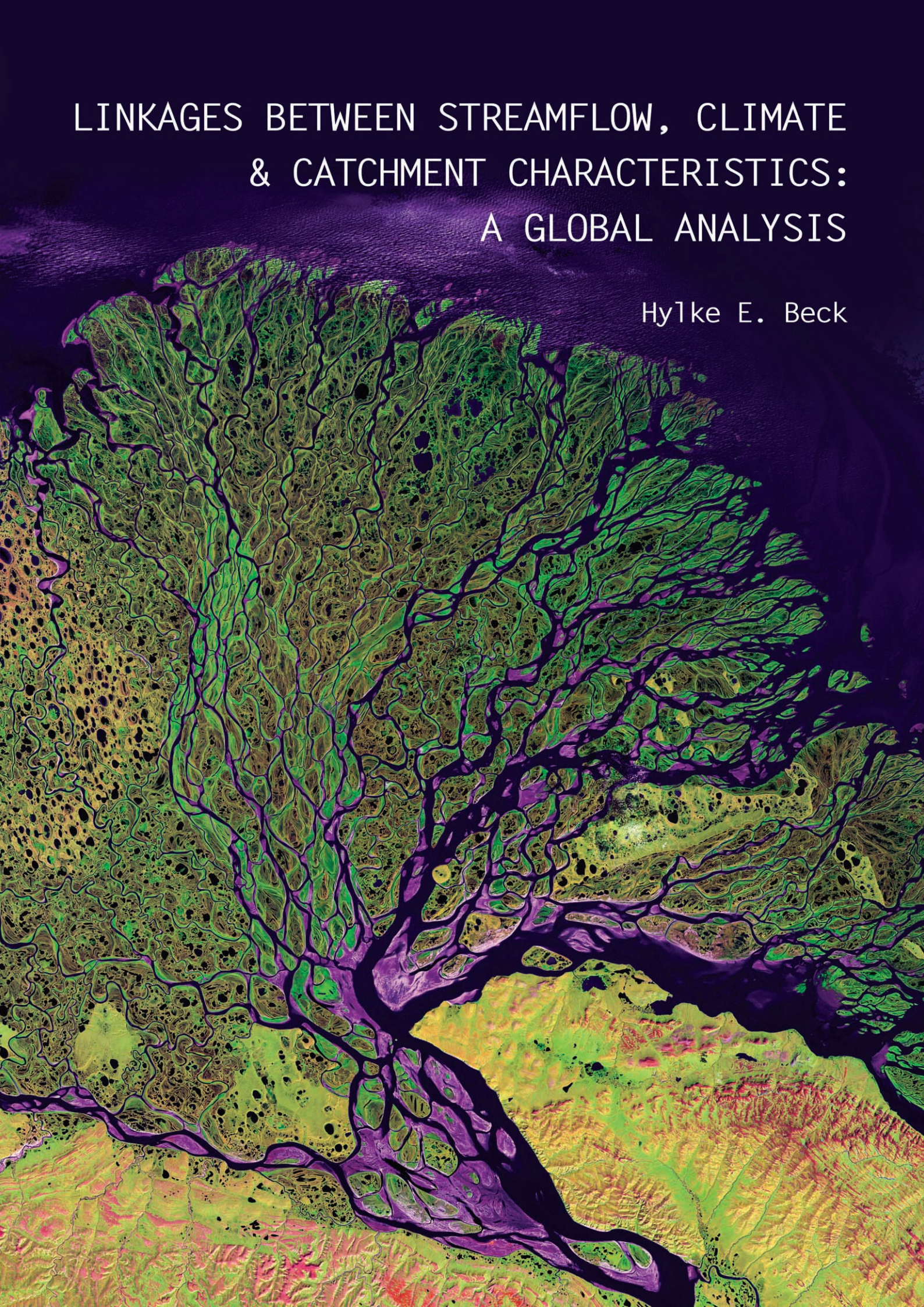
LINKAGES BETWEEN STREAMFLOW, CLIMATE & CATCHMENT CHARACTERISTICS: A GLOBAL ANALYSIS

Hylke E. Beck



Linkages between streamflow, climate, and catchment characteristics: a global analysis

Hylke E. Beck

Cover page: Image of the Lena River Delta acquired on 27 February, 2000, by the Enhanced Thematic Mapper plus (ETM+) sensor on-board the Landsat-7 satellite. Image provided by the USGS EROS Data Center Satellite Systems Branch as part of the Landsat Earth as Art Image series.

English title: Linkages between streamflow, climate, and catchment characteristics: a global analysis (PhD thesis, VU University Amsterdam)

Nederlandse titel: Verbanden tussen rivierafvoer, klimaat en karakteristieken van stroomgebieden: een globale analyse (Academisch proefschrift, Vrije Universiteit Amsterdam)

This study was funded by Deltares (Delft, the Netherlands) and carried out at VU University Amsterdam (the Netherlands).

Keywords: AVHRR; baseflow index; forest hydrology; hydrological model calibration; NDVI; neural networks; recession constant; remote sensing; soil moisture; streamflow; ungauged catchments.

VRIJE UNIVERSITEIT

Linkages between streamflow, climate, and catchment characteristics: a global analysis

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. F.A. van der Duyn Schouten,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de Faculteit der Aard- en Levenswetenschappen
op woensdag 16 oktober 2013 om 11.45 uur
in de aula van de universiteit,
De Boelelaan 1105

door

Hylke Edward Beck

geboren te Almere

promotor:	prof.dr. L.A. Bruijnzeel
copromotoren:	dr. J. Schellekens
	dr. R.A.M. de Jeu

reading committee: prof.dr. M.F.P. Bierkens
dr. W. Buytaert
prof.dr. A.J. Dolman
dr. M. Mulligan
prof.dr. S. Uhlenbrook

Contents

1 General introduction **1**
1.1 Background . . . 1
1.1.1 Importance of streamflow . . . 1
1.1.2 Influence of physiography and climate on streamflow . . . 1
1.1.3 The forest-streamflow relationship . . . 3
1.1.4 Streamflow estimation in ungauged catchments . . . 3
1.1.5 The role of satellite data . . . 4
1.2 Thesis objectives and outline . . . 5

2 Improving Curve Number based storm runoff estimates using soil moisture proxies **7**
2.1 Introduction . . . 7
2.2 Data and methods . . . 8
2.2.1 Catchment selection . . . 8
2.2.2 Streamflow and precipitation data . . . 8
2.2.3 AMSR-E soil moisture data . . . 9
2.2.4 TRMM precipitation data . . . 10
2.2.5 The Curve Number model . . . 10
2.2.6 Soil moisture proxies . . . 11
2.2.7 Approach . . . 12
2.3 Results and discussion . . . 12
2.4 Conclusions . . . 13

3 Global evaluation of four AVHRR-NDVI datasets: Intercomparison and assessment against Landsat imagery **19**
3.1 Introduction . . . 19
3.2 Datasets . . . 20
3.2.1 AVHRR-NDVI . . . 20
3.2.2 Landsat-NDVI . . . 22
3.2.3 MODIS-NDVI . . . 24
3.2.4 Land-cover classification map . . . 24
3.3 Methods . . . 24
3.3.1 Intercomparison of AVHRR-NDVI datasets . . . 24
3.3.2 Assessment against Landsat-NDVI . . . 25
3.4 Results . . . 26
3.4.1 Intercomparison of AVHRR-NDVI datasets . . . 26
3.4.2 Assessment against Landsat-NDVI . . . 28
3.5 Discussion . . . 32
3.5.1 Intercomparison of AVHRR-NDVI datasets . . . 32
3.5.2 Assessment against Landsat-NDVI . . . 35
3.6 Conclusion . . . 36

4 The impact of forest regeneration on streamflow in 12 meso-scale humid tropical catchments **39**
4.1 Introduction . . . 39
4.2 Study area . . . 40
4.3 Data . . . 41
4.3.1 Land cover . . . 41
4.3.2 Streamflow . . . 41

4.3.3 Precipitation . . . 41
4.3.4 Minimum and maximum air temperature . . . 45
4.4 Methodology . . . 45
4.4.1 Spatio-temporal interpolation and rescaling of climatic variables . . . 45
4.4.2 Potential evaporation . . . 46
4.4.3 HBV-light model . . . 46
4.4.4 Evaluation of the impacts of land-cover change on streamflows . . . 47
4.5 Results . . . 48
4.5.1 Changes in climatic variables . . . 48
4.5.2 HBV-light model performance . . . 48
4.5.3 Changes in land cover . . . 48
4.5.4 Changes in streamflow characteristics . . . 49
4.5.5 Impacts of land-cover change on streamflows . . . 49
4.6 Discussion . . . 50
4.6.1 Changes in climatic variables . . . 50
4.6.2 HBV-light model performance . . . 50
4.6.3 Impacts of land-cover change on streamflows . . . 50
4.6.4 Field-based estimates of vegetation water use . . . 53
4.6.5 Potential explanations for the lack of relationships . . . 55
4.7 Conclusion . . . 56

5 Global patterns in baseflow index and recession derived from 3520 small catchments 57
5.1 Introduction . . . 57
5.2 Data . . . 58
5.2.1 Observed streamflow . . . 58
5.2.2 Physiographic characteristics . . . 58
5.3 Methodology . . . 59
5.3.1 Computation of BFI and k . . . 59
5.3.2 Physiographic controls of BFI and k . . . 59
5.3.3 Catchment-scale estimation of BFI and k . . . 60
5.3.4 Global maps of BFI and k . . . 61
5.4 Results . . . 61
5.4.1 Physiographic controls of BFI and k . . . 61
5.4.2 Catchment-scale estimation of BFI and k . . . 61
5.4.3 Global maps of BFI and k . . . 62
5.5 Discussion . . . 63
5.5.1 Physiographic controls of BFI and k . . . 63
5.5.2 Catchment-scale estimation of BFI and k . . . 65
5.5.3 Global maps of BFI and k . . . 66
5.6 Conclusion . . . 66

6 Calibration of global hydrological models based on streamflow characteristics 75
6.1 Introduction . . . 75
6.2 Data . . . 76
6.2.1 Observed streamflow . . . 76
6.2.2 HBV-Light model inputs . . . 76
6.2.3 Noah-based streamflow predictions . . . 76
6.2.4 PCR-GLOBWB-based streamflow predictions . . . 77
6.3 Methodology . . . 77
6.3.1 Computation of streamflow characteristics . . . 77
6.3.2 Global maps of streamflow characteristics . . . 77
6.3.3 The HBV-Light model . . . 78
6.3.4 HBV-Light model configuration 1 . . . 78
6.3.5 HBV-Light model configuration 2 . . . 78
6.3.6 HBV-Light model configuration 3 . . . 79
6.3.7 Assessment of the HBV-Light model configurations . . . 79
6.4 Results . . . 80
6.4.1 Global maps of streamflow characteristics . . . 80

6.4.2 Assessment of the HBV-Light model configurations . . . 80
6.5 Discussion . . . 84
6.5.1 Global maps of streamflow characteristics . . . 84
6.5.2 Assessment of the HBV-Light model calibration configurations . . . 85
6.6 Conclusion . . . 86

7 Summary **87**
7.1 Conclusions . . . 87
7.2 Recommendations . . . 88

8 Samenvatting **91**
8.1 Conclusies . . . 91
8.2 Aanbevelingen . . . 93

Acknowledgements **95**

Chapter 1

General introduction

1.1 Background

1.1.1 Importance of streamflow

Streamflow is the spatial integration of runoff and is a major component of the annual catchment water balance. Streamflow is composed of a 'slow' component usually referred to as baseflow, which mainly originates from groundwater storage or other delayed sources (Hall, 1968; Smakhtin, 2001), and a 'quick' component called stormflow or quickflow (Jakeman and Hornberger, 1993). Stormflow can be generated as overland flow, when the soil is saturated (saturation overland flow; Dunne and Black, 1970) or when the rainfall intensity exceeds the soil's infiltration capacity (infiltration-excess overland flow; Horton, 1933), or as subsurface stormflow in the form of rapid lateral flow above an impervious horizon (or bedrock) when the groundwater level increases and subsurface saturated areas become connected (Tromp-van Meerveld and McDonnell, 2006) or via rapid pipeflow through soil pipes and other macropores at shallower depth under (near-)saturated conditions (Jones, 1981; Chappell, 2010). The relative importance of each of the mechanisms is mainly controlled by catchment attributes related to geology, soils, topography, climate, and land cover (e.g., Davis, 1969; Boorman et al., 1995; Zhang et al., 2001; Kirkby et al., 2002; Price, 2011), the level of antecedent soil water content (e.g., Meyles et al., 2003), and rainfall intensity (e.g., Bronstert and Bárdossy, 2003; Chappell et al., 2012).

Streamflow serves a number of essential purposes, such as irrigation, recreation, drinking water, industrial uses, and transport (e.g., Brauman et al., 2007; Quintero et al., 2009). In addition, it is used as a renewable source of energy (Cyr et al., 2011), currently accounting for ∼17 % of the global energy production[1]. Streams further play a key role in the maintenance and regulation of aquatic habitats (Poff et al., 1997). However, streamflow also carries some negative implications for society and environment. Since the turn of the century floods have caused each year on average ∼26 billion US dollars in damage, ∼6000 deaths, and affected ∼110 million people worldwide[2], with developing tropical countries generally the most severely affected (UNISDR, 2011). Furthermore, the transport of pollutants by streams (e.g., Jordan et al., 1997; Kirchner et al., 1999; Kolpin et al., 2004) can adversely affect ecohydrological systems downstream, e.g., in the form of eutrophication of estuaries or lakes with subsequent adverse effects on aquatic biodiversity (Anderson et al., 2002).

Changes in climate, caused by increases in atmospheric greenhouse gases, are seriously affecting the Earth's hydrological cycle (IPCC, 2007; Bates et al., 2008). Continued human-caused emission of greenhouse gases are projected to exacerbate these effects, while continued population growth will further increase both vulnerability to hydrologic disasters and pressure on freshwater resources (Huppert and Sparks, 2006; Vörösmarty et al., 2000). It is thus of great importance to improve our understanding of streamflow timing and quantity around the globe under current and future conditions. Perhaps the greatest obstacle to advancing current understanding is that much of the land surface is ungauged or poorly gauged (Fekete and Vörösmarty, 2007), as demonstrated by Fig. 1.1. Although streamflow estimation in ungauged regions poses a major challenge (Sivapalan et al., 2003), the increasing availability of multi-temporal satellite data provides unique opportunities to rise up to this challenge (Moradkhani, 2008; Van Dijk and Renzullo, 2011; Miralles et al., 2011). Arguably, tropical regions require priority attention in that these have some of the highest population densities and are thus particularly vulnerable to hydrologic extremes (cf. Bradshaw et al., 2007), while at the same time being increasingly poorly gauged in recent decades (cf. Wohl et al., 2012; Fig. 1.1).

1.1.2 Influence of physiography and climate on streamflow

There have been numerous studies examining the relationships between catchment attributes and streamflow characteristics using observations from meso- (1–10 000 km^2) and macro-scale (> 10 000 km^2) catchments (see Li et al. (2013) and Cheng et al. (2012) for recent ex-

[1] The World Bank (`http://water.worldbank.org/topics/hydropower`)

[2] Emergency Events Database (EM-DAT; `http://www.emdat.be`).

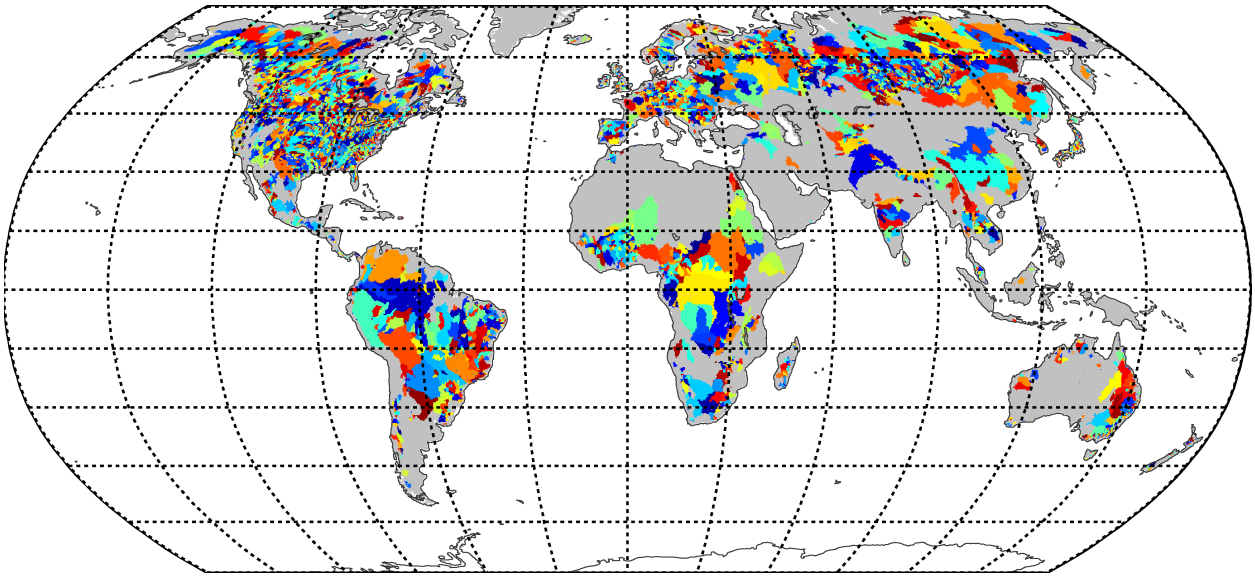


Figure 1.1: Global map showing interstation regions in random colors. Interstation regions are defined as the catchment of a gauging station excluding nested subcatchments of upstream gauges. Ungauged regions are indicated in gray. Ideally, the interstation regions are small, indicating a high station density. Streamflow and catchment boundary data originate from the Global Runoff Data Centre (GRDC; Koblenz, Germany; `http://grdc.bafg.de`), the Model Parameter Estimation Experiment (MOPEX; Schaake et al., 2006), and the Peel et al. (2000) dataset. The global maps in this thesis are presented in the Robinson projection (80°S–80°N and 180°W–180°E) with grid lines at every 15° latitude and 30° longitude.

amples). Streamflow characteristics often used in such studies include: mean annual streamflow; runoff coefficient, defined as the ratio of long-term mean streamflow to precipitation; baseflow index (BFI), defined as the ratio of long-term mean baseflow to total streamflow (Smakhtin, 2001); baseflow recession constant, defined as the rate of baseflow decay (Tallaksen, 1995); and various flow percentiles. Olden and Poff (2003) and Monk et al. (2007) provide a more exhaustive list of streamflow characteristics.

Climate encompasses the atmospheric water supply (precipitation) and demand (potential evaporation), and has long been recognized as the dominant control on the long-term water balance of catchments (Schreiber, 1904; Ol'dekop, 1911; Budyko, 1974). The most often used climate index is the so-called aridity index, defined as the ratio of long-term precipitation to potential evaporation (UNEP, 1992). The aridity index and the runoff coefficient are strongly positively related: the higher the aridity index of a catchment, the higher its runoff coefficient (e.g., Arora, 2002). The relationship between the aridity index and the runoff coefficient has been formalized in the well-known empirical Budyko (1974) equation. Subsequent studies have demonstrated that the runoff coefficient is also related to the phase difference between precipitation and potential evaporation seasonalities (Milly, 1994; Wolock and McCabe, 1999; Potter et al., 2005). Climate indices have further been linked to baseflow recession rates (Tschinkel, 1963; Czikowsky and Fitzjarrald, 2004; Peña-Arancibia et al., 2010; Van Dijk, 2010) and BFI (Mazvimavi et al., 2005; Van Dijk, 2010).

Hydrologic properties of soils and geology control the infiltration, storage, transmission, and release of water within a catchment, and are thus considered as primary controls on baseflow (Farvolden, 1963; Davis, 1969; Tague and Grant, 2004; Price, 2011). Accordingly, many observation-based studies have found good relationships between BFI and indices related to soils (Boorman et al., 1995; Santhi et al., 2008; Ahiablame et al., 2013) or geology (Lacey and Grayson, 1998; Longobardi and Villani, 2008; Bloomfield et al., 2009). These studies consistently indicated that the more permeable the subsurface materials of a catchment, the higher will be its BFI. Soil properties have also been found to exert some influence on the long-term water balance by inducing infiltration-excess runoff (Potter et al., 2005). In addition, in regions where the precipitation and potential evaporation seasonalities are strongly out of phase, the soil storage capacity can influence the long-term water balance by determining how much water can be stored during wet periods and subsequently evaporated during dry periods (Wolock and McCabe, 1999). Although there has been some progress in terms of the development of global-scale harmonized datasets of soils (Sanchez et al., 2009; FAO/IIASA, 2012) and geology (Gleeson et al., 2011; Hartmann and Moosdorf, 2012), there is still a major lack of consistent, detailed, hydrologically-relevant information with respect to soils and geology on a global scale. However, satellite remote sensing provides an as yet largely untapped opportunity to improve the spatial interpolation of soil-profile observations (cf. Mulder et al., 2011). Another important factor to consider is land degradation, the general term encompassing long-term reductions in ecosystem function and land productivity (FAO, 1979). Soil compaction due to heavy ma-

chinery or cattle trampling is a form of land degradation with severe hydrological implications (Batey, 2009), including reductions in dry-season baseflows and increases in the frequency and magnitude of stormflows and sedimentation (Bruijnzeel, 2004). Bai et al. (2008) have made an attempt to quantify land degradation worldwide using remotely sensed vegetation productivity. Although their approach has been heavily criticized (Wessels et al., 2012) it was considered the best global representation of surface degradation status in a recent study addressing the potential impacts of land degradation on storm runoff across the tropics (Peña-Arancibia, 2013).

Topography controls the hydraulic gradient and thus further mediates the streamflow response of a catchment, particularly in steep terrain. Based on theoretical considerations one would expect a negative relationship between mean surface slope and BFI or baseflow recession rate, as more steeply sloping aquifers are expected to drain faster (Brutsaert and Nieber, 1977). However, sensitivity experiments using the widely-used TOPMODEL (Beven and Kirkby, 1979) have demonstrated that this relationship is positive (Wolock et al., 1989), while observation-based studies have found it to be either negative, non-existent, or positive (Haberlandt et al., 2001; Mazvimavi et al., 2005; Zecharias and Brutsaert, 1988; Post and Jakeman, 1996; Brandes et al., 2005). Another often used topograhic index is the topographic wetness index (TWI) of Beven and Kirkby (1979). However, the use of TWI to locate contributing areas within catchments has generally been unsatisfactory (e.g., Jordan, 1994; Seibert et al., 1997; Buttle et al., 2001). On the whole, more work is necessary to improve our understanding of topography-streamflow linkages (cf. Price, 2011), and the near-global 90×90-m resolution digital elevation model derived from Shuttle Radar Topography Mission (SRTM) data (van Zyl, 2001; `http://srtm.csi.cgiar.org/`) should prove useful in this regard. However, the fact that surface topography is not necessarily representative of subsurface flow paths and hydraulic gradients constitutes an important confounding factor (e.g., Beven, 1997). Moreover, topographic indices tend to covary with climate (Peña-Arancibia et al., 2010) and soil properties (Price, 2011), thereby rendering it difficult to isolate causal variables.

Last but not least, it is widely recognized that the type of land use and land cover in a catchment affects streamflow totals and regime. Urbanization typically involves an increase in the impervious area of a catchment, which promotes overland flow and thus increases the peak flow frequency and magnitude (Leopold, 1968; Hollis, 1975; Lee and Heaney, 2003; Jacobson, 2011). Conversely, the presence of extensive lakes, wetlands, and/or reservoirs in a catchment tends to promote baseflow, resulting in a higher BFI (Neff et al., 2005; Ahiablame et al., 2013). The effects of agriculture on streamflow are typically mixed, and depend on management practices (notably irrigation, crop rotation, and degree of soil disturbance and/or surface degradation; cf. Price, 2011; Bruijnzeel, 2004). The forest-streamflow relationship deserves special attention and is discussed in the next section.

1.1.3 The forest-streamflow relationship

The relationship between forest cover and streamflow is a subject of ongoing and intense discussion (Andréassian, 2004; Bruijnzeel, 2004; FAO, 2005; Bradshaw et al., 2007; Van Dijk et al., 2009). Until the 1980s, the common perception was that forests act as giant sponges which soak up water during rainy periods and release it gradually to the streams during dry periods. Reviews of micro-scale (< 1 km^2) experimental catchments have since revealed that forests generally reduce amounts of streamflow due to their higher water consumption and high rainfall interception relative to grassland and agricultural crops (Bosch and Hewlett, 1982; Bruijnzeel, 1990; Sahin and Hall, 1996; Zhang et al., 2001; Brown et al., 2005; Jackson et al., 2005; Buytaert et al., 2006, 2007). However, at the scale of interest to water resource managers and planners (> 100 km^2) the relationship is far less clear (e.g., Buttle and Metcalfe, 2000; Wilk et al., 2001; Robinson et al., 2003; Zhou et al., 2010). The reasons for this may include any (combination of) the following: (1) uncertainties in the land-cover, precipitation, and/or streamflow data; (2) additional catchment climate characteristics overriding the effects of land-cover change; and (3) covariance between streamflow and catchment attributes (Van Dijk et al., 2012). Recent studies have also highlighted the important role played by soil surface conditions in mediating the hydrological impacts of land-cover changes (Bruijnzeel, 2004; Bonell et al., 2010; Ghimire et al., 2013). Post-forest soil disturbance and degradation associated with road building, urbanization and prolonged cropping or overgrazing without remedial measures, may well override the effects of changes in forest cover (and thus vegetation water use) on flows (cf. Bruijnzeel, 2004; Peña-Arancibia, 2013). In the tropics, forest regrowth on abandoned agricultural uplands is increasing in areas experiencing outmigration to urban areas (Wright, 2010) whereas other areas undergo rapid urbanization (Gupta, 2002). Yet, comparatively little is known about the hydrological implications of rapidly growing secondary forests (Bruijnzeel, 2004; Hölscher et al., 2005) or tropical urbanization (Chin, 2006).

1.1.4 Streamflow estimation in ungauged catchments

Several observation-based studies have established (multi-variate) models relating catchment attributes and streamflow characteristics (e.g., Mazvimavi et al., 2005; Brandes et al., 2005; Longobardi and Villani, 2008; Van Dijk, 2010; Peña-Arancibia et al., 2010; Krakauer and Temimi, 2011; Ahiablame et al., 2013). These models provide, in theory, a means to estimate streamflow characteristics in ungauged catchments. However, there are several reasons as to why the applicability of such mod-

els outside the region for which they were originally derived can be expected to be limited. First, the majority of models was typically based on a relatively small number of streamflow gauging stations (< 200), which can lead to less reliable models. Second, the models focused mostly on a particular region and used regional datasets to characterize geology or soils, thereby potentially restricting their larger-scale applicability. Third, most studies did not evaluate the generalization ability of the model using an independent set of catchments, and thus it is difficult to judge the true capability of the models. Finally, different studies reached conflicting conclusions regarding the importance of specific catchment attributes, notably mean surface slope and percentage forest cover.

Macro-scale hydrological models (land surface schemes and global hydrological models) simulate the water and energy balance of the land surface at macro-scales ($> 10\,000$ km^2) and are generally used to assess the hydrological implications of global climate change and human activities (Wood et al., 1997). They have a physically-based representation of the chief processes governing the hydrological cycle with *a priori* estimated parameter values, and are therefore expected to provide reasonably accurate streamflow estimates for ungauged regions. However, several studies have shown that streamflow estimates derived with macro-scale models are generally less accurate than streamflow estimates from hydrological models generally applied at the catchment scale (Duan et al., 2006; Nasonova et al., 2009), mainly due to a lack of calibration of the macro-scale models (Beven, 1989; Duan et al., 2001). Indeed, many present-day macro-scale models are essentially uncalibrated, including Noah-MP (Niu et al., 2011), Mac-PDM (Gosling and Arnell, 2011), the Community Land Model (CLM; Oleson et al., 2010), and PCRaster Global Water Balance (PCR-GLOBWB; Bierkens and van Beek, 2009; Van Beek and Bierkens, 2009). Others have been only crudely calibrated, such as the Variable Infiltration Capacity (VIC) model (Liang et al., 1994; Nijssen et al., 2001) and WASMOD-M (Widén-Nilsson et al., 2007), which all use nearest-neighbor interpolation of calibrated model parameters, and WaterGAP (Döll and Fiedler, 2008), which has been calibrated based on the runoff coefficients of gauged catchments. On the whole, there appears to be considerable room for improvement regarding the calibration of macro-scale hydrological models.

1.1.5 The role of satellite data

Satellite remote sensing offers researchers unparalleled opportunities to collect global-scale data (Melesse et al., 2007; Tang et al., 2009) that can be integrated in hydrological models to improve their streamflow simulation capability (Van Dijk and Renzullo, 2011; Moradkhani, 2008). Earth observing satellites differ in terms of their orbit and the sensors they carry (Kramer, 2002), with the sensor specifications determining which biophysical, climatic, or hydrological variables (and changes therein over time) can be calculated. In recent decades, substantial progress has been made with sensor and product algorithms for estimating such key variables as: precipitation (Stephens and Kummerow, 2007); surface soil moisture (De Jeu et al., 2008); vegetation productivity (Bannari et al., 1995); evapotranspiration (Li et al., 2009); surface topography (van Zyl, 2001); flood inundation area (Smith, 1997); land-cover type (Xie et al., 2008; Hansen and Loveland, 2012); surface energy fluxes (Liang, 2010); snow and ice cover (Dietz et al., 2011); continental-scale groundwater storage (using Gravity Recovery and Climate Experiment (GRACE) data; Ramillien et al., 2008); lake and river water level (Hall et al., 2011).

The usefulness of remotely-sensed precipitation for improving streamflow estimates is well recognized. Several regional studies have shown that using a Tropical Rainfall Measuring Mission (TRMM)-based multi-satellite precipitation product (Huffman et al., 2003) gives comparable results to using data from a sparse raingauge network in terms of streamflow prediction performance (e.g., Su et al., 2008, 2011; Collischonn et al., 2008; Yong et al., 2012). However, remotely-sensed precipitation products tend to be less reliable outside the tropics, in cold and/or mountainous regions, and at finer spatial and temporal scales (e.g., Hirpa et al., 2010; Ward et al., 2011; Peña-Arancibia et al., 2013).

Remotely-sensed surface soil moisture can potentially be used as an indicator of antecedent catchment wetness to improve streamflow estimates. Indeed, regional studies using surface soil-moisture products derived from the now defunct Advanced Microwave Scanning Radiometer for the Earth observing system (AMSR-E) or from the Advanced SCATterometer (ASCAT) have reported promising results (e.g., Crow and Ryu, 2009; Brocca et al., 2010; Koster et al., 2010; Draper et al., 2011). Some important challenges to overcome are the shallow penetration depth of the signal (1–2 cm), the large satellite footprint ($\sim$50 km), and the effects of vegetation, all of which introduce inaccuracies in the observations (De Jeu et al., 2008). In addition, AMSR-E suffers from radio frequency interference, a common problem in the USA, Europe, the Middle East, and Japan (Njoku et al., 2005).

The Normalized Difference Vegetation Index (NDVI; Tucker, 1979) is the most widely used remotely-sensed vegetation productivity index and its main use has been to study the vegetation response to climatic variability (e.g., Gong and Shi, 2003; Piao et al., 2006; Voepel et al., 2011; Jia et al., 2011). Other uses include estimating leaf area index (e.g., Wang et al., 2005; Myneni et al., 1997), estimating biomass and net primary productivity (e.g., Box et al., 1989; Ricotta et al., 1999), assessing long-term vegetation trends (e.g., Pouliot et al., 2009; Anyamba and Tucker, 2005), drought monitoring (Anyamba and Tucker, 2012), and mapping land cover and land use (and changes therein; e.g., Lunetta et al., 2006; Kleyn-

hans et al., 2011; Lu et al., 2003; Bai et al., 2008). On the other hand, there has been only a single study using remotely-sensed NDVI to investigate the vegetation influence on streamflow (Donohue et al., 2010). NDVI derived from the Advanced Very High Resolution Radiometer (AVHRR) is particularly interesting in this regard because of its long historical record (dating back to 1981), although the data are subject to large uncertainties due to a coarse spatial resolution and a lack of on-board calibration devices (Staylor, 1990).

1.2 Thesis objectives and outline

The general objectives of this thesis are to relate streamflow characteristics and catchment physiographic attributes over a wide range of catchment and climatic conditions, and to determine the value of specific satellite remote-sensing products for use in meso- and macroscale hydrological modeling. In order to fulfill these general objectives, five specific objectives have been defined, which are addressed in separate chapters. The specific objectives are to:

1. Assess the possible improvement in stormflow estimates when using soil moisture proxies based on TRMM precipitation, AMSR-E surface soil moisture, gauged precipitation, and observed baseflow for 186 Australian catchments (Chapter 2).
2. Globally evaluate four AVHRR-based NDVI datasets by conducting an intercomparison and by validating them against high-resolution NDVI imagery based on the Landsat-5 Thematic Mapper, which has on-board calibration devices (Chapter 3).
3. Analyze the possible impact of forest regeneration and urbanization on streamflow characteristics for a series of catchments on the island of Puerto Rico, one of the few humid tropical areas for which high quality data on rainfall, streamflow as well as land-cover change are available (Chapter 4).
4. Relate selected catchment physiographic attributes and two important baseflow characteristics (BFI and baseflow recession rate) using a global streamflow dataset consisting of 3520 catchments, and examine the feasability of producing global maps of these baseflow characteristics using an artificial neural network approach (Chapter 5).
5. Examine whether global maps of selected streamflow characteristics (mean annual streamflow, BFI, baseflow recession rate, and two flow percentiles) as derived using artificial neural networks and a global streamflow dataset can be used to calibrate a simple conceptual rainfall-runoff model (Chapter 6).

Finally, a summary of the present findings, chief conclusion reached, and suggestions for possible directions for future research are presented in Chapter 7.

To achieve these specific objectives, use is made of large observational datasets which have become available free of charge thanks to organizations such as the Global Runoff Data Centre (GRDC; Koblenz, Germany; http://grdc.bafg.de), the U.S. Geological Survey (USGS; `http://waterdata.usgs.gov/nwis`), and the United Nations Food and Agriculture Organization (FAO; `http://data.fao.org/`), rather than restricting oneself to case studies in comparatively small homogeneous regions. The aim is to move away from the reporting of the idiosyncrasies of single sites or catchments which has dominated so much of the older hydrological literature, and instead attempt to identify robust, generalizable relationships applicable to different environments worldwide (cf. Jones, 2005; Andréassian et al., 2007; Gupta et al., 2013).

Chapter 2

Improving Curve Number based storm runoff estimates using soil moisture proxies[1]

Abstract. Advances in data dissemination and the availability of new remote sensing datasets present both opportunities and challenges for hydrologists in improving flood forecasting systems. The current study investigates the improvement in SCS Curve Number (CN)-based storm runoff estimates obtained after inclusion of various soil moisture proxies based on additional data on precipitation, baseflow, and soil moisture. A dataset (1980–2007) comprising 186 Australian catchments (ranging from 51 to 1979 km^2 in size) was used. In order to investigate the value of a particular proxy, the observed S was compared to values obtained with different soil moisture proxies using linear regression. An antecedent precipitation index (API) based on gauged precipitation using a decay parameter proved most valuable in improving storm runoff estimates, stressing the importance of high quality precipitation data. An antecedent baseflow index (ABFI) also performed well. Proxies based on remote sensing (TRMM and AMSR-E) gave promising results, particularly when considering the expected arrival of higher accuracy data from upcoming satellites. The 5-day API performed poorly. The inclusion of soil moisture proxies resulted in mean modeled vs. observed correlation coefficients around 0.75 for almost all proxies. The greatest improvement in runoff estimates was observed in drier catchments with low Enhanced Vegetation Index (EVI) and topographical slope (all intercorrelated parameters). The present results suggest the usefulness of incorporating remotely sensed proxies for soil moisture and catchment wetness in flood forecasting systems.

[1]This chapter is an edited version of: Beck, H. E.; de Jeu, R. A. M.; Schellekens, J.; Van Dijk, A. I. J. M., and Bruijnzeel, L. A. Improving curve number based storm runoff estimates using soil moisture proxies. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2(4):250–259, 2009.

2.1 Introduction

Floods are among the most costly and frequent natural disasters in terms of people's suffering and economic losses. In fact, flooding is considered to be one of the prime catastrophic events threatening society in many countries (NRC, 1996). Under global environmental changes as a consequence of land use modifications and increased greenhouse gas emissions the flooding risk may increase (Betts et al., 2007). Flood forecasting systems provide a tool that allows decision makers (and in some cases the general public) to be proactive rather than reactive to flood hazard events. Recent developments in remote sensing and data dissemination present both opportunities and challenges for hydrologists in improving currently used flood forecasting systems since remote sensing offers a means to provide frequent global coverage of such critical hydrological data as precipitation and soil moisture (Entekhabi et al., 1999). The anticipated availability of high-accuracy data at finer spatial and temporal resolutions from ESA's Soil Moisture and Ocean Salinity (SMOS) mission in 2009 (Kerr et al., 2001), NASA's Soil Moisture Active/Passive (SMAP) mission in 2012 (NRC, 2007b), and NASA's Global Precipitation Measurement (GPM) mission in 2013 (Smith et al., 2006) presents another incentive to investigate the application of currently available satellite data.

Most nations have organized and operate special *in-situ* networks devoted to measuring river discharge and precipitation. These networks are increasingly made available to the public via internet. Examples are the National Climatic Data Center (NCDC) for precipitation in the US (see `http://www.ncdc.noaa.gov`) and the Queensland Natural Resources and Mines (NR&M) Data Drill for precipitation in Australia (see `http://www.longpaddock.qld.gov.au/silo/`). Also increasingly available are datasets incorporating multiple networks, such as the Global Precipitation Climatology Centre (GPCC) database (Schneider et al., 2008), the Global Precipitation Climatology Project (GPCP) (Adler et al.,

2003), the Global River Discharge Database (GRDD) (Vörösmarty et al., 1996), and the Global Soil Moisture Data Bank (GSMDB) (Robock et al., 2000). Whilst many of these datasets have their limitations—including the use of many different data-formats, low spatial and/or temporal resolution, or a high cost—improving their accessibility would present major opportunities for hydrologists involved in flood forecasting. The Atmospheric Data Access for the Geospatial User Community (ADAGUC) web portal of the Royal Meteorological Institute of The Netherlands (see `http://adaguc.knmi.nl`) has been created specifically to address these issues and to provide easy access to various datasets in a standardized and user-friendly format. It hosts a series of remote sensing products including AMSR-E soil moisture which is used in the current study.

At the base of many flood forecasting exercises lies the need for a robust model to convert input precipitation to storm runoff from a catchment. The Natural Resources Conservation Service (NRCS, formerly the Soil Conservation Service, SCS) Curve Number (CN) method (USDA, 1986) is widely used to estimate runoff from rainfall amounts. The attraction of the CN model lies in its simplicity (it requires the estimation of a single parameter only, the Curve Number) and in its numerous applications since the 1980s (Choi et al., 2002; Arnold and Fohrer, 2005; Kim and Lee, 2008). Because soil moisture is a key factor determining the partitioning of rainfall into runoff and infiltration (Aubert et al., 2003), soil moisture proxies are calculated to account for a catchment's wetness status prior to the rainfall event in the hope to improve stormflow prediction (USDA, 1986). There is some controversy as to the degree of improvement in storm runoff estimates that may be obtained when using additional (satellite-based) information on soil moisture. Some studies that used only a small number of catchments reported strong improvements in runoff estimates after the addition of remotely sensed soil moisture data (Jacobs et al., 2003; Brocca et al., 2008). Others found an improvement only for a subset of their catchments (Parajka et al., 2005b; Pauwels et al., 2002). This paper investigates the degree in improvement in storm runoff prediction obtained with soil moisture proxies derived from gauged precipitation, gauged stream discharge, AMSR-E soil moisture, and TRMM precipitation. Data from a large number (186) of Australian catchments is used, representing a range of precipitation, streamflow and soil moisture conditions. Through analysis of such a large number of catchments some understanding may be gained as to under what conditions the additional information on soil moisture will improve the runoff estimates.

2.2 Data and methods

2.2.1 Catchment selection

A large dataset (1980–2007) of 186 Australian catchments having daily streamflow and precipitation observations provided by CSIRO Land and Water was used in the present analysis. The catchments range in size from 51–1979 km^2 (mean size 447 km^2). Most catchments are located in the southeastern part of Australia (see Fig. 2.1a). This region is characterized by a temperate oceanic climate with a wet winter and low summer rainfall, whereas the central part of Australia is arid to semi-arid; the northern part has dry winters and relatively wet summers (see `http://www.bom.gov.au/climate/`). Most catchments are water-limited (80% of the catchments has ratio of long-term precipitation to potential evaporation < 1), are dominated by extensive agriculture or native vegetation (mean vegetation cover $35\% \pm 23$), and have for the main part low-relief (only 38% of the catchments has an average slope $> 10°$). Catchments where streamflow was subject to regulation or diversion were not included in the dataset.

2.2.2 Streamflow and precipitation data

Time series of daily precipitation were based on a gridded dataset (see `http://www.bom.gov.au/silo/`) with a spatial resolution of 0.05° based on interpolation over approximately 7200 meteorological stations (see Fig. 2.1b) (Jeffrey et al., 2001). Mean catchment precipitation was calculated from the area-weighted mean precipitation of the cells covered by the catchment.

Time series of streamflow were collected as part of previous studies (Guerschman et al., 2008; Peel et al., 2000). The baseflow component of the discharge is estimated using the Eckhardt recursive filter (Eckhardt, 2008), which low-pass filters the hydrograph according to:

$$Q_{\mathrm{bf}}(t) = \frac{(1-\mathrm{BFI}_{\max})\alpha Q_{\mathrm{bf}}(t-1) + (1-\alpha)\mathrm{BFI}_{\max} Q(t)}{1-\alpha \mathrm{BFI}_{\max}}, \tag{2.1}$$

subject to $Q_{\mathrm{bf}}(t) \leq Q(t)$, where Q_{bf} is the baseflow [m^3 s^{-1}], Q is the total discharge [m^3 s^{-1}], $\mathrm{BFI}_{\max}$ is the maximum value of the baseflow index that can be modeled by the algorithm, α is the baseflow recession constant, and t is time [days]. The α parameter is derived from recession analysis by plotting $Q(t_i)$ vs. $Q(t_i+1)$ for discharge values $Q(t_i)$ that are part of a recession period of at least five days, identified by:

$$Q(t_i-2) > Q(t_i-1) > Q(t_i) > Q(t_i+1) > Q(t_i+2). \tag{2.2}$$

Assuming that the aquifer is a linear reservoir, theoretically all points should follow a straight line through the origin with slope equal to the recession parameter α. In practice, however, there is some scatter in the points. It

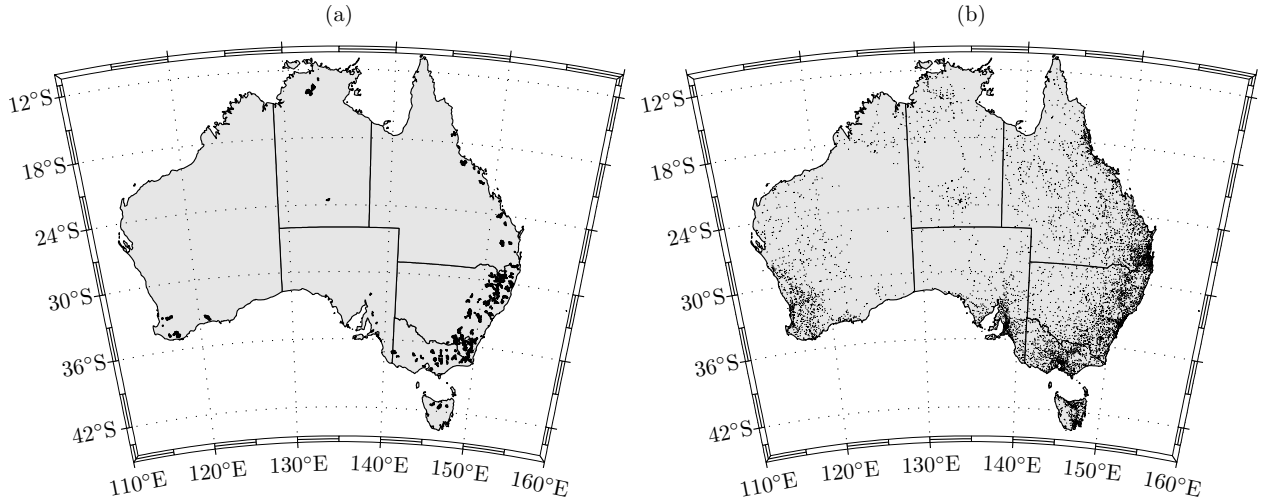


Figure 2.1: Map of Australia showing (a) the geographical distribution of the 186 catchments in the dataset, and (b) the locations of the precipitation gauges used on January 5, 2005. The actual gauges used vary from day to day depending on which are available, but will mostly be similar to this set.

is more probable that the points representing the slower recession represent the true value of α, because (1) the discharge at t_i could still contain direct runoff, and (2) there is probably more groundwater recharge at t_i. In reference (Eckhardt, 2008) the points representing the slowest recession (i.e., the upper limit of the $Q(t_i)$ vs. $Q(t_i + 1)$ plot) are on straight line with slope assumed to be α. However, in the present study the points representing the slowest recession often approach one. To solve this problem common logarithms of both values are taken and α is determined by fitting a model of the form:

$$\log_{10}(Q(t_i)) = \alpha \log_{10}(Q(t_i + 1)) \tag{2.3}$$

through the points exceeding the 90^{th} percentile recession. Since most streams in southern Australia are ephemeral with porous aquifers, following recommendations in (Eckhardt, 2008) BFI_{max} was set to 0.5 for all catchments. By using $\text{BFI}_{\text{max}} = 0.5$ Eq. 2.1 corresponds to the algorithm of Chapman and Maxwell (Chapman and Maxwell, 1996). Fig. 2.2 gives an example of baseflow separated using the described methodology for a single catchment (gauge code 410061) with α determined at 0.9823. The direct component of discharge (Q_{qf}) is calculated from $Q - Q_{\text{bf}}$.

Streamflow data represent totals based on the 24 hours prior to midnight local time, whereas the precipitation data are based on the 24 hours prior to 0900 local time. This 9 hour mismatch between precipitation and runoff days might result in a mismatch between P events and their resulting Q_{qf} in $9/24 = 38\%$ of the events (ignoring the delayed response of Q_{qf}). To reduce the effect of this mismatch the temporal resolution of P and Q_{qf} was resampled to 2 days (at 2 day intervals t_i the mean of $t_i - 1$ and t_i was calculated). This resulted in a decrease in the occurrence of mismatches between P and corresponding Q_{qf} to $9/48 = 19\%$ of the events. Catchments with less than 10 events with $P > 15$ mm after June 2002 (the launch of AMSR-E) were not included in the analysis.

2.2.3 AMSR-E soil moisture data

In this paper Advanced Microwave Scanning Radiometer Earth Observing System (AMSR-E)-based surface soil moisture is used to calculate a soil wetness index (SWI). AMSR-E Level-2A swaths from June 2002 (the launch of the AMSR-E mission) to present were used. Soil moisture values were derived using the Land Parameter Retrieval Model (LPRM) (Owe et al., 2008) from C-band (6.92 GHz) passive microwave brightness temperatures. AMSR-E soil moisture represents the soil moisture in the top 1–2 cm of the soil. Daily (24 hours prior to midnight) mean soil moisture was calculated from the average of all satellite observations in a swath with footprint centers inside the catchment boundaries during that day. The actual number of observations in a swath that were used thus depended on catchment size. On average, AMSR-E passes a location in Australia slightly more often than once per day. Both ascending (daytime, ±13:30 solar time) and descending (nighttime, ±1:30 solar time) passes of AMSR-E were used. Observations where the LPRM residual exceeded zero were excluded.

A validation study of the AMSR-E soil moisture product used in the present study was performed for Australia (Draper et al., 2007). A number of validation sites were in close proximity to the catchments used in the present study (although none were in densely vegetated regions). For these sites the normalized 5-day moving average of AMSR-E C-band ascending and descending soil moisture showed strong correlations with *in-situ* soil moisture time series at 0–7 cm depth (coefficient of determination $r^2 = 0.56$ to 0.87, RMSE = 0.016 to 0.066 $\text{m}^3\ \text{m}^{-3}$ volumetric soil moisture). In general there was good spatial agreement with antecedent precipitation. Radio frequency interference (RFI), a common problem in many areas in Europe and the US, is almost absent in Australia (Njoku et al., 2005).

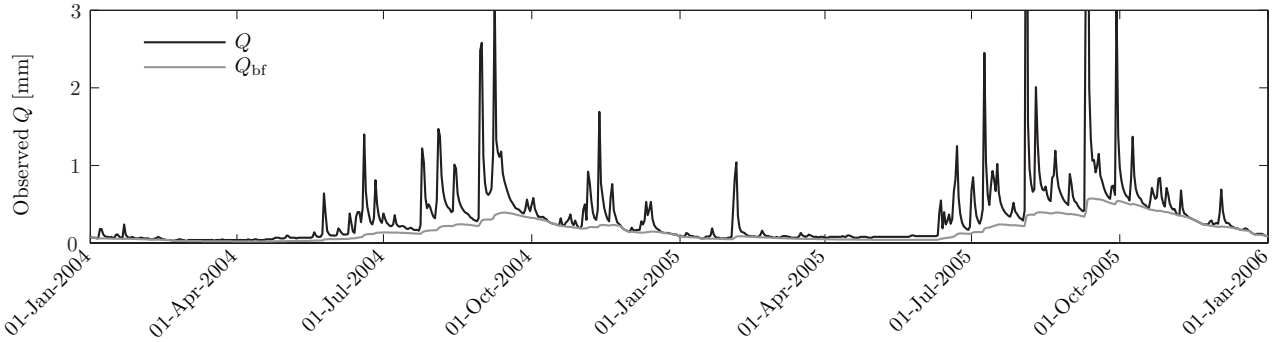


Figure 2.2: Baseflow separation results for catchment 410061 (gauge code). Baseflow was separated from total streamflow using the Eckhardt filter.

2.2.4 TRMM precipitation data

The current study uses the Tropical Rainfall Measuring Mission (TRMM) 3B42RT product, which is available at a lag of about 6 hours, and is a combination of TRMM real-time merged passive microwave (HQ, 3B40RT) and microwave-calibrated infrared (VAR, 3B41RT) (Huffman et al., 2003). For each pixel the HQ value is used if available, and otherwise the VAR value is used. The spatial resolution is 0.25°, the domain with useful data is 45°N–45°S, and the temporal resolution is 3 hours. Daily (24 hours prior to midnight local time) time series for each catchment were calculated from the area-weighted mean of all cells within the catchment. Compared to the TRMM 3B42 (research) product, the 3B42RT product does not include gauged precipitation data.

2.2.5 The Curve Number model

The Curve Number (CN) method is used to predict the amount of direct runoff resulting from large rainfall events in a particular area (USDA, 1986). The method is based on the following relationship:

$$\frac{F}{S} = \frac{Q_{\mathrm{qf}}}{P - I_a}, \tag{2.4}$$

where F is the actual retention, S the potential maximum retention, Q_{qf} the direct runoff, P the precipitation, and I_a the initial abstraction of rainfall by soil and vegetation (all in mm). The following continuity equation is introduced:

$$F = P - I_a - Q_{\mathrm{qf}} \tag{2.5}$$

Combining Eqs. 2.4 and 2.5 eliminates F, yielding:

$$Q_{\mathrm{qf}} = \frac{(P - I_a)^2}{P - I_a + S} \quad \text{for} \quad P > I_a. \tag{2.6}$$

I_a is related to S according to:

$$I_a = \lambda S, \tag{2.7}$$

where λ is the initial abstraction coefficient.

The standard value for λ is 0.2 (Ponce and Hawkins, 1996). However, several more recent studies have shown that the assumption of 0.2 is unusually high, and that values between 0.01 and 0.05 are more realistic. A study using rainfall and runoff data from 307 US catchments or plots found that a value of λ of 0.05 would fit the data much better (Woodward et al., 2003). In an experimental watershed in the Three Gorges Area of China it was found that λ varied from 0.010 to 0.154 with a median of 0.048 (Shi et al., 2009). In an experimental catchment in Attica, Greece, the average value of λ was equal to 0.014, and in a subcatchment equal to 0.037 (Baltas et al., 2007). In yet another study using data from 237 US catchments optimal model performance was found with a value of λ in the order of 0.01 (Mishra et al., 2004). In the present study a fixed value of 0.05 was used for all catchments.

To directly calculate S for a given rainfall and runoff amount the following equation may be used (which is a combination of Eqs. 2.6 and 2.7, rewritten to solve for S):

$$\begin{aligned} S = 0.5\lambda^{-2}[&Q_{\mathrm{qf}} - \lambda Q_{\mathrm{qf}} + 2\lambda P \\ &- \sqrt{Q_{\mathrm{qf}}^2 - 2\lambda Q_{\mathrm{qf}}^2 + \lambda^2 Q_{\mathrm{qf}}^2 + 4\lambda P Q_{\mathrm{qf}}}]. \end{aligned} \tag{2.8}$$

Fig. 2.4b gives an example of S values calculated using Eq. 2.8 ($\lambda = 0.05$) from observed rainfall and direct runoff data shown in Fig. 2.4a.

Because runoff usually varies widely for the same rainfall amount (mainly due to antecedent soil moisture conditions, but also due to spatial variability of rainfall, variability in intensity, measurement errors, and other factors) the observed S varies between storms. The S also appears to be biased towards high values at low storm depths (Hawkins, 1993) because there is not enough runoff. This effect is more evident at higher values of λ because there has to be more rainfall for runoff to occur. To correct this bias a model is fitted with the following equation (from (Hawkins, 1993), rewritten for S):

$$S_{\mathrm{fit}} = S_{\mathrm{inf}} - S_{\mathrm{inf}} \exp(-vP), \tag{2.9}$$

where S_{fit} is the fitted model and P again the amount of precipitation [mm]. S_{inf} (the asymptotic S [mm]) and v are fitting parameters. The S values with the bias

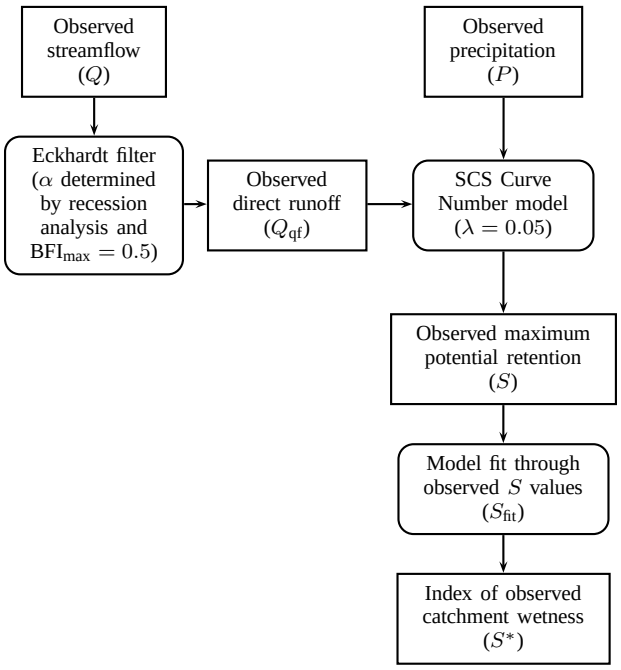


Figure 2.3: Flow-chart showing the derivation of S^* for a single catchment.

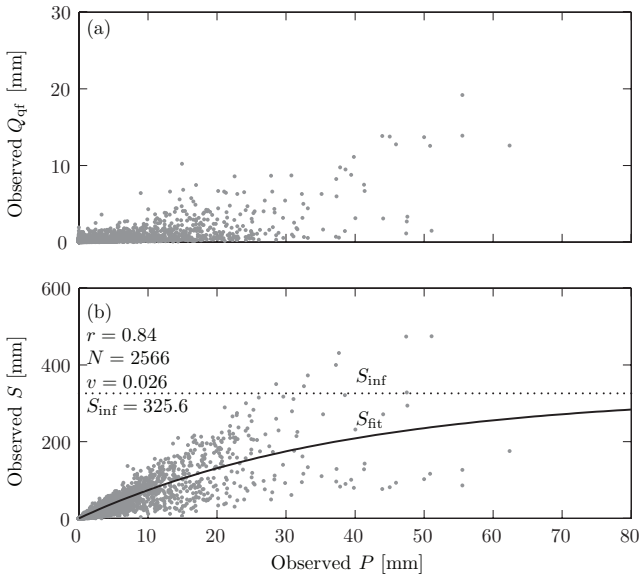


Figure 2.4: Plots for a single catchment (gauge code 410061) of (a) observed precipitation vs. observed direct runoff, and (b) observed precipitation vs. observed S as calculated from observed precipitation and direct runoff using Eq. 2.8. Also shown is the fitted model (S_{fit}) and the asymptote of the fit (S_{inf}). The correlation coefficient r of the fitted model and the number of observations N are also listed. Data from 1980–2007 have been used.

removed are calculated as follows:

$$S^* = S - S_{fit}, \tag{2.10}$$

where S^* may be interpreted as an index of mean catchment wetness or saturated area within a catchment. Ignoring other factors, a low S^* value should indicate a wet catchment with a low infiltration rate and a reduced potential for storage, whereas a high S^* value should indicate a drier catchment where a greater portion of the rainfall infiltrates. Fig. 2.3 summarizes the steps taken in calculating S^* values for a catchment.

2.2.6 Soil moisture proxies

Soil moisture proxies are used to estimate the catchment wetness prior to a rainfall event and adjust Curve Number estimates of runoff accordingly. This section presents the five soil moisture proxies used in this paper (SWI, API_k, $API_{k\ TRMM}$, $API_{5\,day}$, and ABFI). The soil moisture proxies are calculated on a daily basis from June 2002 (the launch of the AMSR-E mission) until 2007 (marking the end of the precipitation and discharge time series).

Soil Wetness Index (SWI)

An exponential moving average filter (Wagner et al., 1999) is used to relate AMSR-E based surface soil moisture to the part of the soil that influences runoff production. The approach is based on the idea that surface soil moisture derived from AMSR-E data fluctuates at a higher frequency (since it pertains to the top few cm only) than the total runoff-generating part of the soil. The Soil Water Index (SWI [%]) is calculated according to:

$$SWI(t) = \frac{\sum_i \theta_{AMSR\text{-}E}(t_i) \exp\left(-\frac{t-t_i}{T}\right)}{\sum_i \exp\left(-\frac{t-t_i}{T}\right)}, \tag{2.11}$$

where $\theta_{AMSR\text{-}E}$ is the AMSR-E-based soil moisture content [% volumetric content] and T is a time lag constant [days]. In the present study T was optimized for each catchment seperately to give the highest linear correlation coefficient r between antecedent SWI and S^*.

Gauge Antecedent Precipitation Index (API_k)

An antecedent precipitation index (API) is often used because it uses readily available precipitation data (Mishra et al., 2004). API_k [mm] assumes a temporal decay constant k to account for soil moisture losses. The following recursive model was used (Linsley et al., 1949):

$$API_k(t) = k\, API_k(t-1) + P(t), \tag{2.12}$$

where P is the gauged precipitation [mm], and k is a decay parameter (< 1) that is inversely related to drainage and evaporation losses and positively related to penetration depth [-]. In literature k is reported to vary between 0.80 and 0.98 (Brocca et al., 2008). In the present study k was optimized separately for each catchment based on the highest linear correlation coefficient r between antecedent API_k and S^*.

TRMM Antecedent Precipitation Index ($\mathbf{API}_{k\ \mathrm{TRMM}}$)

To investigate the performance of a soil moisture proxy based on near real-time satellite precipitation an API is calculated using Eq. 2.12 with TRMM precipitation estimates for P and k equal to the value optimized using gauged precipitation data.

5-day Antecedent Precipitation Index ($\mathbf{API}_{5\,\mathrm{day}}$)

Traditionally the amount of rainfall received in the five days preceding a storm event of interest is used (Chow et al., 1988). This 5-day API is calculated according to:

$$\mathrm{API}_{5\ \mathrm{day}}(t) = \sum_{i=0}^{4} P(t_i), \tag{2.13}$$

where i denotes the ith day before day t.

Antecedent Baseflow Index (ABFI)

In literature an exponential relation between baseflow and groundwater storage has been proposed (Sivapalan et al., 1987). ABFI [-] is therefore calculated from the natural logarithm of the baseflow:

$$\mathrm{ABFI}(t) = \log_e\left(Q_{\mathrm{bf}}(t)\right), \tag{2.14}$$

where Q_{bf} is the baseflow [m^3 s^{-1}] and t is time [days]. The advantage to this method is that it eliminates the need to select additional parameters. However, separation of baseflow from total flow is required. See section 2.2.2 for a description of the separation approach used here.

2.2.7 Approach

For each catchment, observed values of S^* (derived from observed rainfall and direct runoff) for events with $P > 15$ mm were compared to antecedent values of the respective soil moisture proxies. S^* values for day t_i were compared to values of $\mathrm{API}_{k\ \mathrm{TRMM}}$, ABFI, and SWI on day $t_i - 2$ (due to the 2-day resampling of precipitation and direct runoff) and compared to values of API_k and $\mathrm{API}_{5\,\mathrm{day}}$ on day $t_i - 3$ (due to the 2-day resampling and the 9 hour mismatch between measurements of rainfall and streamflow). For each antecedent soil moisture proxy (ASMP) a linear regression is performed of the form:

$$S^* = a\mathrm{ASMP} + b, \tag{2.15}$$

where a and b are fit parameters. The quality of the correlation was taken to indicate the value of the soil moisture proxy for improving runoff estimates. An independent evaluation of the model was also performed using the same catchments but setting T equal to 5 and k equal to 0.971 for all catchments.

The improvement in model performance after including a soil moisture proxy in the CN model was assessed as follows. First, the correlation coefficient of modeled vs. observed Q_{qf} was calculated for the CN model without the inclusion of the proxy by using Eq. 2.6 with the S values equal to S_{fit}. Secondly, to quantify the improvement, the correlation coefficient of modeled vs. observed Q_{qf} was calculated again but now for the CN model with the proxy included by using Eq. 2.6 with S calculated using Eqs. 2.9, 2.10, and 2.15.

2.3 Results and discussion

In the current study for 186 Australian catchments (51–1979 km^2) the performance of five soil moisture proxies in improving runoff estimates is investigated. Table 2.1 shows for each soil moisture proxy (1) the number of catchments having a significant correlation (probability $p < 0.01$) between antecedent values of the soil moisture proxy and S^* (catchment wetness according to observed precipitation and runoff), (2) the mean correlation coefficient r, and (3) the obtained improvement in runoff estimates.

The API_k soil moisture proxy (with k optimized) performed best, with a significant correlation obtained for 77% of the catchments. The ABFI and SWI (with T optimized) proxies performed reasonably well, with significant correlations in 47% and 35% of the catchments, respectively. The $\mathrm{API}_{5\,\mathrm{day}}$ and $\mathrm{API}_{k\ \mathrm{TRMM}}$ proxies exhibited poorer performance with significant correlations in 22% and 23% of the catchments, respectively. The poor performance of $\mathrm{API}_{k\ \mathrm{TRMM}}$ may be due to the small size of the catchments, which on average cover less than a single TRMM cell. Better results can be expected by using the TRMM 3B42 research product, which includes gauge data, but these are not available at near real-time for operational systems. The good performance of API_k stresses the importance of high-quality precipitation data. All soil moisture proxies resulted in large improvements in modeled vs. observed runoff after including the proxy in the model, with correlation coefficients around 0.75 obtained for almost all proxies. Table 2.1 also lists results for API_k ($k = 0.971$), $\mathrm{API}_{k\ \mathrm{TRMM}}$ ($k = 0.971$), and SWI ($T = 5$ days), with slightly poorer results compared to the proxies with optimized parameters. Figs. 2.6, 2.7, and 2.8 show detailed results for a single catchment (gauge code 410061). In this catchment SWI performed particularly well, demonstrating the high potential of SWI as an index of overall catchment wetness status. Our results are similar to those of Parajka et al. (2005b) and Pauwels et al. (2002), who found inclusion of remotely sensed soil moisture improved the model in only a subset of the catchments, but also to Brocca et al. (2008) and Jacobs et al. (2003) who investigated a small number of catchments and found large improvements after the addition of remotely sensed soil moisture. The performance of the various proxies is not interchangeable (when for example ABFI performs well it is likely that API_k also performs well).

Table 2.1: Summary of the results showing the number of catchments with a significant correlation (probability $p < 0.01$) between the soil moisture proxy and S^* (1), the mean correlation coefficient r between the proxy and S^* (3), and the improvement in runoff estimates after including the proxy into the model (4). The standard deviation of r is also given.

	Number of catchments		Mean Q_{qf} observed vs. modeled r	
Soil moisture proxy	with $p < 0.01$	Mean r	Standard SCS-CN	SCS-CN with proxy
$API_{5\,day}$	41 (22%)	−0.62 (±0.12)	0.68 (±0.26)	0.80 (±0.17)
API_k (k optimized for each catchment)	143 (77%)	−0.70 (±0.11)	0.57 (±0.29)	0.75 (±0.18)
API_k ($k = 0.971$)	110 (59%)	−0.68 (±0.11)	0.57 (±0.28)	0.73 (±0.17)
$API_{k\ TRMM}$ (k equal to API_k optimized k)	42 (23%)	−0.62 (±0.11)	0.69 (±0.23)	0.79 (±0.17)
$API_{k\ TRMM}$ ($k = 0.971$)	37 (20%)	−0.60 (±0.11)	0.70 (±0.20)	0.80 (±0.15)
ABFI	88 (47%)	−0.69 (±0.12)	0.55 (±0.30)	0.76 (±0.16)
SWI (T optimized for each catchment)	65 (35%)	−0.71 (±0.12)	0.47 (±0.30)	0.72 (±0.16)
SWI ($T = 5$ days)	47 (25%)	−0.68 (±0.12)	0.43 (±0.30)	0.67 (±0.16)

The SWI time lag constant T and the API_k decay parameter k were optimized for each catchment individually. Fig. 2.5 shows the optimization results for SWI and API_k. Only the catchments with a significant correlation between S^* and antecedent values of the proxy are shown. The value of T resulting in the best correlations was found to be 5 days, implying that 63% of the SWI values are based on data less than 5 days old. Higher values of T were found in some other studies (Brocca et al., 2008; Wagner et al., 1999). The optimum value of k was 0.971, implying that 63% of the API_k values are based on rainfall less than 41 days old. A likely explanation for the difference in lag between SWI and API_k is that remotely sensed soil moisture already contains information on past events.

