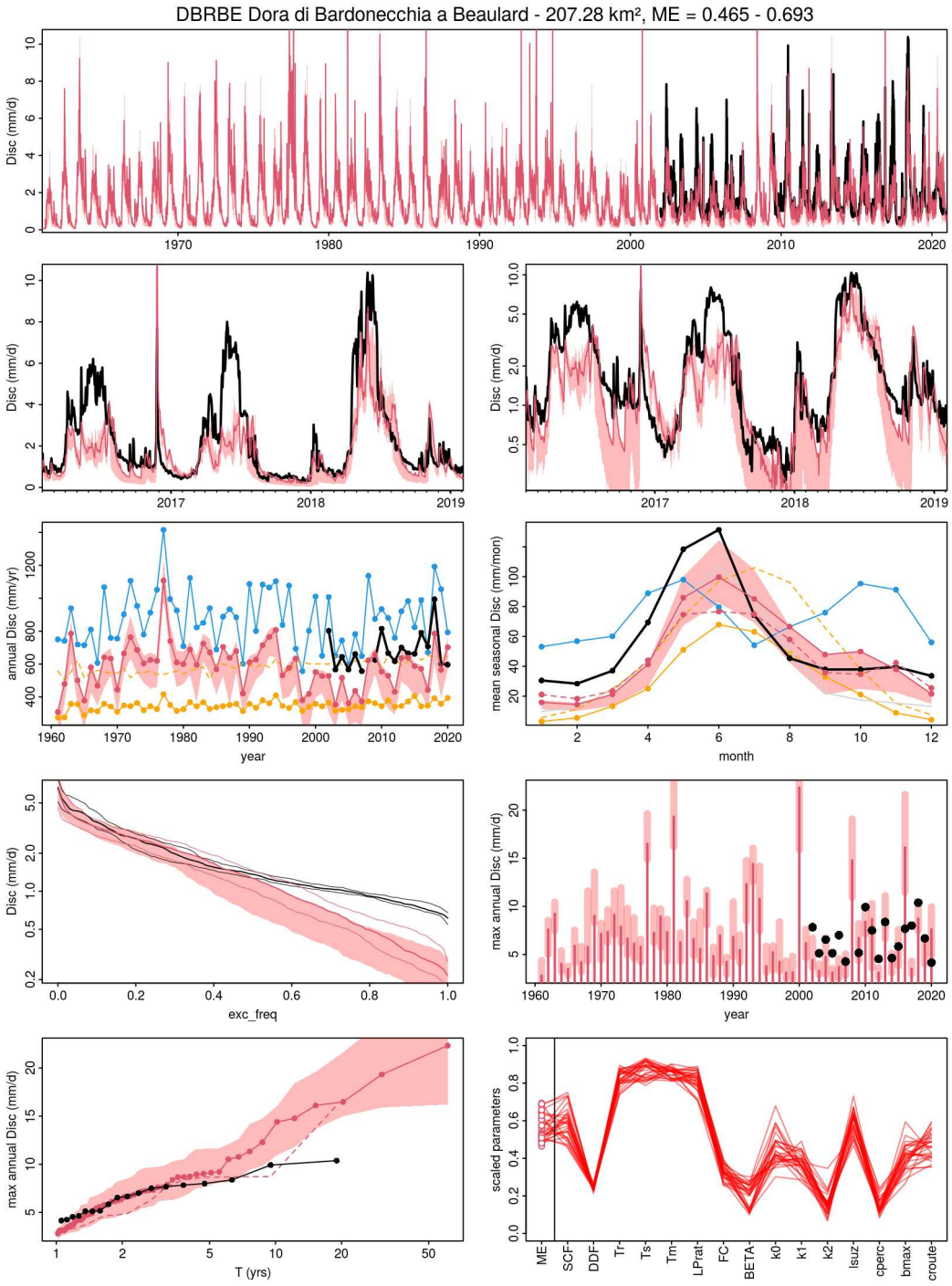


Fig. E.23 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 023.

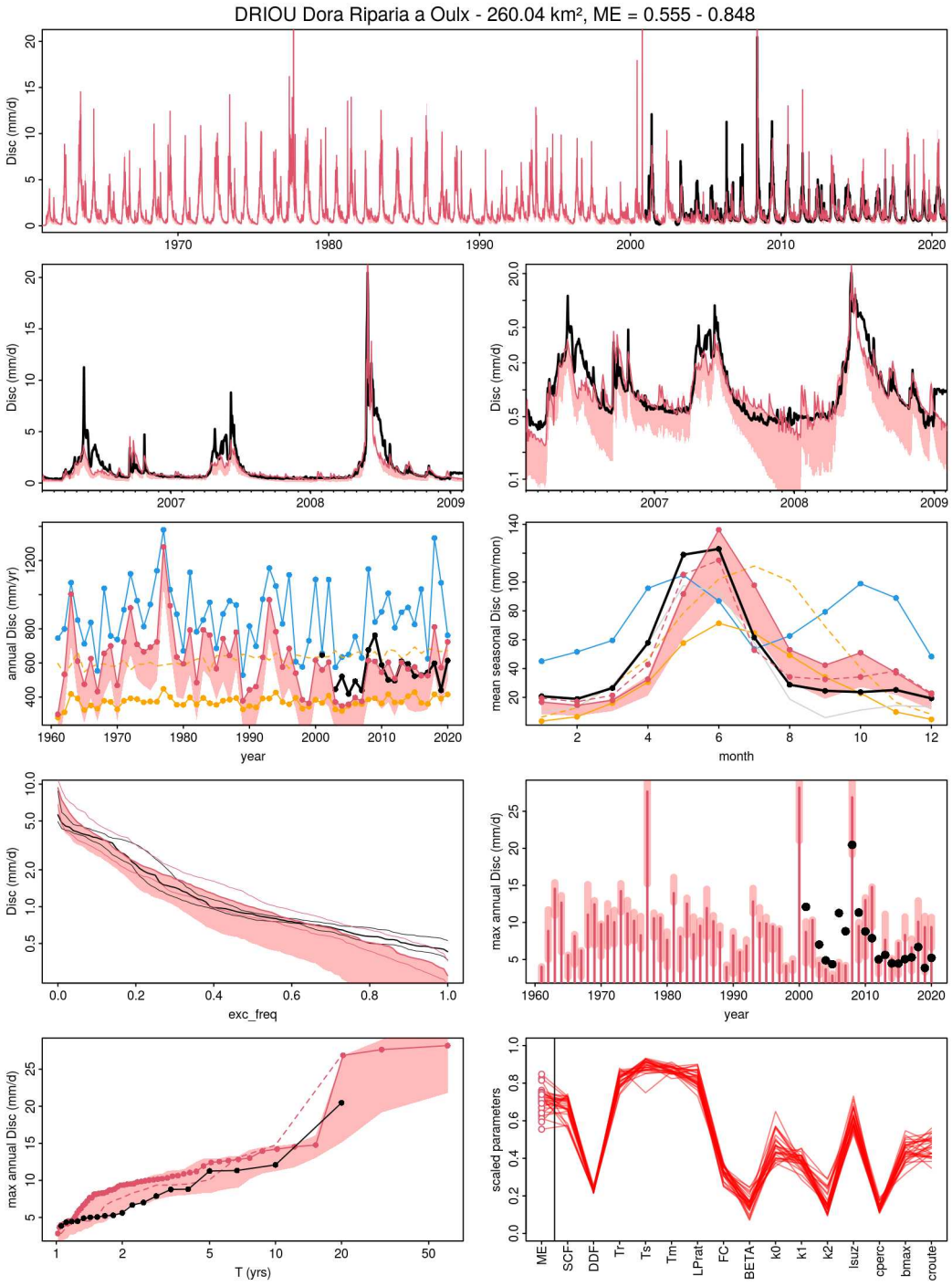


Fig. E.24 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 024.

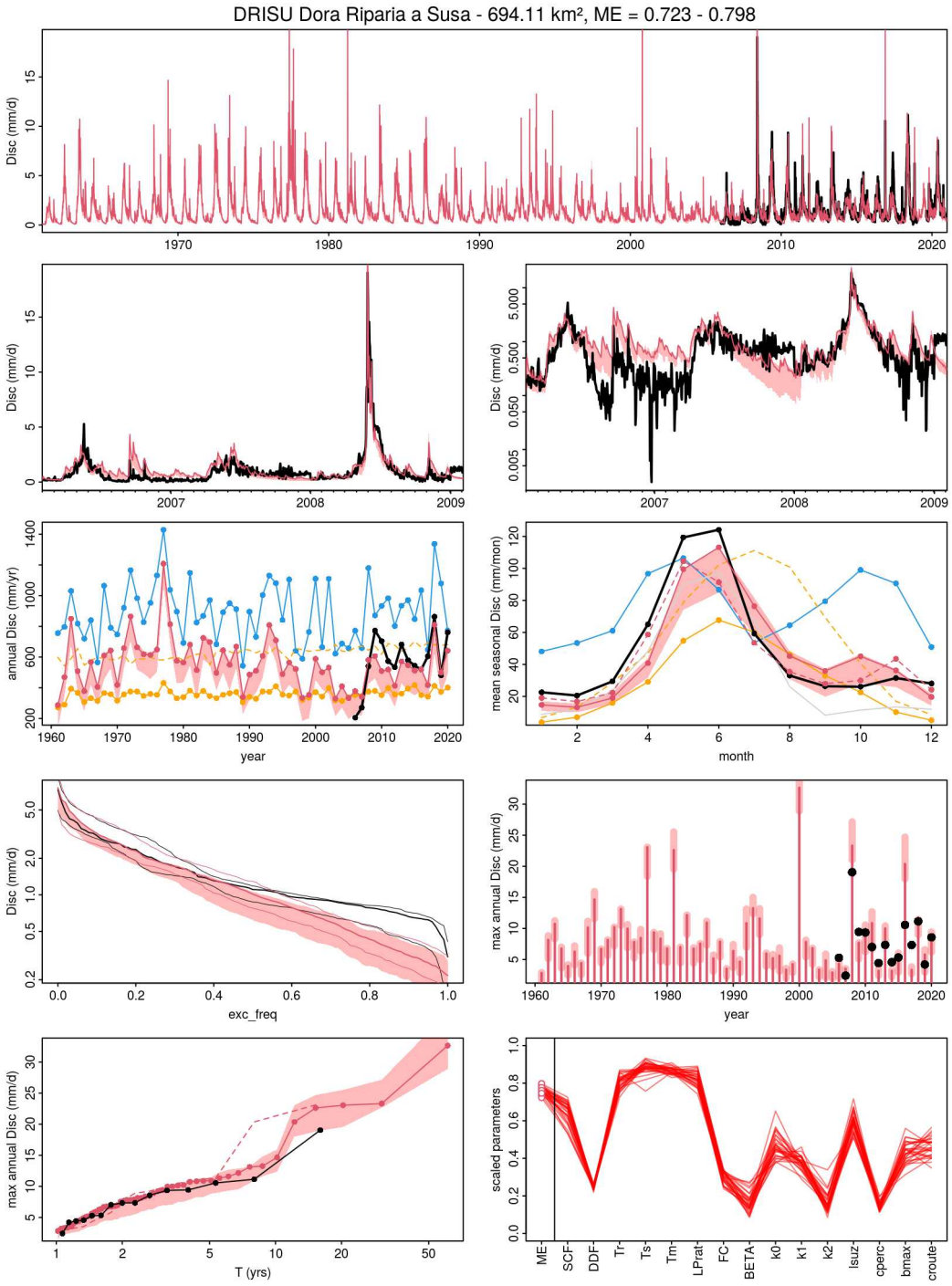


Fig. E.25 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 025.

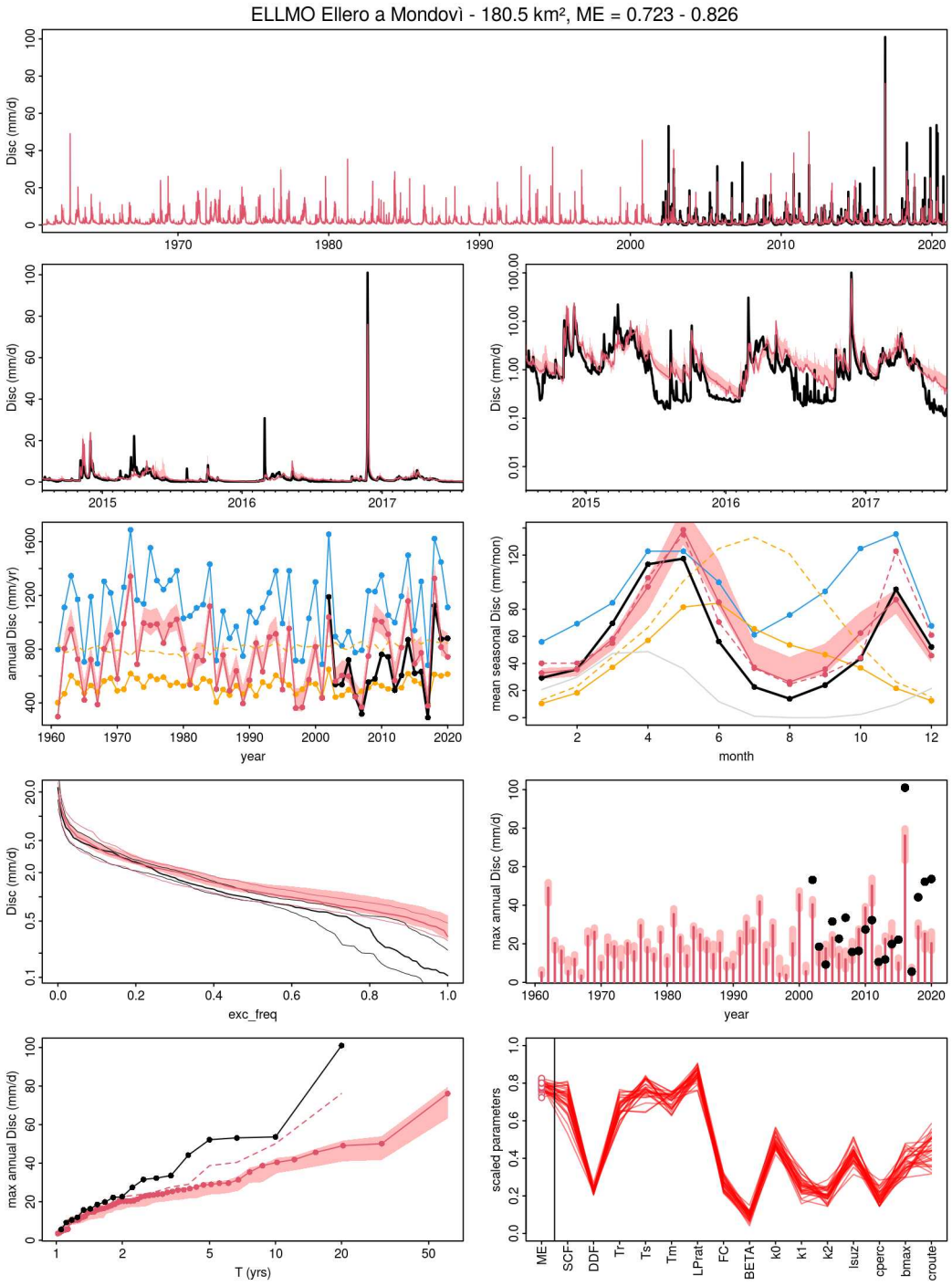


Fig. E.26 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 026.

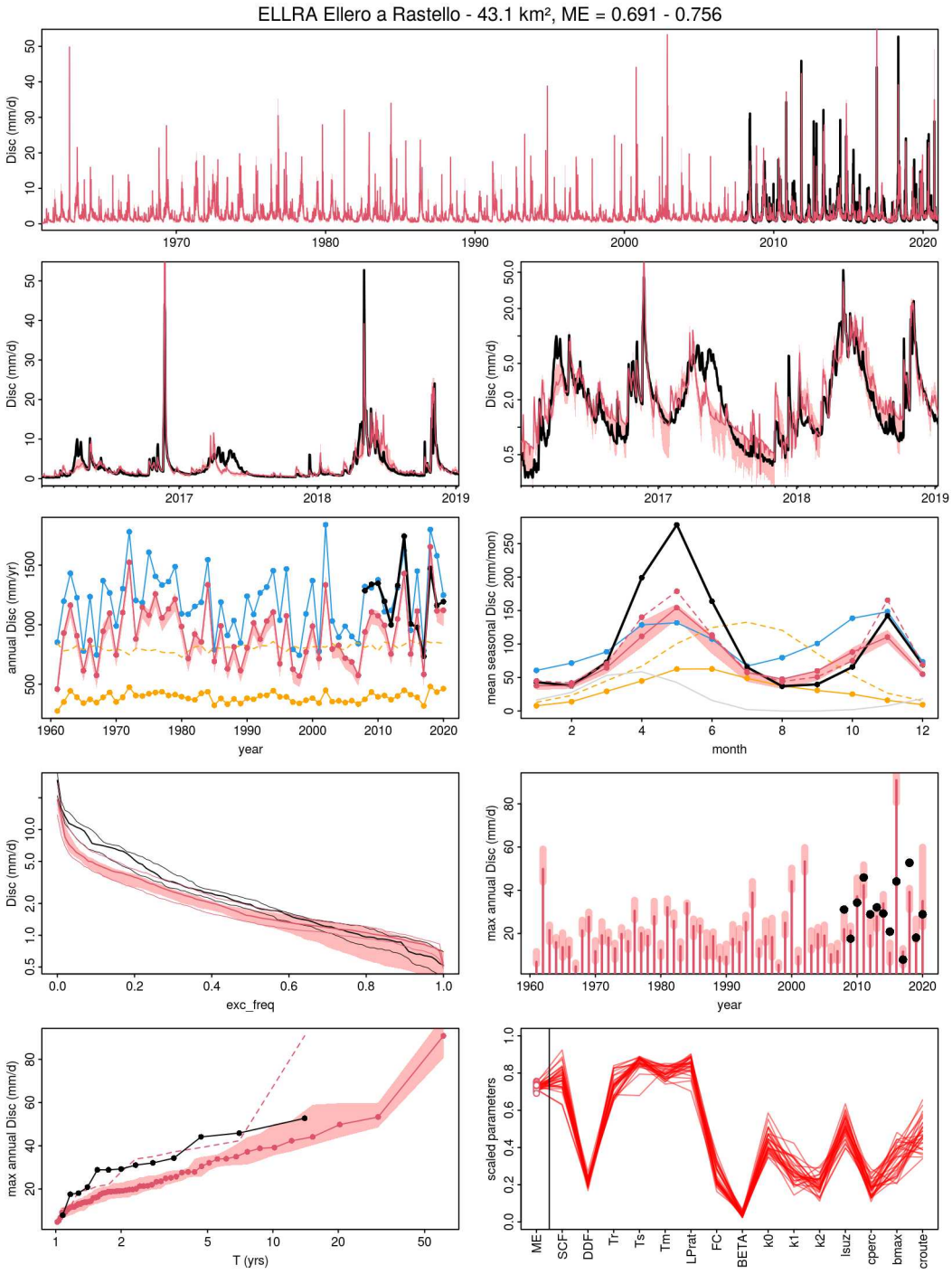


Fig. E.27 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 027.

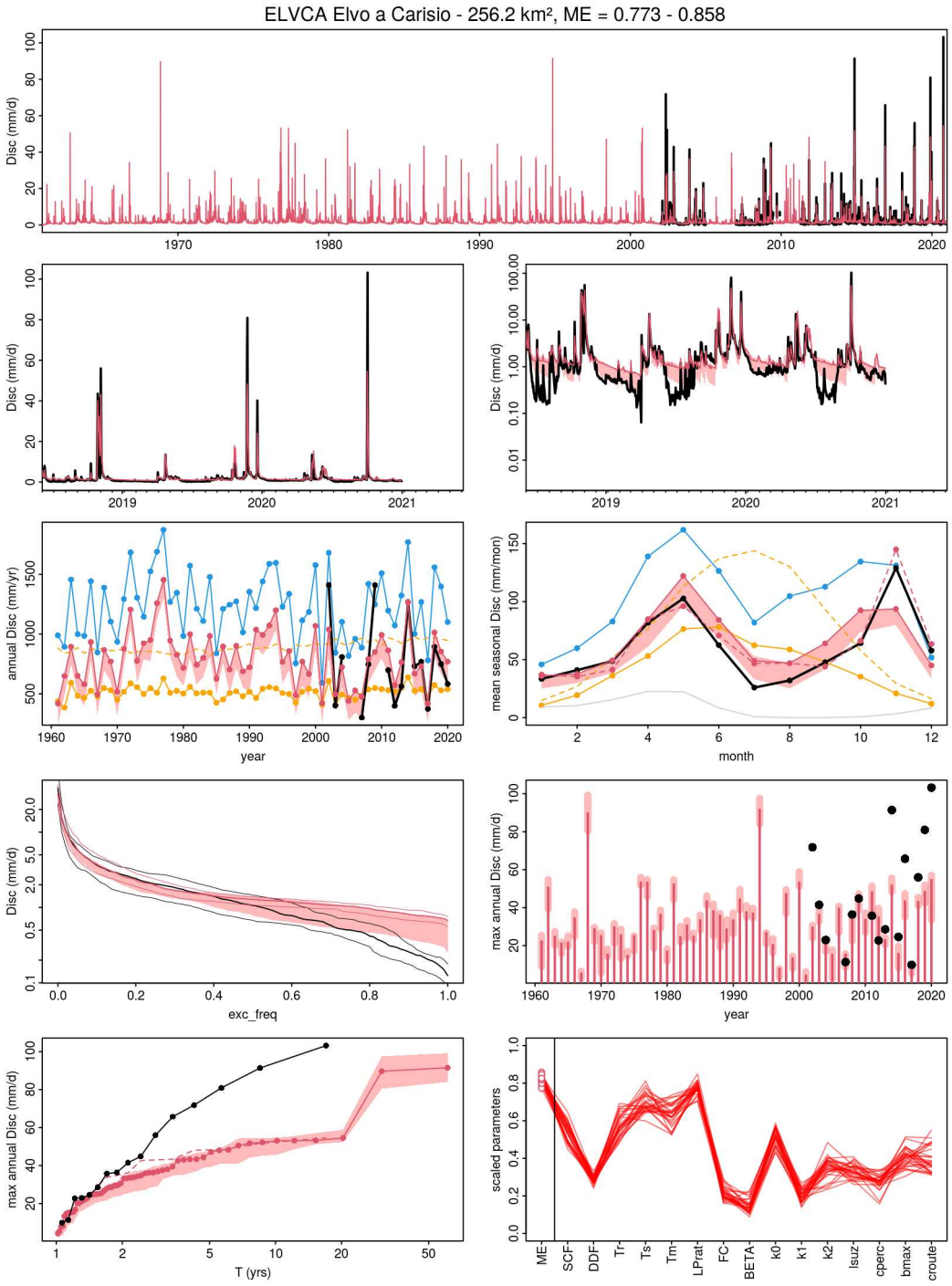


Fig. E.28 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 028.

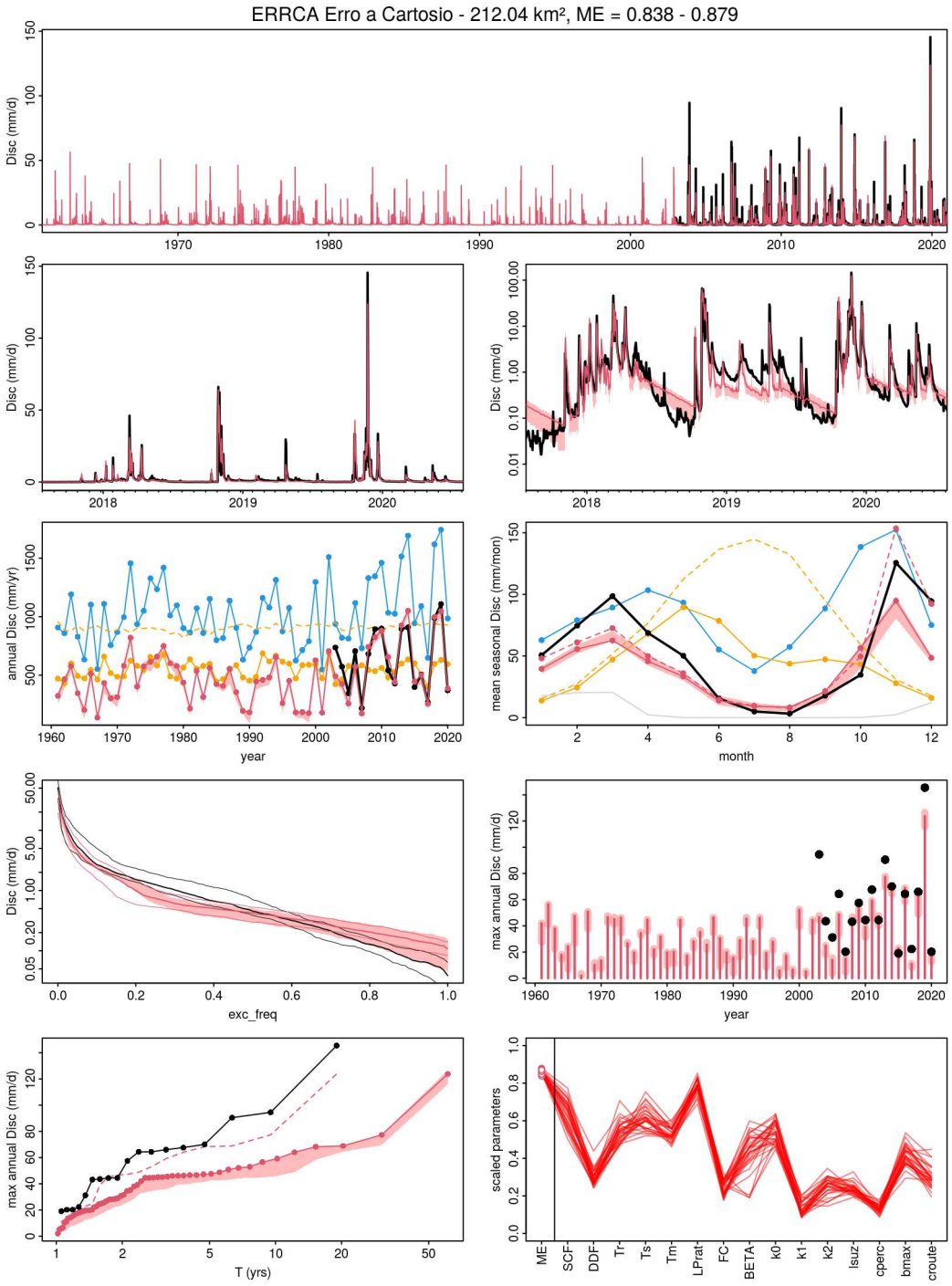


Fig. E.29 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 029.

