Università degli Studi di Padova

Dipartimento di Ingegneria Civile,
Edile ed Ambientale

Master Thesis

WATER BALANCE MODELING IN THE EASTERN UNITED STATES

Supervisor: Prof. Gianluca Botter

Student: Andrea Betterle

Padova - 15 Ottobre 2014

Contents

Introduction **1**

1 Hydroclimatic data and study catchments **5**
1.1 Rainfall and discharges 6
1.2 Potential evapotranspiration 7
1.2.1 MODIS 8
1.2.2 CGIAR 8
1.3 Data management 9
1.4 Study Catchments 10

2 Methods **15**
2.1 The water balance 15
2.1.1 Rainfall partitioning 17
2.2 Water balance models: overview 18
2.3 WB1 20
2.4 WB2 22
2.5 WB3 26
2.6 WB4 32
2.7 WB5 39
2.8 Decomposition of annual water balance 41
2.9 Ranking of the models 43
2.10 Streaflow PDFs and flow regimes 45
2.10.1 Streamflow PDF estimation model 46

3 Results **55**
3.1 Annual Water Balance 56
3.2 Seasonal Water Balance 61

3.2.1 Seasonal decomposition 66
3.3 Ranking of the models 73
3.3.1 Streamflow probability density function 75

4 Conclusions **87**

Bibliography **91**

Introduction

One of the key problems in hydrology is to describe and characterize the spatial and temporal variability of water balance, i.e. the splitting of precipitation into evapotranspiration and runoff at different time scales (e.g. annual or seasonal). A proper understanding of the factors that control the components of the annual water balance provides a basic knowledge of the relations between climatic and hydrologic variables. In addition, as the use of general circulation models to estimate the effects of changing climate on precipitation is becoming increasingly widespread, a proper identification of the factors that affect the annual water balance becomes more urgent for the prediction of the related impacts on human well-being and ecosystems. A good understanding of the variables and processed to which the water cyrcle is most sensitive is the starting point to cope with water management issues such as hydropower production and irrigation, especially in developing countries, where problems concerning water supply might have impacts beyond economy.

Of fresh water available on the Globe, just 1 % is readily accessible to humans (less that the 0,0025% of the total water) being stored in lakes, ponds or rivers. Therefore, quantifying its availability and variability in time and space is a cumbersome and important issue with many ecological and managerial consequences. Streamflows are particularly important in this context because river flows can be measured with a certain accuracy compared to other components of the water balance such as evapotranspiration. Moreover, streamflows are a critical component of the water balance because of the implied value for human uses and aquatic ecosystems.

There have been numerous studies that have examined the factors driving the temporal variability of runoff (e.g. *Wigmosta & Burges, 1997; Xia et. al, 1997*) but fewer studies have examined the factors affecting the underlying spatial variability. *Thorntwaite* (1948) and *Budyko* (1955) used mean annual precipitation and mean annual potential evapotranspiration to identify moisture regimes at global scales.

During the same period *Langbein* (1949) described the physical factors that control the spatial distribution of annual runoff in the conterminous United States and found that climate was the dominant control of the spatial variability of annual runoff. *Budyko* (1974) showed how the competition between available water and available energy is a first order driver of long term annual water balance. More recently, *Milly* (1994) indicated that precipitation characteristics such as storm intensity and frequency, soil texture, vegetation type and density and geomorphology are important determinants of the spatial variability of annual runoff. The aforementioned benchmark studies have identified fundamental concepts of the water balance, stressing the importance of climatic supply (precipitation) and demand (potential evapotranspiration), seasonality in supply and demand and soil-moisture storage for hydrologic studies.

A detailed knowledge of the drivers of the water balance makes more reliable the prediction in areas potentially impacted by climate change. Changes in the overall rainfall depth, intensity and temporal distribution, jointly with alterations in evapotranspiration, might lead to variation in runoff inversely proportional to the buffer capacity of the drainage area, causing shifts in the hydrological regimes, with direct impacts on anthropogenic activities and the ecosphere (*Botter, 2013*). Floods protection measures, reservoirs, diversions and other engineering facilities dimensioned according to specific reference discharges would not be able to face more severe conditions induced by changing climate. Riparian vegetation and riverine ecosystem in general, are strongly influenced by the amount and temporal variability of the discharge (*Doulatyari et al., 2014; Ridolfi et al., 2006; Poff et al., 2007*). Moreover flow variance is strictly related to flow predictability, on which the mobility and colonizing ability of living species depend. In particular, frequency and duration of low discharges determine the amount of carnivores, the amount of physiologically tolerant species and the variation of mortality through disease.

At global scales, climate change has strong impact on the long term variation of water balance and river flow regimes. Among the others, *Arnell et al.* (2010) , evaluating the potential effects of climate change on a series of indicators of hydrological regimes across the global domain, concluded that substantial proportions of the land surface are likely to experience significant changes in hydrological regimes by 2050. In particular, they estimated an increase of annual runoff across 47% of the land surface and a decrease across 36%; an increase of the flood peaks across more than 50% of the land surface and a decrease of drought runoff across 44%;

significant changes in the amplitude of river flows trough the years across 80% of land surface and the shift of the month of maximum runoff earlier, across much of eastern Europe and north America.

The ability to evaluate the magnitude of the shifts of flow regimes not only in response to climate changes, but also as an effect of regulation, is a fundamental task to properly manage water resources. The knowledge of the modifications that dam, diversions, or anthropic infrastructures may produce on downstream flow regimes, proves fundamental to evaluate of the impact on downstream biomes and ecosystem services provided by surface fresh water (*Poff et al., 2007*).

The aim of this thesis is to apply and test the performances of a set of well known water balance models to a selection of catchments belonging to the United States east of the Rocky Mountain, at seasonal and annual time scales. The 39 study catchments used are spreaded throughout the study area, so they involve different climate zones and span a range of sizes from about 40 to 2000 Km^2. Two different datasets of potential evapotranspiration (PET) were applied and an overall ranking of the models was obtained by taking into account their performances and the number of parameters involved (*Akaike, 1974*).

The availability of reliable models at global/continental scales makes feasible a robust estimate of the runoff coefficient, potentially for any ungauged catchment within a given study area, thereby allowing the prediction of the mean discharge based on precipitation and climate data. The same procedure can virtually be applied to any part of the Globe where rainfall and PET data are available, upon calibration of a minimum number of parameters.

Moreover, wherever water balance models are coupled to the stochastic approach for soil moisture and river flow dynamics developed by *Botter et al.*(2007a, 2009) (that links flow-producing rainfall pulses to discharge depending on climate features and hydrological processes), a reliable characterization of flow regimes may be achieved. In particular, the knowledge of the runoff coefficient at seasonal timescale, coupled with the estimate of other parameters describing rainfall properties and recession dynamics during single events, allows for an analytical expression of the probability density function of streamflows (which is equivalent to the flow duration curve). The knowledge of the probability density function (PDF) of the streamflows along a river network is extremely valuable in decision making processes related to water management, flood mitigation and riverine ecosystems. Streamflow distributions provide information about mean daily discharge and its variability, thereby

defining water availability, discharge fluctuations and frequency of high and low flows. As such, flow regimes impact riverine ecosystems and anthropic activities like hydropower production, fishing, navigation, recreation, water supply for civil and industrial uses and land irrigation. Riverine species and riparian vegetation are also sensitive to variability in flow regimes. Impacts on the natural variability of discharges caused by anthropogenic interventions (such as reservoirs and diversions) can be properly assessed only based on the knowledge of pristine flow regimes. Downstream a dam or a weir, in absence of discharge measurements prior to river regulation the goal can be achieved only by using hydrological models.

All these issues suggest the importance of modeling tools to provide a reliable description of the hydrological cycle at catchment scales across global domains, using a narrow set of inputs such as rainfall records, potential evapotranspiration maps and digital terrain models, which are widely available in many regions on the Globe.

Chapter 1

Hydroclimatic data and study catchments

Research conducted in the last century (*Budyko,1974; L'vovich, 1979; Milly, 1994; Voepel et al., 2011*) suggested that the main variables involved in the water balance can be grouped into two categories: climate and landscape variables.

- **Climate variables** include variables such as precipitation, potential evapotranspiration, incoming solar radiation and seasonality, and their time variability.
- **Landscape variables** include latitude, longitude, catchment area, mean elevation, mean slope, soil porosity, soil field capacity, wilting point, saturated hydraulic conductivity, soil water available to plants, rooting depth and soil composition.

It is worth to mention that this classification is quite general provided that some variables of one category might depend on others belonging to a different category. For example soil water available to plants is a function of rooting depth, porosity, field capacity and wilting point and potential evapotranspiration strongly depends on the incoming solar radiation (*Hargreaves, 1985*). Indeed, climate is one of the major long term drivers of landscape and vegetation, and the latter actively interacts with the landscape, at multiple spatial and temporal scales.

All these variables affect the water balance in a very heterogeneous way. The well known semi empirical relation of Budyko identifies in the ratio between potential evapotranspiration (PET) and precipitation a first order driver of the long-term

annual water balance. Afterwards, (*Eagleson, 1978 and Milly, 1993*) showed that both soil water storage capacity and soil water are pivotal in the partition of precipitation into evapotranspiration, deep percolation and surface runoff, thereby potentially explaining the observed departures from the Budyko curve. In these pioneering studies it was also emphasized how seasonality in water supply (precipitation) and demand (evapotranspiration) and their synchronicity, may provide significant deviations from the behaviour predicted by Budyko's approach: large shifts occur when peaks in supply and demand are out of phase (*Budyko and Zubenok, 1961*).

Roughly speaking, the recent literature has summarized the role of climate in the water balance issue through precipitation and evapotranspiration while, among all the landscape variables, the soil water storage capacity proved to be the most influential. Since these variables are considered first order drivers of water balance, they are input variables in most water balance models considered in this study.

1.1 Rainfall and discharges

The data processed in this work include daily rainfall records provided by the 'American National Oceanic And Atmospheric Administration' (NOAA), and daily discharge records provided by the 'United States Geological Survey' (USGS). The time series of these variable typically span several decades. Even though longer time series may be sometimes available, they were neglected due to concerns about the stationarity of the processes involved. Time series shorter than 20 years were avoided in order to increase the robustness and the representativeness of the results.

For this study a brand new database (*db@resQ*), implemented on behalf of Prof. Botter by i4 Consulting S.r.l., an engineering company set in Padova specialized in hydraulics, hydrology, environment and information technology, has been used. The aim of the database is to gather all possible hydrological information coming from external sources (such as the USGS and NOAA ones and virtually many others) and organize them in a consistent manner, allowing the user to export data in an easy and friendly way, ready to be processed by numerical codes for hydrological analysis. A remarkable potential of this tool includes the possibility to couple rainfall and discharge series and download them at once, so as daily time series of different hydrologic variables can be automatically synchronized. The procedure is made easier by the aid of an integrated geographical information system (Google Earth) where the location of the gauging station as well as river networks and PET maps

can be simultaneously displayed. Currently, the database is still at an embryonal stage where most of the data refer to the United States. However its potential is much higher, since more information as geography, topography and climate can easily be incorporated in this tool at Global scales.

1.2 Potential evapotranspiration

Evapotranspiration is the sum of evaporation and plant transpiration from Earth's land and ocean surface to the atmosphere. Evaporation accounts for the movement of water to the air from sources such as soil, canopy and waterbodies. Transpiration accounts for the uptake of water by plants and the subsequent loss as vapor through stomata in the leaves associated to the photosynthetic process. Evapotranspiration is a key factor in the water cyrcle: jointly with precipitation and discharge it represents a key ingredient of the water balance at a catchment scales.

Potential evapotranspiration (PET) is a representation of the climatic conditions and, possibly, of the vegetation cover, of a site. PET is function of the amount of energy available at ground level (solar radiation), temperature, wind, air moisture content, and, possibly, soil cover. Many models have been developed in order to quantify PET, they are characterized by different degrees of accuracy, depending on the number of input information required about the primitive drivers of the evapotranspiration process.

Actual evapotranspiration (ET) is the actual amount of water vapour lost by the land surface and vegetation and accounts for the vegetation characteristics, the growing stage and the lack of soil water.

Since the PET plays a fundamental role in the water balance modeling, it is very important to get reliable potential evapotranspiration data for hydrological analysis. It should be noted that accurate estimates of PET and ET are difficult to obtain: all the models developed to predict the magnitude of these terms starting from climate and landscape variables inevitably introduce errors. Hence, in this study two different global databases have been tested and compared, as detailed in the following sections.

1.2.1 MODIS

The 'MODIS global evapotranspiration Project' (MOD16) available from the Montana University (`http://www.ntsg.umt.edu`) includes a dataset providing PET at 1 km^2 resolution for 1009.03 Million km^2 global vegetated land areas at 8-day, monthly and annual time resolution. The dataset cover the time period from 2000 to 2010 but has been extendet to all the periods for which hyfrologic data were available. The MOD16ET datasets are estimated using Mu et al.'s improved ET algorithm over previous Mu et al.'s paper (2007, 2011). The algorithm is based on the Penman-Monteith equation (*Monteith, 1965*).

Evaporation of water intercepted by the canopy is a very important water flux for ecosystems characterized by high LAI (Leaf Area Index). Canopy conductance for plant transpiration is calculated by using LAI to scale stomatal conductance up to canopy level. The required MODIS data inputs include global land cover type classification (*Friedl et al., 2002*), LAI classification (*Myneni et al., 2002*) and albedo (*Luchet et al., 2000*). The input non-satellite data are NASA's MERRA GMAO daily meteorological analysis from 2000 to 2010. The data are given in tenth of millimeter so as to avoid floating numbers: this trick makes the huge high resolution rasters easier to manage.

1.2.2 CGIAR

The 'CGIAR-CSI Global-Aridity and Global-PET Database' (*Zomer et al., 2007*) is a freely available global PET database (`http://www.cgiar-csi.org`). The potentialevapotranspiration has been modeled using the data available from the WorldClim Global Climate Data (*Hijmans et al., 2005*). The WorldClim, based on a high number of climate observations and SRTM topographical data, is a high-resolution global geo-database (30 arc seconds or ∼1km at equator) of monthly average data (1950-2000) of precipitation and minimum and maximum daily temperature. This set of data is insufficient to fully parametrize physically based models for PET (e.g. the FAO Penmann-Monteith equation), though can be appropriate for simpler temperature-based PET equations. Authors tested the performances of less demanding inputs variables models and concluded that Hargreaves method is the most suitable to model PET globally. This method performed almost as well as the FAO-PM, but required less parameters, with significantly reduced sensitivity to error in climatic inputs (*Hargreaves and Allen, 2003*).

Hargreaves (1985) used mean monthly temperature (T_{mean}), mean monthly temperature range (TD) and mean monthly extra-terrestrial radiation (RA, radiation on top of atmosphere) to calculate mean PET as:

$$PET = 0.0023 \cdot RA(T_{mean} + 17.8)TD \cdot 0.5 \tag{1.1}$$

Average monthly and annual PET (mm) layers at spatial resolution of 30 arc-seconds for the 1950-2000 period are calculated using the Hargreaves method with available layers of monthly average temperature parameters from WorldClim database and extra-terrestrial radiation, computed for specific months using the methodology described by *Allen et al. (1998)*. It should be noted that temperature range (TD) is an effective proxy to describe the effect of cloud cover on the quantity of extra-terrestrial radiation reaching the land surface and, as such, it incorporates the effect of more complex physical processes by using easily available climate data at high resolution.

1.3 Data management

All the former information were included in a geographical information system (ESRI ArcGis 10.0). The exact location of the discharge gauges was determined on a detailed map of the river network of the conterminous United States provided by the NOAA (info: `http://www.nws.noaa.gov/geodata/catalog/hydro/metadata/riversub.htm`; download: `https://www.ncl.ucar.edu/Applications/Data/`). The drainage network upstream of the gauging discharge station was then estimated. With the aid of the GIS system embedded in *db@resQ*, a representative rainfall station, located as close as possible to the center of mass of the catchment area and featured by more than about 20 years of measurements was singled out and associated to each streamflow gauging station selected in the study. The reliability of the use of just one rainfall gauge for each catchment was supported by some test which proved that, at least for the sizes of the basins involved ($< 2000km^2$), the spatial variability of daily rainfall statistics is weak, and the use of a single rainfall station doesn't introduce any remarkable bias in the analysis.

Finally, to calculate a representative value of PET for every catchment, a buffer region of $12km$ around the main channel of each basin was delineated within which the average PET was calculated for every PET map. The plots in Figure 1.1 show a

comparison between the values of PET of the two databases extracted in the former manner for the 39 study catchments. The plots are interesting in order to highlight and compare the results of the two models. Figures 1.1 suggest how the MODIS model gives higher estimates of PET with respect to the CGIAR model, both at seasonal and annual time scale. The departure is quite constant and doesn't increase significantly with increasing values of PET.

1.4 Study Catchments

39 catchments for calibration, as well as 4 catchments for validation, have been considered in this study. They are spreaded quite evenly throughout the eastern part of the United States (east the Rocky Mountains). The size of the watersheds span between 40 and 2000 km^2 and include many different climatic regions. All the study catchments are pristine (unregulated) and not impacted by natural or artificial storages (lakes, ponds, reservoirs). Artificial and natural storages may have strong impact on river flow regimes and significantly modify the intra and inter seasonal flow variability by buffering the effect of input (rainfall) and hiding the natural linkage between rainfall and streamflows. Table 1.1 summarizes the main features of the 43 catchments involved in this study. Capital letters are used to identify the catchments used in the validation phase, while numbers from 1 to 39 mark the catchments used for the models calibration.

Figure 1.2 shows the spatial distribution of the 43 catchments across the Eastern US. On the background the CGIR average annual potential evapotranspiration is shown to represent the underlying heterogeneity of climate regimes.

The northern catchments (marked with a dotted circle) experience relevant snow precipitations during winter. The presence of snow impacts significantly the water balance across seasons, in particular by storing water inside the catchment in winter, (when precipitation occurs) and releasing the stored water in spring, when the snow melting makes the runoff coefficient to rise dramatically. Therefore, in the following application of water balance models at the seasonal scale, the results during winter and spring were disregarded.

Table 1.1: Summary information about the 39 catchments used for the calibration

	USGS Code	Name	Area [km^2]	Stremflow Gauging Station	Rainfall Gauging Station	State
1	02316000	Alapaha River	1717	Alapaha	Alapaha Exp Stn	GA
2	05399500	Big Eau Pleine River	580	Stratford	Stratford 1NW	WI
3	06893100	Blue River	170	Kenneth	Stilwell	KS
4	07261000	Cadron Creek	438	Guy	Greenbrier	AR
5	03419800	Calfkiller River	453	Sparta	Sparta Waste Water Plan	TN
6	07172000	Caney River	1153	Elgin	Cedar Vale 5 SSE	KS
7	07021000	Castor River	1090	Zalma	Zalma 4E	MO
8	03539778	Clear Creek	440	Lancing	Lancing 6 NW	TN
9	02380500	Coosawattee River	611	Ellijay	Ellijay	GA
10	08101000	Cowhouse Creek	1167	Pidkoke	Pidcoke	TX
11	02133500	Drowning Creek	474	Hoffman	Jackson Springs 5 WNW	NC
12	02326000	Econfina River	513	Perry	Perry	FLA
13	01170100	Green River	107	Colrain	Whitingham	MA
14	02143500	Indian Creek	179	Laboratory	Lincoln 4 W	NC
15	02143040	Jakob Fork	67	Ramsey	Casar	ND
16	06339100	Knife River	531	Manning	Fairfield	ND
17	01057000	Little Androscoggin River	190	South Paris	West Paris	ME
18	03109500	Little Beaver Creek	1284	East Liverpool	Millport 3NE	OH
19	07346070	Little Cypress Bayou	1748	Jefferson	Harleton	TX
20	03438000	Little River	632	Cadiz	Hopkinsville	KY
21	04270200	Little Salmon River	239	Bombay	Malone	NY
22	03216500	Little Sandy River	1036	Grayson	Grayson 3 SW	KY
23	06478540	Little Vermillion River	204	Salem	Montrose	SD
24	04115000	Maple River	1124	Maple Rapids	Saint Johns	MI
25	08189500	Mission River	1787	Refugio	Beeville 5 NE	TX
26	03326500	Mississinewa River	1766	Marion	Marion 2 N	IN
27	07144780	Nf Ninnescah River	1847	Caheny	Hutchinson 10 SW	KS
28	03346000	North Fork Embarras River	824	Oblong	Casey	IL
29	08160800	Redgate Creek	45	Columbus	Columbus	TX
30	02034000	Rivanna River	1717	Palmyra	Charlottesville 2 W	VA
31	05060500	Rush River	300	Amenia	Casselton Agronomy	ND
32	06803000	Salt Creek	433	Roca	Roca	NE
33	07185765	Spring River	1101	Carthage	Mt Vernon	MO
34	04021520	Stoney Brook	192	Brookston	Cloquet	MN
35	07375500	Tangipahoa River	1673	Robert	Hammond 5 W	GA
36	01518000	Tioga River	730	Tioga	Covington 2 WSW	GA
37	02462000	Valley Creek	383	Oak Grove	Bankhead Dam	GA
38	01343060	West Canada Creek	668	Wilmurt	Piseco	NY
39	06444000	White River	811	Crawford	Ft Robinson	NE
A	03075500	Youghiogheny River	347	Oakland	Oakland 1 SE	MD
B	03539600	Daddys Creek	360	Hebbertsburg	Crossville mep ap	TN
C	07257006	Big Piney Creek	793	Highway 164	Deer	AR
D	03052500	Sand Run	37	Buckhannon	Buckhannon	WV