An attempt was made to relate the correlation coefficient r of the relationship between antecedent values of a soil moisture proxy and S^* with catchment attributes. We examined catchment size, mean topographical slope, mean saturated conductivity, dominant texture class, mean plant available water content, mean clay content, mean precipitation, mean discharge, mean potential evaporation (ET), mean actual ET as derived from MODIS (Guerschman et al., 2008), mean Enhanced Vegetation Index (EVI) calculated from catchment average MOD34B4 reflectances, and fractions of woody vegetation (from NFI, 1997), non-agricultural land, grazing land, horticulture, and broad acre cropping. Correlations were found in turn with catchment mean precipitation, discharge, ET, topographical slope, and EVI (see Figs. 2.9, 2.10, and 2.11). For most soil moisture proxies and catchment parameters the correlations are significant at the 0.01 level. A study in Austria (Parajka et al., 2005b) found a similar relationship between the standard deviation of the catchment elevation (equivalent to topographical slope) and the correlation coefficient of the relationship between modeled and remotely sensed soil moisture. Due to intercorrelation between the respective parameters it is difficult to identify the primary mechanism behind these correlations. However, there are two possible explanations. Wet catchments (high mean precipitation, discharge, and ET) tend to have more frequent and more intense rainfall events that can be expected to exhibit a more straightforward relationship with runoff amount, leaving little room for improvement and thus resulting in lower correlation coefficients between antecedent values of the proxies and S^*. Fig. 2.12 demonstrates this by showing better model performance at higher ET in (a), and thus less improvement after the addition of a proxy into the model in (b). Poor model performance in drier catchments is also found in other studies (Parajka et al., 2005b; Lidén and Harlin, 2000; Gan and Biftu, 1996; Chiew et al., 1993). Another possibility is that the soil moisture in catchments with high topographical slope and high EVI is generally more spatially heterogeneous, thereby complicating the estimation of catchment-averaged soil moisture. This leads to a soil moisture proxy that is less reliable and thus to poorer performance.

Our results show that large improvements in Curve Number-based estimates of runoff are possible with the addition of soil moisture proxies for most catchments. The improvement is highest in drier catchments with low EVI and topographical slope. Although the remotely sensed proxies currently show lower performance, this could well improve in the near future with the availability of high-accuracy data at finer spatial and temporal resolutions from ESA's Soil Moisture and Ocean Salinity (SMOS) mission in 2009 (Kerr et al., 2001), NASA's Soil Moisture Active/Passive (SMAP) mission in 2012 (NRC, 2007b), and NASA's Global Precipitation Measurement (GPM) mission in 2013 (Smith et al., 2006).

2.4 Conclusions

The following conclusions can be drawn based on the analysis performed in the current study:

1. In 186 Australian catchments (51–1979 km^2) large improvements in the predictive capability of the

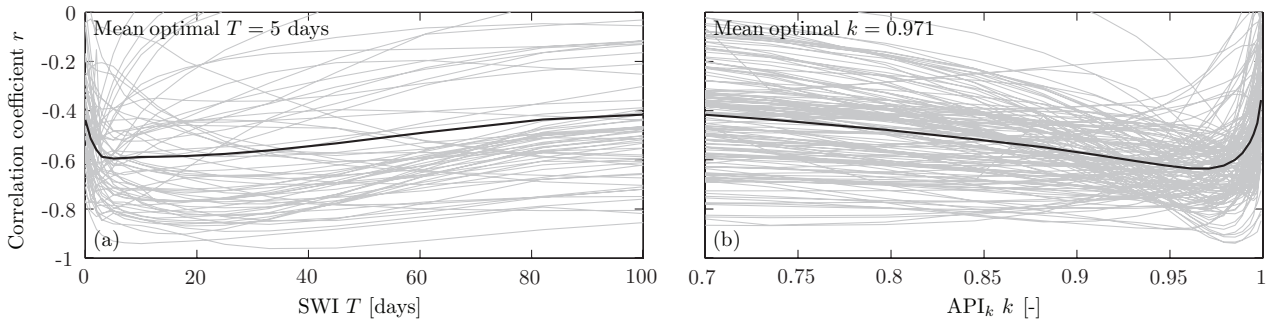


Figure 2.5: The correlation coefficient r between S^* and antecedent SWI for different values of T is shown in (a). The correlation coefficient r between S^* and antecedent API$_k$ for different values of k is shown in (b). Each gray line represents a catchment. The black line represents the mean over all catchments. Only catchments with a significant correlation have been used.

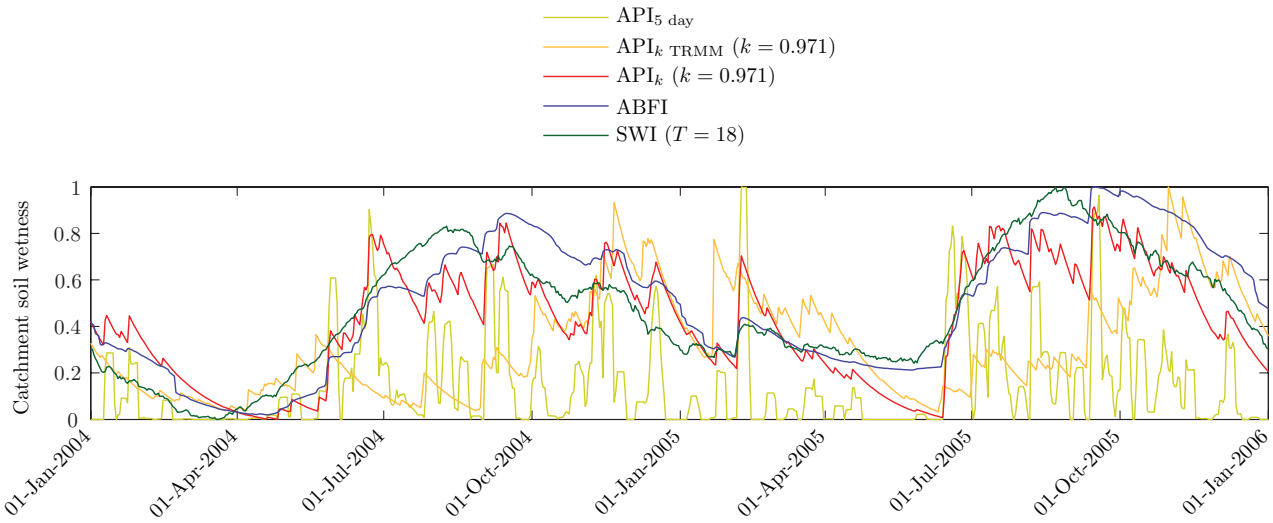


Figure 2.8: Results for catchment 410061 (gauge code) showing the temporal dynamics of the normalized soil moisture proxies.

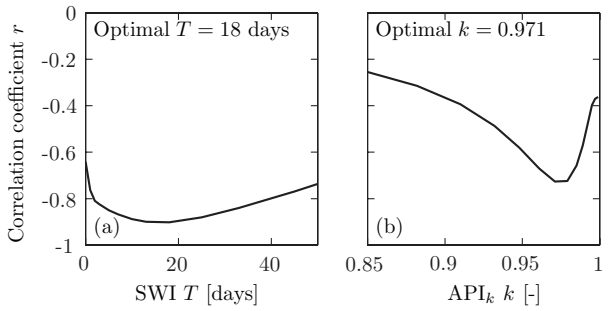


Figure 2.6: Results for catchment 410061 (gauge code) showing the optimization of (a) SWI T and (b) API$_k$ k.

Curve Number model are possible by the use of soil moisture proxies derived from additional data on precipitation, soil moisture, and baseflow. Almost all soil moisture proxies were able to increase the mean modeled vs. observed runoff correlation to around 0.75.

2. API$_k$ performed best, stressing the importance of accurate precipitation data.
3. An optimal value of 5 days was found for the time lag constant T used in the calculation of SWI, and an optimal value of 0.971 was found for the decay parameter k used in the calculation of API$_k$.
4. In dry catchments with low EVI and topographical slope the addition of soil moisture proxies results in large improvements in runoff predictions. Due to covariances between the parameters distinct conclusions about the mechanism behind the correlations are speculative.
5. The soil moisture proxies derived from current satellites (SWI and API$_{k\ TRMM}$) show promising results.

Acknowledgements

The authors wish to thank Juan Pablo Guerschman for assistance in developing the dataset used in this study.

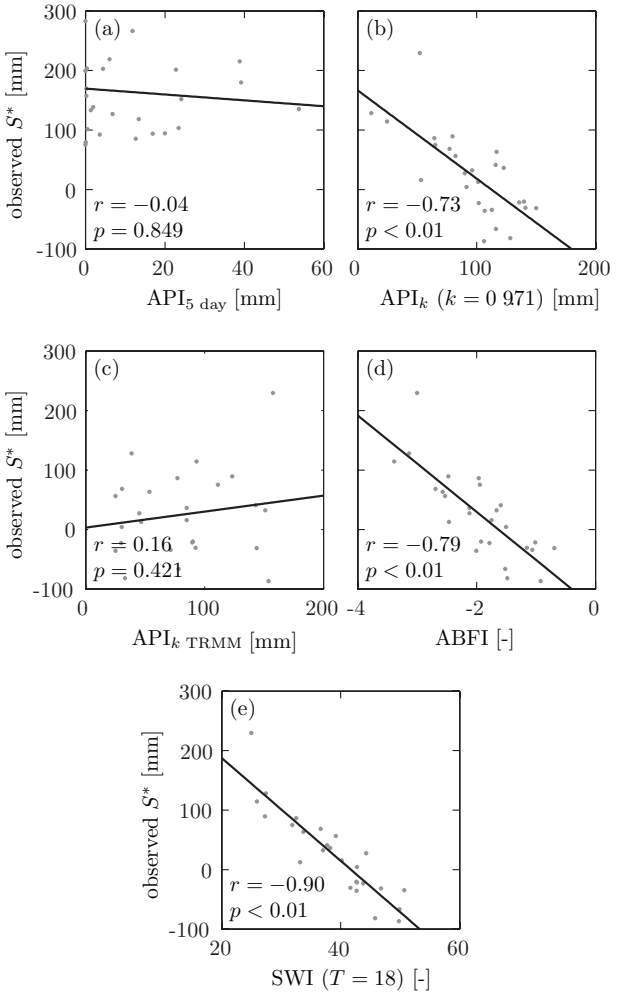


Figure 2.7: Results for catchment 410061 (gauge code) showing the regression between antecedent values of the respective soil moisture proxies and S^*.

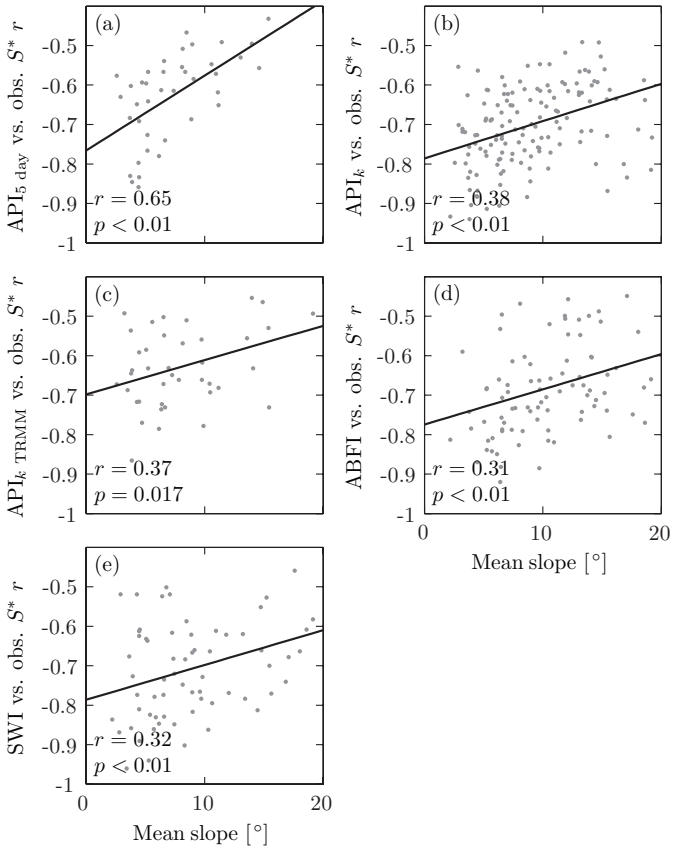


Figure 2.9: The performance of each soil moisture proxy plotted against catchment mean slope. Only the catchments with a significant correlation between the soil moisture proxy and S^* are shown.

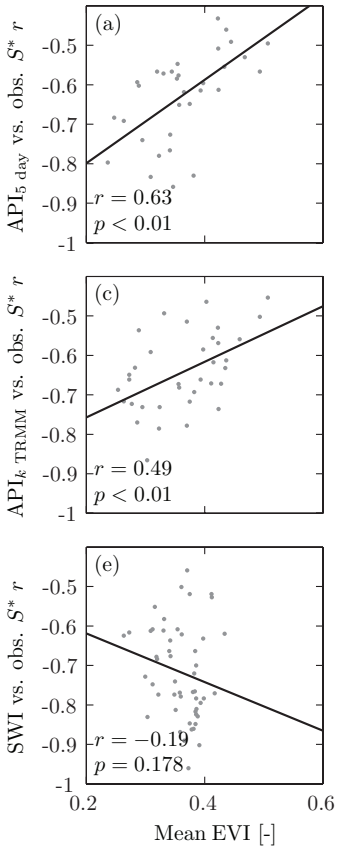


Figure 2.10: The performance of each soil moisture proxy plotted against catchment mean Enhanced Vegetation Index (EVI). Only the catchments with a significant correlation between the soil moisture proxy and S^* are shown.

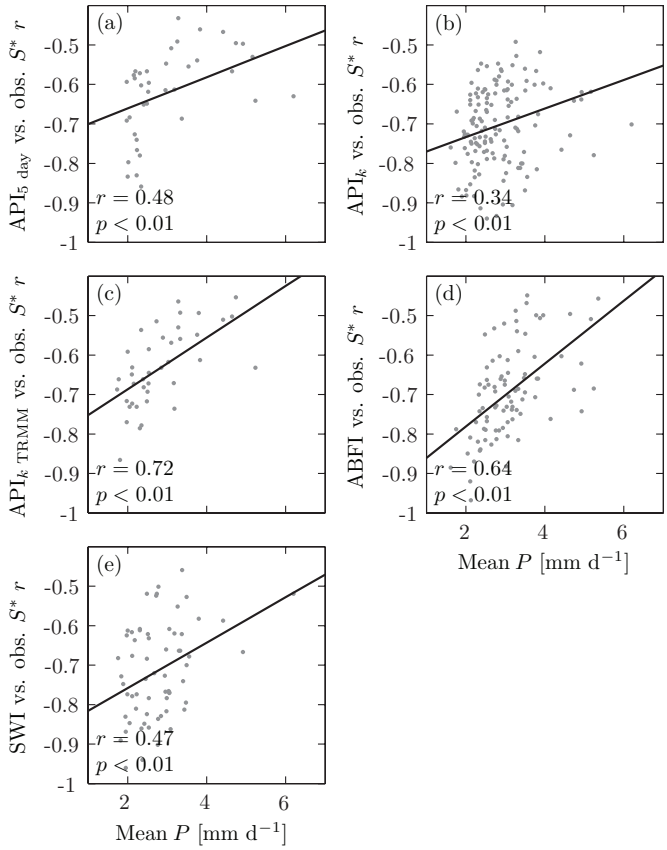


Figure 2.11: The performance of each soil moisture proxy plotted against catchment mean precipitation. Only the catchments with a significant correlation between the soil moisture proxy and S^* are shown.

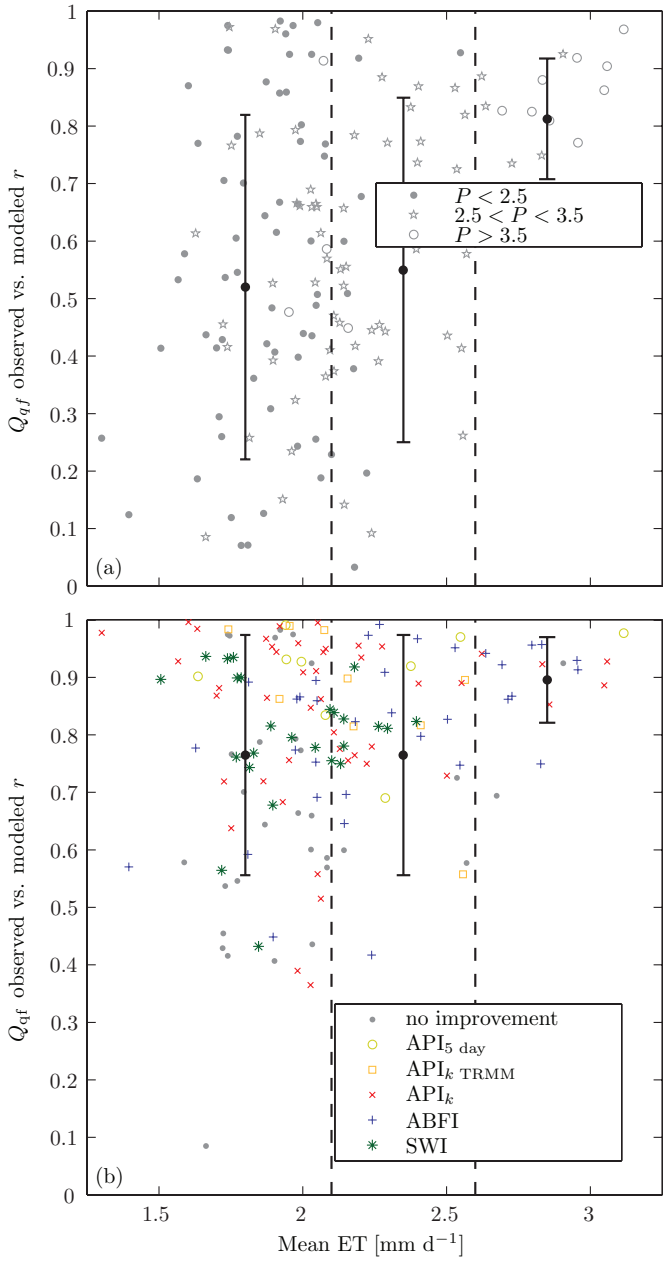


Figure 2.12: (a) Catchment mean evaporation (ET) plotted against the correlation coefficient r of the relationship linking observed and modeled runoff. To show the relationship between ET and P different symbols were used depending on the mean catchment P [mm d^{-1}]. (b) Correlation coefficient r of the relationship linking observed and modeled runoff after the inclusion of the soil moisture proxy that worked best. The error bars show the mean r and standard deviation of the catchments between the dashed lines.

Chapter 3

Global evaluation of four AVHRR-NDVI datasets: Intercomparison and assessment against Landsat imagery[1]

Abstract. Advanced Very High Resolution Radiometer (AVHRR) data with their long-term (1981–current) global coverage at frequent intervals provide unique opportunities to explore vegetation dynamics related to climate variability, climate change, and land-use driven changes of land cover. Several AVHRR-derived Normalized Difference Vegetation Index (NDVI) datasets exist, each based on the AVHRR Global Area Coverage archive but differing in their processing to correct for sensor and atmospheric effects. This paper presents a global comparative analysis for the land surface involving four AVHRR-derived NDVI datasets: (1) Pathfinder AVHRR Land (PAL); (2) Global Inventory Modeling and Mapping Studies (GIMMS); (3) Land Long Term Data Record (LTDR) version 3 (V3); and (4) Fourier-Adjustment, Solar zenith angle corrected, Interpolated Reconstructed (FASIR). Our aims are two-fold: (1) to assess the level of agreement of the medians, trends, and variances, as well as the correlation between the four AVHRR-NDVI datasets from 1982 to 1999; and (2) to independently assess the performance of each AVHRR-NDVI dataset, and that of Moderate Resolution Imaging Spectroradiometer (MODIS) NDVI, using 11 764 Landsat samples of 20×20 km^2 located globally covering every major land-cover type. For the AVHRR-NDVI intercomparison equal medians, variance, and trends, and no correlation between all the respective AVHRR-NDVI datasets were found for 9.9 %, 45.5 %, 48.1 % and 61.6 % of the total land surface, respectively ($p \geq 0.05$). For the four AVHRR-NDVI datasets we found: (1) consistent trends for the tundra and particularly Australia; (2) inconsistent trends for Europe, Africa, and the Sahel; and (3) moderately consistent trends for the rest of the terrestrial land surface including North America and China. The PAL and LTDR V3 datasets lack calibration, as evidenced by the presence of apparent trends in desert areas. In the Landsat-NDVI vs. AVHRR-NDVI comparison of absolute values the LTDR V3 dataset performed best, whereas in the comparison of temporal-change values the GIMMS dataset performed best. In both analyses MODIS-NDVI performed better than any AVHRR-NDVI dataset. The simple average of the four AVHRR-NDVI datasets produced better results than either AVHRR-NDVI dataset alone, indicating that the errors between the datasets are at least partially unrelated. This research emphasizes the implications of AVHRR-NDVI dataset choice for studies assessing the vegetation response to climate change and modelling of the terrestrial carbon balance.

[1]This chapter is an edited version of: Beck, H. E.; McVicar, T. R.; van Dijk, A. I. J. M.; Schellekens, J.; de Jeu, R. A. M., and Bruijnzeel, L. A. Global evaluation of four AVHRR-NDVI datasets: Intercomparison and assessment against Landsat imagery. *Remote Sensing of Environment*, 115(10):2547–2563, 2011.

3.1 Introduction

Amongst the different spectral vegetation indices derived from remotely sensed imagery the Normalized Difference Vegetation Index (NDVI; Tucker, 1979) has been the most widely used. NDVI is calculated as follows:

$$\mathrm{NDVI} = \frac{\rho_{\mathrm{nir}} - \rho_{\mathrm{r}}}{\rho_{\mathrm{nir}} + \rho_{\mathrm{r}}}, \tag{3.1}$$

where ρ_{r} and ρ_{nir} are the spectral reflectance in the red and near-infrared channels, respectively. NDVI has been found to be related to biophysical variables such as Leaf Area Index (LAI; e.g., Wang et al., 2005), fraction of photosynthetically active radiation (fPAR; e.g., Gallo et al., 1985; Goward et al., 1985), foliage projective cover (e.g., Lu et al., 2003), biomass (e.g., Hunt, 1994), and productivity (e.g., Myneni and Williams, 1994; Myneni et al., 1997; Schloss et al., 1999). NDVI has also been widely and successfully used to study vegetation-climate interactions (e.g., McVicar and Jupp, 1998; Ichii et al., 2002; Donohue et al., 2009), for detecting long-term vegetation trends (e.g., Tucker et al., 2001; Eklundh and Olsson, 2003), to model the global carbon balance (e.g., Potter, 1999; Fang et al., 2003), and to assess vegeta-

tion functional characteristics (e.g., Berry and Roderick, 2002; DeFries et al., 2000; Lu et al., 2003).

Long-term time series are required to detect vegetation trends. This renders data collected by the NOAA series of Advanced Very High Resolution Radiometer (AVHRR) sensors (in operation since 1981) important for numerous environmental applications (NRC, 2007a). However, AVHRR lacks reliable onboard calibration devices (Staylor, 1990), suffers from satellite orbital drift (Price, 1991), and has a limited set of relatively broad (compared to modern-day sensors) spectral bands which reduces the accuracy of atmospheric corrections (Tanré et al., 1992). These effects potentially result in unreliable trends (Gutman, 1999). Several global NDVI datasets have been produced to capitalize on the extended temporal coverage, each starting with very similar AVHRR Global Area Coverage (GAC; Kidwell, 1998) data, but differing in their processing streams to reduce sensor, illumination, and atmospheric effects. The four most widely used AVHRR-NDVI datasets are: (1) Pathfinder AVHRR Land (PAL; James and Kalluri, 1994); (2) Global Inventory Modeling and Mapping Studies (GIMMS; Tucker et al., 2005); (3) Land Long Term Data Record (LTDR; Pedelty et al., 2007); and (4) Fourier-Adjustment, Solar zenith angle corrected, Interpolated Reconstructed (FASIR; Los et al., 2000).

Despite some general agreement, global and regional analyses have highlighted important differences, both in terms of absolute AVHRR-NDVI values and trends derived from AVHRR-NDVI series (Table 3.1). Here we extend previous studies by: (1) assessing all four AVHRR-NDVI datasets globally; and (2) performing both absolute-value and temporal-change assessments of the AVHRR-NDVI datasets using Landsat to assess which dataset is the most suited for specific applications. Previously only three datasets have been compared globally (McCloy et al., 2005), whereas a comparison of all four datasets has only been made for the Iberian Peninsula (Alcaraz-Segura et al., 2010b).

Several studies have compared AVHRR-NDVI imagery to NDVI as derived from sensors with better calibration, including Landsat (e.g., Buheaosier et al., 2003; Tucker et al., 2005; Hall et al., 2006; Brown et al., 2006; Tarnavsky et al., 2008; Pouliot et al., 2009; Stellmes et al., 2010), SPOT-VGT (e.g., Tucker et al., 2005; Brown et al., 2006; Fensholt et al., 2006; Tarnavsky et al., 2008; Fensholt et al., 2009; Song et al., 2010), and Moderate Resolution Imaging Spectroradiometer (MODIS; e.g., Huete et al., 2002; Buheaosier et al., 2003; Fensholt, 2004; Venturini et al., 2004; Tucker et al., 2005; Gallo et al., 2005; Kawamura et al., 2005; Bédard et al., 2006; Brown et al., 2006; Tarnavsky et al., 2008; Ji et al., 2008; Fensholt et al., 2009). However, most of these studies were regional in design, only covered a limited number of land-cover types, and only two such studies were conducted with the aim of comparing alternative AVHRR-NDVI datasets (Fensholt et al., 2006; Hall et al., 2006). Additionally, studies using Landsat imagery generally use a small number of Landsat acquisitions (i.e., 1, 8, 3, 21, 1, 10, and 10 Landsat acquisitions for the respective Landsat studies cited at the start of this paragraph).

The present analysis consists of two parts. First, we determine the degree of internal consistency in terms of median, variance, and trend, as well as the degree of correlation between the different AVHRR-NDVI datasets (PAL, GIMMS, LTDR, and FASIR) for the global land surface. This type of analysis provides insight into the differences of results that may arise when using one dataset over another. Our study is the first to use LTDR version 3 (V3), which is expected to bring considerable improvement over previous AVHRR-NDVI datasets (Nagol et al., 2009), including LTDR version 2.

In the second part, we independently assess the performance of the four AVHRR-NDVI datasets (plus MODIS-NDVI) using NDVI values derived from 11 764 Landsat samples of 20×20 km^2 (Potapov et al., 2011) made available by the Food and Agricultural Organization (FAO), located globally covering every major land-cover type. There are numerous instances where two Landsat samples (i.e., a pair) were acquired for the same location, allowing both spatial and temporal validation of the AVHRR-NDVI and MODIS-NDVI datasets. Cloud-free Landsat imagery is suitable to validate AVHRR-NDVI because it represents high-resolution (30 m) instantaneous acquisitions that are less affected by sub-pixel cloud contamination, spatial compositing, and mixed pixels (Hall et al., 2006). Additionally, Landsat does not suffer as much from sensor degradation due to its onboard calibration (achieving an accuracy of ± 5 %; Chander and Markham, 2003). We aim to use MODIS-NDVI data, with its advanced atmospheric correction (Huete et al., 1999; Vermote et al., 2002) to assess the relative performance of the AVHRR-NDVI datasets. The paper is structured around these two parts in Methods (section 3.3), Results (section 3.4), and Discussion (section 3.5), with the Conclusion presented in section 3.6. The datasets used are described in section 3.2.

3.2 Datasets

3.2.1 AVHRR-NDVI

We use four long-term AVHRR-NDVI datasets: (1) PAL (James and Kalluri, 1994); (2) GIMMS (Tucker et al., 2005); (3) LTDR V3 (Pedelty et al., 2007); and (4) FASIR (Los et al., 2000). Table 3.2 provides a detailed description of each dataset. Maximum-value composites of LTDR V3 data at 10-day intervals were calculated, matching the compositing interval of the PAL and FASIR datasets. For all datasets annual means were calculated at 0.5° resolution for the analysis described in section 3.3.1, and monthly means were calculated at their original spatial resolution for the analysis described in section 3.3.2. To restrict the analysis to 'growing season' months, and to avoid large solar zenith angles and influences from snow/ice in winter at the high northern

Table 3.1: Overview of studies comparing AVHRR-NDVI datasets, including key findings of the respective studies. The current study is added for the sake of completeness.

Region (period of study)	Reference	Datasets	Approach	Key findings
Northern hemisphere (1982–1999)	Slayback et al. (2003)	PAL, GIMMS, FASIR, Preflight[a]	Analysis of the impact of dataset choice and processing, on trends.	Large improvement in data quality after calibration and atmospheric correction. The obtained NDVI trends were larger than can be explained by sensor effects. Trends of GIMMS were generally lower than trends from PAL and FASIR, whereas FASIR data had better sensor transitions and stability, and were less noisy than GIMMS and PAL.
North-American boreal forests (1984–2006)	Alcaraz-Segura et al. (2010a)	GIMMS, CCRS[b]	Trends at recently burned sites.	CCRS showed vegetation recovery in all recently burned sites, unlike GIMMS, which appeared to be biased towards negative trends. Possible explanations included drift over-correction by GIMMS, the spatial maximum-value compositing used by GIMMS, and the lower temporal resolution of GIMMS compared to CCRS.
Northern Eurasia (1982–1988 and 1995–1999)	De Beurs and Henebry (2008)	PAL, GIMMS	Correlation of NDVI start of growing season and peak dates with climate indices.	More similarity between AVHRR-NDVI datasets in dates of peak position than in dates for start of season. GIMMS resulted in better correlations than PAL.
Alaskan tundra (1990–1999)	Stow et al. (2007)	GIMMS, USGS[c]	Agreement evaluation of trends.	General patterns were similar, but localized areas of vegetation increases lost by GIMMS.
Eight globally spread sites (1989)	Ouaidrari et al. (2003)	PAL, PAL-II[d], FASIR, LASUR[e]	Comparison of one-year long time series at eight sites ($1° \times 1°$) with different land-cover types.	Time series of PAL-II showed equal or less noise compared to the other datasets.
Iberian Peninsula (1982–1999)	Alcaraz-Segura et al. (2010b)	PAL-II[d], GIMMS, LTDR V2, FASIR	Agreement evaluation of trends.	Spatial consistency over 43 % of the study area across the four datasets. GIMMS most dissimilar to the other datasets.
South America (1982–1999)	Baldi et al. (2008)	PAL, GIMMS, FASIR	Trend agreement evaluation and assessment of plausibility of each dataset at 5 locations.	PAL and FASIR showed the best agreement. GIMMS showed the least plausible results.
Africa (1998–2004)	Fensholt et al. (2006)	PAL, GIMMS	Comparison of continental time series with Satellite Pour l'Observation de la Terre Vegetation (SPOT-VGT) data.	GIMMS showed the lowest Root Mean Square Error and had the best agreement in dynamic range in the validation against SPOT-VGT data, and was therefore considered the more accurate dataset.
Global (1981–2000)	McCloy et al. (2005)	PAL, GIMMS, FASIR	Trend agreement evaluation and temporal correlation.	FASIR and GIMMS better correlated with each other than either dataset with PAL. Anomalies between datasets most pronounced in the high northern latitudes ($> 60°$). None of the datasets can be used for regional studies.
Global (1982–1998)	Hall et al. (2006)	GIMMS, FASIR	Comparison of absolute values, validation against Landsat (three scenes), and time-series comparison.	Large differences between absolute values, especially in the northern latitudes. GIMMS matches best with Landsat top-of-atmosphere reflectance, and FASIR with Landsat surface reflectance. Time series of NDVI anomalies from the mean for the continental regions showed low correlations, especially during 1984–1985 and 1994. dataset selection has large implications for carbon modeling.
Global (1982–1999)	Current study	PAL, GIMMS, LTDR V3, FASIR	Agreement evaluation of trends and validation using 11 764 Landsat-5 samples.	Pronounced trend disagreement was found for Europe, Africa, and the Sahel. Highly consistent trends were found for the tundra and Australia. In the validation against Landsat, LTDR V3 is most accurate in terms of absolute values and GIMMS most accurate in assessing change. Landsat samples were checked against MODIS.

[a] AVHRR-NDVI dataset (1982–1999, half-monthly, 8 km) calibrated using only the NOAA pre-launch coefficients, including aerosol corrections.
[b] Local Area Coverage (LAC) AVHRR-NDVI dataset (1984–2006, 10 days, 1 km) provided by the Canada Centre for Remote Sensing (CCRS) calibrated using post-launch coefficients, corrected for cloud and shadow contamination, water vapor, ozone, and aerosols, and additional calibration applied to account for inter-sensor differences.
[c] LAC AVHRR-NDVI dataset (1990–1999, half-monthly, 1 km) calibrated using only post-launch coefficients produced by the United States Geological Survey (USGS), cross-calibrated with GIMMS to correct for aerosols.
[d] AVHRR-NDVI dataset (1982–1999, 10 days, 8 km) similar to the PAL dataset with the main difference being the inclusion of water-vapor corrections (El Saleous et al., 2000).
[e] Land SUrface Reflectance (LASUR) AVHRR-NDVI dataset (1989–1990, 10 days, 16 km) developed at Centre d'Etudes Spatiales de la BIOsphère, based on the Global Vegetation Index dataset.

latitudes, pixels with $< 0°$ surface air temperature based on a monthly climatology for 1982–1999 were excluded from the annual mean calculation. Average air temperature from the CRU TS 3.0 dataset were used (monthly temporal and 0.5° spatial resolution; downloaded from `http://badc.nerc.ac.uk/browse/badc/cru` in February 2011; Mitchell and Jones, 2005).

3.2.2 Landsat-NDVI

Landsat data (Potapov et al., 2011) made available by the FAO were used for the analysis described in section 3.3.2 to assess the performance of the respective AVHRR-NDVI datasets. The FAO has been reporting on the state of forest resources every 5 to 10 years from 1946 to 2010 through their Forest Resource Assessments (FRA). In order to obtain systematic information on forest cover change, a global Remote Sensing Survey (RSS) is being undertaken for FRA 2010 (Ridder, 2007). The image data for the RSS were provided by the USGS Global Land Survey dataset (Gutman et al., 2008), which consists of a global collection of ortho-rectified cloud-free Landsat image acquisitions around the years 1975, 1990, 2000, and 2005. To reduce computational time the RSS uses a sampling grid consisting of 13 689 sites at each integer terrestrial degree latitude and longitude intersection between 0 and 60°N/S and every second degree intersection between 60° and 75°N/S (Potapov et al., 2011). The RSS consists of 56 218 Landsat-4, -5, and -7 samples that are centered at the intersection points, each covering an area of 20×20 km^2, with often two images being acquired for the same sample location. We only use the 19 978 Landsat-5 images. The data were downloaded from `http://globalmonitoring.sdstate.edu/projects/fao/` in January 2011.

Landsat raw digital numbers were converted to radiance [W m^{-2} sr^{-1} μm^{-1}] and subsequently top-of-atmosphere (TOA) reflectance [unitless] using equations from Chander et al. (2009). Depending on the values listed in the header files supplied with the Landsat samples, two approaches were used for the conversion to radiance. The first approach used minimum and maximum radiance parameters, whereas the second used gain and bias parameters. Samples that used gain and bias parameters that had units of radiance listed as "UNKNOWN" in the header file were discarded from the analysis, as these samples gave suspect results.

NDVI was calculated according to Eq. 3.1 using Landsat bands 3 and 4 TOA reflectance for the ρ_{r} and ρ_{nir} terms, respectively. Prior to the calculation of NDVI the Landsat data were averaged to match the spatial resolution of the AVHRR-NDVI or MODIS-NDVI dataset that was subject of the comparison. Pixels flagged as cloud (using the thermal infrared band) or water were masked, and samples with > 50 % missing values were excluded from our analysis, resulting in a total of 11 764 Landsat samples at 8012 unique sites. The group of samples acquired pre-2000 and post-2000 are used to validate AVHRR-NDVI and MODIS-NDVI, respectively. Fig. 3.1 shows the temporal distribution of the samples, Fig. 3.2 shows the geographical locations, and Table 3.3 lists the distribution amongst the different land-cover classes. It is noted that the Landsat data were not corrected for atmospheric influences, Bidirectional Reflectance Distribution Function (BRDF) effects, or topographical shadowing. The Landsat data are referred to hereafter as the "Landsat samples".

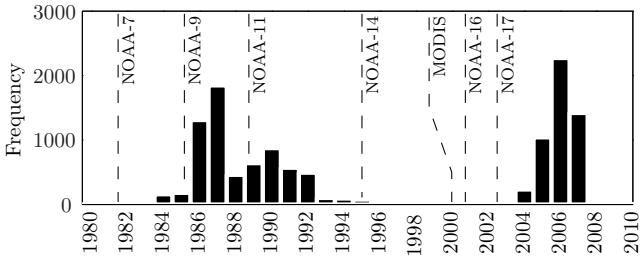


Figure 3.1: Histogram showing the number of Landsat samples used in each year. Also shown are the launch dates of MODIS Terra and the NOAA satellites.

Table 3.3: For each land-cover class the number (N) of Landsat samples and the number of sample pairs of the same location with different acquisition dates are shown. For samples with a mix of classes the most dominant class was used.

	Pre-2000 (N)		Post-2000 (N)	
Land-cover class	Samples	Pairs	Samples	Pairs
All classes	6794	1424	4970	849
Evergreen needleleaf forest	863	282	602	107
Evergreen broadleaf forest	458	39	356	19
Deciduous needleleaf forest	145	41	327	78
Deciduous broadleaf forest	273	62	233	50
Mixed forest	526	154	372	71
Woodland	529	96	549	91
Wooded grassland	324	25	164	9
Closed shrubland	237	43	189	28
Open shrubland	673	93	222	21
Grassland	808	187	586	109
Cropland	1054	295	720	139
Bare ground	725	47	49	4
Mosses and lichens	179	60	601	123

Of the pre-2000 Landsat samples 10.4 % was acquired after June 3 1991, marking the beginning of the Mount Pinatubo volcanic eruption that resulted in a substantial increase in stratospheric aerosol loading. Most of the Landsat samples were acquired in the local summer (i.e., 56.1 % of the samples > 23.5°N/S were acquired in June, July, or August in the northern hemisphere or December, January, or February in the southern hemisphere; 80.1 % of all samples are located > 23.5°N/S). The median difference in acquisition time between Landsat samples having the same location (i.e., Landsat sample pairs) is 1.16 y and 0.76 y for the pre-2000 and post-2000 samples, respectively. The median seasonal difference in acquisi-

Table 3.2: Processing streams for the AVHRR-NDVI datasets used in the present study.

	PAL	GIMMS	LTDR V3	FASIR
Temporal extent	1981–1999	1981–2006	1981–2000	1981–1999
Geographical extent	90°S–90°N	62.85°S–89.22°N	90°S–90°N	70°S–70°N
Data input	AVHRR GAC level-1b data (4.4 × 1.1 km^2).	AVHRR GAC level-1b data (4.4 × 1.1 km^2).	AVHRR GAC level-1b data (4.4 × 1.1 km^2).	PAL dataset.
Data-sensor lineage	07/1981–01/1985 NOAA-7 02/1985–10/1988 NOAA-9 11/1988–08/1994 NOAA-11 01/1995–12/1999 NOAA-14	07/1981–01/1985 NOAA-7 02/1985–10/1988 NOAA-9 11/1988–09/1994 NOAA-11 10/1994–01/1995 NOAA-9[c] 02/1995–10/2000 NOAA-14 11/2000–12/2003 NOAA-16 01/2004–12/2006 NOAA-17	07/1981–01/1985 NOAA-7 02/1985–10/1988 NOAA-9 11/1988–08/1994 NOAA-11 01/1995–12/2000 NOAA-14	Adopted from PAL.
Radiometric calibration	Desert target approach (Rao et al., 1993).	Ocean and clouds vicarious calibration (Vermote and Kaufman, 1995) for NOAA-7 to -14 and pre-launch coefficients for NOAA-16 and -17. Additional calibration using desert targets (Los, 1998).	Ocean and clouds vicarious calibration (Vermote and Kaufman, 1995).	Adopted from PAL, with additional calibration using desert targets (Los, 1998).
Cloud screening	CLAVR-1[a] algorithm.	Channel 5 0°C thermal mask.	Channel 5 0°C thermal mask.	Adopted from PAL.
BRDF[d] correction	None.	Empirical mode decomposition (Pinzon et al., 2005).	BRDF[d] technique also used in MODIS processing (new in V3).	BRDF[d] technique using Ross-Li model (Wanner et al., 1995).
Molecular scattering	Following Gordon et al. (1988).	None.	Using reanalysis data from the NOAA National Center for Environmental Prediction (NCEP).	Adopted from PAL.
Ozone absorption	Using TOMS[b] data.	None.	Using TOMS[a] data.	Adopted from PAL.
Aerosol scattering	None.	El Chichón 04/1982–12/1984 (Sato et al., 1993) and Mount Pinatubo 06/1991–12/1993 (Vermote et al., 1997).	Following Vermote et al. (1997); new in V3.	El Chichón 05/1982–12/1983 (Sato et al., 1993) and Mount Pinatubo 06/1991–12/1994.
Water-vapor absorption	None.	None.	Using reanalysis data from the NOAA NCEP.	None.
Temporal compositing	10-day maximum NDVI.	15-day maximum NDVI.	None, daily values (10-day maximum-NDVI composites are used in the present study).	Adopted from PAL.
Spatial compositing	0.08° maximum NDVI.	0.08° maximum NDVI.	0.05° maximum NDVI.	Adopted from PAL, final data reprojected to 0.125° regular grid.
Noise removal	None.	Kriging interpolation.	Fourier adjustment.	None.
Gap filling	None.	Kriging interpolation.	None.	Replacement of values below the median October value in evergreen needleleaf forests, selection of maximum NDVI in a 3 × 3 pixels moving window in evergreen broadleaf forests, and interpolation during 09/1994–12/1994.
Geo-registration correction	None.	None.	Using an orbital prediction model.	None.
Adjustments	None.	Data were non-linearly rescaled against SPOT-VGT (details not reported).	None.	None.
Data source	No longer available for download, archived copy used (King, 2006).	Downloaded from `ftp://ftp.glcf.umd.edu/glcf/GIMMS/` in December 2009.	Downloaded from `ftp://ltdr.nascom.nasa.gov/pub/f302/AVHRR/allData/Ver3/` in May 2010.	Provided by S.O. Los in March 2010.

[a] Clouds from AVHRR-Phase I (Stowe et al., 1999).
[b] Total Ozone Mapping Spectrometer (McPeters et al., 1998).
[c] Descending node data only.
[d] Bidirectional Reflectance Distribution Function.

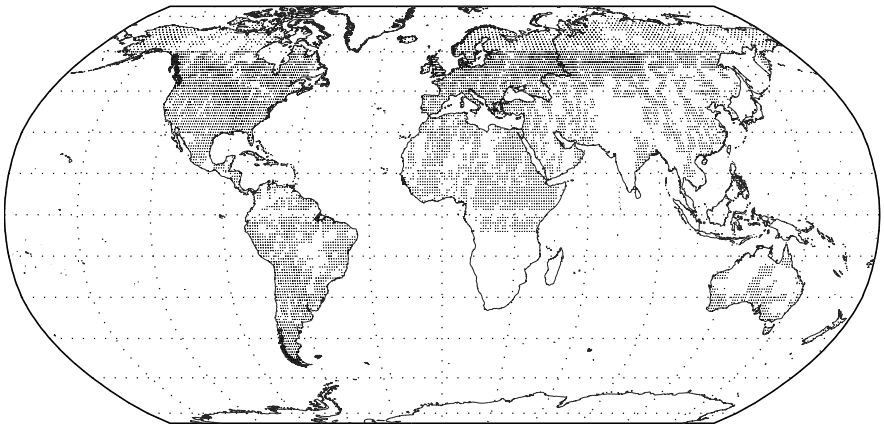


Figure 3.2: The 8012 unique Landsat sampling locations used in the present study.

tion time (i.e., the absolute difference in acquisition day number of the year) between Landsat samples having the same location (i.e., Landsat sample pairs) is 36 and 25 days for the pre-2000 and post-2000 samples, respectively.

3.2.3 MODIS-NDVI

MODIS is a newer generation instrument onboard the Terra satellite (launched December 18, 1999) and the Aqua satellite (launched May 4, 2002). MODIS-NDVI data are used in the analysis described in section 3.3.2 to assess the relative performance of the AVHRR-NDVI datasets. Comparisons between the AVHRR-NDVI datasets and MODIS-NDVI were not performed due to lack of temporal overlap of all four AVHRR-NDVI datasets (cf. Table 3.2). The 0.05° monthly MODIS Collection-5 MOD13C2 product (Huete et al., 1999) was acquired from the NASA Earth Observing System data gateway (`http://eospso.gsfc.nasa.gov/`) in May 2010.

3.2.4 Land-cover classification map

The present study uses the University of Maryland (UMD) Global Land Cover (GLC) classification map (Hansen et al., 2000), which distinguishes fourteen land-cover classes (Fig. 3.3). The UMD GLC classification map is based on AVHRR data acquired between 1981 and 1994 (Hansen et al., 2000).

3.3 Methods

3.3.1 Intercomparison of AVHRR-NDVI datasets

Several analyses were performed globally over the common temporal extent of 1982–1999 and the common geographical extent of 62.85°S–70°N (Table 3.2) to assess the consistency amongst the four AVHRR-NDVI datasets (PAL, GIMMS, LTDR V3, and FASIR). Although many metrics exist to describe the phenology of vegetation growth using NDVI (e.g., Reed et al., 1994), the present study focuses on annual mean NDVI as it is the most commonly used metric in long-term vegetation studies. In the calculation of the annual means pixels with $< 0°$ surface air temperature based on a monthly climatology for 1982–1999 were excluded from the analysis (cf. section 3.2.1). Prior to the analysis, data were downsampled by averaging at 0.5° resolution to expedite the computations. To assess the consistency between the AVHRR-NDVI datasets three sets of analyses were performed:

1. For each dataset and each 0.5° pixel the median, variance, and trend were calculated from time series comprised of 18 z-score transformed annual mean values (one value for each year from 1982 to 1999). The resulting maps of median, variance, and trend were subsequently averaged for each 0.5° latitude band and each land-cover class. To remove potential biases between the datasets, prior to the analysis the data were z-score transformed by:

$$\mathrm{z} = \frac{x - \mu}{\sigma}, \tag{3.2}$$

   where z is the z-score of the pixel, x is the NDVI value of the pixel, and μ and σ are the temporal mean and temporal standard deviation of the dataset, respectively. For each dataset μ and σ were calculated from the monthly pixels that have non-missing values for all four datasets. It is noted that the z-score transformation does not change the skewness or kurtosis of the data, or the sign of trends calculated from time series.
2. For each dataset and each 0.5° pixel the median, variance, and trend were calculated from time series comprised of 18 annual mean values (one value for each year from 1982 to 1999). The spatial correlation (indicated by the coefficient of determination, R^2) between all pairs of the datasets was calculated for the median, variance, and trend, resulting in a

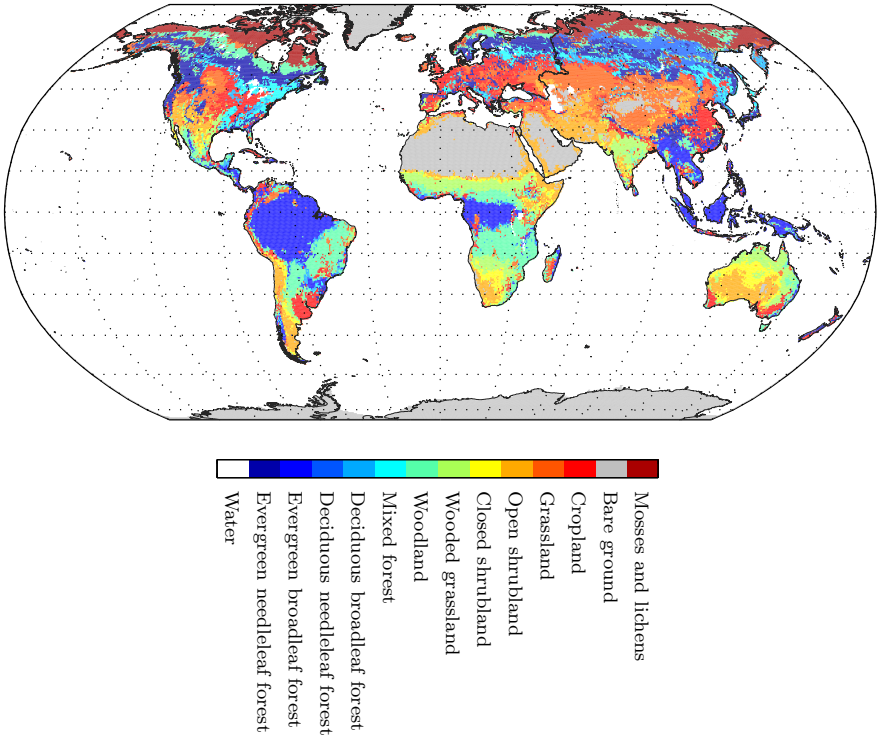


Figure 3.3: Map of the UMD GLC classification.

correlation matrix of six unique R^2 values per statistic. Due to spatial autocorrelation in the data conventional tests of the probability, indicated by the p-value, will yield incorrect results. We therefore used a modified t-test (Clifford et al., 1989) that reduces the degrees of freedom to obtain corrected probability levels.

3. For each 0.5° pixel the agreement between the datasets in terms of median, variance, and trend, and the degree of correlation, was assessed from four groups comprised of 18 z-score transformed (using Eq. 3.2) annual mean values (one value for each year from 1982 to 1999 for all four datasets). The following statistical tests were performed on a per-pixel basis to test four relevant hypotheses (resulting in four p-value maps, one for each statistical test):
   (a) The Kruskal-Wallis test (Kruskal and Wallis, 1952) was used for calculating a p-value to test the null hypothesis that the medians of the datasets are equivalent.
   (b) Levene's test (Levene, 1960) was used to obtain a p-value to test the null hypothesis of equal variances.
   (c) The Kruskal-Wallis test (Kruskal and Wallis, 1952) was used again to test the null hypothesis that the trends of the datasets are equal. For each dataset from all pairs of the time series comprised of 18 annual mean values 'trends' were calculated, resulting in 153 'trends' per dataset. The Kruskal-Wallis test was subsequently used to calculate a p-value to asses the equality in medians between the four groups of 153 'trends'.
   (d) The null hypothesis that there is no correlation between the datasets (Helsel and Hirsch, 2002, section 8.2) was assessed by computing a mean p-value for the correlation matrix between all pairs of datasets.

For all analyses a significance level of 0.05 was used, i.e., if one of the analyses yields a p-value < 0.05, the null hypothesis is rejected. Pixels identified as water in the original datasets were ignored, and if a 0.5° pixel had > 50 % of the original dataset's pixels identified as water in one of the four AVHRR-NDVI datasets, the 0.5° pixel was excluded from analysis.

3.3.2 Assessment against Landsat-NDVI

Two kinds of analyses were performed to evaluate the accuracy of the AVHRR-NDVI and MODIS-NDVI datasets. The first compares AVHRR-NDVI and MODIS-NDVI with Landsat-NDVI in terms of absolute values. The second compares the difference according to a Landsat sample-pair (i.e., two 20×20 km^2 Landsat samples acquired for the same location at different dates) with the difference according to the AVHRR-NDVI and MODIS-NDVI datasets. In both comparisons the mean NDVI of the Landsat sample is calculated and for the AVHRR-NDVI and MODIS-NDVI datasets the mean NDVI of the pixels covered by the 20×20 km^2 sample is calculated. To account for differences in the temporal resolution, monthly means of the AVHRR data were calculated prior to the analysis (at each dataset's original spatial resolution) and the monthly image nearest to

the Landsat sample acquisition date was selected. Only Landsat samples were used for which all four AVHRR-NDVI datasets have non-missing values (i.e., the same samples were used for all four AVHRR-NDVI datasets).

The present study uses the Root Mean Square Difference (RMSD) statistic to quantify the accuracy of the AVHRR-NDVI and MODIS-NDVI datasets. The RMSD is calculated over all data and for each land-cover class separately. In order to remove systematic differences between the AVHRR-NDVI or MODIS-NDVI datasets and Landsat-NDVI that may inflate the RMSD, the AVHRR-NDVI and MODIS-NDVI data are linearly rescaled to Landsat-NDVI in the calculation of the RMSD. The regression model:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i \tag{3.3}$$

is used, where $\hat{y}$ are the rescaled AVHRR-NDVI or MODIS-NDVI values, x are the observed AVHRR-NDVI or MODIS-NDVI values, and i denotes the ith observation pair. The model parameters $\hat{\beta}_0$ and $\hat{\beta}_1$ are estimated using the least-squares method by minimizing the residual:

$$e_i = y_i - \hat{y}_i, \tag{3.4}$$

where e are the residuals and y are the observed Landsat-NDVI values. The Landsat-NDVI values are plotted on the ordinate (or y-) axis and the AVHRR-NDVI or MODIS-NDVI values on the abscissa (or x-) axis, thereby avoiding erroneous estimates of $\hat{\beta}_0$ and $\hat{\beta}_1$ (Piñeiro et al., 2008). The RMSD is calculated as:

$$\text{RMSD} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} e_i^2}, \tag{3.5}$$

where N is the number of observation pairs, and the summation is over $i = 1, 2, \ldots, N$. The units of RMSD are Landsat-NDVI. The bias, reflecting the systematic difference between AVHRR-NDVI or MODIS-NDVI and Landsat-NDVI, is calculated by:

$$\text{Bias} = \frac{1}{N}\sum_{i=1}^{N}(x_i - y_i). \tag{3.6}$$

The units of the bias are Landsat-NDVI. Finally, the coefficient of determination (R^2) is calculated by:

$$R^2 = 1 - \frac{\sum_{i=1}^{N} e_i^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2}. \tag{3.7}$$

To assess whether a dataset performs significantly better than another, we use the t-test to calculate probabilities for assessing the null hypothesis of equal means between the absolute residuals ($|e|$; Eq. 3.4) obtained by each dataset.

A final analysis involves calculating the average of the four AVHRR-NDVI datasets in an effort to produce a better AVHRR-NDVI dataset. The fitted parameters $\hat{\beta}_0$ and $\hat{\beta}_1$ (i.e., the intersect and slope, respectively) are applied to the each AVHRR-NDVI dataset and mean AVHRR-NDVI values are calculated. The result is again regressed against Landsat-NDVI, and the RMSD is calculated. This analysis provides insight into the relationship between the errors of the four AVHRR-NDVI datasets, because if the errors are at least partially independent, averaging should result in a more reliable AVHRR-NDVI dataset that obtains a lower RMSD in the comparison with Landsat-NDVI.

3.4 Results

3.4.1 Intercomparison of AVHRR-NDVI datasets

Fig. 3.4 shows the 1982–1999 trend as calculated from the annual mean NDVI for the PAL, GIMMS, LTDR V3, and FASIR datasets. Consistent amongst the different datasets is the increasing mean global trend, but regionally large discrepancies are apparent. The PAL and LTDR V3 datasets show positive trends over several assumed radiometrically stable sites (e.g., the "bare ground" class in Fig. 3.3, most notably the Sahara).

Fig. 3.5 shows for the four AVHRR-NDVI datasets for 1982–1999 the median, variance, and trend calculated from annual mean time series of z-score transformed NDVI, averaged for each 0.5° latitude band and each land-cover class. Figs. 3.5a and b also show Landsat-NDVI values, which were converted to z-scores using Eq. 3.2 and the mean of the coefficients used for the AVHRR-NDVI z-score transformation (Table 3.4). Fig. 3.5a shows similar median values between the AVHRR-NDVI datasets, although north of 60°N the GIMMS dataset appears positively biased compared to the other datasets, which is reflected in the "mosses and lichens" class (corresponding to the tundra) in Fig. 3.5b. The Landsat-NDVI values confirm that the GIMMS dataset overestimates NDVI in this region. Landsat-NDVI values are consistently higher than AVHRR-NDVI values, which is likely due to the Landsat data consisting primarily of summer acquisitions (i.e., 56.1 % of the Landsat samples > 23.5°N/S were acquired in the local summer), whereas the AVHRR-NDVI data are annual means. In terms of variance (Fig. 3.5c) the LTDR V3 dataset has exceptionally high values from 40°S to 30°S, and over the whole latitudinal range the PAL and LTDR V3 datasets have consistently higher values than the GIMMS and FASIR datasets. Similarly, in Fig. 3.5e over most of the latitudinal range the PAL and LTDR V3 datasets show higher trends than the GIMMS and FASIR datasets. However, at latitudes south of 45°S the GIMMS dataset has higher trends (Fig. 3.5e) due to strong trends found in southern Chile (cf. Fig. 3.4b). Furthermore, an increase in trends is observed for the

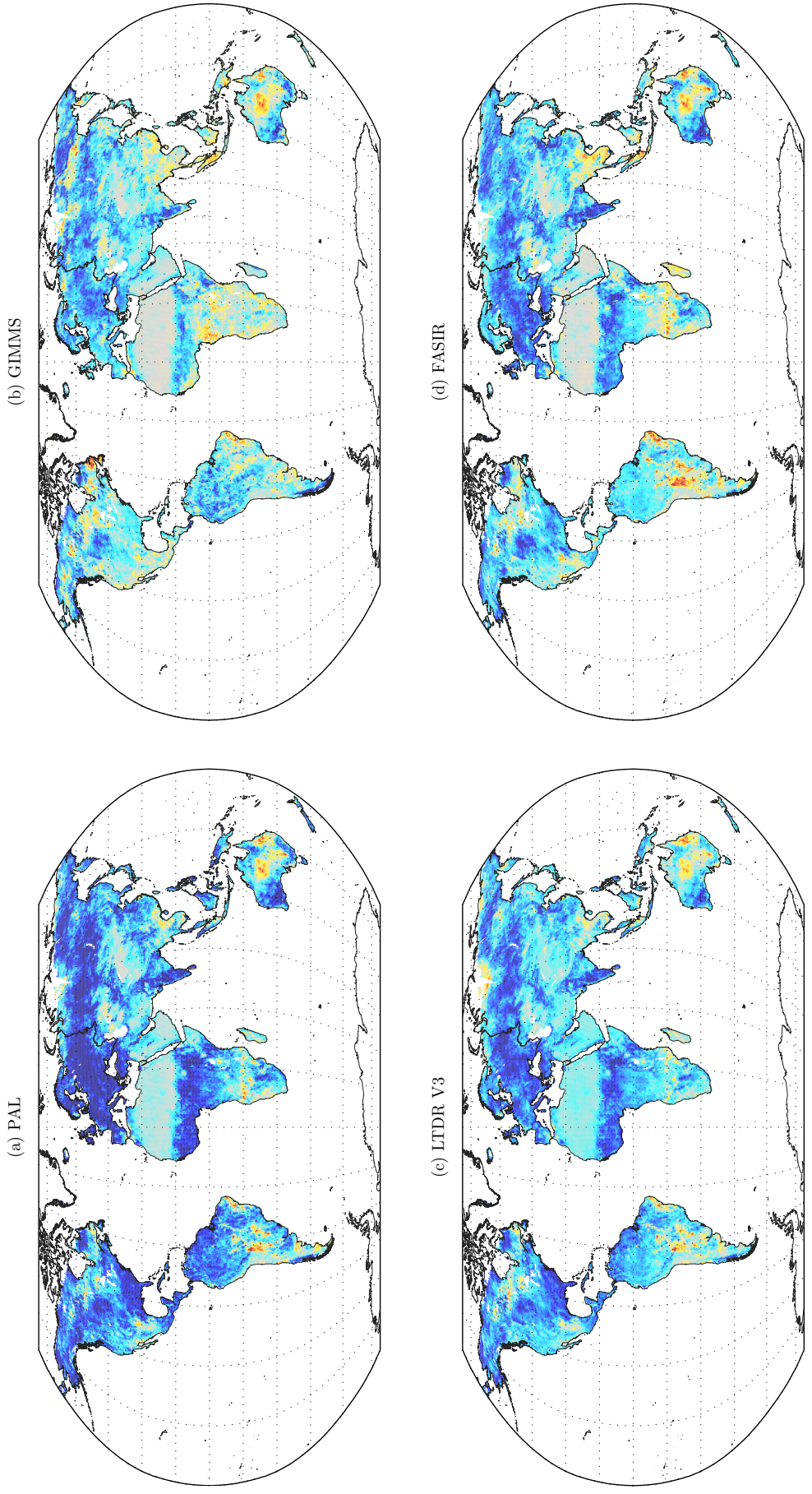


Figure 3.4: The 1982–1999 linear trend of annual mean NDVI time series for (a) the PAL, (b) GIMMS, (c) LTDR V3, and (d) FASIR datasets. All trends (i.e., both significant and insignificant trends at the 0.05 level) are shown.

Table 3.4: Means (μ) and standard deviations (σ) [NDVI units] which served for z-score transformation of the AVHRR-NDVI datasets (Eq. 3.2). Only pixels that have non-missing values for all four datasets were used in the calculation.

	PAL	GIMMS	LTDR V3	FASIR
μ	0.34	0.41	0.37	0.39
σ	0.20	0.22	0.17	0.22

Table 3.5: Coefficient of determination (R^2) of the spatial correlation of median, variance, and trend of annual mean NDVI time series between the AVHRR-NDVI datasets. All correlations were significant ($p < 0.05$; probabilities were corrected for autocorrelation).

		PAL	GIMMS	LTDR V3
Median	GIMMS	0.92		
	LTDR V3	0.96	0.89	
	FASIR	0.96	0.90	0.94
Variance	GIMMS	0.34		
	LTDR V3	0.47	0.21	
	FASIR	0.47	0.25	0.28
Trend	GIMMS	0.34		
	LTDR V3	0.80	0.36	
	FASIR	0.73	0.38	0.69

GIMMS dataset north of 67°N, which is also reflected in the "mosses and lichens" class (corresponding to the tundra) in Fig. 3.5f. Other than the latter no obvious biases are seen in the individual land-cover classes (Fig. 3.5f).

Table 3.5 shows R^2 values of the correlation between spatial maps of median, variance, and trend values for all pairs between the PAL, GIMMS, LTDR V3, and FASIR datasets, calculated from annual means of 1982 to 1999. All correlations were significant at the 0.05 level (significance levels were corrected for autocorrelation). In terms of median values the correlation pairs with the GIMMS dataset had the lowest R^2 values, possibly due to the aforementioned higher values of the GIMMS dataset observed north of 60°N. Similarly, in terms of variance and trends the correlation pairs with the GIMMS dataset show the lowest R^2 values.

Fig. 3.6 shows the results of the per-pixel assessment for the four z-score transformed PAL, GIMMS, LTDR V3, and FASIR datasets from 1982–1999 time series of the null hypotheses of equal medians, equal variance, equal trends, and no correlation. This type of analysis identifies specific regions where the four datasets show agreement/disagreement. Equal medians, equal variance, equal trends, and no correlation were found in 9.9 %, 45.5 %, 48.1 %, and 38.4 % of the land surface, respectively, using a significance level of 0.05 (Table 3.6). Most of the agreement in terms of median and trend is found in open-canopy land-cover types (grass/shrubs), whereas for the agreement in terms of variance, and the correlation level, no such relationship is found. However, the agreement in the "evergreen broadleaf forest" class (corresponding to the tropical forests), is notably lower for all statistics.

The same analysis is performed again for the same period for the two datasets that had negligable trends in deserts and thus seemed most stable: GIMMS and FASIR. Whilst the spatial patterns of agreement/disagreement are similar to the analysis involving all four AVHRR-NDVI datasets (cf. Fig. 3.6), equal medians, equal variance, equal trends, and no correlation were found in 23.7 %, 81.9 %, 76.2 %, and 32.2 % of the land surface, respectively, again using a significance level of 0.05 (Table 3.6). Thus, there is considerable improvement in the agreement in terms of median, variance, trend, and in the correlation, when only using the GIMMS and FASIR datasets compared to the analysis involving all four datasets, but the agreement in the "evergreen broadleaf forest" class (corresponding to the tropical forests) is still markedly lower compared to the other land-cover types.

Table 3.7 shows trends in annual mean NDVI time series for 1982 to 1999 for selected regions. Exceptional trend agreement was found for Australia, whereas good agreement for Asia and the tundra, and poor agreement for Europe, Africa, and the Sahel; cf. Fig. 3.6c. Moderate trend agreement was obtained for the remaining regions (Table 3.7). AVHRR-NDVI trends for all regions and all datasets, except the FASIR dataset in the Libyan desert, are positive.

3.4.2 Assessment against Landsat-NDVI

Fig. 3.7 presents scatter plots of the absolute-values comparison of NDVI from AVHRR and MODIS vs. NDVI from Landsat. The R^2 and RMSD statistics for the scatter plots of the four AVHRR-NDVI datasets against the Landsat-NDVI data ranged between 0.87–0.90 and 0.066–0.076, respectively (Fig. 3.7a–d). The same statistics for the MODIS-NDVI vs. Landsat-NDVI comparison were 0.90 (R^2) and 0.057 (RMSD); see Fig. 3.7e. Fig. 3.8 shows scatter plots of the comparison of changes in NDVI based on AVHRR-NDVI and MODIS-NDVI vs. Landsat-NDVI. The R^2 and RMSD statistics for the scatter plots of the four AVHRR-NDVI datasets against the Landsat-NDVI change data ranged between 0.55–0.61 and 0.077–0.083, respectively (Fig. 3.8a–d). The same statistics for the MODIS-NDVI vs. Landsat-NDVI change comparison were 0.62 (R^2) and 0.073 (RMSD); see Fig. 3.8e. Table 3.8 lists RMSD values of the linear least-squares regression for the entire dataset and for each separate land-cover type. Weighted-mean RMSD values (based on the number of samples in each land-cover class to remove potential bias in the data), are also calculated for each dataset. Based on the weighted-mean RMSD in the absolute comparison the LTDR V3 dataset is the best performing AVHRR-NDVI dataset (RMSD of 0.066; Ta-

Table 3.6: For each land-cover class the percentage of pixels with $p \geq 0.05$ (i.e., failure to reject the null hypothesis) for the null hypothesis of equal median (eq. median), equal variance (eq. var.), equal trend (eq. trend), and no correlation (no corr.), between all four datasets (PAL, GIMMS, LTDR V3, and FASIR) from 1982 to 1999 and the two most 'stable' datasets (GIMMS and FASIR), again from 1982 to 1999. Data were z-score transformed prior to the analysis. For pixels with a mix of classes the most dominant class was used. To improve interpretability, values higher than 50 % are shown in bold typeset.