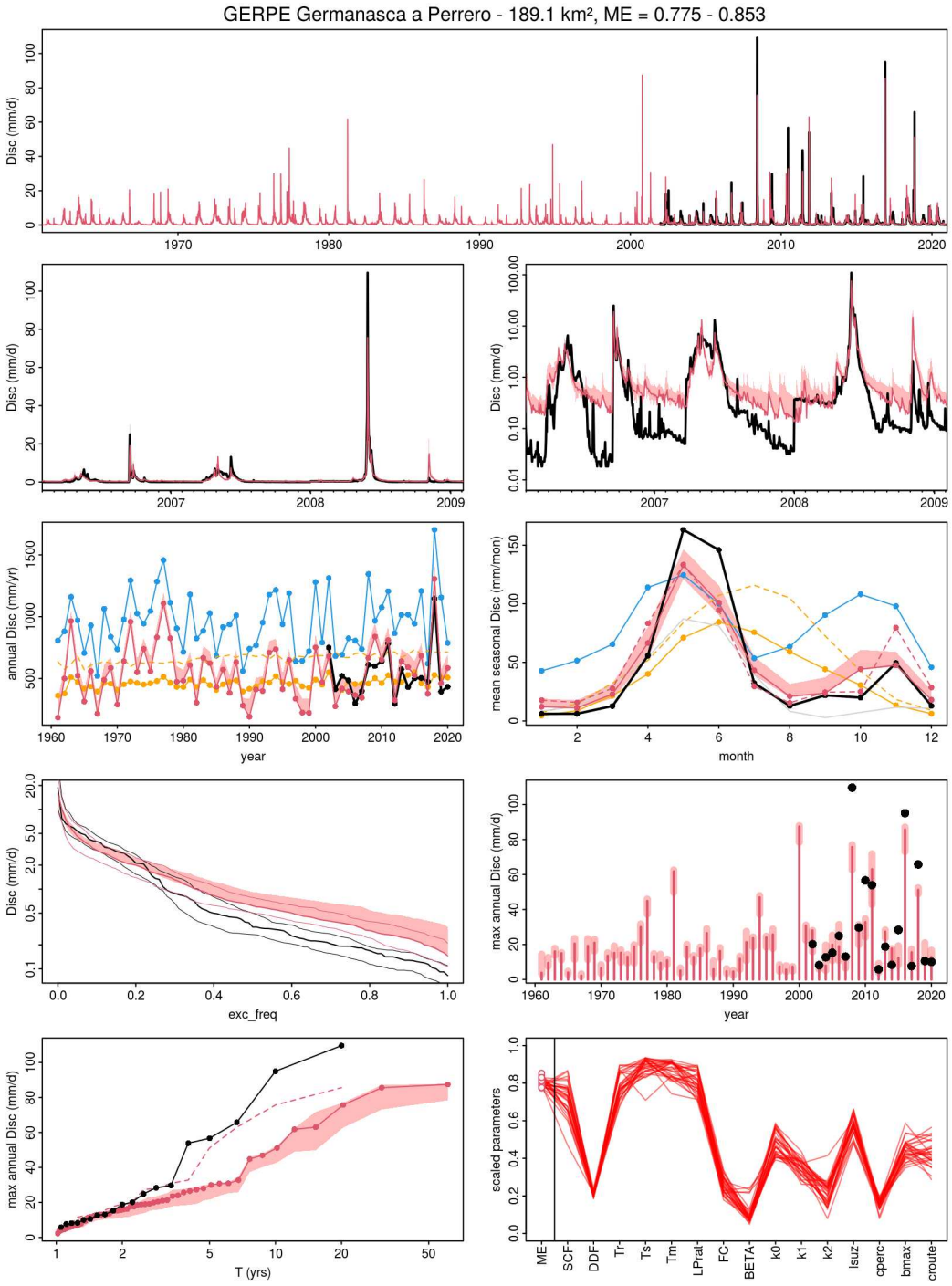


Fig. E.30 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 030.

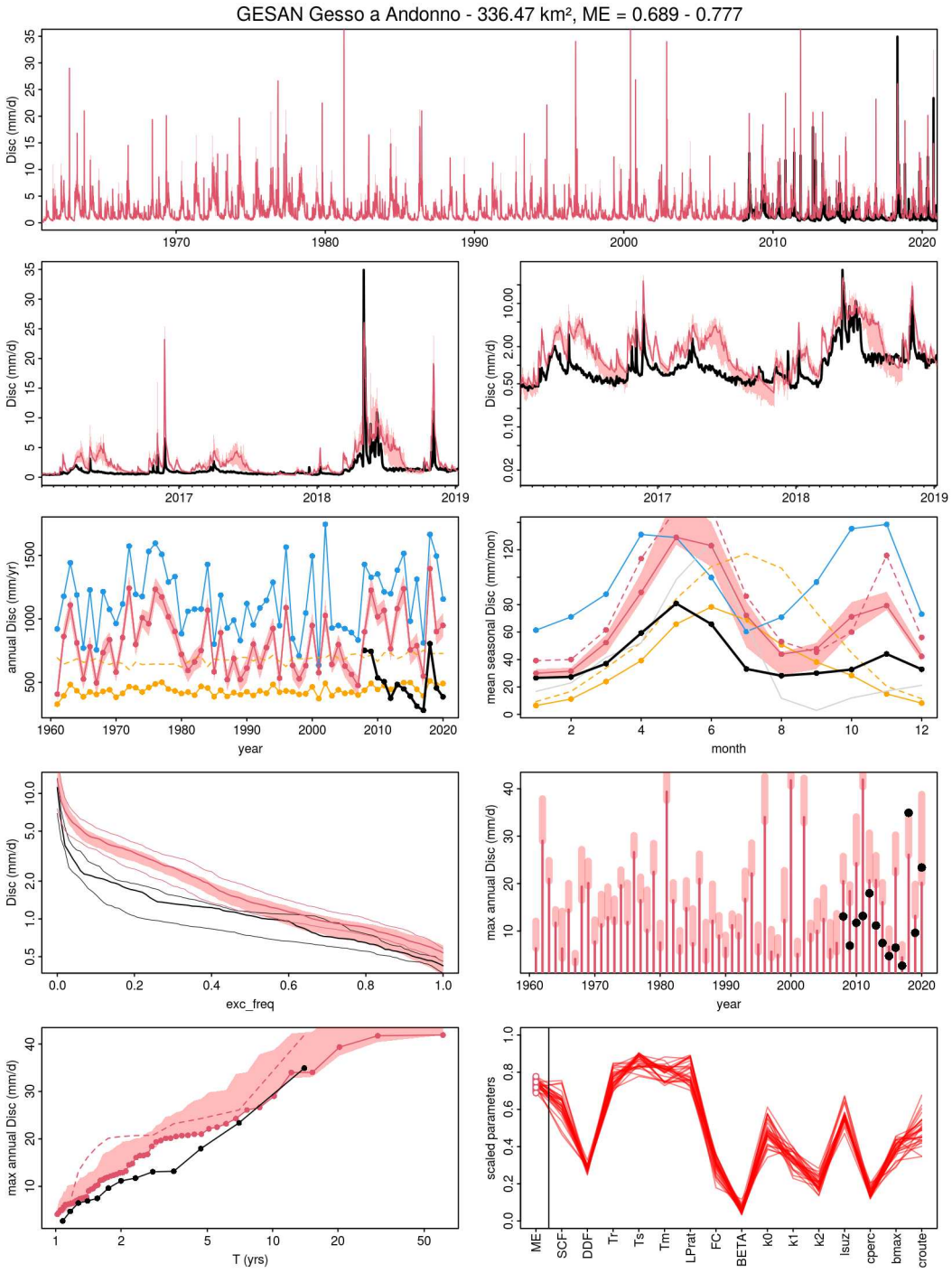


Fig. E.31 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 031.

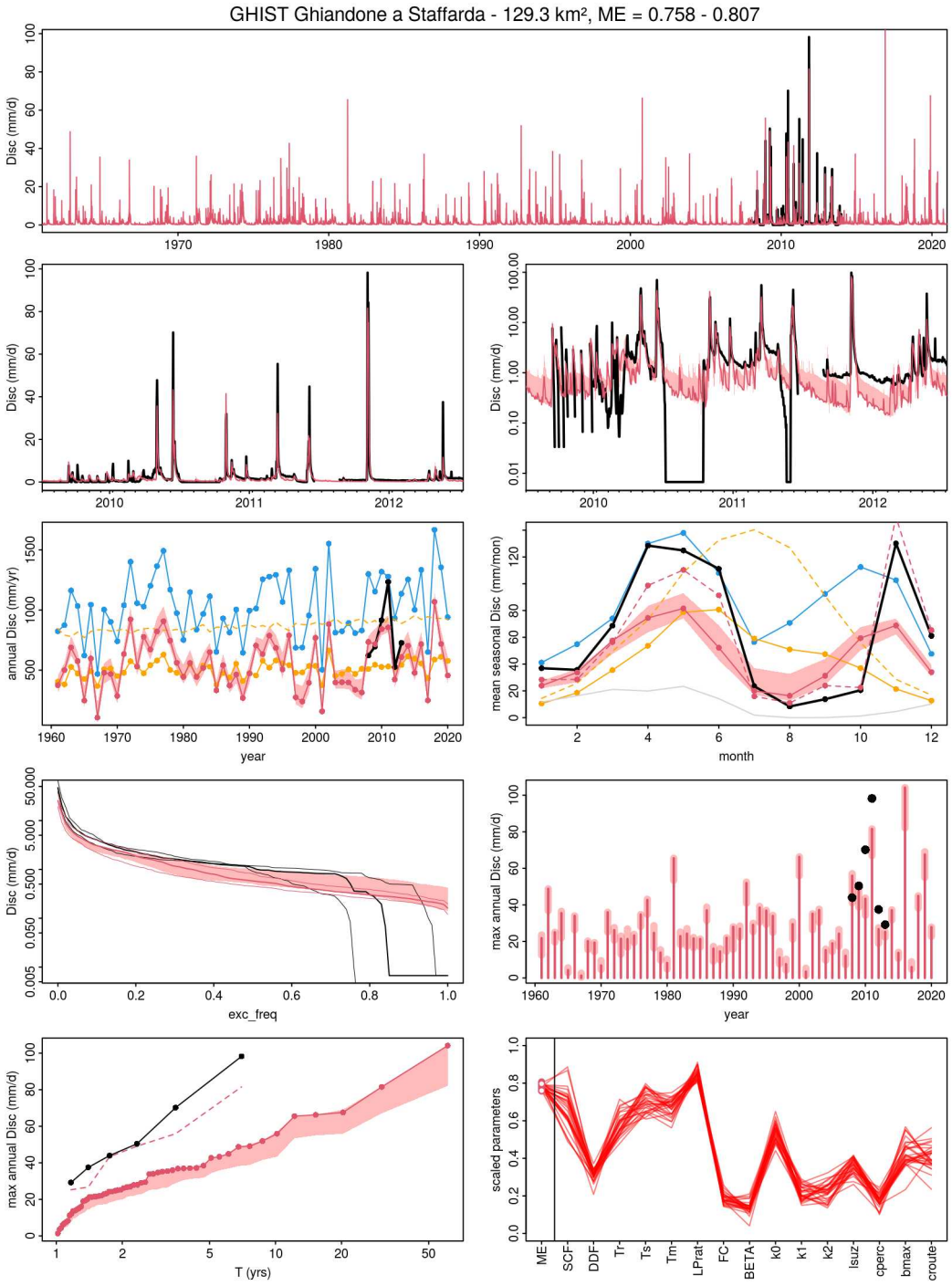


Fig. E.32 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 032.

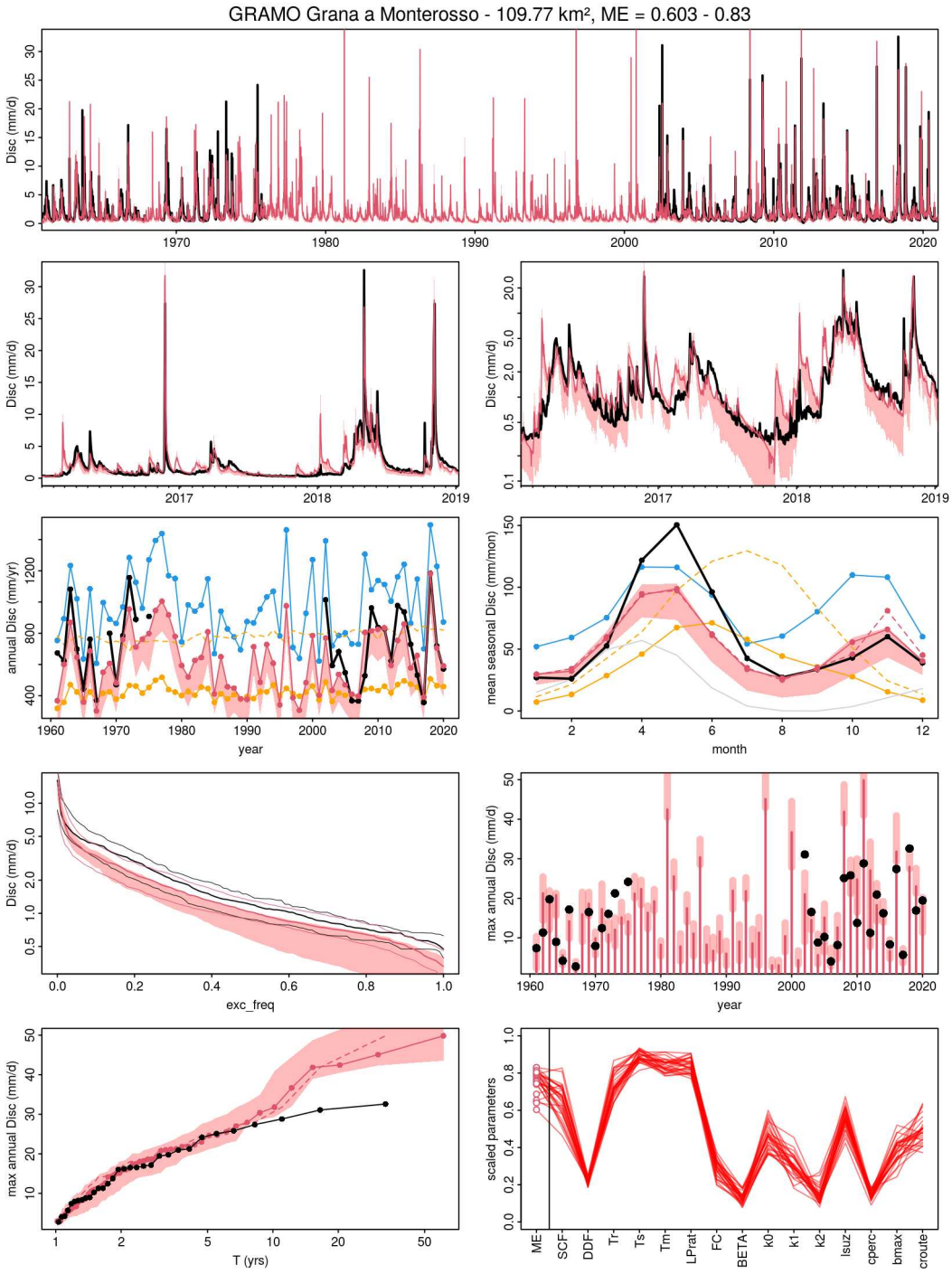


Fig. E.33 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 033.

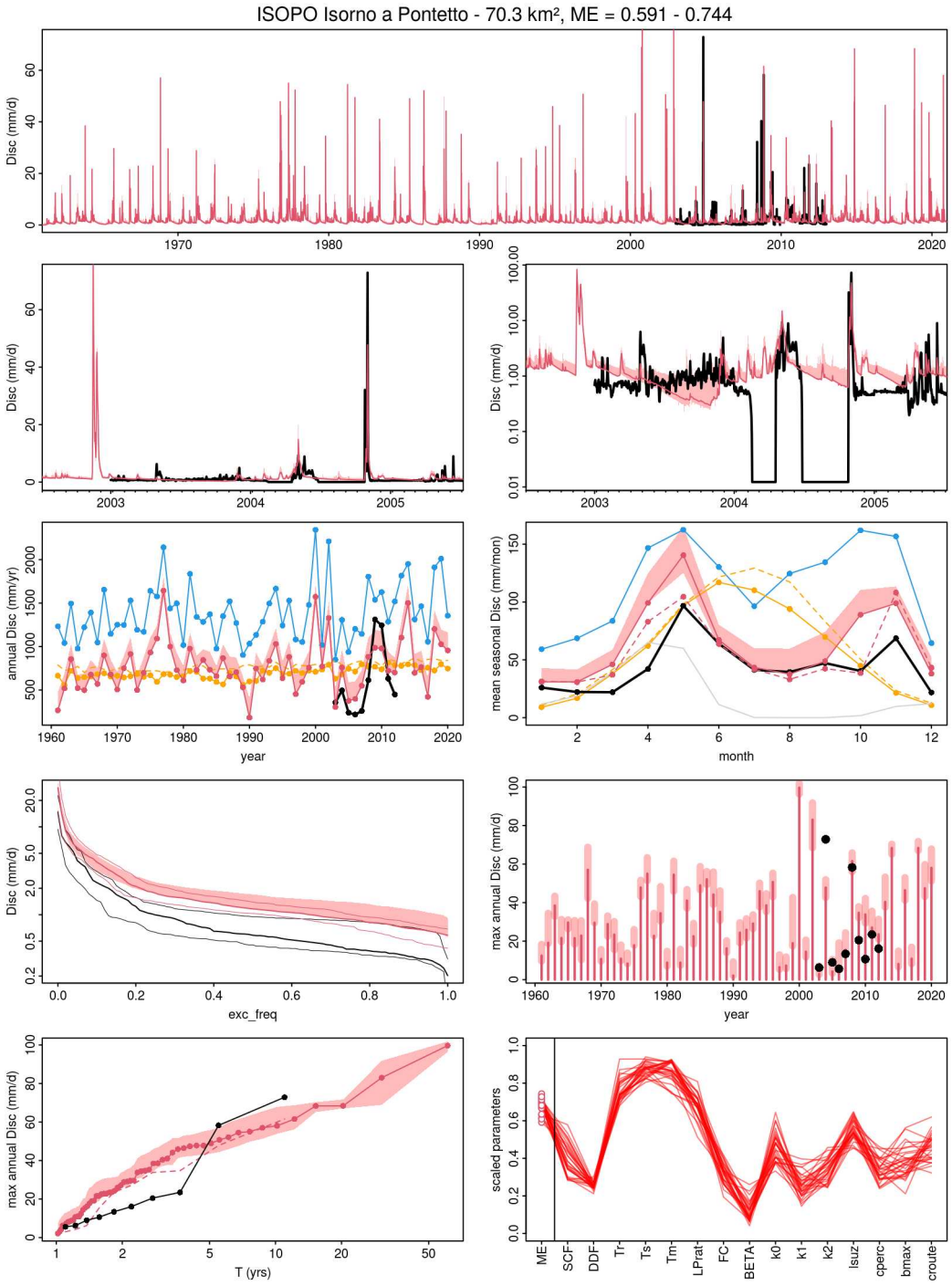


Fig. E.34 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 034.

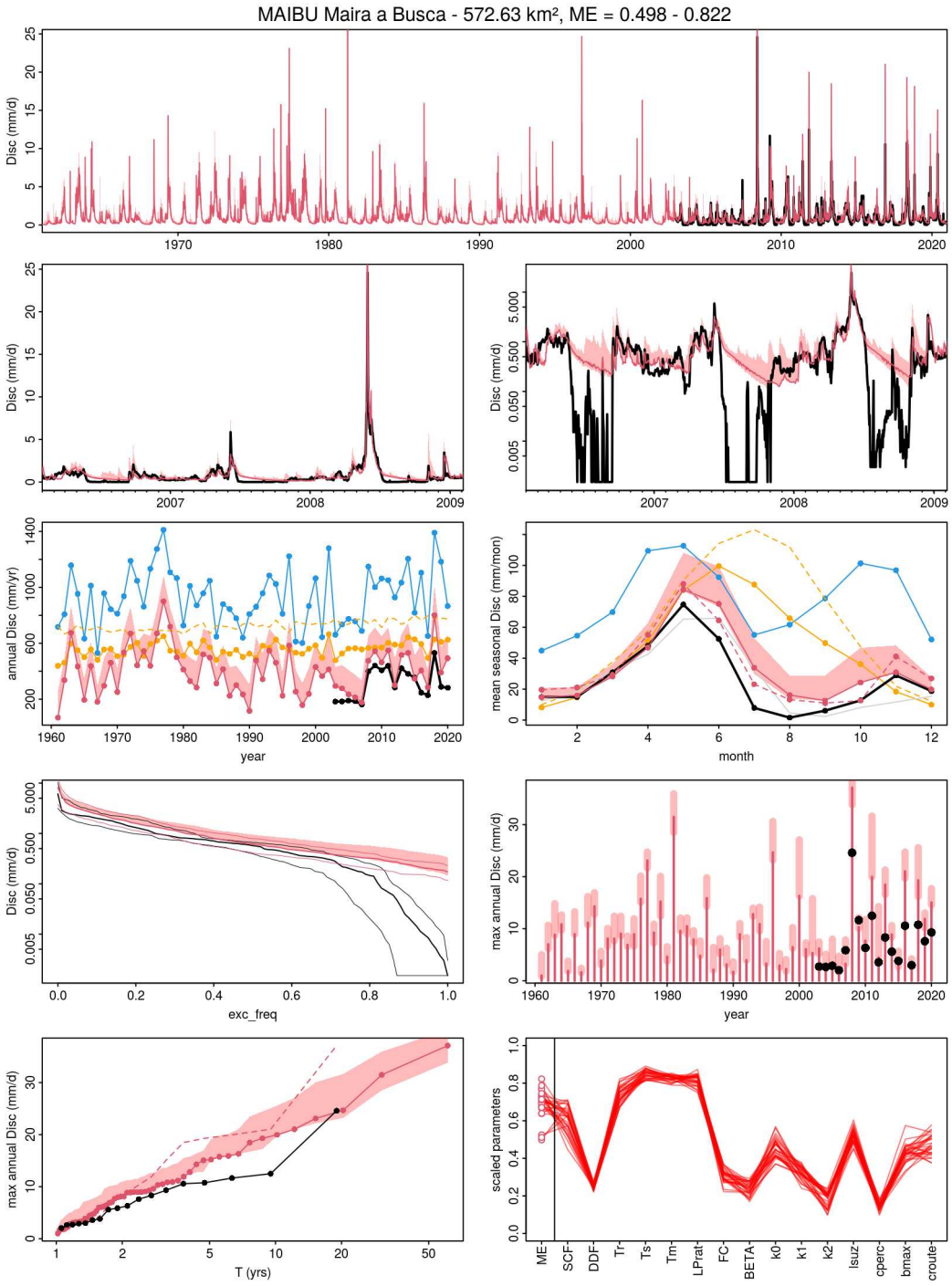


Fig. E.35 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 035.