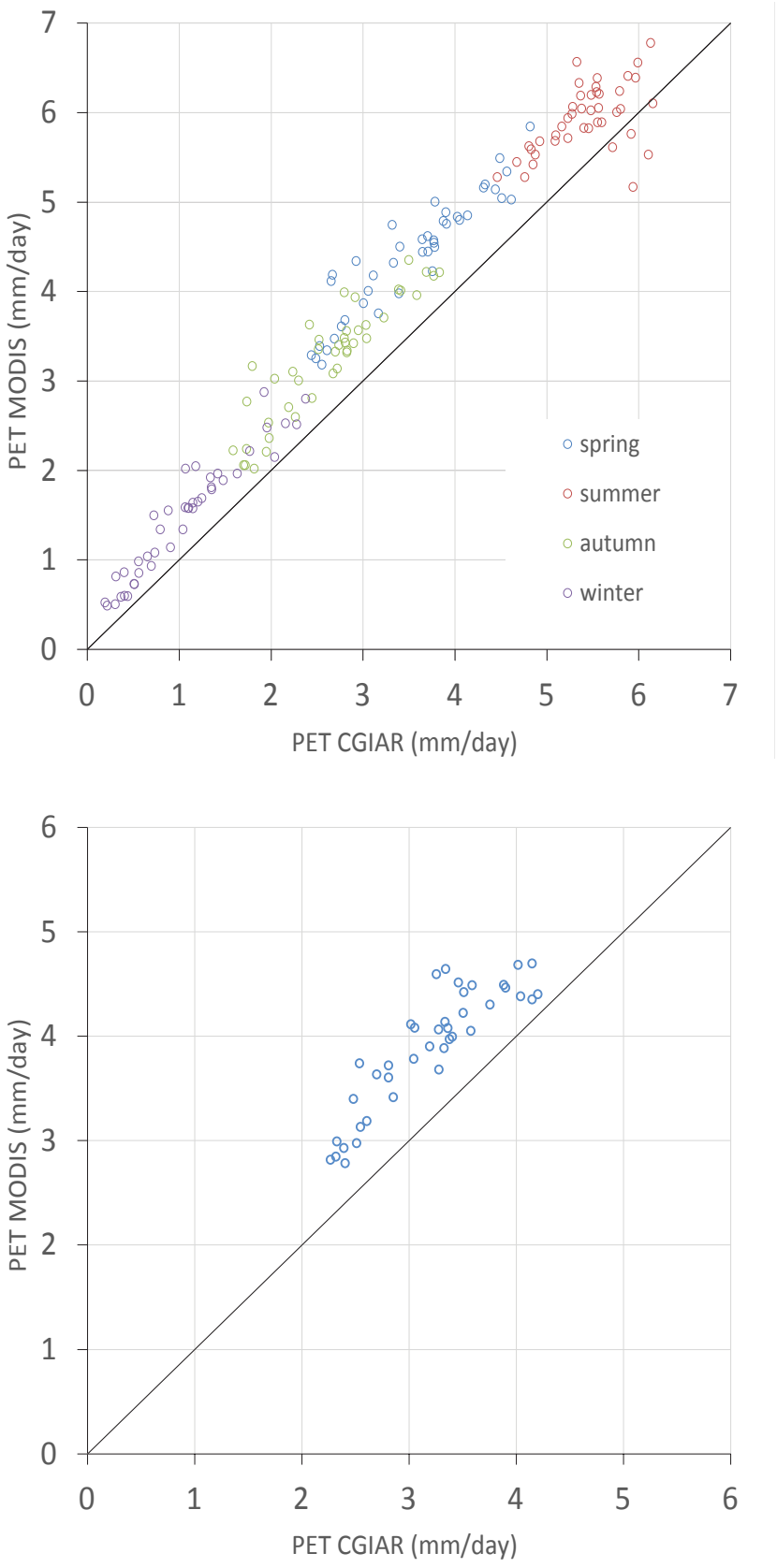

Figure 1.1: *PET comparison*

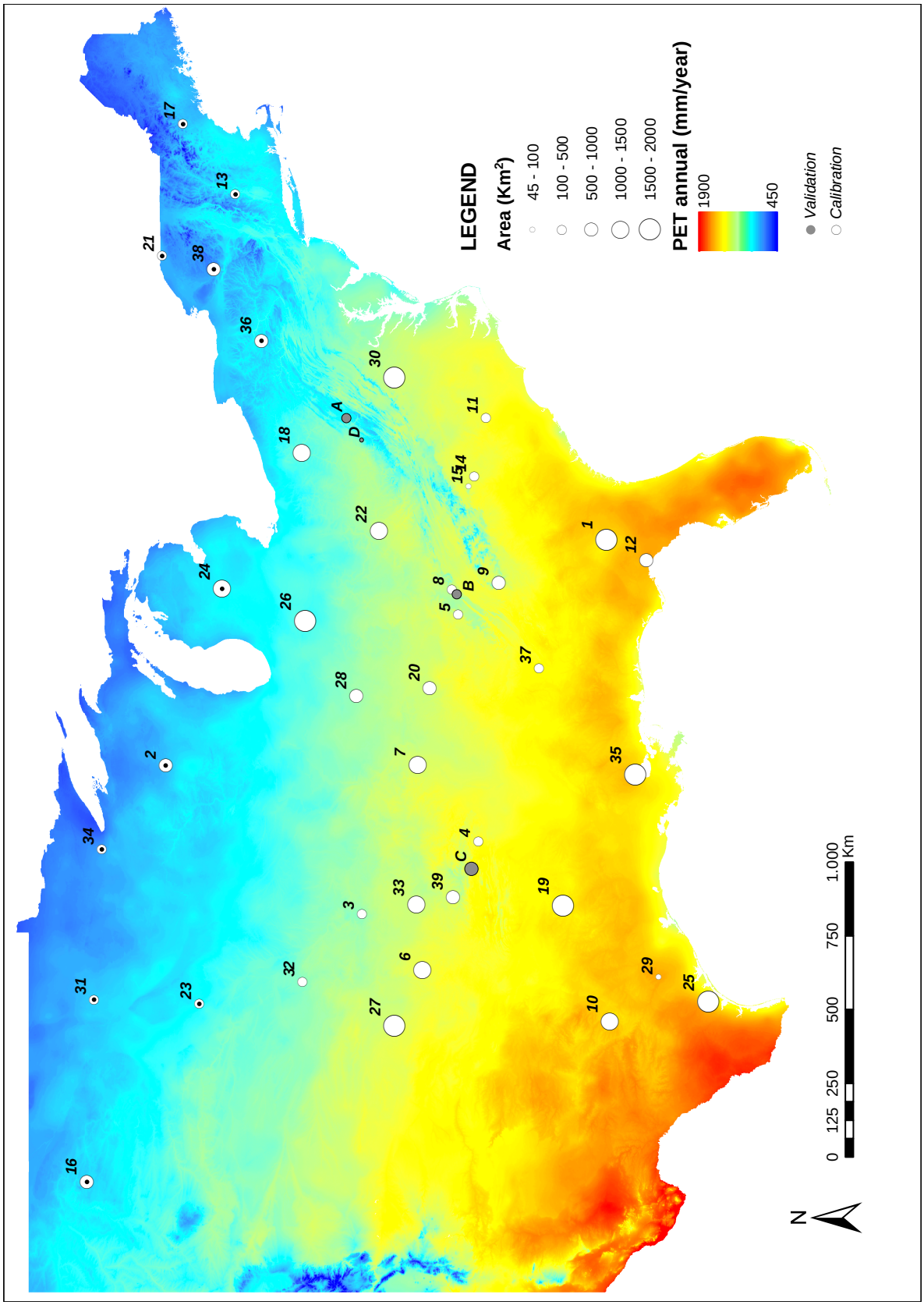

Figure 1.2: *Map of the catchments.*

Chapter 2

Methods

2.1 The water balance

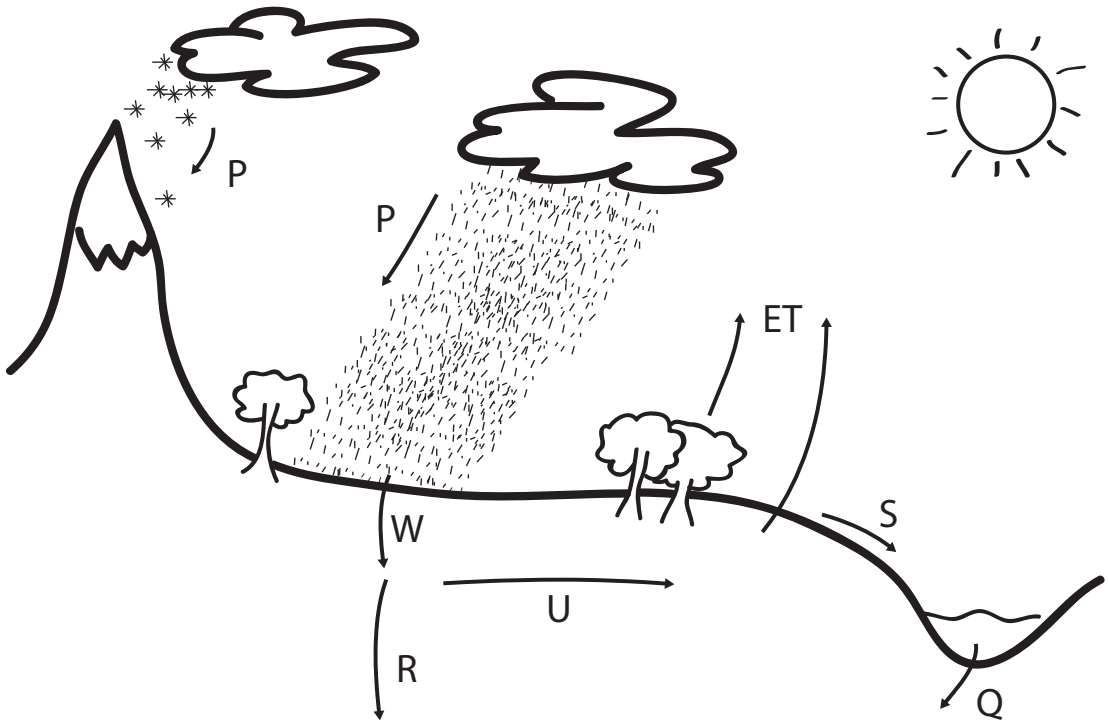

Figure 2.1: *Water balance scheme*

In hydrology water balance equations can be used to describe the flow of water across a given system. Such a system can be constituted by different types of

hydrological domains, such as a soil column or an entire drainage basins. Since mass is a conservative quantity in low energy physical processes (not involving nuclear interactions), a mass balance equation can be easily established for any well defined control volume. When the control volume is represented by a catchment, the water balance equation can be written as:

$$\Delta W = (\langle P \rangle - \langle Q \rangle - \langle ET \rangle - \langle R \rangle)\Delta T \ , \tag{2.1}$$

were ΔW is the difference of water stored inside the system in the time interval ΔT, $\langle P \rangle$ is the average precipitation during ΔT (the only input of the system) while $\langle Q \rangle , \langle ET \rangle , \langle R \rangle$, are the time averages of the streamflow at the catchment outlet, evapotranspiration and recharge (i.e. the system's outputs over the time interval ΔT).

Precipitation potentially includes both snowfall and rainfall, even though snow dynamics are not explicitly taken into account in the water balance models investigated in this thesis. Discharge through the catchment outlet is the sum of the subsurface slow flow and the quick response due to the surface runoff ($Q = U + S$). The wetting component of precipitation i.e. $W = P - S$ is defined as the fraction of rainfall infiltrating into the soil. Water infiltrated can be stored for long times in the soil if it's matric potential is low enough, thereby becoming prone to be kept by plant roots and evapotranspired. Since ET is driven by climate and vegetation conditions, vegetation plays a key role in water balance. Alternatively, the infiltrated water can be released to the drainage network as subsurface runoff (U) or recharge (R). Figure 2.2 provides a scheme of the water balance partitioning described above.

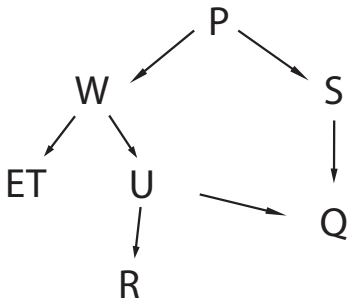

Figure 2.2: *Water fluxes*

The accumulation term, ΔW, includes water stored in the soil and in surface

water bodies. All the models that have been tested in this study explicitly consider the storage of water as soil moisture. Other factors like lakes, artificial storages and snow packs may dramatically impact the water balance, especially at the seasonal time scale. To circumvent this issue, catchments with ponded areas, lakes or artificial basins, aw well as areas (or seasons) strongly affected by snow dynamics, have not been considered.

In agreement with most of the recent literature, the models considered in this thesis assume that all the non-vaporized wetting component W is released as streamflow or accumulated as soil moisture. This means that recharge is neglected. Hence, the analysis should be limited to catchments with a reduced inter-catchment groundwater exchange.

Under these assumptions, and further assuming that both at annual and seasonal time scales the intra-seasonal change of storage is negligible with respect to the underlying input-output flows, Equation (2.1) simplifies into:

$$\langle P \rangle = \langle Q \rangle + \langle ET \rangle \tag{2.2}$$

which is the basis for most physically-based models used in this study.

2.1.1 Rainfall partitioning

The problem of estimating the amount of surface runoff and, hence, the partitioning of rainfall between wetting and surface runoff is a major issue in hydrological modeling. The presence of surface runoff reduces the amount of water infiltrating into the soil and, as a consequence, the actual evapotranspiration (ET) and the base flow (U). The result is a quicker response of a catchment to rainfall and an increase of the runoff coefficient, as surface flows are directly conveyed to the stream without being processed by the soil matrix and vegetation.

From a physical perspective the occurrence of surface runoff strongly depends on topography, geology, soil texture, vegetation, and climate (e.g. frequency and intensity of rainfall events, evapotranspiration rates etc.). Surface runoff is promoted by steep saturated slopes or impervious areas without vegetation in response to prolonged and intense rainfall events.

Some of the models discussed in this work require the prior knowledge of surface runoff, while in other cases such information is not required because surface runoff processes are not explicitly modeled. However, empirical partition models to identify

S from P and Q can be applied to both types of models, so as a water balance equation analogous to eq. (2.1) can be written in terms of wetting, infiltration and base flow, while surface runoff volumes are assumed to entirely contribute to streamflows.

A simple and effective method to perform the partitioning is represented by the following analytical one parameter recursive filter (*Lyne and Hollick, 1979*):

$$\begin{cases} U_k = aU_{k-1} + \frac{1-a}{2}(Q_k + Q_{k-1}) \\ U_k \leq Q_k \end{cases} \tag{2.3}$$

where U and Q are, respectively, the slow component of the streamflow and the overall streamflow itself ($Q = S + U$), while the subscript k identifies the k-th step in the time series(e.g. the k-th day from the beginning of the time series). The value assigned to a is 0.925, in agreement with the literature (*Sivapalan et al., 2011; Voepel et al. 2011*).

The filter singles out the base flow U cutting the highest and steepest peaks of the hydrograph, assuming that the quick response of the catchment caused by surface runoff. This simple model overcomes quite effectively all the difficulties lead by a proper physically based description of the phenomenon. Though, significant overestimation of surface flows may be introduced in presence of fast subsurface flows. Figure 2.3 shows an example application of the separation model. It should be noted that this method can not be applied in the absence of streamflow measurements.

2.2 Water balance models: overview

Five existing water balance models were tested and compared by analyzing their ability to predict observed runoff coefficients in the Eastern United States. The models considered include empirical, semiempirical and physically-based models. Each model has a different number of variables and parameters, which were calibrated in order to maximize model performances. In doing that, physically meaningful bonds for the model parameters have been imposed. Provided that the outcomes of the

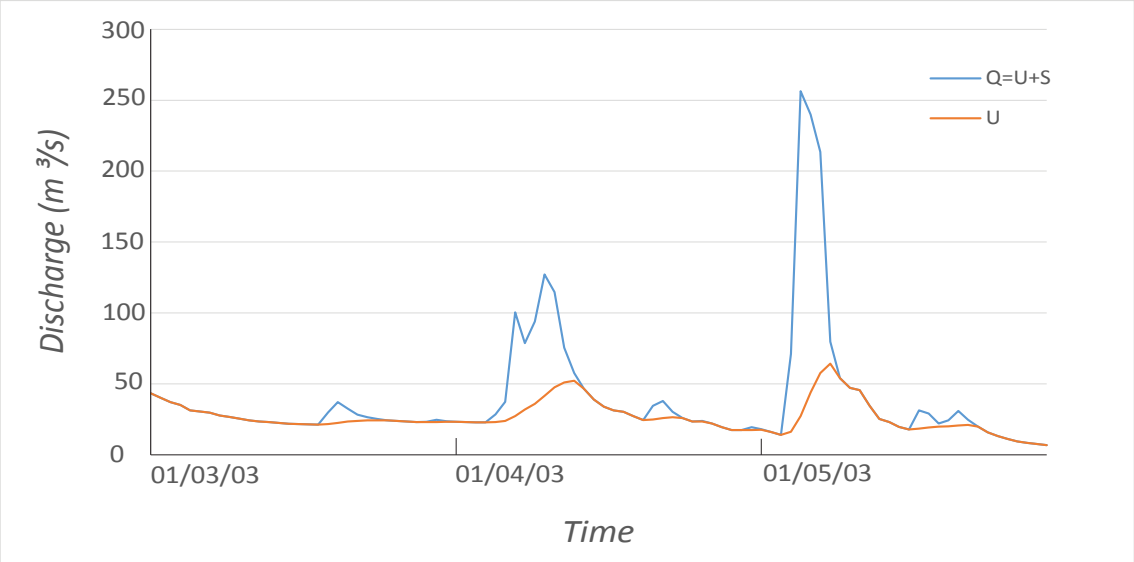

Figure 2.3: *Filter performances in spring (Calfkiller River)*

research is a regional-scale prediction of flow regimes, the model parameters were assumed to be spatially homogeneous within the entire study region, so as the calibrated parameters can be potentially exported to any catchment within the Eastern United States.

Table 2.2 presents the 5 models included in this study and the corresponding reference papers where they are described:

Table 2.1: *Models*

Code	**Relevant references**	**Type**	**Number of parameters**
WB1	Budyko, 1974	Empirical	0
WB2	Porporato et al., 2004	Physically-based	1-4
WB3	P.C.D. Milly, 1994	Physically-based	2
WB4	Sivapalan et al., 2011	Semi-empirical	4
WB5	Trabucco et al., 2010	Physically-based	0

Model performances have been evaluated by properly taking into account the different number of parameters of each model (see section 2.9).

In this study the uppercase symbols (Q, P, ET etc.) denote instantaneous values of hydrological variables, while brackets ($\langle Q \rangle, \langle P \rangle, \langle ET \rangle$ etc.) are used to denote their temporal averages on a specific time interval (ΔT), such as a season or a year, according to the following definition.

$$\langle \bullet \rangle := \frac{1}{\Delta T} \int_{\Delta T} \bullet(t) dt \tag{2.4}$$

where $\bullet$ is a generic function of time. Since spatial variability within a single catchment was not explicitly taken into account (except in one of the models), the hydrological fluxes are referred to spatial averages within the catchment area, otherwise mentioned.

The basic concepts and the physical processes described by each model are presented in the next sections. Some models displays common features and similar working hypothesis. For example, they all assume that vegetation cover is sufficiently extensive such as direct evaporation from the soil need not be considered; however they can differ in crucial features of the formulation or even in the model typology they belong to (e.g. conceptual vs empirical).

2.3 WB1

M. I. Budyko (1920-2001) was a Russian climatologist and one of the founders of physical climatology. He pioneered studies on global climate and calculated the Earth temperature through simple physically-based equilibrium model in which the incoming solar radiation absorbed by the Earth's system is balanced by the energy emitted to space as thermal energy. Budyko's grounbreaking book, *'Heat balance of the Earth's surface'* published in 1956 transformed climatology from a qualitative into a quantitative physical science. These new methods were quickly adopted by climatoliogist around the world. For what concerns the water balance, his subsequent book *'Climate and Life' (1974)* played a key role in delineating for the first time the drivers of annual water balance at catchment scales. In his work he presents a semi-empirical relation which looks at the ratio between the annual average ac-

tual evapotranspiration and the annual average rainfall ($\langle ET \rangle / \langle P \rangle$) as a non-linear function of the 'Dryness Index' (D_I), defined as the ratio between annual average potential evapotranspiration and the annual average rainfall ($\langle PET \rangle / \langle P \rangle$). This approach, considers the dryness index as the first order driver of the water balance and is a direct consequence of Budyko's former works in as much it combines the energy supply and demand into a single index. The amount of energy supplied is expressed as the maximum amount of water that can be vaporized under the observed climate conditions, while the demand is the effective availability of moisture that could potentially be vaporized, which is again a function of the climate.

Budyko estimated the ratio $\langle ET \rangle / \langle P \rangle$ as a function of D_I by fitting the former quantities for hundreds of basins all around the World. The (annual) actual average evapotranspiration, $\langle ET \rangle$, was simply computed, based on rainfall and discharge data, as $\langle ET \rangle = \langle P \rangle - \langle Q \rangle$.

These quantities are temporally averaged values throughout the year and the relation between the ratio $\langle ET \rangle / \langle P \rangle$ and D_I is assumed to hold in the long term.

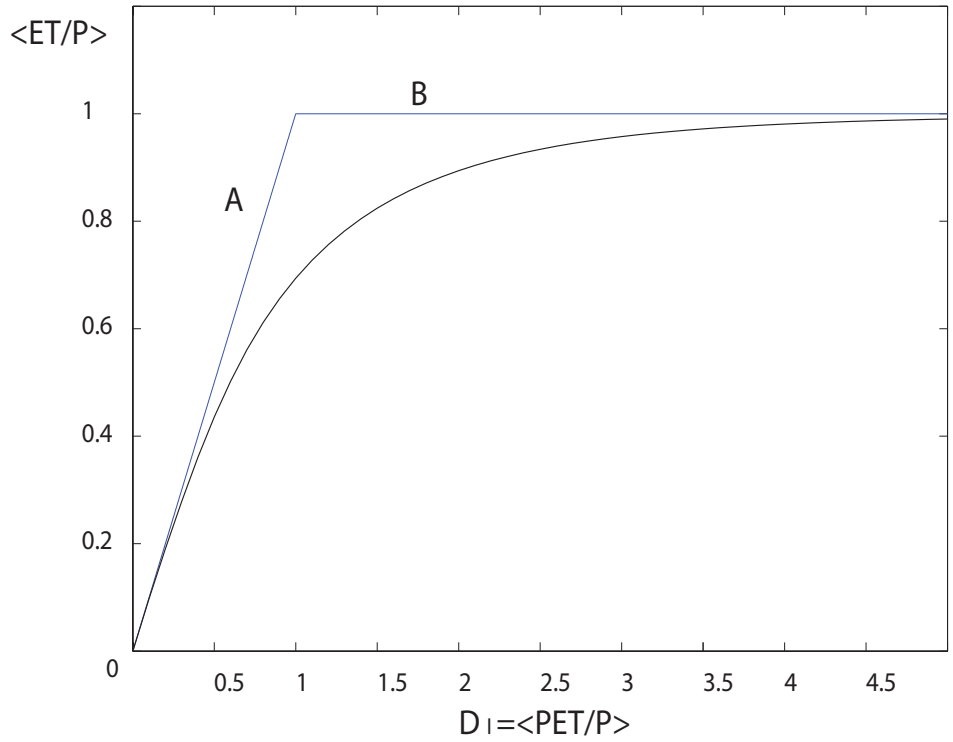

Figure 2.4: *Semiempirical Budyko's curve*

Figure 2.4 displays the semi empirical relation between $\langle ET \rangle / \langle P \rangle$ and D_I de-

veloped by Bydyko, whose equation can be written as:

$$\frac{\langle ET \rangle}{\langle P \rangle} = \left[D_I (1 - e^{-D_I}) \, tanh \left(\frac{1}{D_I} \right) \right]^{0.5} \tag{2.5}$$

where *tanh* is the hyperbolic tangent.

The two straight lines A and B show two theoretical end-members:

- The line A represents the 'Energy limited upper bound': the actual evapotranspiration must be always lower than the potential evapotranspiration regardless of the crop considered, provided that water is tightly bounded to dry soils.
- The line B represents the 'Water limited upper bound': the vaporized water can never exceed the supply (precipitation), no matter how large is D_I.

The Budyko curve represents a very simple and effective way to estimate the annual runoff coefficient, based on rainfall and PET data, through the following relationship:

$$\frac{\langle Q \rangle}{\langle P \rangle} = 1 - \frac{\langle ET \rangle}{\langle P \rangle} = 1 - D_I \left[D_I (1 - e^{-D_I}) \, tanh \left(\frac{1}{D_I} \right) \right]^{0.5} \tag{2.6}$$

The model was included in this study because it is parameter free and is considered a sound milestone in the scientific community.

MODEL PARAMETERS: In this model the only variable involved is D_I, which depends on rainfall and potential evapotranspiration. Since rainfall is measured in climatic stations, while the PET used in this study comes from either the MODIS or the CGIR datasets, no parameter need to be calibrated.

2.4 WB2

This model follows a physically-based minimalistic approach, where the soil-plant-atmosphere system is represented by focusing on the temporal fluctuations of water availability in soil. The water stored within the soil is seen as a stochastic state variable that governs the water balance at a catchment scale.

The terrestrial hydrologic cycle is a manifold system whose understanding requires a massive amount of observations, simple or detailed models and demanding

theoretical/numerical analysis: these tools are needed to overcome the inherent difficulties of dealing with a complex non linear system with essential stochastic components and feedbacks at multiple timescales.

The awareness in the complexity of developing a model for the water balance led to a simplified approach where the role of the parameters that control soil moisture dynamics clearly emerges, offering a theoretical framework whose generality could surpass that of more complicated models that require cumbersome numerical simulations. The model offers a low dimensional description of key hydrological processes where the dominant deterministic (and possibly non linear) components are separated from high-dimensional (i.e. stochastic) environmental forcing.

The aim of this model is to offer a very parsimonious yet realistic representation of the water balance that captures the essential components: the water holding capacity of the soil, which is a function of the soil and root characteristics and is responsible for the threshold-like non linearity that triggers deep infiltration and surface runoff; the soil-moisture dependence of evapotranspiration and photosynthesis; and the intermittency and unpredictability of rainfall, whose variability in terms of both frequency and depth of events proves to be crucial.

Soil moisture dynamics are interpreted and modeled at daily time scales, by conceptualizing the soil as a reservoir with a finite storage capacity intermittently filled by rainfall events in the form of random pulses featured by random depth. Water losses occur via evapotranspiration, deep infiltration and surface runoff. Vertical gradients of soil moisture are neglected, assuming that the propagation of the wetting front and the soil moisture redistribution over rooting zone are negligible at the daily timescale employed by the formulation.

The water balance equation is recasted here in terms of the relative soil moisture s (dimensionless) vertically averaged over the rooting depth Z_r $[L]$. Accordingly, the total volume of soil water per unit ground area at a give time t is $s(t)nZ_r$ $[L]$, where n is the vertically averaged soil porosity (volume of voids/total volume). The model assumes that when s exceeds a given threshold s_1, the rainfall in excess is lost by vertical drainage. The empirical parameter s_1 depends on the type of soil and is typically comprised between the so-called field capacity (i.e. the soil moisture level below which drainage is negligible) and complete saturation ($s = 1$). In practice, s_1 defines the soil holding capacity. As all the incoming precipitation is assumed to infiltrate within the soil up to saturation, the present approach is effective when the Dunne or saturation-from-below mechanism of runoff formation is dominant

compared to the Hortonian runoff (rainfall intensity exceeding the soil saturated hydraulic conductivity); this is often the case for vegetated surfaces with negligible topography and absence of soil crusting.

Evapotranspiration, ET $[L/T]$ is assumed to decrease linearly from a maximum value (potential evapotranspiration PET) under well watered conditions ($s = s_1$) to 0 at the wilting point ($s = s_w$). The reduction of evapotranspiration with decreasing soil moisture is a well-established fact that can be ascribed to increased resistance to soil water transport within the soil-plant-atmosphere continuum when soil water potential is reduced. While a marked non-linearity characterizes the evapotranspiration-soil moisture relationship pointwise, at larger scales the temporal variability and spatial heterogeneity of hydrological processes tends to significantly linearize such relation (*Wetzel and Chang, 1987; Crow and Wood, 2002*)

Rainfall input, $P(t)$ $[L/T]$, is modeled as a marked Poisson process with frequency λ_p $[T^{-1}]$. Each rain event carries a random depth of rainfall, exponentially distributed with mean α $[L]$. Such a model has been shown to provide a simple yet realistic representation of rainfall at the daily timescale for different hydroclimatic regimes (*Milly, 1993; Rodriguez-Iturbe et al., 1999*). According to this model, the cumulated rainfall amount during a time interval ΔT is $\lambda_P \alpha \Delta T$.

Accordingly to the modeling sheme described above, the soil moisture balance equation can thus be written as:

$$nZ_r \frac{ds}{dt} = P(t) - ET[s(t)] - L[s(t), t] \tag{2.7}$$

Because of the forcing term $P(t)$, Equation (2.7) is a stochastic differential equation that requires a solution in probabilistic terms.

If $x = (s - s_w)/(s_1 - s_w)$ is the normalized soil moisture and $w_0 = (s_1 - s_w)nZ_r$ the maximum soil water storage available to plants, the governing quantities of the process are w_0, α, λ_p and PET. According to dimensional analysis, these quantities can be grouped into two dimensionless numbers: $\gamma = w_0/\alpha$ and $\lambda_P/\eta = (\lambda_p w_0)/\langle PET \rangle$ (or $D_I = (\gamma\eta)/\lambda_p = \langle PET \rangle / \langle P \rangle$), being D_I the Budyko's dryness index, η the normalized evapotranspiration loss under well watered conditions ($\eta = \langle PET \rangle / w_0$) and $\langle P \rangle$ the mean rainfall rate ($\langle P \rangle = \alpha\lambda_p$). From a physical perspective, this implies that the terrestrial water balance is governed by three factors: i) the ratio between the soil storage capacity and the mean rainfall input per event γ; ii) the ratio between the maximum evapotranspiration and the mean rainfall rate (i.e. the

dryness index D_I) and iii) the ratio between the rate of occurrence of rainfall events and the maximum evapotranspiration rate λ_p/η. Such dimensionless groups define the interaction of the most important climate, soil, and vegetation parameters in controlling soil moisture dynamics.

Following *Rodriguez-Iturbe et al.(1999)*, the master equation of the probability density function (PDF) of x can solved analytically for steady-state conditions. The result is a truncated Gamma distribution:

$$p(x) = \frac{N}{\eta} x^{\left(\frac{\lambda_p}{\eta}-1\right)} e^{-\gamma x} \tag{2.8}$$

Where N is the normalization constant whose analytical expression is:

$$N = \frac{\eta \gamma^{\frac{\lambda_p}{\eta}}}{\Gamma(\lambda_p/\eta) - \Gamma(\lambda_p/\eta, \gamma)} \tag{2.9}$$

Being $\Gamma(\cdot)$ and $\Gamma(\cdot,\cdot)$ the complete and incomplete Gamma functions (*Abramowitz and Stegun, 1964*). The mean effective relative soil moisture is expressed as:

$$\langle x \rangle = \frac{\lambda_p - N e^{-\gamma}}{\eta \gamma} \tag{2.10}$$

Therefore, the normalized water balance can be written as

$$1 = \frac{\langle ET \rangle}{\langle P \rangle} + \frac{\langle L \rangle}{\langle P \rangle} = D_I \langle x \rangle + \frac{\langle L \rangle}{\langle P \rangle} \tag{2.11}$$

where $\langle ET \rangle = \langle x \rangle \langle PET \rangle$

Equations (2.11) and (2.10), describe the partitioning of the rainfall input into evapotranspiration and deep infiltration plus runoff as a function of the governing climate, soil and vegetation parameters. While initially conceived at a point spatial scale, this model can be also interpreted at catchment scale (*Settin et al., 2007*).

Under the assumption that all the water accounted in the loss term L crosses the control section as discharge Q, the mean runoff coefficient can be written as:

$$\frac{\langle Q \rangle}{\langle P \rangle} = \frac{\eta \gamma^{\frac{\lambda_p}{\eta}} e^{-\gamma}}{\lambda_p \left(1 - \Gamma(\lambda_p/\eta, \gamma)\right)} = \frac{D_I \gamma^{\frac{\gamma}{D_I}} e^{-\gamma}}{\gamma \left(1 - \Gamma(\gamma/D_I, \gamma)\right)} \tag{2.12}$$

where D_I is the dryness index.

The plot represented in Figure 2.5 compares the theoretical solution of the water balance model described above for different values of the parameter $\gamma = w_0/\alpha$.