	PAL, GIMMS, LTDR V3, and FASIR [%]				GIMMS and FASIR [%]			
Land-cover class	Eq. median	Eq. var.	Eq. trend	No corr.	Eq. median	Eq. var.	Eq. trend	No corr.
All classes	9.9	45.5	48.1	38.4	23.7	**81.9**	**76.2**	32.2
Evergreen needleleaf forest	17.2	33.6	41.4	29.8	35.4	**83.5**	**80.3**	41.9
Evergreen broadleaf forest	2.1	17.4	41.4	**81.2**	12.0	**55.3**	**69.3**	**82.6**
Deciduous needleleaf forest	21.2	**76.7**	**54.5**	26.5	27.6	**81.0**	**72.6**	33.6
Deciduous broadleaf forest	18.8	39.2	39.9	43.5	31.8	**89.5**	**73.2**	**50.3**
Mixed forest	30.0	30.7	30.8	29.9	44.7	**90.3**	**75.4**	40.2
Woodland	8.2	44.7	**53.1**	49.1	19.7	**83.8**	**70.7**	**50.1**
Wooded grassland	11.1	44.1	46.4	38.9	22.3	**81.9**	**61.0**	33.6
Closed shrubland	9.2	**68.5**	**79.7**	27.1	24.7	**93.0**	**88.0**	6.7
Open shrubland	7.0	**67.5**	**67.7**	28.5	27.8	**93.2**	**87.5**	8.2
Grassland	11.1	**67.3**	**62.9**	15.1	25.1	**91.1**	**84.0**	12.7
Cropland	11.9	37.2	33.6	31.5	28.7	**80.6**	**63.4**	33.3
Bare ground	0.8	24.0	15.8	**71.9**	15.2	**71.5**	**82.3**	11.0
Mosses and lichens	1.5	**67.6**	**72.2**	11.1	8.8	**82.6**	**83.1**	11.6

Table 3.7: For selected regions the annual trend from 1982 to 1999 in annual mean NDVI and the percentage of pixels with equal trends (eq. trend; i.e., $p \geq 0.05$ for the null hypothesis of equal trends) between all four AVHRR-NDVI datasets (PAL, GIMMS, LTDR V3, and FASIR). To improve interpretability, values higher than 50 % are shown in bold typeset.

	Trend [10^{-3} NDVI units yr^{-1}]				
Region	PAL	GIMMS	LTDR V3	FASIR	Eq. trend [%]
Globe	2.36	1.06	1.80	1.37	48.1
North America	2.53	1.04	2.10	1.44	47.2
South America	1.80	1.08	1.26	0.54	**52.8**
Europe	4.27	1.91	2.98	2.54	30.5
Asia	2.37	1.23	1.73	1.55	**56.8**
China	1.70	0.73	1.44	1.17	**51.5**
Africa	1.71	0.40	1.58	0.95	27.2
Australia	1.34	0.56	0.78	0.94	**87.4**
Sahel[a]	2.02	1.23	2.05	1.68	26.9
Hot spots of defor.[b]	1.92	0.66	1.22	0.47	48.7
Libyan desert[c]	0.29	0.02	0.53	−0.01	0.0
Tundra[d] (Eurasia)	2.40	1.71	1.42	1.69	**79.4**
Tundra[d] (N-A[e])	1.71	1.73	1.47	2.15	**62.6**
Boreal for.[f] (Eurasia)	3.69	1.60	2.51	2.10	48.6
Boreal for.[f] (N-A[e])	2.46	0.50	1.63	0.90	48.6

[a] Delineated using the World Wildlife Fund (WWF) terrestrial ecoregions map (downloaded from `http://www.worldwildlife.org/science/ecoregions/item1267.html`; Olson et al., 2001) by selection of the "Sahelian *Acacia* savanna" region.

[b] Tropical hot spots of deforestation (defor.) according to Achard et al. (2002).

[c] The south-eastern part of the Libyan desert (21°N–23°N and 28°E–29°E; Rao et al., 1993).

[d] Delineated using the WWF terrestrial ecoregions map by selection of all regions having "tundra" in the name.

[e] North America.

[f] All forest (for.) classes (from "evergreen needleleaf forest" to "mixed forest"; see Fig. 3.3) within the boreal biome, which was delineated using the WWF terrestrial ecoregions map (more details at `http://globalmonitoring.sdstate.edu/projects/gfm/boreal/data.html`).

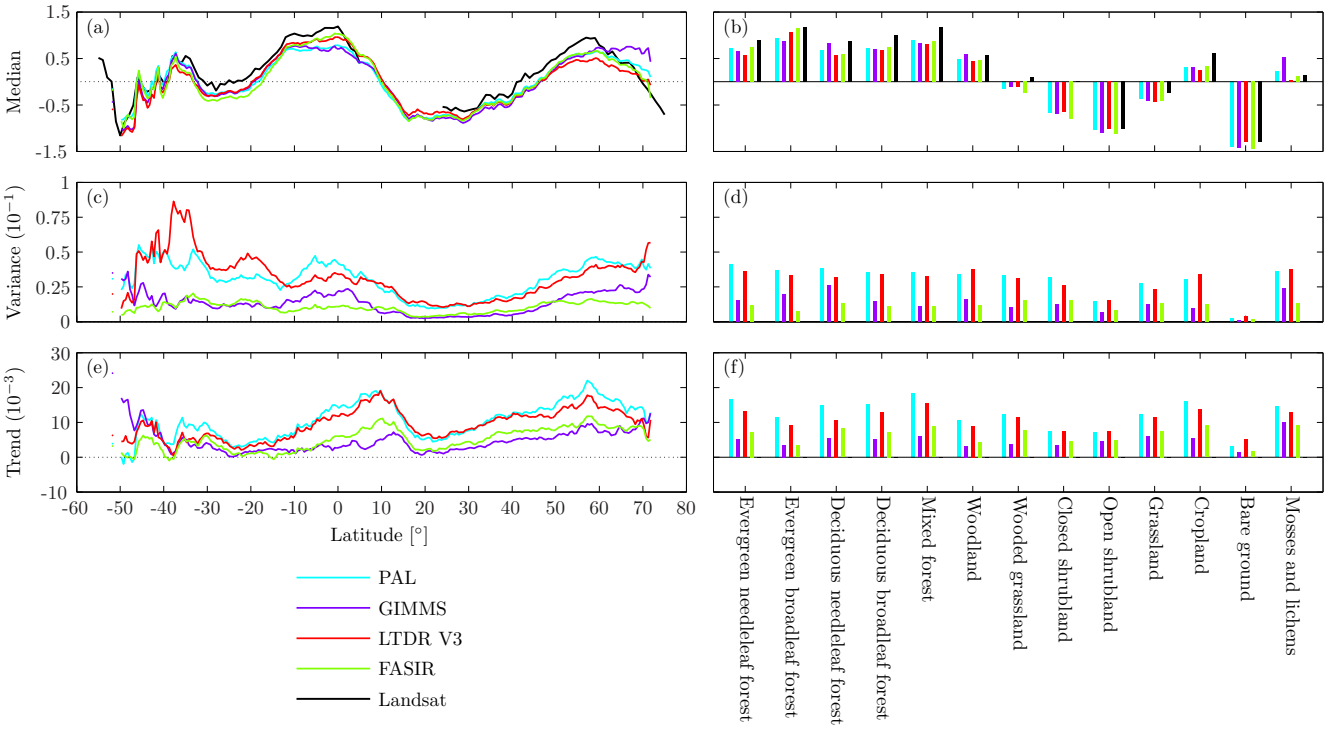


Figure 3.5: Median (a–b), variance (c–d), and trend (e–f) of annual mean NDVI time series for the PAL, GIMMS, LTDR V3, and FASIR datasets for 1982 to 1999, with (a), (c), and (e) showing means for each 0.5° latitude band and (b), (d), and (f) means for each land-cover class. For (a), (c), and (e) only 0.5° latitude bands that have data for more than 10 pixels were shown. The NDVI data have been z-score transformed prior to the analysis, and the values are therefore unitless. To revert the values back to NDVI units Eq. 3.2 and the coefficients listed in Table 3.4 can be used. The Landsat-NDVI values in (a–b) were converted to z-scores using the mean of the coefficients used for the AVHRR-NDVI z-score transformation (Table 3.4). In (a) Landsat-NDVI values were only plotted if the number of samples per 0.5° latitude band exceeded 50 % of the number of terrestrial integer longitudinal degrees, whereas in (b) Landsat-NDVI values were plotted if there were more than 400 Landsat samples in the land-cover class. Landsat samples pre-2000 as well as post-2000 were used. For pixels with a mix of classes the most dominant class was used.

ble 3.8) and FASIR worst (RMSD of 0.075; Table 3.8), whereas in the change comparison GIMMS performs best (RMSD of 0.076; Table 3.8) and FASIR worst (RMSD of 0.083; Table 3.8). All datasets appear to perform worse for closed-canopy land-cover types (i.e., from "evergreen needleleaf forest" to "woodland" in Table 3.8). In both the absolute and change comparisons for both instances of "all classes" and for almost every individual land-cover class MODIS-NDVI performs better than the four AVHRR-NDVI datasets (Table 3.8), confirming the robustness of the calibration of the Landsat-NDVI samples.

The question arises as to whether the relatively small differences between RMSD values represent statistically significant differences. To answer this question Table 3.9 lists p-values for the null hypothesis of equal means of the absolute residuals ($|e|$; Eq. 3.4) obtained in the comparisons between NDVI from AVHRR or MODIS vs. NDVI from Landsat. Using a significance level of 0.05, in the absolute-values comparison of NDVI from AVHRR or MODIS vs. NDVI from Landsat the GIMMS and FASIR datasets did not perform significantly different ($p \geq 0.05$; Table 3.9), whereas all the other combinations did differ significantly. In the change comparison the GIMMS dataset performed significantly better than only the FASIR dataset ($p < 0.05$; Table 3.9). The MODIS dataset performed significantly better than all datasets, except the GIMMS dataset in the change comparison ($p \geq 0.05$; Table 3.9).

The slope and intersect values of the regression equations shown in Fig. 3.7a–d and 3.8a–d were applied and mean AVHRR-NDVI values were calculated (not shown). An RMSD of 0.062 ($N = 6794$) was obtained for the regression comparing absolute values, which is an improvement in performance of 6 % compared to the RMSD of 0.066 obtained by LTDR V3, the best performing individual AVHRR-NDVI dataset (cf. Table 3.8). In the change comparison an RMSD of 0.073 ($N = 1424$) was obtained, which is a 5 % improvement in performance compared to the RMSD of 0.077 obtained by GIMMS, the best performing AVHRR-NDVI dataset in this respect (cf. Table 3.8). However, in both cases there was no significant difference (at the 0.05 level) between the means of the absolute residuals ($|e|$; Eq. 3.4) of the best performing AVHRR-NDVI dataset and the 'simple average' AVHRR-NDVI dataset.

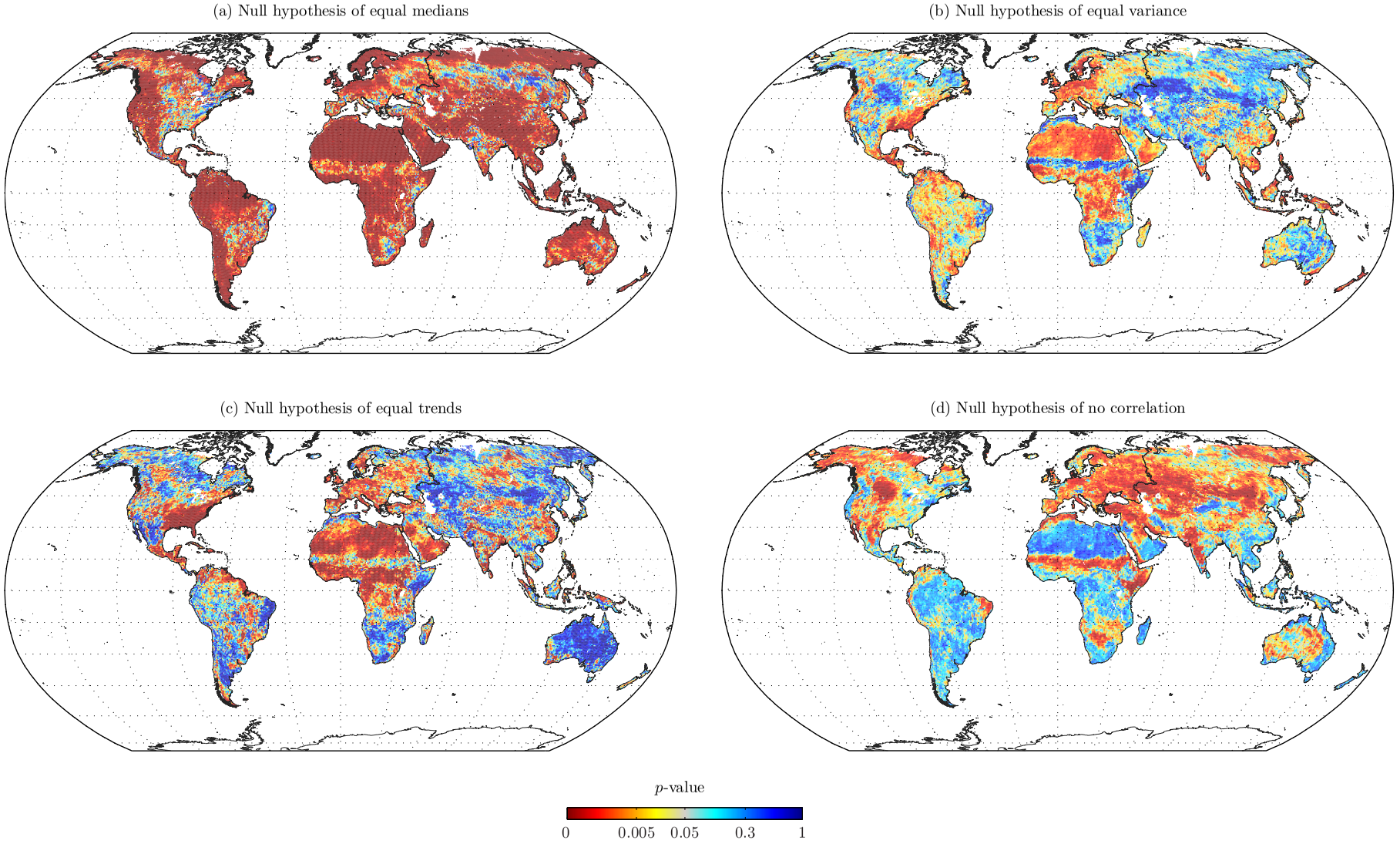


Figure 3.6: The p-values (probabilities) of the null hypothesis of (a) equal medians, (b) equal variance, (c) equal trends, and (d) no correlation, of annual mean time series for 1982–1999 amongst the PAL, GIMMS, LTDR V3, and FASIR datasets. The data have been z-score transformed prior to the analysis. The colormap was chosen such that blue indicates regions where there was failure to reject the null hypothesis ($p \geq 0.05$).

Table 3.8: RMSD statistics of the linear regression of mean NDVI derived from Landsat vs. the respective AVHRR-NDVI datasets and the MODIS-NDVI dataset, and RMSD statistics of the linear regression of the difference in mean NDVI as derived from Landsat vs. the respective AVHRR-NDVI datasets and the MODIS-NDVI dataset. The AVHRR-NDVI datasets were compared to the pre-2000 Landsat-NDVI samples, whereas the MODIS-NDVI dataset was compared to the post-2000 Landsat-NDVI samples; the number of samples in the Landsat-NDVI dataset pre-2000 and post-2000 are described in Table 3.3. To improve interpretability, RMSD values lower than 0.070 are shown in bold typeset. Units are Landsat-NDVI. The smaller the number the better the performance.

	RMSD of absolute-values comparison					RMSD of change comparison				
Land-cover class	PAL	GIMMS	LTDR V3	FASIR	MODIS	PAL	GIMMS	LTDR V3	FASIR	MODIS
All classes (weighted-mean)[a]	**0.069**	0.074	**0.066**	0.075	**0.055**	0.080	0.076	0.078	0.083	0.071
All classes (reg. with all data)[b]	0.070	0.076	**0.066**	0.075	**0.057**	0.080	0.077	0.079	0.083	0.073
Evergreen needleleaf forest	0.071	**0.068**	**0.068**	0.079	**0.058**	0.086	0.071	0.083	0.094	0.072
Evergreen broadleaf forest	0.072	0.073	**0.065**	**0.065**	**0.048**	0.099	0.098	0.093	0.111	0.081
Deciduous needleleaf forest	0.075	0.084	**0.069**	0.084	**0.068**	0.088	0.074	0.086	0.083	0.080
Deciduous broadleaf forest	0.081	0.085	0.076	0.089	**0.063**	0.079	0.073	0.083	0.089	0.077
Mixed forest	**0.070**	0.070	**0.066**	0.076	**0.057**	0.083	0.079	0.081	0.084	**0.070**
Woodland	0.073	0.082	0.074	0.077	**0.047**	0.097	0.090	0.089	0.095	**0.067**
Wooded grassland	**0.060**	0.075	**0.052**	**0.064**	**0.039**	0.075	**0.049**	**0.063**	**0.061**	**0.034**
Closed shrubland	**0.067**	**0.065**	**0.061**	0.074	**0.045**	**0.055**	0.073	**0.057**	**0.056**	**0.035**
Open shrubland	**0.041**	**0.040**	**0.040**	**0.044**	**0.028**	**0.037**	**0.043**	**0.042**	**0.040**	**0.024**
Grassland	**0.067**	0.072	**0.066**	0.078	**0.045**	0.072	0.080	**0.070**	0.072	**0.055**
Cropland	0.071	0.076	**0.069**	0.081	**0.057**	0.076	0.076	0.073	0.079	0.076
Bare ground	**0.020**	**0.019**	**0.020**	**0.024**	**0.062**	**0.030**	**0.030**	**0.027**	**0.029**	0.071
Mosses and lichens	0.100	0.105	0.095	0.121	**0.066**	0.101	0.100	0.105	0.109	0.081

[a] Weighted-mean RMSD over all land-cover classes. Weighting is based on the number of observations in each regression.
[b] RMSD from regression using data from all land-cover classes (Figs. 3.7 and 3.8 show the corresponding scatter plots).

Table 3.9: The t-test was used to calculate p-values (probabilities) for the null hypothesis of equal means between absolute residuals ($|e|$; Eq. 3.4) of the datasets, for the absolute-values and temporal-change comparison. This table can be used to assess whether a dataset performs significantly ($p < 0.05$) better or worse than another. The smaller the p-value the greater the likelihood that the datasets perform differently.

	p-values for absolute-values comparison				p-values for change comparison			
	PAL	GIMMS	LTDR V3	FASIR	PAL	GIMMS	LTDR V3	FASIR
GIMMS	< 0.001				0.095			
LTDR V3	< 0.001	< 0.001			0.533	0.292		
FASIR	0.014	0.381	< 0.001		0.352	0.010	0.121	
MODIS	< 0.001	< 0.001	< 0.001	< 0.001	0.008	0.227	0.033	< 0.001

3.5 Discussion

3.5.1 Intercomparison of AVHRR-NDVI datasets

Previous comparative research of AVHRR-NDVI datasets was primarily conducted at a regional scale using only a subset of the available datasets (Table 3.1). Accordingly, only a brief discussion of relevant studies is presented here. Our results are in accordance with those obtained in similar research for the Iberian Peninsula for 1982 to 1999 (Alcaraz-Segura et al., 2010b). Using an 8-km pixel size and a significance level of 0.05, Alcaraz-Segura et al. (2010b) found a similar trend for each of the four datasets over 43 % of their study area while distinguishing positive, negative, and no trend. The present study applied some additional statistical rigor to the trend agreement analysis but used data averaged to 0.5° spatial resolution; our results show consistent trends for 19 % of the Iberian Peninsula ($p \geq 0.05$; Fig. 3.6c). Alcaraz-Segura et al. (2010b) also reported trends derived from the GIMMS dataset to be the least similar to those obtained using the other AVHRR-NDVI datasets. This conforms with Table 3.5, which has the lowest R^2 values for correlation pairs with the GIMMS dataset, and a visual interpretation of Fig. 3.4, showing a distinctly different pattern for the GIMMS dataset. McCloy et al. (2005) analyzed the intra-annual correlation between the PAL, GIMMS, and FASIR datasets for 1982 to 1999, and found disagreement in the high northern latitudes (> 60°N). In contrast, the inter-annual correlation analysis amongst the four AVHRR-NDVI datasets presented here obtained mostly significant ($p < 0.05$) correlations for

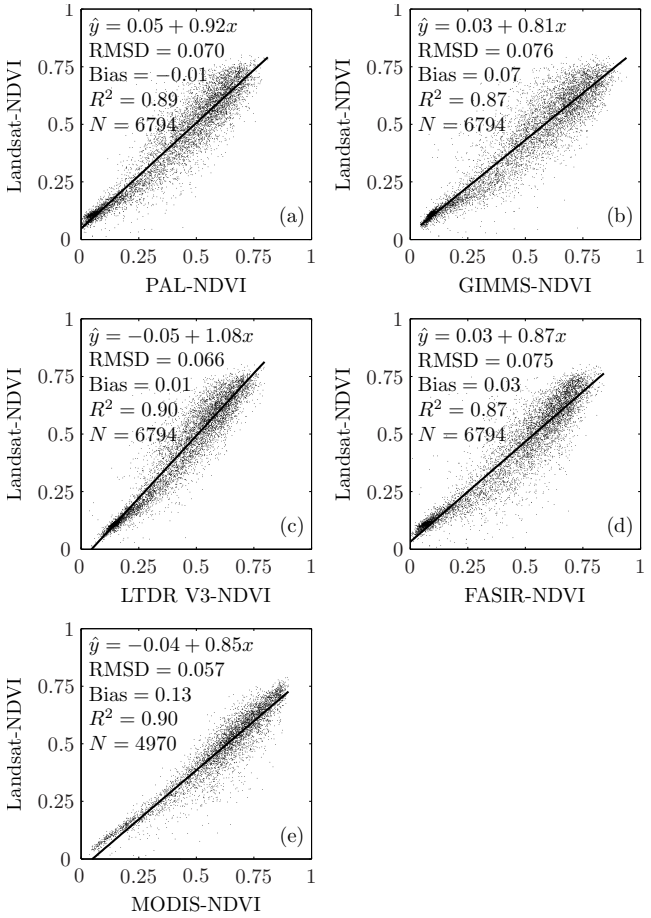


Figure 3.7: Scatter plots of mean NDVI derived from Landsat imagery vs. the (a) PAL, (b) GIMMS, (c) LTDR V3, (d) FASIR, and (e) MODIS datasets. Also shown are the linear regression equations and the best-fit line. The RMSD and bias have the units of the ordinate (or y-) axis.

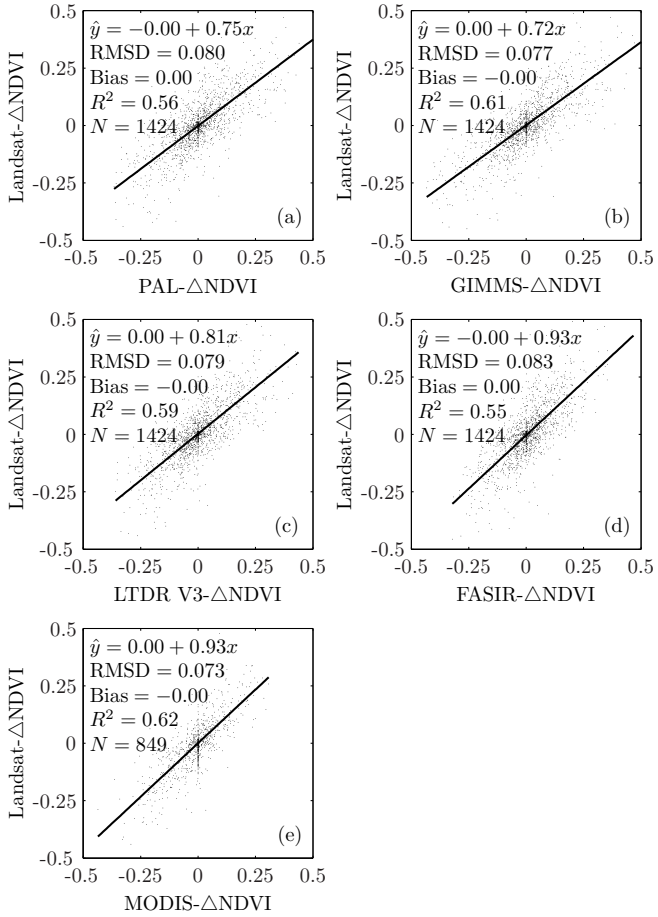


Figure 3.8: Scatter plots of the difference in mean NDVI between imagery acquired at different moments in time of Landsat vs. the (a) PAL, (b) GIMMS, (c) LTDR V3, (d) FASIR, and (e) MODIS datasets. Also shown are the linear regression equations and the best-fit line. The RMSD and bias have the units of the ordinate (or y-) axis.

this region. Since our analysis focuses on the variation between years, rather than within years, our results are affected by sensor degradation and orbital drift and may therefore be more suitable for studies involving long-term vegetation dynamics.

Hall et al. (2006) analyzed the difference in terms of absolute NDVI values between the GIMMS and FASIR datasets for the globe, and found large differences, with higher values found for the GIMMS dataset north of $\sim 50°$N, whereas the FASIR dataset had higher values in the tropical and boreal forests. Our results, based on z-score transformed data (Table 3.4), confirm their findings as we obtained different median values ($p < 0.05$) for most of the global land surface (Fig. 3.6a), considerably higher z-scores for the GIMMS dataset north of 60°N (Fig. 3.5a), and slightly higher z-scores for the FASIR dataset in the "evergreen broadleaf forest" and "evergreen needleleaf forest" classes (tropical and boreal forests, respectively); see Fig. 3.5b. The GIMMS dataset showed similar or higher z-scores in the other classes (Fig. 3.5b). The higher z-scores of the GIMMS dataset north of 60°N are likely due to the matching of GIMMS to SPOT-VGT data (Hall et al., 2006; Tucker et al., 2005). The higher z-scores found by the FASIR dataset in tropical forests are likely due to the selection of maximum NDVI in a 3×3 pixels moving window, and in boreal forests due to the replacement of values below the median October value (Hall et al., 2006; Table 3.2).

Slayback et al. (2003) compared the PAL, GIMMS, and FASIR datasets for the northern hemisphere for 1982–1998 and found: (1) the FASIR dataset to be less noisy in the temporal domain; and (2) increases in all three datasets for each globally averaged 10° latitude band investigated between 5°N–75°N. The present results agree with (1) and show FASIR to be less noisy, although in the spatial domain, than the PAL and GIMMS datasets, and similar to the LTDR V3 dataset (Fig. 3.4). Our analysis takes (2) a step further, as we show trends in z-scores for entire terrestrial latitudinal extent in Fig. 3.5e, and that positive trends are found in all four AVHRR-NDVI datasets north of 14°S.

Table 3.6 shows slightly poorer agreement between z-scores of the four AVHRR-NDVI datasets in terms of trends in dense-canopy cover types. The lesser agreement in the trends is likely attributable to BRDF vari-

ability, which is more pronounced in forests (Bicheron and Leroy, 2000), and to the non-linearity and saturation of NDVI in higher biomass ecosystems such as forests (Huete et al., 1997). The BRDF effect is exacerbated in boreal forests where large solar zenith angles occur frequently (Goward et al., 1991). In the tropics the agreement and the correlation is considerably worse (Brown et al., 2006), possibly due to water-vapor effects and contamination by clouds (Moulin et al., 1997), and differences in gap-filling and interpolation techniques between the datasets (cf. Table 3.2). It should be noted that NDVI is linearly related to fPAR (e.g., Sellers, 1985; Lu et al., 2003), and if one is interested in deriving LAI from remote sensing the simple ratio (defined as $\frac{\rho_{\mathrm{nir}}}{\rho_{\mathrm{r}}}$) may be more suitable (Lu et al., 2003). It is further noted that lack of correlation is desirable in deserts and tropical forests if none show changes over time (Slayback et al., 2003).

Trend estimates from previous studies (mostly single dataset, based either on a Local Area Coverage (LAC) AVHRR-NDVI dataset or on the PAL or GIMMS datasets) generally agree with our results, although here we quantify the uncertainty in the regional trends based on the per-pixel agreement of trends between the four differently processed AVHRR-NDVI datasets, thereby allowing previous studies to be put into context. Differences could arise from restriction of this analysis to 'growing season' months by rejection of pixels with $< 0°$ air temperature based on a monthly climatology. Table 3.7 presents trend estimates for 1982 to 1999 for the four AVHRR-NDVI datasets for selected regions. For most regions the trends derived from the GIMMS and FASIR datasets are the least pronounced and those from the PAL and LTDR V3 datasets the strongest (noting that the PAL and LTDR V3 datasets also show positive trends in desert sites considered to be radiometrically stable; cf. Fig. 3.4). For the entire globe (e.g., Tateishi and Ebata, 2004; Xiao and Moody, 2005) the mean increase in NDVI ranges from 1.06×10^{-3} to 2.36×10^{-3} NDVI units yr^{-1} (Table 3.7). Highly consistent trends between the respective AVHRR-NDVI datasets were found for Australia (Donohue et al., 2009), mostly consistent trends were found for Asia (e.g., Jeong et al., 2009; Park and Sohn, 2010; De Beurs and Henebry, 2004), the tundra, and the boreal forests, whereas less agreement was obtained for Europe (e.g., Stöckli and Vidale, 2004; Julien et al., 2006), Africa, and the Sahel (e.g., Eklundh and Olsson, 2003; Herrmann et al., 2005); see Table 3.7 and Fig. 3.6c. The lower agreement found in Africa and the Sahel is due to the large surface area occupied by deserts and sparse vegetation in which the PAL and LTDR V3 datasets have slightly positive trends (cf. Fig. 3.4). In North America (e.g., Hicke et al., 2002; Reed, 2006) moderate agreement between datasets is seen, as is also the case for South America (e.g., Paruelo et al., 2004; Sobrino et al., 2006). For all datasets the strongest greening trends are found for Europe (ranging from 1.91×10^{-3} to 4.27×10^{-3} NDVI units yr^{-1}; Table 3.7). The trends for almost all regions and all datasets are positive (Table 3.7), indicating increasing global vegetative cover. This suggests that the three primary climatic constraints on vegetation growth (temperature, radiation, and water availability; Nemani et al., 2003) have changed to create more favourable conditions for vegetation growth from 1982 to 1999, both globally and regionally. Additionally, for both agricultural and forest plantations their improved management, breeding, increasing levels of nutrient application, and expansion may also be contributing factors to positive NDVI trends in these areas.

The North American boreal forests exhibits less greening than the North American tundra, as evidenced by the most 'stable' datasets (GIMMS and FASIR; cf. Table 3.7). This conforms with previous studies, based on the GIMMS dataset (Goetz et al., 2005; Verbyla, 2008; Beck et al., 2011) and on a 1-km LAC dataset (Jia et al., 2003), that reported strong positive trends in the North American tundra. However, the studies based on the GIMMS dataset indicated negative trends for the North American boreal forests (Goetz et al., 2005; Verbyla, 2008; Beck et al., 2011), whereas our results show a weak positive trend of 0.50×10^{-3} NDVI units yr^{-1} for the GIMMS dataset (Table 3.7; albeit interior boreal forest areas in North America do exhibit negative trends, see Fig. 3.4b). It should be noted that for the GIMMS dataset (which is the AVHRR-NDVI dataset that best captures changes in the Landsat-NDVI samples; cf. section 3.5.2) the difference in trends found between the North American tundra and boreal forests is greater compared to the other AVHRR-NDVI datasets, as also observed in Fig. 3.5e where the GIMMS dataset showed relatively higher trends north of 67°N. Contrary to North America, based on the GIMMS and FASIR datasets, the Eurasian tundra and boreal forests appear to be greening at similar rates (cf. Table 3.7).

It is interesting to note that the 'hot spots' of tropical deforestation (Achard et al., 2002; Table 3.7) show an increase in NDVI for all four datasets. Possible reasons for this include: (1) rapid regrowth (Lugo and Scatena, 1995; Steininger, 1996) that quickly returns the NDVI to pre-clearing values; and (2) the spatial maximum-value compositing used by the datasets (Table 3.2) which favourably selects intact patches of forest corresponding with high NDVI values (Alcaraz-Segura et al., 2010a; Stow et al., 2007). To optimally detect deforestation hot spots, instead of looking at trends alone, image differencing techniques (e.g., Ingram and Dawson, 2005) using AVHRR-NDVI data have been implemented to identify such hot spots. Clearly, when using AVHRR-NDVI datasets to identify changes from full forest cover to converted or regenerating vegetation types (i.e., as opposed to a total clear-cut) considerable caution needs to be applied.

The positive trends in NDVI derived from the PAL and LTDR V3 datasets for the Libyan desert (Table 3.7), which is considered to be radiometrically stable

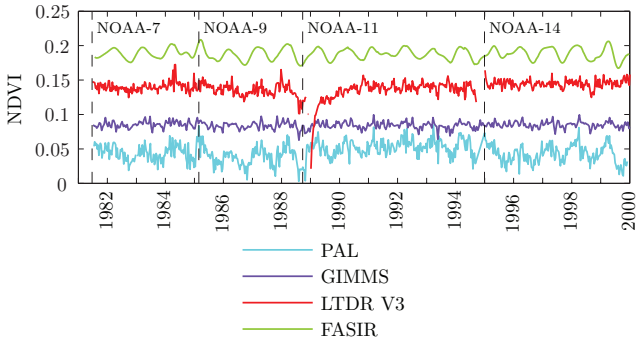


Figure 3.9: Mean NDVI time series for the south-eastern part of the Libyan desert (21°N–23°N and 28°E–29°E; Rao et al., 1993). The temporal resolution of the plotted PAL, GIMMS, LTDR V3, and FASIR time series are 10 days, 15 days, 10 days, and 10 days, respectively. The original LTDR V3 data were provided at a daily time step and were maximum-value composited to 10 days for the present study. To improve interpretability, the FASIR data were offset by +0.13 NDVI units. Also shown are the NOAA launch dates, which do not necessarily correspond with the sensor transitions of each AVHRR-NDVI dataset as listed in Table 3.2. The abscissa marks January 1st of each year.

(e.g., Staylor, 1990; Kaufman and Holben, 1993; Rao et al., 1993), indicates problems with the radiometric calibration of these datasets. Closer inspection of LTDR V3 10-day maximum-value time series for the Libyan desert (Fig. 3.9) revealed two causes: (1) at the beginning of the sensor switches from NOAA-9 to -11 (per Nov./1988) and from NOAA-11 to -14 (per Jan./1995) NDVI is under- and over-estimated, respectively; and (2) the presence of a positive trend (significant at the 0.05 level even when excluding the years 1988, 1989, and 1995, corresponding to the years affected by the aforementioned errors after the sensor switches). A second problem with the LTDR V3 dataset are the exceptionally high variance in annual means observed from 40°S to 30°S (Fig. 3.5c). Since the LTDR dataset is currently under development such problems may be solved during future releases. Problems with the calibration of PAL have been reported previously (e.g., De Beurs and Henebry, 2004).

Lastly, although the focus of this paper is on trends, we emphasize that trends almost never fully encompass all the variation in time series. Although our results show trends amongst the AVHRR-NDVI datasets to be significantly similar (at the 0.05 level) over more than half of the global land surface, this does not imply that they are the same in an absolute sense. Further, our definition of "similar" depends on the significance level chosen (0.05 in the present study). Our analysis advises on regions where one should be cautious about using a particular AVHRR-NDVI dataset. Clearly, the large regional differences obtained between the AVHRR-NDVI datasets (Fig. 3.6 and Table 3.7) highlight the importance of dataset choice for modelling the carbon balance of the terrestrial biosphere, as previously reported by Hall et al. (2006), and for studying the vegetation response to climate change.

3.5.2 Assessment against Landsat-NDVI

The availability of the FAO RSS database of Landsat imagery has enabled the first comprehensive global validation of AVHRR-NDVI datasets using an independent data source of 11 764 samples covering all terrestrial globally important land-cover types. The Landsat imagery is ideal for this evaluation as it is nearly devoid of temporal and spatial biases, due to the onboard calibration. The robustness of the calibration procedure is confirmed by the low RMSD values obtained in the comparison between MODIS-NDVI and Landsat-NDVI (Table 3.8) which lends confidence to the results obtained in the Landsat validation of AVHRR-NDVI. However, there is a discrepancy in the results, in that GIMMS has the second highest of the weighted-RMSD across all classes (0.074; Table 3.8) for the absolute-values comparison between the four AVHRR-NDVI datasets, and yet GIMMS also has the lowest weighted-RMSD across all classes (0.076; Table 3.8) for the change comparison when compared to the other three AVHRR-NDVI datasets. A possible explanation is that the GIMMS dataset contains spatial biases in terms of absolute values, but not in terms of temporal variance. Such biases in absolute values, that are not accounted for by the land-cover classification map used in the present study, have also been reported in Brown et al. (2006). The non-linear rescaling and offset of the GIMMS dataset to SPOT-VGT in the processing (Tucker et al., 2005) might be responsible, although details on the procedure have not been reported. Nevertheless, for studies using absolute AVHRR-NDVI values the LTDR V3 dataset is recommended, whereas for studies involving change, or trend analysis, the GIMMS dataset is recommended. However, since the strength of the AVHRR-NDVI datasets lies in their long temporal span, the ultimate test of performance for the datasets lies in the temporal-change analysis. Table 3.9. It should be noted that future releases of the LTDR V3 dataset might perform substantially better, as issues with the sensor calibration have been identified in Version 3 (cf. section 3.5.1). The higher RMSD values obtained using the FASIR dataset (Table 3.8) might be due to the lower spatial resolution (cf. Table 3.2). The poorer performance exhibited in closed-canopy land-covers (cf. Table 3.8) is likely due to the non-linearity and saturation effect of NDVI in higher biomass ecosystems such as forests (Sellers, 1985), exacerbated by excessive cloud contamination and water-vapor absorption in tropical forests (Moulin et al., 1997), large solar zenith angles in boreal forests (Goward et al., 1991), and differences in gap-filling and interpolation techniques between the datasets (cf. Table 3.2 and section 3.5.1).

A caveat of this analysis is the extrapolation of each dataset's performance in assessing temporal-change (i.e., the difference in NDVI between two dates) to the performance in assessing trends (i.e., the tendency of NDVI based on a time series at regular intervals). This extrapolation is probably justified since the analysis is based

on a large number of observations (i.e., 1424 and 849 observations for the AVHRR-NDVI datasets and MODIS-NDVI, respectively; Fig. 3.8), and because most Landsat samples were acquired in summer (i.e., 56.1 % of the Landsat samples $> 23.5°$N/S were acquired in the local summer), thereby limiting the influence of seasonality on the results. However, most of the Landsat samples only cover the NOAA-9 and -11 operational years, with results thus mainly dependent on the processing of these sensors.

Sensor and atmospheric effects are known to affect the reflectance signal reaching the satellite (e.g., Cracknell, 1997). The AVHRR-NDVI datasets (and the GIMMS dataset in particular) rely on maximum-value compositing to minimize the atmospheric effects (Holben, 1986). Due to the use of a similar length of the maximum-value compositing period in the datasets (15 days in the GIMMS dataset and 10 days in the other three AVHRR-NDVI datasets) the present study focuses on the quality of other aspects in the processing chain. Previous studies have summarized the potential influences of the atmospheric and sensor effects on the computed NDVI (Table 3.10). Total potential atmospheric errors for each AVHRR-NDVI dataset were calculated using the listed potential errors, assuming that applying a correction for a particular effect means there is no residual error due to that effect, amounting to 0.32, 0.41, 0.00, and 0.12 NDVI units for the PAL, GIMMS, LTDR V3, and FASIR datasets, respectively (ignoring the sign of the potential errors). The LTDR V3 dataset thus has the lowest total potential atmospheric induced error and the GIMMS dataset the highest. This is consistent with the absolute-values comparison of AVHRR-NDVI vs. Landsat-NDVI, but not with the change comparison (Table 3.8 and Figs. 3.7a–d and 3.8a–d), where the GIMMS dataset had the lowest RMSD. This could suggest that the GIMMS dataset's radiometric calibration and BRDF correction results in more accurate and consistent data than the methods applied to the LTDR V3 and FASIR datasets (cf. Table 3.2). It should be noted that the global application of BRDF models is precarious, as one needs to address the balance between parsimony and complexity, whilst accounting for potentially dubious input data controlling key variables (Verstraete et al., 1996).

Previously several studies have compared AVHRR-NDVI with Landsat-NDVI, but these studies were designed for specific applications and their small sample sizes need to be considered. Hall et al. (2006) compared Landsat surface reflectance NDVI (based on three Landsat scenes) with NDVI from the FASIR and GIMMS datasets and found mean absolute errors of 0.116 and 0.142 NDVI units, respectively, which is consistent with the results of the absolute-values analysis (Table 3.8). Brown et al. (2006) compared NDVI derived from 21 Landsat scenes (over land-cover types ranging from tundra to tropical forest) with NDVI from SPOT-VGT, MODIS, Sea-viewing Wide Field-of-view Sensor (SeaWiFS), and the GIMMS dataset, resulting in R^2 values of 0.63, 0.70, 0.62, and 0.81, respectively. These results imply optimal performance using the GIMMS dataset, contrary to our findings (Table 3.8). Pouliot et al. (2009) compared NDVI time series from 10 Landsat images acquired at the same location with NDVI derived from a 1-km LAC AVHRR-NDVI dataset, obtaining a mean absolute difference of < 7 %, which is about the same order of magnitude as the RMSD values found in our study (Table 3.8).

By averaging across the PAL, GIMMS, LTDR V3 (10-day maximum-value composited), and FASIR datasets a lower overall RMSD was obtained in the comparison with Landsat-NDVI, with the average of the datasets thus likely producing the best AVHRR-NDVI dataset currently available. This indicates that the errors in the datasets are to a certain degree unrelated, which is surprising as all four datasets are based on the AVHRR GAC level-1b archive (Kidwell, 1998). There are two possible explanations: the different correction schemes for sensor and atmospheric effects (cf. Table 3.2) introduce independent errors in the respective datasets; or the maximum-value compositing technique used in the datasets, where each dataset may select a different 4.4×1.1 km^2 pixel from the GAC archive to represent the final pixel value. Two other studies also suggested information may be lost due to maximum-value compositing (Alcaraz-Segura et al., 2010a; Stow et al., 2007).

3.6 Conclusion

We investigated the level of agreement for the period 1982–1999 between four AVHRR-NDVI datasets (PAL, GIMMS, LTDR V3, and FASIR). Additionally, these four AVHRR-NDVI datasets and MODIS-NDVI were validated against 11 764 Landsat-NDVI images sampling all terrestrial globally important land-cover types. The following summarizes the main conclusions reached herein:

1. There is general agreement amongst the four AVHRR-NDVI datasets in terms of trend over 48.1 % of the land surface. Equal medians, equal variance, and good correlations between the datasets were found over 9.9 %, 45.5 %, and 61.6 % of the land surface, respectively.
2. Mostly consistent trends are found for Australia and the tundra, whereas inconsistent trends are found for Europe, Africa, and the Sahel. Moderate agreement is found for the remaining regions.
3. The evaluation against Landsat-NDVI suggests that the LTDR V3 dataset is the most accurate in terms of absolute values and the GIMMS dataset the most accurate for assessing temporal change. MODIS-NDVI performed better than any AVHRR-NDVI dataset.
4. The average of the four AVHRR-NDVI datasets produces the best results when compared against Landsat-NDVI. The errors in the AVHRR-NDVI

Table 3.10: The magnitude of atmospheric and sensor effects (including BRDF) on the NDVI. Table adapted from Donohue et al. (2008).

		Correction applied to dataset?			
Source of variability	Likely error on NDVI	PAL	GIMMS	LTDR V3	FASIR
Atmospheric effects					
Molecular scattering	−0.23 over dense vegetation (El Saleous et al., 1997)	Yes	No	Yes	Yes
Ozone absorption	+0.06 over sparse vegetation (El Saleous et al., 1997)	Yes	No	Yes	Yes
Aerosol scattering	−0.20 over dense vegetation (Tanré et al., 1992)	No	Yes	Yes	Yes
Water-vapor absorption	−0.12 over sparse vegetation (El Saleous et al., 1997)	No	No	Yes	No
Sensor effects					
Radiometric calibration	Varies depending on processing and sensor	Yes	Yes	Yes	Yes
Orbital drift	Varies depending on processing and sensor	No	Yes	Yes	Yes

datasets thus appear to be to a certain degree unrelated.

Acknowledgments

The main part of this research was conducted when the senior author was a visiting scientist at CSIRO Land and Water, Canberra, Australia. The authors wish to acknowledge the generous support of Mr Garth Warren, Dr Randall Donohue, Dr Luigi Renzullo (all from CSIRO Land and Water), Mr Yi Liu (University of New South Wales, Sydney, Australia), and Dr Diego Miralles (VU University Amsterdam, Amsterdam, The Netherlands). We are grateful to Scott Goetz (Associate Editor of Remote Sensing of Environment) and three anonymous reviewers for their careful critiques, suggestions, and comments, which helped us to considerably improve the manuscript. We also thank Dr Sietse Los (Swansea University, Swansea, United Kingdom) for providing the FASIR dataset, the GIMMS team members for providing the GIMMS dataset, the LTDR team members for providing the LTDR V3 dataset, and the FAO RSS team members for providing the Landsat samples.

Chapter 4

The impact of forest regeneration on streamflow in 12 meso-scale humid tropical catchments[1]

Abstract. Although regenerating forests make up an increasingly large portion of humid tropical landscapes, little is known of their water use and effects on streamflow (Q). Since the 1950s the island of Puerto Rico has experienced widespread abandonment of pastures and agricultural lands, followed by forest regeneration. This paper examines the possible impacts of these secondary forests on several Q characteristics for 12 meso-scale catchments (23–346 km^2; mean precipitation 1720–3422 $mm\,yr^{-1}$) with long (33–51 yr) and simultaneous records for Q, precipitation (P), potential evaporation (PET), and land cover. A simple spatially-lumped, conceptual rainfall-runoff model that uses daily P and PET time series as inputs (HBV-light) was used to simulate Q for each catchment. Annual time series of observed and simulated values of four Q characteristics were calculated. A least-squares trend was fitted through annual time series of the residual difference between observed and simulated time series of each Q characteristic. From this the total cumulative change $\hat{A}$ was calculated, representing the change in each Q characteristic after controlling for climate variability and water storage carry-over effects between years. Negative values of $\hat{A}$ were found for most catchments and Q characteristics, suggesting enhanced actual evaporation overall following forest regeneration. However, correlations between changes in urban or forest area and values of $\hat{A}$ were insignificant ($p \geq 0.389$) for all Q characteristics. This suggests there is no convincing evidence that changes in the chosen Q characteristics in these Puerto Rican catchments can be ascribed to changes in urban or forest area. The present results are in line with previous studies of meso- and macro-scale (sub-)tropical catchments, which generally found no significant change in Q that can be attributed to changes in forest cover. Possible explanations for the lack of a clear signal may include: errors in the land-cover, climate, Q, and/or catchment boundary data; changes in forest area occurring mainly in the less rainy lowlands; and heterogeneity in catchment response. Different results were obtained for different catchments, and using a smaller subset of catchments could have led to very different conclusions. This highlights the importance of including multiple catchments in land-cover impact analysis at the meso scale.

4.1 Introduction

Tropical regions have experienced extensive changes in land use and land cover during the last few decades (Lepers et al., 2005). Continuously rising demands for crop land and timber have led to substantial deforestation in many regions (Drigo, 2005), and although the global tropical deforestation rate remains high at $13\,Mha\,yr^{-1}$ (FAO, 2006), forest regrowth on abandoned agricultural land is increasing, particularly in Latin America (Aide and Grau, 2004; Hecht, 2010) and South-East Asia (cf. Fox et al., 2000; Xu et al., 1999). Because these "secondary forests" account for approximately one-third of the total tropical forest area (Brown and Lugo, 1990; Hölscher et al., 2005), understanding the impact of forest regrowth on water yield is important for water resources management and planning purposes (Giambelluca, 2002; Bruijnzeel, 2004) and the development of viable "Payments for Ecosystem Services" schemes (Landell-Mills and Porras, 2002; Lele, 2009). However, despite this recognized importance, little is known of the water use of secondary tropical forests, although there are indications of enhanced water use during the period of most active biomass accumulation (Giambelluca, 2002; Juhrbandt et al., 2004; Hölscher et al., 2005).

The relationship between forest cover and streamflow (Q) is subject to a long-standing and ongoing discussion (Andréassian, 2004), also in the tropics (Bruijnzeel, 2004; Calder, 2005). The influence of forest cover change

[1]This chapter is an edited version of: Beck, H. E.; Bruijnzeel, L. A.; van Dijk, A. I. J. M.; McVicar, T. R.; Scatena, F. N., and Schellekens, J. The impact of forest regeneration on streamflow in 12 meso-scale humid tropical catchments. *Hydrology and Earth System Sciences*, 17(7):2613–2635, 2013.

on flooding is particularly contentious (e.g. FAO, 2005; Bradshaw et al., 2007; Van Dijk et al., 2009) whereas the effect of forestation on tropical dry-season flows is also under debate (Calder, 2005; Scott et al., 2005). The general contention is that the *net* effect of an increase in forest cover on dry-season flow depends on the "trade-off" between *increases* in Q due to enhanced soil water recharge on the one hand (as forestation generally increases soil macroporosity and infiltration characteristics; Ilstedt et al., 2007; Bonell et al., 2010; Zimmermann et al., 2006, 2010; Hassler et al., 2011), and *decreases* in soil water reserves and Q on the other hand due to the higher water use of trees compared to crops, pasture, or scrubs (Bruijnzeel, 1989, 2004; Jackson et al., 2005; Scott et al., 2005). Reviews of micro-scale ($< 1\,\mathrm{km}^2$) experimental catchment studies (e.g. Jackson et al., 2005; Brown et al., 2005) – mostly conducted outside the tropics and in non-degraded settings where soil infiltration characteristics are not likely to be improved substantially by forestation – suggest that the increase in vegetation water use is indeed more important, and thus an increase in forest cover commonly leads to a decrease in both total and dry-season Q. However, although direct experimental evidence of the "infiltration trade-off hypothesis" (Bruijnzeel, 1989; Bonell et al., 2010) is missing due to a lack of comprehensive studies, demonstrated reductions in amounts of headwater- or hillslope stormflow production after reforesting severely degraded land in various parts of the tropics (e.g. Chandler and Walter, 1998; Zhou et al., 2002; Zhang et al., 2004) should be large enough to overcome the associated increases in forest water use (Chandler, 2006; cf. Bruijnzeel, 2004; Scott et al., 2005; Zhou et al., 2010). Indeed, as long-term Q records for large, once degraded catchment areas are becoming available, evidence of improved baseflows (Q_{bf}) following large-scale land rehabilitation is beginning to be documented (Wilcox and Huang, 2010; Zhou et al., 2010).

The tropical island of Puerto Rico provides a unique opportunity to study the impacts of natural forest regeneration on Q at the meso-catchment scale (1–10 000 km^2), as high-quality and long-term hydroclimatic records and sequential land-cover data are available. Since the 1950s, Puerto Rico has seen widespread secondary forest regrowth on abandoned pastures, agricultural land (mostly sugar cane), and coffee plantations (Thomlinson et al., 1996; Aide et al., 2000; cf. Del Mar López et al., 1998). Although previous work on the relationship between land-cover change and Q using lumped meso- and macro-scale catchment data has experienced some difficulty demonstrating unequivocal results (e.g. Van Dijk et al., 2012), possibly stronger conclusions may be obtained by selecting catchments within similar regions (in this case Central Puerto Rico; cf. Peña-Arancibia et al., 2010). On the whole, one would expect marked drops in total Q and Q_{bf} during forest recovery in areas where the general extent of soil surface degradation before land abandonment was limited and soil structural characteristics (and thus infiltration opportunities) therefore remained relatively unaffected by forest regeneration (cf. Aide et al., 1996; Zou and Gonzalez, 1997). For catchments that experienced advanced soil degradation prior to agricultural abandonment, major declines in the volumes of both total Q and quickflow (Q_{qf}) would be expected during forest regrowth due to much improved infiltration and retention capacities (cf. Chandler and Walter, 1998; Zhou et al., 2002). The direction and magnitude of the change in Q_{bf} will depend on the trade-off between the changes in vegetation water use and infiltration associated with forest regeneration (Bruijnzeel, 1989; Scott et al., 2005).

The following hypotheses are tested here: (1) there is a negative relationship between the area under regenerating forest and the change in total Q (i.e. $Q_{\mathrm{qf}} + Q_{\mathrm{bf}}$); (2) Q_{qf} shows a negative relationship with area under regenerating forest and a positive one with area under urbanization; and (3) depending on the trade-off between the changes in vegetation water use and infiltration associated with forest regrowth, Q_{bf} shows either a negative, no, or a positive relationship with the area under regenerating forest. Specific objectives are to quantify the effects of forest regeneration and urbanization on total Q, Q_{qf}, and Q_{qf}.

4.2 Study area

Puerto Rico is an island with a tropical maritime climate, located in the north-eastern Caribbean occupying $\sim 8870\,\mathrm{km}^2$. The geology is dominated by the volcanic Cordillera Central, with a few major outcrops of plutonic rock (mostly granodiorite), and karstic limestones towards the far north and south (Olcott, 1999). Soils developed in the volcanic substrates are largely clayey Ultisols with a rapidly diminishing saturated hydraulic conductivity (K_{s}) with depth (Schellekens et al., 2004) whereas the granodiorites produce less clayey Ultisols with a less pronounced K_{s} profile (Kurtz et al., 2011). Island-wide mean annual precipitation (P) is $\sim 1700\,\mathrm{mm\,yr}^{-1}$ (Daly et al., 1994, 2003). There is a moderate P seasonality, with the three driest months of the year (January–March) receiving on average $\sim 300\,\mathrm{mm}$ in total and the three wettest months of the year (September–November) receiving on average $\sim 775\,\mathrm{mm}$ in total (Daly et al., 1994, 2003). The northern and eastern portions of the island receive $\sim 30\,\%$ more P due to the rising of the moisture-bearing trade-winds against the slopes of the central mountain range (Calvesbert, 1970; García-Martinó et al., 1996; Daly et al., 2003).

During the second half of the 20th century, socio-economic changes in Puerto Rico led to migration from (upland) rural areas and to (lowland) ubanization (Dietz, 1986). The associated abandonment of pastures and agricultural fields allowed secondary forests to develop over increasingly large areas as time progressed (Thomlinson et al., 1996; Grau et al., 2003; Helmer, 2004; Parés-

Ramos et al., 2008; Fig. 4.1). The dynamics of this forest recovery are well-documented, both in terms of its expansion over time, and forest composition and structure (Aide et al., 1996, 2000; Chinea and Helmer, 2003; Grau et al., 2003). Tree density typically reaches a peak between 25 and 35 years after abandonment whereas species richness and forest structure resemble those of old-growth forest after ca. 40 yr of regeneration (Aide et al., 2000). Total actual evaporation (ET_a) of mature upland forest in the maritime tropical climate of Puerto Rico is high compared to tropical continental sites (Schellekens et al., 2000), mostly because of enhanced wet-canopy evaporation rates (Holwerda et al., 2006, 2012), and the same may well apply to the island's secondary forests (cf. Giambelluca, 2002).

The locations of the 12 catchments examined here and the respective changes in land cover over the period 1951–2000 (see also section 4.3.1) are shown in Fig. 4.1. The size of the catchments ranges from 23 to 346 km^2 (median size 42 km^2), mean annual P varies between 1720 and 3422 $mm\,yr^{-1}$ (median value 2021 $mm\,yr^{-1}$), and the length of simultaneous P, potential evaporation (PET), and Q records between 33 and 51 yr (median duration 44 yr; Table 4.1).

4.3 Data

4.3.1 Land cover

Land-cover maps were obtained for 1951, 1978, 1991, and 2000 (Fig. 4.1). The maps for 1951 (Brockmann, 1952; Kennaway and Helmer, 2007) and 1978 (Ramos and Lugo, 1994) were derived from aerial photography. Although the maps for 1951 and 1978 were rasterized at a resolution of ∼30 and ∼11 m, respectively, the actual mapping resolution used by the photo interpreters is estimated at ∼300 and ∼50 m, respectively. The 1991 and 2000 maps[2] (∼30-m resolution; Helmer and Ruefenacht, 2005) were derived from Landsat data using regression-tree modelling and histogram matching. All land-cover maps were reprojected to a common 0.0001° (∼11 m) geographical grid by nearest-neighbour interpolation. To accommodate the different classification schemes used in the respective mapping exercises, each land-cover class was assigned to a generalized class (Table 4.2). For each catchment the net changes in urban and forest areas from the start of the simultaneous P, PET, and Q records (Table 4.1) until 2000 were calculated by linear interpolation of urban- and forest-area time series.

4.3.2 Streamflow

All available daily Q records were downloaded from the US Geological Survey (USGS) National Water Information System[3] in December 2012, resulting in an initial dataset of 111 gauging sites. Catchment areas were derived for each site using the PCRaster software (Wesseling et al., 1996) and the USGS National Elevation Dataset (NED) digital elevation model (∼30-m resolution). The following criteria for inclusion here were applied: (1) the USGS published estimate of catchment area deviated by $< 10\,\%$ from our computed catchment area; (2) the length of the Q record between 1950 and 2005 was > 30 yr; and (3) the catchment was not subject to flow regulation or affected by major anthropogenic water extraction. The latter was assessed using annual USGS Water-Data Reports[4] and a map of water supply intakes in the Luquillo Experimental Forest (Crook et al., 2007; an island-wide map of intakes is lacking). The final dataset comprised 12 catchments (Fig. 4.1 and Table 4.1). For the conversion of measured discharge [$feet^3\,s^{-1}$] to areal mean Q [$mm\,d^{-1}$] the computed catchment area was used. The following four Q characteristics were calculated, on an annual basis, to study changes in Q through time: (1) the annual 95th percentile (percent time not-exceeded) daily Q (Q_{p95} [$mm\,d^{-1}$]; indicative of peak flows); (2) the annual mean Q (Q_{tot} [$mm\,yr^{-1}$]; indicative of total water yield); (3) the annual 5th percentile daily Q (Q_{p5} [$mm\,d^{-1}$]; indicative of low flows); and (4) the annual mean dry-season (January–March) flow (Q_{dry} [$mm\,yr^{-1}$]).

For four catchments the daily Q strongly exceeded the daily P (see section 4.4.1) on one or more days, indicating errors in the Q and/or P data. To prevent such errors from biasing the results, for some catchments parts of the Q record were excluded from the analysis. Specifically, for the Bauta catchment data for 1996–1998 were excluded, for the Grande de Loíza catchment data prior to 1961 were excluded, for the Fajardo catchment data for 1989 were excluded, and for the Inabón catchment data for 1975 were excluded.

4.3.3 Precipitation

Daily P data from the Global Historical Climatology Network-Daily (GHCN-D) database[5] (Gleason, 2002) and from the El Verde station in north-eastern Puerto Rico[6] were used. After quality checks, the entire record from the Cerro Maravilla station (included in the GHCN-D database) and 1991–1994 data from the El Verde station were excluded from the analysis. Stations with a record > 20 yr were selected from the GHCN-D database, resulting in a dataset comprising 70 stations (including the El Verde station). Figure 4.2 presents the number of P observations available for each day between 1955 and 2010. In addition, a map[7] of mean annual P derived using the Precipitation-elevation Regressions on

[2] Downloaded from `https://www.sas.upenn.edu/lczodata/` in June 2011.

[3] Accessible at `http://waterdata.usgs.gov/nwis`.

[4] Available at `http://wdr.water.usgs.gov/`.

[5] Downloaded from `ftp://ftp.ncdc.noaa.gov/pub/data/ghcn/daily/` in December 2010.

[6] Downloaded from `http://luq.lternet.edu/data/lterdb14/data/evrain.htm` in July 2011.

[7] Downloaded from `http://www.wcc.nrcs.usda.gov/ftpref/support/climate/prism/` in September 2011.

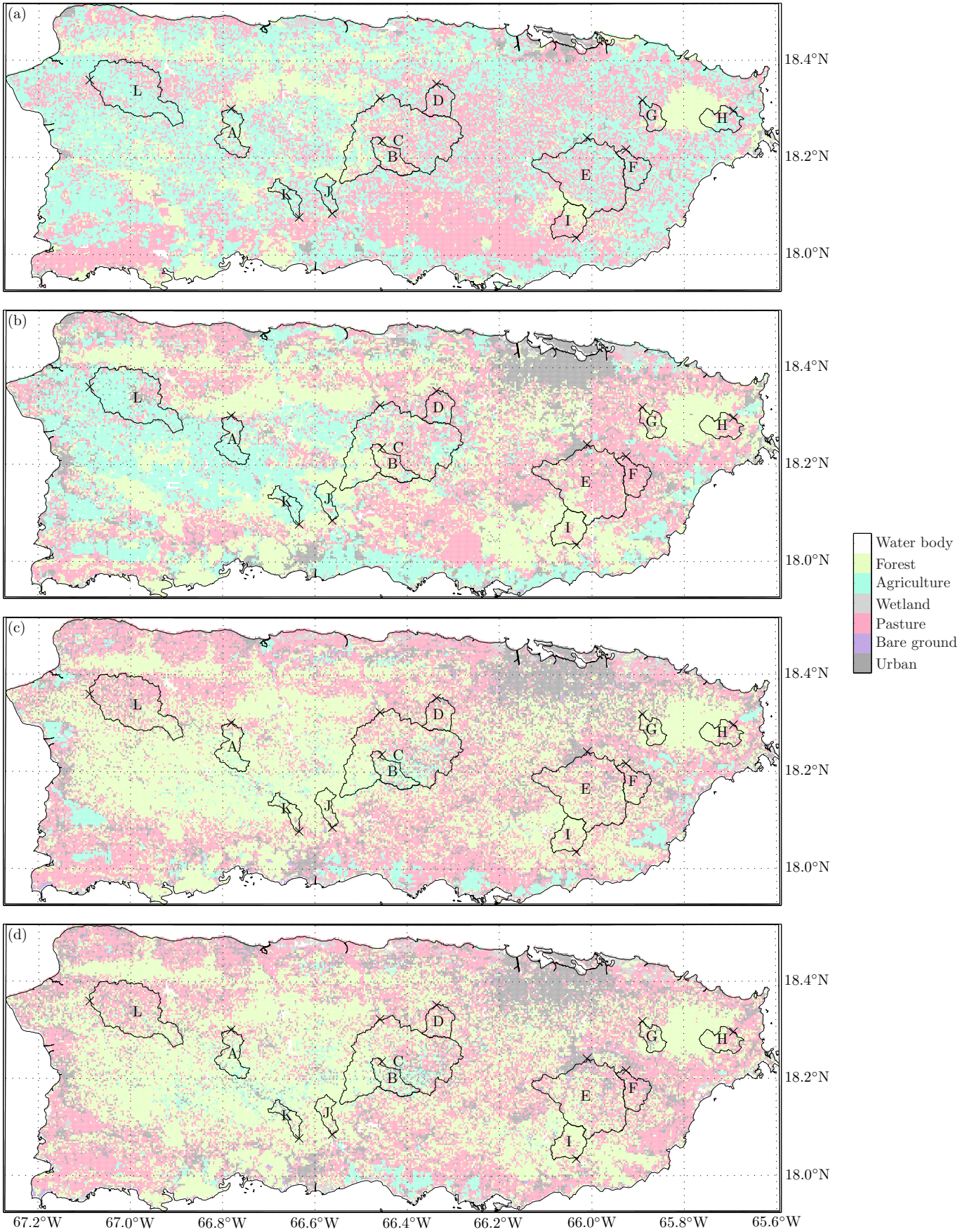


Figure 4.1: Maps showing the locations of the study catchments in Puerto Rico with the generalized land-cover classifications for: (a) 1951; (b) 1978; (c) 1991; and (d) 2000. Maps for 1951 and 1978 courtesy of E.H. Helmer and O.M. Ramos (USDA-IITF), respectively. Catchments identified by letters and stream gauges by crosses. All maps in this paper are presented in the Albers equal area conic projection with latitudinal limits 17.93°N–18.52°N and longitudinal limits 67.27°W–65.59°W.

Table 4.1: Background details and basic hydrological characterization of the catchments used in the present study. The catchments are identified by the letters A–L in the header.

	A: Tanamá	B: Bauta	C: Grande de Manatí	D: Cibuco	E: Grande de Loíza	F: Valenciano	G: Canóvanas	H: Fajardo	I: Grande de Patillas	J: Inabón	K: Portugués	L: Culebrinas
Catchment details												
USGS ID	028000	034000	035000	038320	055000	056400	061800	071000	092000	112500	115000	147800
Gauge latitude [°N]	18.301	18.236	18.323	18.352	18.242	18.217	18.318	18.296	18.035	18.086	18.078	18.360
Gauge longitude [°W]	66.782	66.455	66.459	66.336	66.010	65.925	65.889	65.692	66.034	66.562	66.634	67.089
Catchment area[a] [km^2]	47.1	43.5	345.9	39.8	233.0	41.5	26.5	39.3	46.3	24.5	22.9	177.1
Catchment mean elevation [m a.s.l.]	571	734	578	247	265	170	461	272	442	558	568	167
Catchment elongation ratio[b] [-]	0.66	0.57	0.69	0.84	0.73	0.72	0.71	0.69	0.76	0.61	0.52	0.62
Catchment mean slope [°]	16.2	21.5	19.6	13.1	13.3	8.9	14.4	16.0	20.1	24.1	22.8	10.0
Record[c] start year	1960	1970	1957	1970	1960	1971	1968	1962	1966	1965	1965	1968
Record[c] end year	2010	2010	2010	2006	2010	2010	2010	2009	2010	2010	1997	2010
Record[c] length [yr]	51	34	51	35	51	40	43	48	45	44	33	43
Dominant crop(s) in 1951[d]	c	c	c	c, sc, p	c, sc, f	c, sc, f	c, sc, f	sc	c, sc, f	c	c	c, sc
Dominant geology[e]	v, gd, ls	v	v, gd	v, gd	v, gd, q	gd	v	v, gd, q	v, gd	v, gd	v, gd	v, ls
Long-term catchment means of hydroclimatic parameters calculated for record[c] period												
Baseflow recession constant (k_{bf} [mm d^{-1}])	0.956	0.942	0.949	0.935	0.940	0.927	0.943	0.919	0.925	0.926	0.929	0.959
Precipitation (P [mm yr^{-1}])	2102	2038	2004	1827	1963	2039	3422	2875	1720	1781	1875	2178
Total streamflow (Q_{tot} [mm yr^{-1}])	953	812	653	642	845	1067	942	1529	1185	681	743	1509
Baseflow (Q_{bf} [mm yr^{-1}])	680	389	341	320	420	394	463	656	601	441	388	673
Quickflow (Q_{qf} [mm yr^{-1}])	273	423	312	322	426	673	480	872	584	241	354	836
$P - Q$ [mm yr^{-1}]	1149	1225	1351	1185	1118	972	2480	1347	535	1100	1132	669
Potential evaporation (PET [mm yr^{-1}])	1487	1407	1456	1535	1542	1548	1399	1449	1514	1502	1519	1658
Runoff coefficient (Q_{tot}/P [-])	0.453	0.399	0.326	0.351	0.431	0.523	0.275	0.532	0.689	0.383	0.396	0.693
Aridity index (P/PET [-])	1.42	1.46	1.38	1.18	1.28	1.33	2.46	1.97	1.15	1.18	1.23	1.31
Baseflow index[f] (Q_{bf}/Q_{tot} [-])	0.714	0.479	0.522	0.499	0.496	0.369	0.491	0.429	0.507	0.647	0.522	0.446

[a] Computed catchment area; see section 4.3.2.
[b] Calculated by dividing the diameter of a circle having the same area as the catchment, by the maximum length of the catchment (Schumm, 1956).
[c] Record is defined as the period of simultaneous P, PET, and Q data. Several catchments have gaps in their record.
[d] After Brockmann (1952) and Kennaway and Helmer (2007). Explanation of abbreviations used: c, shade coffee; sc, sugar cane; f, fruits and orchards; p, pineapple. See section 4.3.1 for details.
[e] After Olcott (1999). Explanation of abbreviations used: v, mixed volcanic; gd, granodiorite; ls, limestone; q, Quaternary alluvium. See section 4.2.
[f] Separated using a forward- and backward-recursive filter (Van Dijk, 2010).