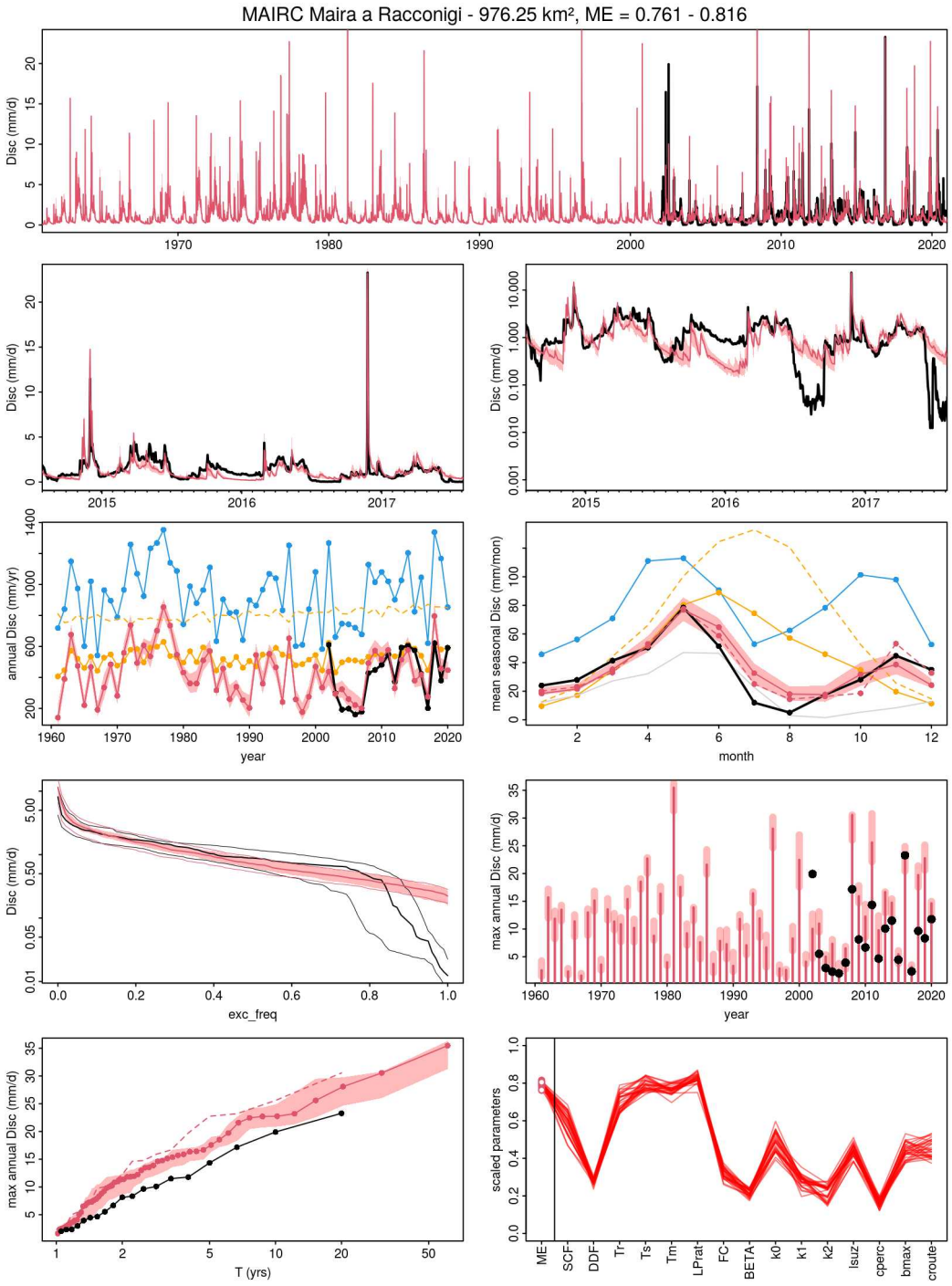


Fig. E.36 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 036.

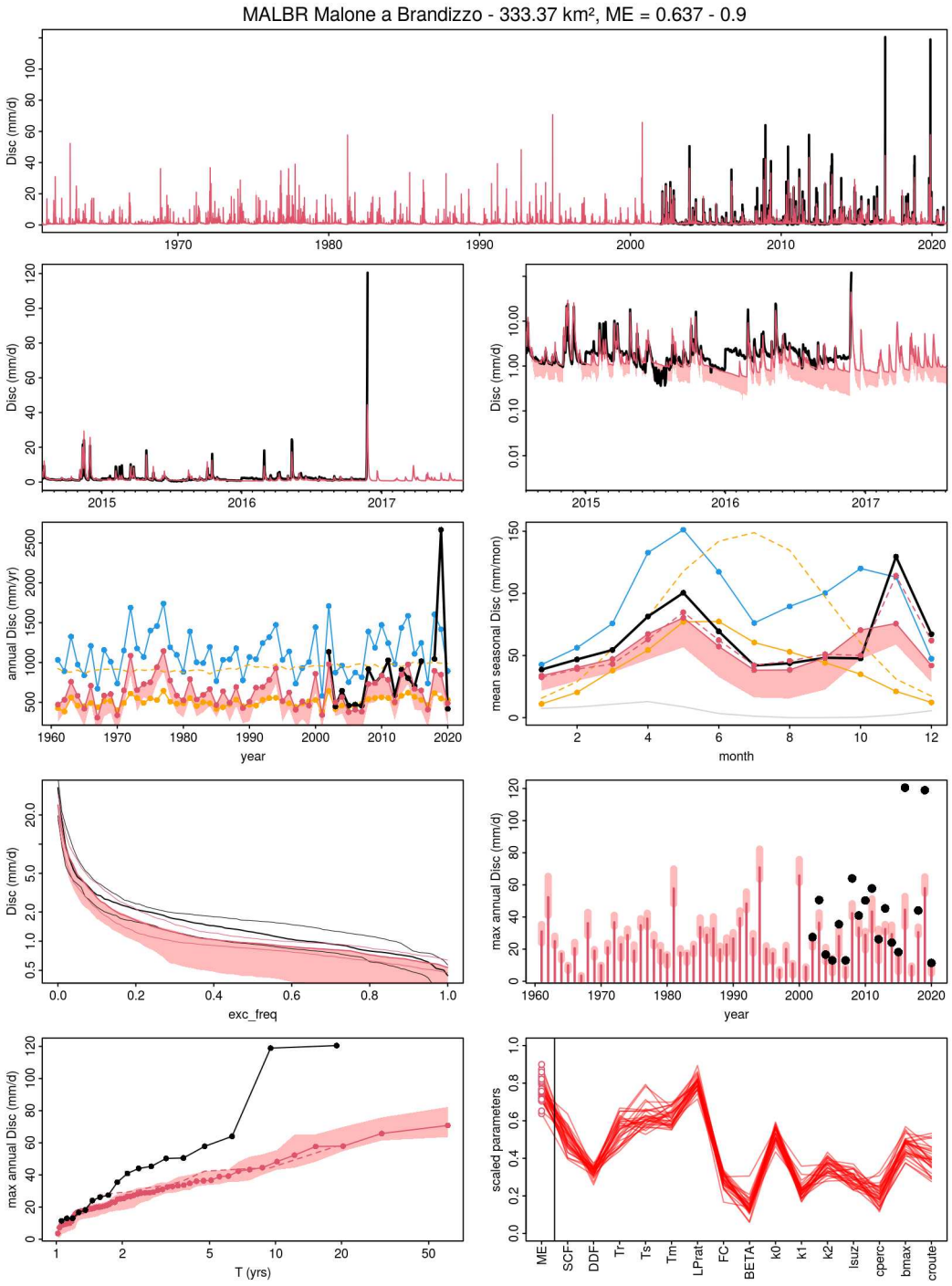


Fig. E.37 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 037.

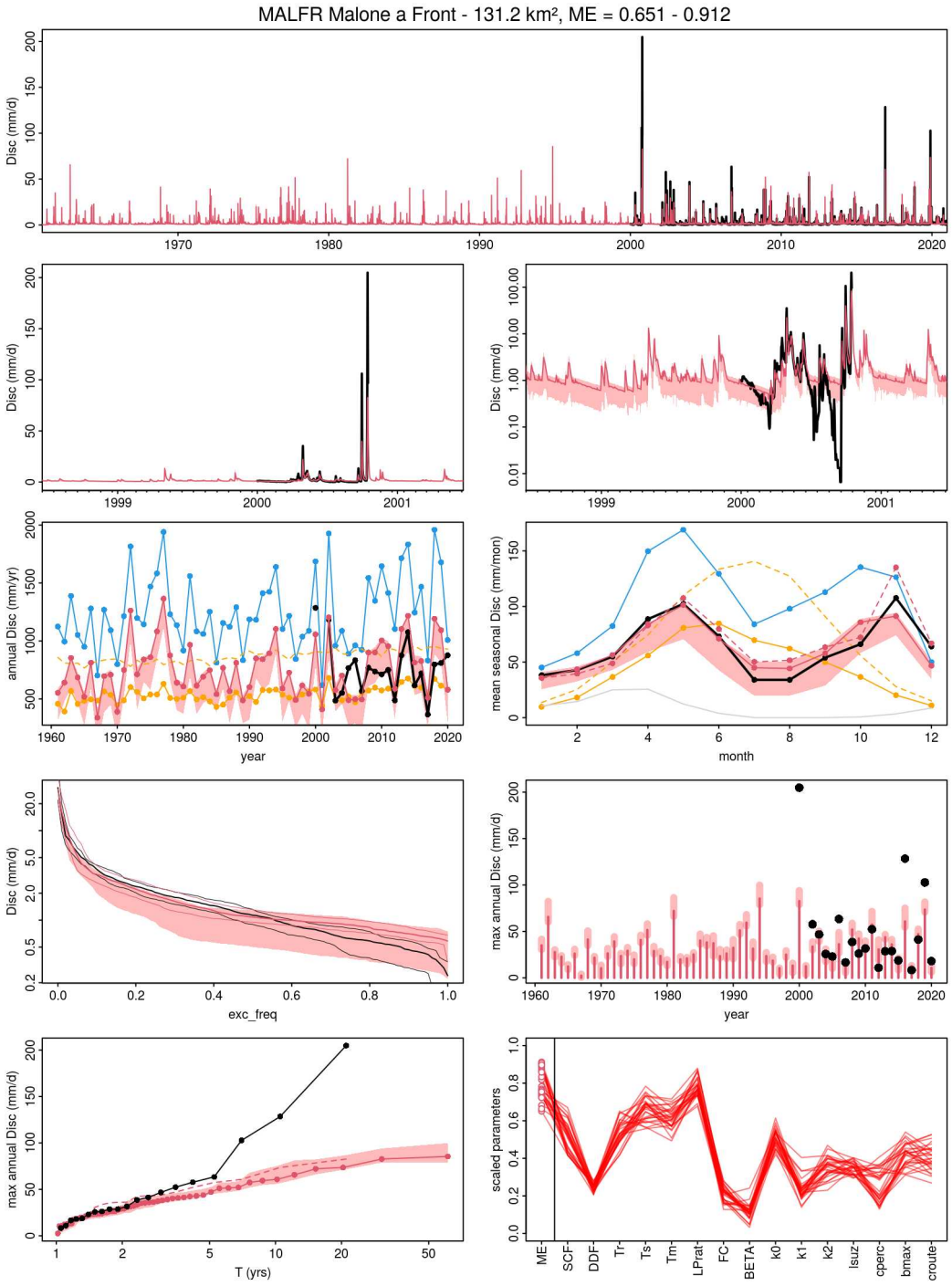


Fig. E.38 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 038.

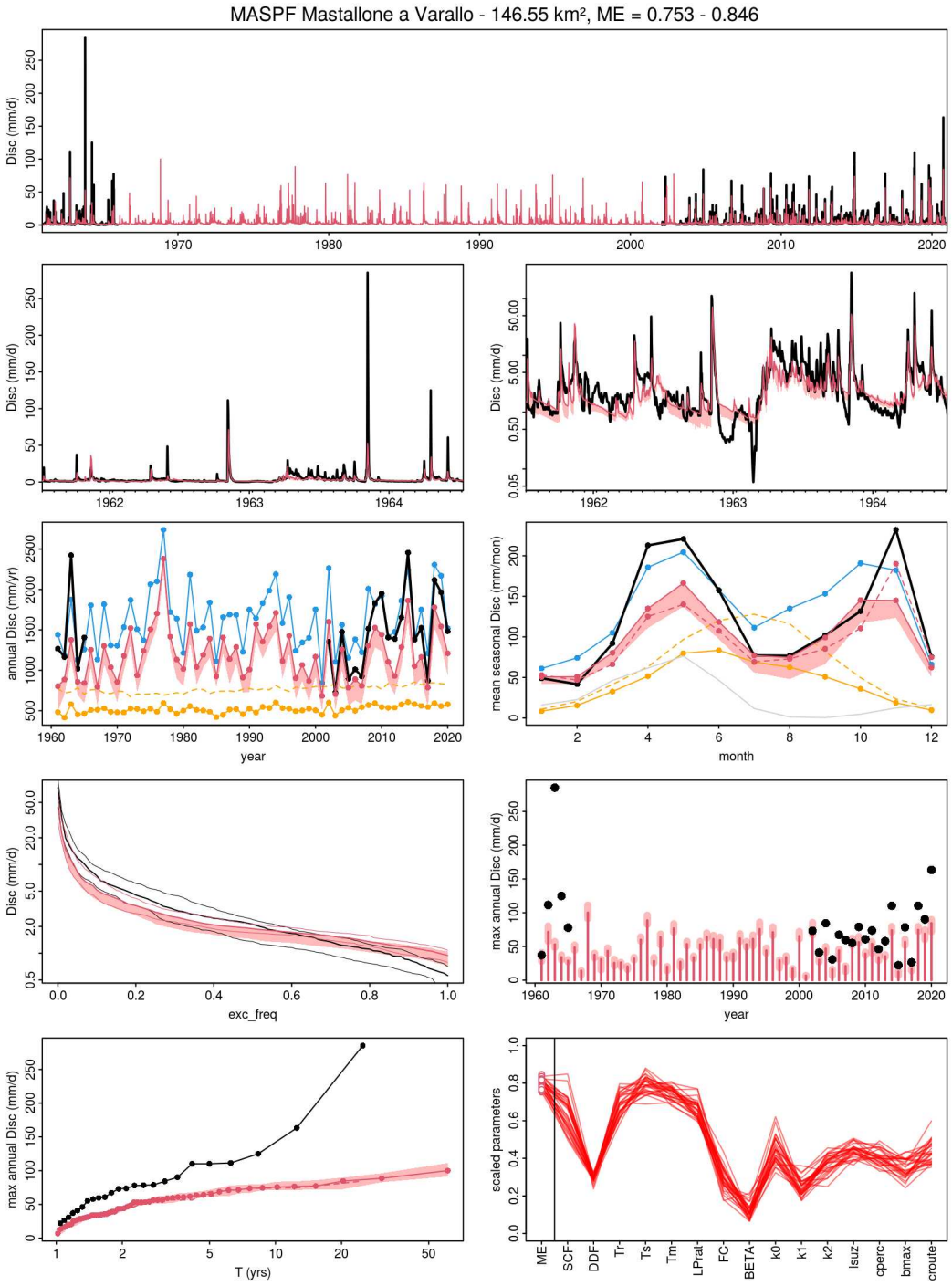


Fig. E.39 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 039.

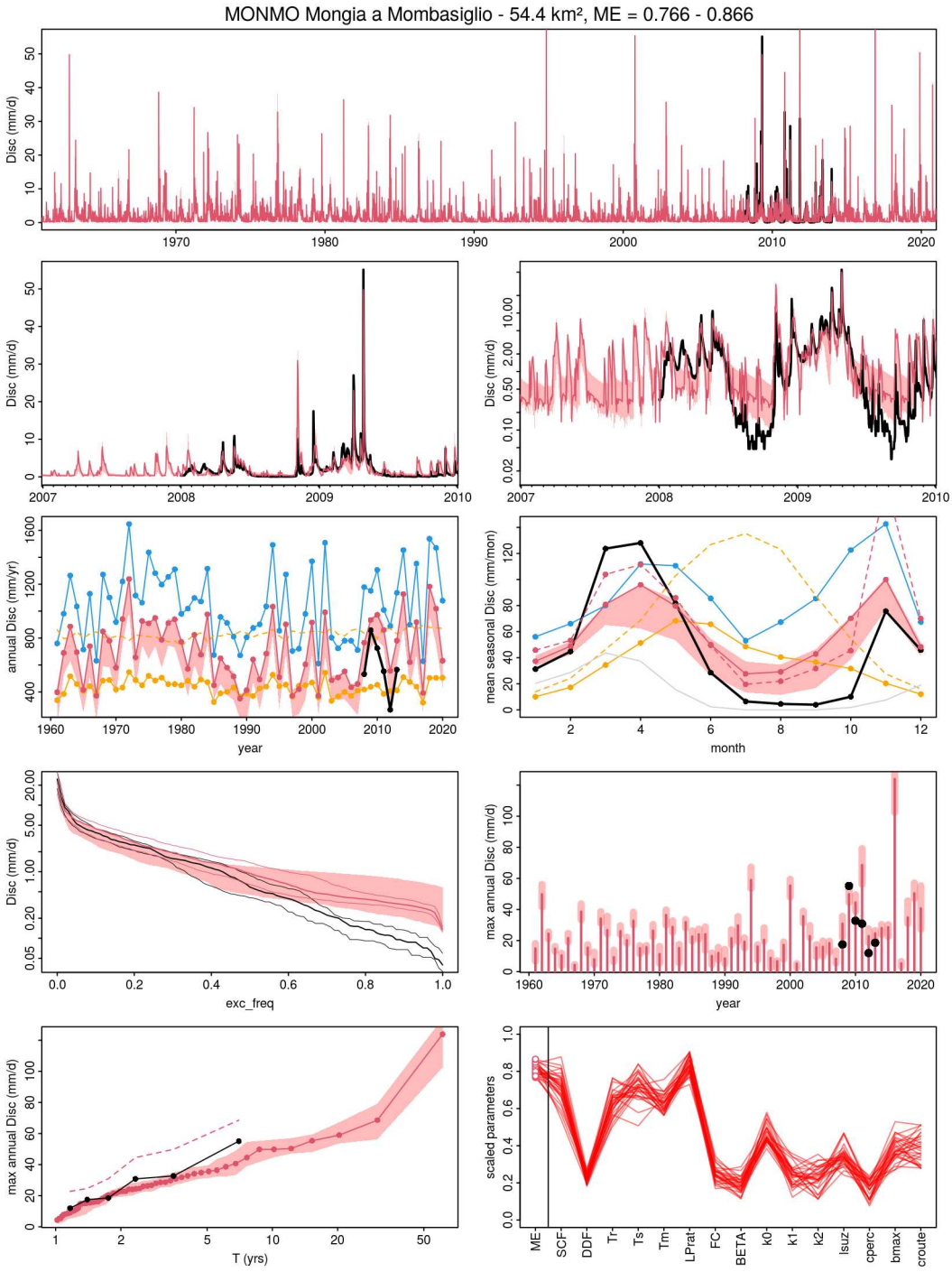


Fig. E.40 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 040.

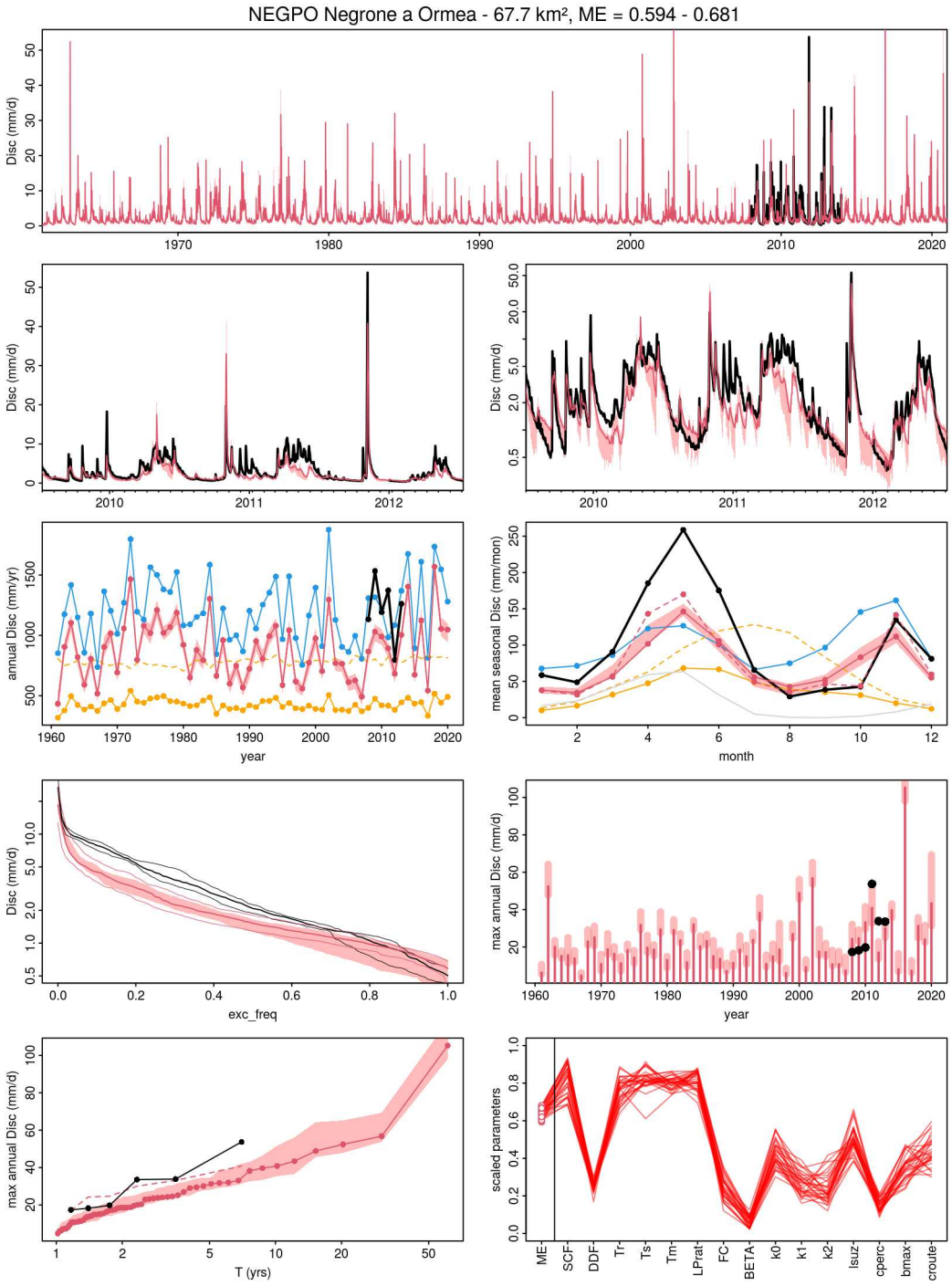


Fig. E.41 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 041.

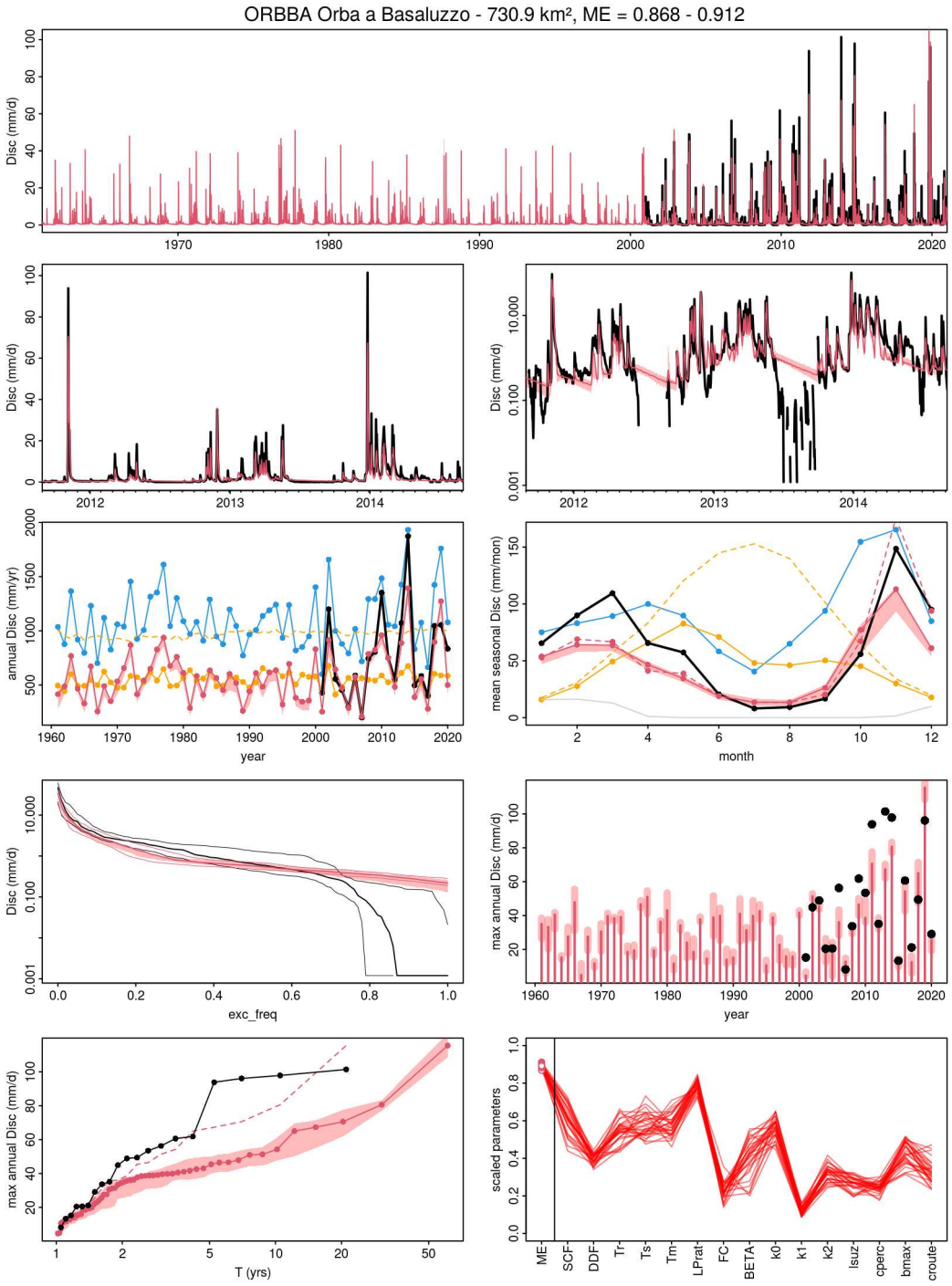


Fig. E.42 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 042.