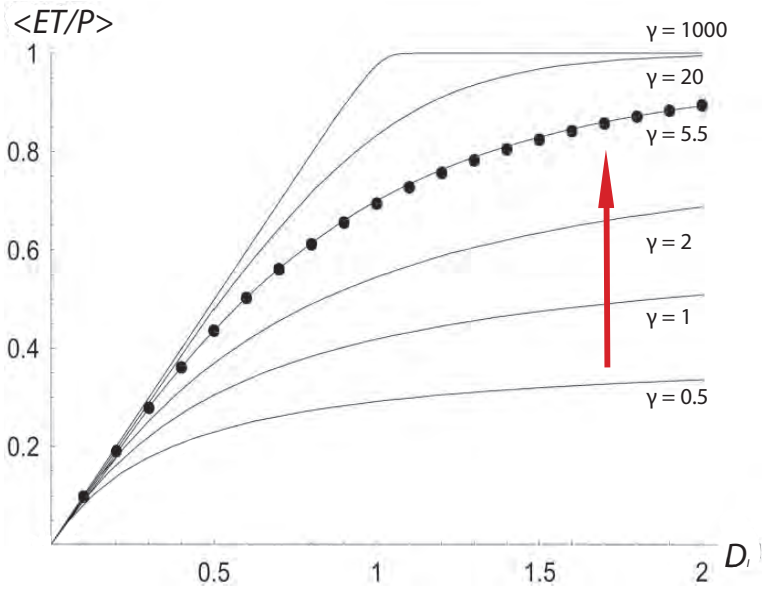

Figure 2.5: *Fraction of total rainfall lost by evapotranspiration as a function of Budyko's dryness index for different values of the parameter* γ

MODEL PARAMETERS: The key variable of the model is D_I, that is a function of the measured average rainfall characteristics (embodied in λ_p and α), and of the potential evapotranspiration, whose value is provided by one of the two PET datasets (i.e. CGIR or MODIS). Being $\gamma = w_0/\alpha$, the only parameter of the model is $w_0 = (s_1 - s_w)nZ_r$. In this study the calibration was performed on Z_r, assuming fixed values for s_1, s_w and n (s_1=*0.5*, s_w=*0.2*, n=*0.35*).

2.5 WB3

Model WB3 is based on the hypothesis that the long-term water balance is determined by the local interaction of fluctuating water supply (precipitation) and demand (potential evapotranspiration), mediated by water storage in the soil. Hence, the approach is similar to that underlying WB2. Peculiarity of the model is the adoption of an idealized representation of time and space variability of the most relevant hydrologic variables, providing a simple water balance equation for a given

area under spatially uniform climate and vegetation conditions. The partitioning of average annual precipitation into evapotranspiration and runoff is assumed to depend on seven dimensionless factors:

- The ratio of spatially-averaged annual potential evapotranspiration to average annual precipitation (dryness index D_I);
- The mean numbers of precipitation events per year $< N >$ (which is a frequency factor analogous to λ_P in WB2);
- The ratio of spatially averaged plant-available water holding capacity, w_0, of the soil to the annual average precipitation, $\alpha < N >$;
- The shape parameter of the Gamma distribution describing spatial variability of storage capacity, k;
- Simple measures of seasonality of mean precipitation intensity, storm arrival rate and potential evapotranspiration.

The model focuses on the following aspects of the water balance: in humid areas (dryness index < 1) the dominant factor producing runoff is the excess of annual precipitation over annual potential evapotranspiration, even though runoff caused by variability of supply and demand over time may be also significant; in arid regions (dryness index > 1), instead, most of runoff is caused by forcing variability over time. The model represents the first attempt to provide a physically-based explanation of the departure of Budyko's curve from the two theoretical asymptotes, and understand the reason of the considerable scatter of observations around that curve. From a physical basis, the most likely reason underlying the observed gap from the horizontal asymptote, (i.e. the runoff coefficient), has to be sought in two related characteristics of land surface: its finite water storage capacity and its finite permeability. For example, if the water storage capacity of soil is too small, temporary excesses of water supply will be lost as surface runoff, leading to an increase of the runoff coefficient even though the dryness index of the area exceeds 1. Finite-permeability effects are involved in two different ways,both of which increase runoff at the expense of evapotranspiration: i) if precipitation rates exceed the rates at which water can infiltrate the soil, the runoff will occur regardless of the underlying long-term water and energy supplies; ii) if potential evapotranspiration rates exceed the rates at which water within the root zone can travel the short

distances to plant roots (or to the surface of bare soil), then evapotranspiration may fall below its water and energy supply limits.

The working hypothesis is that water balance can be described as the simple interaction of water supply, demand, and finite soil water storage. Variabilities in time associated to both seasonality and storminess are included in the formulation, as well as spatial variability of storage capacity. Conversely, finite permeability effects are ignored.

By explicitly resolving intraseasonal, interseasonal and spatial variabilities, the approach avoids the introduction of empirical parameters. The conceptual simplicity of the approach allows the development of analytic solutions for particular cases and provides a dimensionless formulation of the problem even in its most general case.

The model is here presented in its general formulation, which doesn't have a close analytical solution and can be solved by means of cumbersome Monte Carlo simulations. However, once additional assumptions about the time variability of the forcing factor are introduced, it is possible to get analytical solutions, which are easy to be implemented and included in a computer code.

The starting point for the analysis is a local storage model for the water balance of the root zone. The average water balance is then obtained by integration of a continuous water balance equation with respect to time. This integration recognizes both the deterministic seasonal variability and the random, intraseasonal variability of atmospheric forcing of the surface. The areal mean of the long-term balance is finally achieved by integration in space. The latter is performed using a distribution function to describe spatial variability of surface characteristics but ignoring spatial variability of climate.

The uppercase symbols P, PET, Q denote the instantaneous specific (per unit of area) water fluxes associated to precipitation, potential evapotranspiration, and root zone drainage (discharge). These variables fluctuate at daily (and shorter) time scales because of the random nature of atmospheric processes, but also display clear seasonal cycles. Angle brackets (e.g. $\langle P \rangle$) are used to denote temporal averages.

In agreement with WB2, the reference control volume is bounded above by the soil-atmosphere interface and has sufficient vertical extent to contain essentially all of the water readily available to vegetation uptake and transpiration. The vertical extent corresponds approximately to the average depth of rooting of the predominant plants, which is typically around 1m. The horizontal extent of the control volume is sufficiently large to reflect the effect of horizontal root zone water fluxes, induced by

soil heterogeneity and topographic curvature. The mass balance of water for such a control volume, expressed in terms of equivalent liquid water depth and volumetric flux rates, is:

$$\frac{dw}{dt} = P - ET - Q \tag{2.13}$$

Where $w = Z_r n\, s(t)$ is the depth of water stored and P is the rate of infiltration of liquid precipitation, which is assumed to be equal to precipitation itself, implying that the presence of frozen precipitation, snowmelt and snowpack are implicitly ignored. The following assumptions are then introduced:

1. The soil is sufficiently permeable to allow all liquid precipitation and snowmelt to infiltrate;

2. All soil soil water stored at potentials greater than the permanent wilting point is readily depleted at the potential evapotranspiration rate;

3. All water stored in excess of a well-defined field capacity is rapidly removed from the control volume by drainage;

4. No drainage occurs when the average soil moisture content falls below the field capacity.

For a well-developed vegetation cover the storage capacity w_0 may be interpreted as a depth integral over the root zone of the difference between the volumetric moisture contents of the soil at field capacity (ns_1) and at the wilting point (ns_w), being n the average porosity of the soil within the root zone. Let $r(z)$ denote the fraction of area at depth z that is affected by the root system of the vegetation; in principle, this fraction depends in a complicated way on rooting density, hydraulic properties of the soil and timescale (i.e., seasonal versus storm/inter-storm) of the uptake process. Typically, it is assumed that $r(z)$ steps down from 1 to 0 at some well-defined rooting depth Z_r, in which case the water holding capacity w_0 is

$$w_0 = nZ_r(s_1 - s_w) \tag{2.14}$$

which is the parameter that controls the water balance in this formulation.

Variability over time The dominant mechanism that controls the seasonality of climate is the periodicity of the solar irradiance normal to the top of the atmosphere. At extratropical locations, this produces a strong signal with a dominant period of 1 year in most climatic regions. Therefore is assumed that:

$$P(t) = \langle P \rangle (1 + \delta_p \sin \omega t) + \xi_p(t) \tag{2.15}$$

$$ET(t) = \langle ET \rangle (1 + \delta_e t \sin \omega t) \tag{2.16}$$

Where δ_p and δ_{et} are the ratios of the amplitudes of the annual harmonics to the annual averages of P and PET and $\xi_p(t)$ is the random component of the signal due to the natural stochasticity of the rainfall events. Conversely, the random component of the evapotranspiration is disregarded. With $2\pi/\omega$ equal to one year, these expressions capture the essential features of the annual land surface hydrologic forcing outside the tropics. Concerning the random component of the rainfall signal, it is assumed that precipitation arrives in discrete events that we shall call storms, that the arrival of these storms in time is a Poisson process, and that the amount of precipitation in any storm is governed by the exponential distribution. The mean storm arrival rate is allowed to vary seasonally with only the annual harmonic retained:

$$N(t) = \langle N \rangle (1 + \delta_N \sin \omega t) \tag{2.17}$$

Where $\langle N \rangle = 365\lambda_p$ (being λ_p the rainfall frequency in day^{-1}) represents the average number of rainy days per year. The expected value of storm depth at any time of the year is simply $\langle P \rangle / \langle N \rangle$.

Variability over space To characterize the catchment water balance, it is desirable to integrate Equation (2.13) in space. In doing that, spatial variability of both climatic and soil factors must be considered. In this analysis, however, the variability over space of the statistics of point atmospheric forcing is ignored, thus the local values of P, PET, N are everywhere equal to their areal means. It is well known, however, that the soil hydraulic characteristics vary greatly at relatively small scales. The nonlinear dependence of water balance on w_0 suggests the need

for explicit consideration of spatial variability of w_0. It is assumed that the distribution of water-holding capacity within a given area, $f_w(w_0)$, is given by a Gamma distribution. This distribution is very flexible and analytically tractable:

$$f_w(w_0) = \frac{\lambda(\lambda w_0)^{k-1} e^{-\lambda w_0}}{\Gamma(k)} \tag{2.18}$$

According to eq. (2.18) , the mean of w_0 is k/λ and its coefficient of variation is $k^{-0.5}$. The spatial mean of any function of w, $Z(w)$ (such as evapotranspiration and runoff) are assumed to be spatially averaged over the density function (2.18).

An analytical solution of the general water balance problem formulated above has not been found. However, when the variability over time of atmospheric forcing is ignored or is limited to either the seasonal or random components alone, analytic solutions can be derived. The methodology required to obtain the general solution by Monte Carlo simulation is not included herein and can be found in the related paper. The trivial analytical solutions in the cases of reduced temporal variability of the hydrological forcings and large storm arrival rate ($\delta_p = \delta_N = \delta_e t = 0; \langle N \rangle \to \infty$) are not presented. The analytical solution presented below and used in this study, instead, refers to the case where seasonal cycles of climatic variables are neglected. ($\delta_p = \delta_N = \delta_e t = 0$).

The analytical solution of the water balance equation in the simple case where spatial variability of soil properties is neglected ($k \to \infty$) reads:

$$\frac{\langle Q \rangle}{\langle P \rangle} = 1 - \frac{e^{\langle PET \rangle \langle N \rangle (1 - D_I^{-1})} - 1}{e^{\langle PET \rangle \langle N \rangle (1 - D_I^{-1})} - D_I^{-1}} \tag{2.19}$$

Where D_I is the dryness index. Instead, the more general result in the case where spatial variability of soil storage capacity is considered (which is the analytical solution adopted in the following analysis) reads:

$D_I < 1$

$$\frac{\langle Q \rangle}{\langle P \rangle} = 1 - (1 - D_I) \sum_{j=0}^{\infty} \left[1 + j\gamma (D_I^{-1} - 1) k^{-1}\right]^{-k} {D_I}^j \tag{2.20}$$

$D_I > 1$

$$\frac{\langle Q \rangle}{\langle P \rangle} = 1 - (1 - D_I) \sum_{j=0}^{\infty} \left[1 + (j+1)\gamma(1 - D_I^{-1})k^{-1}\right]^{-k} {D_I}^{-j} \qquad (2.21)$$

MODEL PARAMETERS: According to this model, the water balance is ruled by $\langle P \rangle = \langle N \rangle \alpha, \langle PET \rangle, w_0 = nZ_r(s_1 - s_w)$ and k, being $\gamma = w_0/\alpha$. The rainfall data and PET datasets allow a direct estimate of $\alpha, \langle N \rangle$ and PET, while the parameter that need to be calibrated are Z_r and k, assuming $s_1 = 0.5$, $s_w = 0.2$, $n = 0.35$.

2.6 WB4

This model is based on an annual water balance which is performed through a a two-stage partitioning: first, annual precipitation is decomposed into quick flow and soil wetting and, subsequently, the resulting wetting is partitioned into slow flow and vaporization. The analytical functional relationships are fitted at each stage to the measured values in order to produce parametric expressions for the components of the water balance, (namely quick flow, slow flow and vaporization). The end point is a dimensionless reinterpretation of two former models previously developed by *L'vovich (1979)* and *Ponce and Shetty (1995a, 1995b).*

The L'vovich approach is partly empirical since it is based on empirical analyzes of rainfall-runoff data for the characterization of the annual water balance. On the other hand, it goes further than Budyko by explicitly including the partitioning of annual precipitation into its major components of storage, release by quick flow and slow flow, and evapotranspiration (combining bare soil evaporation, interception loss and plant water uptake). Hence, *L'vovich* approach is suitable to characterize catchment at the annual time scale and can be defined as a *functional* approach (*Wagener et al.; 2007*). About twenty years later, *Ponce and Shetty* developed an analytical formulation of the *L'vovich* approach and provided some mathematical relations to express the main components of the water balance as a function of four physically sound parameters. These parameters need to be calibrated on measured data as will be detailed afterwards.

The rainfall partitioning into quick flow, slow flow and vaporization used by L'vocich follows the schematization shown in Figure 2.1, assuming that the recharge is negligible ($R = 0$). First, precipitation P is partitioned into a quick flow component (S) and an infiltration (termed catchment wetting, W). Then, the result-

ing wetting is further partitioned into a slow.flow component (U), and an energy-dependent vaporization component (evaporation plus transpiration, ET). Both the quick-flow and slow-flow components need to be combined to yield the total discharge in the stream ($Q = U + S$). Mathematically the two-stage hydrologic partitioning described above can be written as:

$$P = S + W \tag{2.22}$$

$$W = U + ET \tag{2.23}$$

The combined annual water balance, neglecting carryover of storage between consecutive years, can then be written as:

$$\langle P \rangle = \langle ET \rangle + \langle Q \rangle \tag{2.24}$$

$$\langle Q \rangle = \langle S \rangle + \langle U \rangle \tag{2.25}$$

L'vovich implemented this theory in a large number of catchments in many ecoregions of the world. This was done by assembling continuous data on precipitation and streamflow and then applying a base flow separation procedure to partition total streamflow into a slow-flow (e.g., base flow) component and a quick-flow (e.g., surface flow) component. In particular, the *Lyne and Hollick* (1979) algorithm, (see Equation (2.3)), was adopted in order to single out the slow component of streamflow. By aggregating all these quantities to the annual scale, L'vovich was then able to estimate $\langle P \rangle, \langle Q \rangle, \langle S \rangle$ and $\langle PET \rangle$ for every year of record, then, by difference, $\langle W \rangle$ and $\langle ET \rangle$. Based on this partitioning he was able to estimate the empirical relationships between $\langle S \rangle$ and $\langle P \rangle$, $\langle W \rangle$ and $\langle P \rangle$, $\langle U \rangle$ and $\langle W \rangle$, as well as the dependence of $\langle ET \rangle$ on $\langle W \rangle$. He finally presented the results in the form of normographs and tables and highlighted regional differences in the relationships between the different ecoregions of the world.

Despite some scattering, the empirical relationships obtained by L'vovich exhibit some common, universal patterns. A general trend of the empirical relations obtained by L'vovich are schematically shown in Figure 2.6.

As per the partitioning of precipitation into surface runoff and wetting, there seems to be a threshold value of annual precipitation that must be satisfied before

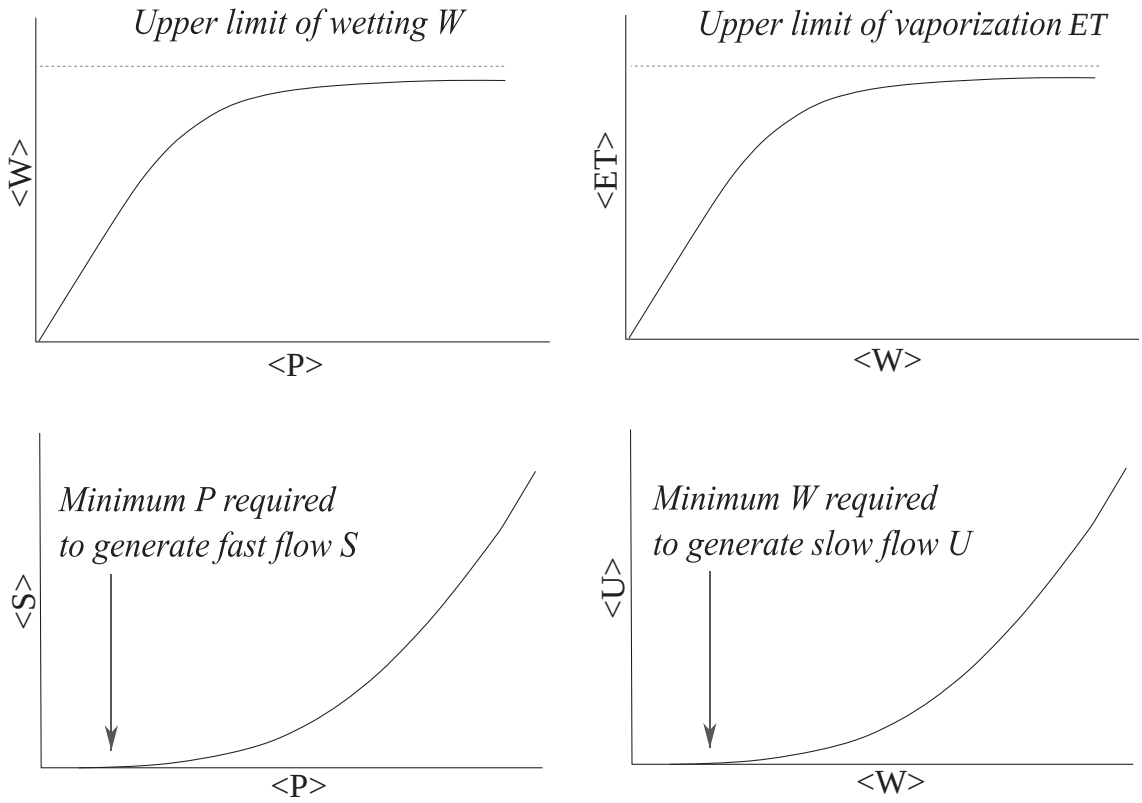

Figure 2.6: *Curves illustration typical trends of preipitation partitioning*

any quick flow can be observed; all precipitation up to this threshold becomes catchment wetting (of canopy, surface and soil). On the other hand, there appears to be an upper limit to wetting: with increasing precipitation the wetting approaches this upper limit, and the quick flow increases accordingly. In the asymptotic limit, quick flow grows at the same rate as precipitation. Moreover, the empirical analyses of *L'vovich* highlighted a similar pattern in the splitting of W into ET and U in the second stage partitioning of the soil wetting into vaporization and slow flow, as in Figure 2.6. Once again, there appear to be a threshold value of wetting that must be satisfied before there is any slow flow; all wetting up to this limit goes into vaporization. There seems to be an upper limit to the vaporization as well: with increasing wetting, vaporization approaches this upper limit, and the slow flow increases accordingly. At this limit slow flow grows at the same rate as wetting. This partitioning suggests that with the increase of annual precipitation, control is transferred from storage (wetting) to quick flow, and with increase of wetting, control is transferred from interception and plant water use (vaporization) to subsurface

drainage (slow flow).

In 1995 *Ponce and Shetty* gave to the *L'vocich* empirical approach a sounder analytical background. Inspired by the similarity of the S versus P and U versus W relationships to the (event-scale) Q versus P relationship in the widely used 'Soil Conservation Service Curve Number' runoff generation model (*Soil Conservation Service, 1985*), the two authors suggested a couple of analytical relationships to describe the water balance partitioning. The partitioning of precipitation into wetting and quick flow is described as:

$$if \quad \langle P \rangle < \lambda_s W_p \quad then \quad \langle S \rangle = 0, \quad \langle W \rangle = \langle P \rangle ; \tag{2.26}$$

$$if \;\; \langle P \rangle > \lambda_s W_p \;\; then \;\; \langle S \rangle = \frac{(P - \lambda_s W_p)^2}{\langle P \rangle + (1 - 2\lambda_s) W_p}, \;\; \langle W \rangle = \langle P \rangle - \frac{(\langle P \rangle - \lambda_s W_p)^2}{\langle P \rangle + (1 - 2\lambda_s) W_p} \tag{2.27}$$

Hence:

$$\lim_{\langle P \rangle \to +\infty} \langle S \rangle = \langle P \rangle - W_p \quad and \quad \lim_{\langle P \rangle \to +\infty} \langle W \rangle = W_p \, . \tag{2.28}$$

The partitioning of W into $\langle ET \rangle$ and $\langle U \rangle$ is instead described by:

$$if \quad \langle W \rangle < \lambda_u PET \quad then \quad \langle U \rangle = 0, \quad \langle ET \rangle = \langle W \rangle \tag{2.29}$$

$$if \;\; \langle W \rangle > \lambda_u PET \;\; then \;\; \langle U \rangle = \frac{(\langle W \rangle - \lambda_u PET)^2}{\langle W \rangle + (1 - 2\lambda_u) PET}, \;\; \langle ET \rangle = \langle W \rangle - \frac{(\langle W \rangle - \lambda_u PET)^2}{\langle W \rangle + (1 - 2\lambda_u) PET} \tag{2.30}$$

Hence:

$$\lim_{W \to +\infty} \langle U \rangle = \langle W \rangle - PET, \lim_{W \to +\infty} \langle ET \rangle = PET \, . \tag{2.31}$$

The functional forms suggested in the above equations can be shown to capture the general trends postulated by *L'vovich* even if, of course, these are merely mathematical constructs guided by the empirical data analysis.

In the previous equations the parameters W_p and PET are the upper bounds of $\langle W \rangle$ and $\langle ET \rangle$, which thus represent the potential wetting and the potential evapotranspiration of a catchment, respectively. The threshold values of P and W that

must be exceeded before any flow (quick flow and slow flow) can occur are defined as $\lambda_s W_p$ and $\lambda_u PET$, respectively, where λ_s and λ_u are empirical dimensionless coefficients satisfying that $\lambda_s > 0$ and $\lambda_u < 1$.

At this stage the model needs to be calibrated in order to obtain the values of the four parameters $W_p, PET, \lambda_s, \lambda_s$ and λ_u. W_p and λ_s can be calibrated to minimize the mean square error (MSE) of the difference between the S extracted from runoff hydrographs through the filter provided by Equation (2.3) and the one calculated with the Ponce and Shetty model. This method is repeated for U in the second partitioning by using the values of W obtained as $\langle W \rangle_{calc} = \langle P \rangle_{obs} - \langle S \rangle_{calc}$ in order to calibrate PET and λ_u. It is worth mentioning how this model doesn't need explicitly as input any estimation of potential evapotranspiration: the fraction of the precipitation which is vaporized (i.e. the actual evapotranspiration ET), is obtained from calibration according to the above procedure. This is a major difference of this model with respect to the other models here presented that need as input the PET. The above formulation can be also transposed in nondimensional form. This is motivated by the expectation that the dimensionless formulation will lead to a more compact formulation of the water balance problem, which may reveal general functional relationships common to the behavior of all the catchments in space and in time. Rearranging the terms in equations from (2.26) to (2.31), the following dimensionless expressions of the four water balance fluxes can be obtained (*Sivapalan et al., 2011*):

$$\langle S \rangle^* = \frac{\widetilde{\langle P \rangle}}{1 + \widetilde{\langle P \rangle}} \qquad \langle W \rangle^* = \frac{1}{1 + \widetilde{\langle P \rangle}} \tag{2.32}$$

$$\langle U \rangle^* = \frac{\widetilde{W}}{1 + \widetilde{\langle W \rangle}} \qquad \langle ET \rangle^* = \frac{1}{1 + \widetilde{W}} \tag{2.33}$$

being

$$\widetilde{\langle P \rangle} = \frac{\langle P \rangle - \lambda_s W_p}{(1 - \lambda_s) W_p} \qquad \widetilde{\langle W \rangle} = \frac{\langle PET \rangle - \lambda_u PET}{(1 - \lambda_u) PET} \qquad \widetilde{\langle ET \rangle} = \frac{PET - \lambda_u PET}{(1 - \lambda_s) W_p} \tag{2.34}$$

Where $\widetilde{\langle P \rangle}$ is a rescaled annual precipitation and $\widetilde{\langle W \rangle}$ is a rescaled annual soil wetting. $\widetilde{\langle ET \rangle}$ is a rescaled vaporization limit which may be deemed equivalent

to the concept of normalized potential evapotranspiration. If a new dimensionless coefficient, K, is introduced as:

$$K = \frac{\lambda_s W_p - \lambda_u PET}{(1-\lambda_s)W_p} \tag{2.35}$$

which is a function of all the four Ponce-Shetty parameters, $\widetilde{\langle W \rangle}$ can be rewritten as:

$$\widetilde{\langle W \rangle} = \frac{K + \widetilde{\langle P \rangle} + K\widetilde{\langle P \rangle}}{\widetilde{\langle ET \rangle} + \widetilde{\langle P \rangle}\widetilde{\langle ET \rangle}} \tag{2.36}$$

Even though the nondimensional formulations were derived for the case $\langle P \rangle > \lambda_s W_p$ and $W_p > \lambda_u PET$ only, they are nevertheless valid for $\langle P \rangle < \lambda_s W_p$ and $W_p < \lambda_u PET$ as well, in the sense that in the case $\langle P \rangle <_s W_p, \langle S \rangle^* = 0$ and $\langle W \rangle^* = 1$ and similarly in the case $\langle W \rangle < \lambda_u PET$, $U^* = 0$ and $\langle ET \rangle^* = 1$. Taken together, these formulations satisfy the following conditions

$$\langle S \rangle^* + \langle W \rangle^* = 1 \qquad \langle U \rangle^* + \langle ET \rangle^* = 1 \tag{2.37}$$

which reflects the notion of a competition between S and W in the first-stage partitioning and between U and ET in the second-stage partitioning.