Table 4.2: Generalized land-cover classes as used here and their relation to the specific land-cover classes distinguished in the 1951, 1978, and 1991/2000 land-cover maps. The numbers correspond to the values for the classes used in the original land-cover maps.

Generalized classes	1951 map	1978 map	1991/2000 maps
Urban	1. Urban and coastal sand	12. Urban	1. High-medium density urban 2. Low-medium density urban
Bare ground		10. Rocky areas	20. Salt or mud flats 25. Quarries 26. Coastal sand and rock 27. Bare soil (including bulldozed land)
Pasture	4. Pasture and grass	3. Pasture	5. Pasture, hay or inactive agriculture 6. Pasture, hay or other grassy areas
Wetland	6. Non-forested wetlands	8. Mangroves 9. Non-forested wetland	19. Emergent wetlands including seasonally flooded pasture 21. Mangrove 23. Pterocarpus swamp
Agriculture	2. Herbaceous agriculture and hay 3. Coffee and mixed woody agriculture	2. Agriculture	3. Herbaceous agriculture 4. Active sun coffee and mixed woody agriculture
Forest	5. Forest, woodlands, shrublands, and forested wetlands 12. Unknown	4. Dense canopy forest 1 5. Dense canopy forest 2 6. Less dense canopy forest 7. Woodland and shrubland	7. Drought deciduous open woodland 8. Drought deciduous dense woodland 9. Deciduous, evergreen coastal and mixed forest or shrubland 10. Semi-deciduous and drought deciduous forest on alluvium and non-carbonate substrates 11. Semi-deciduous and drought deciduous forest on karst/limestone (includes semi-evergreen forest) 12. Drought deciduous, semi-deciduous and seasonal evergreen forest on serpentine 13. Seasonal evergreen and semi-deciduous forest on karst 14. Seasonal evergreen and evergreen forest 15. Seasonal evergreen forest with coconut palm 16. Seasonal evergreen and evergreen forest on karst 17. Evergreen forest on serpentine 18. Sierra palm, transitional and tall cloud forest 22. Seaonally flooded savannahs and woodlands 24. Tidally flooded evergreen dwarf-shrubland and forb vegetation 29. Elfin and sierra palm cloud forest
Water body	10. Water/other	11. Water body	0. Water 28. Water (permanent)

Independent Slopes Model (PRISM) method (based on data from 1963–1995; Daly et al., 1994, 2003) was used to ensure reliable long-term, elevation-corrected P for the catchments. The method used to obtain time series of daily P for each catchment is described in section 4.4.1.

4.3.4 Minimum and maximum air temperature

Daily minimum and maximum air temperature ($T_{\min}$ and $T_{\max}$, respectively [°C]) were used to compute daily time series of PET for each catchment. $T_{\min}$ and $T_{\max}$ time series ($>20\,\text{yr}$) are available for 21 stations in Puerto Rico within the GHCN-D database (Gleason, 2002). Figure 4.2 presents the number of daily $T_{\min}$ and $T_{\max}$ observations between 1955 and 2010. Maps of mean annual $T_{\min}$ and $T_{\max}$ from the WorldClim dataset[8] ($\sim$1-km resolution; Hijmans et al., 2005) were also used to ensure reliable long-term, elevation-corrected $T_{\min}$ and $T_{\max}$ for the catchments.

4.4 Methodology

Fig. 4.3 presents a flow diagram summarizing the various steps that were carried out to investigate whether changes in forest and/or urban area have influenced the observed Q characteristics. The various steps will be explained in detail hereafter.

4.4.1 Spatio-temporal interpolation and rescaling of climatic variables

Having reliable catchment-mean time series of the climatic variables (P, $T_{\min}$, and $T_{\max}$) is important to prevent spurious trends in the simulated Q from influencing the results. To calculate time series of the climatic variables from 1955 (marking the start of many records) to 2010 for the catchments an approach was developed that: (1) removed the linear trend from the time series for each station and variable; (2) spatially interpolated the trend and daily-irregular components; and (3) reunited them on a per-pixel basis; after which (4) the time series were rescaled. More specifically, the following steps were carried out for each variable (where X denotes the variable in question):

1. $T_{\min}$ and $T_{\max}$ data were converted from °C to K. For each station with record length $>20\,\text{yr}$ trends were calculated from annual mean X time series using the non-parametric Mann–Kendall statistical test (Kendall, 1975; Mann, 1945) with Sen's estimate of slope (Sen, 1968).
2. For each station daily X time series were de-trended and divided by the station mean such that the new mean is unity.

[8]Downloaded from `http://www.worldclim.org` in September 2011.

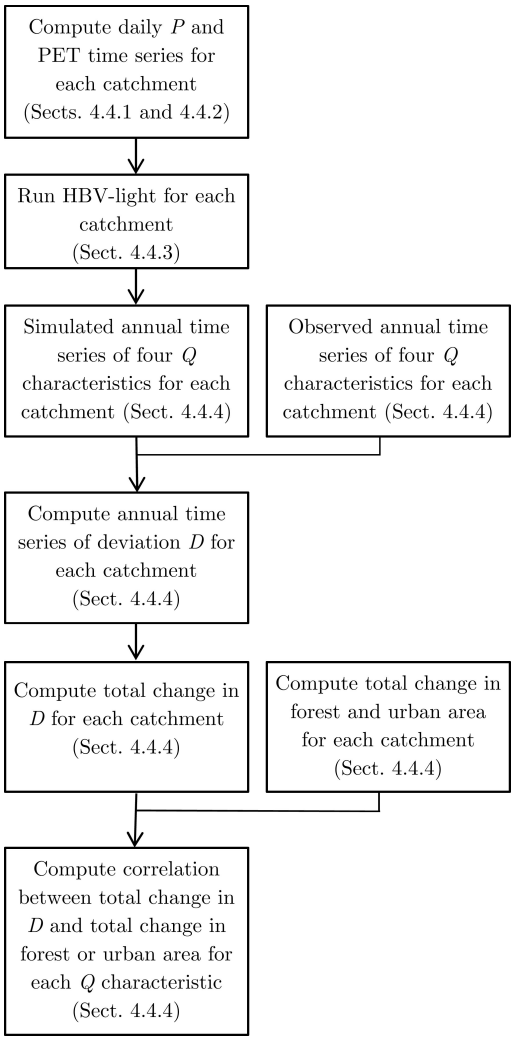


Figure 4.3: Flow diagram summarizing the various steps that were carried out to examine whether changes in forest and/or urban area have influenced the observed Q characteristics.

3. Daily maps with a resolution of 0.02° ($\sim 2\,\text{km}$) were computed from time series of X trend (having a constant value for each station) and from time series of de-trended unity X values using spatial interpolation (see next paragraph).
4. On a per-pixel basis the cumulative integral of the trend was calculated and offset such that the mean is unity, resulting in time series of net X change.
5. By multiplying the time series of net X change by time series of de-trended unity X values, time series of unity X values were obtained.
6. The time series of unity X values were multiplied by a map of elevation-corrected long-term mean X (PRISM for P, and WorldClim for $T_{\min}$ and $T_{\max}$; see sections 4.3.3 and 4.3.4, respectively).
7. Finally, daily X time series were calculated for each catchment by averaging over all the pixels comprising each catchment, and the $T_{\min}$ and $T_{\max}$ time series were re-converted from K to °C.

We employed the computationally efficient inverse-distance weighting (IDW; Shepard, 1968; Dirks et al.,

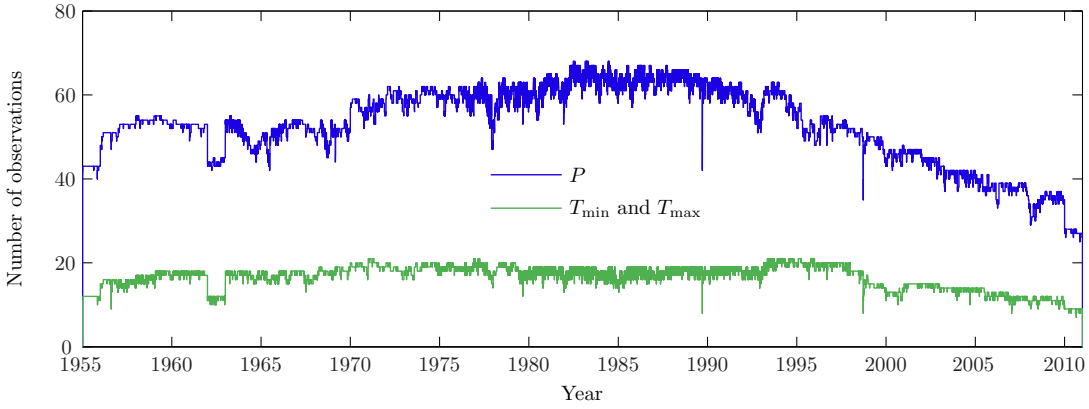


Figure 4.2: Graph showing the number of daily observations for P, $T_{\max}$, and $T_{\min}$. Only stations with records > 20 yr were used. The x-axis marks January 1st of each fifth year.

1998) technique for the spatial interpolation. The IDW technique requires a value for the distance-decay parameter, which controls how much weight is given to nearby stations relative to stations further away. Although it is recognized that this parameter varies per location (e.g., Lu and Wong, 2007), following recommendations by Garcia et al. (2008), and to reduce computational time, here a constant value of 3 was assumed.

The present approach has two advantages over the customary approach of using the nearest station with the longest record. Firstly, information from all nearby stations with records > 20 yr is incorporated. Secondly, our approach ensures unbiased mean annual values by rescaling against elevation-corrected long-term means. On the other hand, there are two caveats. First of all, portions of a time series may originate from different stations, thereby introducing spurious signals if, for instance, they have different seasonal patterns. Furthermore, for P, step six of the approach merely changes the intensity of the series, and does not correct for storm events unsampled by the isolated "point" network of meteorological stations.

4.4.2 Potential evaporation

The present study used the empirical Hargreaves equation (Hargreaves and Samani, 1985) according to guidelines set by the UN Food and Agriculture Organization (FAO; Allen et al., 1998) to assess long-term changes in the evaporative situation of the study catchments. The Hargreaves equation (Hargreaves and Samani, 1985) reads:

$$\mathrm{PET} = 0.0023\sqrt{T_{\max} - T_{\min}} \times \left(\frac{T_{\max} + T_{\min}}{2} + 17.8\right) R_a, \qquad (4.1)$$

where R_a is the extraterrestrial radiation [mm d^{-1}]. R_a was computed as described by Allen et al. (1998). The approach followed to obtain catchment-wide daily mean time series of $T_{\min}$ and $T_{\max}$ is outlined in section 4.4.1. The choice for the empirical Hargreaves method was motivated on the one hand by the lack of available long-term data for the climatic variables required for more physically based methods such as the Penman-Monteith equation (Monteith, 1965), and on the other hand by the better availability of $T_{\min}$ and $T_{\max}$ data. The Hargreaves method has been evaluated successfully against PET based on the Penman–Monteith equation (Trajkovic, 2007) and various other equations (Lu et al., 2005; Sperna Weiland et al., 2011).

4.4.3 HBV-light model

The HBV-light model (Seibert, 2005) was used to simulate Q. HBV-light is a spatially-lumped, conceptual rainfall-runoff model based on the HBV model (Bergström, 1976). HBV-light runs at a daily time step, has two groundwater stores and one unsaturated-zone store, and uses daily time series of P, PET, and T as inputs. T was not relevant in the present case since it is only used to drive the snow model subroutine. Rainfall interception is not estimated explicitly in HBV-light but is implicit in the BETA, FC, and UZL parameters. Although the model has been used predominantly in temperate-zone catchments (e.g., Te Linde et al., 2008; Steele-Dunne et al., 2008; Driessen et al., 2010), it has also been used successfully in tropical settings and for a range of catchment sizes, e.g., in Fiji (0.63 km^2; Waterloo et al., 2007), Southern Ecuador (75 km^2; Plesca et al., 2012), and Thailand (12 100 km^2; Wilk et al., 2001). For in-depth discussion of the model, see Seibert (2005). Table 4.3 briefly describes the model parameters and lists the calibration ranges used. We were unable to close the water balance for several catchments after exhausting all possible parameter combinations, probably due to errors in the PRISM P map, Q data and/or catchment boundary data, (non-quantified) water extractions, and/or inter-basin groundwater transfers. Therefore, an additional parameter (PCORR) was introduced that scaled P as required to match observed and predicted Q. Note that HBV-light does not consider groundwater flow within or between catchments.

Model parameters for each catchment were calibrated

using Latin hypercube sampling (LHS; McKay et al., 1979). LHS is a more efficient alternative (Yu et al., 2001) to the commonly used Monte Carlo technique (Metropolis, 1987; Beven, 1993; Seibert, 1999). Using LHS the parameter space (Table 4.3) is split up in n equal intervals. Values for the parameters are generated by sampling each interval just once in a random manner. The model is run n times with random combinations of the parameter values from each interval for each parameter. Here $n = 30\,000$ model runs were used to ensure convergence of the performance criterion. The first 10 yr of the record (Table 4.1) were used as spin-up period to initialize the groundwater stores after which the model was run for the entire period (i.e. the first 10 yr were run twice). The split-sample procedure (Refsgaard, 1997) was used to calibrate the parameters against data from the second half of the period and validate the parameters against data from the first half of the period.

Parameters of hydrological models are commonly calibrated using a composite of objective functions (i.e. a summary statistic incorporating several measures of performance; e.g., Madsen, 2000). Here, three different objective functions that strike a balance between accurate representations of all portions of the hydrograph were used. The first objective function represents the Nash-Sutcliffe efficiency (Nash and Sutcliffe, 1970; NS [−]). The NS is defined as:

$$\mathrm{NS} = 1 - \frac{\sum_{t=1}^{k} \left(Q_\mathrm{o}^t - Q_\mathrm{s}^t\right)^2}{\sum_{t=1}^{k} \left(Q_\mathrm{o}^t - \overline{Q_\mathrm{o}}\right)^2}, \tag{4.2}$$

where Q_s and Q_o represent 3-day mean simulated and observed Q, respectively [mm d^{-1}], t is the time step [−], and the summation is over $t = 1, 2, ..., k$. The second objective function (NS_log [−]) is the NS calculated from log-transformed Q_s and Q_o, thereby giving more weight to low Q values. NS and NS_log were calculated from 3-D mean Q to account for the flashy nature of the streams, which resulted in frequent mismatches between daily peaks of observed and simulated Q, thereby confounding the calibration. The third and final objective function represents the volume error (VE [%]):

$$\mathrm{VE} = 100\,\frac{\overline{Q_\mathrm{s}} - \overline{Q_\mathrm{o}}}{\overline{Q_\mathrm{o}}}. \tag{4.3}$$

For each run, the objective functions were combined to form a single aggregate measure using the combined-rank method (Booij and Krol, 2010).

The main analysis was conducted using the 30 "best" Q simulations. Parameter uncertainty (e.g., Beven, 1993; Seibert, 1997) was quantified as described in section 4.4.4. It should be noted that the K2 parameter in HBV-light was not calibrated, but was set to $1 - k_\mathrm{bf}$ (where k_bf is the baseflow recession constant listed in Table 4.1). In addition, the MAXBAS parameter was not calibrated, but was set to one, since the travel time for the catchments under study is likely to be less than a day (confirmed by exploratory calibration efforts in all catchments). The PCORR parameter was calibrated only for catchments that had absolute VE values $> 20\,\%$ for the calibration period when PCORR was set to one and parameter combinations were exhausted. For these catchments, the range of PCORR values used for the calibration was an initial estimate ± 0.05, where the initial estimate was calculated based on the initially obtained VE values.

4.4.4 Evaluation of the impacts of land-cover change on streamflows

For each catchment and Q characteristic, annual time series of the deviation between observations and simulations were calculated as:

$$D_d^x = 100\,\frac{Q_\mathrm{obs}^x - Q_{\mathrm{sim},d}^x}{\overline{Q_\mathrm{obs}^x}}, \tag{4.4}$$

where D are the deviation time series [%], Q_obs and Q_sim are annual time series of the observed and simulated Q characteristics, respectively, x is the year, and $d = 1, 2, ..., 30$ are the simulations. D integrates the effects of P and PET variability and basin carry-over storage on the Q characteristics, thereby isolating the impact of other factors, notably land-cover change.

For each catchment and Q characteristic, the trend (termed $\hat{a}_\mathrm{J}$ [%]) in the annual time series of D prior to 2000, when land-cover data are available, was calculated using the non-parametric jackknife resampling technique (Quenouille, 1956; Tukey, 1958). From each of the 30 annual time series of D consisting of M values, this procedure takes M subsamples of $M - 1$ values by omitting a different value each time. The trend in annual time series of D prior to 2000 is re-computed for each subsample, using the least-squares method, thus obtaining for each Q characteristic a 30-by-M matrix of trend estimates. The mean of the trend estimates was taken as the final trend estimate ($\hat{a}_\mathrm{J}$). The total change ($\hat{A}$ [%]) in D prior to 2000 is subsequently calculated by:

$$\hat{A} = \hat{a}_\mathrm{J}\, L, \tag{4.5}$$

where L [−] is the length of the record until 2000 (cf. Table 4.1). $\hat{A}$ can be interpreted as the cumulative change in annual observed values of the Q characteristic after accounting for the effects of climate variability and carry-over water storage on the Q characteristic.

The standard error associated with $\hat{a}_\mathrm{J}$ (termed $\hat{\sigma}_\mathrm{J}$ [%]) consists of two parts. The first ($\hat{\sigma}_\mathrm{s}$ [%]) was calculated from the mean dispersion of $\hat{a}$ within the time series of D, and is indicative of observational errors in P, PET, and Q data, non-linearities in the land-cover impact on Q, and/or possible defects in the HBV-light model structure. $\hat{\sigma}_\mathrm{s}$ was calculated using the jackknife technique as

Table 4.3: HBV-Light model parameter units, descriptions, and ranges used for the calibration. The K2 and MAXBAS parameters were not calibrated, but were set to 1-k_{bf} (cf. Table 4.1) and one, respectively. For the Valenciano, Canóvanas, Grande de Patillas, and Culebrinas catchments calibration ranges of 1.2–1.3, 0.7–0.8, 1.4–1.5, and 1.4–1.5, respectively, were used for PCORR. For the remaining eight catchments PCORR was set to one. See section 4.4.3 for details.

Parameter	Units	Description	Minimum	Maximum
BETA	-	Shape coefficient of recharge function	1	6
FC	mm	Maximum soil moisture storage	50	750
K0	d^{-1}	Recession coefficient of upper zone	0.05	0.99
K1	d^{-1}	Recession coefficient of upper zone	0.01	0.50
K2	d^{-1}	Recession coefficient of lower zone	-	-
LP	-	Soil moisture value above which actual $\mathrm{ET_a}$ reaches PET	0.05	0.95
MAXBAS	d	Length of equilateral triangular weighting function	-	-
PERC	$\mathrm{mm\ d}^{-1}$	Maximum percolation to lower zone	0	5
UZL	mm	Threshold parameter for extra outflow from upper zone	0	100
PCORR	-	P correction factor required to close water budget	-	-

follows:

$$\hat{\sigma}_{\mathrm{s}} = \frac{1}{30}\sum_{d=1}^{30}\sqrt{\frac{M-1}{M}\sum_{i=1}^{M}\left(\hat{a}_{di} - \overline{\hat{a}_d}\right)^2}, \qquad (4.6)$$

where $\hat{a}_{di}$ [%] are the estimates of a for subsamples $i = 1, 2, ..., M$ of annual D time series $d = 1, 2, ..., 30$. The second part of the standard error ($\hat{\sigma}_{\mathrm{p}}$ [%]) was calculated from the dispersion of the mean a amongst the 30 time series of D, and is mainly indicative of parameter uncertainty. It was calculated as:

$$\hat{\sigma}_{\mathrm{p}} = \mathrm{std}\left(\overline{\hat{a}_{d=1}},\ \overline{\hat{a}_{d=2}},\ ...,\ \overline{\hat{a}_{d=30}}\right), \qquad (4.7)$$

where std refers to the standard deviation. The two parts were combined to yield the standard error of $\hat{a}_{\mathrm{J}}$ ($\hat{\sigma}_{\mathrm{J}}$ [%]) as follows:

$$\hat{\sigma}_{\mathrm{J}} = \hat{\sigma}_{\mathrm{s}} + \hat{\sigma}_{\mathrm{p}}. \qquad (4.8)$$

The standard error associated with $\hat{A}$ ($\hat{\sigma}_{\hat{A}}$ [%]) is given by:

$$\hat{\sigma}_{\hat{A}} = \hat{\sigma}_{\mathrm{J}}\, L. \qquad (4.9)$$

To examine whether changes in forest and/or urban area influenced the four observed Q characteristics, correlation coefficients were calculated between amounts of change in forest or urban area per catchment and corresponding values of $\hat{A}$. Conventionally a significance level of 0.05 is applied, but this level was adjusted for the number of inferences made (four in the present study) to 0.01 using the Bonferroni procedure (Bland and Altman, 1995).

4.5 Results

4.5.1 Changes in climatic variables

Using the new P dataset (1955–2010), which comprises the records from 70 stations, trends for the island ranging from -0.25 to $+0.41\,\%\,\mathrm{yr}^{-1}$ (mean value $+0.02\,\%\,\mathrm{yr}^{-1}$) were found (Fig. 4.4a). In addition, a strong north-south disparity was observed, with positive P trends mainly identified to the south of the Cordillera Central, and negative trends north of it (Fig. 4.4a). Likewise, the new PET dataset (1955–2010), based on 21 stations, shows similar magnitudes of change, with per-pixel values ranging from -0.30 to $+0.17\,\%\,\mathrm{yr}^{-1}$ (mean value $-0.03\,\%\,\mathrm{yr}^{-1}$), but with a less clear spatial pattern compared to P (Fig. 4.4b).

4.5.2 HBV-light model performance

Table 4.4 lists the calibrated parameter values for the HBV-light model plus objective function scores for the calibration and validation periods for each catchment. Median Nash-Sutcliffe (NS) values of 0.64 and 0.63 (median of 12 values, where each value represents the mean NS value of the 30 best simulations for each catchment) were obtained for the calibration and validation periods, respectively (Table 4.4). The median absolute VE values were 3.1 and 7.2 % (median of 12 values, where each value represents the mean VE value of the 30 best simulations for each catchment) for the calibration and validation periods, respectively (Table 4.4). Catchments F, G, I, and L required optimization of the PCORR parameter (cf. Table 4.4).

4.5.3 Changes in land cover

Figure 4.5 shows the land-cover time series for each catchment. In all catchments forest and urban areas increased markedly between 1951 and 2000 at the expense of pasture and agricultural areas. During the period of simultaneous P, PET, and Q recording (Table 4.1) the net change in forest and urban areas for the study catchments ranged from +2 to +55 % (mean value +26 %) and from +2 to +11 % (mean value +7 %), respectively. The changes in pasture and agricultural areas ranged from −19 to +26 % (mean value −7 %) and from −63 to +9 % (mean value −26 %), respectively.

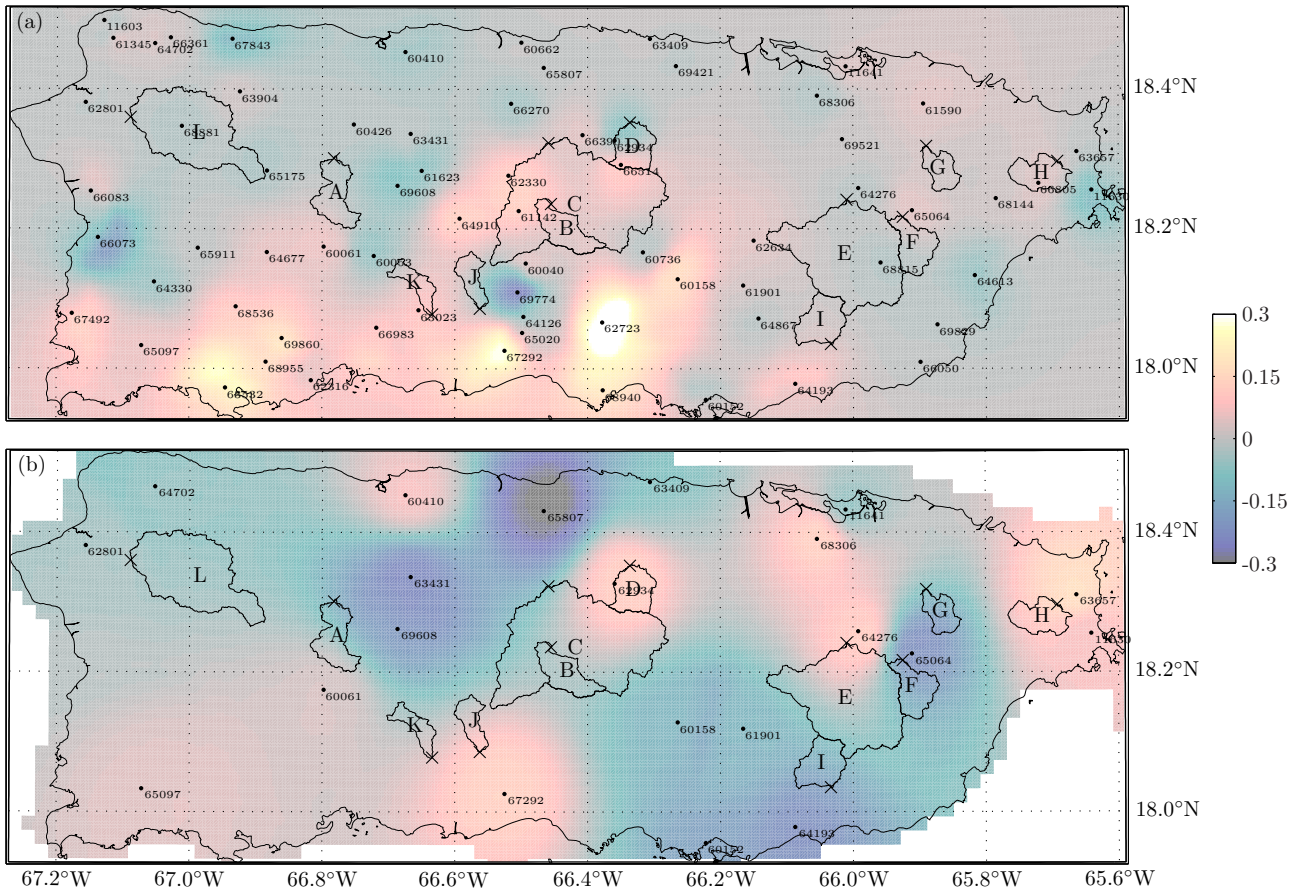


Figure 4.4: The trend [% yr^{-1}] from 1955 to 2010 for annual mean (a) P and (b) PET. Significant as well as insignificant trends are shown. The points mark stations with > 20 yr time series of (a) P, and (b) T_{min} and T_{max} (used to derive PET, see section 4.4.2). The numbers represent the last five digits of the GHCN-D station identification code. Catchments identified by letters and stream gauges by crosses.

4.5.4 Changes in streamflow characteristics

For each catchment and Q characteristic, Fig. 4.6 shows annual time series of the difference between observed and simulated Q as expressed by D (Eq. 4.4) and the corresponding fitted trend lines, whereas Table 4.5 lists values of the total change $\hat{A}$ (Eq. 4.5) and the corresponding error $\hat{\sigma}_{\hat{A}}$ (Eq. 4.9). Catchments with low NS values generally gave higher $\hat{\sigma}_{\hat{A}}$ values (cf. Tables 4.4 and 4.5). In general D_{tot} shows the least temporal variability (Fig. 4.6), which is reflected by low $\hat{\sigma}_{\hat{A}}$ values for $\hat{A}_{tot}$ (Table 4.5). The low temporal variability of Q_{tot} is likely due to the low degree of sampling error associated with the calculation of annual mean Q. Most catchments show progressive decreases over time in their time series of D (Fig. 4.6), as indicated by the mostly negative $\hat{A}$ values (Table 4.5), suggesting decreases in observed Q characteristics that are unrelated to carry-over effects of water storage and climate variability. The most pronounced decreases in Q_{tot} were found for catchments B, C, and J, whereas a moderate increase was found for catchment D. Pronounced decreases in both Q characteristics related to low flows (Q_{p5} and Q_{dry}) were found for catchments A, C, F, G, H, J, K, and L, whereas a (moderate) increase was found only for catchment I. Clear decreases in the Q characteristic related to peak flows (Q_{p95}) were found for catchments B, C, I, and J, whereas strong increases were found for catchments D, H, and L.

4.5.5 Impacts of land-cover change on streamflows

Figure 4.7 shows regressions between the amount of urban and forest area change vs. the cumulative change in annual time series of D prior to 2000 (expressed by $\hat{A}$ in Eq. 4.5) for each Q characteristic. In spite of strong increases in forest and urban area, and pronounced changes in D over time for the Q characteristics of several catchments, all correlations were insignificant ($p \geq 0.389$). Nonetheless, a weak (i.e. non-significant) positive relationship can be observed between changes in forest cover and changes in annual total streamflow Q_{tot}, when excluding catchments C and J (which appear to be outliers; Fig. 4.7c).

Table 4.4: Calibrated parameter values for the single best simulation by the HBV-Light model, and objective-function scores (mean of the 30 best simulations) for the calibration and validation periods, computed from 3-day means. The PCORR parameter was only calibrated for four catchments.

	A: Tanamá	B: Bauta	C: Grande de Manatí	D: Cibuco	E: Grande de Loíza	F: Valenciano	G: Canóvanas	H: Fajardo	I: Grande de Patillas	J: Inabón	K: Portugués	L: Culebrinas
Calibrated parameter values for the simulation with the best performance												
BETA [-]	1.351	1.306	0.940	2.887	2.999	0.862	3.204	4.212	1.056	1.041	1.117	3.633
FC [mm]	995	927	651	493	874	394	918	889	514	987	918	893
K0 [d^{-1}]	0.934	0.521	0.666	0.756	0.611	0.699	0.778	0.752	0.462	0.514	0.337	0.508
K1 [d^{-1}]	0.284	0.353	0.366	0.210	0.441	0.274	0.449	0.270	0.489	0.246	0.323	0.492
LP [-]	0.685	0.328	0.146	0.854	0.888	0.144	0.443	0.901	0.525	0.443	0.420	0.851
PERC [mm d^{-1}]	4.645	2.571	1.697	1.807	1.757	1.490	1.832	2.706	3.303	3.065	4.524	3.077
UZL [mm]	80	3	4	0	7	4	8	0	7	65	2	19
PCORR [-]	1.000	1.000	1.000	1.000	1.000	1.215	0.712	1.000	1.415	1.000	1.000	1.400
Objective-function scores for the calibration period, computed from 3-day means												
NS [-]	0.62	0.63	0.63	0.65	0.77	0.72	0.75	0.65	0.78	0.58	0.46	0.37
NS_{log} [-]	0.60	0.71	0.76	0.42	0.76	0.60	0.56	0.58	0.68	0.56	0.53	0.60
VE [%]	−6.22	4.11	11.97	−1.63	−7.56	1.78	12.63	2.22	−0.69	−0.09	−13.26	−1.01
Objective-function scores for the validation period, computed from 3-day means												
NS [-]	0.50	0.64	0.68	0.71	0.73	0.68	0.63	0.54	0.77	0.62	0.40	0.52
NS_{log} [-]	0.60	0.68	0.60	0.63	0.74	0.63	0.50	0.53	0.67	0.67	0.60	0.77
VE [%]	−1.27	−16.91	−13.65	0.44	−8.83	−3.95	10.81	−7.57	−1.91	−15.64	−6.89	−4.79

4.6 Discussion

4.6.1 Changes in climatic variables

Previous research on long-term changes in P in Puerto Rico, mostly using a limited number of climate stations, has suggested progressive declines in long-term P over 1900–2000 (Larsen, 2000; Van der Molen, 2002), but to the authors' knowledge no comprehensive analysis has been conducted for the entire island. Using the new P dataset (1955–2010) a strong north-south disparity in terms of trends was observed (Fig. 4.4a), which may be attributed to changes in wind patterns induced by changes in sea surface temperature (Comarazamy and González, 2011; cf. Van der Molen et al., 2006). Although trends in PET (1955–2010) were just as strong, they showed a less clear spatial pattern. Nevertheless, these findings reaffirm the importance of accounting for climate variability in studies assessing the effects of land-cover changes on Q.

4.6.2 HBV-light model performance

The HBV-light performance was good for both the calibration and validation periods (Table 4.4), suggesting that the simulated Q for the catchments can be used with confidence. Note that strong land-cover effects would have deteriorated performance statistics for the validation period. Several catchments required optimization of the PCORR parameter (cf. Table 4.4), possibly due to biases in the PRISM P map, uncertainties in the Q and/or catchment boundary data, water extractions, inter-basin groundwater transfers (potentially exacerbated by karst), and/or water recycling (e.g., Ellison et al., 2012), which combined may cancel out or amplify one another. However, the influence of the P scaling on the results is probably limited because the simulated Q was only used to control for climate and storage carry-over effects.

4.6.3 Impacts of land-cover change on streamflows

We were unable to support the three hypotheses because no significant relationships ($p < 0.01$) were found between the change in forest cover or urban area, and the change in the investigated Q characteristics across the 12 catchments (Fig. 4.7). This result agrees with previous meso- to macro-scale catchment studies in the tropics, subtropics, and warm-temperate regions (Table 4.6), which mostly failed to demonstrate a clear relationship between Q and change in forest area. Our use of multiple catchments proved important since many individual catchments showed pronounced changes in flow charac-

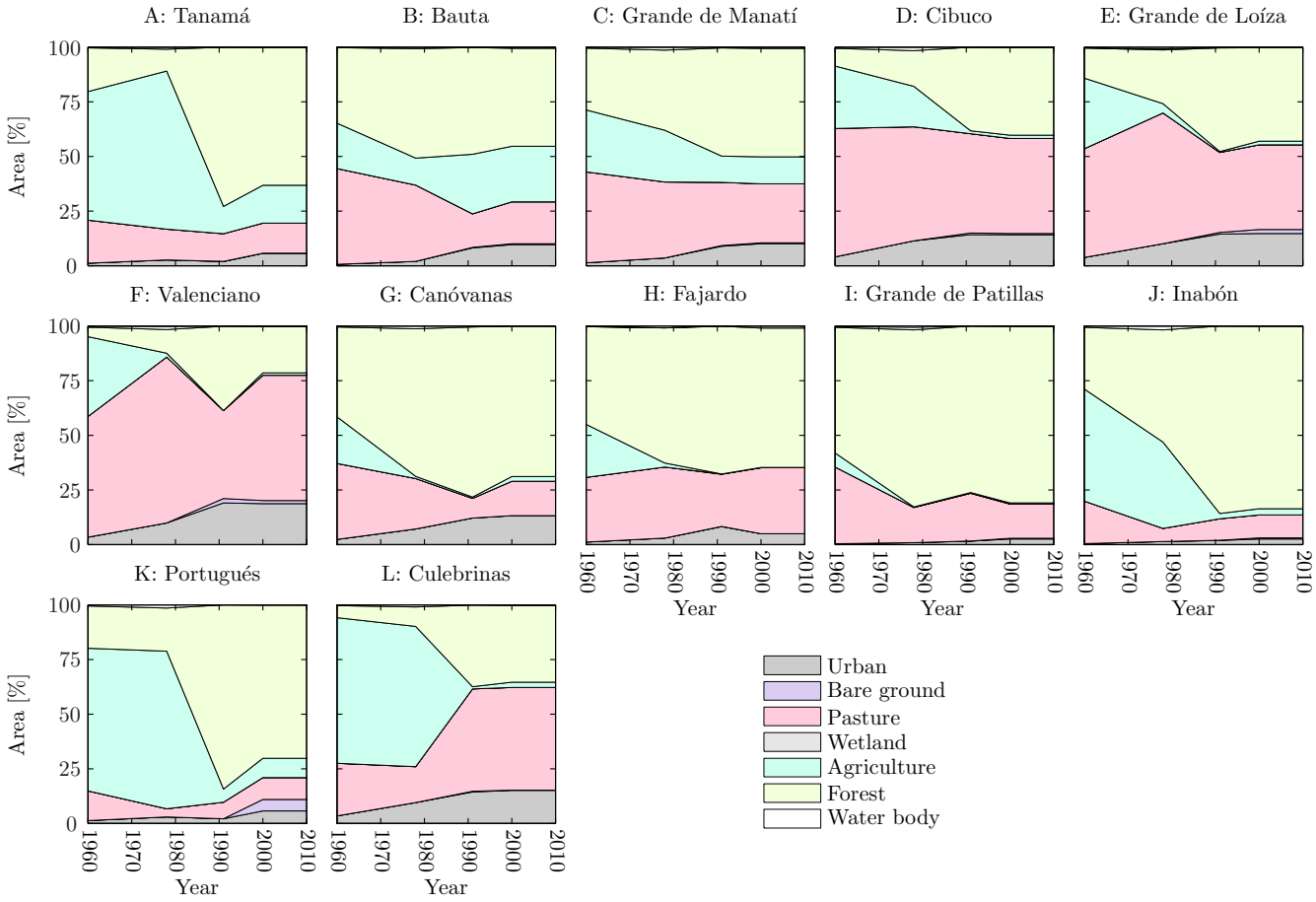


Figure 4.5: Changes with time in area of the dominant land-cover types for the study catchments. Graphs are based on generalized (cf. Table 4.2) land-cover classification maps for the years 1951, 1978, 1991, and 2000.

teristics that might well have been attributed to land-cover change otherwise.

Nevertheless, a weak positive relationship was observed between the change in forest cover and the change in Q_{tot} (Fig. 4.7c), tentatively suggesting that regeneration of forests leads to increases in Q in Puerto Rico. If true, this would imply that in the investigated catchments the increase in Q due to enhanced rainfall infiltration during forest regrowth overrides the decrease in Q associated with the greater water use of the aggrading forests (cf. Bruijnzeel, 1989; Scott et al., 2005). The soils beneath young secondary forests in Puerto Rico show higher bulk density than beneath mature secondary forests (Weaver et al., 1987; Lugo and Scatena, 1995) and are less structured (Lugo and Helmer, 2004). Presumably, this reflects soil compaction by livestock and cropping prior to abandonment and subsequent secondary forest regeneration. However, the general level of soil degradation under pasture in Puerto Rico (cf. Aide et al., 1996) is probably not sufficient to cause widespread occurrence of overland flow, although surface erosion under sugar cane and coffee plantations has been reported to be rampant in parts of the island (e.g., Del Mar López et al., 1998; Smith and Abruña, 1955). Unfortunately, to date, no unequivocal case of a positive relationship between changes in Q and reforested degraded area has been reported in the tropics (cf. Table 4.6), although demonstrated decreases in the amounts of headwater- or hillslope stormflow generation after reforesting severely degraded land (Chandler and Walter, 1998; Zhou et al., 2002; Zhang et al., 2004; Sun et al., 2006) must be considered large enough to overcome the associated increases in forest water use (Chandler, 2006; cf. Bruijnzeel, 2004; Scott et al., 2005). Nevertheless, long-term Q data for large, once degraded but subsequently rehabilitated catchments in sub-humid Texas (Wilcox and Huang, 2010) and the humid Red Soils region of South China (Zhou et al., 2010) have recently indicated gradually increased Q_{bf} over prolonged periods of time following large-scale land rehabilitation and regreening, suggesting soil improvement to be the dominant factor in these cases. Further process-based work is required to substantiate this contention. Similarly, pending the results of hydrological process studies in Puerto Rico's regenerating forests (in particular, infiltration and soil water retention vs. forest water use) the presently obtained positive relationships between the change in forest cover and the change in Q_{tot} may be spurious.

The present investigation is confounded somewhat by modest urbanization occurring simultaneously with forest regeneration in some of the investigated catchments (Fig. 4.5). Since urbanization typically increases the

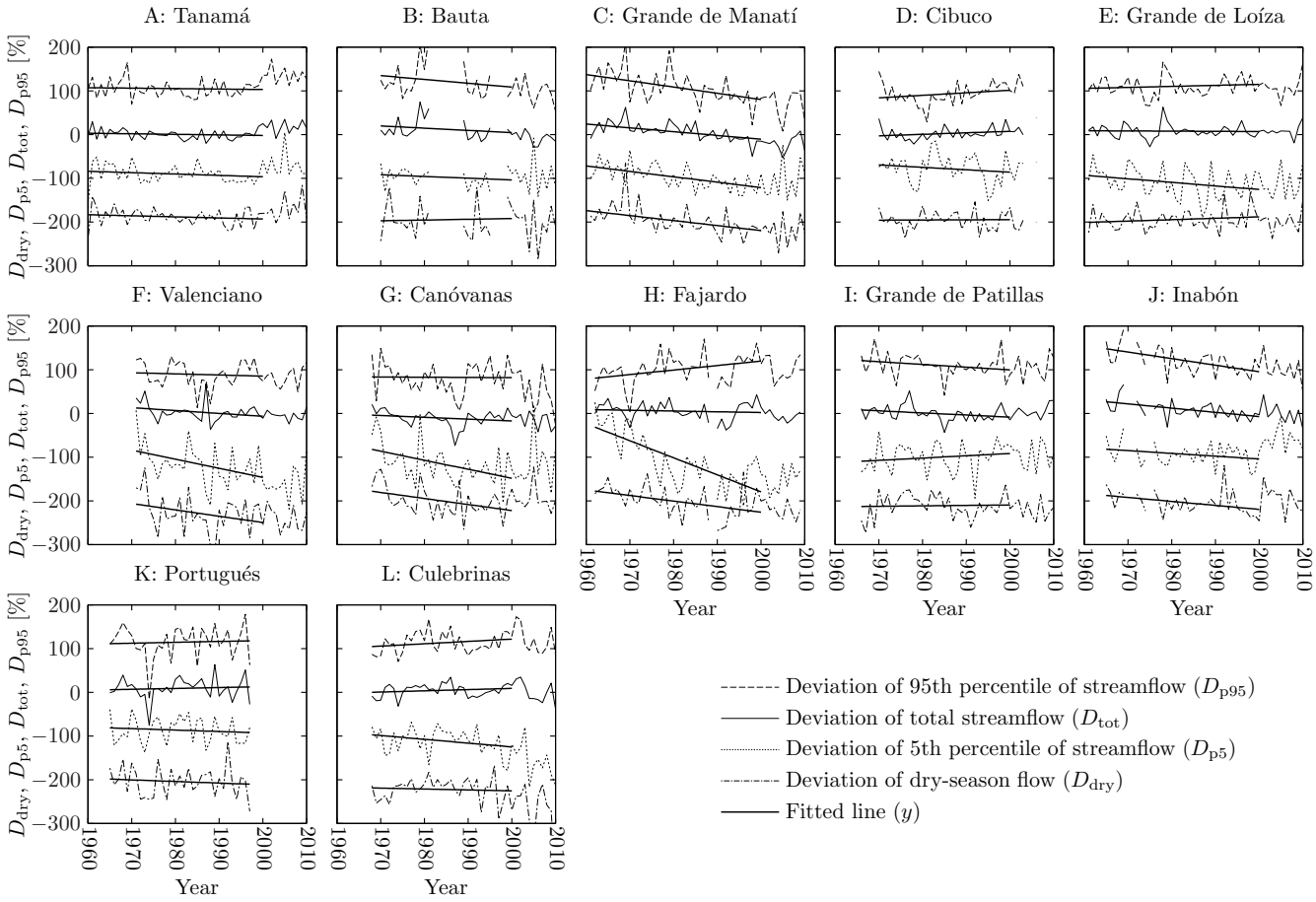


Figure 4.6: Annual time series of D_{p95}, D_{tot}, D_{p5}, and D_{dry} (displayed, respectively, from top to bottom in each panel; calculated using Eq. 4.4) and fitted trend lines. For clarity the D_{p95}, D_{p5}, and D_{dry} time series were offset by +100, −100, and −200 %, respectively.

area of impervious surfaces, thereby enhancing the frequency and intensity of infiltration-excess overland flow, more Q_{qf} is produced (e.g., Harto et al., 1998; DeWalle et al., 2000; Ziegler et al., 2004; Rijsdijk et al., 2007). This, if progressing beyond a critical threshold, can even result in reductions in Q_{bf} (Van der Weert, 1994; Bruijnzeel, 1989, 2004) on top of the reductions incurred already by the higher water use of the regenerating forest (Giambelluca, 2002). However, no significant relationships ($p < 0.01$) between the degree of urbanization and changes in any of the observed Q characteristics in our catchments were identified (Fig. 4.7). This is probably due to insufficient amounts of urbanization occurring (+2 to +11 %, mean value of +7 %; Fig. 4.5). Similarly, studies of the "flashiness" (sensu Baker et al., 2007) of the runoff behaviour of urbanized and forested catchments in Puerto Rico revealed no differences (Ramírez et al., 2009; Phillips and Scatena, 2010). However, Phillips and Scatena (2010) did detect significant differences ($p < 0.05$) in the frequency of stage change (cf. McMahon et al., 2003), possibly indicating locally enhanced Q_{qf} due to urbanization.

Wu et al. (2007) reported a change in Q_{tot} of −25 % (corrected for changes in P using a single P station that was not included in the current study as it did not meet the necessary quality requirements), between 1973–1980 and 1988–1995 for the Río Fajardo catchment in northeastern Puerto Rico (here identified as catchment H). Wu et al. (2007) attributed this negative change in Q to the small amount of forest regeneration occurring during that period (estimated at 8 % based on the interpolated time series shown in Fig. 4.5). For the same catchment, here a similar change in Q_{tot} of −31 % (calculated using D_{tot}) was found between the respective periods. However, a much smaller estimated change of −5 (± 4) % was obtained between the respective periods when taking into account the complete record (48 yr), using the calculated trend in D_{tot} prior to 2000 ($\hat{a}_J$) and the corresponding standard error ($\hat{\sigma}_J$, Eq. 4.8). Thus, Wu et al. (2007) may have overestimated the reduction in Q_{tot} for the Río Fajardo catchment because they only used part of the available record. Additionally, given that the Río Fajardo catchment appears to be an outlier in the present analysis (Figs. 4.7e and 4.7f) possibly due to non-quantified anthropogenic water extractions, any inferences for this catchment should be viewed with caution.

Table 4.5: Cumulative change in the difference between observed and simulated Q characteristics, as expressed by $\hat{A}$ (Eq. 4.5), and corresponding standard errors ($\hat{\sigma}_{\hat{A}}$; Eq. 4.9) for the 12 study catchments. The + and − signs indicate, respectively, increases and decreases in $\hat{A}$. The change covers the start of the record (cf. Table 4.1) until 2000.

	A: Tanamá	B: Bauta	C: Grande de Manatí	D: Cibuco	E: Grande de Loíza	F: Valenciano	G: Canóvanas	H: Fajardo	I: Grande de Patillas	J: Inabón	K: Portugués	L: Culebrinas
$\hat{A}_{p95}$ [%]	-4 ± 13	-27 ± 24	-61 ± 31	$+17 \pm 22$	$+8 \pm 15$	-6 ± 22	-3 ± 31	$+39 \pm 21$	-19 ± 22	-53 ± 19	$+8 \pm 29$	$+19 \pm 16$
$\hat{A}_{tot}$ [%]	-5 ± 8	-15 ± 12	-37 ± 14	$+10 \pm 12$	-2 ± 7	-19 ± 15	-13 ± 13	-6 ± 13	-16 ± 9	-34 ± 12	$+7 \pm 18$	$+9 \pm 10$
$\hat{A}_{p5}$ [%]	-12 ± 17	-13 ± 23	-52 ± 18	-17 ± 22	-33 ± 19	-58 ± 39	-65 ± 35	-147 ± 34	$+17 \pm 19$	-22 ± 19	-12 ± 22	-29 ± 22
$\hat{A}_{dry}$ [%]	-10 ± 16	$+5 \pm 33$	-49 ± 21	$+1 \pm 16$	$+13 \pm 19$	-42 ± 33	-44 ± 23	-48 ± 20	$+3 \pm 21$	-32 ± 17	-13 ± 28	-7 ± 24

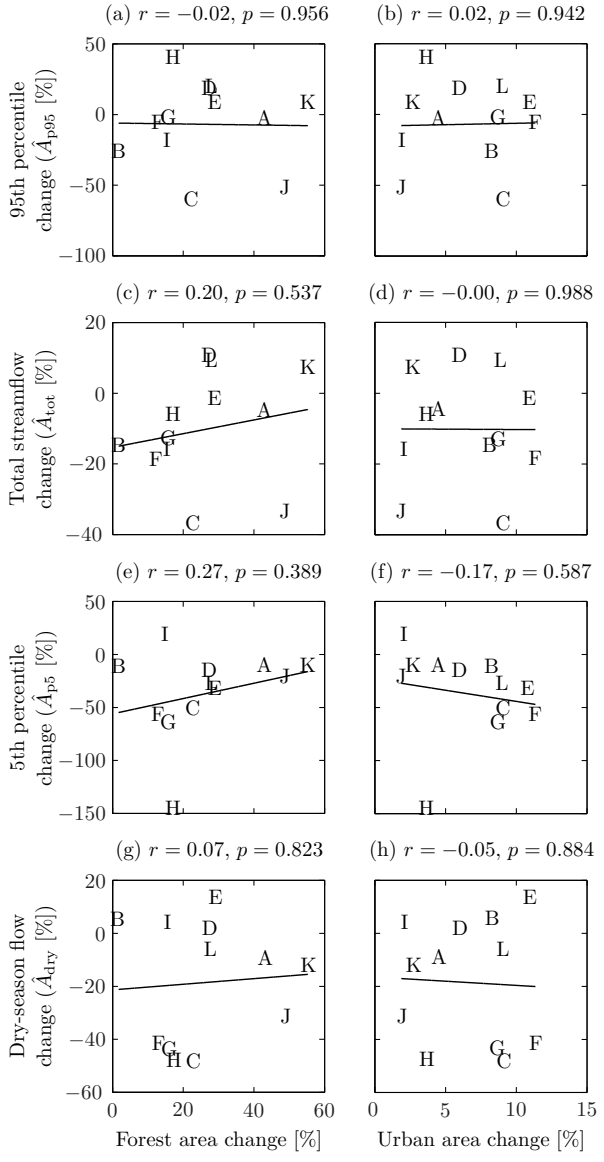


Figure 4.7: Net changes in forest and urban areas plotted against A_{p95} (a–b), Q_{tot} (c–d), Q_{p5} (e–f), and Q_{dry} (g–h) values (Eq. 4.5; Table 4.5). Each letter in the scatter plots represents a different catchment. The change covers the start of the record (cf. Table 4.1) until 2000.

4.6.4 Field-based estimates of vegetation water use

To explore the possible changes in total water yield that may be associated with a conversion from shaded coffee, sugar cane or pasture (the dominant land uses in Puerto Rico in 1951, see section 4.3.1) to (semi-)mature secondary forest (current situation), it is of interest to compare the typical amounts of total water use (ET_a) by the respective vegetation types. Field-based estimates of transpiration plus soil evaporation for shaded coffee and sugar cane in Puerto Rico range between 1.4–3.2 mm d^{-1} (Lin, 2010; Gutiérrez and Meinzer, 1994) and 2.5–4.6 mm d^{-1} (Vázquez, 1970; Fuhriman and Smith, 1951; Goyal and González-Fuentes, 1989; cf. Harmsen, 2003), respectively. The mean soil water uptake of various well-watered (lowland) pastures in northern Puerto Rico was estimated at 2.8 mm d^{-1} (Van der Molen, 2002), whereas for mature upland (Tabonuco) forests, values converge around 3.0 ($\pm$ 0.1) mm d^{-1} (Van der Molen, 2002; Wu et al., 2006). Thus, despite the deeper root systems of the forests (cf. Nepstad et al., 1994), mean soil water use for the agricultural crops under consideration, pasture, and forests is quite similar, possibly due to the relatively rainy climate prevailing in Puerto Rico all year round (Calvesbert, 1970). To this, rainfall interception losses (higher for forest) should be added. The best estimates of rainfall interception for mature Tabonuco forest ($\sim$ 21 % of mean annual P; Holwerda et al., 2006) translate to ca. 1.2 mm d^{-1} for an annual P of 2000 mm yr^{-1} (i.e. the approximate average rainfall for the 12 study catchments), vs. 0.9 mm d^{-1} for shaded coffee (Siles et al., 2010) and 0.2 mm d^{-1} for sugar cane (Leopoldo et al., 1981). Extrapolating the above-mentioned average values to a year, gives mean estimated ET_a totals for shaded coffee, sugar cane, and pasture of, respectively, ca. 1170, 1355, and 1020 mm yr^{-1} vs. ca. 1515 mm yr^{-1} for mature forest. Thus, the full conversion from pasture or crop land to mature forest could potentially change Q_{tot} by about -160 mm yr^{-1} (in the case of sugar cane) to -495 mm yr^{-1} (in the case

Table 4.6: Previous studies on tropical, subtropical, and warm-temperate meso- (≥ 1 km^2) and macro-scale ($\geq 10\,000$ km^2) catchments investigating the impact of changes in forest area on Q, listed alphabetically by country name. Only findings based on time-series analysis were listed. Present study included for completeness. In the "Q change" column "No change" also includes insignificant changes in flows. Where possible the reported change in Q was corrected for P variability. The "No. of gauges" column lists the number of stream gauges used in the analysis. The + and − signs indicate, respectively, increases and decreases in forest area or Q. The mean annual PET values represent reported estimates or Penman-Monteith estimates taken from Fisher et al. (2011) when no estimate was reported. Abbreviations: P, precipitation; PET, potential evaporation; and CO, carry-over water storage

Region (country)	No. of gauges	Area [km^2]	Mean annual P [mm yr^{-1}]	Mean annual PET [mm yr^{-1}]	Forest area change [%]	Time series length [yr]	Corrections applied	Q change [%]	Reference
Comet, Upper Burdekin (Australia)	2	16440, 17299	650, 690	1680, 1930	−45, −25	62, 13	P, PET	No change	Peña-Arancibia et al. (2012)
Tocantins (Brazil)	1	767000	1600	1600	−19	40	P	+24	Costa et al. (2003)
Ji-Paraná (Brazil)	7	33012	1875	1450	−50	23	P	No change	Linhares (2005); Rodriguez et al. (2010)
P. en Nirivilo, C. en El Arrayán (Chile)	2	253, 708	835, 718	1100, 1100	+5, +9[a]	20	P	−43, −32[b]	Little et al. (2009)
Guangdong Province (China)	32[c]	179752	1770	1230	+33	51	P, PET	No change	Zhou et al. (2010)
Dakeng (China)	1	10	1530	1050	−19	27	P, PET	−22[d]	Sun et al. (2008)
Hainan Island (China)	12	7 to 729	1950	1350	−29	25	P	No change[e]	Qian (1983)
Upper Yangtze (China)	4	200000	793	900	+8	41	P	+10[f]	Cheng (1999)
Citarum River (Indonesia)	1	4133	2500	1650	−50	16	P	+11[f]	Van der Weert (1994)
Galas, Kelantan (Malaysia)	2	13100	2000	1500	−3, +6[g]	31	P	No change	Adnan and Atkinson (2011)
Río Fajardo (Puerto Rico)	1	40	2908	1500	+12	23	P	−25	Wu et al. (2007)
Various (Puerto Rico)	12	23 to 346	1720 to 3422	1393 to 1657	+2 to +55	33 to 51	P, PET, CO	No change[h]	Present study
Nilwala (Sri Lanka)	1	380	2188	1500	−35	58	P	No change	Elkaduwa and Sakthivadivel (1999)
Chao Phraya (Thailand)	2	14522, 36000	1150	1500	−45	18, 25	P	No change	Dyhr-Nielsen (1986)
Nam Pong (Thailand)	1	12100	1350	1700	−63	36	P, PET, CO	No change	Wilk et al. (2001)
Upper Tacuarembó (Uruguay)	1	2097	1300	1230	+26	34	P	−38	Silveira and Alonso (2009)
Southern Piedmont (USA)	10	2820 to 19450	1300	1050	+10 to +28	49	P	−4 to −21[h]	Trimble et al. (1987)

[a] Represents the net forest area change of native forest cover (−38 % and −28 %, respectively) and forest plantations (+43 % and +37 %, respectively).
[b] Refers to changes in summer Q.
[c] Mostly nested catchments.
[d] The decrease in Q was explained by the higher ET_a of the regenerating vegetation.
[e] Refers to changes in peak flows only.
[f] Bruijnzeel (2004) argues that these increases in Q should be ascribed to urbanization rather than tree planting.
[g] Represents the net forest area change of native forest cover (−9 % and +2 %, respectively) and forest plantations (+6 % and +4 %, respectively).
[h] No relationship was found between the degree of forest-cover change and the change in Q.

of pasture).

For the catchments examined here a mean modeled change in Q_{tot} of $-86\,\mathrm{mm\,yr^{-1}}$ (standard deviation $\pm 124\,\mathrm{mm\,yr^{-1}}$) was found (calculated from values of $\hat{A}_{\mathrm{tot}}$ and Q_{tot} listed in Tables 4.5 and 4.1, respectively). If it is assumed that this change was solely due to the mean change in forest cover of +26 %, then the +100 % change in forest cover would have resulted in a change in Q_{tot} of about $-331\,(\pm 478)\,\mathrm{mm\,yr^{-1}}$. Although the modeled change in Q_{tot} lies in between the above estimates based on local field studies of $\mathrm{ET_a}$, the very high standard deviation of the estimated modeled change in Q_{tot} suggests that this agreement may be mere chance. Therefore, the question re-emerges as to why analysis of the impacts of land-cover change on Q in meso-scale tropical catchments does not confirm the findings of micro-scale experimental studies, which clearly demonstrate a relationship between change in forest area and change in Q (Bosch and Hewlett, 1982; Bruijnzeel, 1990; Sahin and Hall, 1996; Brown et al., 2005; Jackson et al., 2005).

4.6.5 Potential explanations for the lack of relationships

Upscaling in hydrology is a challenging issue that has received considerable attention (e.g., Peterson, 2000; Rodriguez-Iturbe, 2000; McDonnell et al., 2007; Blöschl and Montanari, 2010). For many studies the lack of a clear relationship between deforestation and change in Q can be explained by the rapid regeneration of tropical forests, where $\mathrm{ET_a}$ quickly reverts back to its original level and possibly exceeds it for decades (Giambelluca, 2002; Bruijnzeel, 2004; Juhrbandt et al., 2004). However, this explanation does not apply to the present study, since the catchments were exploited as pastures or agricultural fields for a sustained period of time prior to their abandonment and subsequent forest regeneration. Another possible reason for the inconclusive results obtained by many studies is the use of seasonal or annual mean Q characteristics only, whereas it is generally preferred to use characteristics related to the frequency and magnitude of Q (Alila et al., 2009, 2010), particularly when evaluating the change in peak flows. Finally, the failure to correct for climatic variability by many studies may also have led to inconclusive or even erroneous outcomes.

The first among the potential explanations for the lack of relationship between land-cover change and Q characteristics in Puerto Rico is covariance between land cover and climate in space, due to landscape ecology and history (cf. Van Dijk et al., 2012), or in time, due to land-cover changes altering the climate (Van der Molen, 2002; Pielke et al., 2007). Several meso-scale atmospheric modelling studies have produced conflicting results regarding the existence of such a relationship for Puerto Rico (Van der Molen et al., 2006; Comarazamy and González, 2011). The contrasting P trends reported here for different parts of the island (Fig. 4.4a) are also unable to provide evidence since similar magnitudes of forest regeneration occur across the island (Fig. 4.1). Also, because observed P and PET time series were input to the HBV-light model, most of the covariance between land-cover change, and P and PET should have been accounted for in the present analysis.

A second potential explanation is that the land-cover, P, PET, Q, and catchment boundary data used here contain more uncertainty than those used in small, commonly well instrumented, experimental studies. Using data for 278 Australian catchments Van Dijk et al. (2012) showed that the introduction of noise to long-term means of observed PET, P, and Q reduced the likelihood of detecting a land-cover signal. Here, as different methodologies were employed to derive the land-cover maps for the different years, and because the maps had different spatial resolutions and classification schemes (see section 4.3.1), it is conceivable that the land-cover time series contained uncertainties. For P a higher station density is probably desirable, particularly in tropical areas where the uncertainty in estimated catchment-wide P is exacerbated by the predominantly convective character of P and the limited spatial extent of individual storm cells (Nieuwolt, 1977; Hastenrath, 1991). Likewise, the PET data should also be viewed with caution, since net radiation, wind speed, and relative humidity, other dominant factors controlling PET magnitude and trends (see McVicar et al., 2012, and references therein), were not explicitly included in its calculation (cf. Eq. 4.1). The error in the Puerto Rico Q data is estimated to be 3–6 % (Sauer and Meyer, 1992). The cumulative error in annual P, PET, and Q was quantified by $\hat{\sigma}_{\hat{A}}$ (Eq. 4.9). Eventhough $\hat{\sigma}_{\hat{A}}$ does not account for the error in the land-cover data or for drift errors in the P, PET, and/or Q data during the study period, values of $\hat{\sigma}_{\hat{A}}$ already constitute a substantial portion of the variance in $\hat{A}$ values (the estimated total change in the observed Q) for all Q characteristics (cf. Table 4.5, and Fig. 4.7). This suggests that errors in land-cover, P, PET, and Q data may have restricted the detection of a relationship between land-cover change and Q characteristics in Puerto Rico. Kundzewicz and Robson (2004) and Chappell and Tych (2012) also suggested that a high degree of observational error may mask the identification of Q change.

A third potential explanation is that the amount of forest area change in the investigated catchments is less than that of most small catchment experimental studies. Based on small catchment experiments the critical threshold value for forest area change beyond which changes in Q can be detected is generally assumed to be ca. 20 % (e.g., Bosch and Hewlett, 1982). Increases in forest area for the investigated catchments were 2 to 55 % (mean 26 %), with increases exceeding 20 % in six of the twelve catchments. However, forest regeneration progressed from the headwaters to the lowlands (Grau et al., 2003) and by the time the records started (cf. Table 4.1) most of the headwaters would have been reforested already. Since most Q is produced in these rainier

headwater areas (García-Martinó et al., 1996) the observed overall change in Q may be far less than expected on the basis of lumped representations of P and forest area.

A fourth potential explanation relates to anthropogenic water extractions from the investigated catchments. Irrigation on agricultural fields from local wells may have increased ET_a, thereby masking the effect of increasing forest cover on Q. However, irrigation of sugar cane and pineapple in Puerto Rico occurs only on the drier southern coastal plains, and most irrigation water originates from lakes outside the study catchments (Molina-Rivera and Gómez-Gómez, 2008). Therefore, irrigation effects can probably be excluded as a possible explanation. However, water withdrawals associated with urbanization may have confounded the results, possibly causing reductions in Q_{bf} that are unrelated to changes in forest cover. Although catchments affected by major water extraction were excluded from the analysis (see section 4.3.2), it is possible that some catchments contain undocumented water intakes.

A fifth potential explanation, proposed by Zhou et al. (2010) and applicable to humid tropical settings, is that ET_a under such conditions is constrained by PET (i.e. energy limited; Calder, 1998), and therefore the expected increases in ET_a due to forest regrowth are not evident in the Q record. However, the interception evaporation component of ET_a can be well in excess of PET, particularly on mountainous maritime islands (Schellekens et al., 1999; Roberts et al., 2005; Holwerda et al., 2006; McJannet et al., 2007; Giambelluca et al., 2009; Holwerda et al., 2012). Moreover, an insignificant correlation was obtained between forest cover change and mean Q during the dry season (January–March; Q_{dry}), when PET is not expected to be a limiting factor (Fig. 4.7g). Hence, this is not seen as a convincing explanation.

Finally, a sixth potential explanation concerns heterogeneity between the study catchments in terms of vegetation cover and land-use history prior to land abandonment (not accounted for by the semi-quantitative classification of the present study), morphology, geology, and soils (McDonnell et al., 2007), influencing the response of Q to changes in forest cover. In addition, the use of lumped values for the climate variables at the catchment scale may not be valid due to spatial heterogeneity of P–PET–Q relationships (e.g., García-Martinó et al., 1996; Blöschl et al., 2007; Donohue et al., 2011).

4.7 Conclusion

Although there was considerable variability in the change in Q across the 12 examined catchments, the correlations between changes in urban or forest area and changes in Q were insignificant ($p \geq 0.389$) for all four Q characteristics. Consequently, there is little evidence to support the hypothesis that there is a relationship between the degree of secondary forest regeneration and the change in Q for the investigated catchments. Nevertheless, most catchments exhibited decreases in Q_{tot}, which may be attributable to enhanced vegetation water use associated with forest regeneration.

The most likely reasons for the lack of relationship between land-cover change and Q characteristics in Puerto Rico include (1) errors in the land cover, climate, Q, and/or catchment boundary data used in the analysis; (2) generation of Q is mostly in the rainier headwater areas that were already forested at the start of Q observations whereas subsequent changes in forest area mainly occurred in the drier lowlands; and (3) heterogeneity in the catchment response. Overall, it seems that the hydrological impacts of forest regeneration in Puerto Rico are currently unpredictable at the mesoscale, and further advances in our understanding are hampered by the quality, availability, and record length of the climatic and flow data. Our findings highlight the importance of catchment-scale analyses using multiple catchments but at the same time confirm the need for additional process studies at all stages of forest regeneration, notably of vegetation water use (both rainfall interception and transpiration) and changes in lowland hillslope hydrological response.

Acknowledgements

We wish to dedicate this paper to the memory of our friend, co-author and colleague Fred Scatena who passed away during the last stage of the writing. The authors further wish to thank Miguel Leon (University of Pennsylvania, Philadelphia, US), Jorge Peña-Arancibia (CSIRO Land and Water, Canberra, Australia), Maarten Waterloo, and Ilja van Meerveld (both from VU University, Amsterdam, The Netherlands) for various forms of support and information. We are also most grateful to Eileen Helmer and Olga Ramos (USDA Forest Service International Institute of Tropical Forestry, Río Piedras, Puerto Rico) for providing the 1951 and 1978 land-cover maps, respectively. Finally, we like to thank Dr Younes Alila (University of British Columbia, Vancouver, Canada) for insightful comments on an earlier version of this paper.