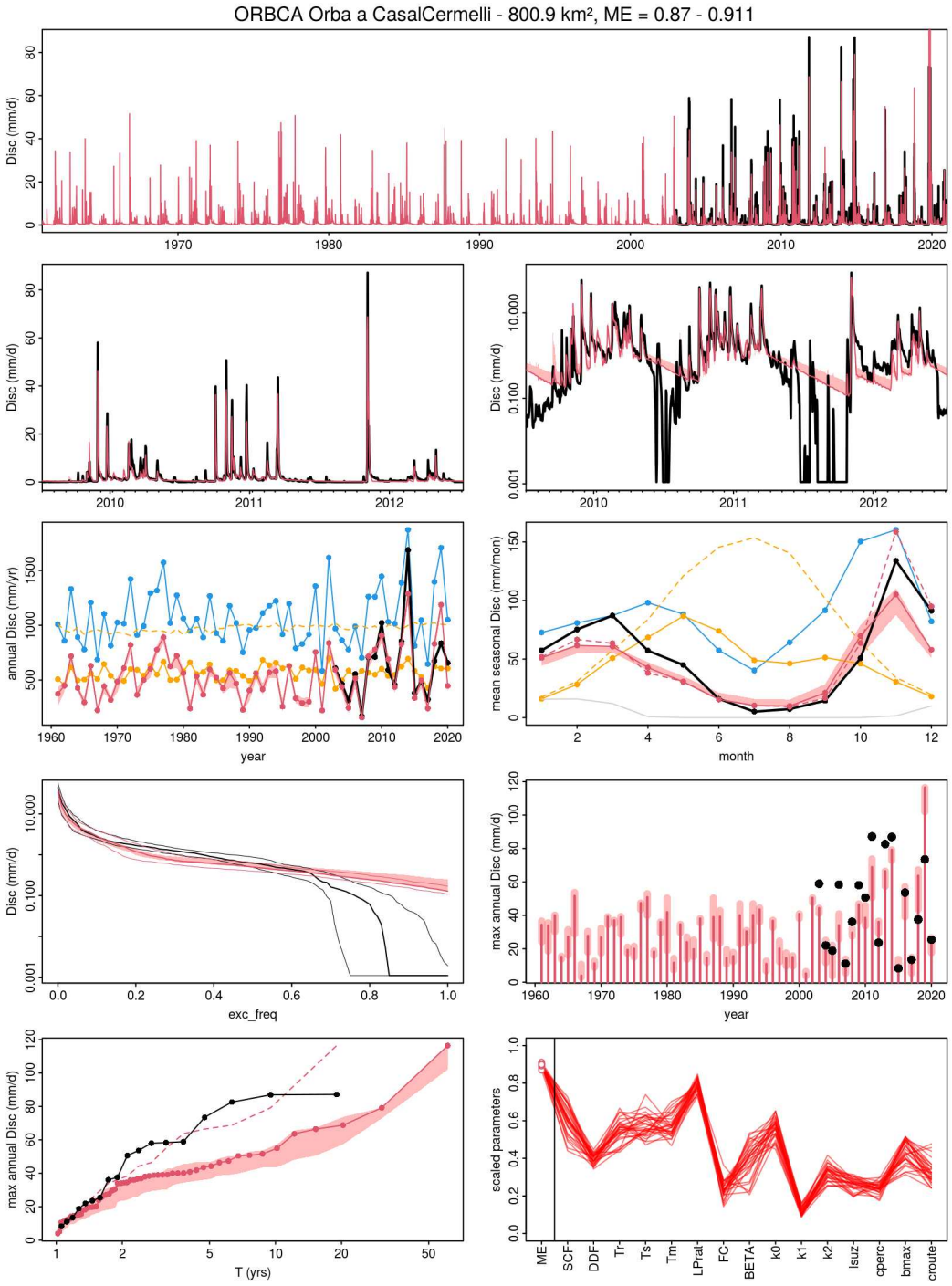


Fig. E.43 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 043.

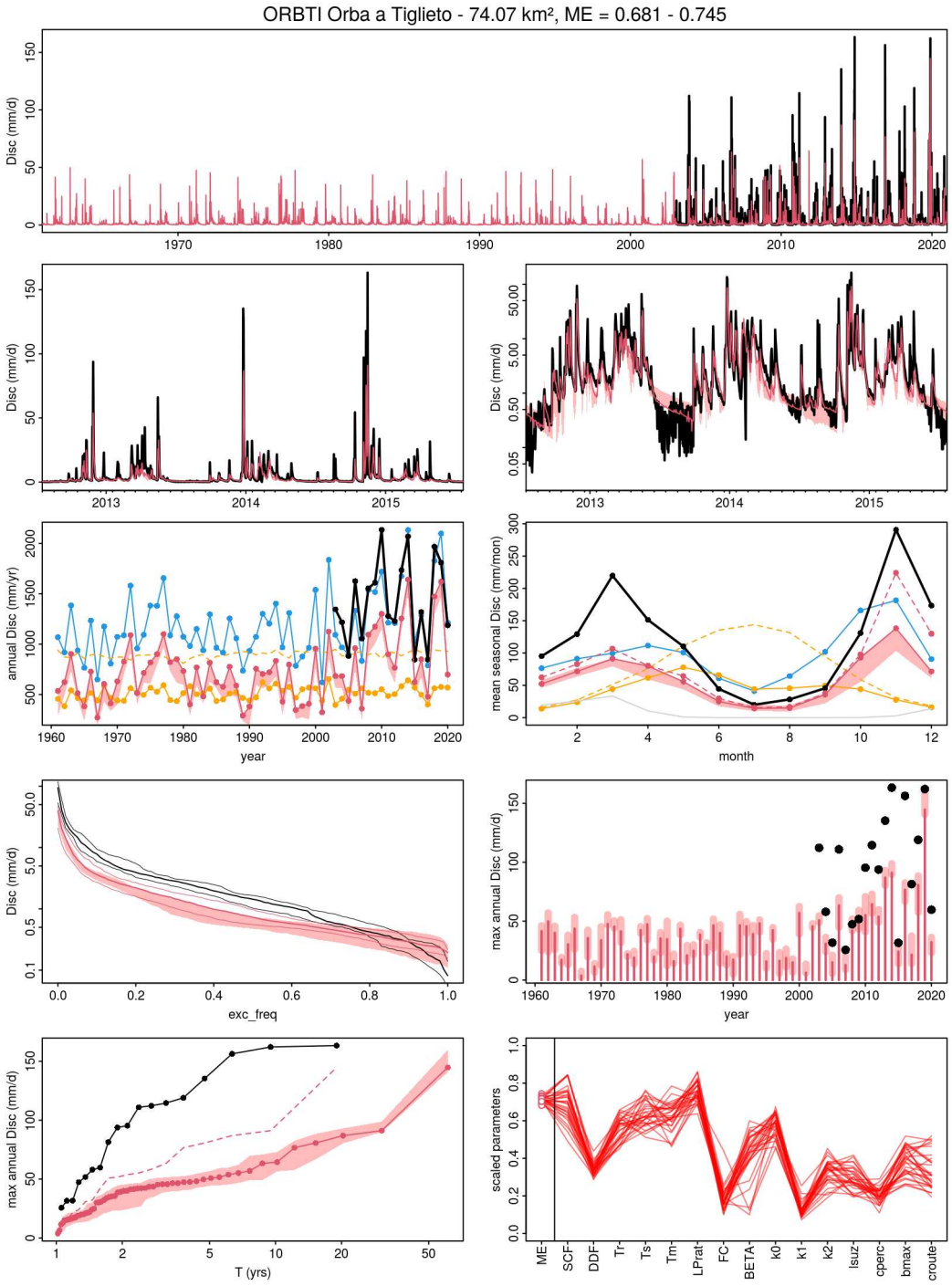


Fig. E.44 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 044.

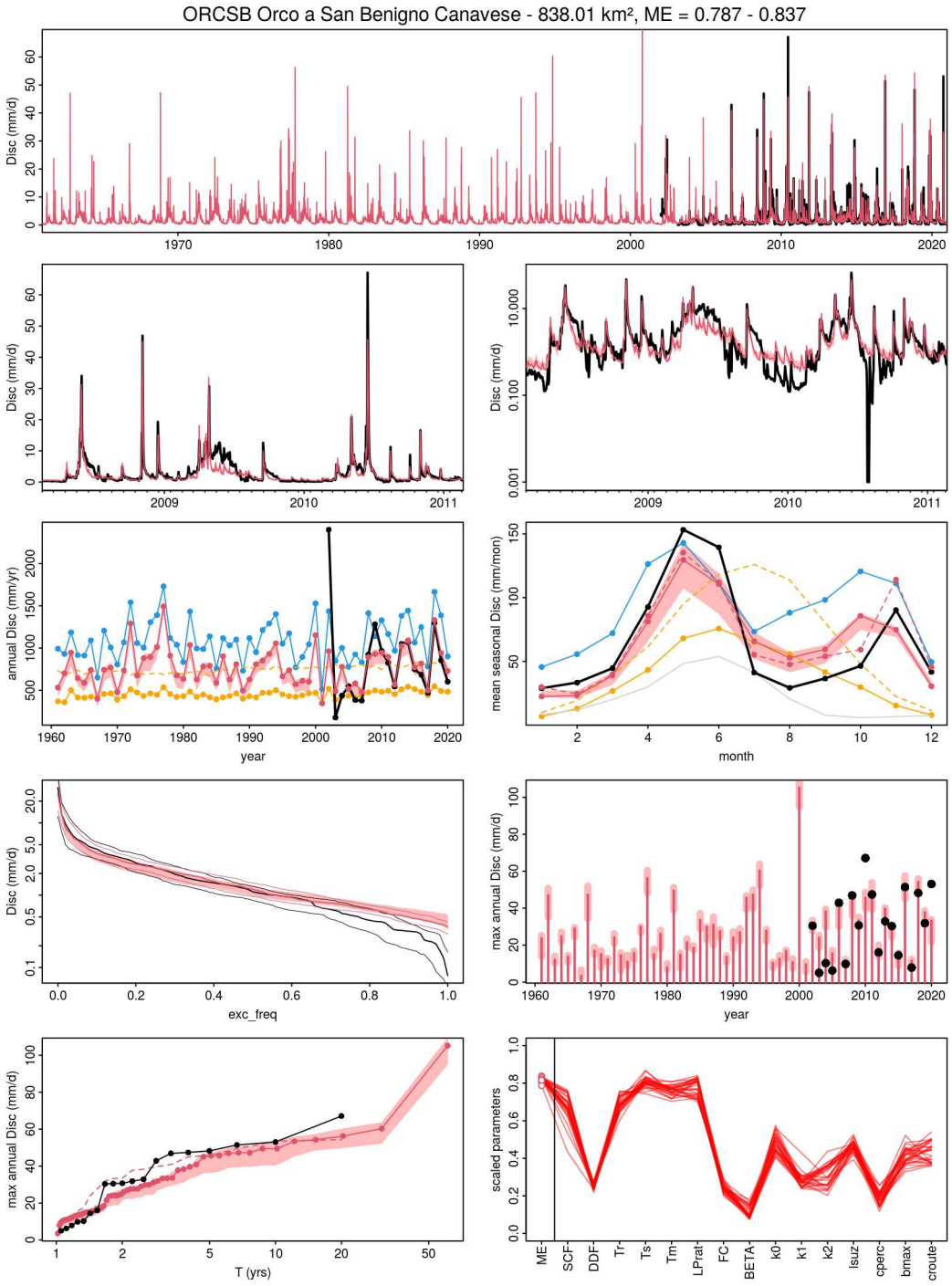


Fig. E.45 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 045.

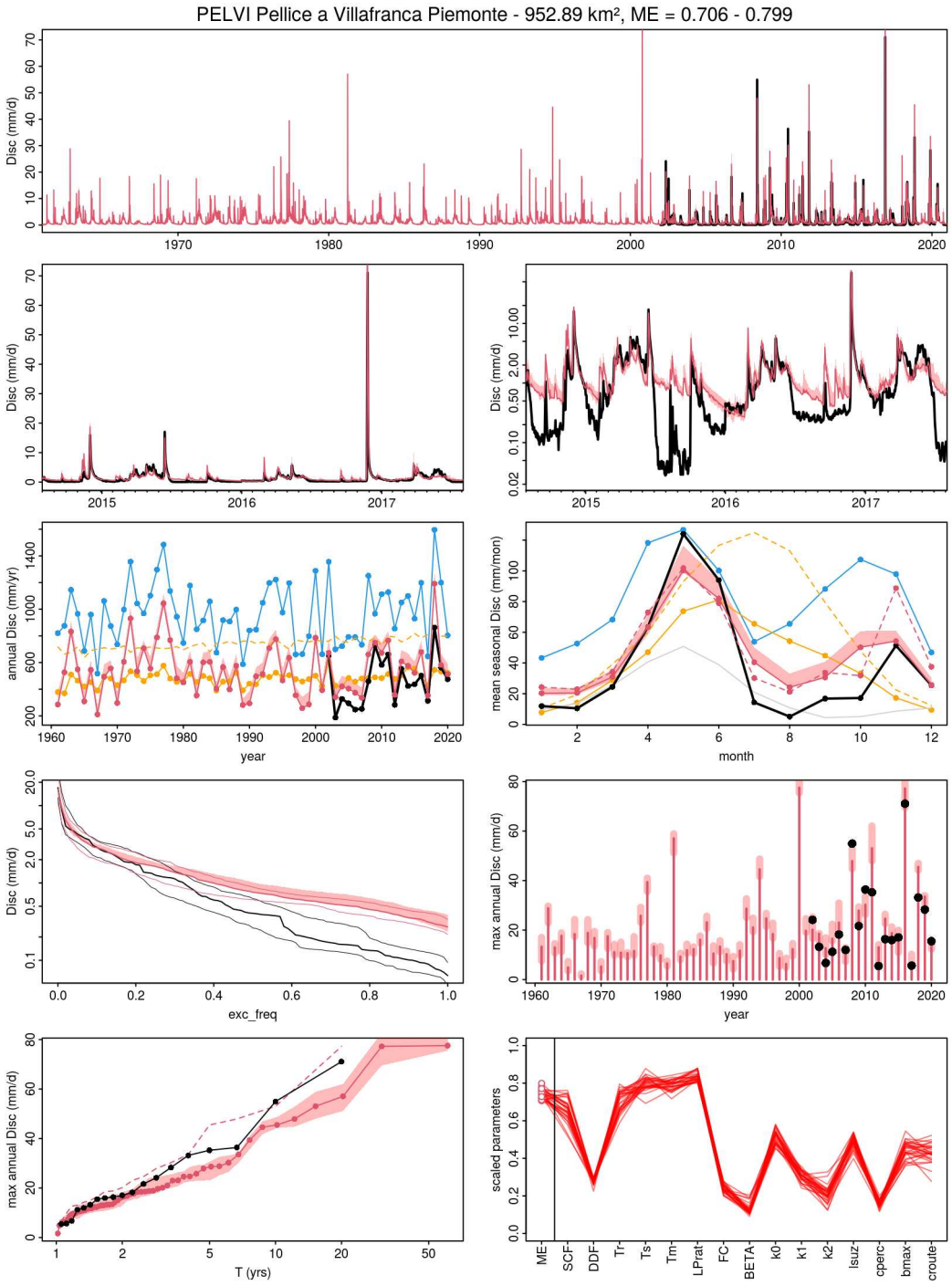


Fig. E.46 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 046.

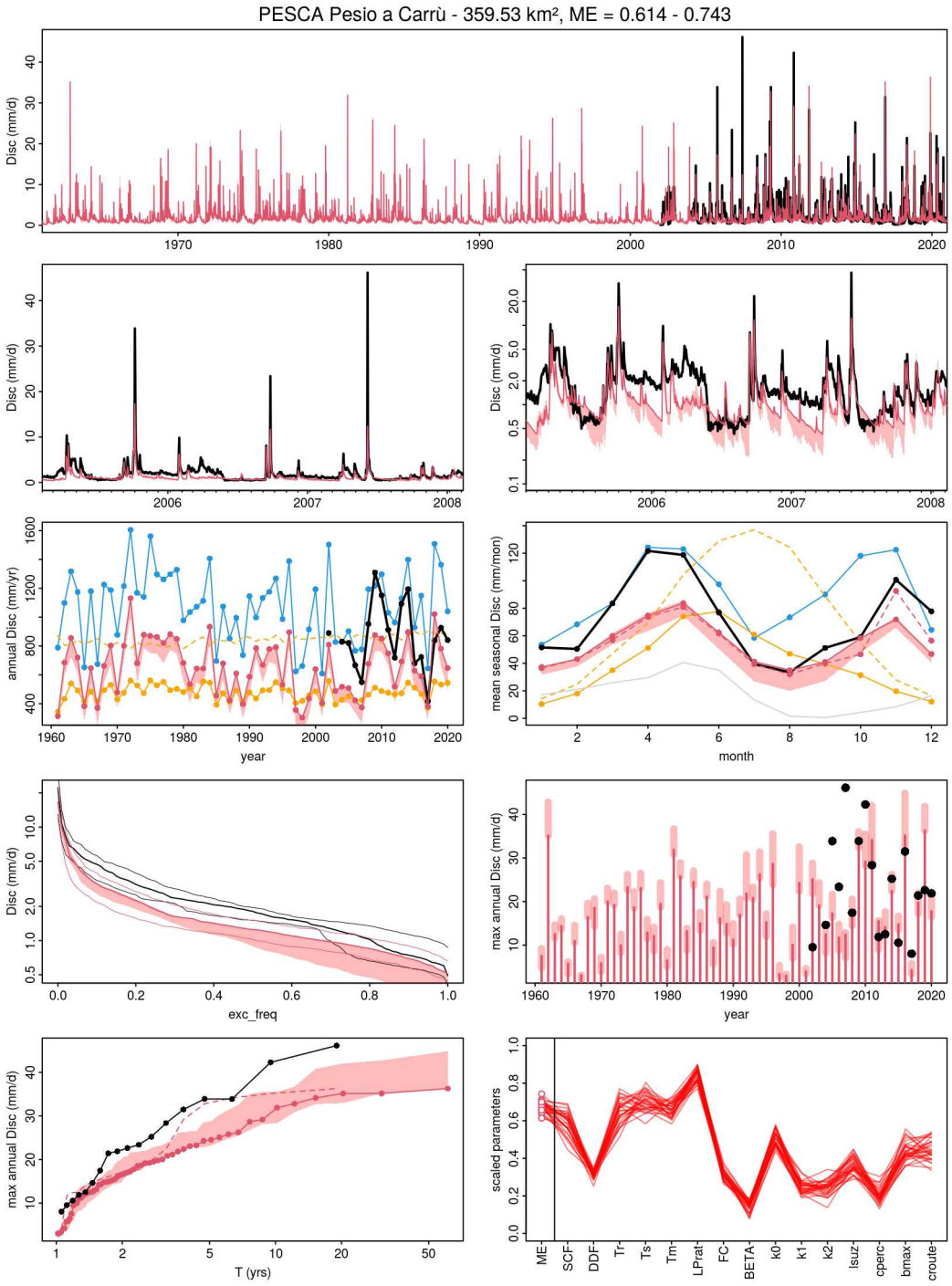


Fig. E.47 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 047.

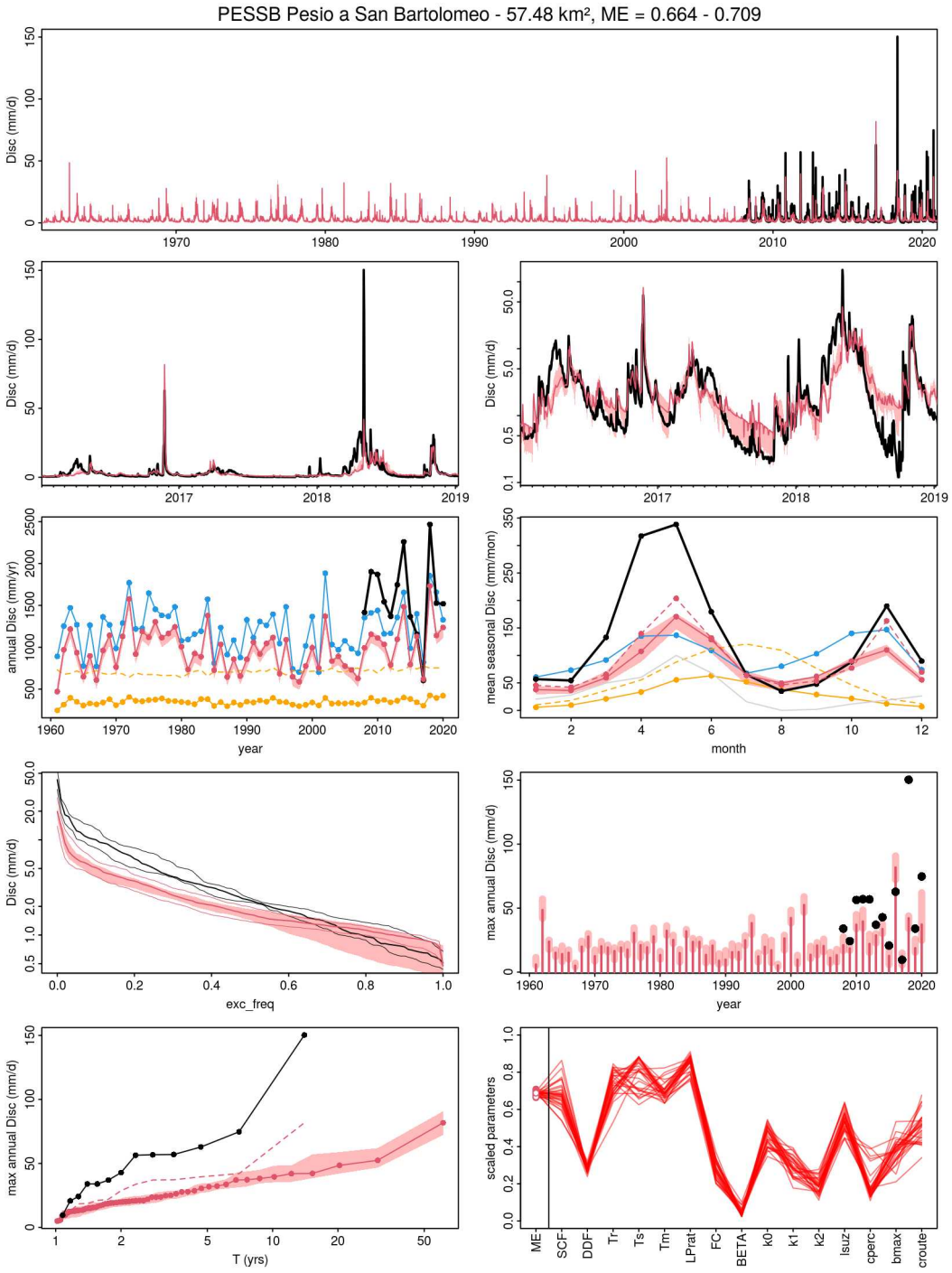


Fig. E.48 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 048.

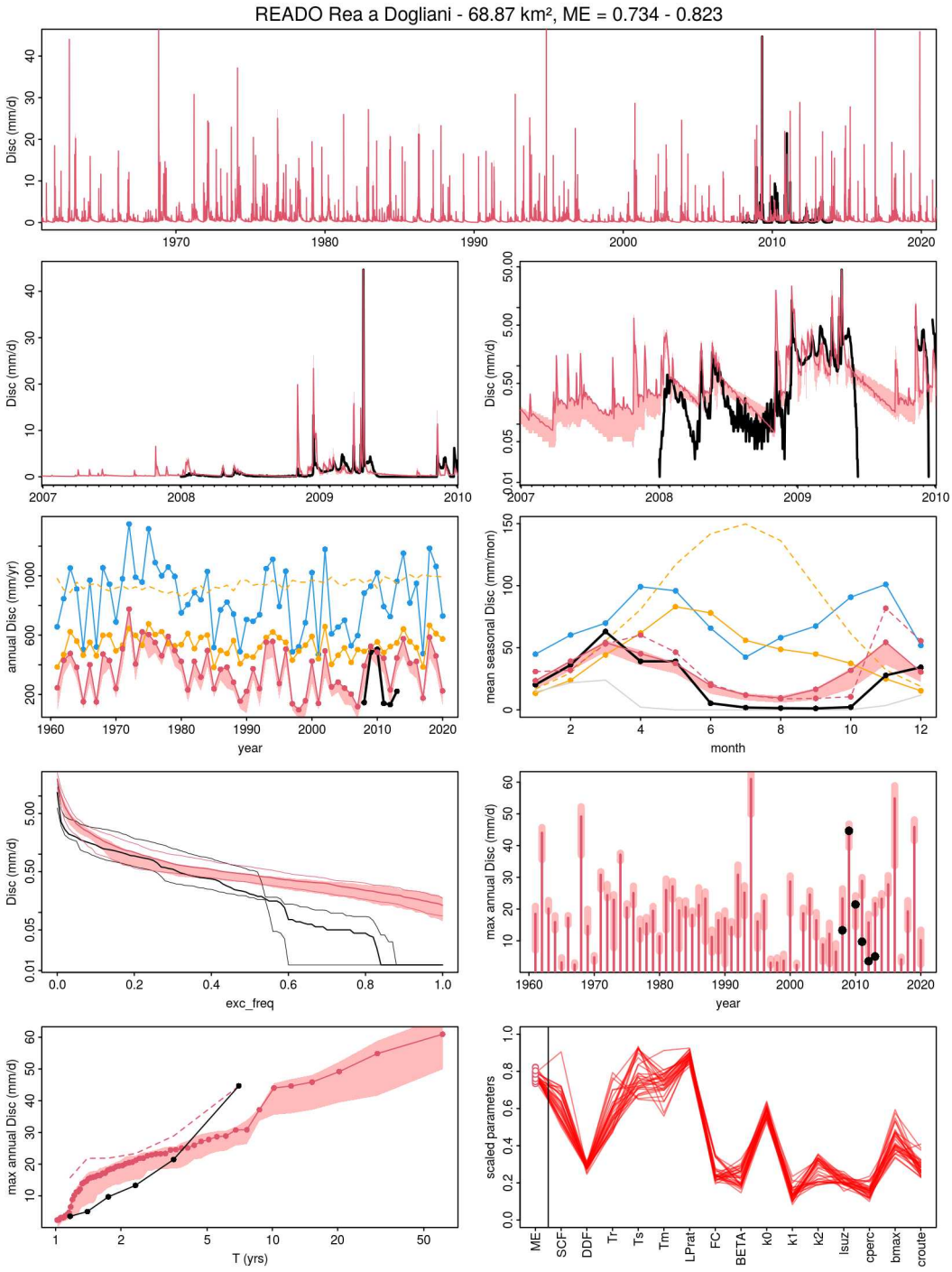


Fig. E.49 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 049.