Finally it's possible to write the runoff coefficient as a function of the dimensionless terms introduced in equations (2.34) and the coefficient K as:

$$\frac{\langle Q \rangle}{\langle P \rangle} = \frac{K^2(1+\widetilde{\langle P \rangle}) + K\widetilde{\langle P \rangle}(2+\widetilde{\langle P \rangle}) + \widetilde{\langle P \rangle}^2(1+\widetilde{\langle ET \rangle})}{(K+\widetilde{\langle P \rangle})(K+\widetilde{\langle P \rangle} + K\widetilde{\langle P \rangle} + \widetilde{V}\widetilde{\langle P \rangle})} \tag{2.38}$$

The former equation can be significantly simplified assuming $K = 0$. Provided that K is in most cases comprised within the range (-0.04,0.04), this assumption seems o be justified for practical applications. It this case the runoff-ratio equation can be expressed as:

$$\frac{\langle Q \rangle}{\langle P \rangle} = \frac{\widetilde{\langle P \rangle}(1+\widetilde{\langle ET \rangle})}{\widetilde{\langle P \rangle} + \widetilde{\langle ET \rangle} + \widetilde{\langle P \rangle}\widetilde{\langle ET \rangle}} = \frac{1+\widetilde{\langle P \rangle}\varphi}{1+\varphi+\widetilde{\langle P \rangle}\varphi} \tag{2.39}$$

where $\widetilde{\langle P \rangle}$ is the scaled annual precipitation and $\varphi = \left\langle \widetilde{ET} \right\rangle / \widetilde{\langle P \rangle}$ is the ratio of vaporization potential to precipitation, and in this sense it is a counterpart of the classical dryness index ($D_I = \langle PET \rangle / \langle P \rangle$). These two indexes, though, are in turn a function of the 4 parameters previously identified ($PET, W_p\lambda_s, \lambda_u$).

MODEL PARAMETERS: The model is open to be be handled in different ways. In the original version the authors calibrate all the aforementioned parameters ($\lambda_s, \lambda_u, W_p, PET$) for every single basin in a pool of catchments located throughout the US. In the framework of this thesis, such approach would lead to an undesired increment in the number of parameters ($M = 4C$ where C=39 is the number of catchments), making the model useless for the prediction in catchments where discharge measurements are lacking. To circumvent this issue, the 4 parameters could be assumed to be spatially uniform across the study catchments. This procedure would lead to a model having just 4 parameters. However, all the catchments would be characterized by the same value of potential evapotranspiration PET, thereby neglecting the natural spatial variability of climate and, thus, making the approach less robust. In order to preserve the spatial variability of evapotranspiration it is possible to include in the model the available esitimate of potential evapotranspiration provided by the MODIS and CGIR datasets. However, to improve model performances, the dataset's values need to be multiplied by a calibrated correction factor ξ. The operation does not impact the number of parameters, provided that the calibrated (partially uniform) PET would be replaced by the correction factor ξ. The latter is the version of the model implemented when the filtering of precipitation is accounted for (4 parameters: λ_s, $lambda_u$, W_p, ξ).

To make the approach applicable for the prediction to ungauged sites, a second issue should be handled. In fact, the partitioning of rainfall based on the filter presented in section 2.1.1 would make the model unusable where streamflows are not measured. Introducing the extra assumption that all precipitation is turned into soil wetting, without the occurrence of any surface runoff, seems to be a reasonable way to circumvent this problem. In the latter case, when the MODI and CGIR datasets are used to account for the observed spatial variability of evapotranspiration in the study catchments, the number of parameters to be calibrated reduces to just one (λ_u), being $< W >=< P >$ and, hence, $\lambda_s = 1$. This is the approach adopted in cases when the decomposition of precipitation into surface runoff and infiltration is

disregarded.

2.7 WB5

WB5 is a spatially distributed global-scale soil water balance. It was implemented on the high-resolution $WorldClim$ and $CGIAR - CSI - PET$ climate dataset (`http://www.worldclim.org/`) using ArcAML (ESRI) as modeling tool. The original cells size is 30 arc-second (about 1 Km at equator). However, in this study, the spatial resolution has been reduced to cells of about 15 Km in order to reduce the sensitivity towards potential errors in the original dataset and to make the analysis computationally easier. The model uses spatially distributed values (temporally average over the period 1950-2000) of monthly average precipitation ($\langle P \rangle$) and monthly potential evapotranspiration ($\langle PET \rangle$) and provides estimates of actual evapotranspiration (ET) and soil water content ($w = nsZ_r$) at monthly time scale. The model is spatially explicit to represent heterogeneity of climate conditions, while vegetation and soil properties are assumed as uniform (characterized by crop coefficient equal to 1, rain interception coefficient equal to 0.15 and maximum soil water content $w_0 = ns_1Z_r$ in the rooting zone equal to 350 mm). Rain interception, which is the process by which precipitation is intercepted by vegetation canopy and litter and then vaporized, is assumed to be an important process in the overall water balance since it reduces the amount of precipitation available to plants. Vegetation interception is a mechanical function of the storage space of vegetation structure and, hence, strongly dependent on the LAI. The authors simply assumed the effective precipitation $\langle P \rangle_e$ as a fraction of the gross precipitation $\langle P \rangle$. The monthly quantity of rain intercepted is proportional overall rainfall as:

$$\langle P \rangle_e = 0.85 \langle P \rangle \tag{2.40}$$

The dependance of actual evapotranspiration on available atmospheric energy (PET), vegetation characteristics, quantity of water available in the soil and soil hydrological properties is modeled through the following equation:

$$\langle ET \rangle = \langle PET \rangle K_{veg} K_{soil} \tag{2.41}$$

where

- K_{veg}: is a vegetation coefficient dependent on vegetation characteristics and stage of growth. Values of K_{veg} used in this model are spatially standardized and assumed to be equal to that of the reference crop ($K_{veg} = 1$) ;

- K_{soil}: Is a soil stress coefficient representing a reduction factor resulting from the limit imposed by the monthly soil water content ($0 < K_{soil} < 1$). The model uses a linear soil moisture stress function evaluated at monthly timescale. The maximum amount of soil water available for ET processes within the plant rooting zone ($w_0 = 350$ mm) is equal to soil water content at field capacity minus soil water content at wilting point times the rooting depth Z_r.

$$w_0 = Z_r n(s_1 - s_w) \tag{2.42}$$

$$K_{soil} = \frac{s}{s_1} \tag{2.43}$$

The dataset used to determine monthly PET globally is the $CGIAR-CSI$ one (see page 8).

The model provides a global raster of temporally averaged monthly actual evapotranspiration values ($\langle ET \rangle$) and soil water content (SWC). Based on a spatially distributed map of monthly precipitation, available at the $WorldClim$ web site (`http://www.worldclim.org/current`), the seasonal runoff coefficients can be calculated as:

$$\frac{\langle Q \rangle}{\langle P \rangle} = 1 - \frac{(\Delta s n Z_r + \langle ET \rangle)}{\langle P \rangle} \tag{2.44}$$

where $\Delta s n Z_r$ is the difference in soil water content between two subsequent months/periods.

MODEL PARAMETERS: WB5 has not been implemented directly in this thesis work since its results are already available on line, and no parameters need to be calibrated. The data provided by the authors include global rasters of actual evapotranspiration at annual and monthly time scale as well as global rasters of monthly soil water stress (defined as the monthly fraction of soil water content available for evapotranspiration processes as a percentage of w_0) . In this thesis work the former

rasters were processed in GIS environment and 12 global raster of mean monthly precipitation were added in order to estimate the average seasonal and annual water balance and runoff coefficients.

2.8 Decomposition of annual water balance

In this thesis, a novel approach has been developed in order to describe the inter-seasonal variability of the water balance. Many of the models presented in section 2.3 to 2.7 are based on hypotheses that hold only at annual time-scale, or they have been implemented only at the annual level. Because of this reason they are best applicable to estimate annual runoff coefficients. To get an estimate of the inter-seasonal variability of streamflow regimes during the year, the knowledge of seasonal average runoff coefficients would be instead desirable. The decomposition in season is a convention and can be established according to the trends of climatic variables during the year, which is a function of the geographical position and, in particular, of the latitude. Mid-latitudes areas show marked seasonality with four different season characterized by specific climatic and meteorological features. Seasonality reduces in tropical areas, which are typically featured by a wet and a dry season. Because of the wide range of locations included in the study area, typical subdivision into four seasons has been adopted, which broadly follows calendar dates, as detailed below:

- **SEASON 1 (SPRING)**: March, April, May;
- **SEASON 2 (SUMMER)**: June, July, August;
- **SEASON 3 (AUTUMN)**: September, October, November;
- **SEASON 4 (WINTER)**: December, January, February;

The general relationship between annual and seasonal runoff coefficients will be discussed in the following. The (average) annual runoff coefficient is defined as:

$$\phi_a = \frac{\langle Q \rangle_a}{\langle P \rangle_a} \tag{2.45}$$

where $\langle Q \rangle_a$ and $\langle P \rangle_a$ are the average streamflow and rainfall during one entire year.

The following relation between the annual average and the seasonal averages holds:

$$\phi_a = \frac{\langle Q \rangle_a}{\langle P \rangle_a} = \frac{\sum_{i=1}^{4} \frac{\langle Q \rangle_i}{\langle P \rangle_i} \langle P \rangle_i}{4 \ \langle P \rangle_i} = \frac{\sum_{i=1}^{4} \phi_i \ \langle P \rangle_i}{4 \ \langle P \rangle_i} \tag{2.46}$$

where $\phi_i =< Q >_i \ / \ < P >_i$ is the runoff coefficient during the season i. Equation (2.8) states that the average annual runoff coefficient is the weighted mean of the seasonal average runoff coefficients. The approach consists of the estimate of the annual runoff coefficient and the subsequent calculation of the seasonal runoff coefficient of a catchment by multiplying the annual average runoff coefficient by a 'Seasonal Multiplication Coefficient' $\psi_i = \phi_i / \phi_a$:

$$\phi_i = \frac{\langle Q \rangle_i}{\langle P \rangle_i} = \frac{\langle Q \rangle_a}{\langle P \rangle_a} \psi_i. \tag{2.47}$$

The values of the seasonal coefficients ψ_i are obtained through calibration based on the 39 catchments analyzed in the study. The values of ψ_i across the study catchments are presented in the box-plots of Figure 2.7.

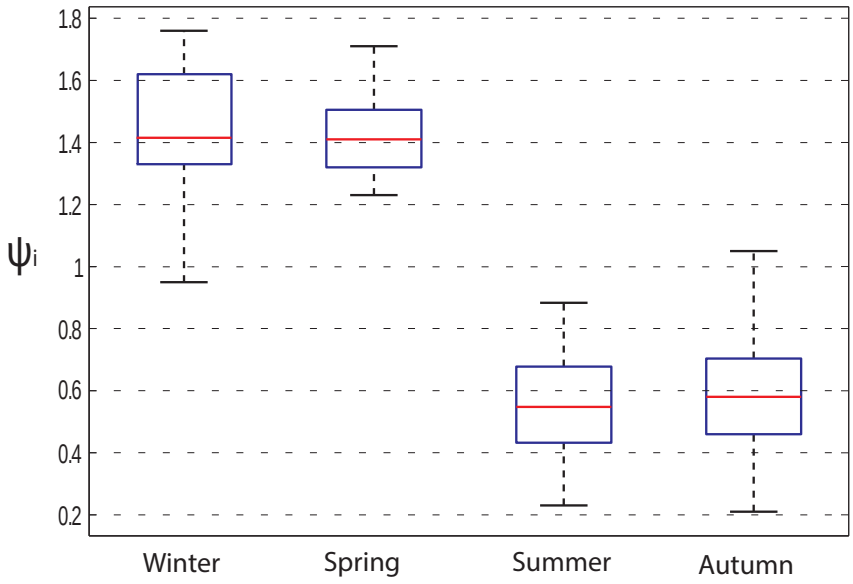

Figure 2.7: *Seasonal runoff coefficients*

The average values of the seasonal coefficients are:

$\psi_{winter} = 1.44$
$\psi_{spring} = 1.42$
$\psi_{summer} = 0.56$
$\psi_{autumn} = 0.59$

As expected, a strong seasonality in the normalized seasonal runoff coefficient is observed. Moreover, given the broad range of hydroclimatic conditions featuring the 39 study catchments, the inter-catchment variability is quite reduced, suggesting the existence of universal patterns across relatively wide areas. Figure 2.7 shows that the variability is higher in winter and autumn. These results suggest that the regionalization procedure of the normalized runoff coefficients described above could be applicable to ungauged catchments belonging to the study area, assuming that the annual trend of the normalized runoff coefficients follows the general pattern exhibited by the set of catchments used for calibration.

2.9 Ranking of the models

The performances of each model have been objectively quantified by means of the *Akaike's information criterion (Akaike, 1973)*. The method gives a rigorous way for model selection based on the maximization of the log-likelihood function between experimental data and model estimates. The goodness of fit of each model is discounted by accounting for the number of parameters that are fitted to observations. The *Alaike's information criterion* (AIC) quantifies the performance of a model through a real number: the lower is the value, the better the model performs. Worth to be mentioned that performances index given by the method doesn't have a stand-alone meaning: it's just a way to rank a set of different models and, hence, just differences of AIC between models are meaningful.

Given a general model that simulates an output variable y as a function of a set of predictors (input) I and a set of M model parameters, ξ_M, the following equation holds:

$$\overline{y} = \overline{y}(I, \xi_M) + \epsilon \tag{2.48}$$

where $\overline{y}$ is the observed response variable, and ϵ is the model error (i.e. the residual). Residuals can be used to evaluate the performance of any set of moldel

parameters. Assuming normally distributed residuals with standard deviation σ, the likelihood of the model parameters is

$$L(\epsilon|\xi_M, \sigma) = \frac{exp\left(\frac{-\epsilon_i^2}{2\sigma^2}\right)}{\sqrt{2\pi\sigma^2}}. \tag{2.49}$$

Equation (2.49) gives a consistency test for the parameter estimates: if the observed residual is an unlikely estimate of ϵ, ξ_M is an unlikely estimate of the true parameters. Parameter estimates that maximize the likelihood are the most likely values for the parameters. Under the assumption that a set of n observations $\overline{y}_i$ are available, and assuming independent residuals, the total likelihood of all the residuals is simply the product of the likelihoods of the individual residuals

$$L(\epsilon) = \prod_{i=1}^{n} L(\epsilon_i|\xi_M, \sigma). \tag{2.50}$$

Accordingly, the log-likelihood can be written as:

$$log\left[L(\epsilon)\right] = \sum_{i=1}^{n} log\left[L(\epsilon_i\,|\xi_M, \sigma)\right]\ , \tag{2.51}$$

where the residuals are the differences between observations and predictions ($\epsilon_i = \overline{y}_i - y_i$). When σ^2 is independent on the other parameters, then the estimates of ξ_M obtained by minimizing Equation (2.51) is the same as those obtained by minimizing the mean square error (MSE)

$$MSE = \frac{\sum_{i=1}^{n} \epsilon_i^2}{n}. \tag{2.52}$$

The general definition of AIC can be written as:

$$AIC := -2log(L) + 2(M+1) \tag{2.53}$$

where $log(L)$ is the log-likelihood function of the residuals and M is the number of parameters of the model under consideration. AIC can be used to compare model

performances if they are not nested. Under the assumption that the variance of the residuals are estimated based on the observations available using the maximum likelihood, the AIC can be computed as:

$$AIC = 2\,n\,MSE + 2(p+1), \tag{2.54}$$

that is the formula used to rank the different water balance compared in this study.

2.10 Streaflow PDFs and flow regimes

Recent studies (*Botter et al. 2007, 2009*) have provided a simple theoretical framework to analytically derive the Probability Density Functions (PDF) of streamflows in an arbitrary section of a river network based on a simple schematization of the main processes involved in the transformation of rainfall into runoff. As detailed hereafter, a limited set of parameters tightly related to the physical processes involved are sufficient to catch the main processes that control transformation of rainfall into streamflows. The potential of the aforementioned method consist in its ability to provide an estimate of the streamflow PDF at any arbitrary location of a river network starting from the primitive information about precipitation and landscape. This ambitious result can be reached once a reliable framework to estimate the parameters involved in the model is identified. The importance of this thesis work lies in the application of the water balance models identified to provide estimates of the most relevant parameters involved in the analytical expression for the streamflow PDF. Two parameters of the above mentioned model, in fact, can be predicted based on the average runoff coefficient of a catchment. Worth to be pointed out how a PDF carries the same amount of information of a duration curve, which is a widespread tool to characterize river flow regimes. Once the soundness of the procedures used to estimate all the parameters involved have been proved, a powerful tool to characterize the streamflow regime at any point of a river network in an arbitrary catchment will be available. A reliable estimate of the average annual or seasonal runoff coefficient, will allow the estimate of the streamflow PDF without the need of discharge data. The ability to forecast the statistical distribution of water availability has remarkable outcomes. From an ecological point of view, the temporal variability of discharges is important for riparian vegetation and riverine ecosystems in general (*Doulatyari et al.,2014; Ridolfi et al.,2006*). Moreover, flow

variance is strictly related to flow predictability on which the mobility and colonizing ability of living species depend. The knowledge of the probability density function of discharge within a river or a stream is useful in dimensioning floods protection measures as levees, diversions, artificial reservoirs for flood control or other structures, such as bridges, which are in close contact with water bodies since they are dimensioned according to certain expected magnitudes of streamflows. The ability to estimate the streamflow distribution in an ungauged reach of a river provides a tool to make sounder hypothesis about the optimal placement of diversions for hydropower production or water supply, thereby addressing water resources management problems. Finally, from a scientific point of view, understanding through modeling river flow regimes how the climate, the weather and basin morphology control the storage and release of water within catchments, gives an additional insight on how climate change may impact the hydrological cycle in future decades.

2.10.1 Streamflow PDF estimation model

One of the main topics in hydrology is the study of how the streamflow variability in river basins is affected by the intertwined action of geomorphological, climatic, meteorological and echohydrological processes operation at catchment scales. The variability in time of streamflows mirrors the stochastic nature of all the underlying forcings and the huge complexity of causal relations, which makes deterministic frameworks unfeasible, often calls for statistical approaches. The model here presented (*Botter et al. 2007, 2009*) is able to provide the steady state probability density function (PDF) of streamflows as a function of few macroscopic rainfall, soil, vegetation and geomorphological features. The lumped approach developed considers spatially averaged parameters, whose values need to be estimated via calibration, simulation or directly from basic geomorphoclimatic data. The model couples a soil moisture dynamic model (*Porporato et al., 2004*) to a streamflow generation model which, in its basic version (*Botter et al., 2007*), postulates a linear relation between soil water storage and discharge (i.e. assuming exponential recession curves). In a more recent version of the model (*Botter et al., 2009*) this hypothesis is relaxed by allowing power laws function to describe the process in a more realistic and general fashion.

The approach moves from the long-term soil water balance in the root zone (*Porporato et al., 2004*) where the competition between deep percolation and evap-

otranspiration processes takes place. The relevant ecohydrological processes occurring therein are described through a simplified approach which considers constant parameters representative of the behavior of a given catchment during a given season: root zone depth (i.e. the depth of the active soil layer), Z_r $[L]$; porosity n and maximum evapotranspiration rate PET $[L/T]$. These values are assumed to be spatially and temporally averaged. The temporal evolution of spatially averaged relative soil moisture in the root zone, $s(t)$ (volume of water divided by volume of voids), is thus seen as the result of the following three processes:

1. stochastic instantaneous increments due to infiltration from rainfall, which is modeled at daily timescales as a Poisson process with frequency λ_p $[T^{-1}]$ and where daily rainfall depths are assumed to be exponentially distributed with parameter $1/\alpha$, being α $[L]$ the average rainfall depth on rainy days;

2. linear losses due to evapotranspiration, occurring in the range of soil moisture comprised between the wilting point s_w and a suitable soil moisture threshold s_1 comprised between field capacity and saturation;

3. instantaneous deep percolation producing effective rainfall and hence runoff (above s_1).

Even if the the model follows a lumped approach, additional studies (*Botter et al.,2007*) have suggested that it can be used to properly predict the PDFs of streamflows also in relatively large catchments ($A = O(10^3 Km^2)$).

According to the above scheme, when s exceeds the threshold s_1 due to infiltration input, the soil moisture content is assumed to instantaneously decrease to s_1 through the release of an effective rainfall pulse which is released as streamflow. Effective rainfall time series resulting from the soil moisture dynamics described above may be approximated, at the daily timescale, by a new marked Poisson process where the net rainfall depths (i.e., the fraction of the incoming water which exceeds the limit s_1) follow an exponential distribution having, as it can be proved, the same shape parameter α^{-1} of the rainfall distribution. The related average runoff frequency λ may be obtained from crossing properties of the threshold s_1 if discharge measurements are available. This thesis provided an alternative method to estimate λ, which is one of the most important parameters involved in the formulation, when no discharge data are available. The method takes advantages of the relation between λ and the average runoff coefficient of a catchment, opening

the way to the use of a broad set of water balance models, as shown in the previous sections. Provided that the rainfall series are characterized by an average frequency λ_p and average depth α while the effective rainfall is an analogous Poisson process characterized by a frequency parameter λ and equal distribution of rain depths, the following relation holds (*Porporato et al.*, 2004):

$$\frac{\langle Q \rangle}{\langle P \rangle} = \frac{\alpha\lambda}{\alpha\lambda_p} = \frac{\eta\gamma^{\lambda_P}\eta e^{-\gamma}}{\lambda_P\left(1-\Gamma(\lambda_P/\eta,\beta^{-1})\right)} \tag{2.55}$$

which exploits the key properties of marked Poisson processes and where $\eta = PET/(nZ_r(s_1 - s_w))$ is the normalized maximum evapotranspiration rate and $\gamma = (nZ_r(s_1 - s_w))/\alpha$ is the ratio between the mean rainfall depth and the soil storage capacity. λ can thus be estimated based on frequency of daily rainfall, λ_p, and the seasonal runoff coefficient $< Q > / < P >$, as provided by rainfall measurements and a given water balance model. The effective rainfall pulses are assumed to propagate throughout the catchment and eventually be released to the channel network as surface/subsurface/groundwater flow. The continuity equation for the excess water volume stored in the catchment, W, can be written as:

$$\frac{dW(t)}{dt} = -\rho(W) + \epsilon_t^{'}(\lambda;\gamma_w)\ , \tag{2.56}$$

where $\rho(W)$ represents the deterministic losses due to streamflows (in general a function of the stored water volume W) and $\epsilon_t^{'}$ is the stochastic noise incorporating the storage increments in correspondence of effective rainfall events:

$$\epsilon_t^{'}(\lambda;\gamma_w) = \sum_{i;t_i<t} \Delta W_i \delta(t - t_i)\ , \tag{2.57}$$

δ being the Dirac delta function. According to the soil moisture model discussed, the arrival times of effective rainfall events, t_i, are distributed according to a Poisson process with rate λ, and the storage increments in correspondence of such runoff events, ΔW_i, are assumed to be instantaneous (at the daily timescale) and follow an exponential distribution with parameter $\gamma_w = 1/(\alpha\, A)$.

In the original version of the model (*Botter et al.,2007*), residence times in the catchment are assumed to be described by an exponentially distributed random vari-

able which is equal to assume the system to behave as a linear reservoir, where Q is proportional to W ($Q = \rho(W) = kW$). The time constant of the system $1/k$ can be also interpreted as the timescale of the hydrologic response of the catchment. However, both observational evidences and theoretical argument suggest that in many cases (e.g., in small deep soil-mantled catchment in humid climates) the relation between Q and W can be nonlinear. In this case, the master equation associated with Equation (2.10.1) reads (*Gardiner, 2004*):

$$\frac{\partial p_w^{'}(W,t)}{\partial t} = \frac{\partial\left[\rho(W)p_w^{'}(W,t)\right]}{\partial W} - \lambda p_w^{'}(W,t) + \lambda \int_{W_L}^{W} p_w^{'}(W-\omega,t)\gamma_w e^{-\gamma_w \omega} d\omega + \lambda p_0(t)\gamma_w e^{-\gamma_w W} \tag{2.58}$$

where $p_0(t)$ is the atom of probability corresponding to the minimum available storage W_L (which is the value of W for which $\rho(W) \to 0$) and $p_w^{'}(W,t)$ is the continuous part of the time-dependent probability density function of W for $W > W_L$ (i.e., $p_w^{'}(W,t)dW$ measures the relative probability of observing a water volume in subsurface/groundwater environment in the interval $[W, W + dW]$). The master Equation (2.58) is a partial differential equation that expresses the probability conservation in a given state for the process under consideration. Following Rodiguez-Iturbe and *Porporato* (2004), *Rodrigue-Iturbe et al.* (1999) and *Botter et al.* (2009) the general steady state solution of Equation (2.58), which is obtained for $t \to \infty$, reads:

$$p_w(W) = c\left\{\frac{1}{\rho(W)} e^{-\gamma_W W + \lambda I(W)} + \frac{e^{-\gamma_W W_L}}{\lambda} e^{\lambda \lim_{x \to W_L} I(x)} \delta(W - W_L)\right\} \tag{2.59}$$

where $I(x) = \int \rho^{-}1$ is the primitive of the reciprocal of the loss function ρ. If we assume $Q = \rho(W)$ a one-to-one function, the relation between W and Q can be inverted (i.e., $W = \rho^{-1}(Q) = r(Q)$) and the probability distribution of the runoff Q can be readily obtained from Equation (2.59) as:

$$p_q(Q) = p_w(aQ^b)\left|\frac{dW}{dQ}\right| = c\frac{r'(Q)}{Q} e^{-\gamma_W r(Q) + \lambda I[r(Q)]} + p_0\delta(Q) \tag{2.60}$$

where $r'(q) = d[r(Q)]/dQ$ and the atom of probability in $W = W_L, p_0$, corresponds to the same atom of probability in $Q = 0$.

Analytical solutions for $p_q(Q)$ can be obtained if we assume that the dependance between storage and discharge is described as a simple power low such as:

$$Q = \rho(W) = K(W - W_0)^h \tag{2.61}$$

where $W_0 = W_L \geq 0$ and $K > 0$. Inverting eq. (2.61) leads to:

$$W = r(Q) = W_0 + aQ^b \tag{2.62}$$

with a and b positive constant satisfying

$$b = h^{-1} \tag{2.63}$$

$$a = K^{-1/h} \tag{2.64}$$

An analytical solution is also available in the case of hyperbolic dependence of Q on W (*Botter et al.*, 2009). The parameter h in the power low function between Q and W characterizes the shape of the curve, being it concave when $h < 1$ and convex when $n > 1$.

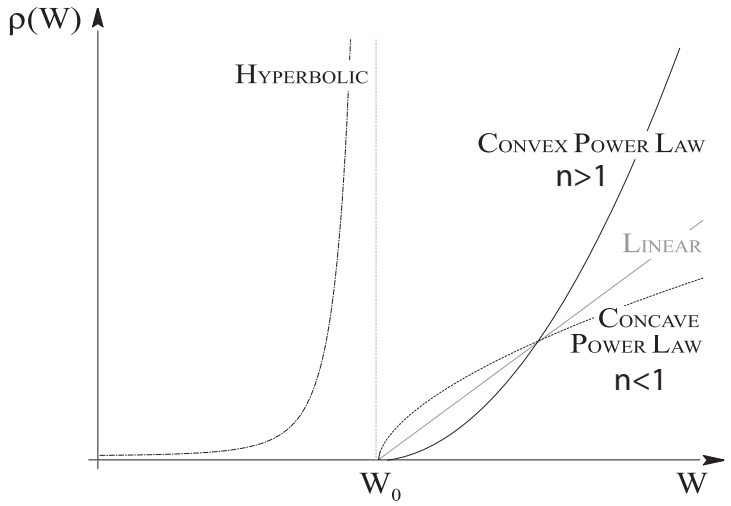

Figure 2.8: *Possible storage-runoff functions*

Non linear power-law models for base flow recessions have been widely employed in the literature because of their simplicity and their ability to fit observations under

a variety of hydro-geological environments. All the three nonlinear relationships between W and Q mentioned here are compatible with the frequently observed power law type decay of Q through time between runoff events:

$$\frac{dQ}{dt} \propto Q^{\nu} \tag{2.65}$$

Note that exponent ν changes depending on h, being $\nu = 2 - 1/h$. At the moment, an additional model capable to estimate the recession parameters starting only from the topography of a catchment (DTM) is available. The tool is currently under testing but encouraging results have already been reached. Such results will be discussed later on in this thesis.