Chapter 5

Global patterns in baseflow index and recession derived from 3520 small catchments[1]

Abstract. Baseflow is the portion of streamflow that is fed by groundwater storage and/or other delayed sources, and is important for water resources management, river ecology, and water quality. Previous studies have constructed models to estimate baseflow characteristics from catchment physiographic characteristics and applied these to ungauged regions. However, these studies generally used a relatively small number of catchments (< 200) and were regional in nature, which may have led to less reliable models having limited applicability elsewhere. This study uses a highly heterogeneous set of 3520 catchments from around the globe to construct reliable, widely applicable models using 18 physiographic characteristics to estimate two important baseflow characteristics: (1) the baseflow index (BFI [-]), defined as the ratio of long-term mean baseflow to total streamflow; and (2) the baseflow recession constant (k [d^{-1}]), defined as the rate of baseflow decay. Regression analysis results revealed that BFI and k were related to several physiographic characteristics, notably mean annual potential evaporation, mean catchment elevation, mean surface slope, fraction of open water, and the mean sand content of the soil. Ensembles of artificial neural networks (obtained by sub-sampling the original set of catchments) were used to estimate the baseflow characteristics from physiographic data. The catchment-scale estimation of the baseflow characteristics demonstrated encouraging performance with R^2 values of 0.82 for BFI and 0.73 for k. Global maps of estimated BFI and k were obtained using global physiographic data as input to the established models. The uncertainty was lowest in North America, Europe, and southeastern Australia, where most catchments were located, and highest in areas with limited streamflow data (arid, semi-arid, and Arctic regions). These global maps should prove useful for various large-scale hydrological applications.

[1]This chapter is an edited version of: Beck, H. E; van Dijk, A. I. J. M; Miralles, D. G.; de Jeu, R. A. M.; Bruijnzeel., L. A.; McVicar, T. R., and Schellekens, J. Global patterns in baseflow index and recession derived from 3520 small catchments. Submitted, 2013.

5.1 Introduction

Baseflow is the portion of streamflow (Q) originating from groundwater storage and/or other delayed sources (Hall, 1968). Knowledge of the baseflow regime is important for a number of purposes: water resources management; aquatic ecosystem preservation; hydropower generation; contaminant transport; and low-flow forecasting (e.g., Campolo et al., 1999; Brauman et al., 2007; Cyr et al., 2011, and references therein). Such knowledge is not directly available for ungauged catchments and hence for most of the terrestrial land surface (Fekete and Vörösmarty, 2007) necessitating regionalization procedures to transfer model parameters or Q characteristics from gauged to ungauged catchments (e.g., Parajka et al., 2005a; Yadav et al., 2007; Oudin et al., 2008; Zhang et al., 2008b). Two important baseflow characteristics are: (1) the baseflow index (BFI [-]), defined as the ratio of long-term mean baseflow to total Q (Smakhtin, 2001); and (2) the recession constant (k [d^{-1}]), defined as the rate of baseflow decay (Vogel and Kroll, 1996).

Several regression-based regionalization studies have established models to estimate BFI or k from catchment physiographic characteristics (e.g., Mazvimavi et al., 2005; Brandes et al., 2005; Longobardi and Villani, 2008; Van Dijk, 2010; Peña-Arancibia et al., 2010; Krakauer and Temimi, 2011; Ahiablame et al., 2013). Geology and soils were generally among the most important physiographic characteristics identified in these studies, although topography, climate, and/or land cover have also proven useful in some cases. Although previously developed models may, in theory, be used to estimate baseflow characteristics for ungauged catchments, they suffer from one or more of the following shortcomings that limit their use in macro- and global-scale applications. First, they were mainly based on a relatively small number of Q gauging stations (< 200), which can lead to less reliable and/or overfitted models. Second, they focused on a regional scale and used regional data sets to characterize geology or soils, thereby potentially restrict-

ing their larger-scale applicability. Third, most studies did not evaluate the generalization ability of the model using an independent set of catchments, and thus it is difficult to judge the true performance of the models. Finally, they reached conflicting conclusions regarding the importance of certain physiographic characteristics, notably mean surface slope and the fraction of forest.

Consequently, current macro- and global-scale baseflow parameterizations, such as those used in land surface schemes (a crucial component of global circulation models) and global hydrological models, are often far from ideal and vary widely (cf. Duan et al., 2001). Some use globally fixed parameters, such as the Community Land Model (CLM; Oleson et al., 2010), Noah-MP (Niu et al., 2011), and Mac-PDM (Gosling and Arnell, 2011). Others use nearest-neighbour interpolation of calibrated parameters, such as the Variable Infiltration Capacity (VIC) model (Liang et al., 1994; Nijssen et al., 2001) and WASMOD-M (Widén-Nilsson et al., 2007). Yet others rely on expert opinion and hydrologic interpretation of global geological data sets, such as WaterGAP (Döll and Fiedler, 2008) and PCR-GLOBWB (Bierkens and van Beek, 2009). Recently, the wider availability of Q and physiographic data, thanks in particular to (on-going) efforts by the U.S. Geological Survey (USGS) and the Global Runoff Data Centre (GRDC; Koblenz, Germany; `http://grdc.bafg.de`) and advances in computing and remote sensing technology, has created the possibility to explore the estimation of BFI and k at larger scales. Examples include Schneider et al. (2007) and Santhi et al. (2008), who investigated the regionalization of BFI for Europe and the conterminous USA, respectively, and Peña-Arancibia et al. (2010), who investigated the regionalization of k across the tropics.

The current study uses a large set of 3520 catchments that covers all continents and a wide range of physiographic conditions, thereby allowing the construction of reliable, more widely applicable models to estimate BFI and k. To our knowledge, the current study is the first attempt to estimate these baseflow characteristics from physiographic data at such global scale using such a large Q data set. Specific objectives are to: (1) analyze the relationships between catchment physiographic characteristics and the selected baseflow characteristics; (2) construct models to estimate the baseflow characteristics from physiographic characteristics and assess their generalization ability; and (3) investigate the feasibility of global-scale estimation of the baseflow characteristics.

5.2 Data

5.2.1 Observed streamflow

Daily observed Q data were derived from three sources. First, Q data for 1862 USA stations that were part of the Model Parameter Estimation Experiment (MOPEX; Duan et al., 2006) were downloaded from the USGS National Water Information System (`http://waterdata.usgs.gov`) and the associated catchment boundaries from the MOPEX webserver (`ftp://hydrology.nws.noaa.gov/pub/gcip/mopex/`). Second, Q data for 4047 stations from the GRDC were considered. Corresponding catchment boundary data were provided by the GRDC. Third and finally, Q data and associated catchment boundaries for 321 Australian stations part of a database compiled by Peel et al. (2000) were used. Together, this resulted in an initial data set comprising 6230 Q gauging stations.

For catchments to be included here three requirements needed to be satisfied. First, due to the importance of channel routing in large catchments (e.g., McGlynn et al., 2004) the catchment area had to be $< 10\,000$ km^2 (cf. Peña-Arancibia et al., 2010). Second, to reduce anthropogenic influences, < 2 % of the catchment was allowed to be urban (using the "artificial areas" class of the GlobCover v2 map; Bontemps et al., 2011) or subject to irrigation (using the Global Irrigated Area Map; `http://www.iwmigiam.org`). Third and finally, to ensure reliable estimates of the baseflow characteristics the Q record length had to be > 10 yr (not necessarily consecutive). This resulted in a set of 3520 Q gauging stations, the locations of which are shown in Fig. 5.1. The 10th percentile, median, and 90th percentile values of the Q record lengths were 21 yr, 50 yr, and 86 yr, respectively. The 10th percentile, median, and 90th percentile values of the catchment areas were 129 km^2, 901 km^2, and 5062 km^2, respectively. All Q data were converted to mm d^{-1} using the catchment areas.

5.2.2 Physiographic characteristics

Table 5.1 lists the physiographic characteristics used as predictors to model BFI and k. Predictor selection was inclusive and guided by previous regionalization studies, expert knowledge, and data availability. Among the selected physiographic characteristics, nine were related to climate, two to topography, three to land cover, and four to soils—bringing the total number of predictors to 18. For the catchment-scale estimation of the baseflow characteristics the full-resolution data were used. However, for the computation of global maps, data with a resolution $< 0.25°$ were re-sampled to $0.25°$ using averaging, whereas data with a resolution $> 0.25°$ were re-sampled to $0.25°$ using bilinear interpolation.

A number of other physiographic characteristics were considered, but not included in the analysis. The topographic wetness index (TWI; Beven and Kirkby, 1979) was not used because a global high-resolution TWI data set is not (yet) available. Drainage density (total length of streams per unit catchment area) was not used either, due to the lack of globally consistent river-network data (cf. Benstead and Leigh, 2012). Permafrost extent (Brown et al., 1997) was tested but not used as its inclusion did not result in better model performance and since its spatial patterns closely matched those of SNOW (cf. Table 5.1). Although four other global data sets pro-

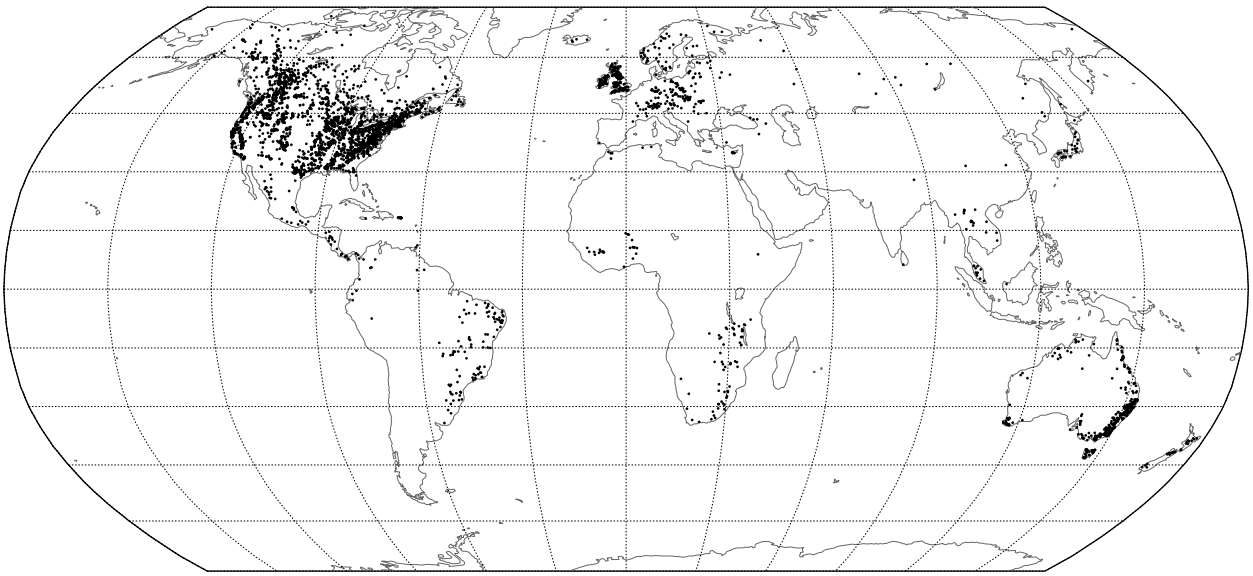


Figure 5.1: Locations of the catchments used here. Each data point represents a catchment centroid ($n = 3520$).

viding information on soils or geology were considered (Dürr et al., 2005; Batjes, 2006; Gleeson et al., 2011; Hartmann and Moosdorf, 2012), the Harmonized World Soil Database (HWSD; FAO/IIASA, 2012) was selected as it has a high resolution of 1 km, is based on a comprehensive collection of soil data, and because strong links have been found between the sand content of the soil (SAND) and BFI (Boorman et al., 1995; Santhi et al., 2008). To improve the HWSD data set it was supplemented with SAND data for the USA as derived from the STATSGO (Wolock et al., 2004) data set, with the silt and clay contents changed accordingly.

5.3 Methodology

5.3.1 Computation of BFI and k

Single values of BFI and k were computed from the Q record of each catchment following Van Dijk (2010). A linear reservoir model was assumed as this is generally considered to be a good approximation (e.g., Chapman, 1999; Fenicia et al., 2006; Van Dijk, 2010). Because Q may still contain quickflow following rainfall events the first five days after the last day with an increasing Q (compared to the previous day) were excluded in the compution of k. See Van Dijk (2010) for further details. It is pertinent to note that the choice of techniques to compute k and BFI may affect the results (cf. Nathan and McMahon, 1990; Vogel and Kroll, 1996; Bullock et al., 1997; Eckhardt, 2008). The distribution of the derived k values for the different catchments had a strong negative negative skew that might confound the modeling exercise. Hence, to make the data better conform to a normal distribution the following logarithmic transformation was applied:

$$k_{\text{trans}} = -\ln\left(1 - k\right), \tag{5.1}$$

where ln refers to the natural logarithm and k_{trans} [-] is the transformed k. The inverse of this transformation is given by:

$$k = 1 - e^{-k_{\text{trans}}}. \tag{5.2}$$

The obtained k was subsequently used to separate the Q record into baseflow and quickflow using a forward- and backward-recursive digital filter (Van Dijk, 2010). Fig. 5.2 gives an example of baseflow computed in this way. The BFI was calculated as the ratio of long-term mean baseflow to total Q, ranging from 0 to 1. Since the distribution of the derived BFI values for the different catchments showed a weak negative skew the following power-transformation was applied:

$$\text{BFI}_{\text{trans}} = \text{BFI}^2, \tag{5.3}$$

where $\text{BFI}_{\text{trans}}$ [-] is the transformed BFI. The inverse of this transformation is given by:

$$\text{BFI} = \sqrt{\text{BFI}_{\text{trans}}}. \tag{5.4}$$

The remaining three methodological sub-headers reflect the original three objectives of this study, and are used to structure the subsequent Results and Discussion sections.

5.3.2 Physiographic controls of BFI and k

Using regression analysis the strength and shape of the relationships between catchment physiographic characteristics and the transformed baseflow characteristics ($\text{BFI}_{\text{trans}}$ and k_{trans}) were evaluated. Linear, exponential, logarithmic, and power functions were fitted by least-squares and the function with the highest coefficient of determination (R^2) was reported. Significance levels (or p values) were not calculated as these may be misleading (Nicholls, 2001), particularly when using such large data sets (Royall, 1986; Johnson, 1999).

Table 5.1: The physiographic characteristics used as predictors of the baseflow index (BFI) and recession constant (k).

Type	Predictor(s)	Description	Source and calculation[a]	Resolution
Climate	AI [-]	Aridity index	$\mathrm{AI} = P/\mathrm{PET}$	0.5°
	$\mathrm{AI_{si}}$ [-]	Aridity index seasonality	$\mathrm{AI_{si}} = \lvert \mathrm{AI_{sum}} - \mathrm{AI_{win}} \rvert / \mathrm{AI}$	0.5°
	P [mm $\mathrm{yr^{-1}}$]	Mean annual precipitation	WorldClim (Hijmans et al., 2005)	∼1 km
	P_{si} [-]	Precipitation seasonality	$P_{\mathrm{si}} = \lvert P_{\mathrm{sum}} - P_{\mathrm{win}} \rvert / P$	∼1 km
	PET [mm $\mathrm{yr^{-1}}$]	Mean annual potential evaporation	Penman-Monteith (Fisher et al., 2011)	0.5°
	$\mathrm{PET_{si}}$ [-]	Potential evaporation seasonality	$\mathrm{PET_{si}} = \lvert \mathrm{PET_{sum}} - \mathrm{PET_{win}} \rvert / \mathrm{PET}$	0.5°
	TA [K]	Mean annual air temperature	WorldClim (Hijmans et al., 2005)	∼1 km
	$\mathrm{TA_{si}}$ [-]	Air temperature seasonality	$T_{\mathrm{si}} = \lvert T_{\mathrm{sum}} - T_{\mathrm{win}} \rvert / T$	∼1 km
	SNOW [mm]	Mean snow-water equivalent depth	GlobSnow L3A prototype with mountains included v1 (mean of 2008–2010; Luojus et al., 2010) for latitudes > 35°N and AMSR-E/Aqua L3 v10 (mean of 2003–2011; Chang and Rango, 2000) for latitudes ≤ 35°N	∼25 km
Topography	ELEV [m asl]	Mean elevation	For the catchment-scale analysis CGIAR-CSI SRTM v4.1 data were used, whereas for the global-scale analysis IIASA-LUS data (Fischer et al., 2008) were used	∼90 m, ∼0.08°
	SLO [°]	Mean surface slope	Idem	∼90 m, ∼0.08°
Land cover	fW [-]	Fraction of open water	GlobCover v2 (Bontemps et al., 2011)	∼300 m
	fTC [-]	Fraction of forest	MODIS MOD44B collection 4 v3 (Hansen et al., 2003)	∼250 m
	NDVI [-]	Mean Normalized Difference Vegetation Index (NDVI; Tucker, 1979)	MODIS MOD13C2 collection 5 (Huete et al., 2002), mean of 2001–2012	0.05°
Soils	GRAV [%]	Mean gravel content	HWSD v2.0 (FAO/IIASA, 2012), mean of topsoil and subsoil values	∼1 km
	SAND [%]	Mean sand content	HWSD v2.0 (FAO/IIASA, 2012), mean of topsoil and subsoil values, supplemented with STATSGO (Wolock et al., 2004) for the USA	∼1 km
	SILT [%]	Mean silt content	HWSD v2.0 (FAO/IIASA, 2012), mean of topsoil and subsoil values	∼1 km
	CLAY [%]	Mean clay content	Idem	∼1 km

[a] The "sum" and "win" subscripts denote Northern Hemisphere 'summer' (April–September) and 'winter' (October–March) climatologic means, respectively.

5.3.3 Catchment-scale estimation of BFI and k

Artificial Neural Networks (ANNs) are flexible, non-paracharacteristic tools able to model complex non-linear relationships between inputs and outputs (Bishop, 1995). ANNs have been used succesfully in many fields of science, including hydrology (ASCE, 2000a,b; Govindaraju and Rao, 2000; Maier and Dandy, 2000). Here, feed-forward ANNs based on the multi-layer perceptron (MLP; Bishop, 1995) with one hidden layer were used to estimate $\mathrm{BFI_{trans}}$ and k_{trans} from the physiographic data. The inputs (i.e., catchment physiographic data) and outputs (i.e., $\mathrm{BFI_{trans}}$ and k_{trans}) were standardized using the means and standard deviations of the catchment values. The MLP models were trained using the Levenberg-Marquardt (LM) algorithm (Levenberg, 1944; Marquardt, 1963) in combination with the mean-squared error performance function. The LM algorithm is considered to be one of the most efficient learning algorithms (Hagan, 1994).

The 10-fold cross-validation procedure (Shao, 1993) was used to estimate the generalization ability of the established MLP models. This procedure randomly partitions the original set of 3520 catchments into ten subsamples, of which each comprises 10 % of the catchments ($n = 312$). For ten iterations, each time a different subsample of catchments was used to independently test the model's performance, the other 90 % were randomly partitioned further into a training subset consisting of 75 % of the catchments ($n = 2337$) and a validation subset consisting of 15 % of the catchments ($n = 467$). For each iteration, the MLP model was trained on the training subset, while to prevent overfitting, the training process was stopped once the error for the validation subset started to increase (Sarle, 1995; Bishop, 1995). The generalization ability of the MLP models was derived by averaging R^2 and root mean square error (RMSE) statistics computed for each cross-validation iteration from the testing subset of catchments. The optimal number of neurons in the hidden layer was determined by trial and error, based on R^2 and RMSE values obtained for the testing subsets. The number of neurons in the hidden layer was set at 30, as no further performance improvement was gained beyond 30 neurons. Using such a high number of neurons in the hidden layer avoids the con-

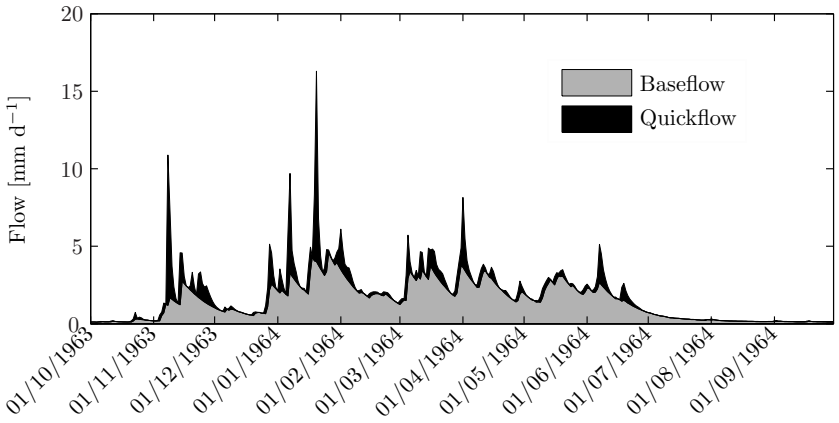


Figure 5.2: Computed baseflow and quickflow at USGS gauge no. 14307700 (Jackson Creek near Tiller, OR, USA; catchment area 394 km^2). The computed quickflow is the observed Q minus the computed baseflow. BFI = 0.66 and k = 0.94 d^{-1} for this catchment based on the complete record from 1955–1986.

vergence to local minima (Tetko et al., 1995). As a last step, the output was de-standardized using the mean and standard deviation of the catchment values.

5.3.4 Global maps of BFI and k

Global physiographic data were used as input to the established MLP models, producing ten maps of $\text{BFI}_{\text{trans}}$ and k_{trans} (0.25° resolution). These maps were combined into single maps of $\text{BFI}_{\text{trans}}$ and k_{trans} by calculating the per-pixel median of the ten $\text{BFI}_{\text{trans}}$ or k_{trans} values, respectively. Next, to correct for the statistical phenomenon of regression toward the mean (Galton, 1886; Bland and Altman, 1994), the regression equations describing the catchment-scale relationships between median estimated vs. observed values of BFI and k were applied to the maps of $\text{BFI}_{\text{trans}}$ and k_{trans}, respectively. The maps were subsequently transformed back to BFI and k using Eqs. 5.2 and 5.4, respectively. The estimation uncertainty was quantified by calculating the per-pixel standard deviation of the ten $\text{BFI}_{\text{trans}}$ and k_{trans} values. Pixels attributed as ice (using the WWF terrestrial biomes map v2; Olson et al., 2001; see Fig. 5.3) or open water (using GlobCover v2; Bontemps et al., 2011) were excluded. To better understand the spatial patterns of the global estimates, for each 0.25° latitude band the 90th percentile, median, and 10th percentile of the global BFI and k estimates were computed. Additionally, for each WWF terrestrial biome the medians of the global BFI and k estimates were computed.

5.4 Results

5.4.1 Physiographic controls of BFI and k

Figs. 5.4 and 5.5 show scatterplots of catchment-mean values of the 18 selected physiographic characteristics vs. observed values of the transformed baseflow characteristics ($\text{BFI}_{\text{trans}}$ and k_{trans}, respectively). The relationships were all rather weak ($R^2 \leq 0.21$) and often characterized by high degrees of non-linearity and/or heteroscedasticity (i.e., uneven variability). Among the nine climate predictors, AI_{si}, PET, PET_{si}, TA, and SNOW were moderately well related to $\text{BFI}_{\text{trans}}$ (Figs. 5.4b, 5.4e, 5.4f, 5.4g, and 5.4i, respectively), whereas AI, AI_{si}, and PET were moderately well related to k_{trans} (Fig. 5.5a, 5.5b, and 5.5e, respectively). The relationships between ELEV or SLO and $\text{BFI}_{\text{trans}}$ or k_{trans} were positive (Figs. 5.4j, 5.5j, 5.5k, and 5.5k, respectively). Among the three land-cover indices, fTC and NDVI were related to neither $\text{BFI}_{\text{trans}}$ nor k_{trans} (Figs. 5.4m, 5.5m, 5.4n, and 5.5n, respectively), whereas fW was positively related to $\text{BFI}_{\text{trans}}$ (Fig. 5.4l). However, the relationship between fW and $\text{BFI}_{\text{trans}}$ was highly heteroscedastic, demonstrating high variability at low fW and low variability at high fW. Among the four soil indices, moderate (positive) relationships were obtained with SAND (Figs. 5.4p and 5.5p).

5.4.2 Catchment-scale estimation of BFI and k

Table 5.2 shows mean R^2 and RMSE values (mean of ten cross-validation iterations) obtained by the MLP models for the training and testing subsets of catchments. The mean training and testing R^2 values for $\text{BFI}_{\text{trans}}$ are 0.73 and 0.65, respectively, whereas the corresponding values for k_{trans} are somewhat poorer at 0.62 and 0.52 (Table 5.2). Fig. 5.6 shows scatterplots of estimated vs. observed values of $\text{BFI}_{\text{trans}}$ and k_{trans}, including the linear regression line. The estimated values are the median estimates of the ten MLP models (one for each cross-validation iteration). The associated R^2 values are 0.82 and 0.73 for the respective baseflow characteristics (Fig. 5.6), and thus are substantially higher than the mean training R^2 values (Table 5.2). The models tend to overestimate (underestimate) low (high) values of $\text{BFI}_{\text{trans}}$ and k_{trans} due to the statistical phenomenon of regression toward the mean (Galton, 1886; Bland and Altman,

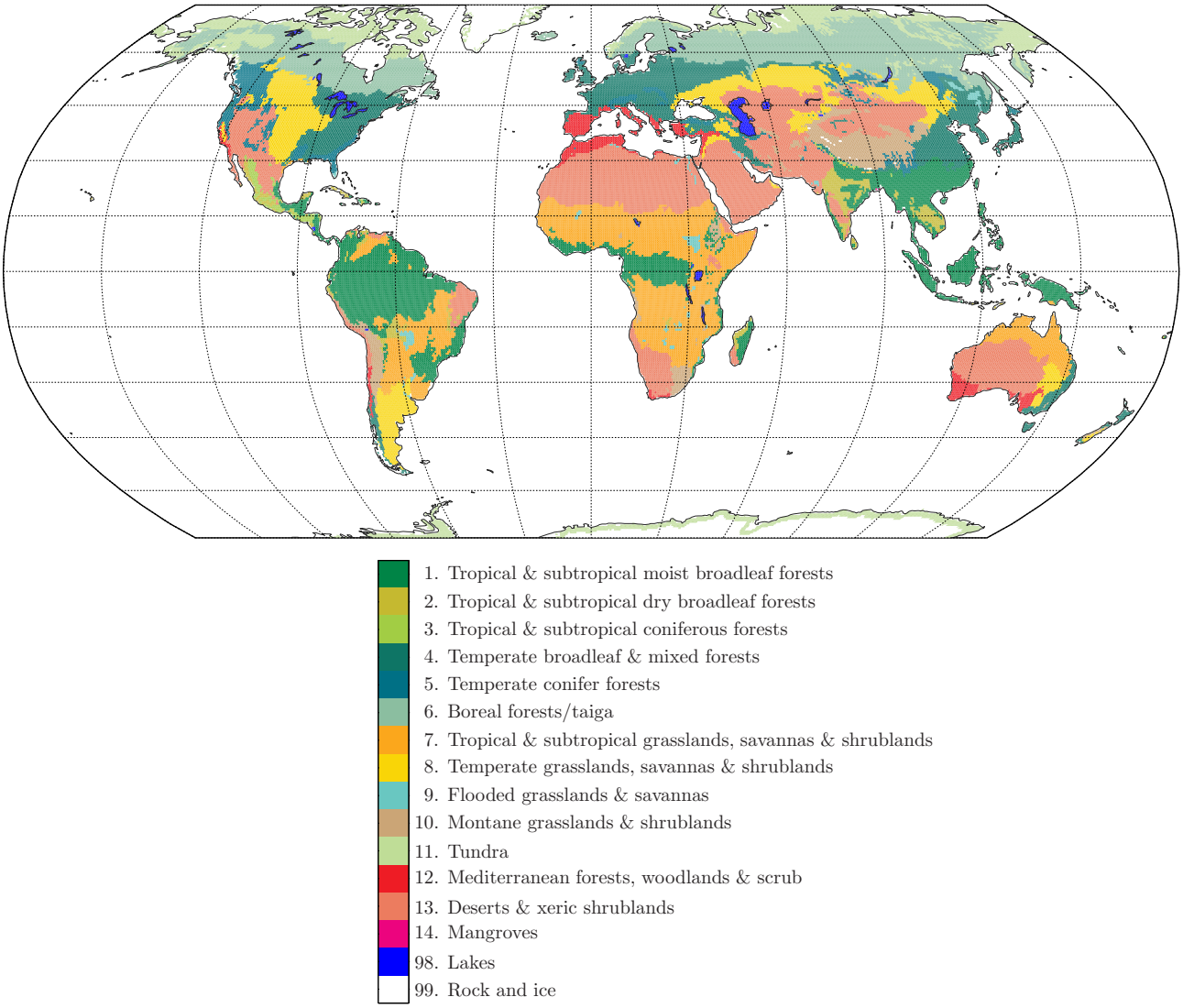


Figure 5.3: The WWF terrestrial biomes map (Olson et al., 2001).

Table 5.2: Mean R^2 and RMSE statistics obtained for the training and testing subsets of catchments.

Baseflow characteristic	Mean R^2		Mean RMSE [-]	
	Training	Testing	Training	Testing
BFI_{trans}	0.73	0.65	0.12	0.13
k_{trans}	0.62	0.52	0.36	0.41

1994).

5.4.3 Global maps of BFI and k

Figs. 5.7a and 5.8a present global maps of BFI and k, respectively. These were produced in turn, using global physiographic data as input to the ten trained MLP models (one for each cross-validation iteration), calculating the per-pixel median and applying the regression equations shown in Fig. 5.6 to correct for the regression toward the mean phenomenon, and back-transforming the result. Higher BFI and k are found in the tropics, in the tundra-taiga zone, and for mountain ranges (e.g., the Chilean Andes, the Rocky Mts, and the Himalayas). To facilitate the comparison of spatial patterns, Figs. 5.7b and 5.8b show the observed values. The BFI and k estimates appear to be relatively unbiased and the maps generally agree well with the observations (Figs. 5.7 and 5.8).

Fig. 5.9 shows for each 0.25° latitude band the median and degree of dispersion of the global BFI and k estimates. High values of the baseflow signatures are generally found north of ∼40°N and between ∼15°S and ∼7°N (Fig. 5.9). There is a markedly lower spread in k values north of ∼60°N (Fig. 5.9). Table 5.3 lists for each baseflow characteristic and terrestrial biome the median of the global maps. The variability in median BFI between biomes is relatively small, with values ranging from 0.53 (deserts & xeric shrublands) to 0.78 (temperate conifer forests). Values for k range from 0.74 (deserts & xeric shrublands) to 0.95 (boreal forests/taiga, tundra, and rock and ice).

Fig. 5.10 shows for the transformed baseflow char-

Table 5.3: Median estimated BFI and k for the WWF terrestrial biomes.

Biome[a]	BFI [-]	k [d^{-1}]
1. Tropical & subtropical moist broadleaf forests	0.72	0.94
2. Tropical & subtropical dry broadleaf forests	0.59	0.89
3. Tropical & subtropical coniferous forests	0.66	0.94
4. Temperate broadleaf & mixed forests	0.74	0.93
5. Temperate conifer forests	0.78	0.94
6. Boreal forests/taiga	0.75	0.95
7. Tropical & subtropical grasslands, savannas & shrublands	0.67	0.88
8. Temperate grasslands, savannas & shrublands	0.67	0.87
9. Flooded grasslands & savannas	0.70	0.90
10. Montane grasslands & shrublands	0.69	0.90
11. Tundra	0.71	0.95
12. Mediterranean forests, woodlands & scrub	0.60	0.86
13. Deserts & xeric shrublands	0.53	0.74
99. Rock and ice	0.74	0.95

[a] Fig. 5.3 shows the biome map. Only biomes with > 60 pixels of 0.25° resolution with a value are shown here.

acteristics (BFI_{trans} and k_{trans}) the uncertainty of the global estimates, as computed from the estimation spread of the ten trained MLP models. For both baseflow characteristics a lower uncertainty is found in North America, Europe, and southeastern Australia, whereas greater uncertainty is associated with arid, semi-arid, and Arctic regions (Fig. 5.10).

5.5 Discussion

5.5.1 Physiographic controls of BFI and k

Table 5.4 provides an overview of previous studies of (non-transformed) BFI or k regionalization. Although most of these studies were regional in nature, three had a continental scope (Gustard and Irving, 1994; Schneider et al., 2007; Santhi et al., 2008), and one covered the entire tropics (Peña-Arancibia et al., 2010). The most commonly used climate-related indices were mean annual precipitation (P; Lacey and Grayson, 1998; Haberlandt et al., 2001; Mazvimavi et al., 2005; Longobardi and Villani, 2008; Peña-Arancibia et al., 2010; Krakauer and Temimi, 2011), aridity index (AI; Lacey and Grayson, 1998; Mwakalila et al., 2002; Van Dijk, 2010; Peña-Arancibia et al., 2010), and mean annual potential evaporation (PET; Lacey and Grayson, 1998; Van Dijk, 2010). The studies generally reported a positive relationship of BFI and k with AI, a negative relationship with PET, and an inconsistent relationship with P. This is all in agreement with the present results (Figs. 5.4a, 5.5a, 5.4c, 5.5c, 5.4e, and 5.5e, respectively). However, AI was only weakly related to BFI_{trans} and the relationship between AI and k_{trans} was characterized by a high degree of heteroscedasticity (Figs. 5.4a and 5.5a, respectively). The present results therefore suggest that catchments with a high evaporative demand dry out faster after rainfall events, resulting in flows dominated by short-duration events (i.e., low BFI and k).

The relationships obtained here between catchment-mean surface slope (SLO) and BFI_{trans} or k_{trans} were positive (Figs. 5.4k and 5.5k, respectively), which is in agreement with several other regionalization studies (Lacey and Grayson, 1998; Longobardi and Villani, 2008; Peña-Arancibia et al., 2010; Van Dijk, 2010; Krakauer and Temimi, 2011) and a sensitivity experiment using TOPMODEL (Wolock et al., 1989). However, the relationships were rather weak and somewhat heteroscedastic (Figs. 5.4k and 5.5k), indicating that topography is not a major control of baseflow, particularly in catchments with gentle slopes. Consequently, the use of TOPMODEL-based runoff parameterizations in such catchments may be inappropriate (cf. Beven, 1997; Li et al., 2011). Conversely, several other regionalization studies found negative relationships between SLO and BFI (Haberlandt et al., 2001; Mazvimavi et al., 2005) or k (Zecharias and Brutsaert, 1988; Post and Jakeman, 1996; Brandes et al., 2005), although these studies used only a limited number of catchments (≤ 52). Additionally, the positive relationships found here (and by others) seems to contradict general drainage theory, which predicts a negative relationship between SLO and BFI or k, based on the premise that more steeply sloping aquifers drain faster (Brutsaert and Nieber, 1977; Zecharias and Brutsaert, 1988; Vogel and Kroll, 1992). There are two potential explanations for this discrepancy. The first is that SLO may be a poor proxy for the aquifer flow gradient. However, this explanation fails to clarify why a postive relationship was obtained. The second is that SLO (which is based on relatively high-resolution SRTM data, cf. Table 5.1) acts as a surrogate for hydrologic characteristics of soils and geology in the absence of more detailed data on substrates. Topography is one of the primary influences on pedogenesis (Price, 2011), with soils forming

Table 5.4: Overview of BFI or k regionalization studies, listed in order of the number of Q gauges used.

Baseflow characteristic	Reference	Number of gauges	Region	Climate	Model[a]	R^{2}[b]	RMSE[b]	Predictor(s)[c]
BFI	Mwakalila et al. (2002)	15	Tanzania	Semi-arid	LIN	0.89	-	AI, catchment fraction underlain by permeable geology
	Ahiablame et al. (2013)	22	Indiana, USA	Temperate	LOG	0.91	-	fW, catchment fractions of 2 soil infiltration capacity classes
	Haberlandt et al. (2001)	25	NE Germany	Temperate	LIN	0.87	0.07	SLO, TWI, soil hydraulic conductivity, P
	Longobardi and Villani (2008)	28	Southern Italy	Mediterranean	LIN	0.68	0.17	Catchment fraction underlain by permeable geology
					LIN	0.80	-	Catchment fraction underlain by permeable geology, P, ELEV, SLO, fTC
	Bloomfield et al. (2009)	44	Thames Basin, UK	Temperate	LIN	0.89	0.09	Catchment fractions of six hydrogeological classes
	Mazvimavi et al. (2005)	52	Zimbabwe	Subtropical	LIN	0.75	0.09	P, drainage density, 75th percentile of slopes in the catchment
					ANN	0.77	0.02	P, drainage density, 10th percentile of slopes in the catchment
	Schneider et al. (2007)	103	Europe	Temperate	LIN	-	-	Catchment fractions of 29 soil classes
	Lacey and Grayson (1998)	114	VIC, Australia	Temperate	LIN	0.72	-	Catchment area, ELEV, PET, fTC, catchment fraction underlain by sedimentary rock, P, channel length
	Van Dijk (2010)	183	SE Australia	Temperate, subtropical	EXP	0.34	0.16	PET
	Boorman et al. (1995)	575	UK	Temperate	LIN	0.79	0.09	Catchment fractions of 29 soil classes
	Neff et al. (2005)	959	Great Lakes Basin, North America	Temperate	EXP	0.53	0.11	Surficial geology, fW
	Gustard and Irving (1994)	1530	Europe	Temperate	LIN	0.46	-	Catchment fractions of nine soil classes
	Current study[d]	3520	Global	Various	ANN	0.73	0.10	AI, $\mathrm{AI_{si}}$, P, P_{si}, PET, $\mathrm{PET_{si}}$, TA, $\mathrm{TA_{si}}$, SNOW, ELEV, SLO, fW, fTC, NDVI, GRAV, SAND, SILT, CLAY
	Santhi et al. (2008)	∼8600	Conterminous USA	Arid, semi-arid, temperate, subtropical	LIN	0.79	-	SAND, elevation range of catchment
k	Post and Jakeman (1996)	16	VIC, Australia	Temperate	LIN	0.53	-	SLO
					LIN	0.41	-	Catchment shape
	Brandes et al. (2005)	24	Pennsylvania, USA	Temperate	POW	0.80	0.02	Drainage density, soil hydraulic conductivity, SLO
	Demuth and Hagemann (1994)	57	SW Germany	Temperate	POW	0.70	-	Catchment fractions of 14 hydrogeological classes
	Krakauer and Temimi (2011)	61	USA	Temperate	LIN	0.3–0.5	-	Longitude, soil infiltration capacity, latitude, channel length, fTC, P
	Hughes (1997)	134	South Africa	Subtropical, mediterranean	LIN	0.25	-	Rainfall seasonality classes, modeled estimates of recharge, geological index
	Peña-Arancibia et al. (2010)	167	Entire tropics	(Sub)tropical	LOG	0.49	-	AI
					EXP	0.49	-	P
	Van Dijk (2010)	183	SE Australia	Temperate, subtropical	POW	0.27	-	AI
	Current study[d]	3520	Global	Various	ANN	0.62	0.03	AI, $\mathrm{AI_{si}}$, P, P_{si}, PET, $\mathrm{PET_{si}}$, TA, $\mathrm{TA_{si}}$, SNOW, ELEV, SLO, fW, fTC, NDVI, GRAV, SAND, SILT, CLAY

[a] Lists the type of (multi-variate) regression equation used. Abbreviations used: LIN, linear; POW, power; LOG, logarithmic; EXP, exponential; and ANN, artificial neural network.
[b] Dash indicates the information was not reported in the study or is not directly comparable to other studies. The RMSE for k is expressed in units d^{-1}.
[c] Lists the predictors included in the model only and not all predictors considered. See Table 5.1 for descriptions of the predictor variables.
[d] The current study is added for the sake of completeness. The R^2 of the current study represents the mean training R^2 (cf. Table 5.2), whereas the RMSE represents the mean RMSE computed from back-transformed observed and estimated values of the baseflow characteristics for the training subsets. Note that the RMSE in Table 5.2 was computed from transformed observed and estimated values of the baseflow characteristics.

on steep slopes often being more permeable than their counterparts on gentle slopes (e.g., Ciolkosz et al., 1989; Janeau et al., 2003; Soulsby and Tetzlaff, 2008). Since permeable soils favor higher BFI and k relative to less permeable soils (Boorman et al., 1995) this could explain the positive relationships obtained between SLO and the baseflow characteristics. A counter argument could be that bedrock usually occurs at more shallow depths on steep slopes than on gentler slopes, although bedrock is not necessarily impermeable (Davis, 1969; Tromp-van Meerveld et al., 2007) and can contribute a considerable portion of total Q (Uchida et al., 2003).

Small-scale experimental studies demonstrate a clear relationship between the fraction of forest (fTC) and annual water yield (Bosch and Hewlett, 1982; Brown et al., 2005; Jackson et al., 2005), but this relationship does not always hold at the meso- to large-catchment-scale (e.g., Wilk et al., 2001; Zhou et al., 2010; Van Dijk et al., 2012; Peña-Arancibia et al., 2012). Similarly, no evidence of catchment-scale relationships between fTC and the transformed baseflow characteristics was found here (Figs. 5.4m and 5.5m) nor in several other BFI regionalization studies (Demuth and Hagemann, 1994; Mazvimavi et al., 2005; Longobardi and Villani, 2008). Although other regionalization studies did find relationships between fTC and BFI (Lacey and Grayson, 1998) or k (Brandes et al., 2005; Peña-Arancibia et al., 2010; Krakauer and Temimi, 2011), the direction and strength of the relationships varied (cf. Price, 2011). By contrast, our results between the fraction of open water (fW) and $\text{BFI}_{\text{trans}}$ or k_{trans} were positive, albeit somewhat heteroscedastic (Figs. 5.4l and 5.5l, respectively). This suggests that when large areas of a catchment are occupied by lakes, wetlands, or reservoirs (i.e., $f\text{W} > 0.08$) the flow is delayed (i.e., high BFI and k). Similar results were obtained in BFI regionalization studies for the Great Lakes region of North America (Neff et al., 2005) and Indiana, USA (Ahiablame et al., 2013).

Soils and geology are undoubtedly two of the dominant controls of baseflow (Farvolden, 1963; Davis, 1969; Tague and Grant, 2004) as confirmed by the fact that 16 of the 20 considered regionalization studies incorporated one, or more, indices related to soils or geology into their models (Table 5.4). Here, four indices related to the soil were used. Moderate (positive) relationships with the mean sand content of the soil (SAND) were found (Figs. 5.4p and 5.5p), in agreement with BFI regionalization studies for the conterminous USA (Santhi et al., 2008) and the UK (Boorman et al., 1995). The weaker relationship found here (Fig. 5.4p) is most likely due to the relatively poor quality of the global HWSD data set compared to the more regional soil data sets used in the two previously mentioned studies.

5.5.2 Catchment-scale estimation of BFI and k

Most regionalization studies used multi-variate linear regression (Table 5.4) and non-transformed values of the baseflow characteristics (BFI and k). However, the negative skew of the BFI and k distributions found here suggests the need for transformation of the data to achieve greater normality. Additionally, the highly non-linear relationships between the predictors and the (transformed) baseflow characteristics obtained here (Figs. 5.4 and 5.5; cf. Van Dijk, 2010; Peña-Arancibia et al., 2010) suggest the usefulness of ANNs. Many regionalization studies have used a relatively small number of catchments (< 100; Table 5.4), which may have led to less reliable and/or overfitted models. In general, studies with < 100 catchments obtained higher training R^2 values than studies based on a larger number of catchments (≥ 100; Table 5.4). For $\text{BFI}_{\text{trans}}$, the training R^2 value obtained here (0.73; Table 5.2) falls in the upper range of values reported for other studies with ≥ 100 catchments (0.34–0.79; Table 5.4). For k_{trans}, the training R^2 value obtained here (0.62; Table 5.2) is much higher than the corresponding values reported for other studies with ≥ 100 catchments (0.25–0.49; Table 5.4). However, it is pertinent to note that two of these studies (Van Dijk, 2010; Peña-Arancibia et al., 2010) incorporated only one (climate-related) predictor in their model, whereas better results might have been achieved using multiple predictors.

Assessing a model's performance on an independent data set (i.e., to provide an estimate of generalization) is a crucial aspect of model development. Here, mean testing R^2 values of 0.65 and 0.52 were obtained for $\text{BFI}_{\text{trans}}$ and k_{trans}, respectively (Table 5.2). We consider these results to be acceptable given the large geographic spread and the wide range of geology, soils, topography, climate, and land use covered by the catchments. Only three other (BFI) regionalization studies conducted a generalization assessment of their established models (Haberlandt et al., 2001; Schneider et al., 2007; Bloomfield et al., 2009). We strongly urge all future researchers to explicitely report generalization assessment statistics in their papers. Of these three studies, only one reported the associated statistical measures (Haberlandt et al., 2001). Haberlandt et al. (2001) used multi-variate linear regression, ordinary kriging, and external drift kriging to estimate BFI from ten physiographic characteristics for 25 catchments located in the German part of the Elbe River Basin. Based on a leave-one-out cross-validation they obtained RMSE values of 0.09, 0.11, and 0.08 for the respective approaches. Their RMSE values are comparable to the mean RMSE value of 0.11 computed here from back-transformed observed and estimated BFI for the testing subsets (noting that the value of 0.13 in Table 5.2 was derived from transformed BFI values).

The performance statistics obtained for $\text{BFI}_{\text{trans}}$ were better than those for k_{trans} (cf. Table 5.2 and Fig. 5.6), in

agreement with previous studies using non-transformed values (Table 5.4). The better performance of BFI may be attributable to several factors. First, k has been found to vary somewhat seasonally in response to changes in actual evaporation (Czikowsky and Fitzjarrald, 2004). Moreover, if the baseflow recession of a catchment is non-linear, the assumption of a linear reservoir to derive k as used here will lead to different estimates of k depending on the flow rate (Wittenberg, 1999; Krakauer and Temimi, 2011). Additionally, k is calculated from low-flow periods in the Q record that are subject to relatively lower instrument precision than intermediate flows (Carter, 1963) and are affected by rating-curve uncertainty (Tomkins, 2013). Finally, k suffers from a greater sampling error than BFI as it is computed from only parts of the overall Q record.

The R^2 values of the scatterplots exceed the mean training R^2 values (cf. Fig. 5.6 and Table 5.2) because the estimated values in the scatterplots represent the mean of the ten MLP models (one for each cross-validation iteration). Ensemble modeling (i.e., using the outputs from multiple models or from different realizations of the same model) is widely used in atmospheric and climate sciences and is known to typically improve predictive accuracy (e.g., Tebaldi and Knutti, 2007; Wandishin et al., 2001). Additionally, several studies have reported that using ensembles of neural networks improves the accuracy compared to using single neural networks (Hansen and Salamon, 1990; Tetko et al., 1995).

5.5.3 Global maps of BFI and k

To the best of our knowledge our study is the first attempt to estimate BFI and k globally. Santhi et al. (2008) produced a BFI map for the conterminous USA by interpolating the BFI values of ∼8600 catchments. The BFI map of Santhi et al. (2008) and our newly derived map (Fig. 5.7a) exhibit very similar spatial patterns, although the latter contains considerably more detail. Neff et al. (2005) presented a BFI map for the Great Lakes region of North America based on exponential relationships with the fraction of open water (fW) and a geological index. Comparing the BFI map of Neff et al. (2005) with our map (Fig. 5.7a) reveals that the mean BFI for this region is similar at ∼0.75. Bullock et al. (1997) produced a BFI map for Southern Africa by assigning the BFI values computed from the Q records of ∼650 gauging stations to the associated catchments. Some of the catchments used by Bullock et al. (1997) were rather large (8 % of the catchments were $> 10\,000$ km^2 and 2 % $> 100\,000$ km^2), which may have led to inflated BFI values due to channel routing effects. Nevertheless, the BFI map of Bullock et al. (1997) and the corresponding part of our map (Fig. 5.7a) agree well in terms of spatial patterns—both maps showed markedly higher BFI values north of ∼15°S. Lee et al. (2006) produced a BFI map for Taiwan by interpolation of BFI values computed from the Q records of 174 gauging stations. Again, our map and that of Lee et al. (2006) showed very similar spatial patterns. Similarly, the BFI maps of Haberlandt et al. (2001) for the German part of the Elbe River Basin and our map agreed well; both maps give a mean BFI of ∼0.75 and place the highest BFI values in the central part of the Elbe.

Peña-Arancibia et al. (2010) produced a k map for the entire tropics extending between 35°S and 30°N based on an exponential relationship with long-term mean P. The most notable difference between the current map (Fig. 5.8a) and the map of Peña-Arancibia et al. (2010) is that the latter only shows high k values (i.e., $k > 0.95$) near the equator, whereas our newly produced map indicates such high k values between ∼15°S and ∼7°N (Fig. 5.9). Visual inspection suggests that our current map (Fig. 5.8a) agrees slightly better with the observed values for the catchments used by Peña-Arancibia et al. (2010) and the observed values for the catchments used here (Fig. 5.8b).

The global maps of the estimation uncertainty for BFI and k exhibit similar patterns (Fig. 5.10). The generally greater uncertainty in arid, semi-arid, and Arctic regions (Fig. 5.10) is due to application of the MLP models outside the physiographic domain of the catchment data. Additional efforts are recommended to validate and/or improve the present results for these regions. In light of this, the declining number of Q gauging stations in operation around the globe is cause for concern (Stokstad, 1999; Shiklomanov et al., 2002; Fekete and Vörösmarty, 2007).

The global maps of BFI and k produced here are likely useful for a wide range of large-scale hydrological applications, including the diagnosis and parameterization of land surface schemes and global hydrological models, water resource assessments, catchment classification, and groundwater recharge estimation. However, some important characteristics and limitations should be noted. First, the maps reflect flows under natural, unregulated conditions. Second, in cold regions the main source of baseflow is a combination of snow and glacier melt and not groundwater. Third, the maps are not representative of flows in catchments $> 10\,000$ km^2 due to the increasing importance of channel routing effects at larger scales. This means that the estimated BFI and k should be used for Q from contributing catchments and not Q routed through larger basins. Finally, although the maps have a resolution of 0.25°, their effective resolution is lower due to the lower resolution of some of the original input data (cf. Table 5.1).

5.6 Conclusion

This study is the first attempt to estimate two important baseflow characteristics (BFI and k) globally. A highly heterogeneous set of 3520 catchments was used to construct widely applicable models relating physiographic characteristics to BFI and k. The main findings are:

1. Since the BFI and k distributions showed negative skewness, a data transformation was needed to better approximate a normal distribution, required to avoid bias in the estimates. The relationships between catchment physiographic characteristics and the transformed baseflow characteristics (BFI_{trans} and k_{trans}, respectively) were often highly non-linear and heteroscedastic. Among the catchment physiographic characteristics pertaining to climate, the aridity index seasonality (AI_{si}), mean annual potential evaporation (PET), PET seasonality (PET_{si}), mean annual air temperature (TA), and mean snow-water equivalent depth (SNOW) were best related to BFI_{trans}, whereas the aridity index (AI), AI_{si}, and PET were best related to k_{trans}. The positive relationships found between mean surface slope (SLO) and BFI_{trans} or k_{trans} seem to contradict classical drainage theory and may represent a spurious relationship due to underlying patterns in soil hydrology and hydrogeology. Among the predictors pertaining to land cover, the fraction of forest (fTC) and the mean Normalized Difference Vegetation Index (NDVI) were related to neither BFI_{trans} nor k_{trans}, whereas the fraction of open water (fW) showed moderate (positive) relationships with both BFI_{trans} and k_{trans}. Positive but weak relationships were found between the mean sand content of the soil (SAND) and both BFI_{trans} and k_{trans}.
2. The non-linear relationships obtained between catchment physiographic characteristics and BFI_{trans} or k_{trans} justified the use of artificial neural networks to estimate BFI_{trans} and k_{trans}. It proved possible to satisfactorily estimate BFI_{trans} and k_{trans} from catchment physiographic data, yielding training R^2 values of 0.73 and 0.62, respectively, although high (low) values of BFI_{trans} and k_{trans} were slightly underestimated (overestimated) due to regression toward the mean. It was found that averaging the estimates of the ten multi-layer perceptron models (one for each cross-validation iteration) resulted in more accurate estimates. The results further show that artificial neural networks can be considered a viable and perhaps better alternative to the commonly used multi-variate linear regression.
3. Global maps of BFI and k were obtained by using global physiographic data as input to the established models and back-transforming the result. The BFI and k values showed higher uncertainty in comparatively data-poor arid, semi-arid, and Arctic regions, and lower uncertainty in the more data-rich North America, Europe, and southeastern Australia. The global maps will prove useful for a variety of large-scale hydrological applications, although further validation of the maps is recommended, particularly in poorly gauged and ungauged regions.

Acknowledgments

The first author was supported by Deltares (Delft, The Netherlands). The Global Runoff Data Centre (GRDC; Koblenz, Germany) is thanked for providing the flow and catchment boundary data used in the present study. We are also most grateful to Dr Tom Gleeson (University of British Columbia, Vancouver, Canada) for providing the global permeability map and Dr Jens Hartmann (Universität Hamburg, Hamburg, Germany) for providing the GLiM data set.

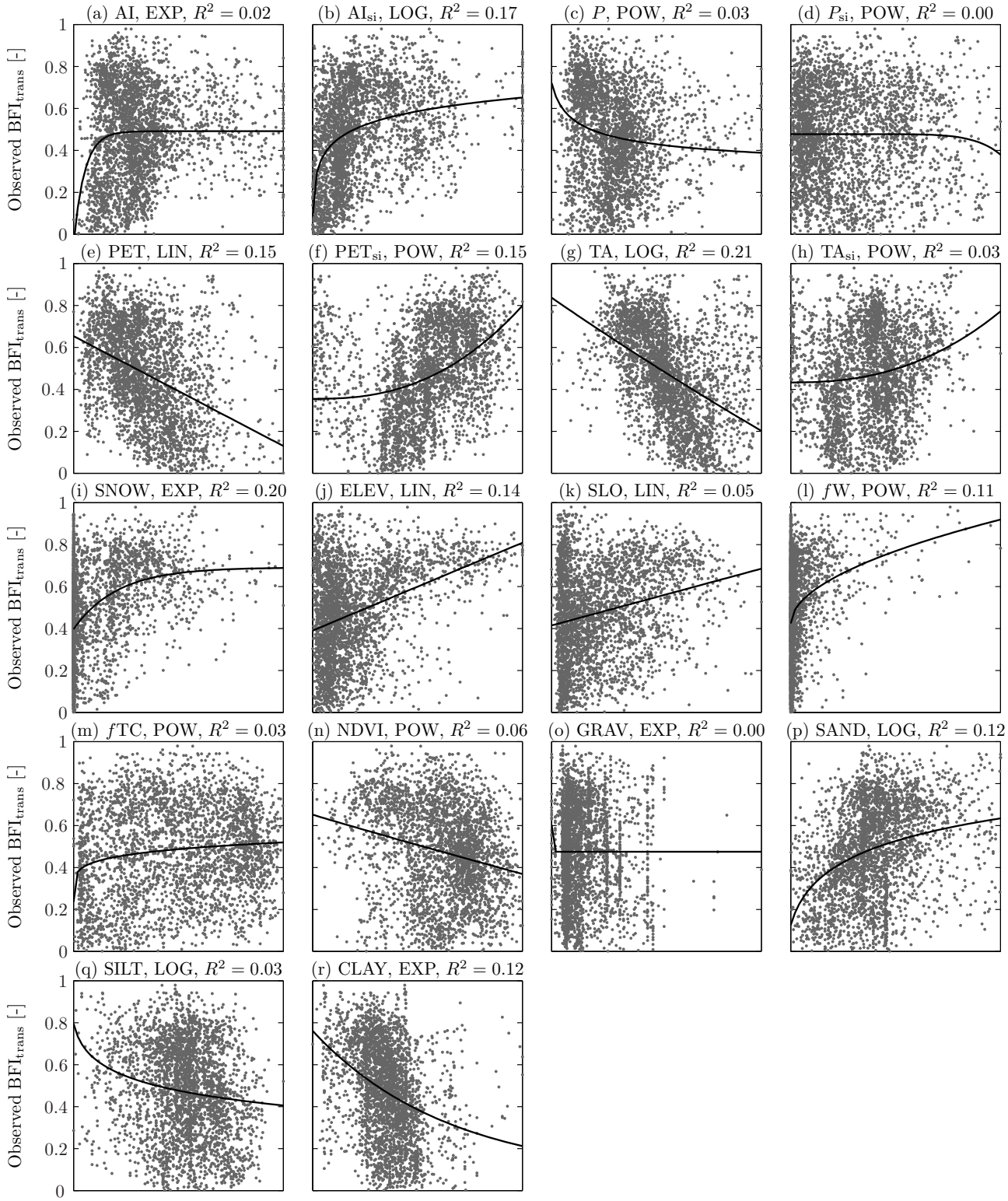


Figure 5.4: Scatterplots of catchment-mean values of the physiographic characteristics (along the x-axis) vs. the observed BFI_{trans} (along the y-axis), including the best-fit regression line. Each data point represents a catchment ($n = 3520$). The x-axis range is from the minimum to maximum value of the data (not shown). Table 5.1 lists descriptions of the predictor variables. Abbreviations referring to the type of regression equation: EXP, exponential; LIN, linear; LOG, logarithmic; and POW, power.

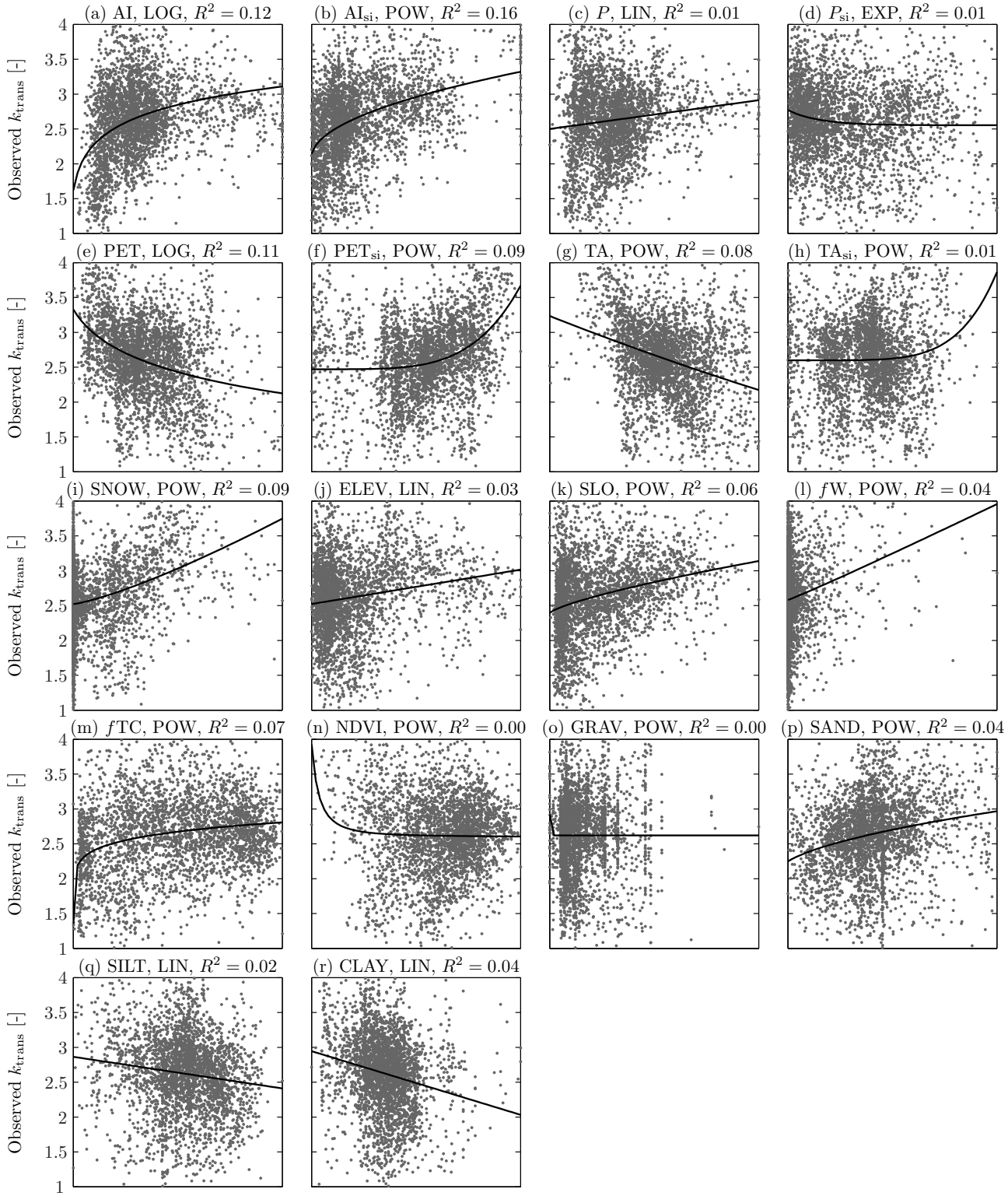


Figure 5.5: Scatterplots of catchment-mean values of the physiographic characteristics (along the x-axis) vs. the observed k_{trans} (along the y-axis), including the best-fit regression line. Each data point represents a catchment ($n = 3520$). The x-axis range is from the minimum to maximum value of the data (not shown). Table 5.1 lists descriptions of the predictor variables. Abbreviations referring to the type of regression equation: EXP, exponential; LIN, linear; LOG, logarithmic; and POW, power.

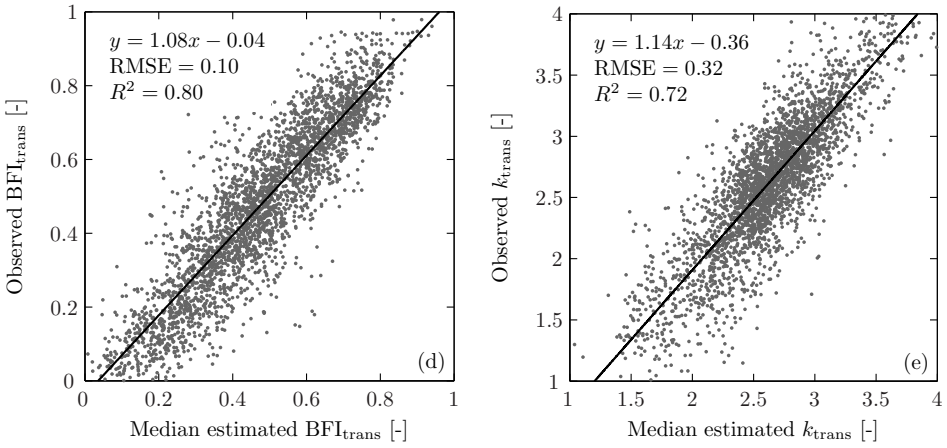


Figure 5.6: Scatterplots of estimated vs. observed values of (a) $\mathrm{BFI}_{\mathrm{trans}}$ and (b) k_{trans}, including the linear regression line. Each data point represents a catchment ($n = 3520$). The estimated values are the median estimates of the ten multi-layer perceptron models used (one for each cross-validation iteration). For each iteration the training, validation, and testing subsets were included. Scatterplots using the non-transformed baseflow characteristics (BFI and k) were not made as being based on non-normally distributed data these would result in non-robust regressions.

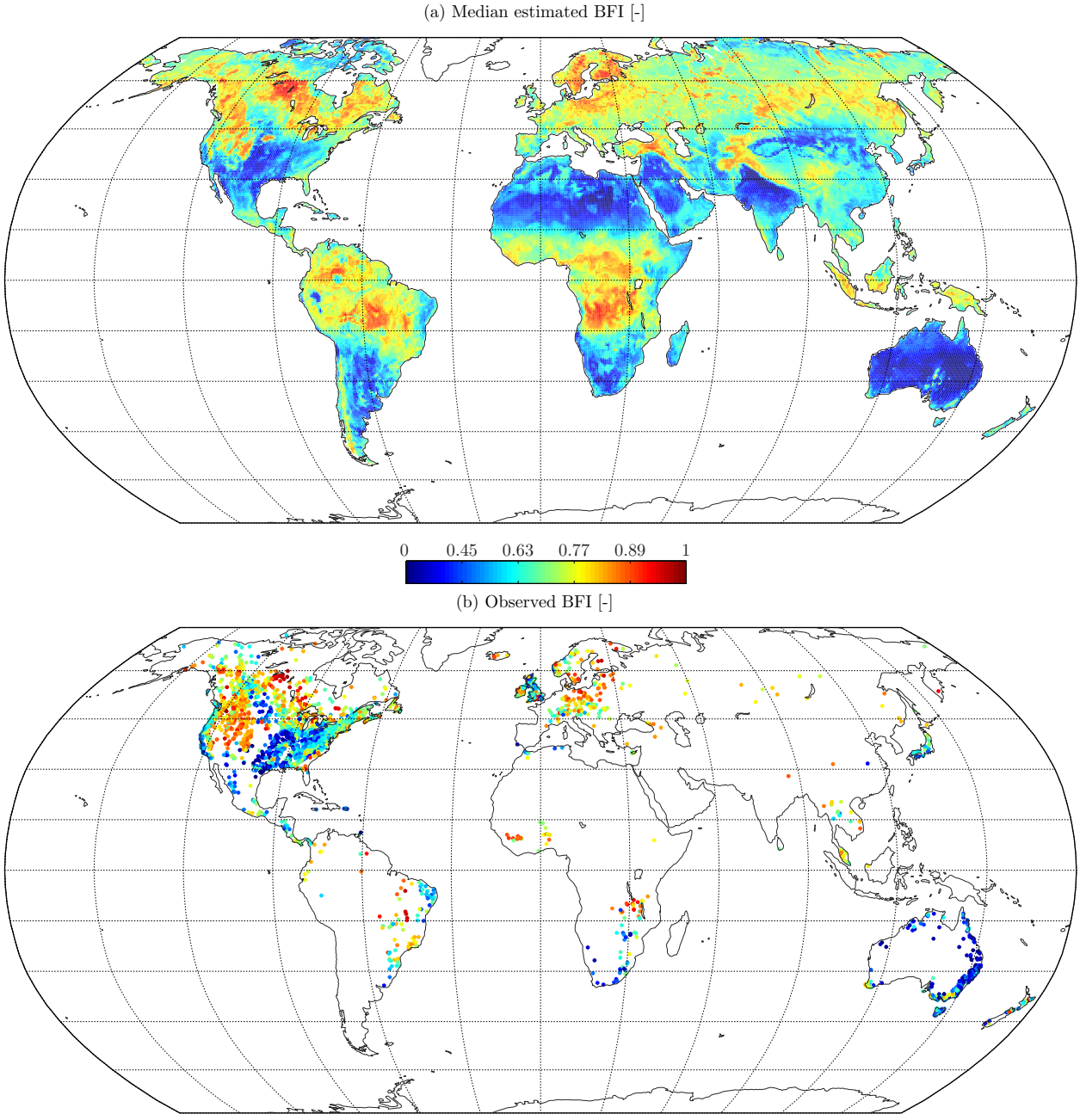


Figure 5.7: Global maps of (a) median estimated and (b) observed BFI. The estimated values in (a) are the back-transformed medians of the ten cross-validation iterations. Each data point in (b) represents a catchment centroid ($n = 3520$).