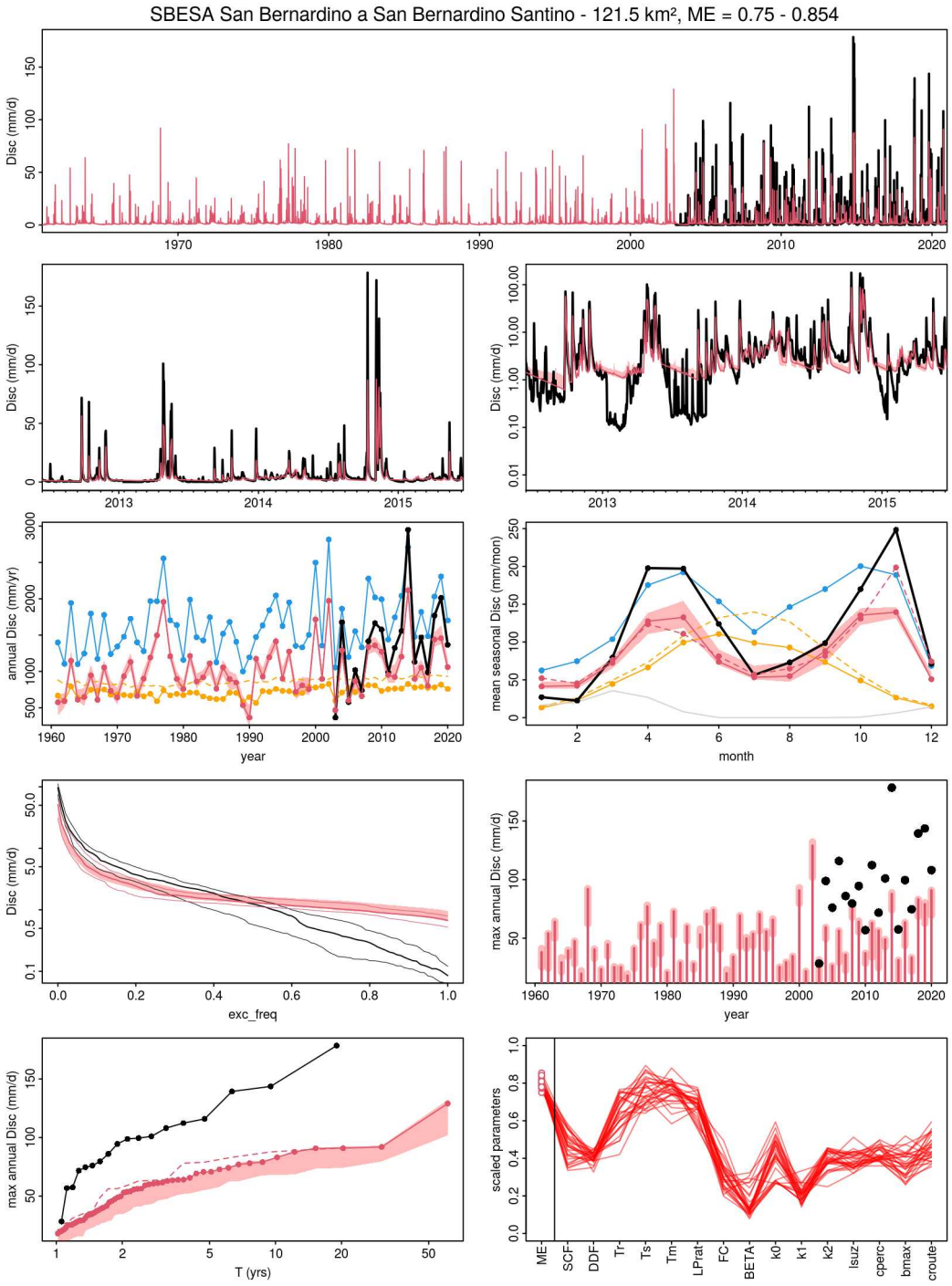


Fig. E.50 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 050.

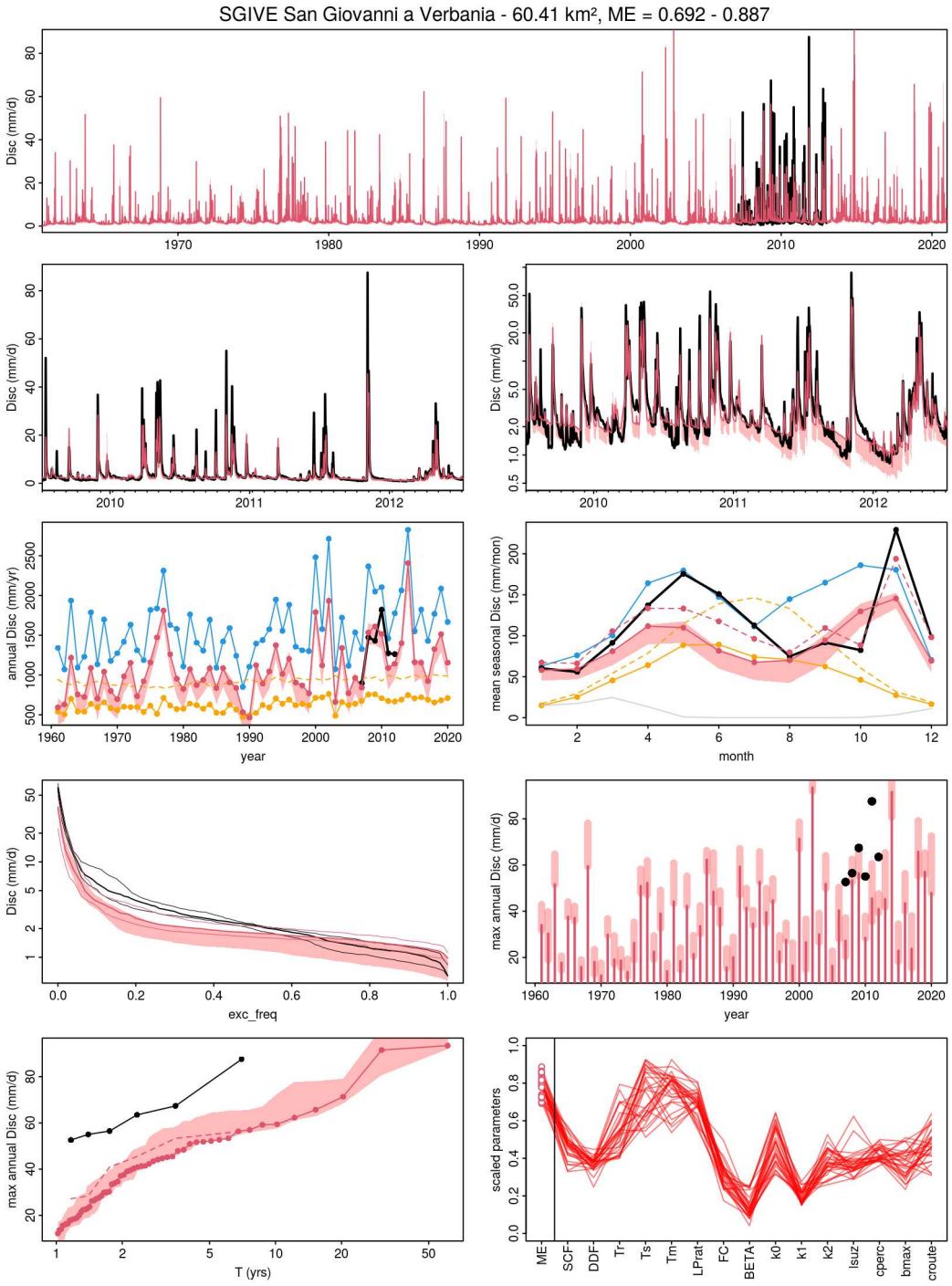


Fig. E.51 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 051.

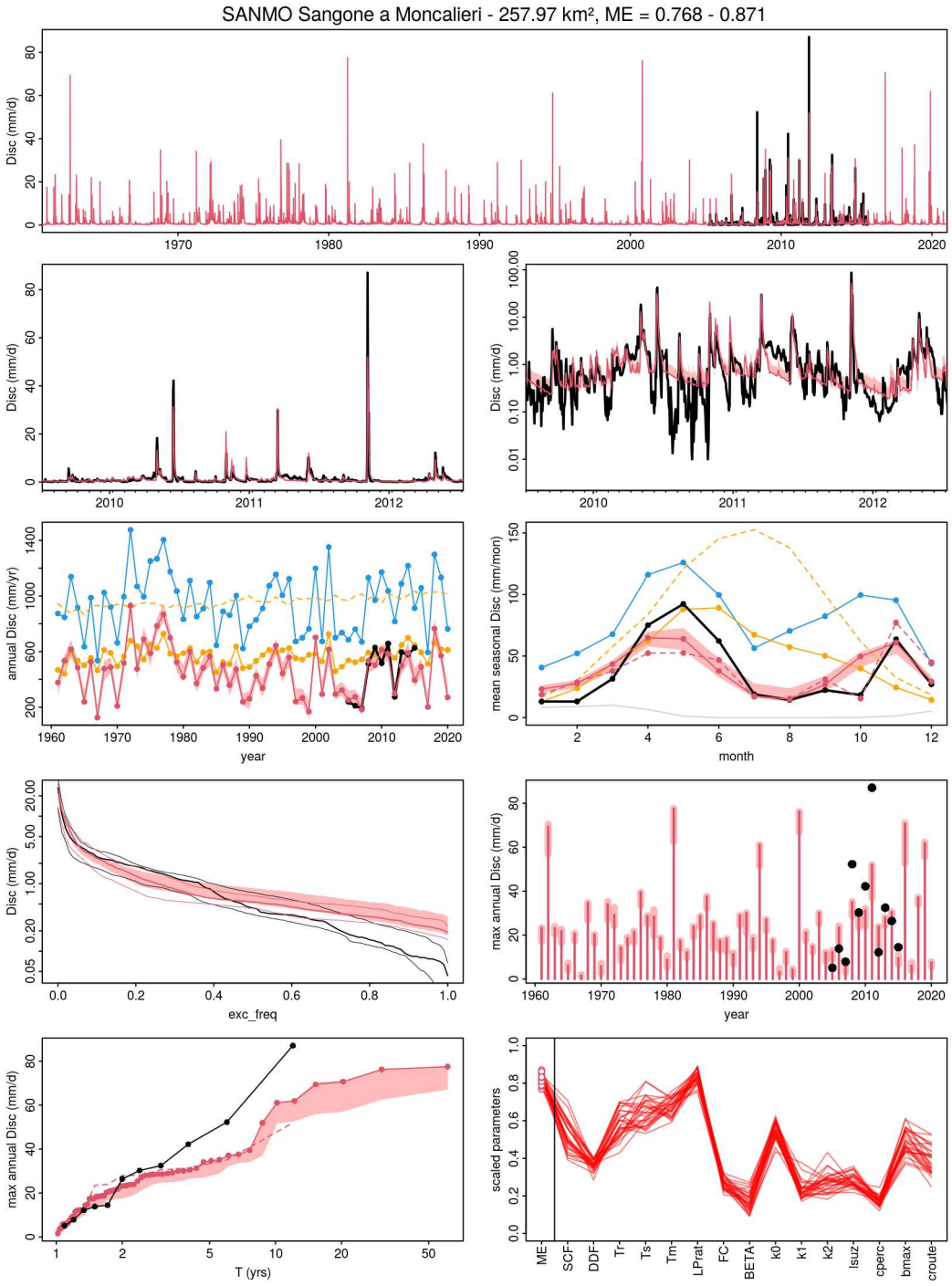


Fig. E.52 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 052.

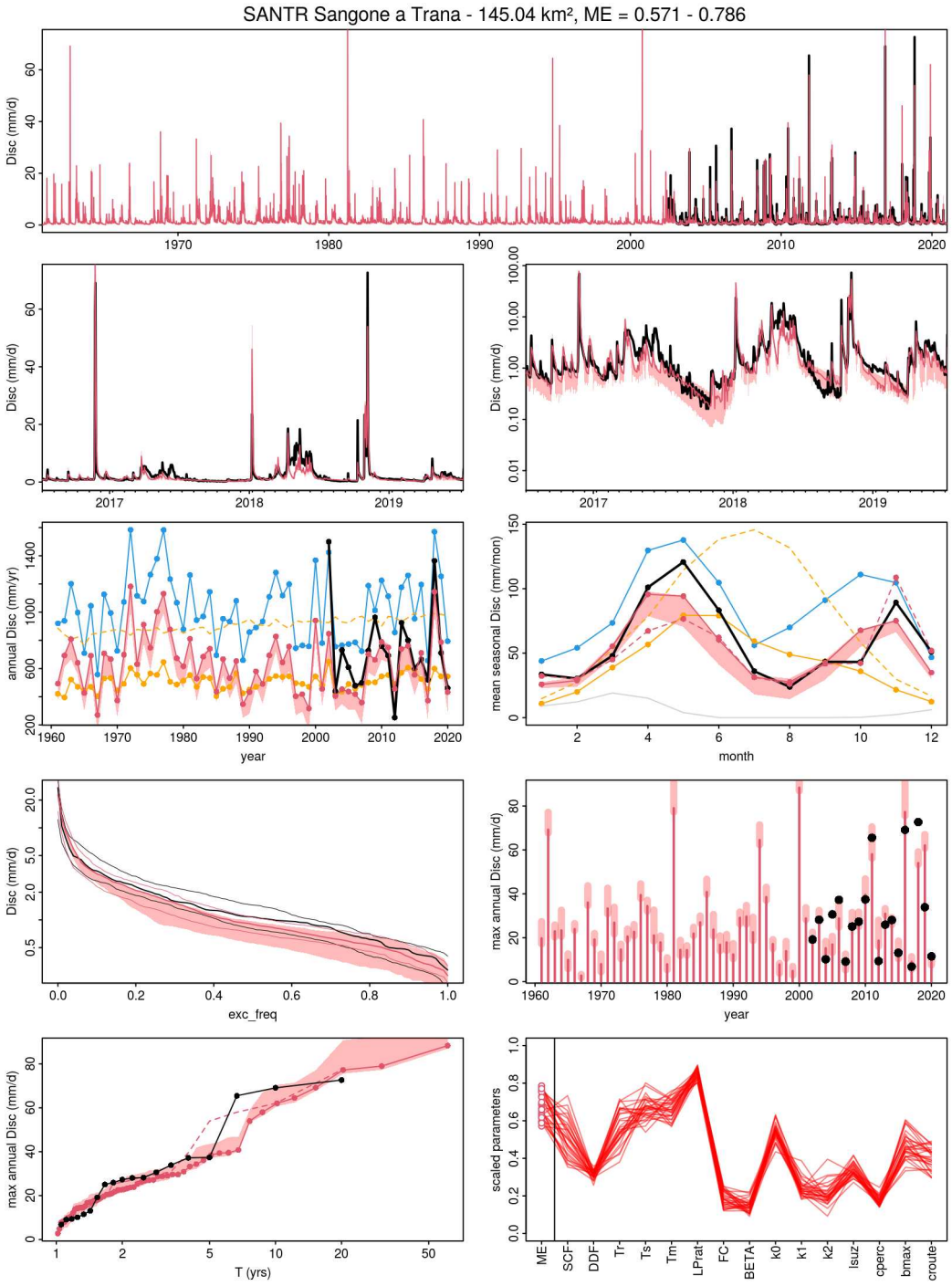


Fig. E.53 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 053.

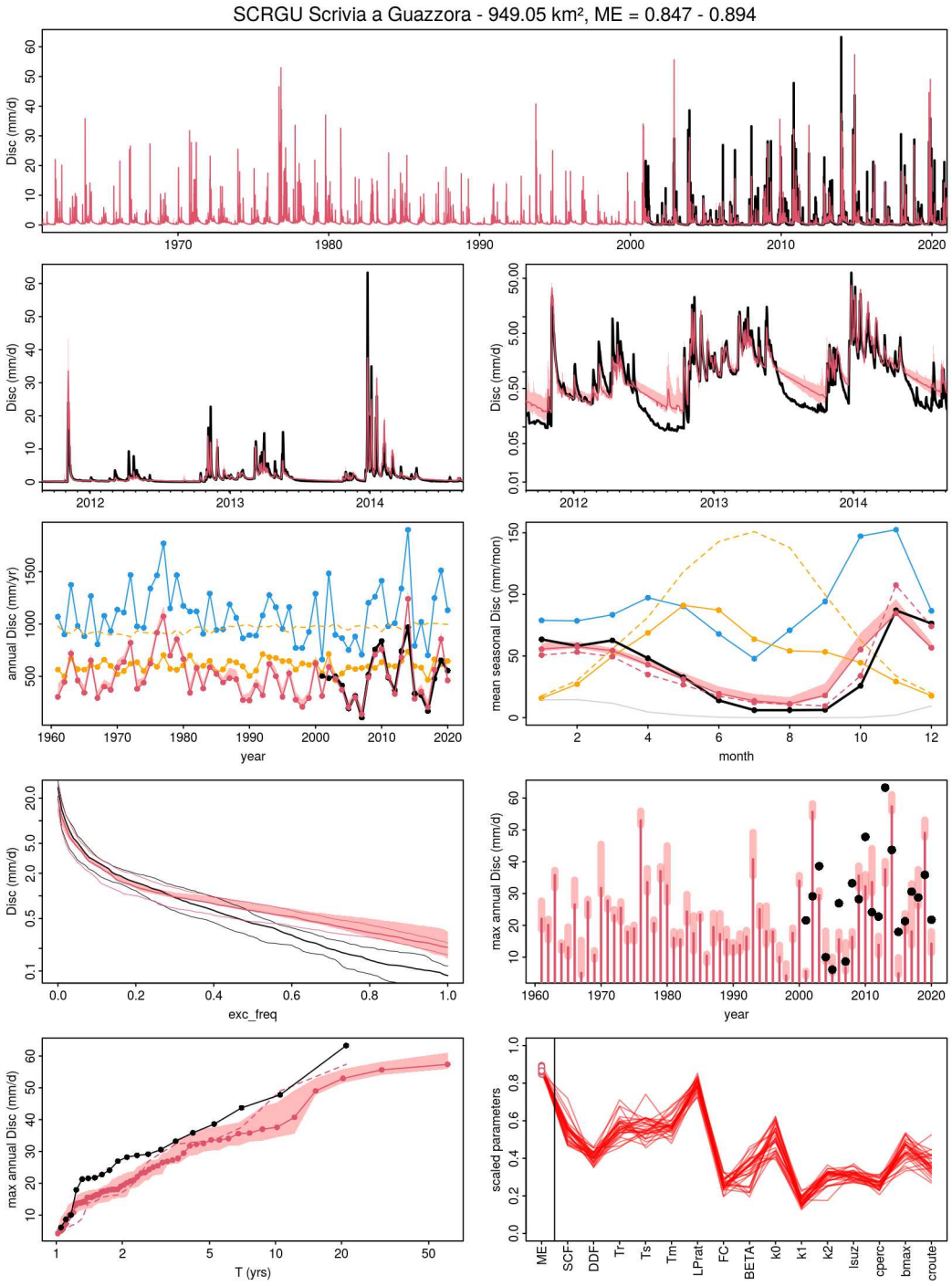


Fig. E.54 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 054.

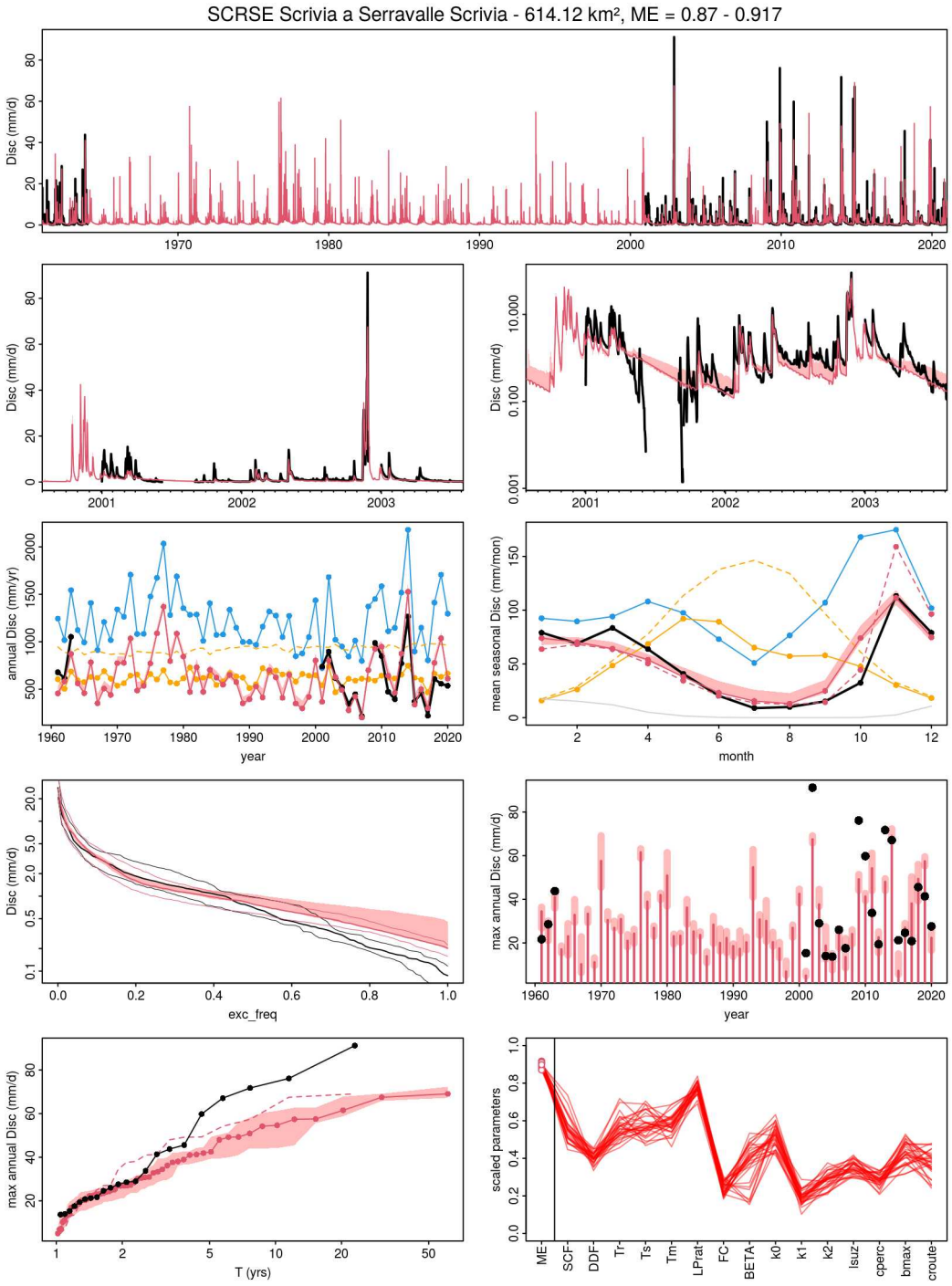


Fig. E.55 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 055.