In conclusion, the analytical solution of Equation (2.60) in the two cases of power law dependance reads:

$$p_q(Q) = C_1 \left[\frac{1}{Q^{2-b}} \, exp\left(-\gamma_w a Q^b + \frac{\lambda a b}{b-1} Q^{b-1} \right) + \frac{1}{\lambda a b} \delta(Q) \right] \qquad n < 1 \tag{2.66}$$

$$p_q(Q) = \frac{C_2}{Q^{2-b}} \, exp\left(-\gamma_w a Q^b - \frac{\lambda a b}{1-b} \frac{1}{Q^{1-b}} \right) \qquad n > 1 \tag{2.67}$$

The Cumulative Density Function (CDF), which is another wide spread way to depict streamflow regimes, is related to the PDF as it follows in Equation (2.68):

$$CDF(Q) = \int_{Q}^{\infty} p(Q') dQ' \tag{2.68}$$

that is the probability to observe a streamflow magnitude larger that Q.

Figure 2.9 shows, as an explanatory example, the results of a simulation of the processes described so far in case of linear storage-discharge relationship (n=1).

This simulation is performed by means of the linear relation between runoff and storage and assumes typical values of the meteorologic and climatic forcings ($\lambda_p = 0.3\, d^{-1}$, $\alpha = 0.42\, cm$, $PET = 0.35\, cm/d$) as well as typical soil, vegetation and transport parameters ($n = 0.55$, $Z_r = 30\, cm$, $s_w = 0.18$, $s_1 = 0.6$; $k = 0.6\, d^{-1}$). Panel (a) shows how rainfall is modeled as a marked Poisson process. The resulting temporal distribution of rainfall events at daily time is highly heterogeneous,

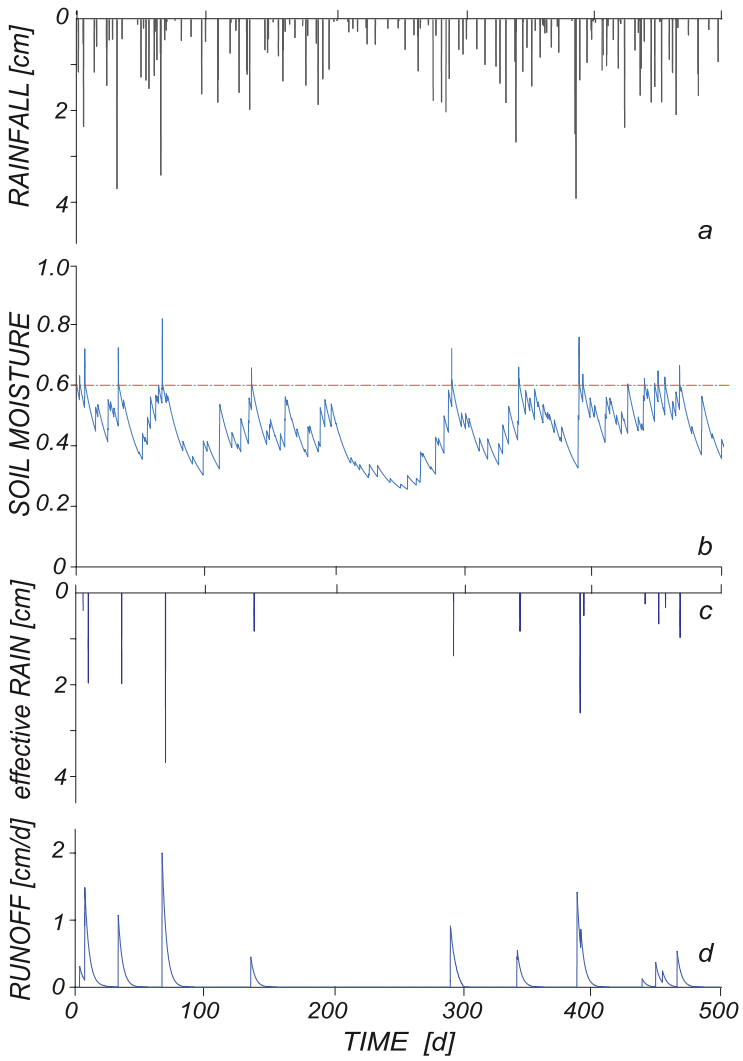

Figure 2.9: *Streamflow generation simulation*

as commented by the frequency parameter λ_p. The marks are the rainfall depths, whose distribution is assumed to be exponential with parameter $1/\alpha$. Daily rainfall pulses drive the soil moisture dynamics (b). When the soil moisture level in the upper soil layer exceeds the critical value s_1, the surplus is assumed to be immediately released as subsurface/groundwater flow. For soil moisture content (s) in the range between s_1 and s_w the moisture depletion rate via evapotranspiration is assumed to be proportional to the product between the relative soil moisture and the potential evapotranspiration, leading to an exponential decrease in time of w. Effective rainfall pulses, which generates the runoff are, again, characterized by a marked Poisson process of frequency parameter λ and exponentially distributed in-

tensities of parameter $1/\alpha$ (same as the rainfall). Finally, time series (d), shows how the water stored as W is released at a rate proportional to the water stored itself, leading to an exponential time decay of each effective rainfall pulse. The knowledge of the hydrograph (which comes form the convolution in time of the catchment's response to the single pulse), allows to statistically characterize the streamflow regime by means of its probability density function.

Chapter 3

Results

This chapter details the results achieved by applying the water valance models presented in the former sections to the 39 study catchments described in Table 1.1. All the models, (except than WB5), have been implemented and calibrated in a *MATLAB R*2010*a* environment.

The following notation has been used in order to uniquely identify each model and the set of possible variants adopted. Each model is labeled by a string which is composed by three main parts, separated by dots.

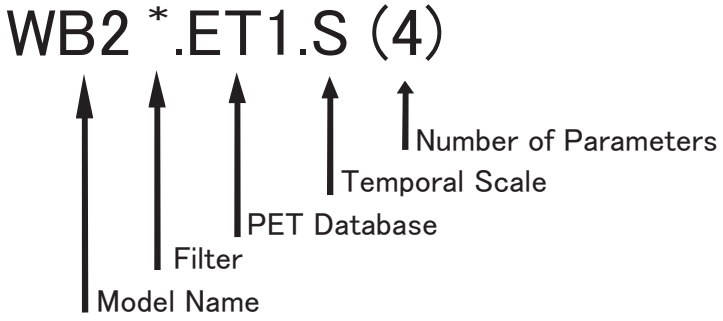

Figure 3.1: *Nomenclature*

The first three characters of the string refer to the specific water balance model. A star is then used to indicate whether precipitation is partitioned into surface runoff and wetting by means of a filtering procedure prior to the water balance model application. The lack of the star implies that all the incoming precipitation is assumed to infiltrate and, therefore, takes part to soil moisture dynamics. After

the first dot, there is a code referring to the potential evapotranspiration dataset used in the model calibration: ET1 refers to the $CGIR$ while ET2 refers to the $MODIS$ (see subsection 1.2). The character right after the second dot specifies if the model is applied at seasonal (S) or annual (A) timescales. The flag Sc denotes instead that the seasonal runoff coefficients have been estimated by applying the Seasonal Coefficients ψ_i (Subsection 2.8) to an average annual runoff coefficient obtained through an annual water balance model. Finally, the number in brackets specifies, if necessary, the numbers of model parameters that are calibrated based on the available ET datasets and observed rainfall and discharge time series in the study catchments. Many of the models under consideration include the average rooting dept Z_r as a key parameter. Z_r rules the maximum soil moisture storage capacity $w_0 = nZ_r(s_1 - s_w)$. Hence, for the sake of convenience and without any loss of generality, s_w, s_1 and n are assumed to be constant throughout the all simulations (and equal to 0.2, 0.5 and 0.35, respectively). The departures of modeled results from measurements is quantified by means of the Mean Square Error (MSE), defined by Equation (2.52).

3.1 Annual Water Balance

The results presented in this section refer to the application of the five water balance models at annual time scale. The tables shown in Figure 3.2 summarize the parameters involved (with and without the use of a filter to identify surface runoff), and the calibration values that optimize model performances.

The plots in Figures 3.3 and 3.4 show the performances of the calibrated models when the use of the filter is avoided. On the y-axis the modeled value of the runoff coefficient is shown, while the observed value, calculated as the ratio between the average annual precipitation and runoff ($\phi = \langle Q \rangle \,/\, \langle P \rangle$), is shown on the x-axis.

The table shown in Figure 3.2 shows that all the models perform better if associated with the ET1 potential evapotranspiration dataset. The difference related to the use of different potential evapotranspiration datasets is especially striking for WB1, where the MSE almost doubles if the ET2 dataset is used. As per the application of the filter, model performances improve when rainfall is partitioned into surface runoff and wetting, with the exception of WB4. The result is surprising because WB4 is the only model which was originally conceived to explicitly take into account for surface runoff. Because of this reason a deeper insight on the physical

Unfiltered rainfall

Model	ET	Parameters	MSE
WB1	ET1	-	0.0112
	ET2	-	0.0214
WB2	ET1	Z_r=420 mm	0.0121
	ET2	Z_r=300 mm	0.0157
WB3	ET1	Z_r=900 mm	0.0161
		k=0.525	
	ET2	Z_r=700 mm	0.0203
		k=0.525	
WB4	ET1	λ_u=0.2	0.0079
	ET2	λ_u=0.2	0.0097
WB5	ET1	-	0.0271

Filtered rainfall

Model	ET	Parameters	MSE
WB1*	ET1	-	0.007
	ET2	-	0.0092
WB2*	ET1	Z_r=735 mm	0.0066
	ET2	Z_r=480 mm	0.009
WB3*	ET1	$Z_{r=}$=1200 mm	0.0124
		k=0.0124	
	ET2	Z_r=950 mm	0.0147
		k=0.0147	
WB4*	ET1	λ_u=0.8	0.0109
		λ_s=0.015	
		W_p=30000 mm	
		ζ=5.5	
	ET2	λ_u=0.8	0.0113
		λ_s=0.015	
		W_p=27000 mm	
		ζ=4.9	

Figure 3.2: *Parameters of the annual water balance models*

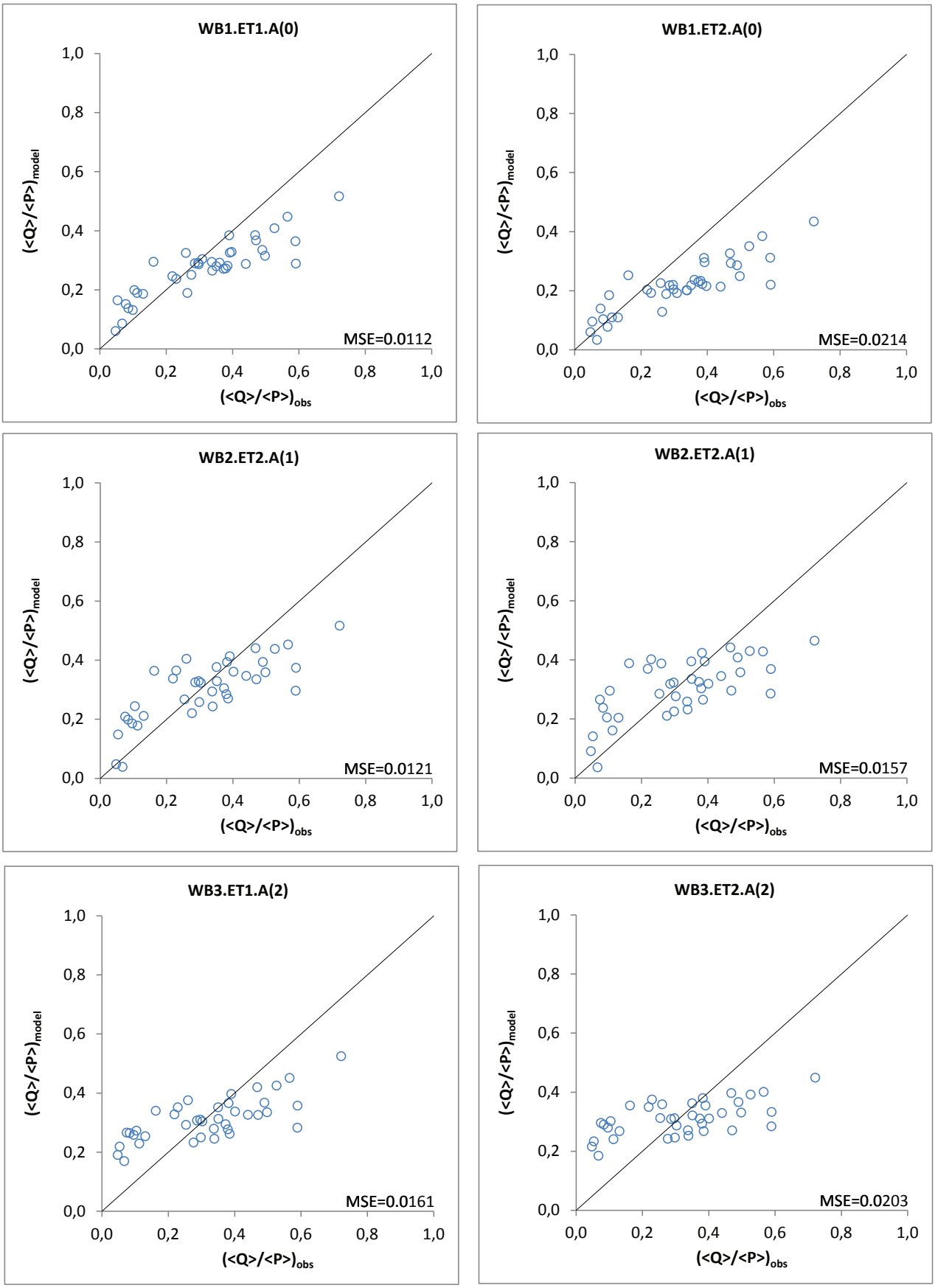

Figure 3.3: *Annual runoff ratio (without filter) (1/2)*

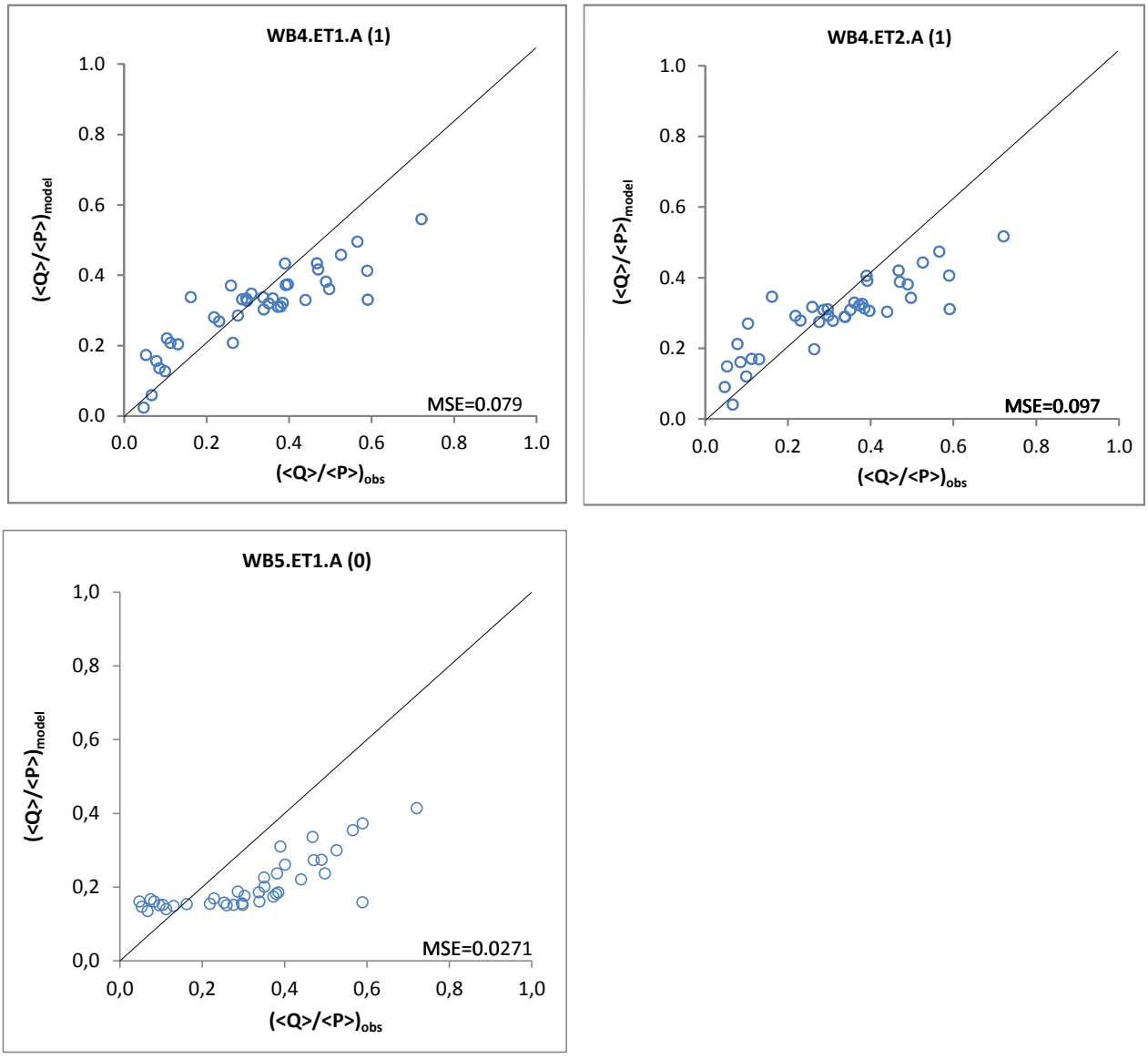

Figure 3.4: *Annual runoff ratio (without filter) (2/2)*

reasons leading to this counter-intuitive outcome would be worth.

If a rainfall filtering is performed, ET1 still provides better results, but the improvement is less visible. In fact, the presence of surface runoff reduces the amount of water processed by soil moisture dynamics, which is impacted by PET.

The scatter plots corresponding to the cases where preliminary partitioning of precipitation is performed, are presented in Figure 3.5 and 3.6. WB5 is excluded in this case because the model hasn't been directly implemented in this thesis work (provided that previous results of the model are available online at the global scale) and no parameters have been calibrated. Moreover WB5 was implemented based on the CGIR potential evapotranspiration database. Thus, the MODIS dataset has not been used in this case.

The scatter plots in Figure 3.5 and 3.6 show that the filter increase the modeled runoff coefficients in the catchments featured by highest values of ϕ, where the models without any filter systematically underestimate observed ϕ. The improvement of performances is particularly evident for WB3. The dispersion around the 45 degree line of the points in WB3* is reduced with respect to WB3 for both the PET datasets. However, the model WB3 doesn't seem to catch the inter-catchment variability of the runoff coefficient properly in none of these cases. For WB4, the optimal value of the parameter λ_u is lower when the filter is not applied. The result is consistent because lower values of λ_u imply that the catchment is more prone to release water as subsurface flow. In fact, when the filtering procedure is not performed, the wetting component equals the entire amount of precipitation, and evpotranspiration is not able to buffer the strong wetting pulses as efficiently as it occurs when the application of the filter reduces infiltration.

Despite its good performances when the filter is disregarded, WB4 is not the best model if the filtering procedure is applied. Concerning the optimal parameters of WB4*, the calibration procedure leads to extremely high values of potential wetting ($W_p = 27000 - 30000\, mm$) and potential evapotranspiration ($\xi = 4.9 - 5.5$). These values are deemed unrealistic from a physical point of view. In particular, $\xi \approx 5$ means that the performances of WB4* are optimized when PET is about five times larger the one estimated by the MODIS and CGIR datasets.

Figure 3.4 shows how the model WB5 is unable to catch the variance of ϕ for the most arid catchments (say, $\phi < 0.4$). Moreover, the runoff coefficient is strongly underestimated for the most humid catchments. WB1 proves quite effective in predicting the average runoff coefficient at annual timescale, especially in association

with ET1 and the filtering procedure. WB2 achieves good performances, given the reduced number of calibrated parameters. In particular, when WB2 is coupled with the filter and the potential evapotranspiration dataset ET1, the mean squared error is very low and the optimal rooting depth Z_r is 735 mm, which is physically meaningful. Overall WB4 seems to be the best model in order to estimate the average annual water balance in the study area in the absence of the filter, even though its performances are just slightly better than those of WB1 and WB2, that have a reduced number of parameters. The systematic underestimation of the runoff coefficient for humid catchments in the absence of a filter to separate surface runoff from infiltration may be a consequence of the hypothesis that all the incoming rainfall infiltrate in the soil and takes part to soil moisture dynamics. In fact, the highest values of ϕ might refer to catchments characterized by quite impervious soil, or by humid climate conditions and high soil moisture values. Both scenarios are likely to promote surface runoff at the expense of infiltration, thereby increasing the the observed runoff coefficients.

The increase of rooting depth when the filter is applied is a consequence of the increased amount of water directly converted into streamflow. In order to comply with the measured values of ϕ, the slow streamflow component (due to subsurface flow) needs to be reduced by increasing the rooting depth, and the soil water holding capacity, so as to remove an higher amount of water via evapotranspiration. Finally, it is worth to note how the calibration of the annual models led to reasonable values of Z_r in all cases ($500 < Z_r < 1000$), in agreement with previous studies (*Allen et al., 1998*).

3.2 Seasonal Water Balance

In this section the performances of the models at seasonal time scale are discussed. First are shown the results of the models applied at a seasonal time scale, either in the case where incoming rainfall is partitioned by means of the filter or in the case where all the precipitation is assumed to infiltrate into the soil. Then, the results obtained by coupling annual water balance models to the 'Seasonal coefficients' introduced in Section 2.8 are discussed.

The scatterplots shown in Figures 3.9 and 3.10 compare modeled and measured values of the seasonal runoff coefficients for the 39 study catchments. In this case each scatter plot contains 134 points (the seasons affected by snow dynamics, namely

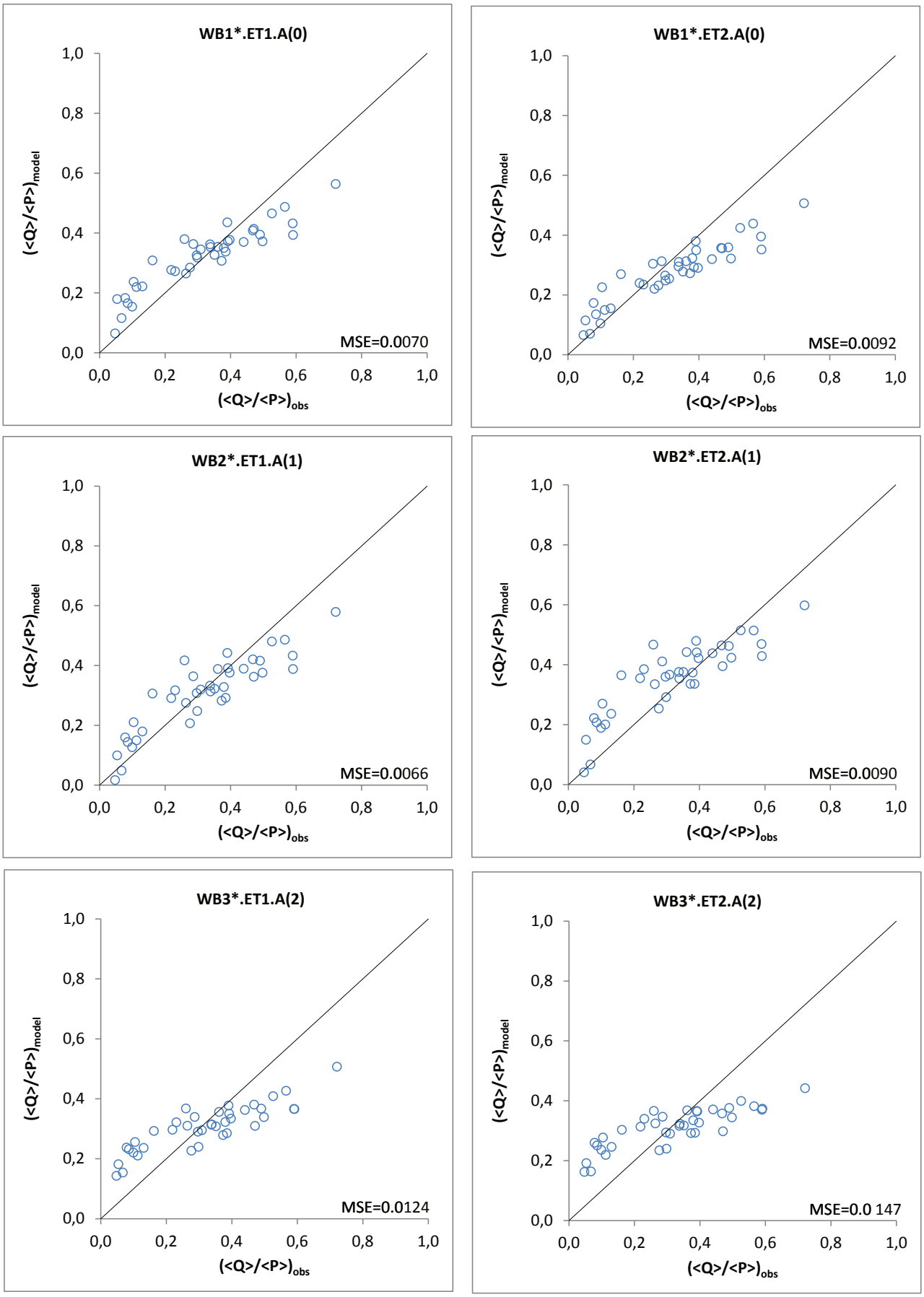

Figure 3.5: *Annual runoff ratio (with filter)(1/2)*

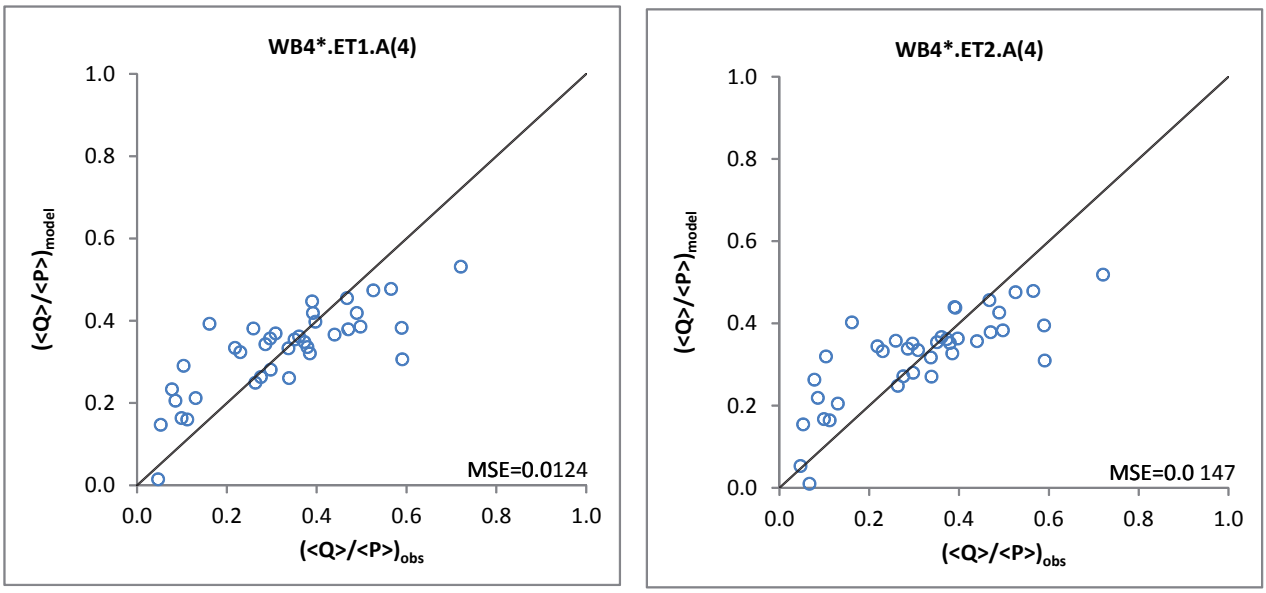

Figure 3.6: *Annual runoff ratio (with filter)(2/2)*

winter and spring, in northern catchments have been neglected). The performances of WB1 at seasonal scale are less brilliant than those at annual time scale. Even though the absence of parameters is an appealing feature of this model, WB1 doesn't seem to be robust enough to estimate the seasonal water balance in the study catchments. This fact is not surprising since the Budyko model has been designed to describe just long term average annual water balance and is not able to capture fluctuations at finer time scales. However, $WB1$ outperforms $WB5$, whose performances are judged to be unacceptable at seasonal timescale.