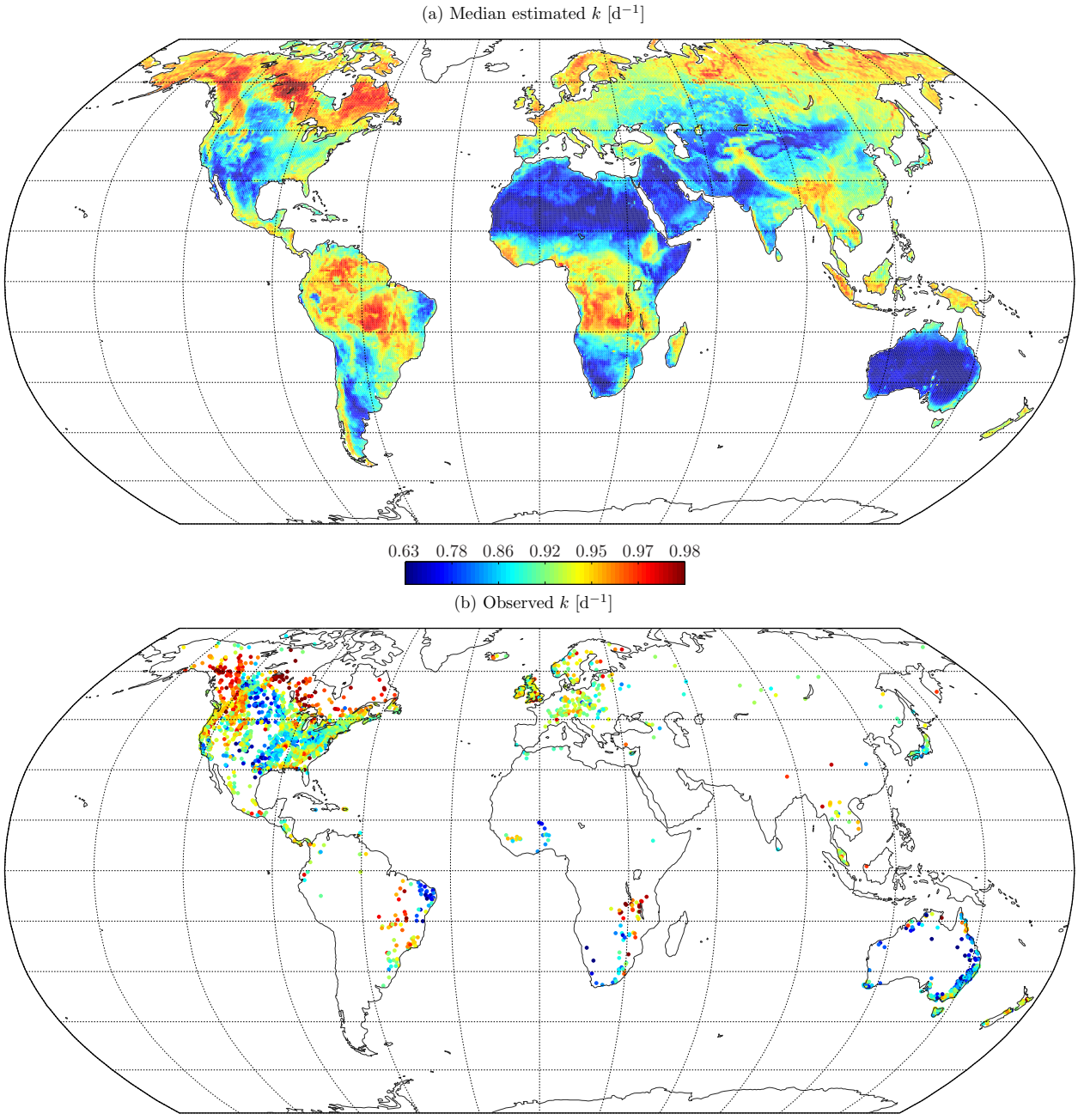


Figure 5.8: Global maps of (a) median estimated and (b) observed k. The estimated values in (a) are the back-transformed medians of the ten cross-validation iterations. Each data point in (b) represents a catchment centroid ($n = 3520$).

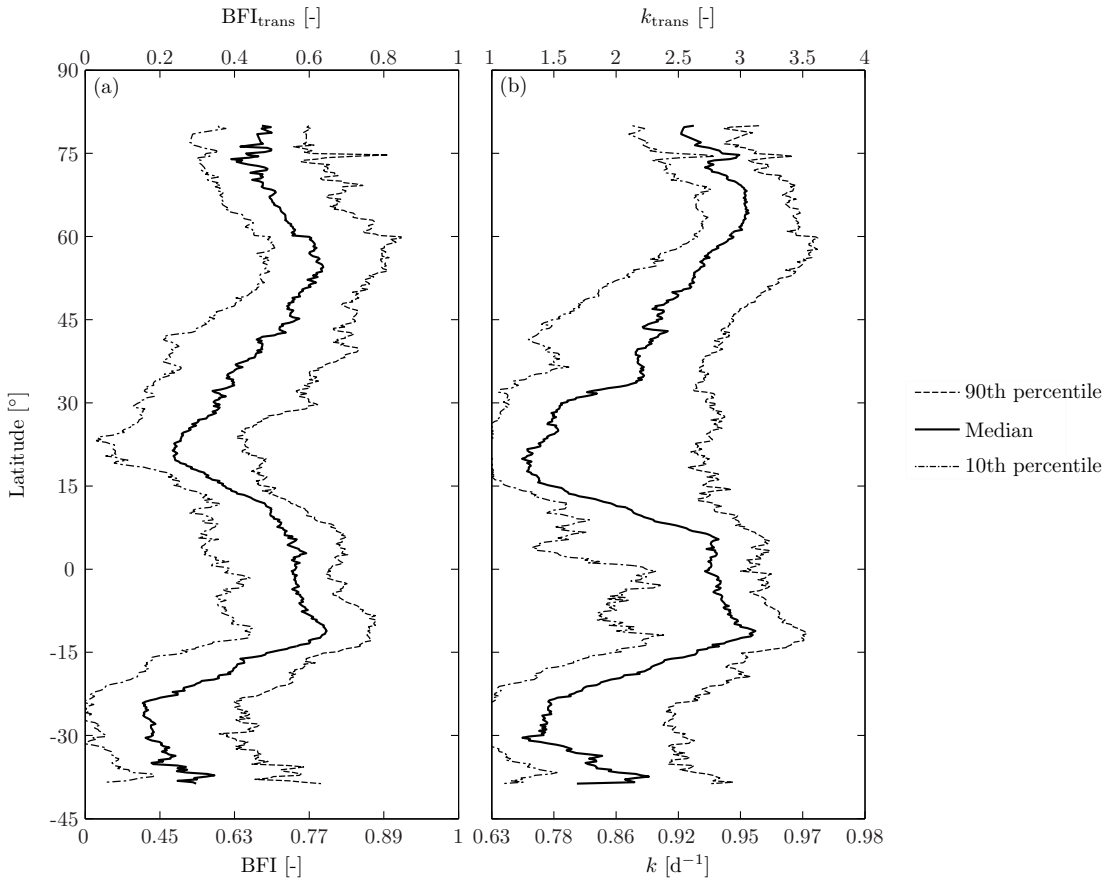
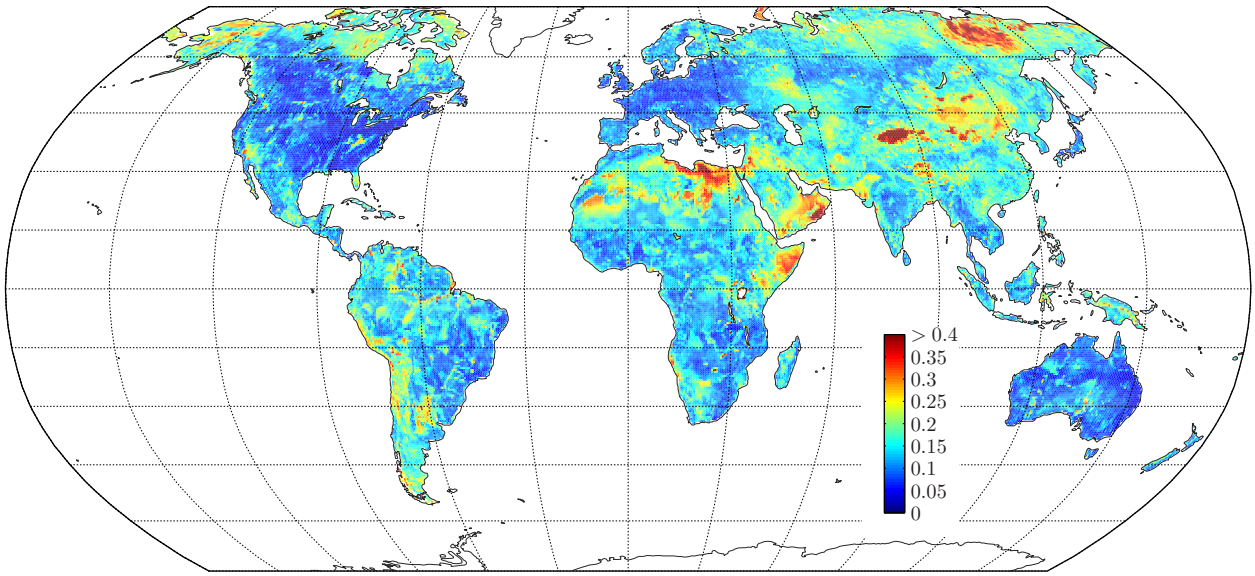


Figure 5.9: The 90th percentile, median, and 10th percentile values for (a) BFI and BFI$_{trans}$, and (b) k and k_{trans} for each 0.25° latitude band derived from the global maps. Values are plotted only if there are > 60 pixels of 0.25° resolution with a value within the latitudinal band.

(a) BFI_{trans} uncertainty [-]



(b) k_{trans} uncertainty [-]

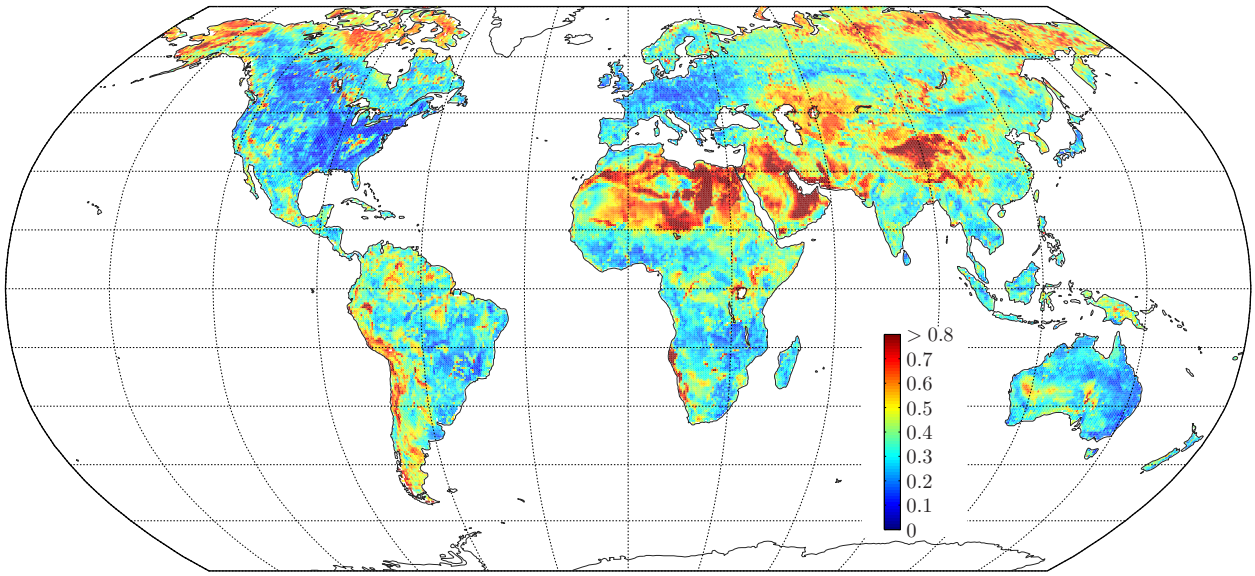


Figure 5.10: Global maps of the uncertainty of (a) BFI_{trans} and (b) k_{trans}. The data represent the per-pixel standard deviation of the ten estimates (one for each cross-validation iteration).

Chapter 6

Calibration of global hydrological models based on streamflow characteristics

Abstract. Streamflow (Q) estimation in ungauged catchments is perhaps the most fundamental challenge faced by hydrologists. A promising approach to calibrate hydrological models in ungauged catchments is to use estimated Q characteristics to identify acceptable model parameter sets. This study is the first to test this approach at the global scale using a large, heterogeneous catchment set. First, observed Q data from 3320 catchments were used to establish neural-network models to derive global maps of five selected Q characteristics, each describing a different aspect of the hydrograph. All Q characteristics could be estimated satisfactorily, with mean training R^2 values ranging from 0.62 to 0.86. Next, a conceptual rainfall-runoff model (HBV-Light) was calibrated in a lumped fashion for 200 independent catchments for 2003–2007 using values of the Q characteristics derived from the newly produced maps. A substantial improvement in the simulated Q characteristics was noted, which, in turn, led to improvements in most of the traditional Q performance measures as computed from simulated and observed Q time series. The comparison between the newly produced maps of the respective Q characteristics and estimates derived from two macro-scale hydrological models (Noah and PCR-GLOBWB) suggests that the maps can be employed to diagnose the runoff parameterization of the models. The methodology further offers unique possibilities for the diagnosis and/or calibration of future macro-scale hydrological models with high spatial resolution (down to 1 km).

6.1 Introduction

Understanding the spatio-temporal variability of streamflow (Q [mm d^{-1}]) is important for water resources management, for reliable drought and flood forecasting, for hydropower and irrigation systems, for maintaining aquatic habitats, and for sediment and contaminant transport (e.g., Brauman et al., 2007; Quintero et al., 2009; Cyr et al., 2011). Calibration is a crucial step in hydrological model application to obtain accurate Q estimates for a catchment. It involves the identification of one or more acceptable model parameter sets, typically by evaluating objective functions using observed and modeled Q (e.g., Madsen, 2000). However, observed Q data are unavailable for ungauged catchments and hence over the majority of the Earth's land surface (Fekete and Vörösmarty, 2007), thus requiring recourse to other approaches. One of the most commonly used approaches is the transfer of model parameters from gauged to ungauged catchments (Kim and Kaluarachchi, 2008, and references therein), based on geographic or physiographic proximity or using multivariate regression (He et al., 2011). However, this approach has had limited success due to the equifinality problem, where different model parameter sets may lead to similar results (Beven, 1993). An alternative approach that has been gaining popularity in recent years is the transfer of Q characteristics (e.g., Olden and Poff, 2003) from gauged to ungauged catchments and to use these to identify acceptable model parameter sets (Wagener and Montanari, 2011). This approach has three advantages: (1) it avoids the problem of equifinality; (2) it is model-independent; and (3) Q characteristics are more meaningful than model parameters. Several studies have demonstrated the usefulness of this approach (Yadav et al., 2007; Zhang et al., 2008a; Castiglioni et al., 2010; Lombardi et al., 2012; Pinheiro and Naghettini, 2012), although these studies had a regional scope and used a relatively small number of catchments (≤ 30).

Macro-scale hydrological models (land surface schemes and global hydrological models) are important tools designed to simulate the water and energy balance of the land surface at continental or global scales (Wood et al., 1997). They have a physically-based representation of the chief processes governing the water cycle and a large number of *a priori* estimated parameters describing physical characteristics of the land surface. Such macro-scale models are expected to provide reasonably accurate Q estimates for ungauged regions owing to their

physical basis. However, comparisons between hydrological models and macro-scale models have revealed Q estimates from macro-scale models to be relatively inaccurate (Duan et al., 2006; Nasonova et al., 2009), largely due to sub-par calibration (Beven, 1989; Duan et al., 2001). Some macro-scale models are uncalibrated, such as Noah-MP (Niu et al., 2011), Mac-PDM (Gosling and Arnell, 2011), and the Community Land Model (CLM; Oleson et al., 2010). Others have been crudely calibrated only, such as the Variable Infiltration Capacity (VIC) model (Liang et al., 1994; Nijssen et al., 2001) and WASMOD-M (Widén-Nilsson et al., 2007), which both use nearest-neighbor interpolation of calibrated model parameters, and WaterGAP (Döll and Fiedler, 2008), which has been calibrated for gauged catchments based on the runoff coefficient (the ratio of long-term Q to precipitation).

Recently, thanks to (on-going) Q data collection and verification efforts by the Global Runoff Data Centre (GRDC; Koblenz, Germany; `http://grdc.bafg.de`), global maps of two important baseflow-related Q characteristics were produced using neural-network (NN) models and global physiographic data (Beck et al., 2013b). Such maps present a unique opportunity to calibrate hydrological models for the entire land surface including ungauged regions. The current paper tests whether the global maps of five Q characteristics can be used to calibrate a simple conceptual rainfall-runoff model (HBV-Light). The employed methodology consists of two stages. First, the respective maps are derived and their global patterns are compared to estimates derived from two commonly used macro-scale hydrological models. Second, the improvement in the simulated Q obtained by the HBV-Light model is quantified and the possibility of calibrating the two macro-scale models using the respective maps is explored.

6.2 Data

6.2.1 Observed streamflow

The observed daily Q data used here originate from three sources. First, Q data from the 1862 Model Parameter Estimation Experiment (MOPEX; Schaake et al., 2006) catchments located in the USA were downloaded from the US Geological Survey (USGS) National Water Information System (`http://waterdata.usgs.gov`). Catchment boundaries associated with the US Q data were obtained from the MOPEX webserver (`ftp://hydrology.nws.noaa.gov/pub/gcip/mopex/US_Data/`). Second, Q data from 4047 stations around the world from the GRDC streamflow dataset were considered. Third and finally, Q and associated catchment boundary data for 321 Australian stations compiled by Peel et al. (2000) were considered. Together this resulted in an initial dataset comprising 6230 Q stations.

For a catchment to be included several requirements had to be satisfied. First, to ensure that the used catchments were relatively undisturbed, < 2 % of the catchment area was allowed to be urban (using the "artificial areas" class of the GlobCover v2 map; Bontemps et al., 2011) or subject to irrigation (using the Global Irrigated Area Map; `http://www.iwmigiam.org`). Second, catchments needed to have an area of $< 10\,000$ km^2 to avoid any effects of channel routing. This resulted in a dataset comprising 3520 catchments. From this set, 200 catchments with a temporal Q data coverage of > 90 % during 2003–2007 (i.e., the time period for which the HBV-Light model was calibrated) were randomly selected and used for the evaluation of the HBV-Light model configurations, whereas the remaining 3320 catchments were employed to derive the global maps of the various Q characteristics. Fig. 6.1 shows the locations of the catchments. The Q data were converted to areal mean runoff in mm d^{-1} using the corresponding catchment areas.

6.2.2 HBV-Light model inputs

Daily time series of precipitation (P [mm d^{-1}]), air temperature (T [°C]), and net radiation (R_n [W m^{-2} d^{-1}]) were calculated for the catchments to run the HBV-Light model. For P the daily 0.25° gauge-based Climate Prediction Center (CPC) Unified v1.0/RT dataset (1979–2012; Xie et al., 2007; Chen et al., 2008) was used, and for T the daily 1° satellite-based Atmospheric Infrared Sounder (AIRS) AIRX3STD v005 dataset (2002–2011; Olsen, 2007). Gaps in the T record were filled using linear interpolation. Long-term means of CPC P were linearly transformed to WorldClim data (Hijmans et al., 2005), whereas long-term means of AIRS T were offset to WorldClim data. The WorldClim data represent high-resolution (1 km), elevation-corrected, long-term means. For R_n the daily 1° satellite-based NASA/GEWEX Surface Radiation Budget (SRB) v3.0/v3.1 dataset (1983–2007; `http://gewex-srb.larc.nasa.gov`) was used. All data were bilinearly interpolated to 0.25°. Catchment-mean time series of P, T, and R_n were calculated for 2003–2007 (i.e., the common time period for all datasets) for the 200 evaluation catchments.

6.2.3 Noah-based streamflow predictions

Simulated Q data from the Noah land surface scheme version 2.7.1 (Schaake et al., 1996; Ek et al., 2003) developed at the National Centers for Environmental Prediction (NCEP) were obtained from the NASA Goddard Earth Science (GES) Data and Information Service Center (DISC; `http://disc.sci.gsfc.nasa.gov/hydrology/data-holdings`). The model was driven by the Global Land Data Assimilation System (GLDAS; Rodell et al., 2004). The data have a 3-hourly temporal and 0.25° spatial resolution and cover the period 2000–present. Catchment-mean Q time series for the 200 evaluation catchments were computed for 2003–2007. Although simulated Q data from three other land surface

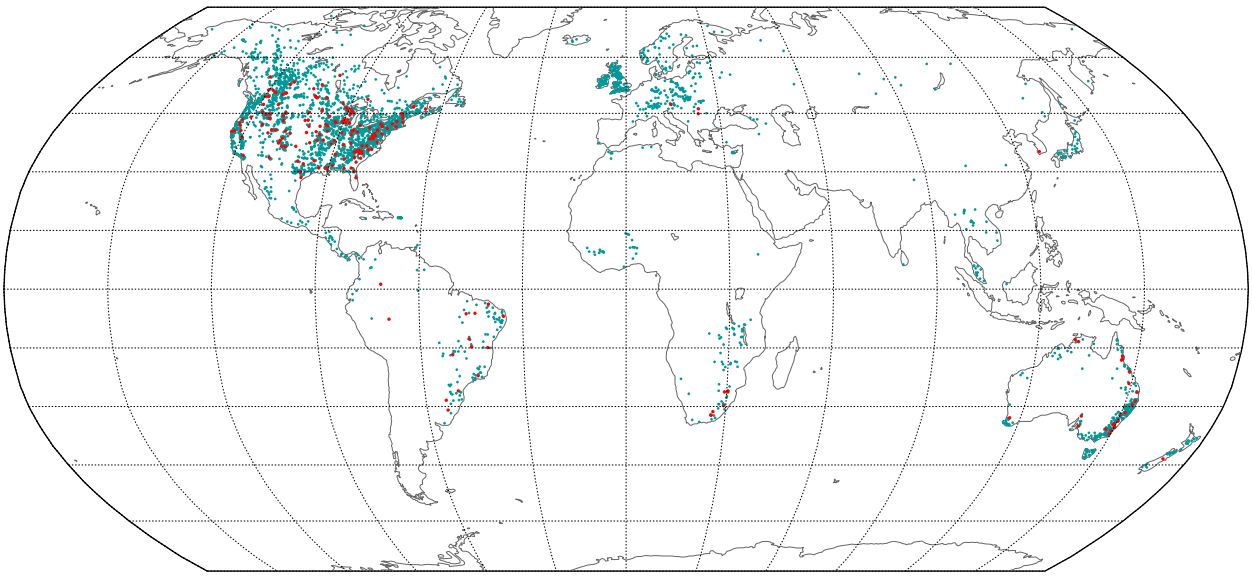


Figure 6.1: The locations of the catchments used here. Each data point represents a catchment centroid ($n = 3520$). Blue catchments were used to produce the global maps ($n = 3320$), whereas red catchments were used to assess the HBV-Light model configurations ($n = 200$).

schemes driven by GLDAS were available, these data were less suitable due to their 1° spatial resolution.

6.2.4 PCR-GLOBWB-based streamflow predictions

Simulated Q data from the PCRaster Global Water Balance (PCR-GLOBWB) global hydrological model (Bierkens and van Beek, 2009; Van Beek and Bierkens, 2009) developed at Utrecht University in the Netherlands were used. The model was driven by the ERA-Interim reanalysis dataset (Dee et al., 2011). The data have a daily temporal and 0.5° spatial resolution and cover the period 2003–2010. Catchment-mean Q time series for the 200 evaluation catchments were computed for 2003–2007 as well.

6.3 Methodology

6.3.1 Computation of streamflow characteristics

The five selected Q characteristics were: (1) mean annual runoff (MAR [mm yr^{-1}]); (2) 1st percentile (probability of non-exceedance) runoff (R1 [mm d^{-1}]); (3) 99th percentile runoff (R99 [mm d^{-1}]); (4) the baseflow index (BFI [mm d^{-1}]), defined as the ratio of long-term mean baseflow to total Q; and (5) the baseflow recession constant (k [d^{-1}]), defined as the rate of baseflow decay. The Q characteristics were chosen so as to represent the most important aspects of the hydrograph. Only five metrics were selected to allow the results of each to be discussed in sufficient detail. All Q characteristics were calculated from daily continuous Q time series (either simulated or observed), with BFI and k computed following Van Dijk (2010) with the "window size" set at 5 days. To normalize the distributions of the respective Q characteristics, values were transformed according to:

$$\begin{aligned} \mathrm{MAR}_{\mathrm{trans}} &= \ln(\mathrm{MAR}), \\ \mathrm{R1}_{\mathrm{trans}} &= \mathrm{R1}^{1/3}, \\ \mathrm{R99}_{\mathrm{trans}} &= \ln(\mathrm{R99}), \\ \mathrm{BFI}_{\mathrm{trans}} &= \mathrm{BFI}^2, \text{ and} \\ k_{\mathrm{trans}} &= -\ln(1-k), \end{aligned} \tag{6.1}$$

where the trans-subscript denotes the respective transformed values of the Q characteristic under consideration [-].

6.3.2 Global maps of streamflow characteristics

Global observation-based maps (0.125° spatial resolution) of the five Q characteristics were derived following Beck et al. (2013b), who used a highly heterogeneous set of 3548 catchments with areas $< 10\,000$ km^2 around the globe to construct neural-network (NN) models involving 18 physiographic characteristics to estimate BFI and k. The catchment set was divided into training and testing subsets for ten cross-validation iterations. For each iteration a NN model was trained using the training subset of catchments and independently evaluated using the testing subset of catchments, whereas for each subset R^2 and RMSE values were computed. Global maps of distributed BFI and k were subsequently obtained using global physiographic data as input for the ten NN models and calculating the per-pixel median. Uncertainty estimates were calculated as the per-pixel standard deviation (see Beck et al. (2013b) for details). Note that in

this study the 200 evaluation catchments were excluded when deriving the global maps.

The newly produced maps were compared to estimates of the corresponding Q characteristics derived from the Noah and PCR-GLOBWB models. In addition, the MAR maps were compared to a MAR map computed using the Budyko formula (Budyko, 1974) with long-term mean P taken from the WorldClim dataset (Hijmans et al., 2005) and long-term mean Penman-Monteith potential evaporation from the Fisher et al. (2011) dataset. For these comparisons, all maps were averaged to 1° resolution, and for each 1° latitudinal band and method the median values of the respective Q characteristics were calculated.

6.3.3 The HBV-Light model

HBV-Light (Seibert, 2005) is a simple conceptual rainfall-runoff model run here in a spatially-lumped fashion for the 200 evaluation catchments for 2003–2007. The HBV-Light model was chosen because of its parsimony and proven effectiveness under a wide range of physiographic conditions (Te Linde et al., 2008; Steele-Dunne et al., 2008; Driessen et al., 2010; Beck et al., 2013a). The model runs at a daily time step, has two groundwater stores and one unsaturated-zone store, and requires daily time series of P, potential evaporation (PET [mm d^{-1}]), and T as inputs. The stores were initialized by running the model three times consecutively without re-initializing the stores in between runs. Table 6.1 describes the model parameters and lists the calibration ranges used. The routing parameter (MAXBAS) was set at 1 day and the snow-refreezing parameters (CFR and CWH) were both set to 0.

PET was calculated using the Priestley-Taylor (P-T) equilibrium equation (Priestley and Taylor, 1972), a less input-intensive version of the Penman-Monteith equation (Allen et al., 1998) with the aerodynamic term removed and an empirical constant (α [-]) added. The P-T equation performed among the best in a comparison of six methods to calculate potential evaporation (Lu et al., 2005). The P-T equation reads:

$$\mathrm{PET} = \frac{\alpha}{\lambda}\left(R_\mathrm{n} - G\right)\frac{\Delta}{\Delta+\gamma}, \tag{6.2}$$

where G is the soil heat flux [MJ m^{-2} d^{-1}], Δ is the slope of the saturation vapor pressure-temperature curve [kPa $°C^{-1}$], and λ is the latent heat of vaporization [MJ kg^{-1}]. Time series of G were computed as $0.05 \times R_\mathrm{n}$, time series of λ and Δ were computed from T following Allen et al. (1998), and α was set at 1.26.

6.3.4 HBV-Light model configuration 1

Three different HBV-Light model configurations were used. In the first, the HBV-Light model was run in a lumped fashion for 2003–2007 for the 200 evaluation catchments using 2000 candidate parameter sets. The latter were identified by Latin hypercube sampling (LHS; McKay et al., 1979) of uniform *a priori* distributions. LHS is an improved variant (Yu et al., 2001) of the commonly used Monte Carlo technique (Metropolis, 1987; Beven, 1993; Seibert, 1999) that splits up the parameter space in 2000 equal intervals from which values for the parameters were generated by randomly sampling each interval just once. Table 6.1 gives the sampled range for each parameter. The median simulated Q (referred to hereafter as HBV-1) was calculated, reflecting the performance of the HBV-Light model without calibration.

6.3.5 HBV-Light model configuration 2

In the second HBV-Light model configuration the model parameters were calibrated based on values of the respective Q characteristics derived from the newly produced global maps (see section 6.3.2). Specifically, the following steps were carried out:

1. The HBV-Light model was run in a lumped fashion for 2003–2007 for the 200 evaluation catchments using 2000 candidate parameter sets identified by LHS.
2. The five Q characteristics were computed for each of the 2000 Q simulations. Additionally, catchment-mean values of the five Q characteristics were derived from the global maps.
3. All Q characteristics were transformed following Eq. 6.1 to improve the normality and standardized to allow intercomparison following:

$$Z_i = \frac{X_{\mathrm{trans}\ i} - \overline{X_i}}{s_i}, \tag{6.3}$$

where Z_i and $X_{\mathrm{trans}\ i}$ are, respectively, the standardized and transformed values of Q characteristic i [-], $\overline{X_i}$ is the mean of the transformed global map of Q characteristic i (4.65, 0.22, 1.09, 0.46, and 2.40 for the respective Q characteristics), and s_i is the standard deviation of the transformed global map of Q characteristic i (1.67, 0.24, 1.32, 0.20, 0.78 for the respective Q characteristics).

4. Standardized upper and lower uncertainty bounds (w_i [-]) were calculated as:

$$w_i = \frac{e_i}{2s_i}, \tag{6.4}$$

where e_i is the uncertainty estimate associated with the global map of Q characteristic i (see section 6.3.2).

5. For each of the 2000 Q simulations the following aggregate objective function was evaluated:

$$L_n = \sum_{i=1}^{5} \max\left(\left|Z_i^\mathrm{mod} - Z_i^\mathrm{ref}\right| - w_i, 0\right), \tag{6.5}$$

where L_n is the aggregate objective function associated with simulation n [-], $i = 1, ..., 5$ are the Q

Table 6.1: HBV-Light model parameter units, descriptions, and calibration ranges.

Parameter	Units	Description	Minimum	Maximum
TT	°C	Threshold temperature when precipitation is simulated as snowfall	−2.5	2.5
CFMAX	mm °C^{-1} d^{-1}	Melt rate of the snowpack	0.5	5
SFCF	-	Snowfall correction factor to account for undercatch	1	1.5
BETA	-	Shape coefficient of recharge function	0.1	6
FC	mm	Maximum water storage in the unsaturated-zone store	50	700
K0	d^{-1}	Additional recession coefficient of upper groundwater store	0.05	0.99
K1	d^{-1}	Recession coefficient of upper groundwater store	0.01	0.8
K2	d^{-1}	Recession coefficient of lower groundwater store	0.001	0.15
LP	-	Soil moisture value above which actual evaporation reaches PET	0.05	1
PERC	mm d^{-1}	Maximum percolation to lower zone	0	5
UZL	mm	Threshold parameter for extra outflow from upper zone	0	100

characteristics [-], and the mod and ref superscripts denote the standardized and transformed values of Q characteristic i as derived, respectively, from the model and the global maps [-]. It follows from Eq. 6.5 that the score for an individual objective function is 0 when the difference between the value of the simulated Q characteristic and the value derived from the global map is $< w_i$, whereas it increases when the difference between the value of the simulated Q characteristic and the value derived from the global map is $> w_i$. It is noted that the utopian solution (i.e., the simulation n yielding $L \approx 0$) will not always be attainable as the individual objective functions may conflict with each other.

6. The 50 Q simulations with the lowest values of L were selected and the median Q was computed. The resulting Q simulation is referred to hereafter as HBV-2. The performance difference between HBV-1 and HBV-2 can be attributed to the use of the global maps of the respective Q characteristics.

Note that the calibration framework presented here can also be applied to other hydrological models and ensembles of hydrological models.

6.3.6 HBV-Light model configuration 3

In the third HBV-Light model configuration the model parameters were calibrated using values of the Q characteristics derived from observed Q data. The same steps as in HBV-Light model configuration 2 were carried out, except that the Z^{ref} values were derived from observed Q data for 2003–2007 for each catchment. Additionally, the upper and lower uncertainty bounds w_i were set to 0 for all Q characteristics. The resulting Q simulation, referred to hereafter as HBV-3, represents the performance obtained by calibrating the HBV-Light model in the gauged situation. The performance difference between HBV-2 and HBV-3 can be attributed to inaccuracies in the global maps.

6.3.7 Assessment of the HBV-Light model configurations

The effectiveness of the HBV-Light model calibration procedure using the newly derived global maps of the various Q characteristics was assessed by comparing estimated (from the global NN maps, HBV-1, HBV-2, HBV-3, Noah, and PCR-GLOBWB) and observed values of the Q characteristics for 2003–2007. All values were transformed using Eq. 6.1 and standardized using Eq. 6.3. The correlation coefficient (R [-]) and the mean difference (D [-]) between estimated and observed values of the Q characteristics were calculated. Significance levels (or p values) were not calculated as these may be misleading (Anderson et al., 2000; Nicholls, 2001).

In addition, for each catchment five commonly used objective functions were evaluated using daily simulated (from HBV-1, HBV-2, HBV-3, Noah, and PCR-GLOBWB) and observed Q to assess the 'real-world' performance of the different methods. The first is the Nash-Sutcliffe efficiency (Nash and Sutcliffe, 1970; NS [-]):

$$\text{NS} = 1 - \frac{\sum_{t=1}^{g}(Q_{\text{o}}^t - Q_{\text{s}}^t)^2}{\sum_{t=1}^{g}(Q_{\text{o}}^t - \overline{Q_{\text{o}}})^2}, \tag{6.6}$$

where Q_{s} and Q_{o} are 3-day mean simulated and observed Q, respectively [mm d^{-1}], t is the time step [-], while the summation is over $t = 1, 2, \ldots, g$. A 3-day averaging period was used to account for the flashy nature of many streams, which could have resulted in mismatches between daily peaks of observed and simulated Q. The second objective function ($\text{NS}_{\log}$ [-]) is the NS efficiency computed from $\log_{\text{e}}$-transformed 3-day Q_{s} and Q_{o} to give more weight to low Q values. The third is the NS efficiency computed from monthly mean observed and simulated Q. The fourth is the long-term bias (B [%]):

$$B = 100\frac{\overline{Q_{\text{s}}} - \overline{Q_{\text{o}}}}{\overline{Q_{\text{o}}}}. \tag{6.7}$$

The fifth and final is the absolute long-term bias ($|B|$ [%]).

6.4 Results

6.4.1 Global maps of streamflow characteristics

Table 6.2 presents mean R^2 and RMSE values obtained for the catchment-scale prediction of the five (transformed) Q characteristics (MAR_{trans}, $R1_{trans}$, $R99_{trans}$, BFI_{trans}, and k_{trans}). Mean training R^2 values of 0.86, 0.76, 0.77, 0.73, and 0.62 were obtained for the respective Q characteristics (Table 6.2). Fig. 6.2 shows scatterplots of median estimated vs. observed values of the (transformed) Q characteristics, including the corresponding linear regression lines. The median estimated values represent the median estimates of the ten NN models (one for each cross-validation iteration; see section 6.3.2). The scatterplots use transformed Q characteristics to avoid non-robustness issues in the regression analysis. The associated R^2 values are 0.89, 0.82, 0.83, 0.81, and 0.71 for the respective Q characteristics (Fig. 6.2), and are thus somewhat higher than the mean training R^2 values (Table 6.2).

Global maps of the Q characteristics were derived from global physiographic data using the NN models. Fig. 6.3 shows for each 1° latitudinal band median estimates of the five Q characteristics as derived from the NN models and the two macro-scale hydrological models (Noah and PCR-GLOBWB). Also shown are MAR values estimated using Budyko's formula (panel (a) only). Fig. 6.4 presents global maps of MAR as derived from the NN models, Noah, and PCR-GLOBWB. The agreement in terms of MAR among the methods is highest between ∼20°S and ∼10°N (i.e., the tropics; Fig. 6.3a and 6.4). Noah generally produces slightly lower and PCR-GLOBWB slightly higher MAR relative to the global NN map (Fig. 6.3a and 6.4). The global NN map produces somewhat higher MAR than the other methods between 15°N and 30°N (which includes the Sahara; Fig. 6.3a and 6.4). PCR-GLOBWB and the global NN map exhibit similar latitudinal patterns for R1, whereas Noah shows relatively little latitudinal variability in R1 (Fig. 6.3b). There is general agreement among the methods in terms of R99 latitudinal patterns (Fig. 6.3c). Noah agrees well with the global NN map in terms of BFI south of ∼45°N, but produces markedly lower BFI relative to the global NN map north ∼45°N (i.e., in snow-dominated regions; Fig. 6.3d). PCR-GLOBWB shows relatively little latitudinal BFI variability and generally produces somewhat lower BFI than the global NN map (Fig. 6.3d). Relative to the global NN map Noah produces lower k-values north of 60°N, whereas PCR-GLOBWB indicates higher k-values across the entire latitudinal range (Fig. 6.3e).

6.4.2 Assessment of the HBV-Light model configurations

Fig. 6.5 presents scatterplots of estimated (from the global NN maps, HBV-1, HBV-2, HBV-3, Noah, and PCR-GLOBWB) against observed values of the various Q characteristics for the 200 (independent) evaluation catchments, including the best-fit regression lines. The correlation coefficients (R) and mean differences (D) associated with the scatterplots are listed in Table 6.3. Among the six methods, HBV-3 (which used the same observed Q for both calibration and validation) performed best with $R \geq 0.69$ and $|D| \leq 0.30$, followed closely by the global NN maps with $R \geq 0.65$ and $|D| \leq 0.36$ (Table 6.3). HBV-2 (which used the global NN maps for calibration) performed well, with $R \geq 0.36$ and $|D| \leq 0.43$. HBV-1 (which was uncalibrated) showed fair performance, yielding $R \geq -0.15$ and $|D| \leq 0.64$. PCR-GLOBWB also showed fair performance, with $R \geq 0.06$ and $|D| \leq 2.19$, whereas Noah performed poorly with $R \geq 0.02$ and $|B| < 1.23$ (Table 6.3). MAR estimated using Budyko's formula against observed MAR showed an R of 0.79 and a D of -0.02 (corresponding scatterplot not shown).

Since the data were standardized the D values in Table 6.3 can be compared directly between the five Q characteristics. Absolute D values higher than 0.5 reflect a bias of more than half a standard deviation in the estimated relative to the observed values, indicating major biases in the estimates. These were found for Noah for MAR, R1, R99, and BFI with D values of -0.50, -1.23, -0.59, and -0.82 respectively, and for PCR-GLOBWB for BFI and k, with D values of -1.29, and 2.19, respectively. Among the five Q characteristics, MAR and R99 generally exhibited the highest R, whereas R1 and k generally exhibited the lowest R (Table 6.3). This result was rather consistent among the six methods.

Table 6.4 shows the median objective function scores evaluated between simulated and observed Q for each catchment. HBV-2 (which used the global NN maps for calibration) outperformed HBV-1 (which was uncalibrated) for all but one of the objective functions (Table 6.4). Relative to HBV-1, HBV-2 reduced the absolute bias ($|B|$) from 37.60 % to 26.65 % and increased the monthly Nash-Sutcliffe efficiency (NS) from 0.20 to 0.23. HBV-3 (which used the same observed Q for both calibration and validation) performed best among the various methods, obtaining a $|B|$ value of 19.17 % and a monthly NS value of 0.35. PCR-GLOBWB performed slightly less than HBV-1, yielding a $|B|$ value of 51.69 % and a monthly NS value of -0.10, followed by Noah with a $|B|$ value of 66.05 % and a monthly NS value of -0.24.

Table 6.2: Mean R^2 and RMSE values obtained for the catchment-scale prediction of the transformed Q characteristics.

Q characteristic	Mean R^2		Mean RMSE [-]	
	Training	Testing	Training	Testing
MAR_{trans}	0.86	0.82	0.48	0.54
$R1_{trans}$	0.76	0.67	0.12	0.15
$R99_{trans}$	0.77	0.70	0.52	0.61
BFI_{trans}	0.73	0.65	0.12	0.13
k_{trans}	0.62	0.52	0.36	0.41

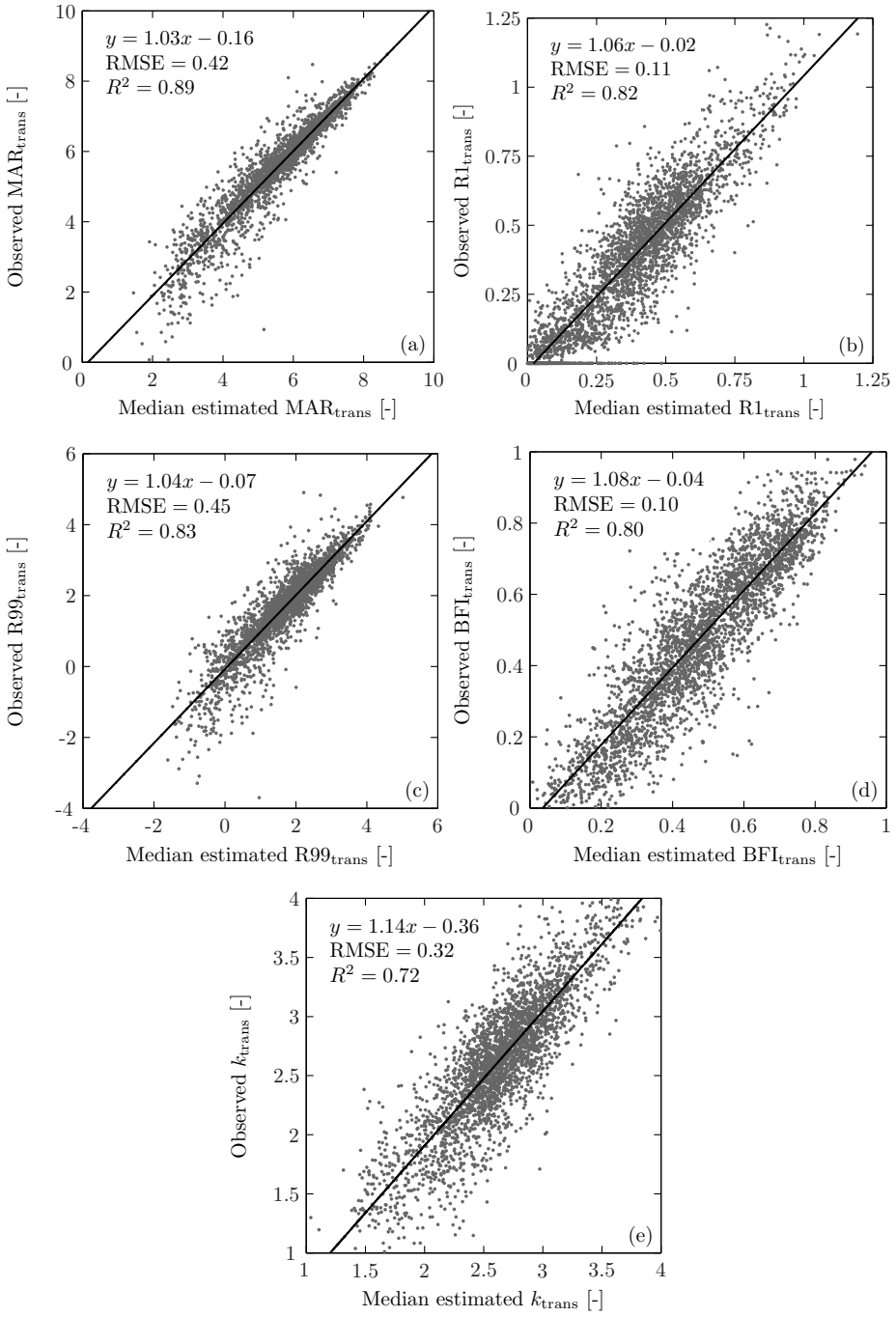


Figure 6.2: Scatterplots of estimated vs. observed values of (a) MAR_{trans}, (b) $R1_{trans}$, (c) $R99_{trans}$, (d) BFI_{trans}, (e) k_{trans} for the analyzed catchments, including the linear regression line. Each data point represents a catchment ($n = 3320$). The estimated values are the median estimates of the ten NN models used (one for each cross-validation iteration). For each iteration the training, validation, and testing subsets of catchments were included.

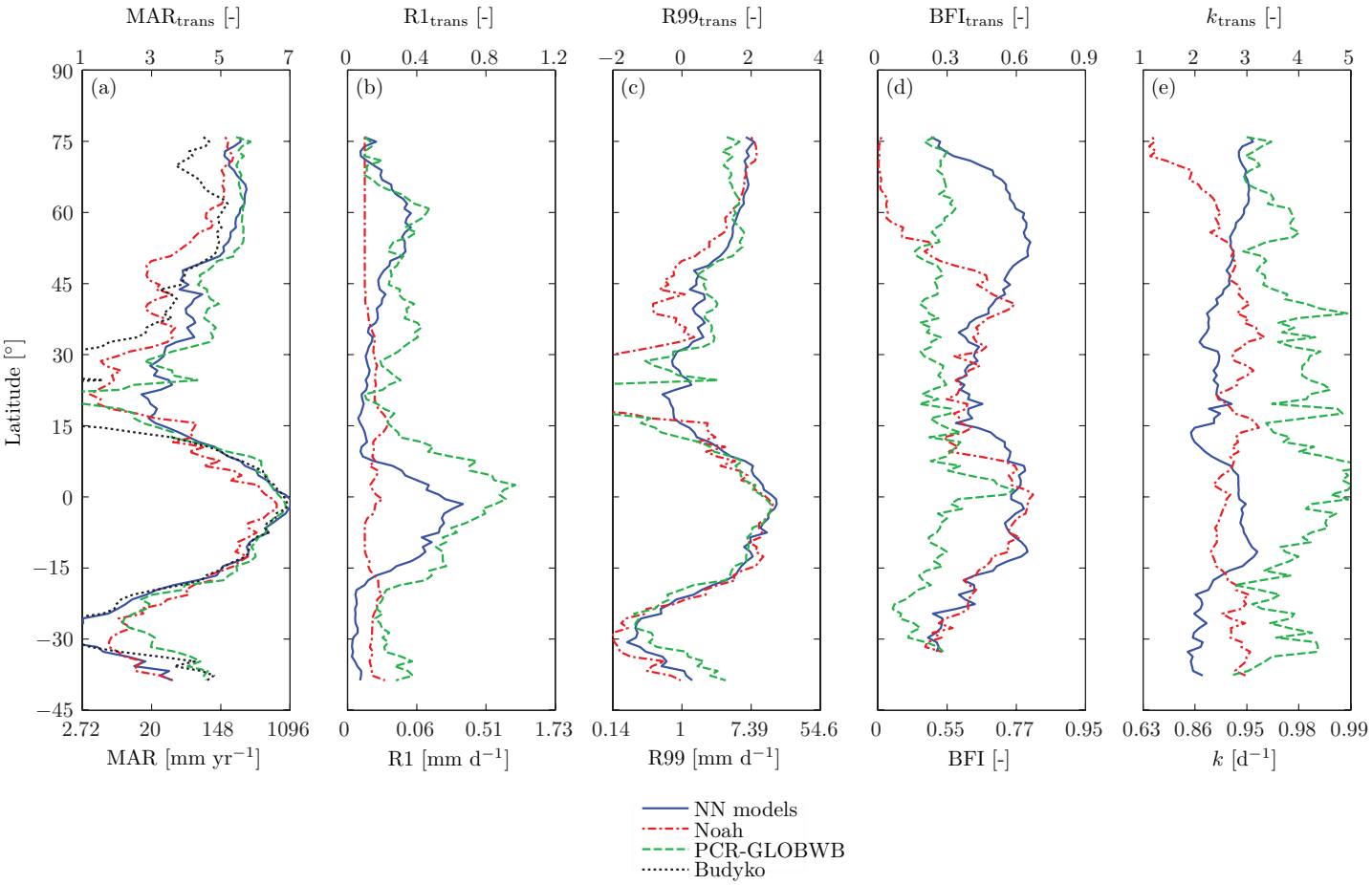


Figure 6.3: For each 1° latitudinal band the median (a) MAR, (b) R1, (c) R99, (d) BFI, and (e) k estimates as derived from the global NN maps, the Noah model, and the PCR-GLOBWB model. Also shown in (a) are MAR estimates derived using Budyko's formula. Values are plotted only if there are > 15 pixels of 1° resolution with a value within the latitudinal band. Noah and PCR-GLOBWB had several 1° resolution pixels without runoff which prevented the computation of BFI and k. Therefore, these pixels were excluded for all three methods (the global NN maps, Noah, and PCR-GLOBWB) from the median BFI and k calculation in (d) and (e). Consequently, the patterns of BFI and k differ from those shown in Fig. 9 of Beck et al. (2013b)

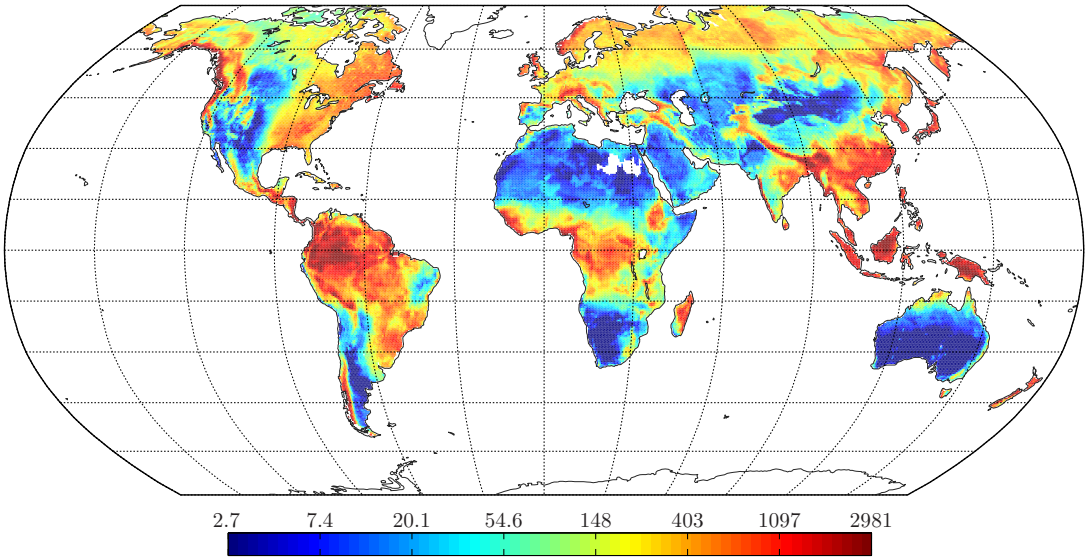
Table 6.3: Correlation coefficient (R) and mean difference (D) as calculated between estimated and observed Q characteristics. The estimated and observed Q characteristics were transformed using Eq. 6.1 and standardized using Eq. 6.3 prior to calculation of the statistics. Accordingly, the D values are intercomparable. Each statistic is based on values for the 200 evaluation catchments.

Q characteristic	NN models		HBV-1		HBV-2		HBV-3		Noah		PCR-GLOBWB	
	R	D	R	D	R	D	R	D	R	D	R	D
MAR	0.91	0.06	0.84	−0.05	0.91	0.06	0.96	−0.00	0.74	−0.50	0.64	−0.07
R1	0.75	−0.22	0.51	−0.64	0.69	−0.43	0.83	−0.30	0.25	−1.23	0.29	−0.11
R99	0.87	0.29	0.80	−0.09	0.86	0.11	0.93	−0.05	0.67	−0.59	0.45	−0.12
BFI	0.78	−0.36	0.34	0.48	0.74	−0.00	0.95	0.27	0.02	−0.82	0.06	−1.29
k	0.65	−0.28	−0.15	−0.10	0.36	−0.12	0.69	0.16	0.14	0.13	0.08	2.19

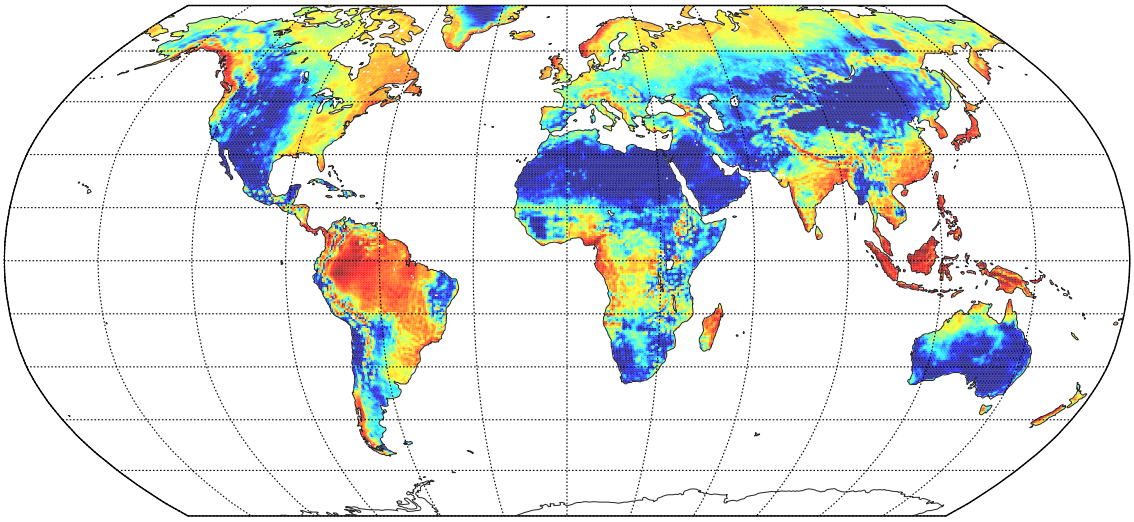
Table 6.4: Medians of objective functions evaluated between simulated and observed Q for each catchment and method. The global NN maps are not shown here since they represent long-term means of the Q characteristics and not continuous Q time series.

Objective function	HBV-1	HBV-2	HBV-3	Noah	PCR-GLOBWB
3-day NS	0.19	0.13	0.29	−0.13	−0.23
3-day NS$_{log}$	−0.44	0.08	0.25	−2.31	−1.03
Monthly NS	0.20	0.23	0.35	−0.24	−0.10
B [%]	−26.50	−2.22	−7.01	−60.42	−23.63
$\|B\|$ [%]	37.60	26.65	19.17	66.05	51.69

(a) NN models MAR [mm yr^{-1}]



(b) Noah MAR [mm yr^{-1}]



(c) PCR-GLOBWB MAR [mm yr^{-1}]

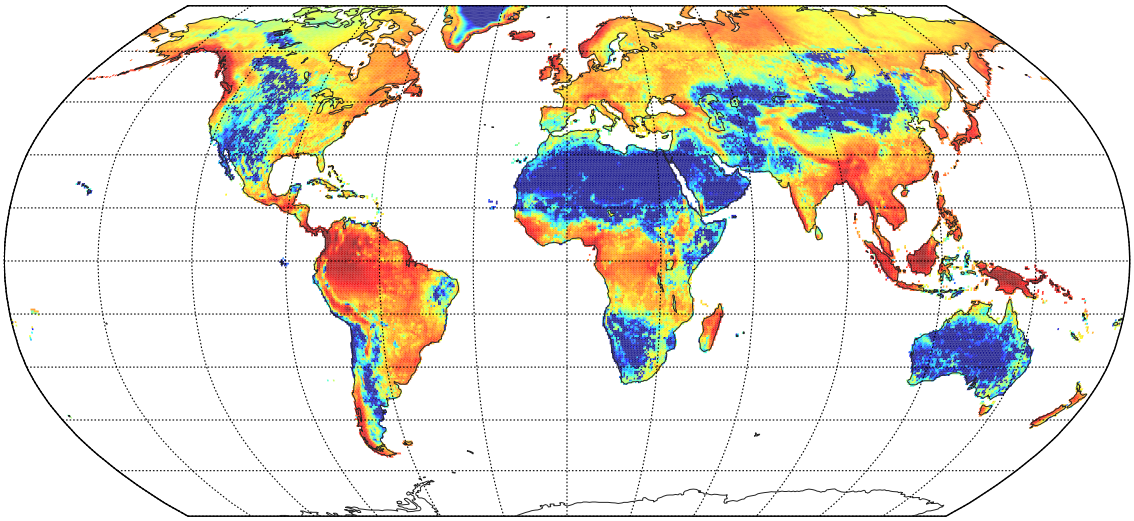


Figure 6.4: MAR as derived from (a) the global NN maps, (b) the Noah model, and (c) the PCR-GLOBWB model.

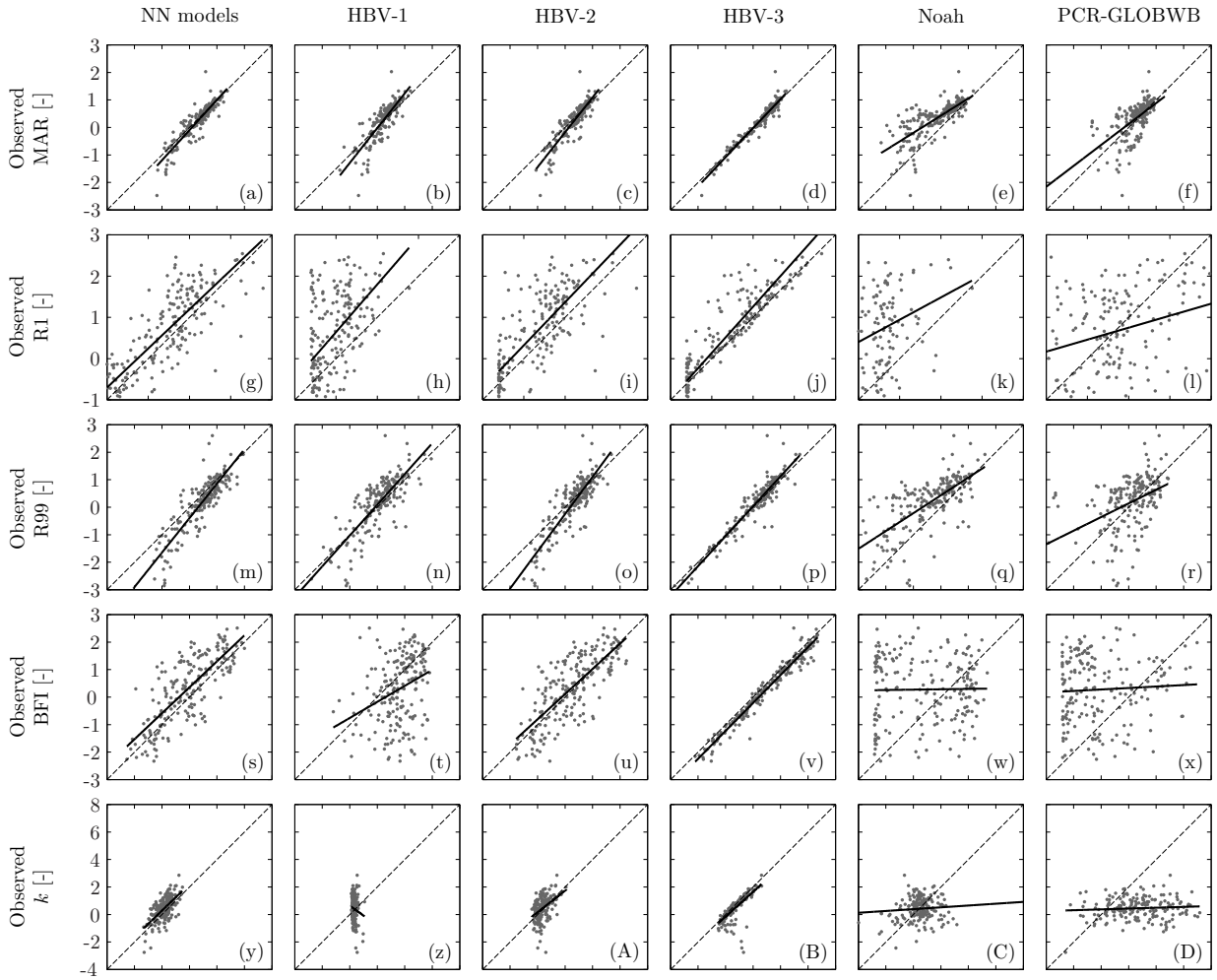


Figure 6.5: Scatterplots of estimated (from the global NN maps, HBV-1, HBV-2, HBV-3, Noah, and PCR-GLOBWB) against observed Q characteristics. Estimated values are plotted along the x-axis and observed values along the y-axis. In each scatterplot the x-axis and y-axis have the same range. The estimated and observed Q characteristics were transformed using Eq. 6.1 and standardized using Eq. 6.3. Each data point represents a catchment ($n = 200$). The dashed line is the 1:1 line and the solid line is the best-fit regression line. Statistics associated with the best-fit regression line are listed in Table 6.3.

6.5 Discussion

6.5.1 Global maps of streamflow characteristics

The neural-network (NN) models were able to successfully estimate the five Q characteristics from catchment physiographic data. The performance of $\mathrm{MAR_{trans}}$ was best with a mean training R^2 of 0.86 and a mean testing R^2 of 0.82 (Table 6.2). The mean testing R^2 value greatly exceeds the R^2 range of 0.55–0.70 obtained in the validation of (non-transformed) MAR from 14 macroscale land-surface models and six Budyko-type models against observed Q for 150 large catchments (>10 000 km^2) around the globe by Zhou et al. (2012). To our knowledge, no studies have estimated R1 and R99 from catchment physiographic data. The obtained mean training R^2 of 0.73 obtained for $\mathrm{BFI_{trans}}$ is in the upper range of training R^2 values obtained by various regional studies that estimated (non-transformed) BFI from physiographic data (e.g., Boorman et al., 1995; Lacey and Grayson, 1998; Neff et al., 2005; Santhi et al., 2008; Van Dijk, 2010; Ahiablame et al., 2013). Although the mean training R^2 of 0.62 obtained for k_{trans} was the lowest among the Q characteristics examined (Table 6.2), it is still in the upper range of training R^2 values obtained by other more regional studies (Post and Jakeman, 1996; Brandes et al., 2005; Demuth and Hagemann, 1994; Krakauer and Temimi, 2011; Peña-Arancibia et al., 2010; Van Dijk, 2010).

Global maps of the five Q characteristics were produced using the established NN models from global physiographic data. MAR is arguably the most important Q characteristic, and yet there are substantial differences between the respective methods (NN models, Noah, and PCR-GLOBWB; Figs. 6.3a and 6.4). The Noah-based results exhibited little variability in terms of R1 over the entire latitudinal range (Fig. 6.3b), possibly because no topography-related information is used in the model's parameterization. This essentially suggests Noah is unable to simulate flows in mountainous and humid-tropical environments satisfactorily. Noah fur-

ther appears to produce excessive quickflow in the Arctic domain (Fig. 6.3d), which is in agreement with the findings of Slater et al. (2007), who used observed Q of four large Arctic catchments (>1 680 000 km^2). The overestimation of quickflow under Arctic conditions is likely due to excessive reduction of the soil infiltration capacity under freezing conditions in the version of Noah used here (cf. Niu et al., 2011). By contrast, PCR-GLOBWB predicted volumes of quickflow that were too high across the entire latitudinal range relative to the NN models and Noah (Fig. 6.3d), suggesting that a re-evaluation of the model's runoff subroutine might be in order. The poor agreement for k between the three methods (Fig. 6.3e) and the relatively poor performance of the NN model for k (Table 6.2 and Fig. 6.2e) suggests the estimation of k to be the most challenging among the five Q characteristics examined here.

6.5.2 Assessment of the HBV-Light model calibration configurations

The global NN maps of the five Q characteristics discussed in the previous sub-section have the potential to globally improve the parameterization of hydrological models. This was tested using a simple conceptual rainfall-runoff model (HBV-Light) calibrated for 200 (independent) evaluation catchments using values of the Q characteristics derived from the global NN maps (HBV-2). HBV-2 showed substantially better performance than the HBV-Light model without calibration (HBV-1) for all Q characteristics (Fig. 6.5 and Table 6.3), demonstrating the effectiveness of the calibration procedure. The improved representation of the Q characteristics in HBV-2 led, in turn, to improvements in the medians of four traditional Q performance measures (3-day NS_{log}, monthly NS, B, and $|B|$) but also to a deterioration in the median of one measure (3-day NS; Table 6.4). The deterioration in the median 3-day NS is due to the high sensitivity of the NS efficiency value to peakflows (cf. Eq. 6.6). Although HBV-1 simulated lower peakflows compared to HBV-2, HBV-2 generally yielded a larger value of the numerator in Eq. 6.6 resulting in a lower NS because of mismatches between observed and simulated peakflows, mostly due to the relatively poor quality of the P data. HBV-3 exhibited superior performance compared to all other models (HBV-1, HBV-2, Noah, and PCR-GLOBWB) because it was calibrated using Q characteristics derived from the same observed Q data also employed for the validation. HBV-3 obtained $R \geq 0.69$ and $|B| \leq 0.27$ between simulated and observed values of the Q characteristics (Table 6.3), suggesting that the HBV-Light model is capable of capturing a wide range of Q characteristics by exploring the *a priori* parameter space. This further suggests that the aggregate objective function (Eq. 6.5) is capable of identifying the 'optimal' solution for most catchments.

Other studies demonstrating that Q characteristics estimated from catchment physiographic characteristics can be used to calibrate hydrological models had a regional focus (Yadav et al., 2007; Zhang et al., 2008a; Castiglioni et al., 2010; Lombardi et al., 2012; Pinheiro and Naghettini, 2012). These studies used ≤ 30 catchments and multi-variate linear regression to estimate the Q characteristics from physiographic data. However, larger catchment sets reveal that the relationships are often decidedly non-linear (cf. Van Dijk, 2010; Peña-Arancibia et al., 2010; Beck et al., 2013b), necessitating the use of non-linear models such as neural networks. In addition, these regional studies did not transform the Q characteristics to make their distributions approximate normality, which may confound the estimation of Q characteristics from physiographic data and the calibration exercise (cf. Parada and Liang, 2010). However, several of the studies (Yadav et al., 2007; Zhang et al., 2008a; Castiglioni et al., 2010) provide Q uncertainty ranges, which can be calculated here from the 50 selected Q simulations (see section 6.3.5 step six).