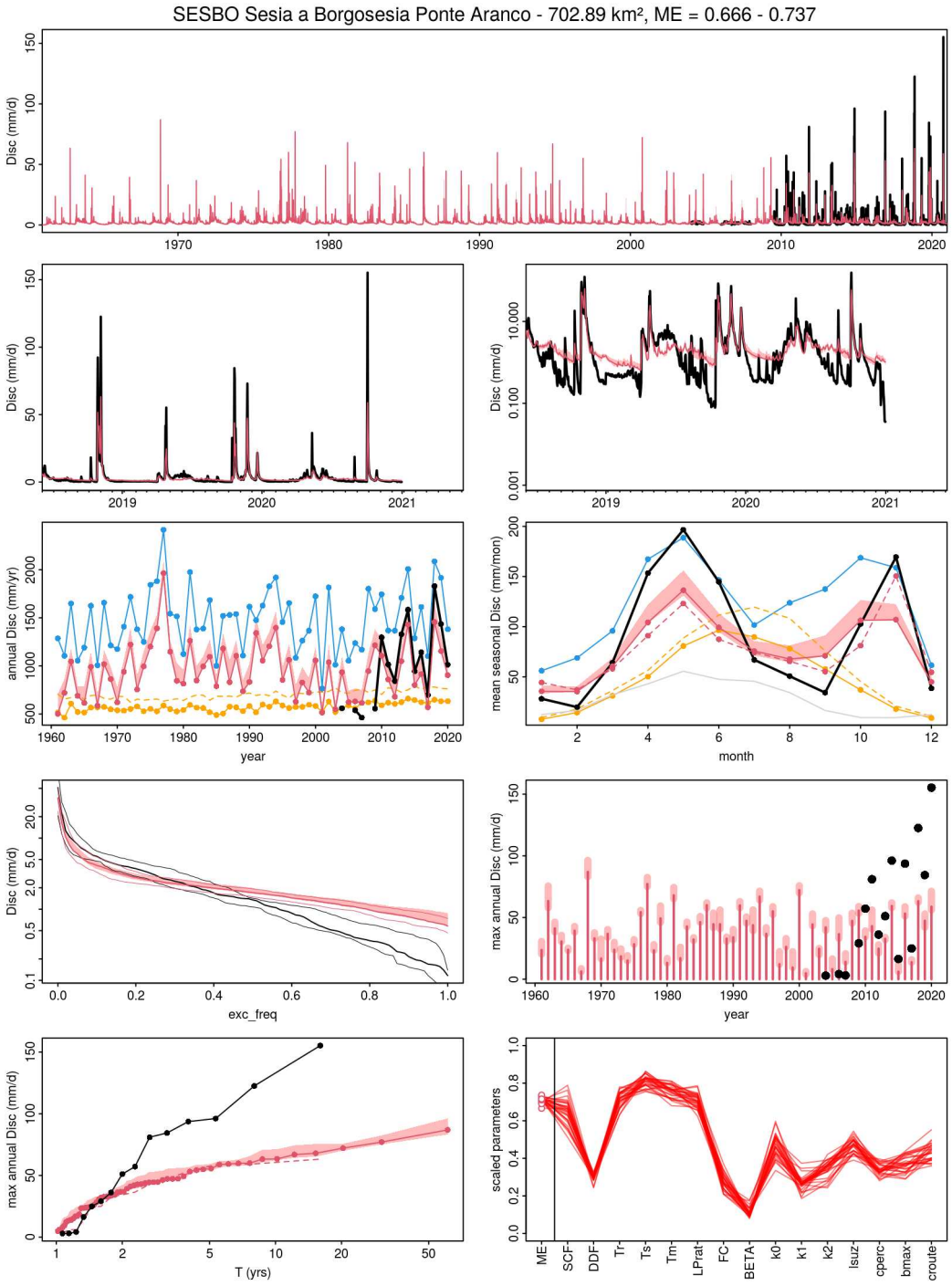


Fig. E.56 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 056.

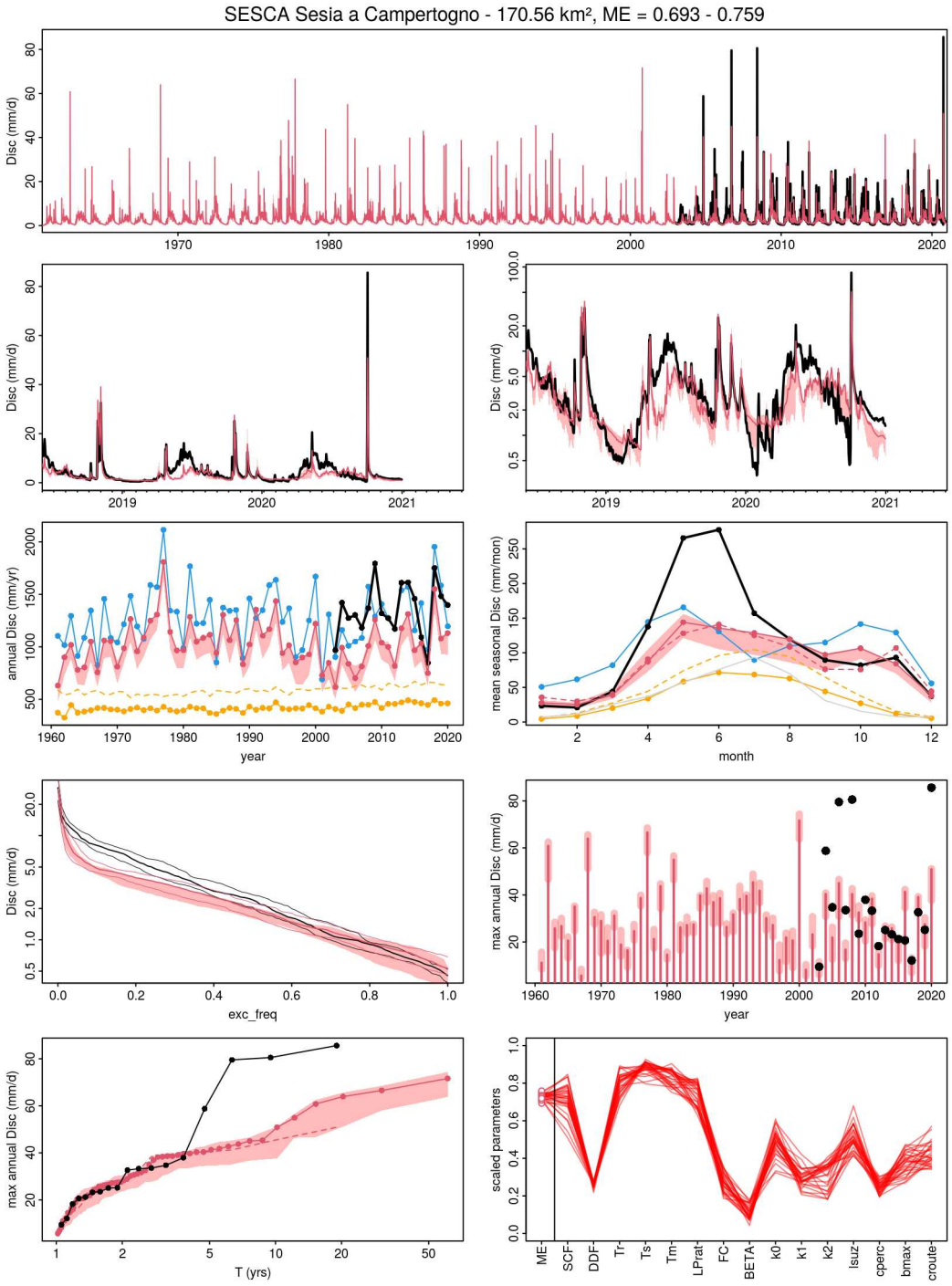


Fig. E.57 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 057.

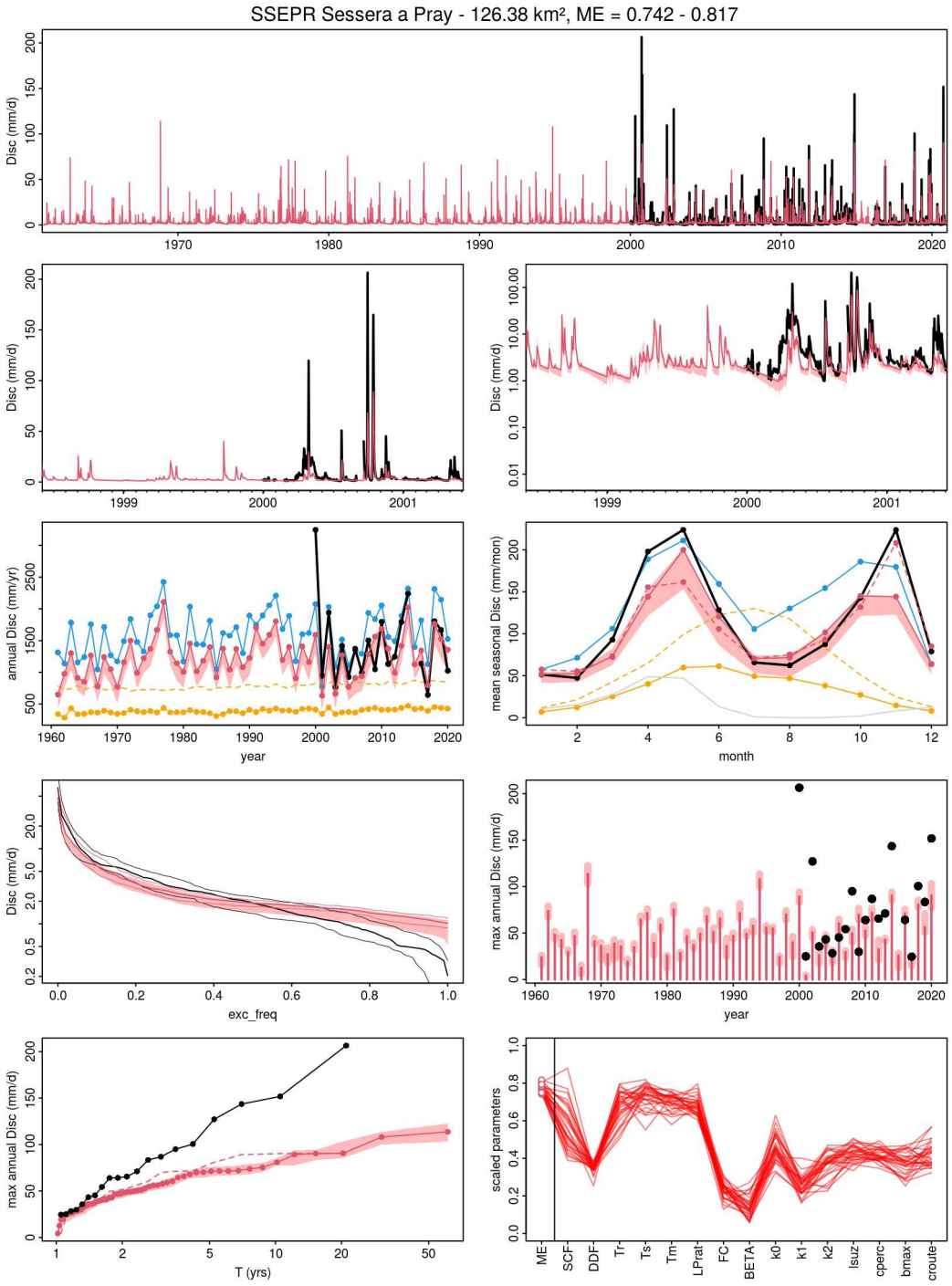


Fig. E.58 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 058.

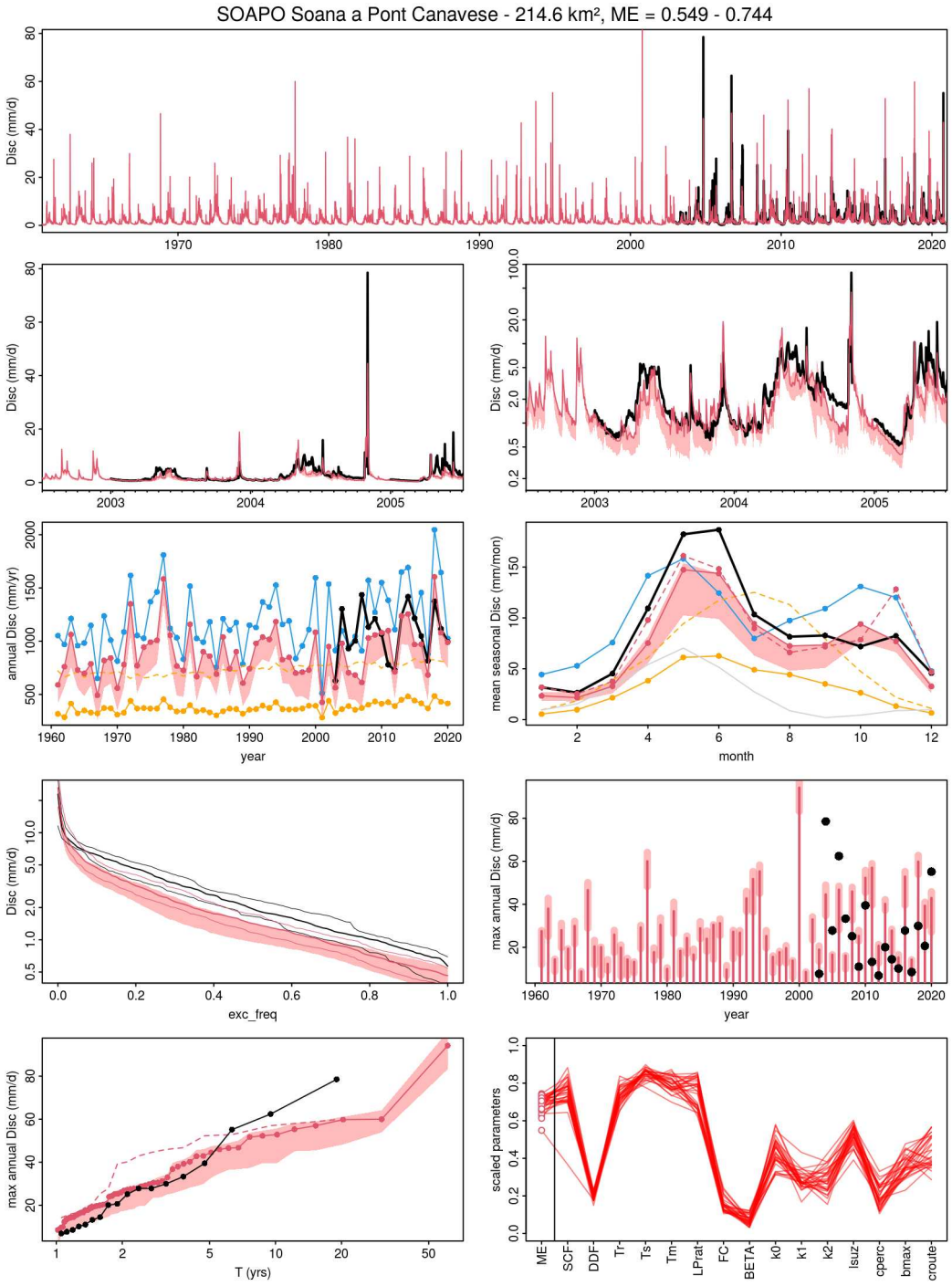


Fig. E.59 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 059.

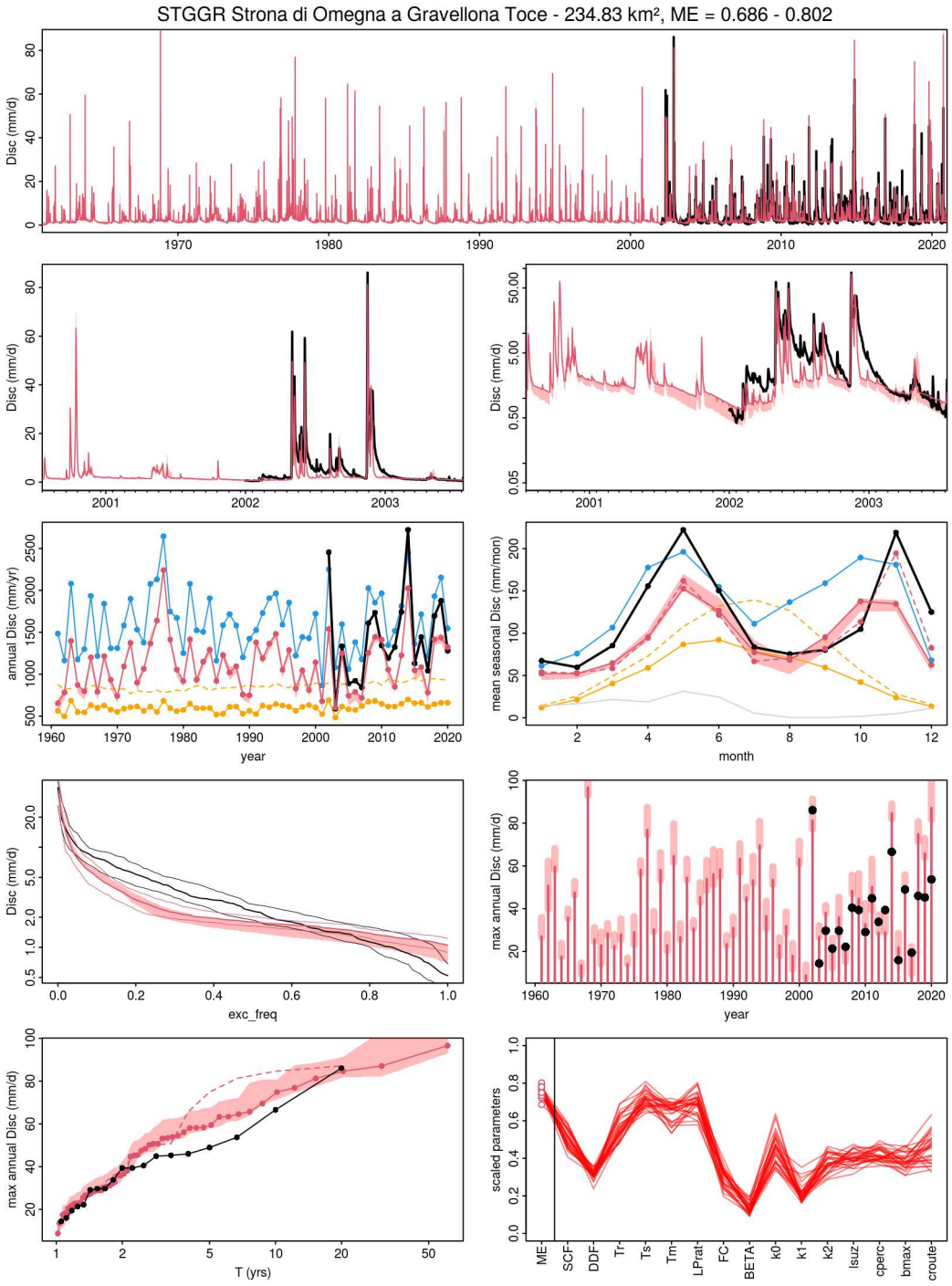


Fig. E.60 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 060.

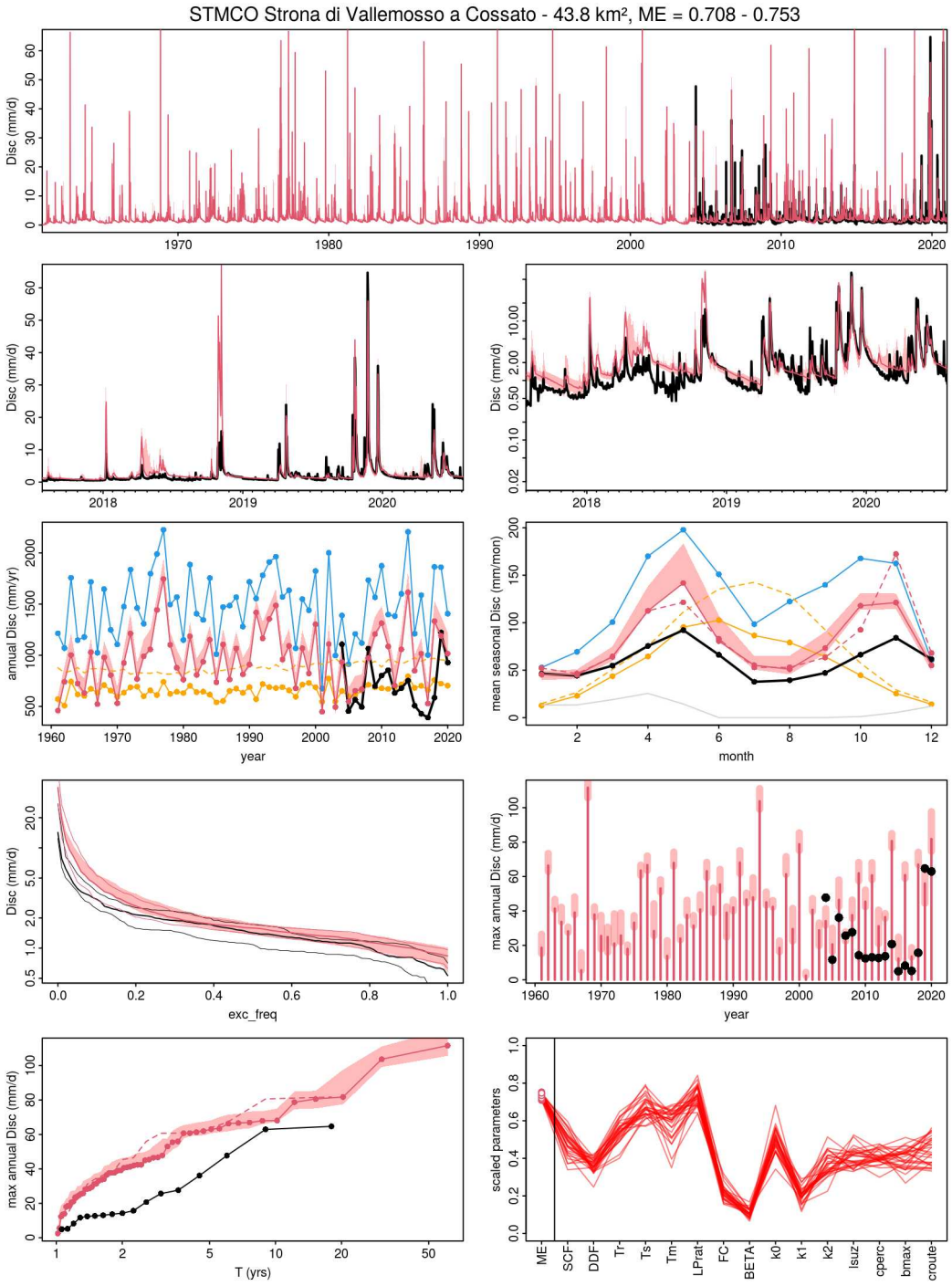


Fig. E.61 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 061.

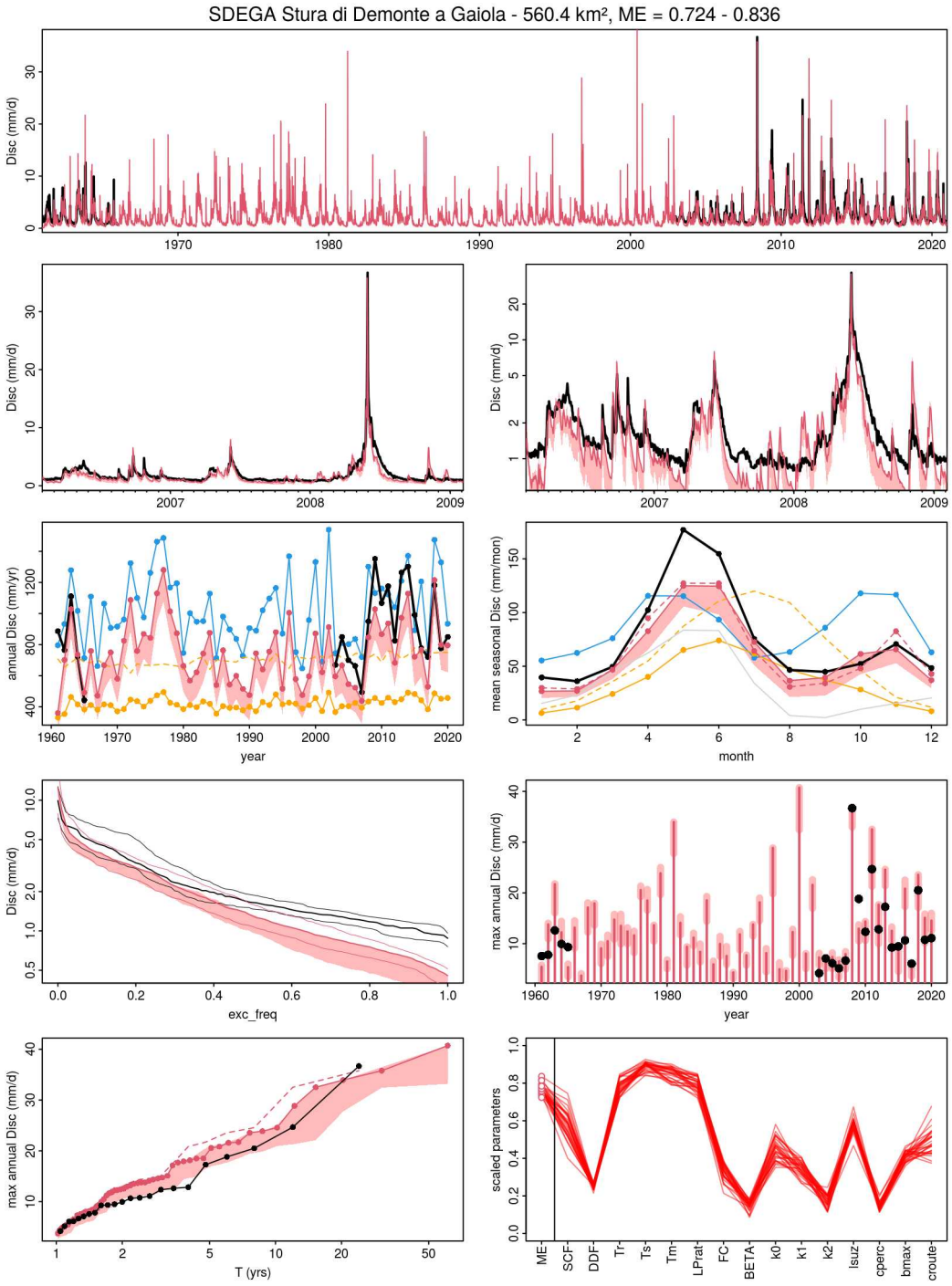


Fig. E.62 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 062.