Despite some scattering, $WB1$ and $WB2$ are not affected by systematic biases, implying that the models don't tend to systematically overestimate (or underestimate) the seasonal observed runoff coefficients. Among the models applied at the seasonal scale $WB2$ outperforms the others. Its performances further improve by adding additional parameters to differentiate among the various seasons. The parameter in this case is the rooting depths Z_r, which is assumed to be a function of the season. A time variant rooting depth implies that the hydrologically active soil layer changes its depth according to the external meteorologic and climatic forcings. The model has been tested in three different scenarios:

1. a single value of Z_r has been considered as representative of the average rooting

Unfiltered rainfall

MODEL	PET	Parameters	MSE
WB1	ET1	-	0.0232
	ET2	-	0.0226
WB2	ET1	$Z_{r\text{-}year}$ = 615 mm	0.0231
		$Z_{r\text{-}summer,spring}$ = 450 mm	0.0161
		$Z_{r\text{-}autumn,\ winter}$ = 1500 mm	
		$Z_{r\text{-}winter}$ = 1500 mm	0.0134
		$Z_{r\text{-}spring}$ = 240 mm	
		$Z_{r\text{-}summer}$ = 570 mm	
		$Z_{r\text{-}autumn}$ = 1500 mm	
	ET2	$Z_{r\text{-}year}$ = 435 mm	0.0219
		$Z_{r\text{-}summer,spring}$ = 330 mm	0.0177
		$Z_{r\text{-}autumn,\ winter}$ = 900 mm	
		$Z_{r\text{-}winter}$ = 570 mm	0.0133
		$Z_{r\text{-}spring}$ = 195 mm	
		$Z_{r\text{-}summer}$ = 510 mm	
		$Z_{r\text{-}autumn}$ = 975 mm	
WB5	ET1	-	0.0461

Filtered rainfall

MODEL	PET	Parameters	MSE
WB1*	ET1	-	0.0205
	ET2	-	0.0148
WB2*	ET1	$Z_{r\text{-}year}$ = 1170 mm	0.0184
		$Z_{r\text{-}summer,spring}$ = 630 mm	0.0150
		$Z_{r\text{-}autumn,\ winter}$ = 1500 mm	
		$Z_{r\text{-}winter}$ = 1500 mm	0.0138
		$Z_{r\text{-}spring}$ = 405 mm	
		$Z_{r\text{-}summer}$ = 870 mm	
		$Z_{r\text{-}autumn}$ = 1500 mm	
	ET2	$Z_{r\text{-}year}$ = 750 mm	0.0147
		$Z_{r\text{-}summer,spring}$ = 510 mm	0.0112
		$Z_{r\text{-}autumn,\ winter}$ = 1500 mm	
		$Z_{r\text{-}winter}$ = 1500 mm	0.0092
		$Z_{r\text{-}spring}$ = 300 mm	
		$Z_{r\text{-}summer}$ = 750 mm	
		$Z_{r\text{-}autumn}$ = 1500 mm	

Figure 3.7: *Parameters of the seasonal water balance models*

depth along the entire year;

2. two different values of Z_r have been considered, one for spring and summer and one for winter and autumn;

3. a different value of Z_r has been assumed for every season.

In all cases the calibrated rooting depths display, similarly to what happens at the annual scale, higher values when the rainfall partitioning is performed. In such a case, to agree with the observed runoff coefficients, the slow flow component of the streamflow has to be reduced to compensate the surface runoff that is directly conveyed to the drainage network thanks to the filtering. In general, the calibrated rooting depths are lower when the ET2 potential evapotranspiration dataset (MODIS) is used. In fact, the MODIS dataset tends to provide higher values of potential evapotranspiration with respect to ET1 (Figure 1.1). Higher values of PET lead to higher soil moisture depletion rate and, hence, lower values of Z_r are required to buffer the infiltration. Because of the same reason, the shallow rooting depth displayed in Figure 3.7 during Spring (for both the PET datasets, and both in the the presence and absence of the filter) is likely to be a consequence of the higher role of surface runoff during Spring, possibly due to the short and intense rainfall events that typically characterize this season at latitudes such as those involved in the study. Moreover, the effect is possibly enhanced by the relatively high values of $< PET >$ and $< P >$ during this season.

An upper bound has been imposed to Z_r (1500 mm) during the calibration procedure for physical reasons. Figure 3.8 shows, as an example, the MSE as a function of the rooting depth for every season in the calibration of WB2 at seasonal scale. The plot evidences that an upper limit for Z_r can be hardly detected by calibration, a feature common to many hydrological models.

The scatterplots of WB2.ET1.S(4) and WB2.ET2.S.(4) shown in Figure 3.10 suggest that WB2 can provide a good estimate of the seasonal water balance across the study catchments, either coupled with the MODIS or CGIR potential evapotranspiration datasets.

The cases in which the filtering procedure is applied prior to the water balance models are presented in the scatterplots of Figures 3.11 and 3.12.

The filter increases the modeled runoff coefficient, leading to increased performances in all the models. Again, WB2 is the best performing model, as it happens

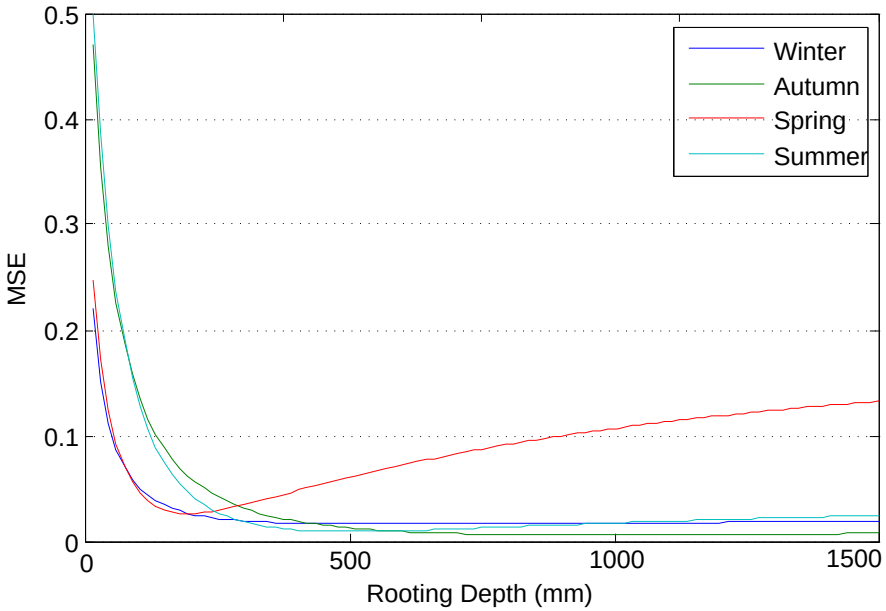

Figure 3.8: *MSE as a function of the season and Z_r in the calibration of WB2.ET2.S.(4)*

in the case without the rainfall partitioning. Nevertheless, there are some concerns regarding the use of the rainfall partitioning in this context. The filter adopted (Section 2.1.1), can't be used to predict the water balance in ungauged catchments, as the infiltration is estimated from streamflow time series. Moreover, there are concerns about the amount of the incoming rainfall directly turned into surface runoff by the filter. Figure 2.3 highlights how, during high discharge events, a relevant part of rainfall is assumed to bypass the soil system, possibly leading to overestimated runoff coefficients. Overall, the consistency of the estimate of fast surface runoff provided by the filter still needs to be assessed (e.g. by using tracer data).

3.2.1 Seasonal decomposition

In order to exploit the improved accuracy of the water balance models at annual timescale we seek to reproduce the seasonal variability of the average runoff coefficients by coupling annual water balance models and the 'Seasonal coefficients' ψ_i, as detailed in section 2.8. In this manner, each point of the annual scatterplots is splitted into four points corresponding to the runoff coefficients during the different seasons of the year. The Table 3.13 summarizes the results of the method, and the calibrated values of the model parameters.

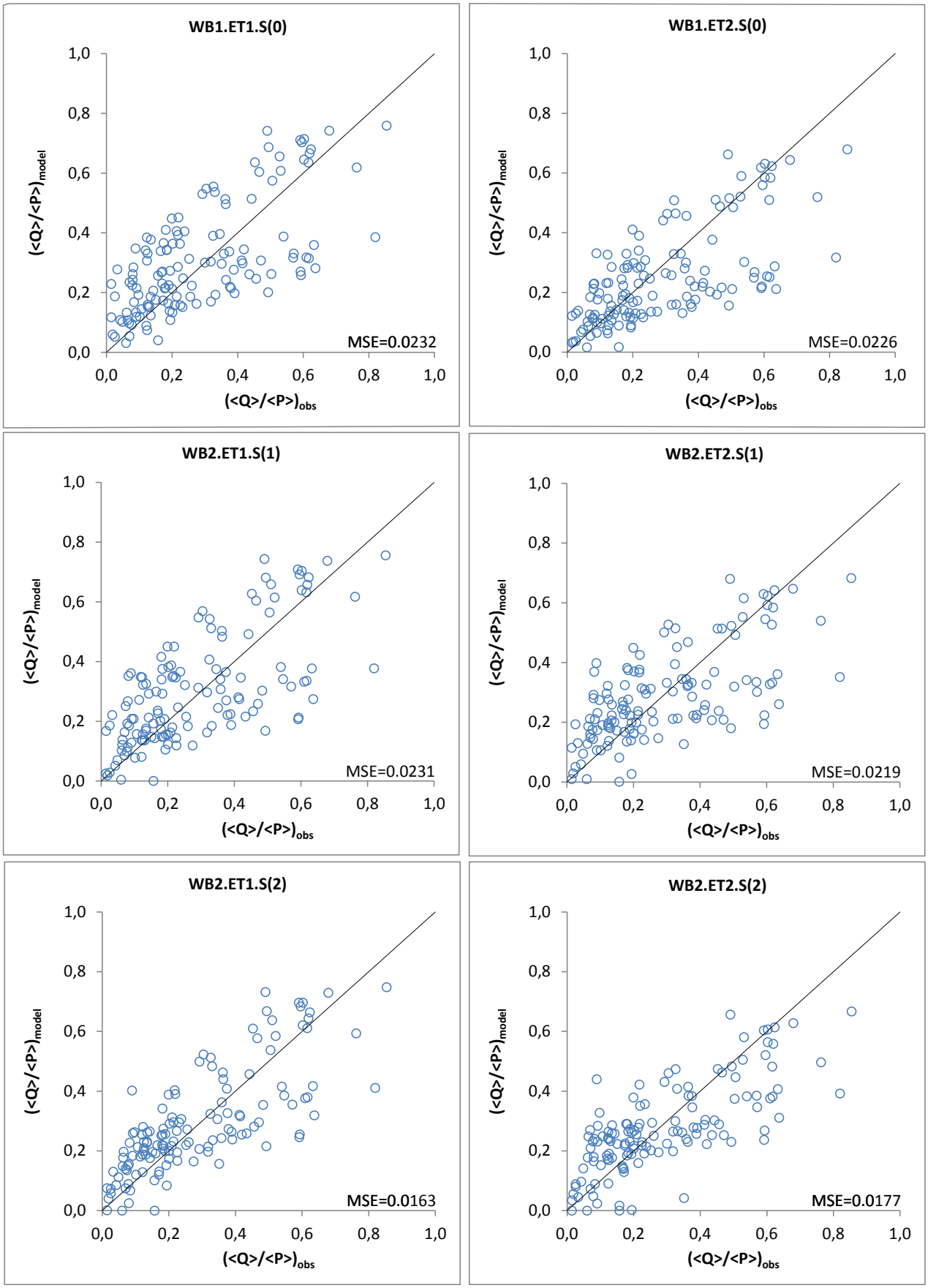

Figure 3.9: *Seasonal runoff ratio (without filter) (1/2)*

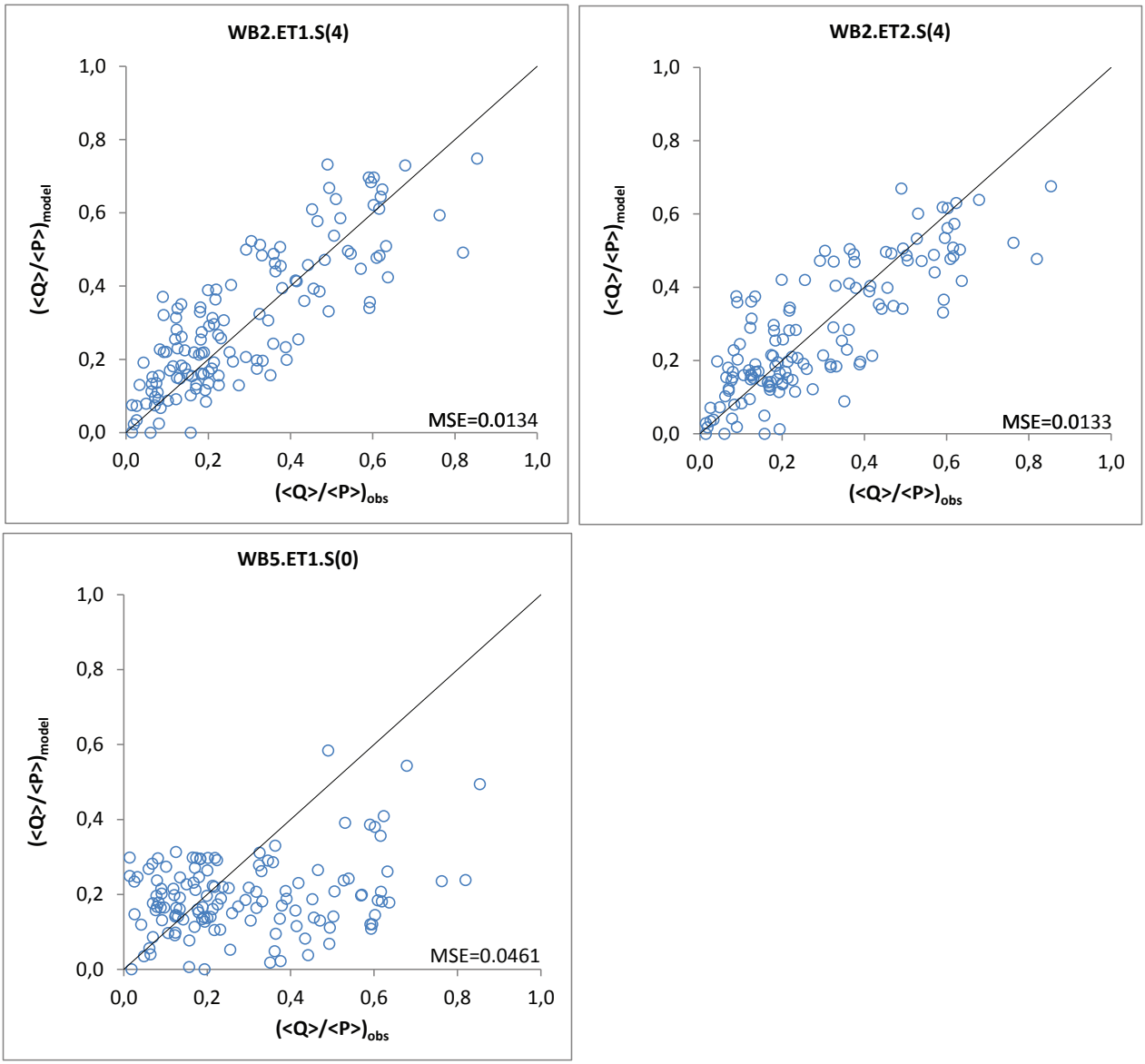

Figure 3.10: *Seasonal runoff ratio (without filter) (2/2)*

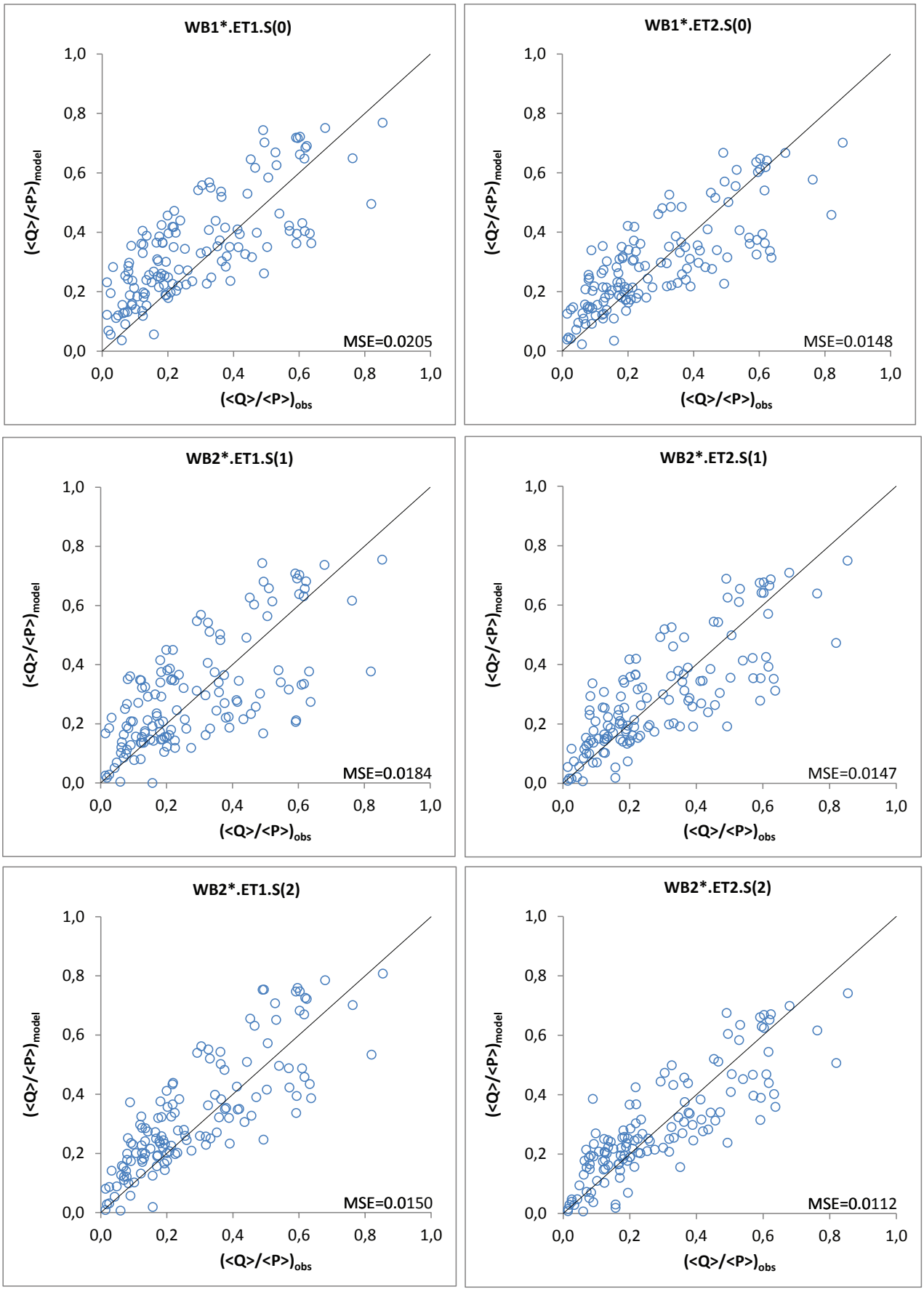

Figure 3.11: *Seasonal runoff ratio (with filter) (1/2)*

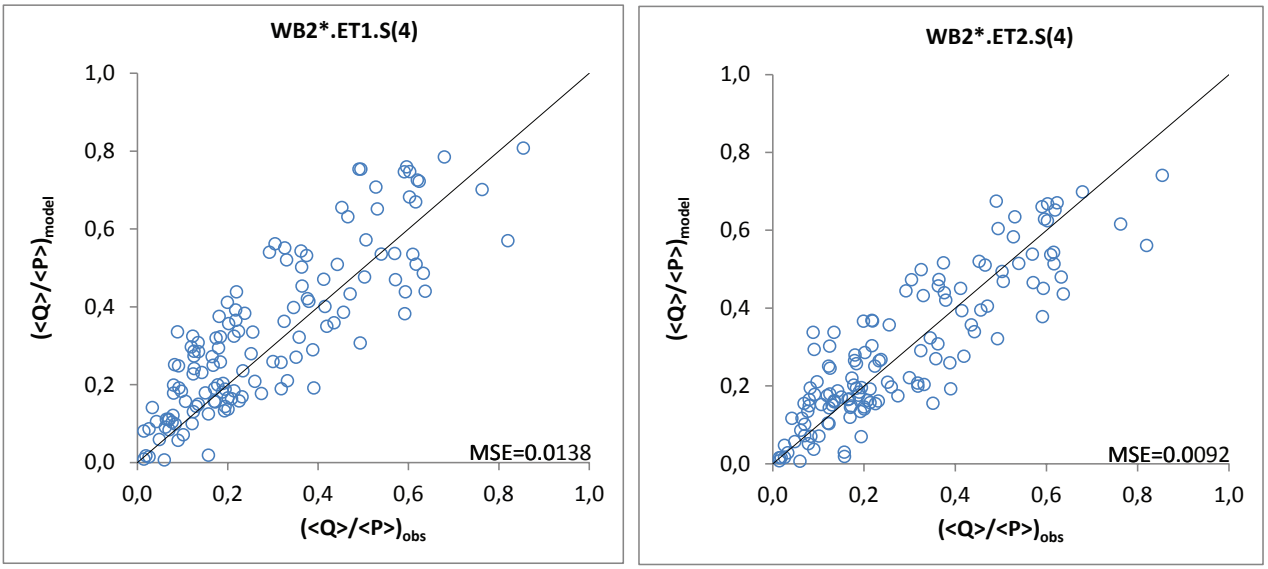

Figure 3.12: *Seasonal runoff ratio (with filter) (2/2)*

Unfiltered Rainfall

MODEL	PET	Parameters ψ_{winter}= 1.44 ψ_{spring}= 1.42 ψ_{summer}= 0.56ψ_{autumn}=0.59	MSE
WB1	ET1	-	0.028
	ET2	-	0.036
WB2	ET1	$Z_{r\text{-}year}$ = 420 mm	0.014
	ET2	$Z_{r\text{-}year}$ = 300 mm	0.018
WB3	ET1	$Z_{r\text{-}year}$ = 900 mm , K = 0.0525	0.018
	ET2	$Z_{r\text{-}year}$ = 700 mm , K = 0.0525	0.021
WB4	ET1	λ_u=0.2	0.028
	ET2	λ_u=0.2	0.030
WB5	ET1	-	0.027

Filtered Rainfall

MODEL	PET	Parameters ψ_{winter}= 1.44 ψ_{spring}= 1.42 ψ_{summer}= 0.56ψ_{autumn}=0.59	MSE
WB1*	ET1	-	0.029
	ET2	-	0.031
WB2*	ET1	$Z_{r\text{-}year}$ = 735mm	0.010
	ET2	$Z_{r\text{-}year}$ = 480mm	0.015
W3*	ET1	$Z_{r\text{-}year}$ = 1200 mm , K = 0.0124	0.014
	ET2	$Z_{r\text{-}year}$ = 950 mm , K = 0.0147	0.016
WB4*	ET1	λ_u=0.8 ,λ_s =0.015 , W_p=30000 mm , ζ=5.5	0.0316
	ET2	λ_u=0.8 ,λ_s =0 .015, W_p=27000 mm , ζ=4.9	0.0315

Figure 3.13: *Parameters of the seasonal decomposition of the annual water balance models*

The seasonal coefficients (namely ψ_{sp}, ψ_{su}, ψ_a, ψ_w) have been obtained directly from the data of rainfall and discharge as described in Section 2.8. The overall performances of this method are acceptable and comparable to the performances of the models WB1 to WB5 directly applied at the seasonal timescale. Despite the size of the study area and the ensuing heterogeneity of climate conditions, the inter-catchment variability of the seasonal coefficients ψ is quite small, thereby suggesting the potential suitability of regionalization approaches where the estimates of these coefficients could be extended to ungauged networks. Moreover, a clear gap between the warmer and colder seasons is clearly visible. These evidences seem to support the soundness of the approach. The emerging pattern of the ratio between average seasonal runoff coefficient and average annual runoff coefficient in catchments spread within a vast area, like the Eastern US (Figure 2.7), could represent a good starting point for further studies. The method based on the seasonal coefficients could be accused of having the same limits of the models based on the rainfall partitioning through of the filtering procedure because they both need discharge time series in order to be applied. However, for the very reason that the dispersion of the calculated seasonal coefficients is reasonably small even for wide areas such as the one involved in this study, a regionalization approach seems to be a reliable option for the application to ungauged sites. In this context, reducing the size of the domain, and grouping catchments featured by similar climate conditions, will further reduce the inter-catchment heterogeneity of the seasonal coefficients, increasing the robustness of the approach.

Figure 3.13 shows that the best performing models where the seasonal decomposition of runoff is adopted are WB2 and WB3, whose performances slightly increase when the filter is applied. In particular WB2, coupled with ET1, outperform the other models in both the filtered and unfiltered case. A selection of the scatterplots that represent the performances of the models based on the seasonal coefficients are presented in Figure 3.14.