In light of the slightly lower performance of the two (uncalibrated) macro-scale models examined (Noah and PCR-GLOBWB) in terms of both Q characteristics (Fig. 6.5 and Table 6.3) and traditional performance measures calculated from observed and simulated daily continuous streamflow time series (Table 6.4), one should bear in mind that these models were not specifically developed to provide continuous Q time series for such small catchments at a daily time step. Further, they may perform considerably better at other Q characteristics or for catchments not examined here. It must also be emphasized that other macro-scale models are likely to exhibit similar performance. Nevertheless, our results indicate that Noah generally underestimated MAR, R1, and R99 (i.e., produced too little flow overall; see Figs. 6.5e, 6.5k, and 6.5q, respectively, and Table 6.3), which has been confirmed by Zaitchik et al. (2010) using observed Q data for 66 large catchments (> 19 000 km^2) around the globe. PCR-GLOBWB generally underestimated BFI (i.e., produced excess quickflow; Fig. 6.5x and Table 6.3) and overestimated k (i.e., the baseflow recedes too slowly; Fig. 6.5D and Table 6.3). PCR-GLOBWB further produced relatively unbiased MAR (Table 6.3), although the scatter was large (Fig. 6.5f). However, this scatter is likely to average out in larger catchments, as shown by Van Beek and Bierkens (2009) using observed Q data for 19 large catchments (> 65 000 km^2) worldwide. Since the global NN maps appear to outperform both models (Fig. 6.5 and Table 6.3) it is conceivable that the global NN maps can be employed to refine the parameterization and/or structure of Noah and PCR-GLOBWB. The global NN maps should further prove useful for macro-scale hydrological models having spatial resolutions down to 1 km that are anticipated in the near future (Wood et al., 2011), since the global NN maps have a resolution limited only by the resolution of the input data used.

Further research should focus on expanding the observed Q dataset in data-poor regions (i.e., outside the

USA, Europe, and Australia), on identifying disinformative data in the observed Q dataset (cf. Kauffeldt et al., 2013), and on refining and/or expanding the selection of Q characteristics. For example, the addition of a Q characteristic related to the seasonal timing of flows could improve our ability to estimate the arrival of spring snowmelt in mountainous or sub-Arctic catchments. It should be noted that to warrant the inclusion of a particular Q characteristic in the calibration of a hydrological model the estimates from the NN model need to be better than the initial estimates from the hydrological model under consideration. Furthermore, the global NN map of MAR may be corrected for bias using MAR as estimated for interstation regions from observed Q data of very large catchments (cf. Fekete et al., 1999, 2002).

6.6 Conclusion

The present study is the first to present an approach to globally improve the parameterization of hydrological models using Q characteristics. The main findings are as follows:

1. Global maps of five Q characteristics (MAR, R1, R99, BFI, and k) were successfully produced using neural-network (NN) models based on observed Q and physiographic data for 3320 catchments worldwide (Table 6.2 and Fig. 6.2). Global patterns of mean annual runoff (MAR) as derived from the NN models, the Noah model, and the PCR-GLOBWB model demonstrated relatively low agreement (Fig. 6.3 and 6.4a). It was further found that Noah showed little latitudinal variability in terms of the 1st percentile runoff (R1; Fig. 6.3b) and produced more quickflow in northern latitudes relative to the NN models (Fig. 6.3d), whereas PCR-GLOBWB produced more quickflow over most of the land surface relative to the NN models (Fig. 6.3d).
2. A simple conceptual rainfall-runoff model (HBV-Light) was calibrated for 200 independent catchments using values of the five examined Q characteristics as derived from the newly produced maps (HBV-2). The simulated Q characteristics improved considerably relative to the uncalibrated HBV-Light model (HBV-1; Fig. 6.5 and Table 6.3). This, in turn, led to improvements in most of the traditional performance measures, particularly in those related to the long-term water balance (Table 6.4). Noah and PCR-GLOBWB performed somewhat less than the global NN maps in terms of specific Q characteristics (Fig. 6.5 and Table 6.3), suggesting that the global NN maps can be used to improve the parameterization and/or structure of these models.

Acknowledgments

The first author was supported by Deltares (Delft, The Netherlands). The Global Runoff Data Centre (GRDC; Koblenz, Germany) is thanked for providing observed Q and catchment boundary data. We are also most grateful to Frederiek Sperna Weiland (Deltares) for providing the PCR-GLOBWB Q data and the NASA DISC and GLDAS teams for providing the Noah Q data.

Chapter 7

Summary

7.1 Conclusions

The main findings of the thesis can be summarized as follows (the respective items listed below correspond to the specific objectives defined in the introductory chapter):

1. In Chapter 2 different antecedent soil moisture proxies representing catchment wetness status were tested for their ability to improve Curve Number-based peakflow predictions for 186 unregulated Australian catchments. The examined antecedent soil moisture proxies were based on gauged precipitation, baseflow (obtained by separating the observed streamflow record), TRMM precipitation, and AMSR-E surface soil moisture. It was found that the 5-day antecedent precipitation index computed from gauge precipitation, the generally recommended soil moisture proxy for use with the Curve Number model, performed considerably worse than an antecedent precipitation index based on gauged precipitation using an optimized decay factor. It was further found that the latter performed better than the soil moisture proxies based on AMSR-E surface soil moisture and TRMM precipitation, demonstrating the continued importance of a sufficiently dense precipitation-gauge network. The greatest improvements in model runoff prediction performance were typically obtained for arid catchments, possibly reflecting the larger variability in catchment wetness levels occurring under such conditions compared to more humid climates.
2. In Chapter 3 the degree of agreement between four widely used AVHRR-NDVI datasets (PAL, GIMMS, LTDR V3, and FASIR) for assessing vegetation productivity was examined. In addition, these datasets and the NDVI based on the more modern MODIS instrument were globally validated against 11 764 high-resolution (∼30 m) NDVI samples (20×20 km^2) derived from the Landsat-5 Thematic Mapper, which has on-board calibration devices. The trends in mean NDVI were positive for all continents and all four AVHRR-NDVI datasets, with trends derived from the PAL dataset generally being the strongest and the ones from the GIMMS dataset generally the weakest. Significantly equal trends between the datasets were found over 48 % of the total land surface only, which is a low proportion considering that each dataset was based on the same AVHRR Global Area Coverage archive. The LTDR V3 and PAL showed trends in desert areas devoid of vegetation and were thus improperly calibrated. The validation against Landsat-derived NDVI values indicated that the LTDR V3 dataset is the most accurate in terms of absolute NDVI values, whereas the GIMMS dataset is the most accurate in terms of NDVI change over time. However, NDVI derived from the MODIS instrument strongly outperformed all AVHRR-NDVI datasets examined. As such, it is considered unlikely that these AVHRR-NDVI datasets presently can be used to obtain accurate global information on land cover or land use change, as the data are subject to a high degree of uncertainty. In addition, they have a coarse spatial resolution (0.05° for the LTDR V3 dataset and 0.08° for the other AVHRR-NDVI datasets) and employ the maximum-value compositing technique, which favourably selects patches with high NDVI values within a pixel. Finally, it was found that taking the simple average of the four AVHRR-based datasets resulted in NDVI values which compared better to Landsat-NDVI than all of the AVHRR-NDVI datasets individually, indicating that the errors in respective datasets are to a certain degree unrelated.
3. In Chapter 4 the possible impacts of forest regrowth on streamflow characteristics were examined for 12 meso-scale (23–346 km^2) catchments on the humid tropical island of Puerto Rico. Long-term records of precipitation and potential evaporation were used to drive a simple conceptual rainfall-runoff model (HBV-Light), producing simulated streamflow that integrated the effects of carry-over water storage between successive years as well as climate variability during the streamflow observation period. For each catchment, simulated and observed time series of four streamflow characteristics were calculated: (1) the annual 95th percentile (i.e., the percentage of time that this level of flow is not exceeded) of daily streamflow (indicative of peak flows); (2) the annual mean streamflow (indicative of total water

yield); (3) the annual 5th percentile daily streamflow (indicative of low flows); and (4) the annual mean dry-season (January–March) streamflow. In addition, trends in the deviations between simulated and observed values for the respective streamflow metrics were evaluated. These represented the change in observed streamflow characteristics after taking the effects of water storage carry-over and climate variability into account. However, no clear relationships were found between changes in streamflow characteristics and changes in either forested or urban area per catchment. These findings are in line with previous studies of meso- and macro-scale (sub-)tropical catchments, which generally found no significant change in streamflow that could be attributed to changes in forest area. Possible explanations for the presently found lack of a clear relationship include: (1) data errors (notably for precipitation); (2) the changes in forest area occur mainly in the less rainy lowlands (thereby having a less pronounced effect on overall streamflow relative to the same change in forest cover if effected over the wetter upland parts of the catchments); and (3) heterogeneity among the catchments in terms of the streamflow response to forest regrowth (due to differences in vegetation cover and/or land-use history not being accounted for by the semi-quantitative classification used, or because of differences in morphology, geology, and/or soils between the catchments).

4. In Chapter 5 relationships between 18 catchment physiographic characteristics (related to soils, topography, climate, and land cover) and two important baseflow characteristics were analyzed using a diverse set of 3520 unregulated catchments worldwide. Previous studies have typically used < 200 catchments and regional datasets, resulting in less reliable relationships with potentially limited applicability elsewhere. The used baseflow characteristics were: (1) the baseflow index (BFI), defined as the ratio of long-term mean baseflow to total streamflow; and (2) the baseflow recession constant (k), defined as the rate of baseflow decay. The two baseflow characteristics proved to be related to several physiographic characteristics, notably mean annual potential evaporation, mean catchment elevation, mean surface slope, fraction of open water, and the mean sand content of the soil. The relationships were generally highly non-linear and heteroscedastic (i.e., showing variable scatter). Artificial neural network ensembles were subsequently used to estimate the two baseflow characteristics for the catchments. Mean training R^2 values of 0.73 and 0.62 were obtained for BFI and k, respectively, suggesting that artificial neural network ensembles provide a viable alternative to the commonly used multi-variate linear regression approach. In addition, global maps of the two baseflow characteristics were produced using global physiographic data as input to the established artificial neural network ensembles. These maps offer unique opportunities for macro-scale hydrological studies, including the diagnosis and parameterization of macro-scale hydrological models (land surface schemes and global hydrological models), water resource assessments, catchment classification, and groundwater recharge estimation (see also below).
5. In Chapter 6 global maps of five important streamflow characteristics were produced using the methodology developed in Chapter 5 in combination with streamflow and physiographic data for 3320 undisturbed catchments around the globe. The streamflow characteristics were: (1) mean annual runoff (indicative of total water yield); (2) annual 1st percentile of daily streamflow (indicative of low flows); (3) annual 99th percentile of daily streamflow (indicative of peakflows); (4) BFI; and (5) k. It was found that the global patterns of the newly produced maps differed considerably from equivalent maps as derived from two commonly used macro-scale hydrological models (Noah and PCR-GLOBWB). Next, the HBV-Light rainfall-runoff model was calibrated using values of the streamflow characteristics derived from the newly produced maps for 200 independent catchments. This resulted in substantial improvements in the simulated streamflow characteristics as compared to values obtained with the uncalibrated HBV-Light model. These improvements in streamflow characteristics, in turn, led to improvements in most of the traditional model performance measures calculated from observed and simulated daily continuous streamflow time series, including the long-term bias and the Nash-Sutcliffe efficiency. These findings demonstrate that the newly produced maps of the respective streamflow characteristics examined can be employed to improve the parameterization and/or structure of hydrological models at a global scale.

7.2 Recommendations

A recurring feature of each of the chapters of this thesis has been the use of large observational datasets in order to arrive at more confident conclusions. As discussed more fully in the introductory chapter, such large observational datasets have become increasingly available owing to recent advances in remote-sensing technology, processing power, and data dissemination. Arguably, in each of the chapters, the use of a data-subset only, might well have led to very different conclusions (see, e.g., the Discussion section in Chapter 4 on streamflow response to tropical regrowth in Puerto Rico). Naturally, at the global scale, the use of these large observational datasets is a *conditio sine qua non* when trying to bet-

ter understand how the hydrological cycle will respond to global climate change and human-induced changes in land cover and land use or land degradation (cf. Jones, 2005; Andréassian et al., 2007; Peña-Arancibia, 2013; Gupta et al., 2013). In addition, using all available data (including outliers) is likely to present a more balanced picture to potential end-users (Andréassian et al., 2007).

In Chapter 2 it was found that the TRMM-based precipitation estimates and AMSR-E-based surface soil moisture data used as proxies for catchment-scale wetness status (and thus runoff response to rainfall) showed poor predictive performance compared to ground-based measured precipitation. Better results may be achieved when using the successor to TRMM, the Global Precipitation Measurement (GPM) mission (Smith et al., 2004) which is to be launched in early 2014. The GPM mission is targeted to provide precipitation data with significantly improved accuracy, spatial coverage and resolution, and at an average revisit time of less than three hours in many regions of the world. Further improvements in estimating catchment-scale wetness status may be expected using surface soil-moisture products derived from the recently launched European Space Agency (ESA) Soil Moisture Ocean Salinity (SMOS) mission (Kerr et al., 2001) or from the National Aeronautics and Space Administration (NASA) Soil Moisture Active/Passive (SMAP) mission (Entekhabi et al., 2010) which is to be launched in 2015. Arguably, the development of remotely-sensed surface soil moisture products based on observations from multiple sensors (e.g., Liu et al., 2011; Dorigo et al., 2012) is particularly promising.

The newly derived observation-based maps of various important streamflow characteristics developed in Chapters 5 and 6 will be made available free of charge for downloading, and should be useful for a variety of macro-scale hydrological applications. Amongst the potential applications is a global diagnosis of macro-scale hydrological models (land surface schemes and global hydrological models), as demonstrated in Chapter 6 where the global patterns of the newly derived maps were compared with equivalent maps derived from two macro-scale models (Noah and PCR-GLOBWB). The new maps further offer unique opportunities for the parameterization of such macro-scale hydrological models, including the type of 'hyper-resolution' models that are anticipated in the near future (Wood et al., 2011). The streamflow characteristic maps may also be used for general water resource assessments and for comparing the behavior of different catchments (i.e., catchment classification; Wagener et al., 2007). In addition, the global streamflow map offers possibilities for correcting precipitation biases in mountainous and/or humid regions due to topographic biases in gauge placement, interpolation errors, and/or wind-induced snowfall undercatch (cf. Adam et al., 2006). Furthermore, the global streamflow and BFI maps can be combined to provide a first estimate of groundwater recharge for any location around the globe (cf. Szilagyi et al., 2003).

Last but not least, the methodology adopted in Chapters 5 and 6 to produce global maps of various important streamflow characteristics proved highly efficient and flexible, and deserves to be used to derive additional global maps of other important streamflow characteristics. For example, the addition of a streamflow characteristic related to the timing of flows could be important to estimate the arrival of spring snowmelt in mountainous and/or sub-Arctic catchments. The methodology could perhaps also be employed to estimate global patterns of other key biophysical, hydrological, or climatic variables.

Chapter 8

Samenvatting

8.1 Conclusies

De belangrijkste bevindingen van dit proefschrift kunnen als volgt worden samengevat (de onderstaande onderwerpen betreffen achtereenvolgens de specifieke doelstellingen zoals gedefinieerd in het inleidende hoofdstuk):

1. In hoofdstuk 2 worden verschillende bodemvochtindicatoren getest op hun vermogen om voorspellingen van piekafvoeren gebaseerd op de zogeheten Curve Number methode voor 186 niet-gereguleerde Australische stroomgebieden te verbeteren. De onderzochte bodemvochtindicatoren waren gebaseerd op de grootte van gemeten neerslag, basisafvoer (verkregen door het scheiden van de gemeten afvoer in basis- en snelle afvoeren met behulp van een digitaal filter), TRMM-gebaseerde neerslag en AMSR-E-gebaseerd oppervlakkig bodemvochtgehalte. Het bleek dat de vijfdaagse neerslagindex berekend op basis van de gemeten neerslag, de bodemvochtindicator die standaard aanbevolen wordt voor gebruik met het Curve Number model, aanzienlijk slechter presteert dan een neerslagindex gebaseerd op de gemeten neerslag met een geoptimaliseerde vervalfactor. Bovendien werd vastgesteld dat laatstgenoemde variabele beter presteert dan de bodemvochtindicatoren gebaseerd op AMSR-E oppervlaktebodemvocht en TRMM neerslag, wat het belang van een dicht neerslagmeetnetwerk benadrukt. De grootste verbeteringen in piekafvoervoorspellingen werden over het algemeen verkregen voor stroomgebieden in relatief droge gebieden, mogelijk vanwege de grotere variabiliteit in bodemvocht onder dergelijke condities in vergelijking met vochtiger klimaten.
2. In hoofdstuk 3 wordt de mate van overeenstemming tussen vier AVHRR-NDVI datasets (PAL, GIMMS, LTDR V3 en FASIR), die veelvuldig gebruikt worden als indicator van de mate van productiviteit van de vegetatie in een gebied, onderzocht. Daarnaast werden deze vier datasets, en NDVI-waarden gebaseerd op het modernere MODIS satelliet-instrument, wereldwijd gevalideerd ten opzichte van 11 764 hoge-resolutie (∼30 m) NDVI afbeeldingen (20×20 km^2) gebaseerd op de Landsat-5 Thematic Mapper sensor, die voorzien is van kalibratieapparatuur. De trends in de jaarlijkse NDVI-waarden waren positief voor alle continenten voor alle vier de onderzochte AVHRR-NDVI datasets, waarbij de trends afgeleid van de PAL dataset over het algemeen het meest uitgesproken en die van de GIMMS dataset over het algemeen het minst uitgesproken waren. Gelijk opgaande trends voor de verschillende datasets werden gevonden voor 48 % van het totale landoppervlak, wat als laag beschouwd dient te worden gezien het feit dat elke dataset gebaseerd is op hetzelfde AVHRR Global Area Coverage archief. De LTDR V3 en PAL datasets vertoonden trends voor woestijngebieden die verstoken waren van vegetatie en deze datasets zijn derhalve niet goed gekalibreerd. De validatie ten opzichte van NDVI gebaseerd op Landsat gaf aan dat de LTDR V3 dataset het meest nauwkeurig is in termen van absolute NDVI-waarden en dat de GIMMS dataset het meest nauwkeurig is in termen van temporele variaties in NDVI. De NDVI afgeleid van het MODIS instrument presteert echter beter dan alle andere onderzochte AVHRR-NDVI datasets. Gezien de hoge onzekerheid die inherent is aan de AVHRR-NDVI datasets is het onwaarschijnlijk dat deze datasets gebruikt kunnen worden om betrouwbare en accurate wereldwijde informatie over veranderingen in landbedekking of landgebruik te verkrijgen. Daarnaast hebben de gegevens een grove ruimtelijke resolutie (0.05° voor de LTDR V3 dataset en 0.08° voor de overige AVHRR-NDVI datasets) en maken deze gebruik van de zogenoemde 'maximale waarde' compositietechniek, die bij voorkeur stukken grond met hoge NDVI-waarden binnen een pixel selecteert. Ten slotte bleek dat het nemen van het gemiddelde van de vier AVHRR datasets resulteerde in NDVI-waarden die beter overeenkwamen met de Landsat-NDVI dataset dan elke AVHRR-NDVI dataset afzonderlijk, hetgeen suggereert dat de fouten in de AVHRR-NDVI datasets in zekere mate onafhankelijk van elkaar zijn.
3. In hoofdstuk 4 worden de mogelijke effecten van boshergroei op rivierafvoerkarakteristieken onderzocht met gegevens voor 12 middelgrote (23–

346 km^2) stroomgebieden op het tropische eiland Puerto Rico. Langdurige metingen van neerslag en potentiële verdamping werden gebruikt om een eenvoudig conceptueel neerslag-afvoermodel (HBV-Light) te sturen, wat resulteerde in gesimuleerde afvoer die rekening houdt met de effecten van klimaatschommelingen alsmede wateropslag in de bodem tussen opeenvolgende jaren. Voor elk stroomgebied werden gesimuleerde en gemeten tijdreeksen van vier afvoerkarakteristieken berekend: (1) het jaarlijkse 95ste percentiel (dat wil zeggen het percentage van de tijd dat dit afvoerniveau niet overschreden wordt) van de dagelijks afvoer (indicatief voor piekafvoeren); (2) de gemiddelde jaarlijkse afvoer (indicatief voor de totale afvoer); (3) het jaarlijkse vijfde percentiel van de dagelijkse afvoer (indicatief voor de basisafvoer); en (4) de gemiddelde jaarlijkse afvoer gedurende het droge seizoen (januari–maart). Verder werden trends in de afwijking tussen gesimuleerde en gemeten waarden voor de verschillende afvoerkarakteristieken berekend, welke de verandering in de waargenomen afvoerkarakteristieken representeren die niet het gevolg zijn van de effecten van klimaatvariabiliteit of wateropslag in de bodem. Er werd echter geen duidelijke relatie gevonden tussen veranderingen in de respectievelijke afvoerkarakteristieken en veranderingen in het areaal per stroomgebied aan bos of stedelijk gebied. Dit is in overeenstemming met voorgaande studies op meso- en macroschaal van (sub)tropische stroomgebieden, die over het algemeen geen significante afvoerveranderingen vonden die konden worden toegeschreven aan veranderingen in bosareaal. Mogelijke verklaringen voor het ontbreken van een duidelijke relatie in de hier onderzochte stroomgebieden zijn: (1) fouten in de gemeten tijdreeksen, met name in de neerslagmetingen; (2) veranderingen in het bosareaal die vooral plaatsvonden in de minder regenachtige laaglanden van de onderzochte stroomgebieden, wat een minder uitgesproken effect op de afvoer heeft dan wanneer dezelfde verandering in bosareaal plaats zou vinden in de nattere hooggelegen delen van de stroomgebieden; en (3) verschillen tussen de stroomgebieden in termen van de afvoerrespons op boshergroei, als gevolg van verschillen in de gebruikte vegetatieclassificeringen op verschillende tijdstippen (1951, 1978, 1991 en 2000) en daarmee in de karakterisering van het landgebruik met de tijd), of als gevolg van verschillen in geomorfologie, geologie en/of bodems tussen de onderzochte stroomgebieden.

4. In hoofdstuk 5 zijn de relaties tussen 18 fysiografische karakteristieken (met betrekking tot bodems, topografie, klimaat en landbedekking) en twee belangrijke basisafvoer-karakteristieken geanalyseerd met behulp van een gevarieerde dataset bestaande uit 3520 niet-gereguleerde stroomgebieden wereldwijd. Eerdere studies gebruikten doorgaans < 200 stroomgebieden en veelal meer regionale fysiografische datasets, wat in het algemeen resulteerde in minder betrouwbare relaties tussen afvoerkarakteristieken en terreinvariabelen met bovendien een beperkte toepasbaarheid elders. De in de huidige studie gebruikte basisafvoer-karakteristieken waren: (1) de basisafvoerindex (BFI), gedefinieerd als de verhouding tussen de gemiddelde basisafvoer en de totale afvoer; en (2) de basisafvoer-vervalconstante (k), gedefinieerd als de vervalsnelheid (recessie) van de basisafvoer. Beide basisafvoer-karakteristieken bleken goed gerelateerd te zijn aan een aantal gebiedskarakteristieken, waaronder de gemiddelde jaarlijkse potentiële verdamping, de gemiddelde stroomgebiedshoogte, de gemiddelde helling van het oppervlak, de fractie ingenomen door oppervlaktewater (meren, moerassen, e.d.) en de gemiddelde zandfractie van de bodem. De relaties waren doorgaans sterk niet-lineair en heteroscedastisch (dat wil zeggen, met variabele spreiding). Ensembles van kunstmatige neurale netwerken werden vervolgens getraind om waarden van de twee basisafvoer-karakteristieken voor alle stroomgebieden te schatten uit combinaties van fysiografische variabelen. Gemiddelde training R^2 waarden van 0.73 en 0.62 werden verkregen voor respectievelijk BFI en k, wat suggereert dat ensembles van kunstmatige neurale netwerken een goed alternatief kunnen bieden voor de doorgaans gebruikte multivariabele lineaire regressietechnieken. Daarnaast werden wereldwijde kaarten van beide basisafvoer-karakteristieken geproduceerd door globale fysiografische data als invoer te gebruiken voor de getrainde neurale netwerken. Deze kaarten bieden unieke mogelijkheden voor macro-hydrologische studies, waaronder het diagnosticeren en parametriseren van macro-hydrologische modellen, het beheer van watervoorraden, het classificeren van stroomgebieden en het schatten van de aanvulsnelheid van het grondwater (zie tevens hieronder).
5. In hoofdstuk 6 zijn wereldwijde kaarten van vijf belangrijke afvoerkarakteristieken geproduceerd met behulp van de methodologie ontwikkeld in hoofdstuk 5 en afvoer- en fysiografische gegevens voor 3320 niet gereguleerde stroomgebieden uit de hele wereld. De bestudeerde afvoerkarakteristieken waren: (1) de gemiddelde jaarlijkse afvoer (indicatief voor de totale afvoer); (2) het jaarlijkse eerste percentiel van de dagelijks afvoer (indicatief voor de basisafvoer); (3) het jaarlijkse 99ste percentiel van de dagelijks afvoer (indicatief voor de piekafvoer); (4) de BFI; en (5) de k. Het bleek dat de wereldwijde patronen van de nieuwe kaarten aanzienlijk verschilden van gelijksoortige kaarten gebaseerd op twee veelgebruikte macro-hydrologische modellen (Noah en PCR-GLOBWB). Vervolgens werd het HBV-Light neerslag-afvoermodel gekalibreerd met behulp van waarden van de afvoerkarakteristieken

afgeleid uit de nieuwe kaarten voor 200 onafhankelijke stroomgebieden. Dit resulteerde in een aanzienlijke verbetering in de gesimuleerde afvoerkarakteristieken vergeleken met waarden gebaseerd op het ongekalibreerde HBV-Light model. Deze verbeteringen in afvoerkarakteristieken leidden vervolgens tot verbeteringen in de meeste traditionele modelprestatieindicatoren berekend op basis van dagelijkse gemeten en gesimuleerde afvoertijdreeksen, zoals de Nash-Sutcliffe efficiëntie. Deze bevindingen tonen aan dat de nieuwe kaarten van de afvoerkarakteristieken kunnen worden gebruikt om de parametrisering en/of structuur van hydrologische modellen op wereldschaal verder te verbeteren.

8.2 Aanbevelingen

Een terugkerend kenmerk van de verschillende hoofdstukken van dit proefschrift is het gebruik van grote observationele datasets om tot robuustere conclusies te komen. Zoals uitgebreider besproken in het inleidende hoofdstuk, komen zulke grote observationele datasets steeds meer beschikbaar als gevolg van recente ontwikkelingen in satelliettechnologie ('remote sensing'), computerrekenkracht en gegevensverbreiding. In elk van de hoofdstukken zou het gebruik van slechts een gedeelte van de dataset ongetwijfeld hebben geleid tot heel andere conclusies (zie bijvoorbeeld de discussiesectie van hoofdstuk 4 over het effect van boshergroei op rivierafvoeren in Puerto Rico). Op wereldschaal is het gebruik van deze grote observationele datasets een *conditio sine qua non* wanneer men probeert te begrijpen hoe de hydrologische kringloop zal reageren op de wereldwijde klimaatverandering en door de mens veroorzaakte veranderingen in landbedekking en -gebruik of landdegradatie (Jones, 2005; Andréassian et al., 2007; Peña-Arancibia, 2013; Gupta et al., 2013). Bovendien zal het gebruik van alle beschikbare gegevens (inclusief 'uitschieters') waarschijnlijk een evenwichtiger beeld presenteren aan potentiële eindgebruikers (Andréassian et al., 2007).

In hoofdstuk 2 bleek dat de bodemvochtindicatoren gebaseerd op TRMM neerslag en AMSR-E oppervlaktebodemvochtgehalte slecht presteerden in vergelijking met een bodemvochtindicator gebaseerd op gemeten neerslag. Wellicht kunnen betere resultaten verkregen worden met neerslag gebaseerd op de opvolger van TRMM, de Global Precipitation Measurement (GPM) missie (Smith et al., 2004) die begin 2014 zal worden gelanceerd. De GPM missie zal neerslaggegevens leveren met een aanzienlijk verbeterde nauwkeurigheid, ruimtelijke dekking en resolutie, en zal een gemiddelde terugkeertijd van minder dan drie uur hebben in veel regio's van de wereld. Verdere verbeteringen in het schatten van de vochtstatus van de bodem kunnen worden verwacht door het gebruik van oppervlaktebodemvochtgegevens afkomstig van de recent gelanceerde European Space Agency (ESA) Soil Moisture Ocean Salinity (SMOS) missie (Kerr et al., 2001) of van de National Aeronautics and Space Administration (NASA) Soil Moisture Active/Passive (SMAP) missie (Entekhabi et al., 2010), die in 2015 zal worden gelanceerd. Verder is de ontwikkeling van oppervlaktebodemvochtproducten gebaseerd op gegevens van meerdere sensoren tegelijkertijd (bijvoorbeeld Liu et al., 2011; Dorigo et al., 2012) bijzonder veelbelovend.

De nieuwe op metingen gebaseerde kaarten van diverse belangrijke afvoerkarakteristieken ontwikkeld in de hoofdstukken 5 en 6 zullen kosteloos ter beschikking worden gesteld voor downloading en zullen naar verwachting nuttig blijken voor een verscheidenheid aan macro-hydrologische toepassingen. Onder de mogelijke toepassingen valt de wereldwijde diagnose van macro-hydrologische modellen, zoals eerder aangetoond in hoofdstuk 6 waar de nieuwe kaarten werden vergeleken met gelijksoortige kaarten afkomstig van twee macroschaalmodellen (Noah en PCR-GLOBWB). De nieuwe kaarten bieden verder unieke mogelijkheden voor het parametriseren van dergelijke macro-hydrologische modellen, inclusief zogenoemde 'hyperresolutie' modellen die worden verwacht in de nabije toekomst (Wood et al., 2011). De kaarten van de afvoerkarakteristieken kunnen ook worden gebruikt voor het beheer van watervoorraden en voor het vergelijken van het afvoergedrag van verschillende stroomgebieden (bijvoorbeeld Wagener et al., 2007). Daarnaast biedt de wereldwijde afvoerkaart wellicht mogelijkheden voor het corrigeren van fouten in neerslaggegevens in bergachtige en/of vochtige gebieden door topografische afwijkingen in het meetnetwerk, onderschattingen wegens sneeuw, of interpolatiefouten (Adam et al., 2006). Bovendien kunnen de wereldwijde afvoer- en BFI kaarten worden gecombineerd om de grondwateraanvulling over de hele wereld te schatten (bijvoorbeeld Szilagyi et al., 2003).

De methodiek beschreven in hoofdstukken 5 en 6 om wereldwijde kaarten van diverse belangrijke afvoerkarakteristieken te produceren bleek zeer efficiënt en flexibel en kan ook worden gebruikt om globale kaarten van andere belangrijke afvoerkarakteristieken te produceren. Zo zou de toevoeging van een afvoerkarakteristiek die is gerelateerd aan de seizoenale gang van de afvoer belangrijk kunnen zijn om de periode waarin de sneeuw smelt in bergachtige en/of subarctische stroomgebieden te voorspellen. De methodiek kan wellicht ook gebruikt worden om wereldwijde kaarten van andere belangrijke biofysische, hydrologische of klimatologische variabelen te produceren.

Acknowledgements

First and foremost, I would like to express my sincere appreciation to my promoter, prof.dr. Sampurno Bruijnzeel, for giving me the wonderful opportunity to pursue this PhD, always ready to provide feedback and encouragement and inspiring me to continue to work as a scientific researcher. Muchisimas gracias Samperro, your guidance was invaluable. I would like to offer special thanks to my first co-promotor dr. Jaap Schellekens, for providing the funding and for helping me overcome seemingly unsolvable problems. To my second co-promotor dr. Richard de Jeu, I would like to thank him for recommending me for this position and for being a true source of inspiration and motivation.

I am truly and deeply indebted to dr. Albert van Dijk, my first unofficial supervisor in Australia, for his valuable and constructive suggestions on my research and for inviting me to CSIRO in Australia. I would not have finished my PhD without his willingness to give his time so generously. Advice given by dr. Tim McVicar, my second unofficial supervisor in Australia, has also been a great help during my research. His enthusiasm, useful critiques, and careful editing of my manuscripts cannot go unnoticed. Thanks are also due to the members of the reading committee, prof.dr. Marc Bierkens, dr. Wouter Buytaert, prof.dr. Han Dolman, dr. Mark Mulligan, and prof.dr. Stefan Uhlenbrook, for their time and diligence in assessing this thesis.

I would also like to thank my colleagues at VU University Amsterdam for their support during this PhD endeavour: Jun, Maarten, Artem, Yanjiao, Ilja, Stefanie, Katrin, Lintao, Robert, Martin, Thijs, Héloïse, Angela, Marjan, Roxanna, Berny, Fabricio, Alex, Tiexi, Brett, Wouter, Zaman, Andreas, Diego, Philip, Patricia, and anyone else whom I may have forgotten. To my collegues in Australia, Marcela, Nora, Ben, Csaba, and Jorge, thank you for your friendship and the wonderful trips to the coast. I want to express my gratitude to Ronald from Deltares for accompanying me on my trip to Indonesia. Last but not least, I would like to thank my friends, Jens, Wisse, Ruben, Jeremy, Achille, Bart, Rob, Megan, Florian, Iris, Moustafa, Eva, Gertjan, Wouter, Ellen, Niels, Raymond, Sonia, my parents, my sister, and Edwin, for their enthusiastic encouragement.

References

Achard, F.; Eva, H. D.; Stibig, H.; Mayaux, P.; Gallego, J.; Richards, T., and Malingreau, J. Determination of deforestation rates of the world's humid tropical forests. *Science*, 297 (5583):999–1002, 2002. doi: 10.1126/science.1070656.

Adam, J. C.; Clark, E. A.; Lettenmaier, Dennis P., and Wood, E. F. Correction of global precipitation products for orographic effects. *Journal of Climate*, 19:15–38, 2006. doi: 10.1175/JCLI3604.1.

Adler, R. F.; Huffman, G. J.; Chang, A.; Ferraro, R.; Xie, P.; Janowiak, J.; Rudolf, B.; Schneider, U.; Curtis, S.; Bolvin, D.; Gruber, A.; Susskind, J.; Arkin, P., and Nelkin, E. The version-2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979–present). *Journal of Hydrometeorology*, 4(6):1147–1167, 2003.

Adnan, N. A. and Atkinson, P. M. Exploring the impact of climate and land use changes on streamflow trends in a monsoon catchment. *International Journal of Climatology*, 31(6):815–831, 2011.

Ahiablame, L.; Chaubey, I.; Engel, B.; Cherkauer, K., and Merwade, V. Estimation of annual baseflow at ungauged sites in Indiana USA. *Journal of Hydrology*, 476(1):13–27, 2013.

Aide, T. M. and Grau, H. R. Globalization, migration, and Latin American ecosystems. *Science*, 305(5692):1915–1916, 2004.

Aide, T. M.; Zimmerman, J. K.; Rosario, M., and Marcano, H. Forest recovery in abandoned cattle pastures along an elevational gradient in northeastern Puerto Rico. *Biotropica*, 28(4): 537–548, 1996.

Aide, T. M.; Zimmerman, J. K.; Pascarella, J. B.; Rivera, L., and Marcano-Vega, H. Forest regeneration in a chronosequence of tropical abandoned pastures: Implications for restoration ecology. *Restoration Ecology*, 8(4):328–338, 2000.

Alcaraz-Segura, D.; Chuvieco, E.; Epstein, H. E.; Kasischke, E. S., and Trishchenko, A. P. Debating the greening vs. browning of the North American boreal forest: differences between satellite datasets. *Global Change Biology*, 16(2):760–770, 2010a. doi: 10.1111/j.1365-2486.2009.01956.x.

Alcaraz-Segura, D.; Liras, E.; Tabik, S.; Paruelo, J. M., and Cabello, J. Evaluating the consistency of the 1982–1999 NDVI trends in the Iberian Peninsula across four time-series derived from the AVHRR sensor: LTDR, GIMMS, FASIR, and PAL-II. *Sensors*, 10(2):1291–1314, 2010b. doi: 10.3390/s100201291.

Alila, Y.; Kuraś, P. K.; Schnorbus, M., and Hudson, R. Forests and floods: a new paradigm sheds light on age-old controversies. *Water Resources Research*, 45:W08416, 2009.

Alila, Y.; Hudson, R.; Kuraś, P. K.; Schnorbus, M., and Rasouli, K. Reply to Comment by Jack Lewis et al. on "Forests and floods: A new paradigm sheds light on age-old controversies". *Water Resources Research*, 46:W05802, 2010. doi: 10.1029/2009WR009028.

Allen, R. G.; Pereira, L. S.; Raes, D., and Smith, M. *Crop evapotranspiration — Gguidelines for computing crop water requirements. FAO Irrigation and Drainage Paper 56*. FAO (Food and Agriculture Organization of the United Nations), Rome, 1998. URL `http://www.fao.org/docrep/X0490E/X0490E00.htm`.

Anderson, D. M.; Glibert, P. M., and Burkholder, J. M. Harmful algal blooms and eutrophication: nutrient sources, composition, and consequences. *Estuaries*, 25(4):704–726, 2002.

Anderson, D. R.; Burnham, K. P., and Thompson, W. L. Null hypothesis testing: problems, prevalence, and an alternative. *The Journal of Wildlife Management*, 64(4):912–923, 2000.

Andréassian, V. Waters and forests: from historical controversy to scientific debate. *Journal of Hydrology*, 291(1–2):1–27, 2004.

Andréassian, V.; Lerat, J.; Loumagne, C.; Mathevet, T.; Michel, C.; Oudin, L., and Perrin, C. What is really undermining hydrologic science today? *Hydrological Processes*, 21(20):2819–2822, 2007.

Anyamba, A. and Tucker, C. J. Analysis of Sahelian vegetation dynamics using NOAA-AVHRR NDVI data from 1981–2003. *Journal of Arid Environments*, 63(3):596–614, 2005.

Anyamba, A. and Tucker, C. J. Historical perspectives on AVHRR NDVI and vegetation drought monitoring. In Wardlow, B. D.; Anderson, M. C., and Verdin, J. P., editors, *Remote Sensing of Drought: innovative monitoring approaches*, chapter 2. CRC Press, Boca Raton, FL, USA, 2012.

Arnold, J. G. and Fohrer, N. SWAT2000: current capabilities and research opportunities in applied watershed modelling. *Hydrological Processes*, 19(3):563–572, 2005.

Arora, V. K. The use of the aridity index to assess climate change effect on annual runoff. *Journal of Hydrology*, 265(1–4):164–177, 2002.

ASCE, Task Committee on Application of Artificial Neural Networks in Hydrology. Artificial neural networks in hydrology. I: Preliminary concepts. *Journal of Hydrologic Engineering*, 5(2): 115–123, 2000a.

ASCE, Task Committee on Application of Artificial Neural Networks in Hydrology. Artificial neural networks in hydrology. II: Hydrologic applications. *Journal of Hydrologic Engineering*, 5 (2):124–137, 2000b.

Aubert, D.; Loumagne, C., and Oudin, L. Sequential assimilation of soil moisture and streamflow data in a conceptual rainfall-runoff model. *Journal of Hydrology*, 280(1):145–161, 2003. doi: 10.1016/S0022-1694(03)00229-4.

Bai, Z. G.; Dent, D. L.; Olsson, L., and Schaepman, M. E. Proxy global assessment of land degradation. *Soil Use and Management*, 24(3):223–234, 2008. doi: 10.1111/j.1475-2743.2008.00169.x.

Baker, D. B.; Richards, R. P.; Loftus, T. T., and Kramer, J. W. A new flashiness index: characteristics and applications to midwestern rivers and streams. *Journal of the American Water Resources Association*, 40(2):503–522, 2007.

Baldi, G.; Nosetto, M. D.; Aragón, R.; Aversa, F.; Paruelo, J. M., and Jobbágy, E. G. Long-term satellite NDVI data sets: Evaluating their ability to detect ecosystem functional changes in South America. *Sensors*, 8(9):5397–5425, 2008. doi: 10.3390/s8095397.

Baltas, E. A.; Dervos, N. A., and Mimikou, M. A. Technical note: Determination of the SCS initial abstraction ratio in an experimental watershed in Greece. *Hydrology and Earth System Sciences*, 11:1825–1829, 2007.

Bannari, A.; Morin, D.; Bonn, F., and Huete, A. R. A review of vegetation indices. *Remote Sensing Reviews*, 13(1–2):95–120, 1995.

Bates, B.; Kundzewicz, Z. W.; Wu, S., and Palutikof, J., editors. *Climate change and water*. Technical Paper of the Intergovernmental Panel on Climate Change. IPCC Secretariat, Geneva, Switzerland, 2008. URL `http://www.ipcc.ch/pdf/technical-papers/climate-change-water-en.pdf`.

Batey, T. Soil compaction and soil management — a review. *Soil Use and Management*, 25(4):335–345, 2009.

Batjes, N. H. *ISRIC-WISE derived soil properties on a 5 by 5 arc-minutes global grid (version 1.0)*. ISRIC - World Soil Information, 2006.

Beck, H. E.; Bruijnzeel, L. A.; van Dijk, A. I. J. M.; McVicar, T. R.; Scatena, F. N., and Schellekens, J. The impact of for-

est regeneration on streamflow in 12 meso-scale humid tropical catchments. *Hydrology and Earth System Sciences*, 17(7):2613–2635, 2013a.

Beck, H. E; van Dijk, A. I. J. M; Miralles, D. G.; de Jeu, R. A. M.; Bruijnzeel., L. A.; McVicar, T. R., and Schellekens, J. Global patterns in baseflow index and recession derived from 3166 small catchments. *submitted to Water Resources Research*, 2013b.

Beck, P. S. A.; Juday, G. P.; Alix, C.; Barber, V. A.; Winslow, S. E.; Sousa, E. E.; Heiser, P.; Herriges, J. D., and Goetz, S. J. Changes in forest productivity across Alaska consistent with biome shift. *Ecology Letters*, 14(4):373–379, 2011. doi: 10.1111/j.1461-0248.2011.01598.x.

Bédard, F.; Crump, S., and Gaudreau, J. A comparison between Terra MODIS and NOAA AVHRR NDVI satellite image composites for the monitoring of natural grassland conditions in Alberta, Canada. *Canadian Journal of Remote Sensing*, 32(1):44–50, 2006.

Benstead, J. P. and Leigh, D. S. An expanded role for river networks. *Nature Geoscience*, 5(10):678–679, 2012.

Bergström, S. Development and application of a conceptual runoff model for Scandinavian catchments. PhD thesis, SMHI Reports RHO 7, Swedish Meteorological and Hydrological Institute (SMHI), Norköping, Sweden, 1976.

Berry, S. L. and Roderick, M. L. Estimating mixtures of leaf functional types using continental-scale satellite and climatic data. *Global Ecology and Biogeography*, 11(1):23–39, 2002. doi: 10.1046/j.1466-822X.2002.00183.x.

Betts, R. A.; Boucher, O.; Collins, M.; Cox, P. M.; Falloon, P. D.; Gedney, N.; Hemming, D. L.; Huntingford, C.; Jones, C. D.; Sexton, D. M. H., and Webb, M. J. Projected increase in continental runoff due to plant responses to increasing carbon dioxide. *Nature*, 448:1037–1041, 2007.

Beven, K. J. Changing ideas in hydrology — The case of physically-based models. *Journal of Hydrology*, 105(1–2):157–172, 1989.

Beven, K. J. Prophecy, reality and uncertainty in distributed hydrological modelling. *Advances in Water Resources*, 16(1):41–51, 1993.

Beven, K. J. TOPMODEL: a critique. *Hydrological Processes*, 11 (9):1069–1085, 1997.

Beven, K. J. and Kirkby, M. J. A physically based, variable contributing area model of basin hydrology. *Hydrological Sciences Bulletin*, 24(1):43–69, 1979.

Bicheron, P. and Leroy, M. Bidirectional reflectance distribution function signatures of major biomes observed from space. *Journal of Geophysical Research Atmospheres*, 105(D21):26669–26681, 2000. doi: 10.1029/2000JD900380.

Bierkens, M. F. P. and van Beek, L. P. H. Seasonal predictability of European discharge: NAO and hydrological response time. *Journal of Hydrometeorology*, 10(4):953–968, 2009.

Bishop, C. M. *Neural Networks for Pattern Recognition*. Clarendon Press, Oxford, UK, 1995.

Bland, J. M. and Altman, D. G. Statistic notes: Regression towards the mean. *BMJ*, 308(6942):1499, 1994.

Bland, J. M. and Altman, D. G. Multiple significance tests: the Bonferroni method. *BMJ*, 310(6973):170, 1995.

Bloomfield, J. P.; Allen, D. J., and Griffiths, K. J. Examining geological controls on baseflow index (BFI) using regression analysis: an illustration from the Thames Basin, UK. *Journal of Hydrology*, 373(1):164–176, 2009.

Blöschl, G. and Montanari, A. Climate change impacts - throwing the dice? *Hydrological Processes*, 24(3):374–381, 2010.

Blöschl, G.; Ardoin-Bardin, S.; Bonell, M.; Dorninger, M.; Goodrich, D.; Gutknecht, D.; Matamoros, D.; Merz, B.; Shand, P., and Szolgay, J. At what scales do climate variability and land cover change impact on flooding and low flows? *Hydrological Processes*, 21(9):1241–1247, 2007.

Bonell, M.; Purandara, B. K.; Venkatesh, B.; Krishnaswamy, J.; Archaya, H. A. K.; Singh, U. V.; Jayakumar, R., and Chappell, N. The impact of forest use and reforestation on soil hydraulic conductivity in the Western Ghats of India: implications for surface and sub-surface hydrology. *Journal of Hydrology*, 391 (1–2):47–62, 2010.

Bontemps, S.; Defourny, P., and van Bogaert, E. GlobCover 2009, products description and validation report. Technical report, ESA GlobCover project, available at: `http://ionia1.esrin.esa.int`, 2011.

Booij, M. J. and Krol, M. S. Balance between calibration objectives in a conceptual hydrological model. *Hydrological Sciences Journal*, 55(6):1017–1032, 2010.

Boorman, D. B.; Hollist, J. M., and Lilly, A. Hydrology of soil types: a hydrologically based classification of the soils of the United Kingdom. Technical Report 126, Institute of Hydrology, Wallingford, UK, 1995. URL `http://www.ceh.ac.uk/products/publications/hydrologypublications.html`.

Bosch, J. M. and Hewlett, J. D. A review of catchment experiments to determine the effect of vegetation changes on water yield and evapotranspiration. *Journal of Hydrology*, 55(1–4):3–23, 1982.

Box, E. O.; Holben, B. N., and Kalb, V. Accuracy of the AVHRR vegetation index as a predictor of biomass, primary productivity and net co_2 flux. *Vegetatio*, 80(2):71–89, 1989.

Bradshaw, C. J. A.; Sodhi, N. S.; Peh, K. S. H., and Brook, B. W. Global evidence that deforestation amplifies flood risk and severity in the developing world. *Global Change Biology*, 13(11):2379–2395, 2007.

Brandes, D.; Hoffmann, J. G., and Mangarillo, J. T. Base flow recession rates, low flows, and hydrologic features of small watersheds in Pennsylvania, USA. *Journal of the American Water Resources Association*, 41(5):1177–1186, 2005.

Brauman, K. A.; Daily, G. C.; Duarte, T. K., and Mooney, H. A. The nature and value of ecosystem services: an overview highlighting hydrologic services. *Annual Review of Environment and Resources*, 32:67–98, 2007. doi: 10.1146/annurev.energy.32.031306.102758.

Brocca, L.; Melone, F.; Moramarco, T., and Morbidelli, R. Antecedent wetness conditions based on ers scatterometer data. *Journal of Hydrology*, 364(1–2):73–87, 2008. doi: 10.1016/j.jhydrol.2008.10.007.

Brocca, L.; Melone, F.; Moramarco, T.; Wagner, W.; Naeimi, V.; Bartalis, Z., and Hasenauer, S. Improving runoff prediction through the assimilation of the ASCAT soil moisture product. *Hydrology and Earth System Sciences*, 14:1881–1893, 2010. doi: 10.5194/hess-14-1881-2010.

Brockmann, V. W. *The rural land classification program of Puerto Rico*, chapter vii, pages 115–161. Northwestern University Studies in Geography Number 1. Northwestern University, Evanston, Illinois, 1952.

Bronstert, A. and Bárdossy, A. Uncertainty of runoff modelling at the hillslope scale due to temporal variations of rainfall intensity. *Physics and Chemistry of the Earth*, 28(6–7):283–288, 2003.

Brown, A. E.; Zhang, L.; McMahon, T. A.; Western, A. W., and Vertessy, R. A. A review of paired catchment studies for determining changes in water yield resulting from alterations in vegetation. *Journal of Hydrology*, 310(1–4):28–61, 2005.

Brown, J.; Ferrians, O. J.; Heginbottom, J. A., and Melnikov, E. S. Circum-arctic map of permafrost and ground-ice conditions. Version 2. Technical report, National Snow and Ice Data Center, Boulder, Colorado USA, 1997.

Brown, M. E.; Pinzon, J. E.; Didan, K.; Morisette, J. T., and Tucker, C. J. Evaluation of the consistency of long-term NDVI time series derived from AVHRR, SPOT-vegetation, SeaWiFS, MODIS, and Landsat ETM+ sensors. *IEEE Transactions on Geoscience and Remote Sensing*, 44(7):1787–1793, 2006.

Brown, S. and Lugo, A. E. Tropical secondary forests. *Journal of Tropical Ecology*, 6(1):1–32, 1990.

Bruijnzeel, L. A. (De)forestation and dry season flow in the tropics: a closer look. *Journal of Tropical Forest Science*, 1(3):229–243, 1989.

Bruijnzeel, L. A. *Hydrology of moist tropical forests and effects of conversion: a state-of knowledge review*. UNESCO International Hydrological Programme, Paris, 1990. URL `http://unesdoc.unesco.org/images/0009/000974/097405eo.pdf`.

Bruijnzeel, L. A. Hydrological functions of tropical forests: not seeing the soil for the trees. *Agriculture, Ecosystems and Environment*, 104(1):185–228, 2004.

Brutsaert, W. and Nieber, J. L. Regionalized drought flow hy-

drographs from a mature glaciated plateau. *Water Resources Research*, 13(3):637–643, 1977.

Budyko, M. I. *Climate and life*. Academic Press, New York, 1974.

Buheaosier, ; Tsuchiya, K.; Kaneko, M., and Sung, S. J. Comparison of image data acquired with AVHRR, MODIS, ETM+ and ASTER over Hokkaido, Japan. *Advances in Space Research*, 32 (11):2211–2216, 2003. doi: 10.1016/S0273-1177(03)90544-8.

Bullock, A.; Andrews, A. J., and Mngodo, R. Regional surface water resources and drought assessment. In *Southern African FRIEND*, pages 40–93. UNESCO, Paris, 1997.

Buttle, J. M. and Metcalfe, R. A. Boreal forest disturbance and streamflow response, northeastern Ontario. *Canadian Journal of Fisheries and Aquatic Sciences*, 57(2):5–18, 2000.

Buttle, J. M.; Hazlett, P. W.; Murray, C. D.; Creed, I. F; Jeffries, D. S., and Semkin, R. Prediction of groundwater characteristics in forested and harvested basins during spring snowmelt using a topographic index. *Hydrological Processes*, 15(18):3389–3407, 2001.

Buytaert, W.; Célleri, R.; De Bièvre, B.; Cisneros, F.; Wyseure, G.; Deckers, J., and Hofstede, R. Human impact on the hydrology of the Andean páramos. *Earth-Science Reviews*, 79(1–2): 53–72, 2006.

Buytaert, W/; niguez, V. I and De Bièvre, B. The effects of afforestation and cultivation on water yield in the Andean páramo. *Forest Ecology and Management*, 251(1–2):22–30, 2007.

Calder, I. R. Water use by forests, limits and controls. *Tree Physiology*, 18(8–9):625–631, 1998.

Calder, I. R. *Blue revolution: integrated land and water resource management*. EarthScan, London, UK, 2005.

Calvesbert, R. J. Climate of Puerto Rico and US Virgin Islands. Technical Report Climatography of the United States No. 60-52, US Department of Commerce, Environmental Sciences Administration, 1970.

Campolo, M.; Soldati, A., and Andreussi, P. Forecasting river flow rate during low-flow periods using neural networks. *Water Resources Research*, 35(11):3547–3552, 1999.

Carter, R. W. Accuracy of current meter measurements. *Journal of the Hydraulics Division, Proceedings of the American Society of Civil Engineers*, 4(1):105–115, 1963.

Castiglioni, S.; Lombardi, L.; Tot, E.; Castellarin, A., and Montanari, A. Calibration of rainfall-runoff models in ungauged basins: a regional maximum likelihood approach. *Advances in Water Resources*, 33(10):1235–1242, 2010.

Chander, G. and Markham, B. Revised Landsat-5 TM radiometric calibration procedures and postcalibration dynamic ranges. *IEEE Transactions on Geoscience and Remote Sensing*, 41(11): 2674–2677, 2003. doi: 10.1109/TGRS.2003.818464.

Chander, G.; Markham, B. L., and Helder, D. L. Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors. *Remote Sensing of Environment*, 113(5):893–903, 2009. doi: 10.1016/j.rse.2009.01.007.

Chandler, D. G. Reversibility of forest conversion impacts on water budgets in tropical karst terrain. *Forest Ecology and Management*, 224(1–2):95–103, 2006.

Chandler, D. G. and Walter, M. F. Runoff responses among common land uses in the Uplands of Matalom, Leyte, Philippines. *Transactions of the ASAE*, 41(6):1635–1641, 1998.

Chang, A. T. C. and Rango, A. Algorithm theoretical basis document for the AMSR-E snow water equivalent algorithm. Technical report, NASA Goddard Space Flight Center, Greenbelt, MD, US, 2000. URL `http://nsidc.org/data/amsre/pdfs/amsr_atbd_snow.pdf`.

Chapman, T. A comparison of algorithms for stream flow recession and baseflow separation. *Hydrological Processes*, 13(5):701–714, 1999.

Chapman, T. G. and Maxwell, A. I. Baseflow separation - comparison of numerical methods with tracer experiments. In *Institute Engineers Australia National Conference Publ. 96/05*, pages 539–545, 1996.

Chappell, N. A. Soil pipe distribution and hydrological functioning within the humid tropics: a synthesis. *Journal of Hydrology*, 24 (12):1567–1581, 2010.

Chappell, N. A. and Tych, W. Identifying step changes in single streamflow and evaporation records due to forest cover change. *Hydrological Processes*, 26(1):100–116, 2012.

Chappell, N. A.; Bonell, M.; Barnes, C., and Tych, W. Tropical cyclone effects on rapid runoff responses: quantifying with new continuous-time transfer function models. In *Revisiting Experimental Catchment Studies in Forest Hydrology*, IAHS Red Book Series, pages 82–93. IAHS Press, Wallingford, UK, 2012.

Chen, M.; Shi, W.; Xie, P.; Silva, V. B. S.; Kousky, V. E.; Higgins, R. W., and Janowiak, J. E. Assessing objective techniques for gauge-based analyses of global daily precipitation. *Journal of Geophysical Research*, 113:D04110, 2008.

Cheng, G. W. Forest change: hydrological effects in the upper Yangtze river valley. *Ambio*, 28(5):457–459, 1999.

Cheng, L.; Yaeger, M.; Viglione, A.; Coopersmith, E.; Ye, S., and Sivapalan, M. Exploring the physical controls of regional patterns of flow duration curves – part 1: insights from statistical analyses. *Hydrology and Earth System Sciencies*, 16:4435–4446, 2012. doi: 10.5194/hess-16-4435-2012.

Chiew, F. H. S.; Stewardson, M. J., and McMahon, T. A. Comparison of six rainfall and runoff modelling approaches. *Journal of Hydrology*, 147(1–4):1–36, 1993. doi: 10.1016/0022-1694(93)90073-I.

Chin, A. Urban transformation of river landscapes in a global context. *Geomorphology*, 79(3–4):460–487, 2006.

Chinea, J. D. and Helmer, E. H. Diversity and composition of tropical secondary forests recovering from large-scale clearing: results from the 1990 inventory in Puerto Rico. *Forest Ecology and Management*, 180(1–3):227–240, 2003.

Choi, J. Y.; Engel, B. A., and Chung, H. W. Daily streamflow modelling and assessment based on the curve-number technique. *Hydrological Processes*, 16(16):3131–3150, 2002.

Chow, V. T.; Maidment, D. R., and Mays, L. W. *Applied Hydrology*. McGraw-Hill, Singapore, 1988.

Ciolkosz, E. J.; Waltman, W. J.; Simpson, T. W., and Dobos, R. R. Distribution and genesis of soils of the northeastern United States. *Geomorphology*, 2(1–3):285–302, 1989.

Clifford, P.; Richardson, S., and Hémon, D. Assessing the significance of the correlation between two spatial processes. *Biometrics*, 45:123–134, 1989.

Collischonn, B.; Collischonn, W., and Tucci, C. E. M. Daily hydrological modeling in the Amazon basin using TRMM rainfall estimates. *Journal of Hydrology*, 360(1–4):207–216, 2008.

Comarazamy, D. E. and González, J. E. Regional long-term climate change (1950–2000) in the midtropical Atlantic and its impacts on the hydrological cycle of Puerto Rico. *Journal of Geophysical Research: Atmospheres*, 116(D21), 2011. doi: 10.1029/2010JD015414.

Costa, M. H.; Botta, A., and Cardille, J. A. Effects of large-scale changes in land cover on the discharge of the Tocantins River, Southeastern Amazonia. *Journal of Hydrology*, 283(1–4):206–217, 2003.

Cracknell, A. P. *The Advanced Very High Resolution Radiometer*. Taylor and Francis, London, 1997.

Crook, K. E.; Scatena, F. N., and Pringle, C. M. Water withdrawn from the Luquillo Experimental Forest, 2004. Technical Report IITF-GTR-34, U.S. Department of Agriculture, Forest Service, International Institute of Tropical Forestry, San Juan, Puerto Rico, 2007. URL `http://www.fs.fed.us/global/iitf/pubs/iitf_gtr34.pdf`.

Crow, W. T. and Ryu, D. A new data assimilation approach for improving runoff prediction using remotely-sensed soil moisture retrievals. *Hydrology and Earth System Sciences*, 13:1–16, 2009. doi: 10.5194/hess-13-1-2009.

Cyr, J.-F.; Landry, M., and Gagnon, Y. Methodology for the large-scale assessment of small hydroelectric potential: application to the province of New Brunswick (Canada). *Renewable Energy*, 36(11):2940–2950, 2011.

Czikowsky, M. J. and Fitzjarrald, D. R. Evidence of seasonal changes in evapotranspiration in eastern U.S. hydrological records. *Journal of Hydrometeorology*, 5(5):974–988, 2004.

Daly, C.; Neilson, R. P., and Phillips, D. L. A statistical-topographic model for mapping climatological precipitation over

mountainous terrain. *Journal of Applied Meteorology*, 33(2): 140–158, 1994.

Daly, C.; Helmer, E. H., and Quiñones, M. Mapping the climate of Puerto Rico, Vieques and Culebra. *International Journal of Climatology*, 23(11):1359–1381, 2003.

Davis, S. N. *Porosity and permeability of natural materials*, pages 53–89. Academic Press, New York, US, 1969.

De Beurs, K. M. and Henebry, G. M. Trend analysis of the Pathfinder AVHRR Land (PAL) NDVI data for the deserts of Central Asia. *IEEE Geoscience and Remote Sensing Letters*, 1 (4):282–286, 2004. doi: 10.1109/LGRS.2004.834805.

De Beurs, K. M. and Henebry, G. M. Northern annular mode effects on the land surface phenologies of northern Eurasia. *Journal of Climate*, 21(17):4257–4279, 2008. doi: 10.1175/2008JCLI2074.1.

De Jeu, R. A. M.; Wagner, W.; Holmes, T. R. H.; Dolman, A. J.; Van De Giesen, N. C., and Friesen, J. Global soil moisture patterns observed by space borne microwave radiometers and scatterometers. *Surveys in Geophysics*, 29(4–5):399–420, 2008.

Dee et al., D. P. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society Part A*, 137(656):553–597, 2011.

DeFries, R. S.; Hansen, M. C.; Townshend, J. R. G.; Janetos, A. C., and Loveland, T. R. A new global 1-km dataset of percentage tree cover derived from remote sensing. *Global Change Biology*, 6(2):247–254, 2000. doi: 10.1046/j.1365-2486.2000.00296.x.

Del Mar López, T.; Aide, T. M., and Scatena, F. N. The effect of land use on soil erosion in the Guadiana watershed in Puerto Rico. *Caribbean Journal of Science*, 34(3–4):298–307, 1998.

Demuth, S. and Hagemann, I. Estimation of flow parameters applying hydrogeological area information. In Seuna, P.; Gustard, A.; Arnell, N. W., and Cole, G. A., editors, *FRIEND: Flow Regimes from International Experimental and Network Data (Proceedings of the Braunschweig Conference, October 1993), IAHS publication no. 221*, pages 151–157, 1994.

DeWalle, D. R.; Swistock, B. R.; Johnson, T. E., and McGuire, K. J. Potential effects of climate change and urbanization on mean annual streamflow in the United States. *Water Resources Research*, 36(9):2655–2664, 2000.

Dietz, A. J.; Kuenzer, C.; Gessner, U., and Dech, S. Remote sensing of snow — a review of available methods. *International Journal of Remote Sensing*, 33(13):4094–4134, 2011.

Dietz, J. L. *Economic History of Puerto Rico. Institutional Change and Political Development.* Princeton University Press, Princeton, NJ, 1986.

Dirks, K. N.; Hay, J. E.; Stow, C. D., and Harris, D. High-resolution studies of rainfall on Norfolk Island: Part II: Interpolation of rainfall data. *Journal of Hydrology*, 208(3–4):187–193, 1998.

Döll, P. and Fiedler, K. Global-scale modeling of groundwater recharge. *Hydrology and Earth System Sciences*, 12(3):863–885, 2008.

Donohue, R. J.; Roderick, M. L., and McVicar, T. R. Deriving consistent long-term vegetation information from AVHRR reflectance data using a cover-triangle-based framework. *Remote Sensing of Environment*, 112(6):2938–2949, 2008.

Donohue, R. J.; McVicar, T. R., and Roderick, M. L. Climate-related trends in Australian vegetation cover as inferred from satellite observations, 1981–2006. *Global Change Biology*, 15 (4):1025–1039, 2009. doi: 10.1111/j.1365-2486.2008.01746.x.

Donohue, R. J.; Roderick, M. L., and McVicar, T. R. Assessing the differences in sensitivities of runoff to changes in climatic conditions across a large basin. *Journal of Hydrology*, 406(3–4): 234–244, 2011.

Donohue, R.J.; Roderick, M. L., and McVicar, T. R. Can dynamic vegetation information improve the accuracy of Budykos hydrological model? *Journal of Hydrology*, 390(1–2):23–34, 2010.

Dorigo, W.; de Jeu, R.; Chung, D.; Parinussa, R.; Liu, Y.; Wagner, W., and Fernández-Prieto, D. Evaluating global trends (1988–2010) in harmonized multi-satellite surface soil moisture. *Geophysical Research Letters*, 39(18), 2012. doi: 10.1029/2012GL052988.

Draper, C.; Mahfouf, J.-F.; Calvet, J.-C.; Martin, E., and Wagner, W. Assimilation of ASCAT near-surface soil moisture into the SIM hydrological model over France. *Hydrology and Earth System Sciences*, 15(12):3829–3841, 2011.

Draper, C. S.; Walker, J. P.; Steinle, P. J.; de Jeu, R. A. M., and Holmes, T. R. H. Remotely sensed soil moisture over Australia from AMSR-E. In *Proceedings MODSIM 2007*. Christchurch, 2007.

Driessen, T. L. A.; Hurkmans, R. T. W. L.; Terink, W.; Hazenberg, P.; Torfs, P. J. J. F., and Uijlenhoet, R. The hydrological response of the Ourthe catchment to climate change as modelled by the HBV model. *Hydrology and Earth System Sciences*, 14 (4):651–665, 2010.

Drigo, R. Trends and patterns of tropical land use change. In Bonell, M. and Bruijnzeel, L. A., editors, *Forests, water and people in the humid tropics*, pages 9–39. Cambridge University Press, Cambridge, UK, 2005.

Duan, Q.; Schaake, J., and Koren, V. *A Priori* estimation of land surface model parameters. In Lakshmi, V.; Albertson, J., and Schaake, J., editors, *Land Surface Hydrology, Meteorology, and Climate: Observations and Modeling*, number 3 in Water Science and Application, pages 77–94. AGU, Washington, D.C., US, 2001.

Duan, Q.; Schaake, J.; Andréassian, V.; Franks, S.; Goteti, G.; Gupta, H. V.; Gusev, Y. M.; Habets, F.; Hall, A.; Hay, L.; Hogue, T.; Huang, M.; Leavesley, G.; Liang, X.; Nasonova, O. N.; Noilhan, J.; Oudin, L.; Sorooshian, S.; Wagener, T., and Wood, E. F. Model Parameter Estimation Experiment (MOPEX): An overview of science strategy and major results from the second and third workshops. *Journal of Hydrology*, 320(1):3–17, 2006.

Dunne, T. and Black, R. D. An experimental investigation of runoff production in permeable soils. *Water Resources Research*, 6(2): 478–490, 1970.

Dürr, H. H.; Meybeck, M., and Dürr, S. H. Lithologic composition of the Earth's continental surfaces derived from a new digital map emphasizing riverine material transfer. *Global Biogeochemical Cycles*, 19(4), 2005. doi: 10.1029/2005GB002515.

Dyhr-Nielsen, M. Hydrological effect of deforestation in the Chao Phraya basin in Thailand. In *Paper Presented at the International Symposium on Tropical Forest Hydrology and Application*, page 12, Chiang Mai, Thailand, 1986.

Eckhardt, K. A comparison of baseflow indices, which were calculated with seven different baseflow separation methods. *Journal of Hydrology*, 352(1–2):168–173, 2008.

Ek, M. B.; Mitchell, K. E.; Lin, Y.; Rogers, E.; Grunmann, P.; Koren, V.; Gayno, G., and Tarpley, J. D. Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model. *Journal of Geophysical Research*, 108(D22):8851, 2003.

Eklundh, L. and Olsson, L. Vegetation index trends for the African Sahel 1982–1999. *Geophysical Research Letters*, 30(8):1430–1433, 2003. doi: 10.1029/2002GL016772.

El Saleous, N. Z.; Kaufman, Y. J., and Dutton, E. Data preprocessing: Stratospheric aerosol perturbing effect on the remote sensing of vegetation: Correction method for the composite NDVI after the Pinatubo eruption. *International Journal of Remote Sensing*, 15(1):7–21, 1997. doi: 10.1080/02757259709532328.

El Saleous, N. Z.; Vermote, E. F.; Justice, C. O.; Townshend, J. R. G.; Tucker, C. J., and Goward, S. N. Improvements in the global biospheric record from the Advanced Very High Resolution Radiometer (AVHRR). *International Journal of Remote Sensing*, 21(6–7):1251–1277, 2000. doi: 10.1080/014311600210164.