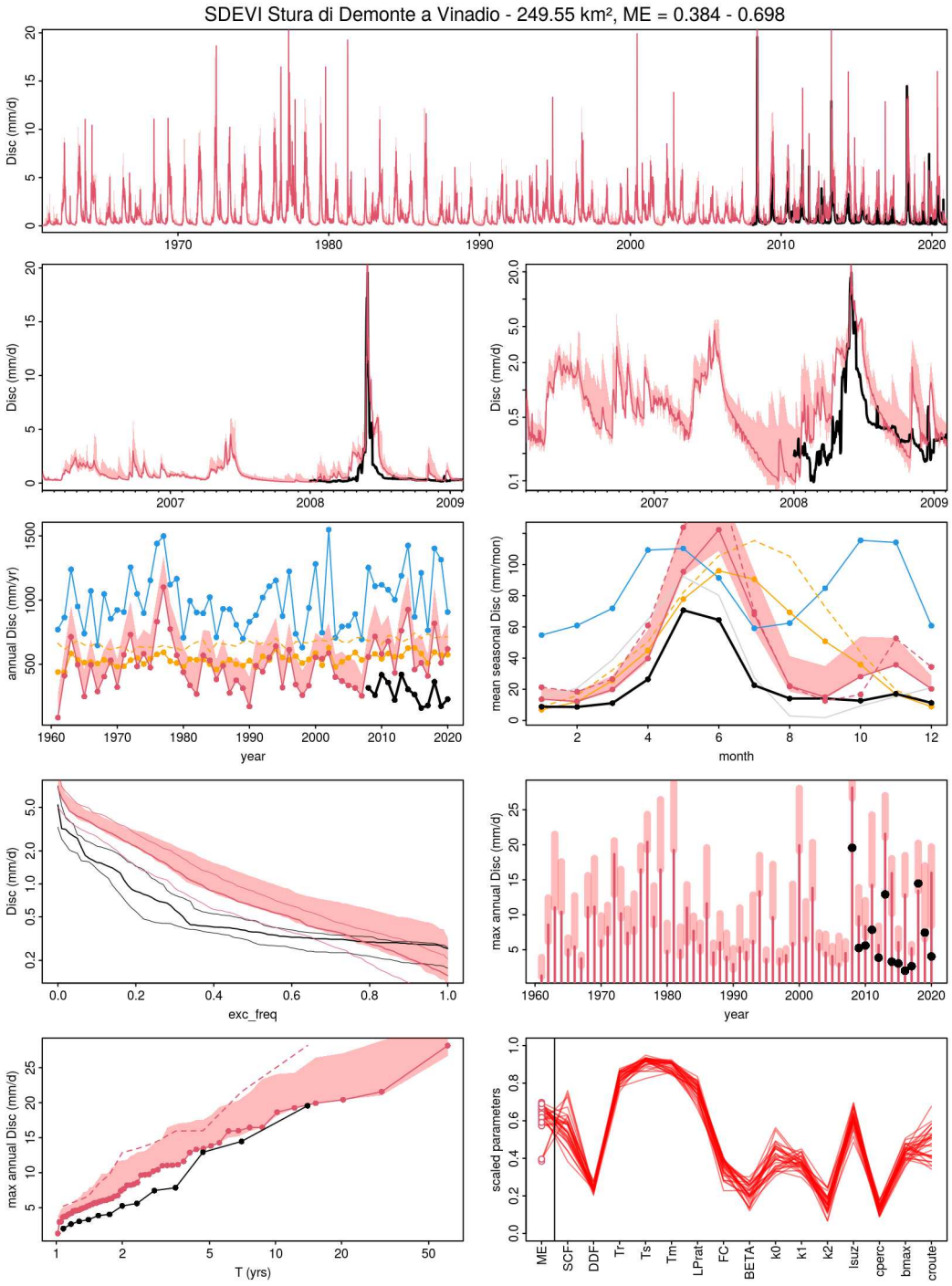


Fig. E.63 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 063.

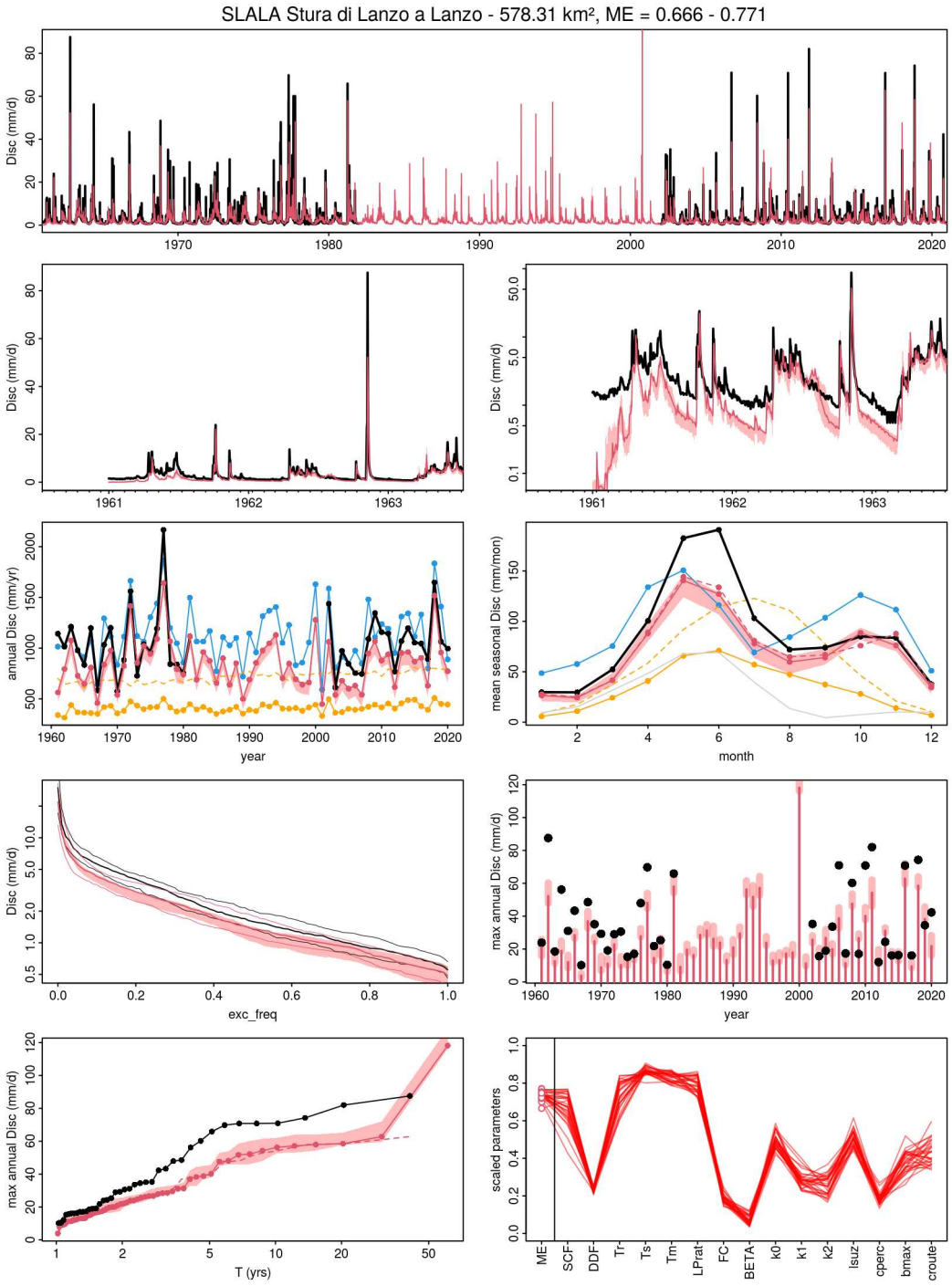


Fig. E.64 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 064.

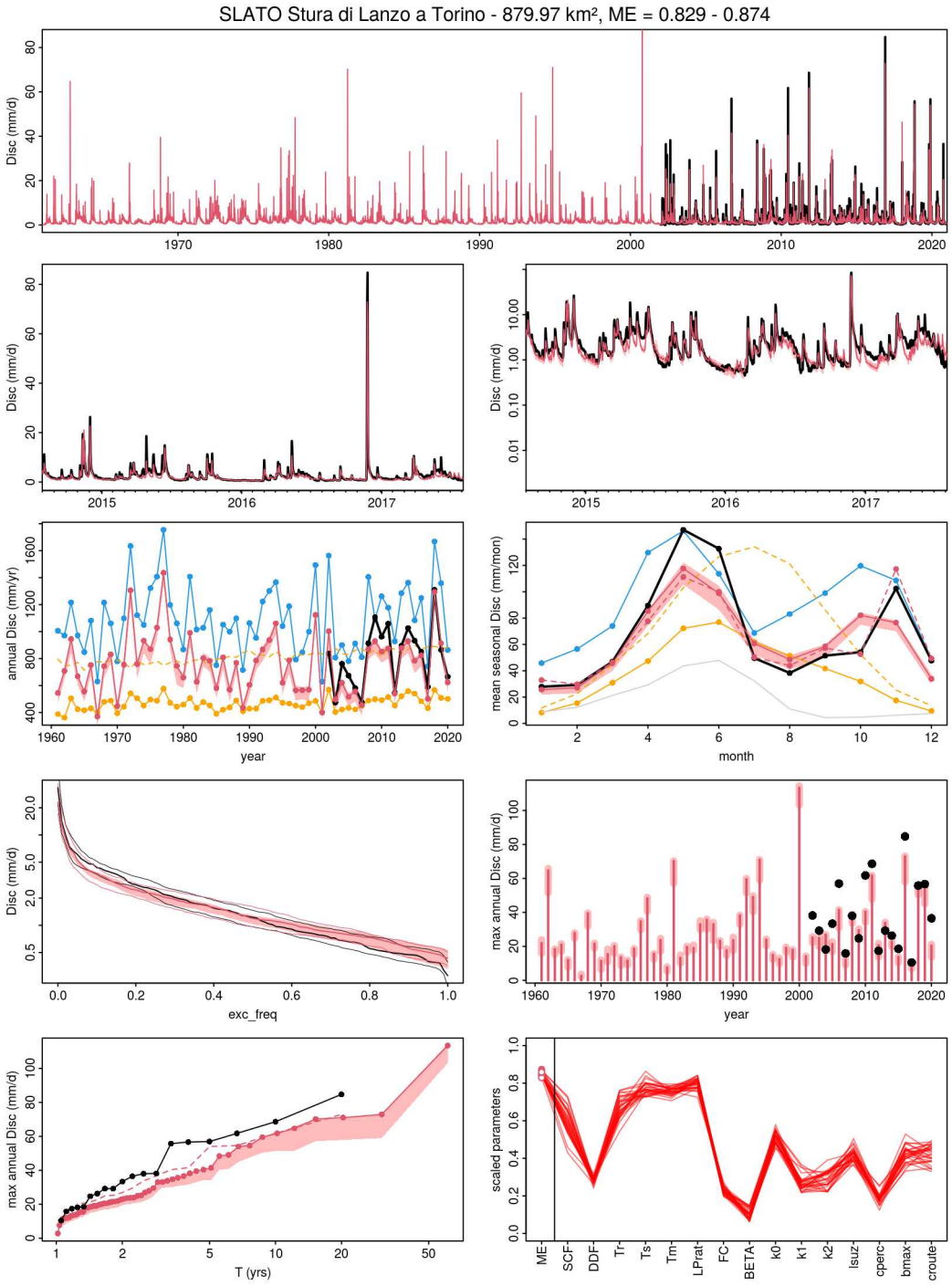


Fig. E.65 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 065.

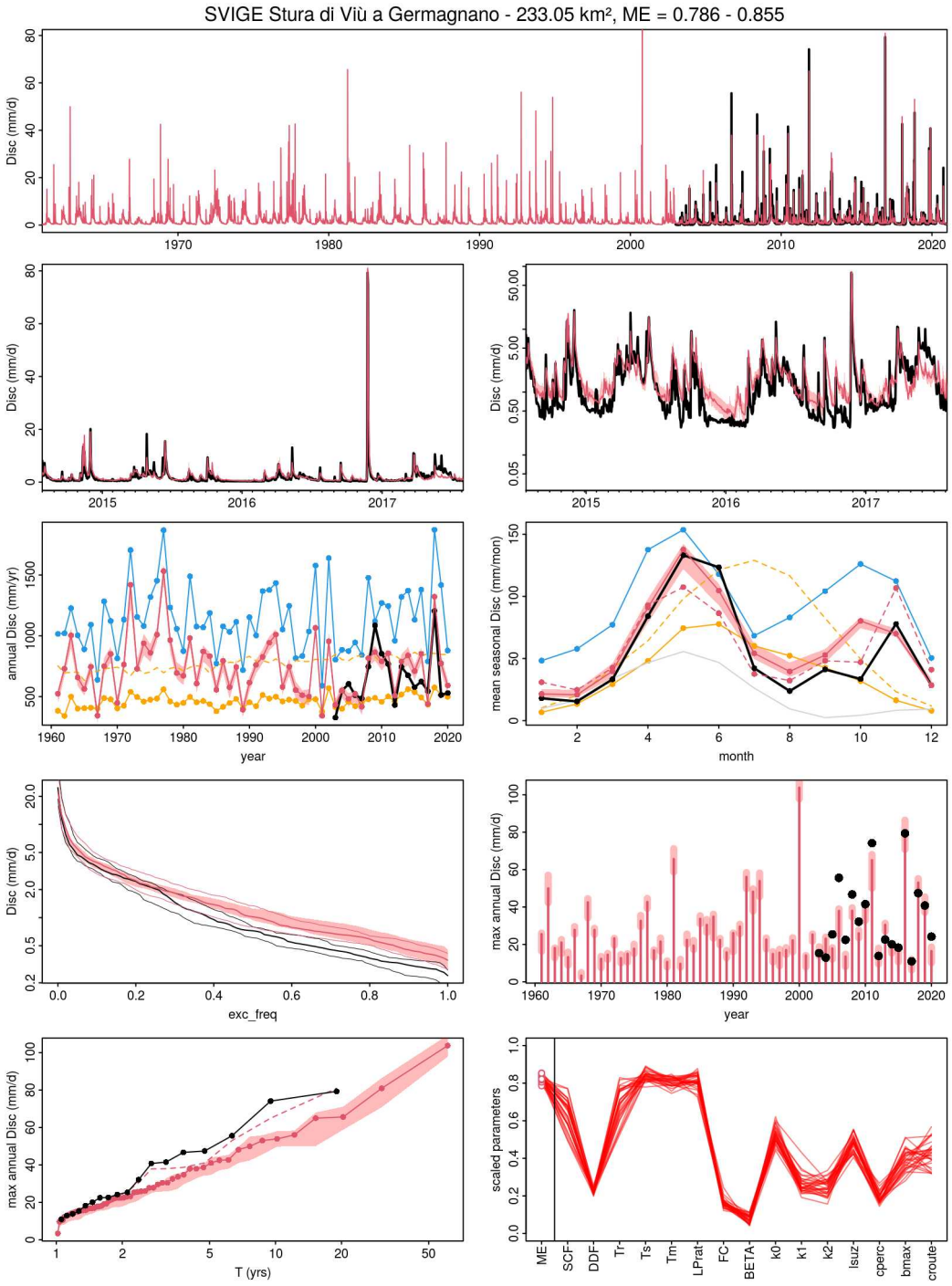


Fig. E.66 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 066.

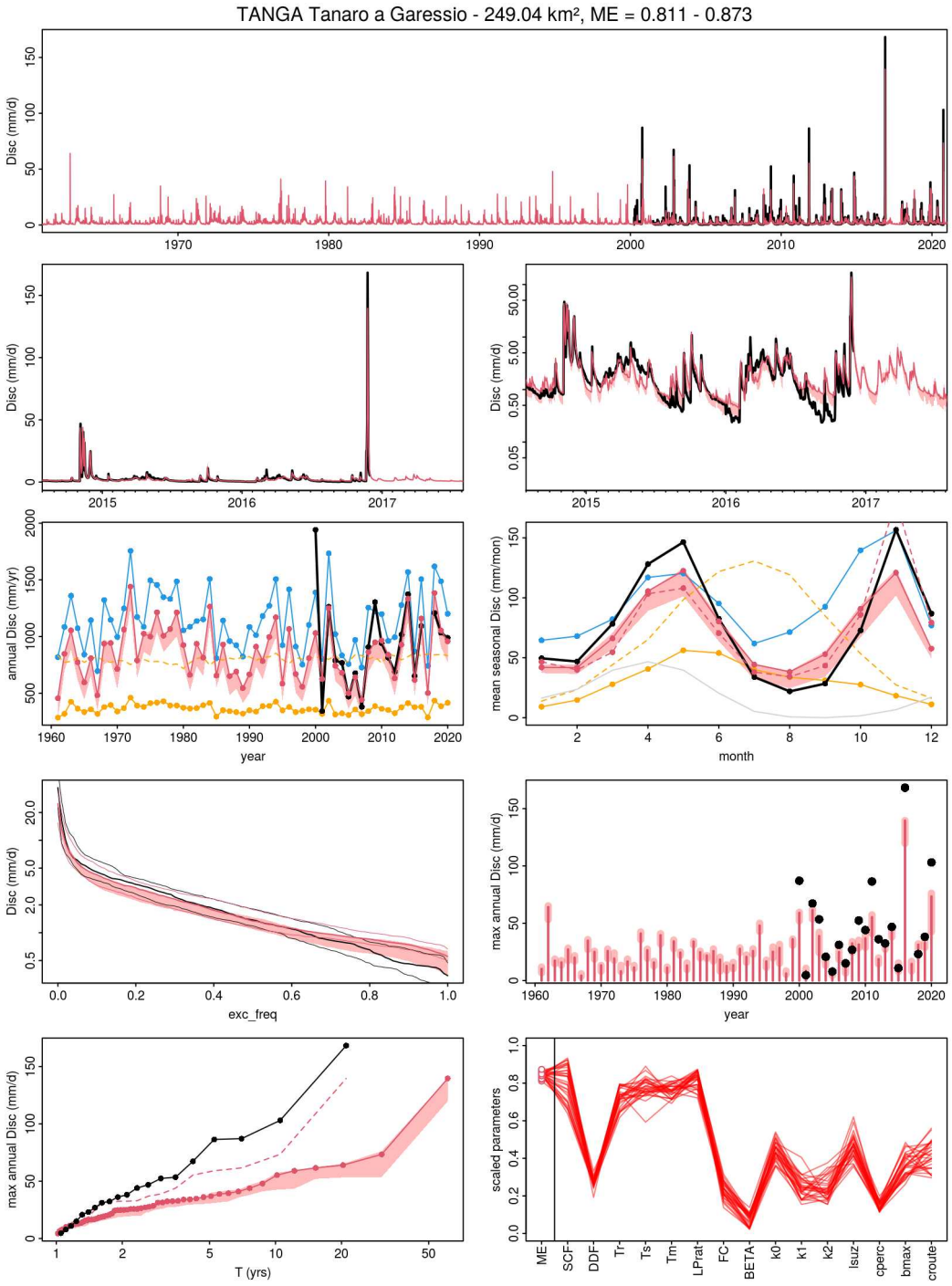


Fig. E.67 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 067.

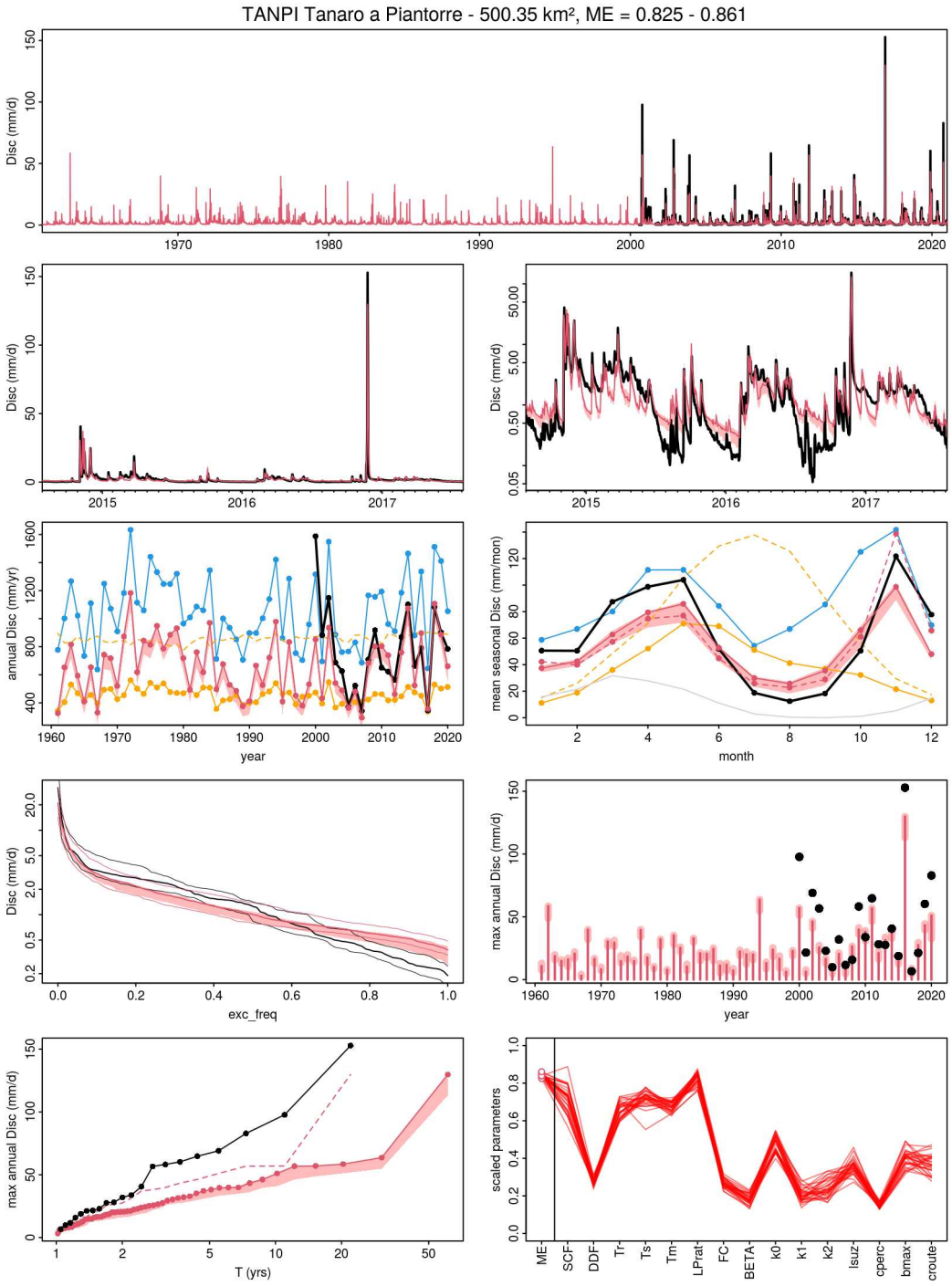


Fig. E.68 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 068.