The performances of the models slightly increase when the rainfall partitioning is preformed prior to the water balance modeling. The first two plots of Figure 3.14 show the performances of WB1 and WB1*. WB3 fails in reproducing the spatial and inter-seasonal variability of the average seasonal water balance across the study catchments, while WB5 strongly underestimates the runoff coefficient in most cases, and its performances are particularly poor in the catchments and seasons featured by reduced discharge.

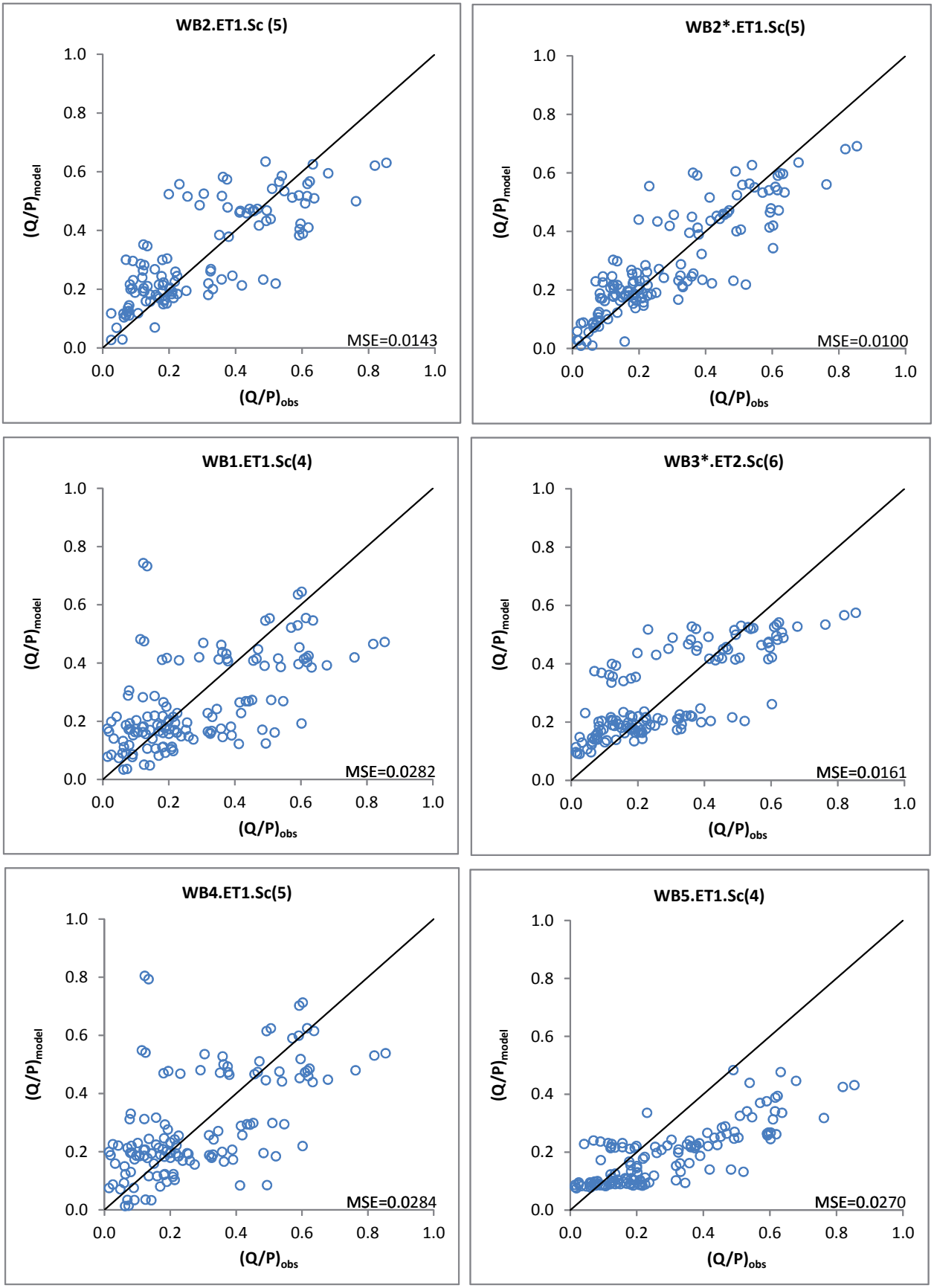

Figure 3.14: *Seasonal runoff ration (seasonal coefficients)*

Many plots display a marked discontinuity, especially when in the original annual water balance model the inter-catchment variability is underestimated (i.e. the model is unable to describe the spatial heterogeneity of the underlying eco-hydrological process). This is a consequence of the calibrated values of the seasonal runoff coefficients having very different values ($\psi_{sp} \approx \psi_{su} \neq \psi_a \approx \psi_w$).

3.3 Ranking of the models

Given the high number of models considered in this study and the number of different versions and combinations of variants involved, an objective ranking of models based on their performances and the number of calibrated parameters proves useful. To this aim, the *Akaike Information Criteria* (Section 2.9), has been used. The charts in Figure 3.15 present a ranking of the annual and seasonal models performed by means of the AIC. ΔAIC represents the difference of AIC between the model and the best performing one. High values of ΔAIC means that the model is far from being the best in terms of MSE, given the number of calibrated params.

Figure 3.15 shows that the models involving the filtering procedure are mostly located in the upper portion of the chart, in particular at the annual timescale. In fact, the models that postulate the complete infiltration of the incoming rainfall tend to underestimate the observed runoff coefficients. WB2 occupies most of the top 10 positions in the ranking, while WB4 and WB1 lie at the bottom of the annual ranking.

The ET1 potential dataset performs better at annual time scale than ET2, while at seasonal time scale an opposite trend is the observed. The explanation might lie in the fact that the algorithm used to estimate PET in the MODIS dataset (ET2) is more sophisticated than the one adopted in the CGIR dataset (ET1) and thus it is able to catch intra-annual fluctuation of climate conditions in a more accurate and robust way. However, the simplicity of the method employed to compute potential evapotranspiration in the CGIR dataset (Section 1.2.2), possibly enhances the robustness of the model during longer time spans.

The models based on the application of the seasonal coefficients performs poorly when WB1 and WB4 are used for the annual water balance. Performances increase for WB2 and WB3. In particular the WB2 variant having the best performances (WB2*.ET1.Sc) ranks as second in the overall seasonal models chart.

Model performances, and the role of different PET datasets are summarized in

ANNUAL MODELS CHART

Rank	Model	Δ AIC
1	WB2*.ET1.A(1)	0.0
2	WB1*.ET1.A(0)	0.3
3	WB4.ET1.A(1)	7.0
4	WB1*.ET2.A(1)	11.0
5	WB2*.ET2.A(1)	12.1
6	WB4.ET2.A(1)	15.0
7	WB1.ET1.A(0)	18.6
8	WB2.ET1.A(1)	23.6
9	WB4*.ET1.A(4)	25.6
10	WB3*.ET1.A(2)	26.6
11	WB4*.ET2.A(4)	27.0
12	WB3*.ET2.A(2)	33.2
13	WB2.ET2.A(1)	33.8
14	WB3.ET1.A(2)	36.8
15	WB1.ET2.A(0)	43.9
16	WB3.ET2.A(2)	45.8
17	WB5.ET1.A(0)	53.1

SEASONAL MODELS CHART

Rank	Model	Δ AIC
1	WB2*.ET2.S (4)	0.0
2	WB2*.ET1.Sc(5)	13.1
3	WB2*.ET2.S (2)	22.3
4	WB2.ET2.S (4)	49.3
5	WB2.ET1.S (4)	50.3
6	WB2*.ET1.S (4)	54.3
7	WB1*.ET2.S(0)	55.7
8	WB2*.ET2.S (1)	56.8
9	WB2.ET1.Sc(5)	61.1
10	WB2*.ET1.S (2)	61.5
11	WB2*.ET2.Sc(5)	62.9
12	WB3*.ET1.Sc(5)	64.0
13	WB2.ET1.S (2)	70.9
14	WB3*.ET2.Sc(6)	78.9
15	WB2.ET2.S (2)	83.6
16	WB2*.ET1.S (1)	86.8
17	WB2.ET2.Sc(5)	90.4
18	WB3.ET1.Sc(6)	93.1
19	WB1*.ET1.S(0)	99.3
20	WB2.ET2.S (1)	110.2
21	WB1.ET2.S(0)	112.4
22	WB3.ET2.Sc(6)	115.8
23	WB1.ET1.S(0)	115.9
24	WB2.ET1.S (1)	117.3
25	WB1.ET1.Sc(4)	150.1
26	WB4.ET1.Sc(5)	153.0
27	WB1*.ET1.Sc(4)	155.2
28	WB4.ET2.Sc(5)	159.0
29	WB1*.ET2.Sc(0)	163.2
30	WB4*.ET2.Sc(8)	172.9
31	WB4*.ET1.Sc(8)	173.3
32	WB1.ET2.Sc(4)	184.3
33	WB5.ET1.S(4)	207.9

a series of histograms presenting the frequency distribution of ΔAIC for different groups of models/variants (Figure 3.16) and potential evapotranspiration datasets (Figure 3.17), both at annual and seasonal time scales.

The performances of every single model, without differentiation between the two PET datasets and the absence/presence of the filter are shown in Figure 3.16. The histograms are complemented with the median value of ΔAIC in order to objectively assess the performances of each approach. The histograms highlight how WB1, WB2 and WB4 exhibit the best performances at annual time scale, having a relevant number of variants featuring low values of ΔAIC. WB5 shows instead poor results. Concerning the seasonal models, it can be noted how WB2 largely outperforms the other models. WB3 displays good performances as well, while the evidence in favor of models WB1, WB4 and WB5 is quite limited.

Figure 3.17 shows the differences between the two potential evapotranspiration datasets. The different shape of the frequency distribution of ΔAIC across the models in the two cases suggest that the CGIR dataset provide more reliable estimates of the potential evapotranspiration at annual time scale in the study area, while at seasonal time scale, the MODIS dataset seems to preform slightly better.

3.3.1 Streamflow probability density function

A reliable water balance model allows the estimate of annual or seasonal runoff coefficients for a catchment not provided with streamflow measurements. Since the catchments used in this study for the model calibration are quite evenly spread throughout a vast area, the resulting calibrated parameters can be assumed as reference values for the eastern US, and thus used to predict streamflow regimes in any other river in this area. Following the approach described in Section 2.10, the knowledge of average annual/seasonal runoff coefficients allows an estimate of the seasonal streamflow probability density function (PDF) at any location of a river network, starting from rainfall data and basic topography information of the drainage basin (digital terrain models).

The application is conceived as follows. Four validation catchments have been selected at random within the study area. Catchments have been selected so as to be pristine (absence of regulation) and with a contributing area smaller that $10^3\,Km^2$. In these catchments, the analytical streamflow distribution provided by equations (2.66) and (2.67), have been compared to the frequency distribution of observed

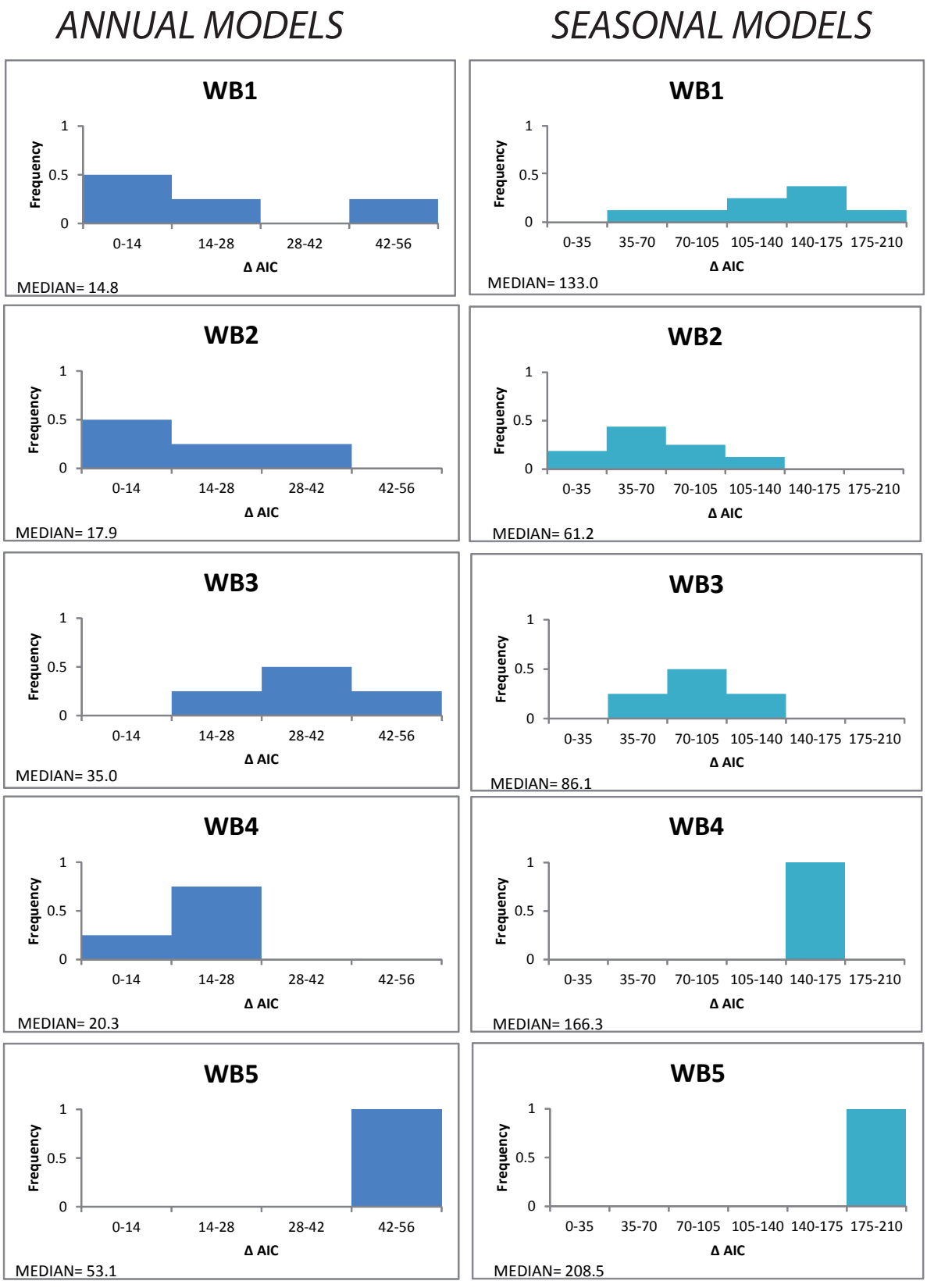

Figure 3.16: *Annul and Seasonal water balance models histograms*

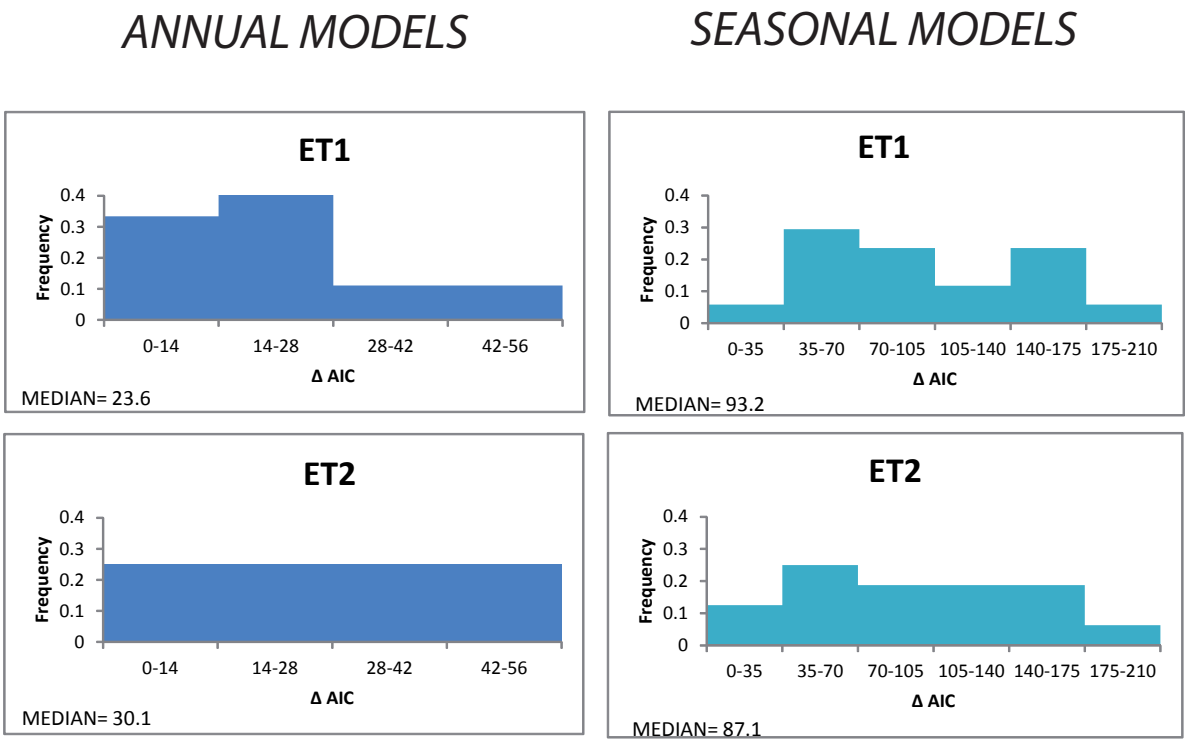

Figure 3.17: *Annual and seasonal PET comparison*

flows. The four parameters of the analytical model have been estimated only based on rainfall, climate and landscape data. In particular:

- γ_w has been estimated based on rainfall data as $\gamma_w = 1/(\alpha\, A)$, where α is the average rainfall depth on rainy days and A is the catchment area;
- a , b have been estimated using a digital terrain map (as detailed in *Biswal and Marani (2010)*);
- λ is estimated using a water balance model as $\lambda = \phi\, \lambda_p$, where $\phi =< Q > / < P >$ is the average seasonal runoff coefficient and λ_p is the frequency of rainy days in a season.

In this section are presented the results obtained by computing *lambda* by means of WB2 (applied at seasonal timescales) coupled with the CGIR potential evapotranspiration dataset without the aid of the filter (WB1.ET2.S). The model is chosen because it is the best performing one among the models without the rainfall partitioning procedure (indeed, the presence of the filter would prevent the framework to be applied in a predictive manner).

The model WB2.ET1.S applied in the 4 validation catchments during all the available seasons, provides the following results:

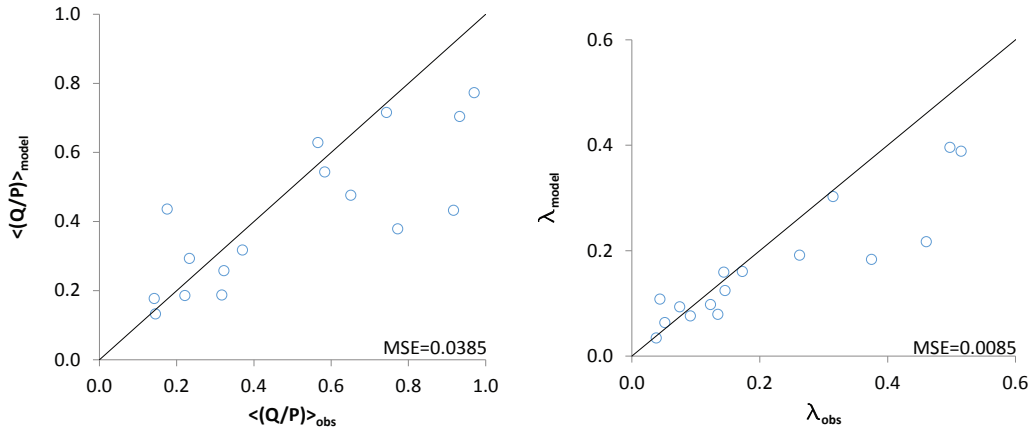

Figure 3.18: *Scatterplot of the seasonal average runoff coefficient and parameter λ for the 4 validation catchments*

Performance are remarkably good in almost all cases. In the left hand side scatterplot of Figure 3.18 it is possible to see two outliers, where the model underestimates the runoff coefficient. It can be noted how the two dots refer to the

Catchment	<Q>/<P> measured	<Q>/<P> observed	λ measured	λ observed
Sand Run (Autumn)	0.32	0.26	0.12	0.10
Sand Run (Spring)	0.77	0.38	0.37	0.18
Sand Run (Summer)	0.22	0.19	0.09	0.08
Sand Run (Winter)	0.93	0.70	0.51	0.39
Youghiogheny River (Autumn)	0.37	0.32	0.15	0.12
Youghiogheny River (Spring)	0.92	0.43	0.46	0.22
Youghiogheny River (Summer)	0.32	0.19	0.13	0.08
Youghiogheny River (Winter)	0.97	0.77	0.50	0.40
Daddy's Creek (Autumn)	0.23	0.29	0.07	0.09
Daddy's Creek (Spring)	0.65	0.48	0.26	0.19
Daddy's Creek (Summer)	0.14	0.18	0.05	0.06
Daddy's Creek (Winter)	0.74	0.72	0.31	0.30
Big Piney Creek (Autumn)	0.18	0.44	0.04	0.11
Big Piney Creek (Spring)	0.58	0.54	0.17	0.16
Big Piney Creek (Summer)	0.15	0.13	0.04	0.03
Big Piney Creek (Winter)	0.57	0.63	0.14	0.16

Figure 3.19: *Seasonal average runoff coefficient and parameter λ for the 4 validation catchments*

spring season of two mountain catchments (Sand Run and Youghiogheny river), very likely affected by snow melting during spring, leading to underestimated runoff coefficients. As it can be seen on the map in Figure 3.20, despite the relativedly low latitude, the two catchments under consideration experience intense snowfall during the year since they are located on the Appalachian Mountains, the second biggest mountain range in the US.

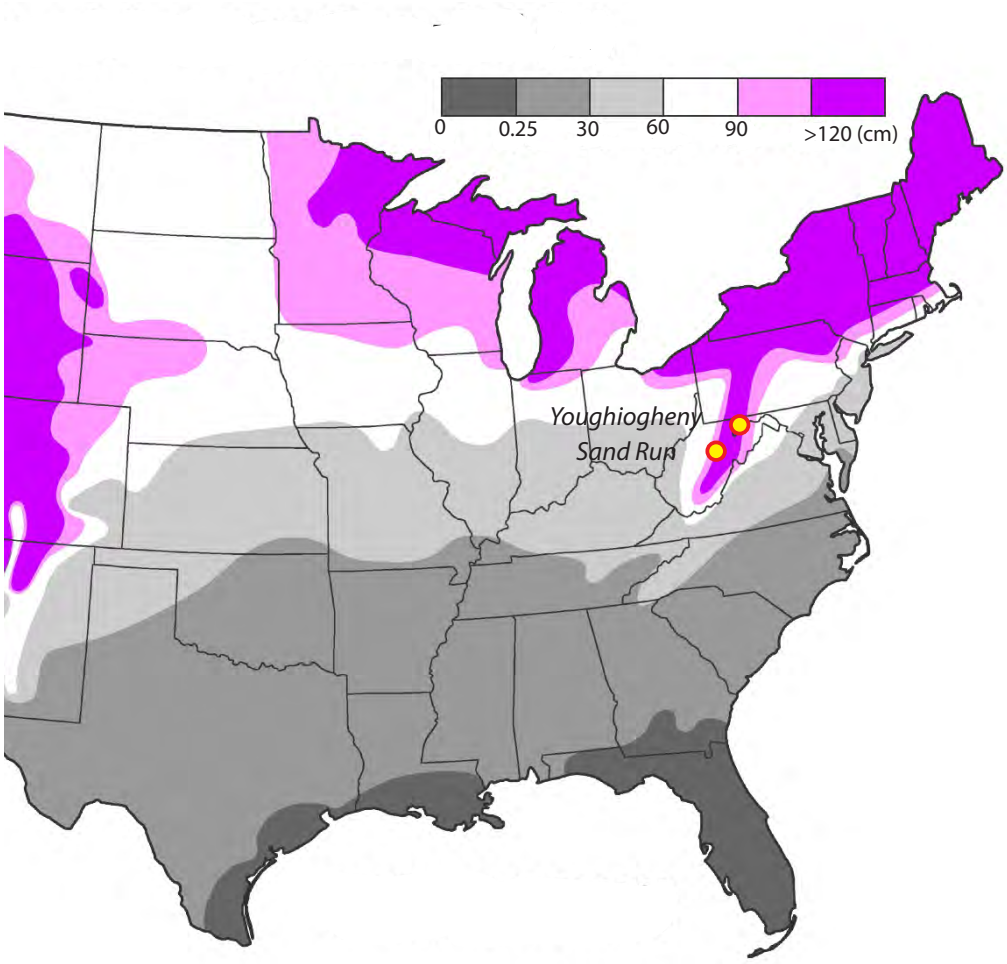

Figure 3.20: *Cumulated average annual snowfall in the Eastern United States (NOAA)*

The final result obtained from the application of the streamflow model led to the results in the plots shown: in Figures 3.21 the seasonal probability density functions obtained for the *Daddy's Creek* are presented, while in Figure 3.22 are shown the PDFs for a selected season (summer) in the other three catchments. The plots include the Cumulative Density Function (CDF) as well, which is the integral of the PDF, namely the probability to observe a streamflow magnitude larger that Q.