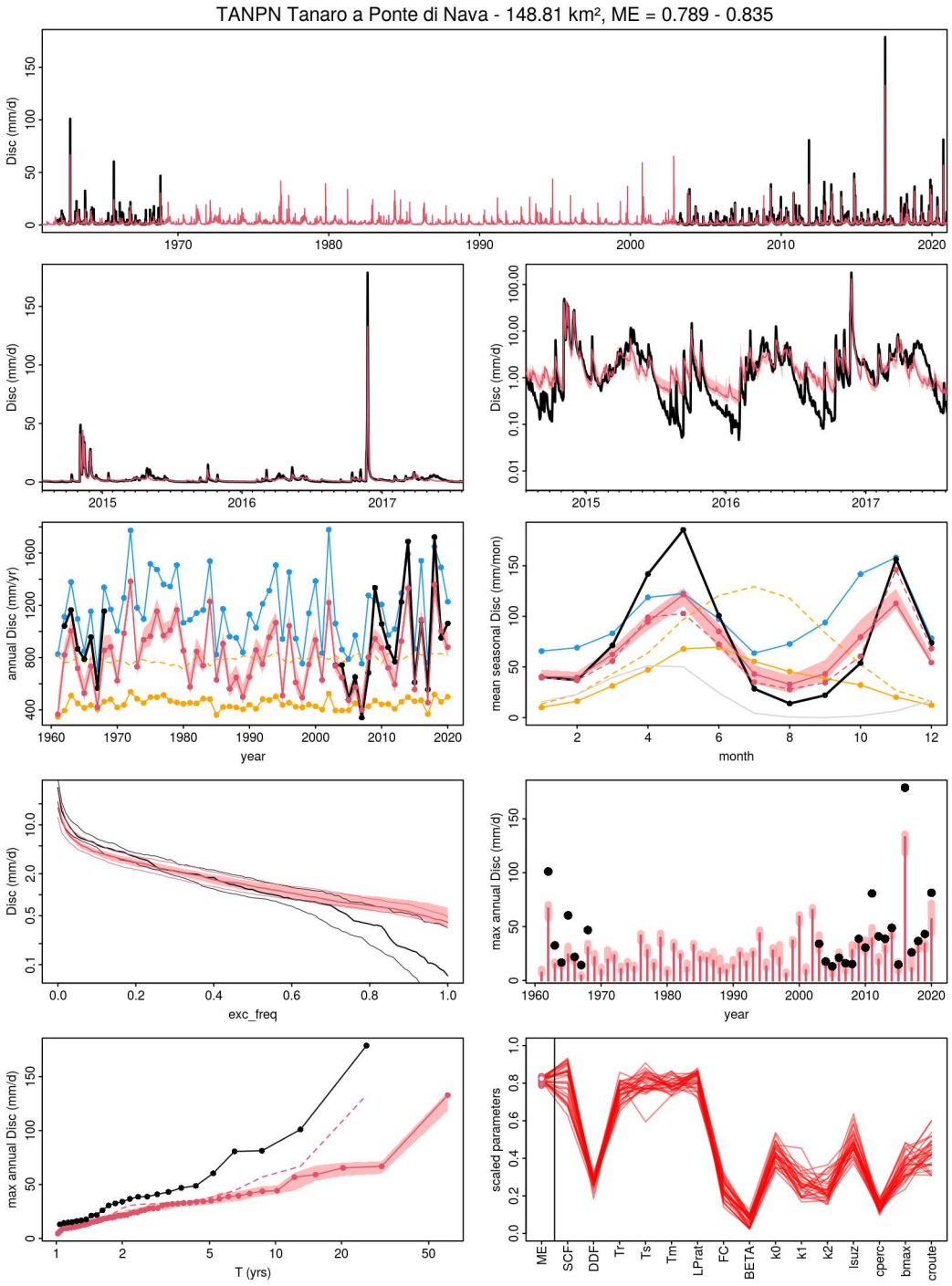


Fig. E.69 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 069.

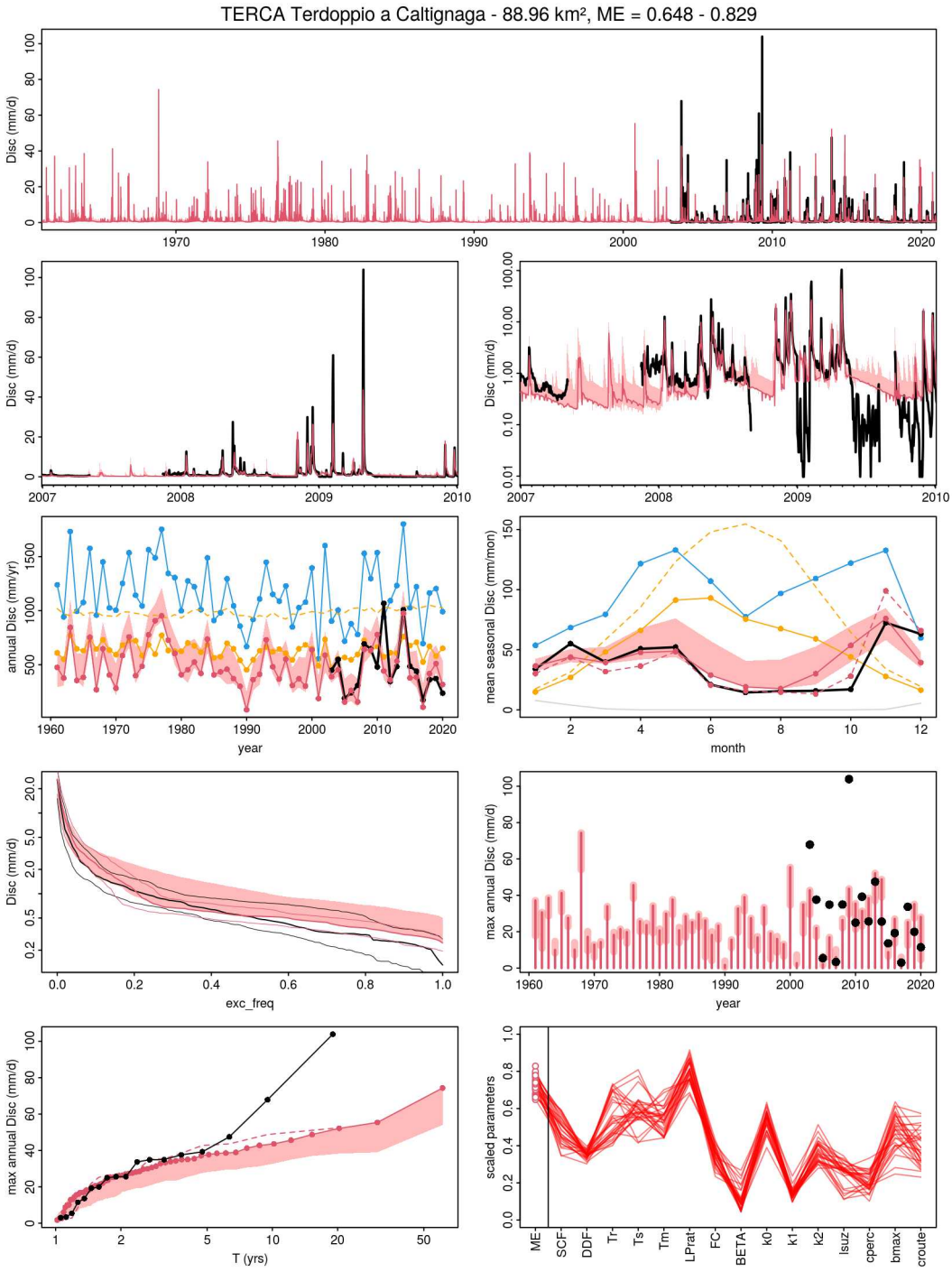


Fig. E.70 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 070.

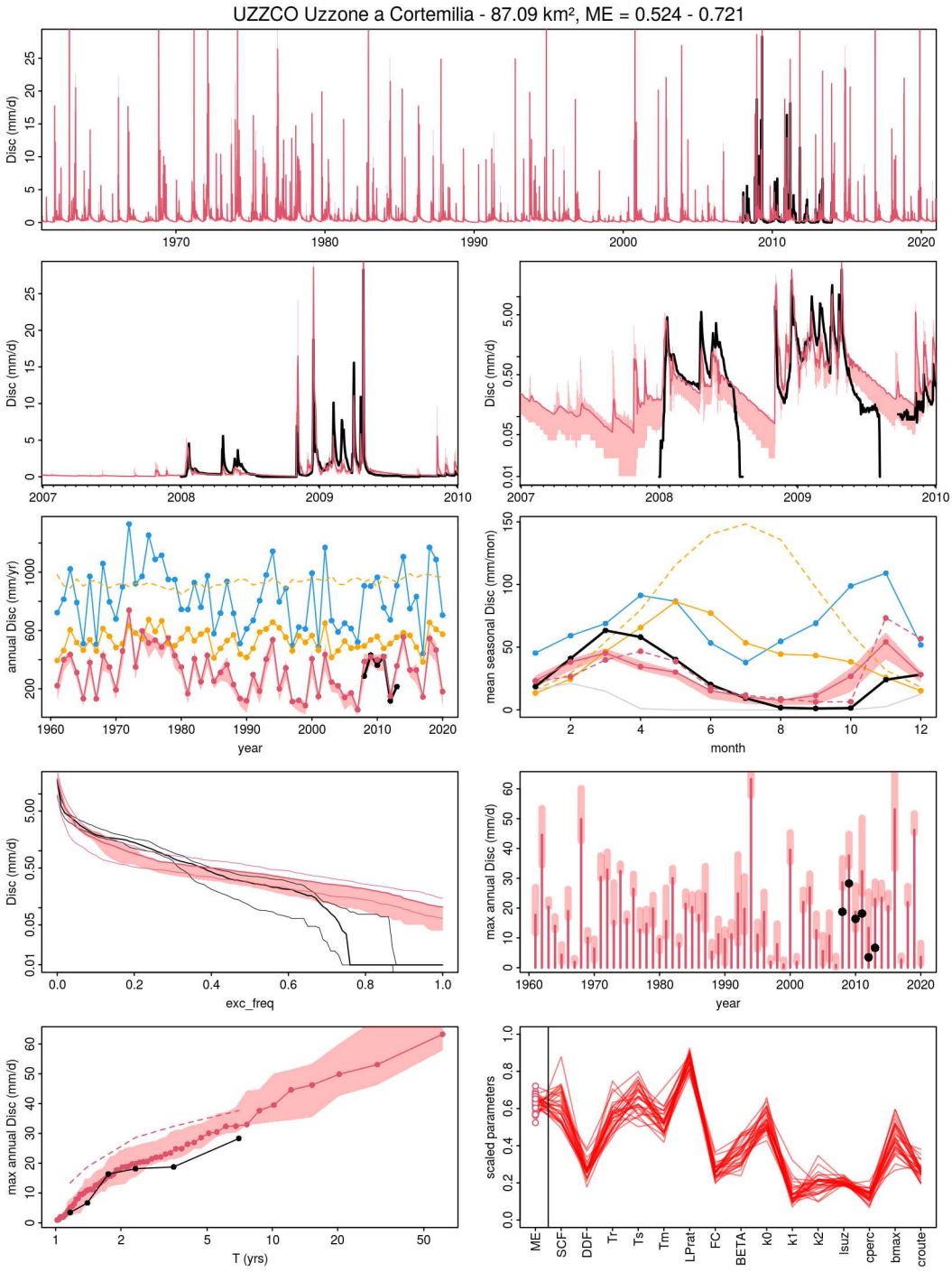


Fig. E.71 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 071.

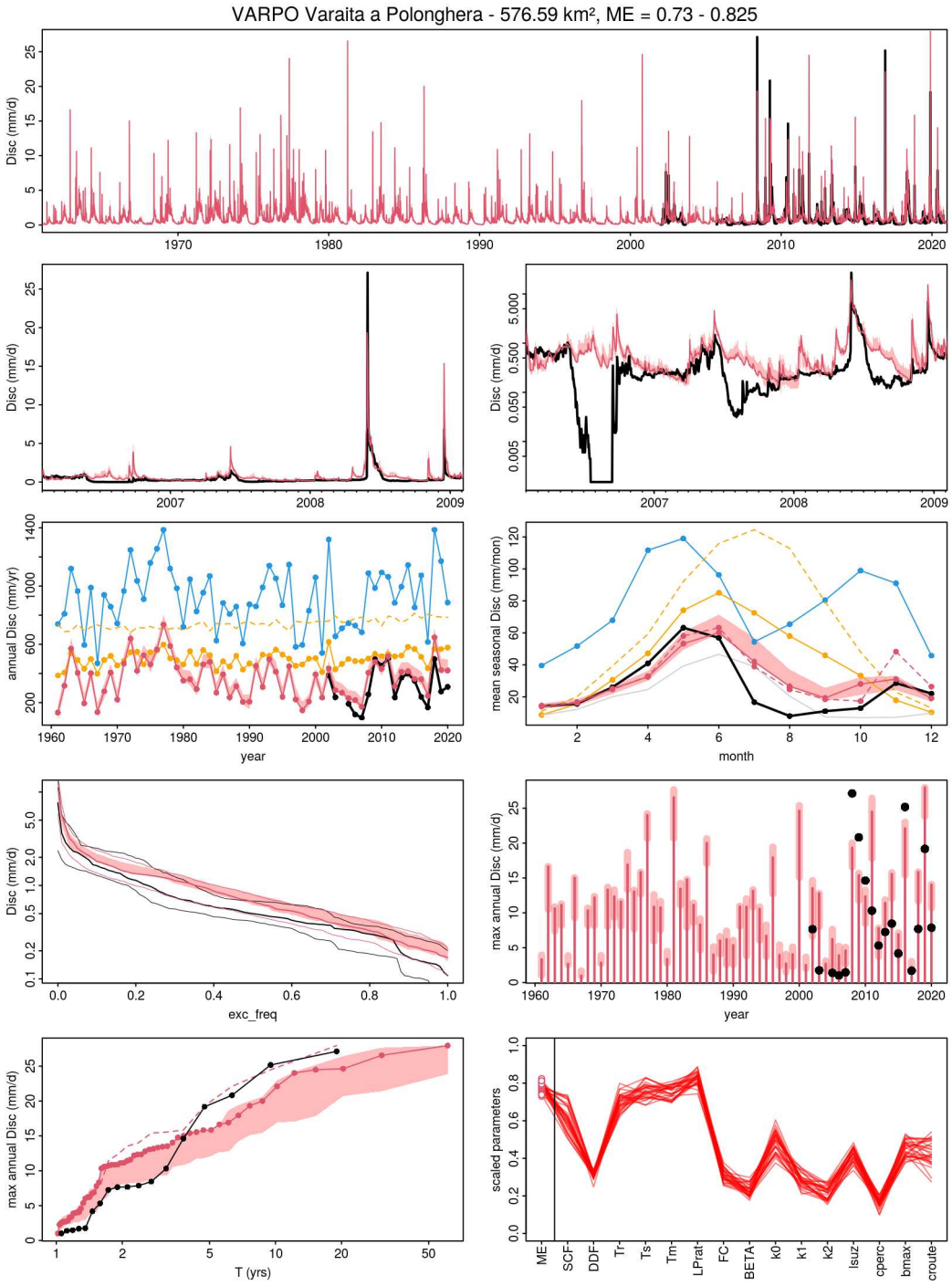


Fig. E.72 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 072.

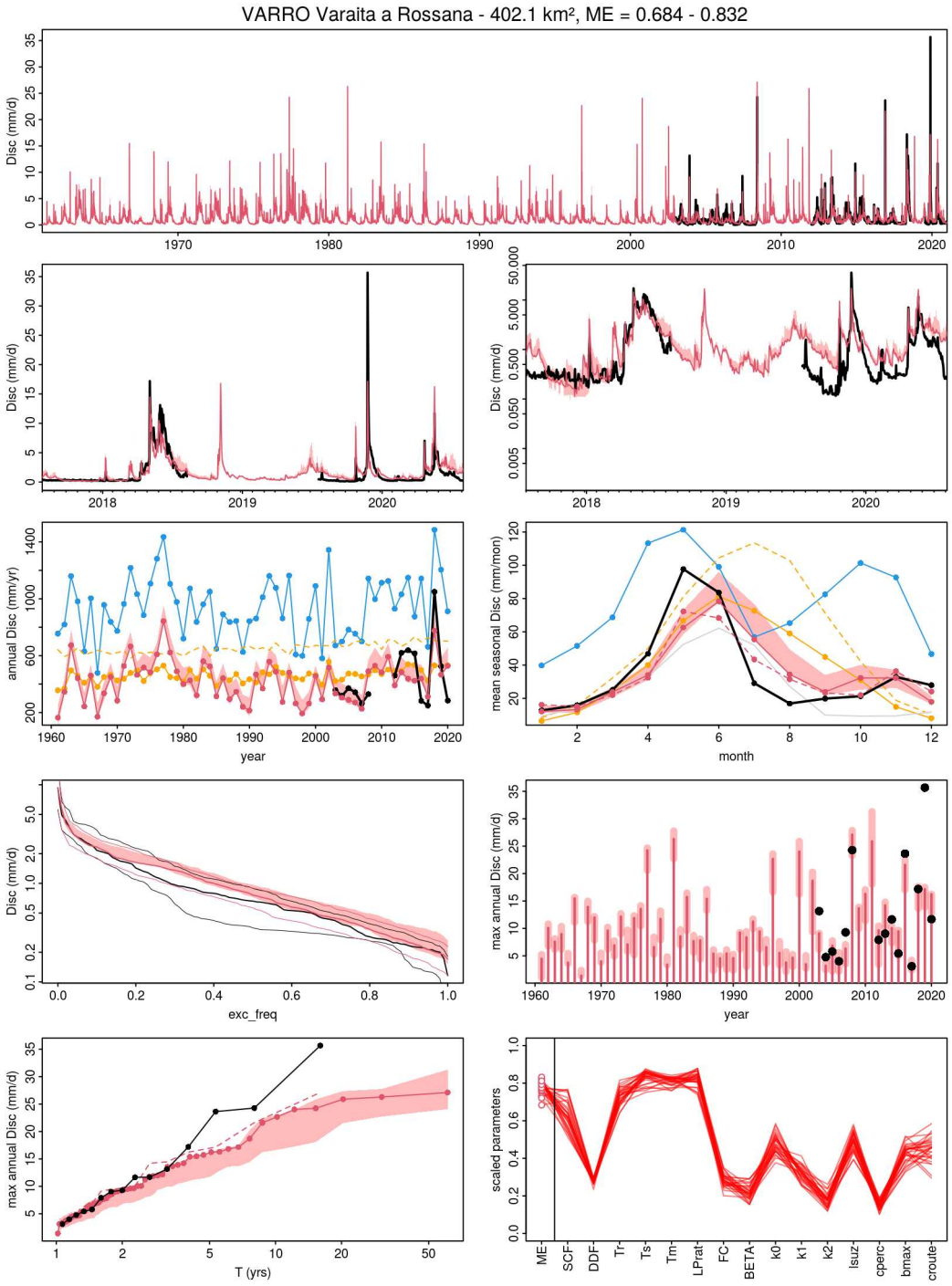


Fig. E.73 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 073.

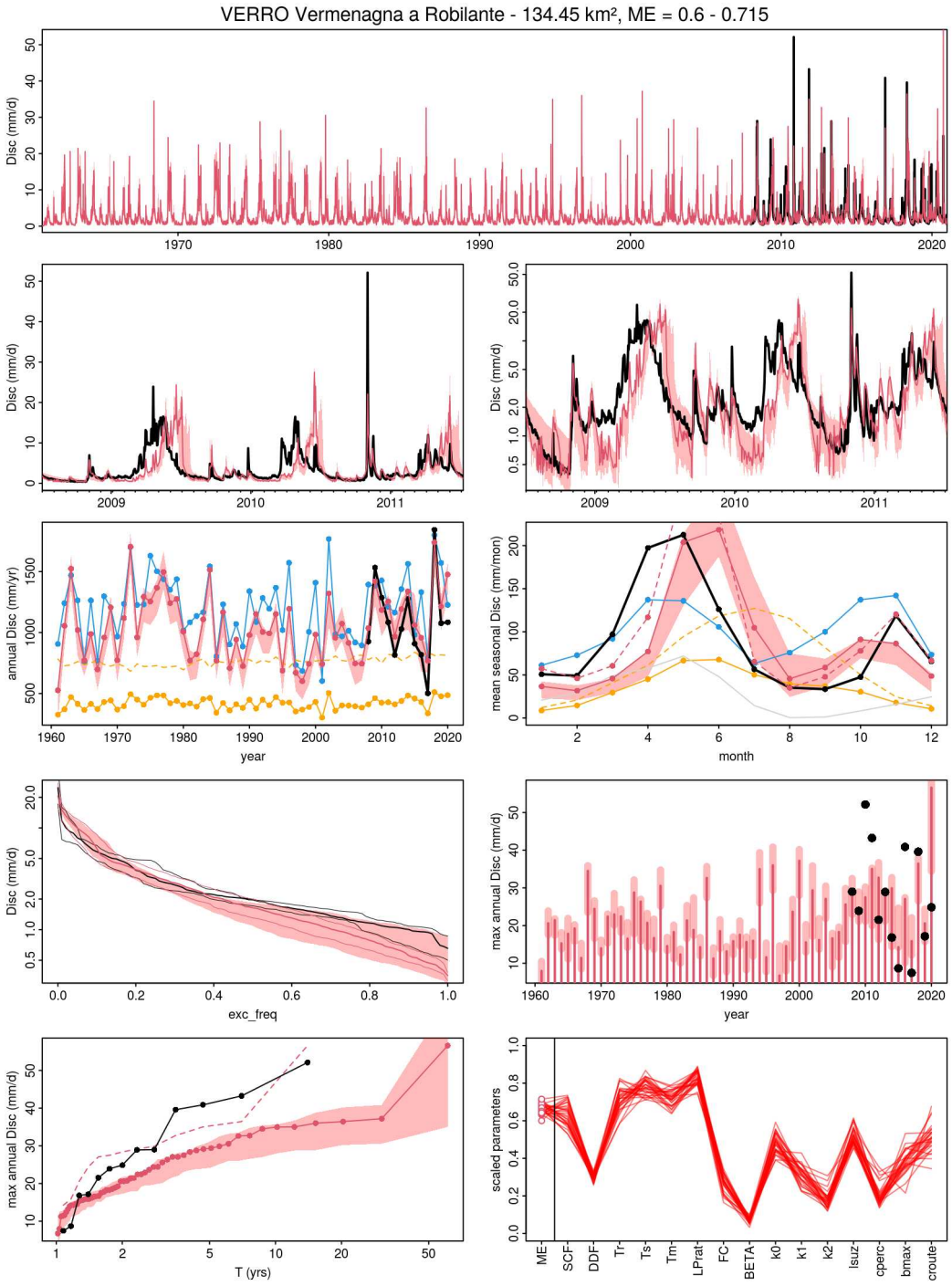


Fig. E.74 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 074.

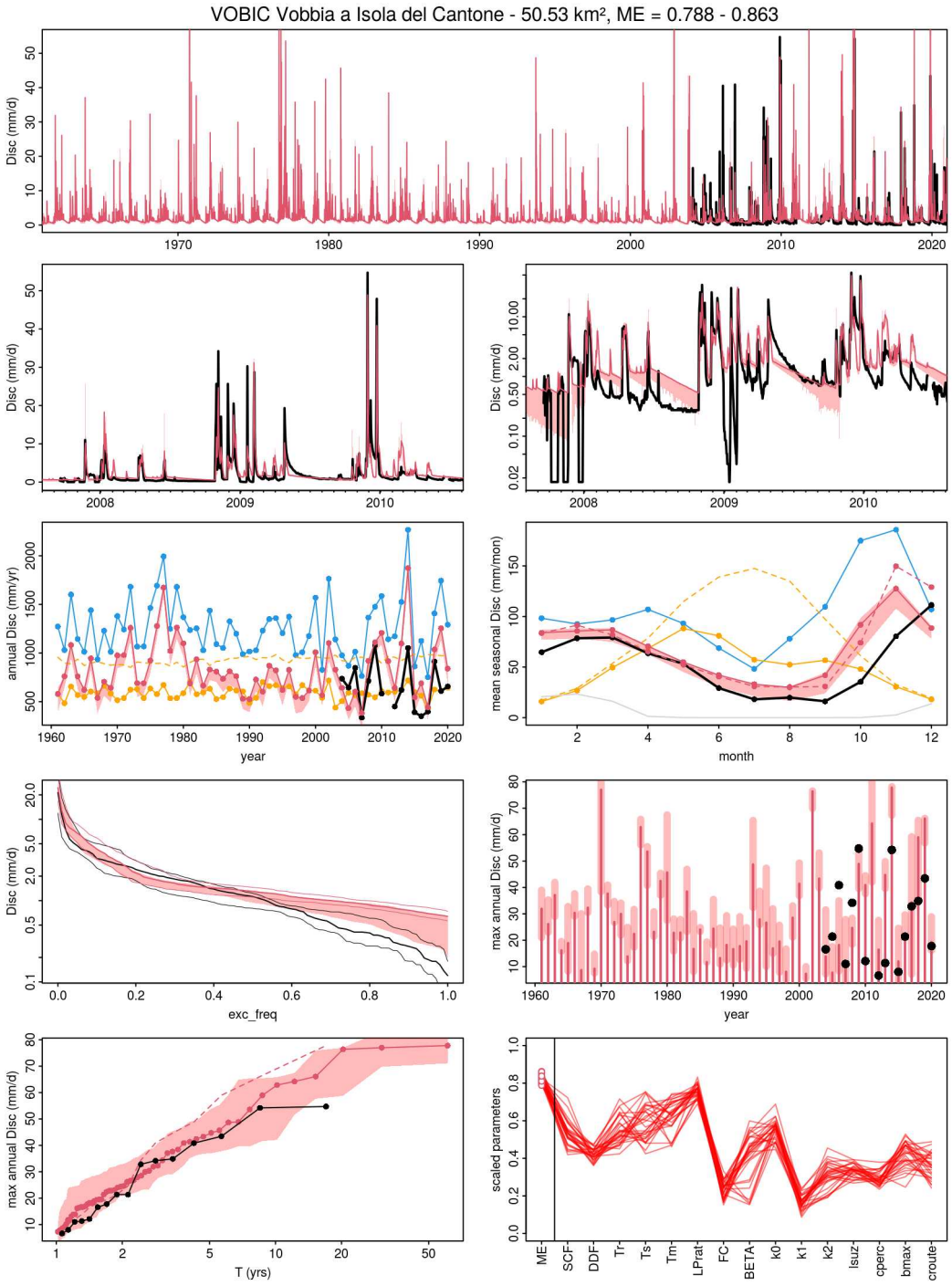


Fig. E.75 Simulated vs. observed discharges with regional PASS parameters obtained by calibration over the period 2000-2010 for TUWmodel, catchment 075.