For what concerns the Daddy's Creek, the modeled PDFs have a strong peak in correspondence of lowest streamflows especially during the warmest seasons of the year. The shape of the streamflow PDF is properly reproduced by the analytical model in all seasons. Though, the model tends to slightly underestimate the high flows, providing lower probability for large discharges compared to the observations. In Spring the catchment displays a tendency to behave similarly to catchments affected by snow dynamics even if it is not affected by snowfalls during cold seasons. The physical reason could be seek in the ability of catchments to store a larger amount of water during cold seasons, and to release such water in Spring. The behavior is a direct consequence of the seasonal streamflow predicted by the water balance model, which underestimates the observed runoff almost in all seasons for all the validation catchments (Figure 3.23)

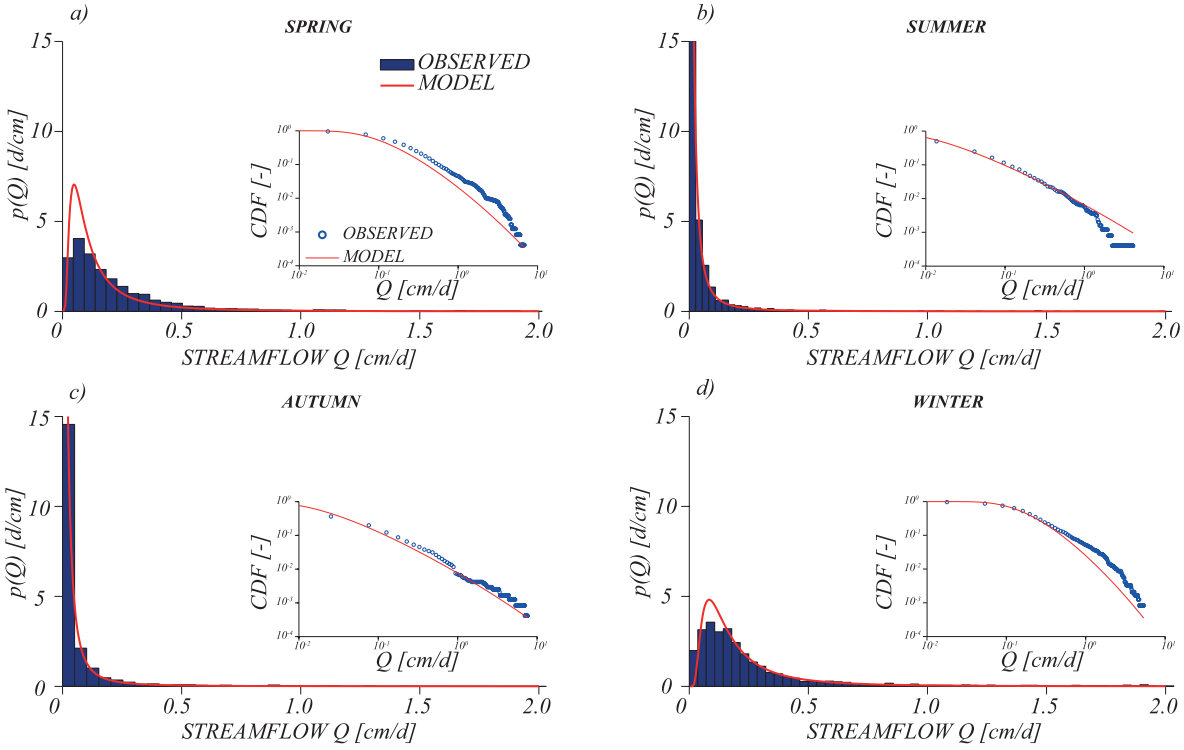

Figure 3.21: *Streamflow PDFs in all seasons (Daddy's Creek)*

Autumn and Summer are the seasons during which model performances are particularly high; the modeled streamflow PDF has an almost perfect agreement with

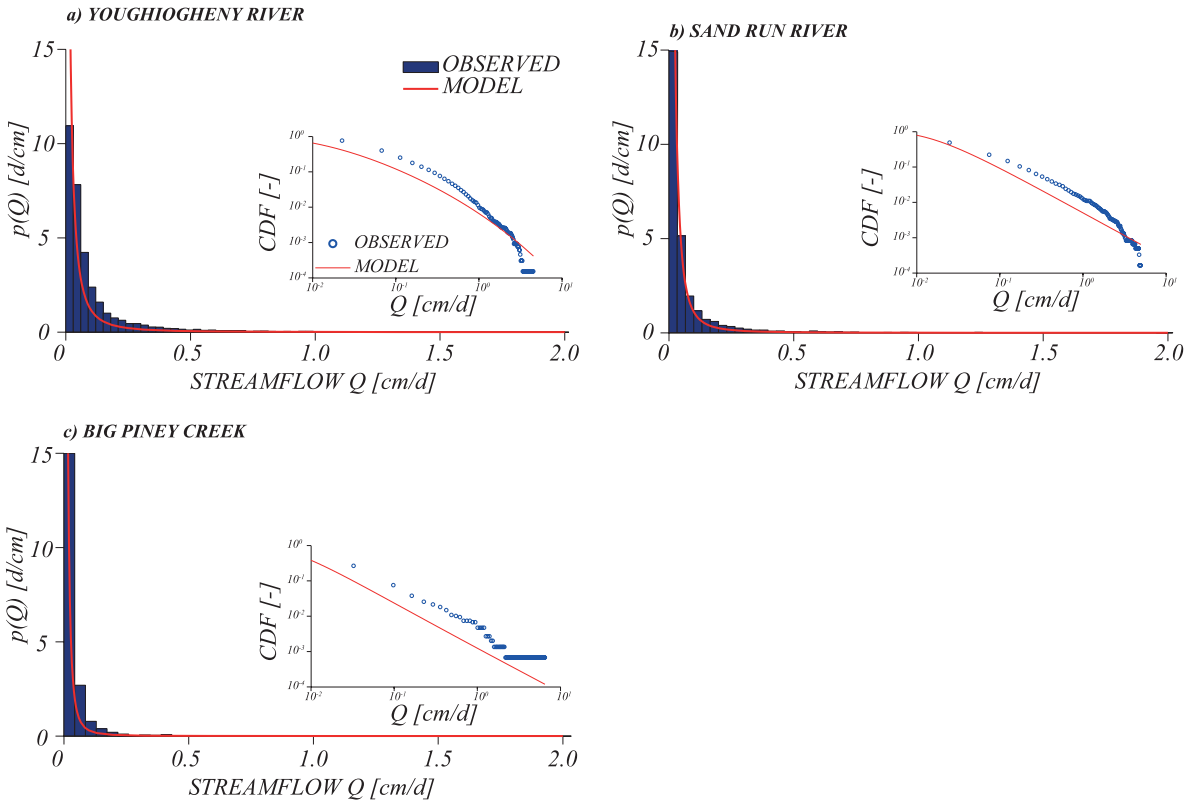

Figure 3.22: *Streamflow PDFs in summer (Youghiogheny River, Sand Run River, Big Piney Creek)*

the observation, except for the slight overestimation of the probability associated to high streamflows in Summer. In winter the model behaves similarly to spring, and tends to underestimate the probability of occurrence of high discharge events.

The plots of Figure 3.22 show the performances of the model in Summer for the remaining validation catchments (Youghiogheny River, Sand Run and Big Piney Creek). The results are in general quite good despite the underestimation of the average seasonal streamflow. The analytical curves reproduce the observations, implying that the streamflow PDF is properly estimated by the model. However, the modeled CDFs are slightly shifted downward as a result of the reduced amount of water available for streamflow generation estimated by the water balance model.

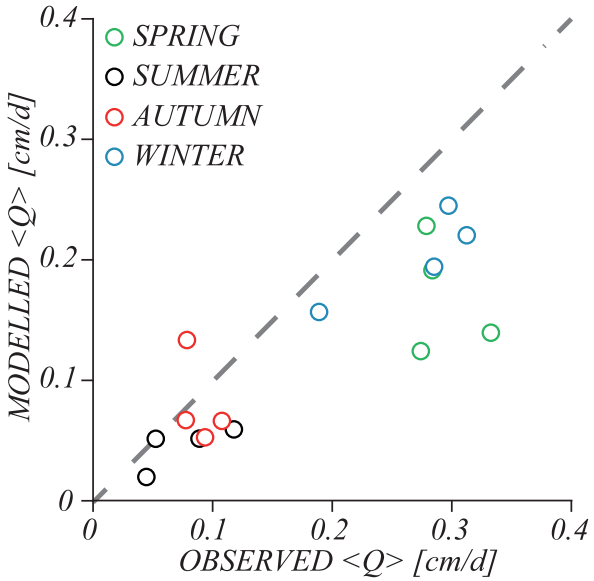

Figure 3.23: *Average seasonal runoff in the 4 validation catchments*

The ability of the model to catch the shift from erratic to persistent streamflow regimes across different seasons seems to be particularly valuable. A persistent flow regime is weakly variable around the mean (low values of the Coefficient of Variation of daily flows, $CV(Q)$) and, hence, more predictable. Such a flow regime is likely to be observed when flow-producing rainffall are frequent enough so that their mean interarrivals are smaller that the mean duration of the flow pulses. The range of stramflows observed between two subsequent events is reduced and a persistent supply of water is guaranteed to the river from the catchment soil. This kind of

regimes are typically observed during humid, cold seasons (as it is in the study case) and/or in slow responding catchments.

Conversely, if the main interarrival between flow producing rainfall events is larger that the mean duration of flow pulses, the range of streamflows observed between two subsequent rainfall events is larger because the river has enough time to dry significantly before the arrival of the next pulse. The result is a flow regime characterized by low discharges with high variance. This kind of regime is defined as " erratic" and it is typical of fast responding catchments during season with sporadic rainfall events or during hot humid seasons, as in the case of all the validation catchments during Summer.

Erratic flow regimes are characterized by monotonically decreasing PDFs: they display high probability in the low flows and a long tail for high flows. Instead, persistent regimes are featured by a bell-shaped probability density function, meaning that intermediate values of streamflows are quite likely to be observed. A graphical characterization of the two flow regimes is presented in Figure 3.24.

The coefficient of variation of daily flows (CV) is a simple indicator of the hydrologic regime of a catchment: high values of CV characterize erratic regimes, while low values characterize persistent regimes. The coefficient of variation is defined as the square root or the variance (σ^2), normalized with respect to the mean (μ), namely $CV(Q) = \sqrt{\sigma^2(Q)/\mu^2(Q)}$.

Figure 3.25 shows a comparison between the estimate of CV(Q) provided by the analytical model and the observed coefficients of variation of daily discharges in the four validation catchments. The plot proves that the model is able to reproduce accurately the streamflow variability and its inter-seasonal dynamics across the study catchments.

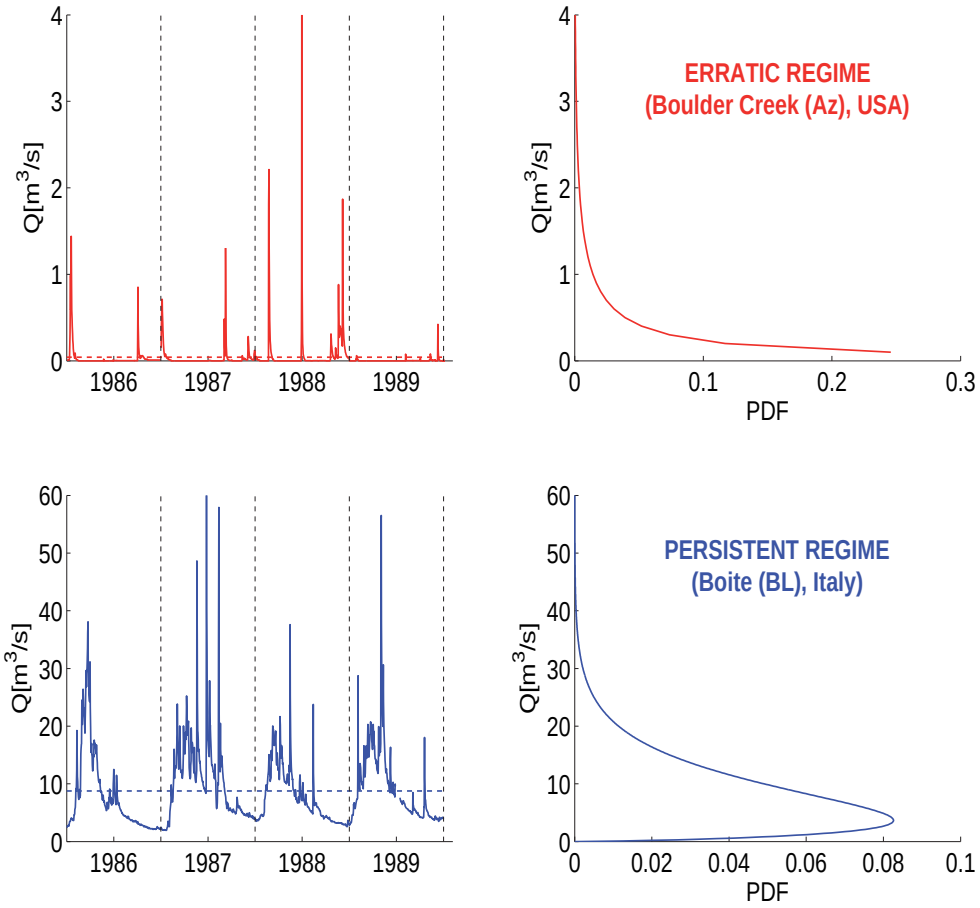

Figure 3.24: *Difference between the stramflow PDFs of erratic and persistent regimes*

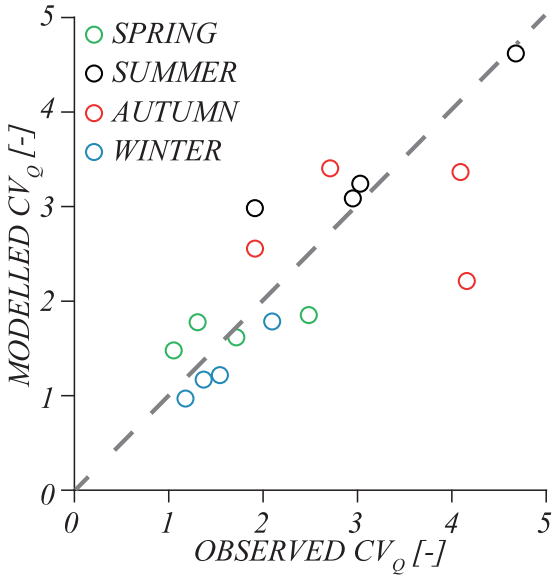

Figure 3.25: *Coefficient of variation of daily discharges in the validation catchments*

Chapter 4

Conclusions

Modeling the water balance is a challenging issue because of the variety and complexity of the underlying physical processes and because of stochasticity of the hydrological variables involved. It is also one of the most important topics in hydrologiy because of the impact of water availability (overall amounts and temporal variability) on human beings and ecosystems.

The aim of this thesis was to identify among a set of water balance models, the best performing model in order to provide an estimation of the annual or seasonal average runoff coefficients in the Eastern United States. The runoff coefficient is used as a key parameter for an analytical model which is able to estimate the streamflow PDF in ungauged catchments starting from rainfall time series and basic topographical information about the contributing catchment.

Five different water balance models (WB1, WB2, WB3, WB4, WB5) were compared, emphasizing their similarities and differences. The models were applied to 39 pristine catchments in the US east of the Rocky Mountains, a vast area featured by various geological, morphological, ecological and hydroclimatological characteristics. Because of the heterogeneity of all these forcing physical variables, the catchments involved in the study can be deemed to be representative of diverse climate regimes.

The importance of evapotranspiration as one of the key fluxes involved in the water balance was highlighted, and the importance of having a robust model able to estimate the magnitude of evapotranspiration fluxes was discussed. Unlike the other terms of the water balance (i.e. precipitation and runoff), that can be easily measured with sufficient accuracy, evapotranspiration process involve many variables highly heterogeneous in time and space and is strictly related to ecological feature difficult to be properly included in a comprehensive approach. Because of these

reasons, two global distributed potential evapotranspiration datasets (CGIR and MODIS) were used and compared. The CGIR dataset was found to provide better results at annual time scale, while the MODIS dataset outperforms the CGIR dataset at seasonal time scales.

In the study it was possible to use and test an embryonal version of a database which is able to manage and organize a huge amount of hydroclimatological data. The database displays great potential for research, water management, weather forecasts and much more, being a versatile instrument to which additional features and datasets from all around the Globe might be included.

In dealing with the water balance modeling, the surface runoff issue was explicitly analyzed. In particular, the estimation of the rainfall fraction that contributes to the fast response of a catchment is a cumbersome issue because its magnitude is tightly related to the rainfall regime (intensity, frequency, duration of events), climatology (dry/wet climate) and to morphological and geological features of the catchments (slopes, type of soil etc.). Hence, a proper estimation of the surface runoff would need a massive amount of information about every catchment and demanding computations, making the problem intractable within the framework of this thesis. A simple a-parametric filter was adopted in order to split the stremflow in its slow and fast components, respectively related to subsurface flow and surface runoff. The filter displayed some limits: because of its simplicity it disregard the physics of the processes involved in the surface runoff generation since it just processes the streamflow signal; moreover, the partitioning of rainfall between deep percolation and surface runoff sometimes provides infiltration rates unrealistically low. Considering that the use of such a filter would prevent the application to ungauged sites, the use of the filter was judged to be unnecessary , and the assumption that infiltration equals the overall rainfall was made.

The models tested can be empirical or physically based and are characterized by different degrees of accuracy in the description of the underlying processes. Every model is characterized by a certain number of parameters that were calibrated in order to achieve the best performances and, as a final result, a ranking of the models was made. To every model was assigned a rank by taking into account not only its departure from the observations, but also the number of calibrated parameters required. Special attention was given in assessing the reliability of these parameters from a physical point of view. As expected, all the models perform better at annual time scale.

At annual time scale the best performing models are WB1, WB2 and WB4. WB1 achieves slightly higher values of MSE compared to WB4 (which is the model having the lowest MSE). However, WB1 has the valuable feature to be parameterless and, hece, to have a wider applicability. WB3 and WB2 have just one calibrated parameter and result in comparable and overall satisfactory performances.

Moreover, it was also shown how the runoff coefficient displays a strong seasonality during the year.

At seasonal time scale WB2 was found to be the best performing model on the 39 study catchments achieving better performances than other models that require more demanding calibrations. The model WB2 was then used (in its version without rainfall partitioning) to predict the streamflow probability density function in 4 validation catchments (Sand Run, Youghiogheny River, Daddy's Creek, Big Piney Creek). The results achieved were encouraging provided that the observed streamflow regimes were properl captured by the models employed.

The whole method (based on the coupling between a water balance model and a geomorphic approach to predict the features of recessions) allowed a reliable estimate of the mean annual or seasonal streamflows and of the occurrence probability of specific discharge ranges. The frequency distribution of streamflows is a valuable indicator in hydrology, especially if this information can be achieved in an arbitrary ungauged section of a river. In fact, the probability density function of streamflows allows the identification of the optimal location of intake facilities for industrial purposes (e.g. hydropower production, power plant cooling etc.) or civil uses (e.g. water supply) and provides useful information to properly dimension water infrastructures. Moreover, the knowledge of the streamflow PDF supports protection plans against floods and natural hazards and helps to evaluating the hydraulic risk in endangered areas. From an ecological perspective, the frequency distribution of different streamflow magnitudes is a fundamental tool in the study of the dynamics of fishes and biomes characterizing riverine and riparian environments which are highly sensitive to streamflow availability and variability. Finally, the approach can be used in the study of stremflow regimes shifts caused by climate change or anthropogenic interventions (artificial dams or reservoirs).

The modeling approach investigated in this thesis is a valuable tool for a wide range of practical and scientific applications such as water resources management, ecological studies and flood risk assessment. The same methodology adopted in this study can be potentially exported to other areas of the Globe and the number of

models assessed can vary according to the data availability. In particular, WB1 and WB5 don't require discharge data to be calibrated and can thus be adopted to model the water balance in areas provided just with precipitation and potential evapotranspiration data. Instead, the availability of a good spatial coverage of streamflow and rainfall gauges allows the use of more sophisticated water balance models and regionalization approach in order to estimate the streamflow regime of ungauged rivers. The estimate of the streamflow probability density function at an arbitrary location of a river based on limited information on climate and landscape is an attractive endpoint with a wide range of consequences. This study has shown that this ambitious goal is now at reach of the scientific community.

Bibliography

[1] Abramowitz M., Stegun, I., A., (1964), Handbook of mathematical functions. *Dover, New York*

[2] Akaike, H. (1974),A new look at the statistical model identification. *Transactions on automatic control 19 (1974), 716-723*

[3] Allen, R., Pereira, L., S., Raes, D., Smith, M., (1998), Crop evapotranspiration - Guidelines for computing crop water requirements *FAO irrigation and drainage paper, 56, FAO, Rome*

[4] Arnell, N.W., Gosling, S.N (2013), The impacts of climate change on river flow regimes at the global scale. *Journal of Hydrology 486(2013) 351-364*

[5] Basso, S. (2013), Un approccio eco-Idrologico per la valutazione dell'effetto degli invasi artificiali sulle risore idriche nel bacino del Piave. *UniversitÃ degli studi di Padova - Dipartimento di ingegneria idraulica, marittima, ambientale e geotecnica.*

[6] Biondi, D., Freni, G., Iacobellis, V., Mascaro, G., Montanari, A. (2012), Validation of hydrological models: Conceptual basis, methodological approaches and a proposal for a code of practice. *Physics and Chemistry of the earth 42, 70-76*

[7] Biswal, B., Marani, M. (2010), Geomorphological origin of recession curves. *Geophysical research letters 37*

[8] Botter, G., Porporato A., Rodriguez-Iturbe, I., Rinaldo, A. (2007), Basin-scale soil moisture dynamics and the probabilistic characterization of carrier hydrologic flows: Slow, leaching-prone components of the hydrologic response. *Water resources research 43*

[9] Botter, G., Porporato A., Rodriguez-Iturbe, I., Rinaldo, A. (2009), Nonlinear storage-discharge relations and catchment streamflow regimes. *Water resources research 45*

[10] Botter, G., Basso, S., Rodriguez-Iturbe, I., Rinaldo, A. (2013), Resilience of river flow regimes. *PNAS 32*

[11] Budyko, M.I. (1955), Atlas of the heat balance. *Izd. GGO, Leningrad*

[12] (1956) Budyko, M., I., Heat valance of the Earth's surface (1956) *U.S. weather bur. Washington D. C.*

[13] Budyko, M., I., Zubenok, L., I., (1961), Determination of the evaporation from the land surface *Izv. AN SSSR, Ser. Geogr., 6*

[14] Budyko, M.I. (1974), Climate and life. *Academic, San Diego, California (1974), 508 pp.*

[15] Burnham, K., P., Andreson, D., R., (2002), Model selection and multimodel inference *Springer-Verlag, New York, pp. 488*

[16] Ceola, S., Bertuzzo, E., Mari, L., Botter, G., Battin, T., Gatto, M., Rinaldo, A.(2004), Light and hydrologic variability controls on stream biofilm dynamics in a flume experiment: a modelling approach.

[17] Doulatyari, B. Basso, S., Schirmer, M., Botter G., (2014), River flow regimes and vegetation dynamics along a river transect. *Adv. Water Resources, 73, 30-43*

[18] Egleson, P. S., (1978), Climate, soil and vegetation, 1, introduction to water valance dynamics *Water resources research, 14*

[19] Friedl, M., A., McIver, D., K., Hodges, J., C., F., Zhang, X., Y., Muchoney, D., Strahler, A., H. (2002), Global land cover mapping from MODIS: algorithms and early results *Remote sensing of environment, 83, 287-302*

[20] Gardiner, C., W., (2004), Handbook of stochastic methods *Springer, Berlin*

[21] Gentile, P., D'Odorico, P., Lintner, B., R., Sivandran, G., Savucci, G., (2012), Interdependence of climate, soil, and vegetation as constrained by the Budyko curve. *Geophysical research letters, 39*

[22] Hargreaves, G. H., & Samani, Z. A., (1985), Reference crop evapotranspiration from temperature. *Applied engineering in agriculture, 1(2), 96-99*

[23] Hargreaves, G., Allen, R., (2003), History and evaluation of hargreaves evapotranspiration equation *Journal of irrigation and drainage engineering, 129, 53-63*

[24] Langbein, W.B. (1949), Annual runoff in the United States. *Geol. Surv. Circ., 52 (1949) 14*

[25] LeRoy Poff, N., Olden, J. D., Merritt, D. M., Pepin, D. M. (2007), Homogeneization of regional river dynamics by dams and global biodiversity implications.

[26] L'vovich, M. I. (1979), *World water resources and their future, 415 pp*

[27] Lyne, V., D., Hollick, M., (1979), Stochastic time-variable rainfall runoff modelling *Hydrology and water resources symposium, Institution of engineers Australia, Perth, 89-92*

[28] Mantovani, A. (2014), Anthropogenic regulation and flow regime alteration at multiple timescales. *UniversitÃ degli studi di Padova - Dipartimento di ingegneria idraulica, marittima, ambientale e geotecnica.*

[29] , Milly, P., C., D., An analytic solution of the stochastic storage problem applicable to soil water (1993), *Water resources research, 29, 3755-3758*

[30] Milly, P.C.D. (1994), Climate, iterseasonal storage of soil water, and the annual water balance.*Adv. Water Resour., (1944)*

[31] Monteith, J., L., (1965), Evaporation and environment *Symposia of the society for experimental biology, 19, 205-224*

[32] Mu, Q., Heinsch, F., A., Zhao M., Running S., W., (2007), Development of a global evapotranspiration algorithm based on MODIS and global meteorology data *Remote sensing of environment, 111, 519-536*

[33] Mu, Q., Zhao M., Running S., W., (2011), Improvements to a MODIS global terrestrial evapotranspiration algorithm *Remote sensing of environment, 115, 1781-1800*

[34] Myneni, R., B., Hoffmann, S., Knyazikhin, Y., Privette, J., L., Glassy, J., Tian, Y., Wang, Y., Song, X., Zhang, Y., Smith, G. R., Lotsch, A., Friedl, M., Morisette, J. t., Votava, P., Nemani, R., R., Running, S., W., (2002), Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data, *Remote sensing of environment, 83, 214-231*

[35] Ponce, V., M., Shetty, A., V., (1995), A conceptual model of catchment water balance. 1. Formulation and calibration. *J. Hydrol. 173, 27-40*

[36] Porporato, A., Daly, E., Rodriguez-Iturbe, I., (2004), Soil water balance and ecosystem response to climate change *The american naturalist, 164*

[37] Puckridge, J. T., Sheldon, F., Walker, K. F., Boulton, A.J. (1998), Flow variability and the ecology of large rivers. *Mar. Freshwater Res.49,55-72*

[38] Ridolfi, L. D'Odorico, P., Laio, F. (2006), Effects of vegetation-water table feedbacks on te stability and resilience of plant ecosystems. *Water resources research 42: doi: 10.1029/2005WR004444. issn: 0043-1397, (2006)*

[39] Rodriguez-Iturbe, I., A., Ridolfi, L., Isham, V., Cox, D., R., (1999), Probabilistic modelling of water balance at a point: the role of climate, soil and vegetation. *Proceedings of the Royal Society of London, 455, 3786-3805*

[40] Sankarasubramanian, A., Vogel, R., M., (2004), Hydroclimatology of the continental United States *Geophysical research letters, 30*

[41] Sivapalan, M., Yaeger, M., A., Harman, C., J., Xu X., Troch, P., A., (2011), Functional model of water balance variability and the catchment scale: 1. Evidence of hydrologic similarity and space-time symmetry *Water resources research, 47*

[42] Thorntwaite, C.W., (1948), An approach toward a rational classification of climate. *Geographical Review 94(1948) 38-55*

[43] Trabucco, A., Daly, E., Rodriguez-Iturbe, I., (2004), Soil water balance and ecosystem response to climate change *The american naturalist, 164*

[44] Voepel, H., Ruddell B., Schumer, R., Troch, P.A., Brooks, P.D., Neal, A., Durick, M., Sivapalan, M. (2011), Quantifying the role of climate and land-

scape characteristics on hydrologic partitioning and vegetation response *Water resources research, 47*

[44] Wagner, W., (2007), Can we model the hydrological impacts of environmental change? *Hydrological processes, 21, 3233-3236*

[46] Wetzel. P., J., Chang, J., (1987), Evapotranspiration from nonuniform surfaces: a first approach for short-term numerical weather prediction*Monthly weather review, 116, 600-622*

[47] Wigmosta, M.S., Burges, S.J. (1997), An adaptative modeling and monitoring approach to describe the hydrologic behavior of small catchments. *Journal of hydrology 202(1977) 48-77*

[48] Zanardo, S. (2011), Catchment-scale transport phenomenta: rainfall intermittency, age of runoff, anthropic catchment management. *Universita' degli studi di Padova - Scuola di dottorato di ricerca in: scienze dell'ingegneria civile e ambientale.*

[49] Zomer, R., J., Trabucco, A., Bossio, D., A., Yuanjie, L., Gupta, D., C., Singh, V., P., (2007), Trees and water: smallholder agroforestry of irrigated lands in northern India *Colombo, Sri Lanka: International water management institute. pp. 45*