Elkaduwa, W. K. B. and Sakthivadivel, R. Use of historical data as a decision support tool in watershed management: a case study of the Upper Nilwala basin in Sri Lanka. Technical Report 26, International Water Management Institute, Colombo, Sri Lanka, 1999.

Ellison, D.; Futter, M. N., and Bishop, K. On the forest cover-water yield debate: from demand- to supply-side thinking. *Global Change Biology*, 18(3):806–820, 2012.

Entekhabi, D.; Asrar, G. R.; Betts, A. K.; Beven, K. J.; Bras, R. L.; Duffy, C. J.; Dunne, T.; Koster, R. D.; Lettenmaier, D. P.; McLaughlin, D. B.; Shuttleworth, W. J.; van Genuchten, M. T.; Wei, M., and Wood, E. F. An agenda for land surface hydrology research and a call for the second international hydrological decade. *Bulletin of the American Meteorological Society*, 80(10):2043–2058, 1999.

Entekhabi, D.; Njoku, E. G.; O'Neill, P. E.; Kellogg, K. H.; Crow, W. T.; Edelstein, W. N.; Entin, J. K.; Goodman, S. D., and others, . The Soil Moisture Active Passive (SMAP) mission. In *Proceedings of the IEEE 98.5*, pages 704–716, 2010. doi: 10.1109/jproc.2010.2043918.

Fang, J.; Piao, S.; Field, C. B.; Pan, Y.; Guo, Q.; Zhou, L.; Peng, C., and Tao, S. Increasing net primary production in China from 1982 to 1999. *Frontiers in Ecology and the Environment*, 1(6):293–297, 2003.

FAO, Food and Agriculture Organization. *A Provisional Methodology for Soil Degradation Assessment*, volume 1. Food and Agriculture Organization of the United Nations, 1979.

FAO, Food and Agriculture Organization. Forest and floods. Drowning in fiction or thriving on facts? Technical report, Food and Agriculture Organization of the United Nations, Rome, 2005. URL `http://www.fao.org/docrep/008/ae929e/ae929e00.htm`.

FAO, Food and Agriculture Organization. Global forest resources assessment 2005. Technical report, Food and Agriculture Organization of the United Nations, Rome, 2006. URL `http://www.fao.org/forestry/fra/fra2005/en/`.

FAO/IIASA, Food and Agriculture Organization/International Institute for Applied Systems Analysis. Harmonized world soil database (version 1.2). Technical report, FAO, Rome, Italy and IIASA, Laxenburg, Austria, 2012.

Farvolden, R. N. Geologic controls on ground-water storage and base flow. *Journal of Hydrology*, 1(3):219–249, 1963.

Fekete, B. M. and Vörösmarty, C. J. The current status of global river discharge monitoring and potential new technologies complementing traditional discharge measurements. In *Predictions in Ungauged Basins: PUB Kick-off (Proceedings of the PUB Kick-off meeting held in Brasilia, 20–22 November 2002), IAHS publication no. 309*, pages 129–136, 2007. URL `http://iahs.info/redbooks/a309/309015.pdf`.

Fekete, B. M.; Vörösmarty, C. J., and Grabs, W. Global, composite runoff fields based on observed river discharge and simulated water balances. Technical Report 22, Global Runoff Data Centre (GRDC) and Federal Institute of Hydrology (BfG), Koblenz, Germany, 1999.

Fekete, B. M.; Vörösmarty, C. J., and Grabs, W. High-resolution fields of global runoff combining observed river discharge and simulated water balances. *Global Biogeochemical Cycles*, 16(3): 15–1, 2002.

Fenicia, F.; Savenije, H. H. G.; Matgen, P., and Pfister, L. Is the groundwater reservoir linear? Learning from data in hydrological modelling. *Hydrology and Earth System Sciences*, 10(1): 139–150, 2006.

Fensholt, R. Earth observation of vegetation status in the Sahelian and Sudanian West Africa: comparison of Terra MODIS and NOAA AVHRR satellite data. *International Journal of Remote Sensing*, 25(9):1641–1659, 2004. doi: 10.1080/01431160310001598999.

Fensholt, R.; Nielsen, T. T., and Stisen, S. Evaluation of AVHRR PAL and GIMMS 10-day composite NDVI time series products using SPOT-4 vegetation data for the African continent. *International Journal of Remote Sensing*, 27(13):2719–2733, 2006. doi: 10.1080/01431160600567761.

Fensholt, R.; Rasmussen, K.; Nielsen, T. T., and Mbow, C. Evaluation of earth observation based long term vegetation trends – intercomparing NDVI time series trend analysis consistency of Sahel from AVHRR GIMMS, Terra MODIS and SPOT VGT data. *Remote Sensing of Environment*, 113(9):1886–1898, 2009. doi: 10.1016/j.rse.2009.04.004.

Fischer, G.; Nachtergaele, F.; Prieler, S.; van Velthuizen, H. T.; Verelst, L., and Wiberg, D. Global agro-ecological zones assessment for agriculture (GAEZ 2008). Technical report, IIASA, Laxenburg, Austria and FAO, Rome, Italy, 2008.

Fisher, J. B.; Whittaker, R. J., and Malhi, Y. ET come home: potential evapotranspiration in geographical ecology. *Global Ecology and Biogeography*, 20(1):1–18, 2011.

Fox, J.; Truong, D. M.; Rambo, A. T.; Tuyen, N. P.; Cuc, L. T., and Leisz, S. Shifting cultivation: a new old paradigm for managing tropical forests. *BioScience*, 50(6):521–528, 2000.

Fuhriman, D. K. and Smith, R. M. Conservation and consumptive use of water with sugar cane under irrigation in the south coastal area of Puerto Rico. *Journal of Agriculture of the University of Puerto Rico*, 35(1):1–45, 1951.

Gallo, K. P.; Daughtry, C. S. T., and Bauer, M. E. Spectral estimation of absorbed photosynthetically active radiation in corn canopies. *Remote Sensing of Environment*, 17(3):221–232, 1985. doi: 10.1016/0034-4257(85)90096-3.

Gallo, K. P.; Jib, L.; Reed, B.; Eidenshink, J., and Dwyer, J. Multi-platform comparisons of MODIS and AVHRR Normalized Difference Vegetation Index data. *Remote Sensing of Environment*, 99(3):221–231, 2005. doi: 10.1016/j.rse.2005.08.014.

Galton, F. Regression towards mediocrity in hereditary stature. *The Journal of the Anthropological Institute of Great Britain and Ireland*, 15:246–263, 1886.

Gan, T. Y. and Biftu, G. F. Automatic calibration of conceptual rainfall-runoff models: Optimization algorithms, catchment conditions, and model structure. *Water Resources Research*, 32 (12):3513–3524, 1996.

Garcia, M.; Peters-Lidard, C. D., and Goodrich, D. C. Spatial interpolation of precipitation in a dense gauge network for monsoon storm events in the southwestern United States. *Water Resources Research*, 44:W05S13, 2008. doi: 10.1029/2006WR005788.

García-Martinó, A. R.; Warner, G. S.; Scatena, F. N., and Civco, D. L. Rainfall, runoff and elevation relationships in the Luquillo Mountains of Puerto Rico. *Caribbean Journal of Science*, 32(4): 413–424, 1996.

Ghimire, C. P.; Bruijnzeel, L. A.; Bonell, M.; Coles, N.; Lubczynski, M. W., and Gilmour, D. A. The effects of sustained forest use on hillslope soil hydraulic conductivity in the Middle Mountains of Central Nepal. *Ecohydrology (in press)*, 2013. doi: 10.1002/eco.1367.

Giambelluca, T. W. Hydrology of altered tropical forest. *Hydrological Processes*, 16(8):1665–1669, 2002.

Giambelluca, T. W.; Scholz, F. G.; Bucci, S. J.; Meinzer, F. C.; Goldstein, G.; Hoffmann, W. A.; Franco, A. C., and Buchert, M. P. Evapotranspiration and energy balance of Brazilian savannas with contrasting tree density. *Agricultural and Forest Meteorology*, 149(8):1365–1376, 2009.

Gleason, B. E. Data documentation for data set 9101. Technical report, National Climatic Data Center, 2002. URL `http://www1.ncdc.noaa.gov/pub/data/documentlibrary/tddoc/td9101.pdf`.

Gleeson, T.; Smith, L.; Moosdorf, N.; Hartmann, J.; Dürr, H. H.; Manning, A. H.; van Beek, L. P. H., and Jellinek, A. M. Mapping permeability over the surface of the earth. *Geophysical Research Letters*, 38(L02401), 2011. doi: 10.1029/2010GL045565.

Goetz, S. J.; Bunn, A. G.; Fiske, G. J., and Houghton, R. A. Satellite-observed photosynthetic trends across boreal North America associated with climate and fire disturbance. *Proceedings of the National Academy of Sciences*, 102(38):13521–13525, 2005. doi: 10.1073/pnas.0506179102.

Gong, D. Y. and Shi, P. J. Northern hemispheric NDVI variations associated with large-scale climate indices in spring. *International Journal of Remote Sensing*, 24(12):2559–2566, 2003.

Gordon, H. R.; Brown, J. W., and Evans, R. H. Exact Rayleigh scattering calculations for use with the Nimbus-7 coastal zone color scanner. *Applied Optics*, 27(5):862–871, 1988.

Gosling, S. N. and Arnell, N. W. Simulating current global river runoff with a global hydrological model: model revisions, validation, and sensitivity analysis. *Hydrological Processes*, 25(7): 1129–1145, 2011.

Govindaraju, R. S. and Rao, A. R. *Artificial Neural Networks in Hydrology.* Kluwer Academic Publishers, Dordrecht, The Netherlands, 2000.

Goward, S. N.; Tucker, C. J., and Dye, D. G. North American vegetation patterns observed with the NOAA-7 Advanced Very High Resolution Radiometer. *Plant Ecology*, 64(1):3–14, 1985. doi: 10.1007/BF00033449.

Goward, S. N.; Markham, B.; Dye, D. G.; Dulaney, W., and Yang, J. Normalized Difference Vegetation Index measurements from the Advanced Very High Resolution Radiometer. *Remote Sensing of Environment*, 35(2–3):257–277, 1991.

Goyal, M. R. and González-Fuentes, E. A. Estimating water consumptive use by sugarcane at four regional sites in Puerto Rico. *Journal of Agriculture of the University of Puerto Rico*, 73(1): 93–95, 1989.

Grau, H. R.; Aide, T. M.; Zimmerman, J. K.; Thomlinson, J. R.; Helmer, E. H., and Zou, X. The ecological consequences of socioeconomic and land-use changes in postagriculture Puerto Rico. *BioScience*, 53(12):1159–1168, 2003.

Guerschman, J.; van Dijk, A.; McVicar, T.; van Niel, T.; Li, L., and Liu, Y. Water balance estimates from satellite observations over the Murray–Darling Basin, 2008. Report to the Australian Government from the CSIRO Murray-Darling Basin Sustainable Yields Project CSIRO: Water for a Healthy Country National Research Flagship.

Gupta, A. Geoindicators for tropical urbanization. *Environmental Geology*, 42(7):736–742, 2002.

Gupta, H. V.; Perrin, C.; Kumar, R.; Blöschl, G.; Clark, M.; Montanari, A., and Andréassian, V. Large-sample hydrology: a need to balance depth with breadth. *Hydrology and Earth System Sciences Discussions*, 10(7):9147–9189, 2013.

Gustard, A. and Irving, K. M. Classification of the low flow response of European soils. In Seuna, P.; Gustard, A.; Arnell, N. W., and Cole, G. A., editors, *FRIEND: Flow Regimes from International Experimental and Network Data (Proceedings of the Braunschweig Conference, October 1993), IAHS publication no. 221*, pages 113–117, 1994.

Gutiérrez, M. V. and Meinzer, F. C. Estimating water use and irrigation requirements of coffee in Hawaii. *Journal of the American Society for Horticultural Science*, 119(3):652–657, 1994.

Gutman, G. G. On the use of long-term global data of land reflectances and vegetation indices derived from the Advanced Very High Resolution Radiometer. *Journal of Geophysical Research*, 104(D6):6241–6255, 1999. doi: 10.1029/1998JD200106.

Gutman, G. G.; Byrnes, R.; Masek, J. G.; Covington, S. J.; Justice, C. O.; Franks, S., and Headley, R. K. Towards monitoring land-cover and land-use changes at a global scale: the global land survey 2005. *Photogrammetric Engineering & Remote Sensing*, 74(1):6–10, 2008.

Haberlandt, U.; Klöcking, B.; Krysanova, V., and Becker, A. Regionalisation of the base flow index from dynamically simulated flow components – a case study in the Elbe River Basin. *Journal of Hydrology*, 248(1–4):35–53, 2001.

Hagan, M. T. Training feedforward networks with the Marquardt algorithm. *IEEE Transactions on Neural Networks*, 5(6):989–993, 1994.

Hall, A. C.; Schumann, G. J. P.; Bamber, J. L., and Bates, P. D. Tracking water level changes of the Amazon basin with spaceborne remote sensing and integration with large scale hydrodynamic modelling: a review. *Physics and Chemistry of the Earth, Parts A/B/C*, 36(7–8):223–231, 2011.

Hall, F.; Masek, J. G., and Collatz, G. J. Evaluation of ISLSCP initiative II FASIR and GIMMS NDVI products and implications for carbon cycle science. *Journal of Geophysical Research*, 111(D22), 2006. doi: 10.1029/2006JD007438.

Hall, F. R. Base-flow recessions—a review. *Water Resources Research*, 4(5):973–983, 1968.

Hansen, L. K. and Salamon, P. Neural network ensembles. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12 (10):993–1001, 1990.

Hansen, M.; DeFries, R. S.; Townshend, J. R. G.; Carroll, M.; Dimiceli, C., and Sohlberg, R. A. Global percent tree cover at a spatial resolution of 500 meters: first results of the MODIS Vegetation Continuous Fields algorithm. *Earth Interactions*, 7 (10):1–15, 2003.

Hansen, M. C. and Loveland, T. R. A review of large area monitoring of land cover change using Landsat data. *Remote Sensing of Environment*, 122:66–74, 2012. doi: 10.1016/j.rse.2011.08.024.

Hansen, M. C.; DeFries, R. S.; Townshend, J. R. G., and Sohlberg, R. Global land cover classification at 1 km resolution using a decision tree classifier. *International Journal of Remote Sensing*, 21:1331–1365, 2000.

Hargreaves, G. H. and Samani, Z. A. Reference crop evapotranspiration from temperature. *Applied Engineering in Agriculture*, 1(2):96–99, 1985.

Harmsen, E. W. Fifty years of crop evapotranspiration studies in Puerto Rico. *Journal of Soil and Water Conservation*, 58(4): 214–223, 2003.

Hartmann, J. and Moosdorf, N. The new global lithological map database GLiM: a representation of rock properties at the Earth surface. *Geochemistry Geophysics Geosystems*, 13(12):1–37, 2012.

Harto, A. B.; Kondoh, A., and Sakura, Y. The effect of land use changes on the water balance in the Ciliwung-Cisadane catchment, West Java, Indonesia. In *Proceedings of the International Symposium on Hydrology, Water Resources and Environment, Development and Management in South East Asia and the Pacific*, pages 121–132, Taegu, Republic of Korea, 1998.

Hassler, S. K.; Zimmermann, B.; van Breugel, M.; Hall, J. S., and Elsenbeer, H. Recovery of saturated hydraulic conductivity under secondary succession on former pasture in the humid tropics. *Forest Ecology and Management*, 261(10):1634–1642, 2011.

Hastenrath, S. *Climate dynamics of the tropics*. Kluwer Academic Publishers, Boston, 1991.

Hawkins, R. H. Asymptotic determination of runoff curve numbers from data. *Journal of Irrigation and Drainage Engineering*, 119 (2):334–345, 1993. doi: 10.1061/(ASCE)0733-9437(1993)119: 2(334).

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

Hecht, S. The new rurality: Globalization, peasants and the paradoxes of landscapes. *Land Use Policy*, 27(2):161–169, 2010.

Helmer, E. H. Forest conservation and land development in Puerto Rico. *Landscape Ecology*, 19(1):29–40, 2004.

Helmer, E. H. and Ruefenacht, B. Cloud-free satellite image mosaics with regression trees and histogram matching. *Photogrammetric Engineering & Remote Sensing*, 71(9):1079–1089, 2005.

Helsel, D. R. and Hirsch, R. M. *Techniques of Water-Resources Investigations Book 4, Chapter A3: Statistical Methods in Water Resources*. U.S. Geological Survey, 2002. URL `http://pubs.usgs.gov/twri/twri4a3/`.

Herrmann, S. M.; Anyamba, A., and Tucker, C. J. Recent trends in vegetation dynamics in the African Sahel and their relationship to climate. *Global Environmental Change Part A*, 15(4):394–404, 2005. doi: 10.1016/j.gloenvcha.2005.08.004.

Hicke, J. A.; Asner, G. P.; Randerson, J. T.; Tucker, C. J.; Los, S. O.; Birdsey, R.; Jenkins, J. C., and Field, C. B. Trends in North American net primary productivity derived from satellite observations, 1982–1998. *Global Biogeochemical Sciences*, 16 (2):1018, 2002. doi: 10.1029/2001GB001550.

Hijmans, R. J.; Cameron, S. E.; Parra, J. L.; Jones, P. G., and Jarvis, A. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25(15): 1965–1978, 2005.

Hirpa, F. A.; Gebremichael, M., and Hopson, T. Evaluation of high-resolution satellite precipitation products over very complex terrain in Ethiopia. *Journal of Applied Meteorology and Climatology*, 49(5):1044–1051, 2010.

Holben, B. N. Characteristics of maximum-value composite images from temporal AVHRR data. *International Journal of Remote Sensing*, 7(11):1417–1434, 1986. doi: 10.1080/01431168608948945.

Hollis, G. E. The effect of urbanization on floods of different recurrence interval. *Water Resources Research*, 11(3):431–435, 1975.

Hölscher, D.; Mackensen, J., and Roberts, J. M. Forest recovery in the humid tropics: changes in vegetation structure, nutrient

pools and the hydrological cycle. In Bonell, M. and Bruijnzeel, L. A., editors, *Forest, water and peaople in the humid tropics*, chapter 25, pages 622–651. Cambridge University Press, United Kingdom, 2005.

Holwerda, F.; Scatena, F. N., and Bruijnzeel, L. A. Throughfall in a Puerto Rican lower montane rain forest: a comparison of sampling strategies. *Journal of Hydrology*, 327(3-4):592–602, 2006.

Holwerda, F.; Bruijnzeel, L. A.; Scatena, F. N.; Vugts, H. F., and Meesters, A. G. C. A. Wet canopy evaporation from a Puerto Rican lower montane rain forest: The importance of realistically estimated aerodynamic conductance. *Journal of Hydrology*, 414–415:1–15, 2012.

Horton, R. E. The role of infiltration in the hydrological cycle. *Trans. Am. Geophys. Union*, 14:446–460, 1933.

Huete, A. R.; Liu, H. Q.; Batchily, K., and van Leeuwen, W. J. D. A comparison of vegetation indices over a global set of TM images for EOS-MODIS. *Remote Sensing of Environment*, 59 (3):440–451, 1997. doi: 10.1016/S0034-4257(96)00112-5.

Huete, A. R.; Justice, C. O., and van Leeuwen, W. MODIS vegetation index (MOD 13) algorithm theoretical basis document version 3. Technical report, University of Arizona, 1999. URL `http://modis.gsfc.nasa.gov/data/atbd/atbd_mod13.pdf`.

Huete, A. R.; Didan, K.; Miura, T.; Rodriguez, E. P.; Gao, X., and Ferreira, L. G. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment*, 83(1):195–213, 2002.

Huffman, G. J.; Adler, R. F.; Stocker, E. F.; Bolvin, D. T., and Nelkin, E. J. Analysis of TRMM 3-hourly multi-satellite precipitation estimates computed in both real and post-real time, 2003. Combined Preprints CD-ROM, 83rd AMS Annual Meeting, Poster P4.11 in: 12th Conf. on Satellite Meteorology and Oceanography.

Hughes, G. O. An analysis of baseflow recession in the Republic of South Africa. Master's thesis, Department of Agricultural Engineering, University of Natal, Pietermaritzburg, South Africa, 1997. URL `http://researchspace.ukzn.ac.za/xmlui/handle/10413/5416`.

Hunt, E. R. Relationship between woody biomass and PAR conversion efficiency for estimating net primary production from NDVI. *International Journal of Remote Sensing*, 15(8):1725–1729, 1994. doi: 10.1080/01431169408954203.

Huppert, H. E and Sparks, R. S. J. Extreme natural hazards: population growth, globalization and environmental change. *Philosophical Transactions of The Royal Society of London, Series A*, 364(1845):1875–1888, 2006.

Ichii, K.; Kawabata, A., and Yamaguchi, Y. Global correlation analysis for NDVI and climatic variables and NDVI trends: 1982–1990. *International Journal of Remote Sensing*, 23(18): 3873–3878, 2002. doi: 10.1080/01431160110119416.

Ilstedt, U.; Malmer, A.; Verbeeten, E., and Murdiyarso, D. The effect of afforestation on water infiltration in the tropics: a systematic review and meta-analysis. *Forest Ecology and Management*, 251(1–2):45–51, 2007.

Ingram, J. C. and Dawson, T. P. Technical note: Inter-annual analysis of deforestation hotspots in Madagascar from high temporal resolution satellite observations. *International Journal of Remote Sensing*, 26(7):1447–1461, 2005. doi: 10.1080/01431160412331291189.

IPCC, Intergovernmental Panel on Climate Change. Climate Change 2007: the physical science basis. In Solomon, S.; Qin, D.; Manning, M.; Chen, Z.; Marquis, M.; Averyt, K. B.; Tignor, M., and Miller, H. L., editors, *Summary for Policymakers. Contribution of working group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, pages 1–18. Cambridge University Press, Cambridge, UK, 2007.

Jackson, R. B.; Jobbágy, E. G; Avissar, R.; Baidya Roy, S.; Barrett, D.; Cook, C. W.; Farley, K. A.; le Maitre, D. C.; McCarl, B. A., and Murray, B. C. Trading water for carbon with biological carbon sequestration. *Science*, 310(5756):1944–1947, 2005.

Jacobs, J. M.; Myers, D. A., and Whitfield, B. M. Improved rainfall/runoff estimates using remotely sensed soil moisture. *Journal of the American Water Resources Association*, 39(2): 313–324, 2003. doi: 10.1111/j.1752-1688.2003.tb04386.x.

Jacobson, C. R. Identification and quantification of the hydrological impacts of imperviousness in urban catchments: a review. *Journal of Environmental Management*, 92(6):1438–1448, 2011.

Jakeman, A. J. and Hornberger, G. M. How much complexity is warrented in a rainfall-runoff model? *Water Resources Research*, 29(8):2637–2649, 1993.

James, M. E. and Kalluri, S. N. V. The Pathfinder AVHRR land data set: an improved coarse resolution data set for terrestrial monitoring. *International Journal of Remote Sensing*, 15(17): 3347–3363, 1994. doi: 10.1080/01431169408954335.

Janeau, J. L.; Planchon, J. P. Bricquet O., and Valentin, C. Soil crusting and infiltration on steep slopes in northern Thailand. *European Journal of Soil Science*, 54(3):543–554, 2003.

Jeffrey, S. J.; Carter, J. O.; Moodie, K. M., and Beswick, A. R. Using spatial interpolation to construct a comprehensive archive of Australian climate data. *Environmental Modelling and Software*, 16(4):309–330, 2001.

Jeong, S. J.; Ho, C. H., and Jeong, J. H. Increase in vegetation greenness and decrease in springtime warming over east Asia. *Geophysical Research Letters*, 36:L02710, 2009.

Ji, L.; Gallo, K. P.; Eidenshink, J. C., and Dwyer, J. Agreement evaluation of AVHRR and MODIS 16-day composite NDVI data sets. *International Journal of Remote Sensing*, 29(16):4839–4861, 2008. doi: 10.1080/01431160801927194.

Jia, G. J.; Epstein, H. E., and Walker, D. A. Greening of arctic Alaska, 1981–2001. *Geophysical Research Letters*, 30(20):HLS 3–1, 2003. doi: 10.1029/2003GL018268.

Jia, L.; Shang, H.; Hu, G., and Menenti, M. Phenological response of vegetation to upstream river flow in the Heihe Rive basin by time series analysis of MODIS data. *Hydrology and Earth System Sciences*, 15:1047–1064, 2011. doi: 10.5194/hess-15-1047-2011.

Johnson, D. H. The insignificance of statistical significance testing. *The Journal of Wildlife Management*, 63(3):763–772, 1999.

Jones, J. *Encyclopedia of Hydrological Sciences*, chapter 121. Intersite comparisons of rainfall-runoff processes. Wiley, Chichester, UK, 2005.

Jones, J. A. A. *The Nature of Soil Piping: a Review of Research*. Geobooks, Norwich, UK, 1981.

Jordan, J. P. Spatial and temporal variability of stormflow generation processes on a Swiss catchment. *Journal of Hydrology*, 153(1–4):357–382, 1994.

Jordan, T. E.; Correll, D. L., and Weller, D. E. Relating nutrient discharges from watersheds to land use and streamflow variability. *Water Resources Research*, 33(11):2579–2590, 1997.

Juhrbandt, J.; Leuschner, C., and Hölscher, D. The relationship between maximal stomatal conductance and leaf traits in eight Southeast Asian early successional tree species. *Forest Ecology and Management*, 202(1–3):245–256, 2004.

Julien, Y.; Sobrino, S. A., and Verhoef, W. Changes in land surface temperatures and NDVI values over Europe between 1982 and 1999. *Remote Sensing of Environment*, 103(1):43–55, 2006. doi: 10.1016/j.rse.2006.03.011.

Kauffeldt, A.; Halldin, S.; Rodhe, A.; Xu, C.-Y., and Westerberg, I. K. Disinformative data in large-scale hydrological modelling. *Hydrology and Earth System Sciences Discussions*, 10(1):487–517, 2013.

Kaufman, Y. J. and Holben, B. N. Calibration of the AVHRR visible and near-IR bands by atmospheric scattering, ocean glint and desert reflection. *International Journal of Remote Sensing*, 14(1):21–52, 1993. doi: 10.1080/01431169308904320.

Kawamura, K.; Akiyama, T.; Yokota, H.; Tsutsumi, M.; Yasuda, T.; Watanabe, O., and Wang, S. Comparing MODIS vegetation indices with AVHRR NDVI for monitoring the forage quantity and quality in Inner Mongolia grassland, China. *Grassland Science*, 51(1):33–40, 2005. doi: 10.1111/j.1744-6961.2004.00006.x.

Kendall, M. G. *Rank Correlation Methods*. Griffin, London, 1975.

Kennaway, T. and Helmer, E. H. The forest types and ages cleared for land development in Puerto Rico. *GIScience & Remote Sensing*, 44(4):356–382, 2007.

Kerr, Y. H.; Waldteufel, P.; Wigneron, J.-P.; Martinuzzi, J.; Font, J., and Berger, M. Soil moisture retrieval from space: the Soil

Moisture and Ocean Salinity (SMOS) mission. *IEEE Transactions on Geoscience and Remote Sensing*, 39(8):1729–1735, 2001.

Kidwell, K. B. *NOAA polar orbiter data user's guide (TIROS-N, NOAA-6, NOAA-7, NOAA-8, NOAA-9, NOAA-10, NOAA-11, NOAA-12, NOAA-13 and NOAA-14)*. National Oceanic and Atmospheric Administration, Washington D.C., 1998. URL `http://www.ncdc.noaa.gov/oa/pod-guide/ncdc/docs/podug/index.htm`.

Kim, N. W. and Lee, J. Temporally weighted average curve number method for daily runoff simulation. *Hydrological Processes*, 22 (25):4936–4948, 2008.

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

King, E. A. Curation of the Pathfinder AVHRR Land Data set at CSIRO. Technical report, CSIRO Marine & Atmospheric Research, Canberra, Australia, 2006. URL `http://www.cmar.csiro.au/e-print/internal/kinge_x2006.pdf`.

Kirchner, J. W.; Feng, X., and Neal, C. Fractal stream chemistry and its implications for contaminant transport in catchments. *Nature*, 403:524–527, 1999. doi: 10.1038/35000537.

Kirkby, M.; Bracken, L., and Reaney, S. The influence of land use, soils and topography on the delivery of hillslope runoff to channels in SE Spain. *Earth Surface Processes and Landforms*, 27(13):1459–1473, 2002.

Kleynhans, W.; Olivier, J. C.; Wessels, K. J.; Salmon, B. P.; van den Bergh, F., and Steenkamp, K. Detecting land cover change using an extended Kalman filter on MODIS NDVI time-series data. *IEEE Geoscience and Remote Sensing Letters*, 8(3):507–511, 2011.

Kolpin, D. W.; Skopec, M.; Meyer, M. T.; Furlong, E. T., and Zaugg, S. D. Urban contribution of pharmaceuticals and other organic wastewater contaminants to streams during differing flow conditions. *Science of the Total Environment*, 328(1–3): 119–130, 2004.

Koster, R. D.; Mahanama, S. P. P.; Livneh, B.; Lettenmaier, D. P., and Reichle, R. H. Skill in streamflow forecasts derived from large-scale estimates of soil moisture and snow. *Nature Geoscience*, 3:613–616, 2010. doi: 10.1038/ngeo944.

Krakauer, N. Y. and Temimi, M. Stream recession curves and storage variability in small watersheds. *Hydrology and Earth System Sciences*, 15(7):2377–2389, 2011.

Kramer, H. J. *Observation of the Earth and Its Environment: survey of Missions and Sensors*. Springer-Verlag, Berlin, Germany, 4th edition, 2002.

Kruskal, W. H. and Wallis, W. A. Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, 47(260):583–621, 1952.

Kundzewicz, Z. W. and Robson, A. J. Change detection in hydrological records - a review of the methodology. *Hydrological Sciences Journal*, 49(1):1–19, 2004.

Kurtz, A. C.; Lugolobi, F., and Salvucci, G. Germanium-silicon as a flow path tracer: application to the Río Icacos watershed. *Water Resources Research*, 47:W06516, 2011. doi: 10.1029/2010WR009853.

Lacey, G. and Grayson, R. Relating baseflow to catchment properties in south-eastern Australia. *Journal of Hydrology*, 204(1–4): 231–250, 1998.

Landell-Mills, N. and Porras, I. T. Silver bullet or fools' gold: a global review of markets for forest environmental services and their impact on the poor. Technical report, International Institute for Environment and Development, London, UK, 2002.

Larsen, M. C. Analysis of 20th century rainfall and streamflow to characterize drought and water resources in Puerto Rico. *Physical Geography*, 21(6):494–521, 2000.

Lee, C.H.; Chen, W.P., and Lee, P.H. Estimation of groundwater recharge using water balance coupled with base-flow-record estimation and stable-base-flow analysis. *Environmental Geology*, 51(1):73–82, 2006.

Lee, J. and Heaney, J. Estimation of urban imperviousness and its impacts on storm water systems. *Journal of Water Resources Planning and Management*, 129(5):419–426, 2003.

Lele, S. Watershed services of tropical forests: from hydrology to economic valuation to integrated analysis. *Current Opinion in Environmental Sustainability*, 1(2):148–155, 2009.

Leopold, L. B. Hydrology for urban land planning — a guidebook on the hydrologic effects of urban land use. Technical report, United States Department of the Interior, Washington, 1968. Geological Survey Circular 554.

Leopoldo, P. R.; de Pádua Sousa, A., and Filho, S. T. Interceptação da áqua de chuva em cultura de cana-de-açúcar. *Brasil Açucareiro*, 98(6):9–16, 1981.

Lepers, E.; Lambin, E. F.; Janetos, A. C.; DeFries, R. S.; Achard, F.; Ramankutty, N., and Scholes, R. J. A synthesis of information on rapid land-cover change for the period 1981–2000. *BioScience*, 55(2):115–124, 2005.

Levenberg, K. A method for the solution of certain problems in least squares. *Quarterly of Applied Mathematics*, 5:164–168, 1944.

Levene, H. *Robust tests for equality of variances*, chapter 25, pages 278–292. Stanford University Press, 1960.

Li, D.; Pan, M.; Cong, Z., and Wood, L. Zhang E. F. Vegetation control on water and energy balance within the Budyko framework. *Water Resources Research*, 49(2):969–976, 2013.

Li, H.; Huang, M.; Wigmosta, M. S.; Ke, Y.; Coleman, A. M.; Leung, L. R.; Wang, A., and Ricciuto, D. M. Evaluating runoff simulations from the Community Land Model 4.0 using observations from flux towers and a mountainous watershed. *Journal of Geophysical Research: Atmospheres (1984–2012)*, 116(D24), 2011. doi: 10.1029/2011JD016276.

Li, Z.-L.; Tang, R.; Wan, Z.; Bi, Y.; Zhou, C.; Tang, B.; Yan, G., and Zhang, X. A review of current methodologies for regional evapotranspiration estimation from remotely sensed data. *Sensors*, 9(5):3801–3853, 2009.

Liang, S. Review on estimation of land surface radiation and energy budgets from ground measurement, remote sensing and model simulations. *IEEE journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 3(3):225–240, 2010.

Liang, X.; Lettenmaier, D. P.; Wood, E. F., and Burges, S. J. A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *Journal of Geophysical Research: Atmospheres (1984–2012)*, 99(D7):14415–14428, 1994.

Lidén, R. and Harlin, J. Analysis of conceptual rainfall and runoff modelling performance in different climates. *Journal of Hydrology*, 238(3–4):231–247, 2000. doi: 10.1016/S0022-1694(00)00330-9.

Lin, B. B. The role of agroforestry in reducing water loss through soil evaporation and crop transpiration in coffee agroecosystems. *Agricultural and Forest Meteorology*, 140(4):510–518, 2010.

Linhares, C. A. *Influência do desmatamento na dinâmica da resposta hidrológica na bacia do Ji-Paraná*. PhD thesis, Brazilian Institute for Space Research, 2005.

Linsley, R. K.; Köhler, M. A., and Paulhus, J. L. H. *Applied Hydrology*. McGraw-Hill, 1949.

Little, C.; Lara, A.; McPhee, J., and Urrutia, R. Revealing the impact of forest exotic plantations on water yield in large scale watersheds in South-Central Chile. *Journal of Hydrology*, 374 (1–2):162–170, 2009.

Liu, Y.; Parinussa, R. M.; Dorigo, W.; de Jeu, R. A. M.; Wagner, W.; van Dijk, A. I. J. M.; McCabe, M. F., and Evans, J. P. Developing an improved soil moisture dataset by blending passive and active microwave satellite-based retrievals. *Hydrology and Earth System Sciences*, 15(2):425–436, 2011.

Lombardi, L.; Toth, E.; Castellarin, A.; Montanari, A., and Brath, A. Calibration of a rainfall-runoff model at regional scale by optimising river discharge statistics: performance analysis for the average/low flow regime. *Physics and Chemistry of the Earth, Parts A/B/C*, 42–44:77–84, 2012. doi: 10.1016/j.pce.2011.05.013.

Longobardi, A. and Villani, P. Baseflow index regionalization analysis in a Mediterranean area and data scarcity context: role of the catchment permeability index. *Journal of Hydrology*, 355 (1–4):63–75, 2008.

Los, S. O. Estimation of the ratio of sensor degradation between NOAA AVHRR channels 1 and 2 from monthly NDVI composites. *IEEE Transactions on Geoscience and Remote Sensing*, 36(1):206–213, 1998. doi: 10.1109/36.655330.

Los, S. O.; Collatz, G. J.; Sellers, P. J.; Malmstrom, C. M.; Pollack, N. H.; DeFries, R. S.; Bounoua, L.; Parris, M. T.; Tucker, C. J., and Dazlich, D. A. A global 9-yr biophysical land surface dataset from NOAA AVHRR data. *Journal of Hydrometeorology*, 1(2): 183–199, 2000.

Lu, G. Y. and Wong, D. W. An adaptive inverse-distance weighting spatial interpolation technique. *Computers & Geosciences*, 34 (9):1044–1055, 2007.

Lu, H.; Raupach, M. R.; McVicar, T. R., and Barrett, D. J. Decomposition of vegetation cover into woody and herbaceous components using AVHRR NDVI time series. *Remote Sensing of Environment*, 86(1):1–18, 2003. doi: 10.1016/S0034-4257(03) 00054-3.

Lu, J.; Sun, G.; McNulty, S. G., and Amatya, D. M. A comparison of six potential evapotranspiration methods for regional use in the southeastern United States. *Journal of the American Water Resources Association*, 41(3):621–633, 2005.

Lugo, A. E. and Helmer, E. H. Emerging forests on abandoned land: Puerto Rico's new forests. *Forest Ecology and Management*, 190(2–3):145–161, 2004.

Lugo, A. E. and Scatena, F. N. Ecosystem-level properties of the Luquillo Experimental Forest with emphasis on the Tabonuco forest. In Lugo, A. E. and Lowe, C., editors, *Tropical Forests: Management and Ecology*, volume 112 of *Ecological Studies*, pages 59–108. Springer-Verlag, New York, 1995.

Lunetta, R. S.; Knight, J. F.; Ediriwickrema, J.; Lyon, J. G., and Worthy, L. D. Land-cover change detection using multi-temporal MODIS NDVI data. *Remote Sensing of Environment*, 105(2):142–154, 2006.

Luojus, K.; Pulliainen, J.; Takala, M.; Lemmetyinen, J.; Derksen, C., and Wang, L. Snow water equivalent (SWE) product guide. Technical report, GlobSnow Consortium, 2010. URL `http://www.globsnow.info/swe/GlobSnow_SWE_product_readme_v1.0a.pdf`.

Madsen, H. Automatic calibration of a conceptual rainfall-runoff model using multiple objectives. *Journal of Hydrology*, 235(3–4):276–288, 2000.

Maier, H. R. and Dandy, G. C. Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. *Environmental Modelling & Software*, 15(1):101–124, 2000.

Mann, H. B. Nonparametric tests against trend. *Econometrica*, 13(3):245–259, 1945.

Marquardt, D. An algorithm for least-squares estimation of nonlinear parameters. *SIAM Journal on Applied Mathematics*, 11 (2):431–441, 1963.

Mazvimavi, D.; Meijerink, A. M. J.; Savenije, H. H. G., and Stein, A. Prediction of flow characteristics using multiple regression and neural networks: a case study in Zimbabwe. *Physics and Chemistry of the Earth, Parts A/B/C*, 30(11):639–647, 2005.

McCloy, K. R.; Los, S.; Lucht, W., and Højsgaard, S. A comparative analysis of three long-term NDVI datasets derived from AVHRR satellite data. *EARSeL eProceedings*, 4(1):52–69, 2005.

McDonnell, J. J.; Sivapalan, M.; Vaché, K.; Dunn, S.; Grant, G.; Haggerty, R.; Hinz, C.; Hooper, R.; Kirchner, J.; Roderick, M. L.; Selker, J., and Weiler, M. Moving beyond heterogeneity and process complexity: a new vision for watershed hydrology. *Water Resources Research*, 43:W07301, 2007. doi: 10.1029/2006WR005467.

McGlynn, B. L.; McDonnell, J. J.; Seibert, J., and Kendall, C. Scale effects on headwater catchment runoff timing, flow sources, and groundwater-streamflow relations. *Water Resources Research*, 40(7), 2004. doi: 10.1029/2003WR002494.

McJannet, D.; Wallace, J.; Fitch, P.; Disher, M., and Reddell, P. Water balance of tropical rainforest canopies in north Queensland, Australia. *Hydrological Processes*, 21(25):3473–3484, 2007.

McKay, M. D.; Conover, W. J., and Beckman, R. J. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2): 239–245, 1979.

McMahon, G.; Bales, J. D.; Coles, J. F.; Giddings, E. M. P., and Zappia, H. Use of stage data to characterize hydrologic conditions in an urbanizing environment. *Journal of the American Water Resources Association*, 39(6):1529–1546, 2003.

McPeters, R. D.; Bhartia, P. K.; Krueger, A. J., and Herman, J. R. *Earth Probe Total Ozone Mapping Spectrometer (TOMS) Data Products User's Guide*. NASA, Goddard Space Flight Center, Greenbelt, Maryland, 1998. URL `http://cedadocs.badc.rl.ac.uk/107/1/earthprobe_userguide.pdf`.

McVicar, T. R. and Jupp, D. L. B. The current and potential operational uses of remote sensing to aid decisions on drought exceptional circumstances in Australia: a review. *Agricultural Systems*, 57(3):399–468, 1998. doi: 10.1016/S0308-521X(98) 00026-2.

McVicar, T. R.; Roderick, M. L.; Donohue, R. J., and Van Niel, T. G. Less bluster ahead? ecohydrological implications of global trends of terrestrial near-surface wind speeds. *Ecohydrology*, 5 (4):381–388, 2012.

Melesse, A. M.; Weng, Q.; Thenkabail, P. S., and Senay, G. B. Remote sensing sensors and applications in environmental resources mapping and modelling. *Sensors*, 7(12):3209–3241, 2007.

Metropolis, N. The beginning of the Monte Carlo method. *Los Alamos Science*, 15:125–130, 1987.

Meyles, E.; Williams, A.; Ternan, L., and Dowd, J. Runoff generation in relation to soil moisture patterns in a small Dartmoor catchment, Southwest England. *Hydrological Processes*, 17(2): 251–264, 2003.

Milly, P. C. D. Climate, soil water storage, and the average annual water balance. *Water Resources Research*, 30(7):2143–2156, 1994.

Miralles, D. G.; Holmes, T. R. H.; de Jeu, R. A. M.; Gash, J. H.; Meesters, A. G. C. A., and Dolman, A. J. Global land-surface evaporation estimated from satellite-based observations. *Hydrology and Earth System Sciences*, 15(2):453–469, 2011.

Mishra, S. K.; Jain, M. K., and Singh, V. P. Evaluation of the SCSCN-based model incorporating antecedent moisture. *Journal of Water Resources Management*, 18:567–589, 2004.

Mitchell, T. D. and Jones, P. D. An improved method of constructing a database of monthly climate observations and associated high-resolution grids. *International Journal of Climatology*, 25: 693–712, 2005. doi: 10.1002/joc.1181.

Molina-Rivera, W. L. and Gómez-Gómez, F. Estimated water use in Puerto Rico, 2005. Technical Report Open-File Report 2008-1286, U.S. Geological Survey, Washington DC, 2008. URL `http://pubs.usgs.gov/of/2008/1286/`.

Monk, W. A.; Wood, P. J.; Hannah, D. M., and Wilson, D. A. Selection of river flow indices for the assessment of hydroecological change. *River Research and Applications*, 23(1):113–122, 2007.

Monteith, J. L. Evaporation and the environment. In *Symposium of the Society of Experimental Biology no. 19*, pages 245–269, Swansea, UK, 1965.

Moradkhani, H. Hydrologic remote sensing and land surface data assimilation. *Sensors*, 8(5):2986–3004, 2008.

Moulin, S.; Kergoat, L.; Viovy, N., and Dedieu, G. Global-scale assessment of vegetation phenology using NOAA/AVHRR satellite measurements. *Journal of Climate*, 10(6):1154–1170, 1997. doi: 10.1175/1520-0442(1997)010⟨1154:GSAOVP⟩2.0.CO;2.

Mulder, V. L.; de Bruin, S.; Schaepman, M. E., and Mayr, T. R. The use of remote sensing in soil and terrain mapping — a review. *Geoderma*, 162(1–2):1–19, 2011.

Mwakalila, S.; Feyen, J., and Wyseure, G. The influence of physical catchment properties on baseflow in semi-arid environments. *Journal of Arid Environments*, 52(2):245–258, 2002.

Myneni, R. B. and Williams, D. L. On the relationship between FAPAR and NDVI. *Remote Sensing of Environment*, 49(3): 200–211, 1994.

Myneni, R. B.; Ramakrishna, R.; Nemani, R. R., and Running, S. W. Estimation of global leaf area index and absorbed PAR using radiative transfer models. *IEEE Transactions on Geoscience and Remote Sensing*, 35(6):1380–1393, 1997. doi:

10.1109/36.649788.

Nagol, J. R.; Vermote, E. F., and Prince, S. D. Effects of atmospheric variation on AVHRR NDVI data. *Remote Sensing of Environment*, 113(2):392–397, 2009. doi: DOI:10.1016/j.rse.2008.10.007.

Nash, J. E. and Sutcliffe, J. V. River flow forecasting through conceptual models part I – A discussion of principles. *Journal of Hydrology*, 10(3):282–290, 1970.

Nasonova, O. N.; Gusev, Y. M., and Kovalev, Y. E. Investigating the ability of a land surface model to simulate streamflow with the accuracy of hydrological models: a case study using MOPEX materials. *Journal of Hydrometeorology*, 10(5):1128–1150, 2009.

Nathan, R. J. and McMahon, T. A. Evaluation of automated techniques for baseflow and recession analysis. *Water Resources Research*, 26(7):1465–1473, 1990.

Neff, B. P.; Day, S. M.; Piggott, A. R., and Fuller, L. M. Base flow in the Great Lakes Basin. Scientific Investigations Report 2005-5217, U.S. Geological Survey, Reston, VA, US, 2005. URL `http://pubs.usgs.gov/sir/2005/5217/`.

Nemani, R. R.; Keeling, C. D.; Hashimoto, H.; Jolly, W. M.; Piper, S. C.; Tucker, C. J.; Myneni, R. B., and Running, S. W. Climate-driven increases in global terrestrial net primary production from 1982 to 1999. *Science*, 300(5625):1560–1563, 2003. doi: 10.1126/science.1082750.

Nepstad, D. C.; de Carvalho, C. R.; Davidson, E. A.; Jipp, P. H.; Lefebvre, P. A.; Negreiros, G. H.; Gustavo, H.; da Silva, E. D.; Stone, T. A.; Trumbore, S. E., and Vieira, S. The role of deep roots in the hydrological and carbon cycles of Amazonian forests and pastures. *Nature*, 372(6507):666–669, 1994.

NFI, National Forest Inventory. National plantation inventory of Australia. Technical report, National Forest Inventory, BRS, Canberra, 1997.

Nicholls, N. Commentary and analysis: the insignificance of significance testing. *Bulletin of the American Meteorological Society*, 82(5):981–986, 2001.

Nieuwolt, S. *Tropical climatology, an introduction to the climates of the low latitudes*. John Wiley and Sons, New York, 1977.

Nijssen, B.; O'Donnell, G. M.; Lettenmaier, D. P.; Lohmann, D., and Wood, E. F. Predicting the discharge of global rivers. *Journal of Climate*, 14(15):3307–3323, 2001.

Niu, G.-Y.; Yang, Z.-L.; Mitchell, K. E.; Chen, F.; Ek, M. B.; Barlage, M.; Kumar, A.; Manning, K.; Niyogi, D.; Rosero, E.; Tewari, M., and Xia, Y. The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. *Journal of Geophysical Research: Atmospheres (1984–2012)*, 116(D12), 2011. doi: 10.1029/2010JD015139.

Njoku, E. G.; Ashcroft, P.; Chan, T. K., and Li, Li. Global survey and statistics of radio-frequency interference in AMSR-E land observations. *IEEE Transactions Geoscience and Remote Sensing*, 43(5):938–947, 2005.

NRC, National Research Council. *Assessment of Hydrologic and Hydrometeorological Operations and Services*, page 51. National Academy Press, 1996.

NRC, National Research Council. *Earth Science and Applications from Space: National Imperatives for the Next Decade and Beyond*, page 200. National Academy Press, 2007a.

NRC, National Research Council. *Earth Science and Applications from Space: National Imperatives for the Next Decade and Beyond (Executive Summary)*, page 35. National Academies Press, 2007b.

Olcott, P. G. Puerto Rico and the U. S. Virgin Islands. In *Ground water atlas of the United States*, chapter HA 730-N. U.S. Geological Survey, Washington DC, 1999. URL `http://pubs.usgs.gov/ha/ha730/ch_n/index.html`.

Ol'dekop, E. M. Ob isparenii s poverknosti rechnykh basseinov (On evaporation from the surface of river basins). *Transactions on Meteorological Observations, University of Tartu 4*, 1911.

Olden, J. D. and Poff, N. L. Redundancy and the choice of hydrologic indices for characterizing streamflow regimes. *River Research and Applications*, 19(2):101–121, 2003.

Oleson, K. W.; Lawrence, D. M.; Bonan, G. B.; Flanner, M. G.; Kluzek, E.; Lawrence, P. J.; Levis, S.; Swenson, S. C., and Thornton, P. E. Technical description of version 4.0 of the Community Land Model (CLM). Technical report, Climate and Global Dynamics Division, National Center for Atmospheric Research, Boulder, CO, US, 2010. URL `http://nldr.library.ucar.edu/repository/assets/technotes/asset-000-000-000-847.pdf`.

Olsen, Edward T. AIRS/AMSU/HSB Version 5 documentation. Technical report, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, US, 2007.

Olson, D. M.; Dinerstein, E.; Wikramanayake, E. D.; Burgess, N. D.; Powell, G. V. N.; Underwood, E. C.; D'amico, J. A.; Itoua, I.; Strand, H. E.; Morrison, J. C.; Loucks, C. J.; Allnutt, T. F.; Ricketts, T. H.; Kura, Y.; Lamoreux, J. F.; Wettengel, W. W.; Hedao, P., and Kassem, K. R. Terrestrial ecoregions of the world: a new map of life on Earth. *BioScience*, 51(11):933–938, 2001.

Ouaidrari, H.; El Saleous, N. Z.; Vermote, E. F.; Townshend, J. R. G., and Goward, S. N. AVHRR Land Pathfinder II (ALP II) data set: evaluation and inter-comparison with other data sets. *International Journal of Remote Sensing*, 24(1):135–142, 2003. doi: 10.1080/01431160305006.

Oudin, L.; Andréassian, V.; Perrin, C.; Michel, C., and Le Moine, N. Spatial proximity, physical similarity, regression and ungaged catchments: a comparison of regionalization approaches based on 913 French catchments. *Water Resources Research*, 44, 2008. doi: 10.1029/2007WR006240.

Owe, M.; de Jeu, R. A. M., and Holmes, T. R. H. Multisensor historical climatology of satellite-derived global land surface moisture. *Journal of Geophysical Research*, 113:F01002, 2008.

Parada, L. M. and Liang, X. A novel approach to infer streamflow signals for ungauged basins. *Advances in Water Resources*, 33(4):372–386, 2010.

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

Parajka, J.; Naeimi, V.; Blöschl, G.; Wagner, W.; Merz, R., and Scipal, K. Assimilating scatterometer soil moisture data into conceptual hydrologic models at the regional scale. *Hydrology and Earth System Sciences Discussions*, 2(6):2739–2786, 2005b.

Parés-Ramos, I. K.; Gould, W. A., and Aide, T. M. Agricultural abandonment, suburban growth, and forest expansion in Puerto Rico between 1991 and 2000. *Ecology and Society*, 13(2):1, 2008.

Park, H. S. and Sohn, B. J. Recent trends in changes of vegetation over East Asia coupled with temperature and rainfall variations. *Journal of Geophysical Research*, 115:D14101, 2010. doi: 10.1029/2009JD012752.

Paruelo, J. M.; Garbulsky, M. F.; Guerschman, J. P., and Jobbagy, E. G. Two decades of Normalized Difference Vegetation Index changes in South America: identifying the imprint of global change. *International Journal of Remote Sensing*, 25(14):2793–2806, 2004. doi: 10.1080/01431160310001619526.

Pauwels, V. R. N.; Hoeben, R.; Verhoest, N. E. C.; de Troch, F. P., and Troch, P. A. Improvement of TOPLATS-based discharge predictions through assimilation of ERS-based remotely sensed soil moisture values. *Hydrological Processes*, 16(5):995–1013, 2002.

Peña-Arancibia, J. L. *Impacts of land use change on dry season flows across the tropics: forests as 'sponges' and 'pumps'*. PhD thesis, King's College London, 2013.

Peña-Arancibia, J. L.; Van Dijk, A. I. J. M.; Mulligan, M., and Bruijnzeel, L. A. The role of climatic and terrain attributes in estimating baseflow recession in tropical catchments. *Hydrology and Earth System Sciences*, 14(11):2193–2205, 2010.

Peña-Arancibia, J. L.; Van Dijk, A. I. J. M.; Guerschman, J. P.; Mulligan, M.; Bruijnzeel, L. A., and McVicar, T. R. Detecting changes in streamflow after partial woodland clearing in two large catchments in the seasonal tropics. *Journal of Hydrology*, 416(1):60–71, 2012.

Peña-Arancibia, J. L.; van Dijk, A. I. J. M.; Renzullo, L. J., and Mulligan, M. Evaluation of precipitation estimation accuracy in reanalyses, satellite products and an ensemble method for regions in Australia and in south and east Asia. *Journal of*

Hydrometeorology (in press), 2013.

Pedelty, J.; Vermote, E. F.; Devadiga, S.; Roy, D.; Schaaf, C.; Privette, J.; Pinheiro, A.; Prince, S.; Justice, C. O.; Nagol, J.; Masuoka, E.; Brown, M. E.; Pinzon, J. E.; Tucker, C. J.; Ju, J., and Liu, J. Generating a long-term land data record from the AVHRR and MODIS instruments. *Geoscience and Remote Sensing Symposium, 2007. IGARSS 2007. IEEE International*, pages 1021–1025, 2007. doi: 10.1109/IGARSS.2007.4422974.

Peel, M. C.; Chiew, F. H. S.; Western, A. W., and McMahon, T. A. Extension of unimpaired monthly streamflow data and regionalisation of parameter values to estimate streamflow in ungauged catchments, 2000. Report prepared for the Australian National Land and Water Resources Audit. Centre for Environmental Applied Hydrology, University of Melbourne, Australia.

Peterson, G. D. Scaling ecological dynamics: Self-organization, hierarchical structure, and ecological resilience. *Climatic Change*, 44(3):291–309, 2000.

Phillips, C. B. and Scatena, F. N. Flashiness indices for urban and rural streams in Puerto Rico. In *AWRA 2010 Summer Specialty Conference*, San Juan, Puerto Rico, 2010. URL `http://www.sas.upenn.edu/lczodata/sites/www.sas.upenn.edu.lczodata/files/ColinPhillips_AWRA%20Flash.pdf`.

Piñeiro, G.; Perelman, S.; Guerschman, J. P., and Paruelo, J. M. How to evaluate models: Observed vs. predicted or predicted vs. observed? *Ecological Modelling*, 216(3–4):316–322, 2008. doi: 10.1016/j.ecolmodel.2008.05.006.

Piao, S.; Mohammat, A.; Fang, J.; Cai, Q., and Feng, J. NDVI-based increase in growth of temperate grasslands and its responses to climate changes in China. *Global Environmental Change*, 16(4):340–348, 2006. doi: 10.1016/j.gloenvcha.2006.02.002.

Pielke, R. A.; Adegoke, J.; Beltrán-Przekurat, A.; Hiemstra, C. A.; Lin, J.; Nair, U. S.; Niyogi, D., and Nobis, T. E. An overview of regional land-use and land-cover impacts on rainfall. *Tellus B*, 59(4923):587–601, 2007.

Pinheiro, V. B. and Naghettini, M. Calibration of the parameters of a rainfall-runoff model in ungauged basins using synthetic flow duration curves as estimated by regional analysis. *Journal of Hydrologic Engineering (in press)*, 2012. doi: 10.1061/(ASCE)HE.1943-5584.0000737.

Pinzon, J. E.; Brown, M. E., and Tucker, C. J. Satellite time series correction of orbital drift artifacts using empirical mode decomposition. In Huang, N., editor, *Hilbert-Huang Transform: Introduction and Applications*, pages 167–86. World Scientific, Hackensack, NJ, 2005.

Plesca, I.; Timbe, E.; Exbrayat, J. F.; Windhorst, D.; Kraft, P.; Crespo, P.; Vachéa, K. B.; Frede, H. G., and Breuer, L. Model intercomparison to explore catchment functioning: Results from a remote montane tropical rainforest. *Ecological Modelling*, 239:3–13, 2012.

Poff, N. L. R.; Allan, D.; Bain, M. B.; Karr, J. R., and Prestegaard, K. L. The natural flow regime. *BioScience*, 47(11):769–784, 1997.

Ponce, V. M. and Hawkins, R. H. Runoff curve number: has it reached maturity? *Journal of Hydrologic Engineering*, 1(1):11–19, 1996.

Post, D. A. and Jakeman, A. J. Relationships between catchment attributes and hydrological response characteristics in small Australian mountain ash catchments. *Hydrological Processes*, 10(6):877–892, 1996. ISSN 1099-1085.

Potapov, P.; Hansen, M. C.; Gerrand, A. M.; Lindquist, E. J.; Pittman, K.; Turubanova, S., and Wilkie, M. L. The global Landsat imagery database for the FAO FRA remote sensing survey. *International Journal of Digital Earth*, 4(1):2–21, 2011. doi: 10.1080/17538947.2010.492244.

Potter, C. S. Terrestrial biomass and the effects of deforestation on the global carbon cycle. *BioScience*, 49(10):769–778, 1999.

Potter, N. J.; Zhang, L.; Milly, P. C. D.; McMahon, T. A., and Jakeman, A. J. Effects of rainfall seasonality and soil moisture capacity on mean annual water balance for Australian catchments. *Water Resources Research*, 41(6), 2005. doi: 10.1029/2004WR003697.

Pouliot, D.; Latifovic, R., and Olthof, I. Trends in vegetation NDVI from 1 km AVHRR data over Canada for the period 1985-2006. *International Journal of Remote Sensing*, 30(1):149–168, 2009. doi: 10.1080/01431160802302090.

Price, J. C. Timing of NOAA afternoon passes. *International Journal of Remote Sensing*, 12(1):193–198, 1991. doi: 10.1080/01431169108929644.

Price, K. Effects of watershed topography, soils, land use, and climate on baseflow hydrology in humid regions: a review. *Progress in Physical Geography*, 35(4):465–492, 2011.

Priestley, C. H. B. and Taylor, R. J. On the assessment of surface heat flux and evaporation using large-scale parameters. *Monthly Wheater Review*, 100:81–92, 1972.

Qian, W. C. Effects of deforestation on flood characteristics with particular reference to Hainan island, China. In *International Association of Hydrological Sciences Publication 140*, pages 249–258, 1983.

Quenouille, M. H. Notes on bias in estimation. *Biometrika*, 43 (3–4):353–360, 1956.

Quintero, M.; Wunder, S., and Estrada, R.D. For services rendered? Modeling hydrology and livelihoods in Andean payments for environmental services schemes. *Forest Ecology and Management*, 258(9):1871–1880, 2009.

Ramillien, G.; Famiglietti, J. S., and Wahr, J. Detection of continental hydrology and glaciology signals from GRACE: a review. *Surveys in Geophysics*, 29(4–5):361–374, 2008.

Ramírez, A.; De Jesús-Crespo, R.; Martinó-Cardona, D. M.; Martínez-Rivera, N., and Burgos-Caraballo, S. Urban streams in Puerto Rico: what can we learn from the tropics? *Journal of the North American Benthological Society*, 28(4):1070–1079, 2009.

Ramos, O. M. and Lugo, A. E. Mapa de la vegetación de Puerto Rico. *Acta Cientifica*, 8(1–2):63–66, 1994.

Rao, C. R. N.; Chen, J.; Staylor, F. W.; Able, P.; Kaufman, Y. J.; Vermote, E. F.; Rossow, W. R., and Brest, C. Degradation of the visible and near-infrared channels of the Advanced Very High Resolution Radiometer on the NOAA-9 spacecraft: assessment and recommendations for corrections. Technical report, NOAA/NESDIS, Washington, D.C., 1993.

Reed, B. C. Trend analysis of time-series phenology of North America derived from satellite data. *GIScience & Remote Sensing*, 43(1):24–38, 2006. doi: 10.2747/1548-1603.43.1.24.

Reed, B. C.; Brown, J. F.; VanderZee, D.; Loveland, T. R.; Merchant, J. W., and Ohlen, D. O. Measuring phenological variability from satellite imagery. *Journal of Vegetation Science*, 5 (5):703–714, 1994. doi: 10.2307/3235884.

Refsgaard, J. C. Parameterisation, calibration and validation of distributed hydrological models. *Journal of Hydrology*, 198(1):69–97, 1997.

Ricotta, C.; Avena, G., and De Palma, A. Mapping and monitoring net primary productivity with AVHRR NDVI time-series: statistical equivalence of cumulative vegetation indices. *ISPRS Journal of Photogrammetry and Remote Sensing*, 54(5–6):325–331, 1999.

Ridder, R. M. Global Forest Resources Assessment 2010: Options and recommendations for a global remote sensing survey of forests. Technical report, Food and Agricultural Organization of the United Nations, 2007. URL `ftp://ftp.fao.org/docrep/fao/010/ai074e/ai074e00.pdf`.

Rijsdijk, A.; Bruijnzeel, L. A., and Sutoto, C. K. Runoff and sediment yield from rural roads, trails and settlements in the upper Konto catchment, East Java, Indonesia. *Geomorphology*, 87(1–2):28–37, 2007.

Roberts, J. M.; Gash, J. H. C.; Tani, M., and Bruijnzeel, L. A. Controls on evaporation in lowland tropical rainforest. In Bonell, M. and Bruijnzeel, L. A., editors, *Forests, water and people in the humid tropics*, pages 287–313. Cambridge University Press, Cambridge, UK, 2005.

Robinson, M.; Cognard-Plancq, A.-L.; Cosandey, C.; David, J.; Durand, P.; Führer, H.-W.; Hall, R.; Hendriques, M. O.; Marc, V.; McCarthy, R.; McDonnell, M.; Martini, C.; Nisbet, T.; O'Dea, P.; Rodgers, M., and Zollner, A. Studies of the impact of forests on peak flows and baseflows: a European perspective. *Forest Ecology and Management*, 186(1–3):85–97, 2003.

Robock, A.; Vinnikov, K. Y.; Srinivasan, G.; Entin, J. K.; Hollinger, S. E.; Speranskaya, N. A.; Liu, S., and Namkhai, A. The global soil moisture data bank. *Bulletin of the American Meteorological Society*, 81(6):1281–1299, 2000.

Rodell, M.; Houser, P.R.; Jambor, U.; Gottschalck, J.; Mitchell, K.; Meng, C.-J.; Arsenault, K.; Cosgrove, B.; Radakovich, J.; Bosilovich, M.; Entin, J.K.; Walker, J.P.; Lohmann, D.; , and Toll, D. The global land data assimilation system. *Bulletin of the American Meteorological Society*, 85(3):381–394, 2004.

Rodriguez, D. A.; Tomasella, J., and Linhares, C. Is the forest conversion to pasture affecting the hydrological response of Amazonian catchments? Signals in the Ji-Paraná Basin. *Hydrological Processes*, 24(10):1254–1269, 2010.

Rodriguez-Iturbe, I. Ecohydrology: a hydrologic perspective of climate-soil-vegetation dynamics. *Water Resources Research*, 36(1):3–9, 2000.

Royall, R. M. The effect of sample size on the meaning of significance tests. *The American Statistician*, 40(4):313–315, 1986.

Sahin, V. and Hall, M. J. The effects of afforestation and deforestation on water yields. *Journal of Hydrology*, 178(1):293–309, April 1996.

Sanchez, P. A.; Ahamed, S.; Carré, F.; Hartemink, A. E.; Hempel, J.; Huising, J.; Lagacherie, P.; McBratney, A. B.; McKenzie, N. J.; de Lourdes Mendonca-Santos, M.; Minasny, B.; Montanarella, L.; Okoth, P.; Palm, C. A.; Sachs, J. D.; Shepherd, K. D.; Vagen, T.; Vanlauwe, B.; Walsh, M. G.; Winowiecki, L. A., and Zhang, G.-L. Digital soil map of the world. *Science*, 325(5941):680–681, 2009.

Santhi, C.; Allen, P. M.; Muttiah, R. S.; Arnold, J. G., and Tuppad, P. Regional estimation of base flow for the conterminous United States by hydrologic landscape regions. *Journal of Hydrology*, 351(1–2):139–153, 2008.

Sarle, W. S. Stopped training and other remedies for overfitting. In *Proceedings of the 27th Symposium on the Interface of Computing Science and Statistics*, pages 352–360, 1995.

Sato, M.; Hansen, J. E.; McCormick, M. P., and Pollack, J. B. Stratospheric aerosol optical depths, 1850–1990. *Journal of Geophysical Research*, 98(D12):22987–22994, 1993. doi: 10.1029/93JD02553.

Sauer, V. B. and Meyer, R. W. Determination of error in individual discharge measurements. Open-file report 92-144, U.S. Geological Survey, Washington DC, 1992. URL `http://pubs.usgs.gov/of/1992/ofr92-144/`.

Schaake, J.; Cong, S., and Duan, Q. The US MOPEX data set. In *Large Sample Basin Experiments for Hydrologyical Model Parameterization: Results of the Model Parameter Experiment — MOPEX*, 2006.

Schaake, J. C.; Koren, V. I.; Duan, Q.-Y.; Mitchell, K., and Chen, F. Simple water balance model for estimating runoff at different spatial and temporal scales. *Journal of Geophysical Research*, 101(3):7461, 1996.

Schellekens, J.; Scatena, F. N.; Bruijnzeel, L. A., and Wickel, A. J. Modelling rainfall interception by a lowland maritime tropical rain forest in northeastern Puerto Rico. *Journal of Hydrology*, 225(3–4):168–184, 1999.

Schellekens, J.; Bruijnzeel, L. A.; Scatena, F. N.; Bink, N. J., and Holwerda, F. Evaporation from a tropical rain forest, Luquillo Experimental Forest, eastern Puerto Rico. *Water Resources Research*, 36(8):2183–2196, 2000.

Schellekens, J.; Scatena, F. N.; Bruijnzeel, L. A.; Van Dijk, A. I. J. M.; Groen, M. M. A., and van Hogezand, R. J. P. Stormflow generation in a small rainforest catchment in the Luquillo Experimental Forest, Puerto Rico. *Hydrological Processes*, 18(3):505–530, 2004.

Schloss, A. L.; Kicklighter, D. W.; Kaduk, J., and Wittenberg, U. Comparing global models of terrestrial Net Primary Productivity (NPP): comparison of NPP to climate and the Normalized Difference Vegetation Index (NDVI). *Global Change Biology*, 5 (1):25–34, 1999. doi: 10.1046/j.1365-2486.1999.00004.x.

Schneider, M. K.; Brunner, F.; Hollis, J. M., and Stamm, C. Towards a hydrological classification of European soils: preliminary test of its predictive power for the base flow index using river discharge data. *Hydrology and Earth System Sciences*, 11 (4):1501–1513, 2007.

Schneider, U.; Fuchs, T.; Meyer-Christoffer, A., and Rudolf, B. Global precipitation analysis products of the GPCC. Internet publication, 2008. URL `ftp://ftp-anon.dwd.de/pub/data/gpcc/PDF/GPCC_intro_products_2008.pdf`.

Schreiber, P. Über die Beziehungen zwischen dem Niederschlag und der Wasseruhrung der Flusse in Mitteleuropa. *Meteorologisches Zeitschrift*, 21(1):441–452, 1904.

Schumm, S. A. Evolution of drainage systems and slopes in badlands at Perth Amboy. *Geological Society of America Bulletin*, 67(5):597–646, 1956.

Scott, D. F.; Bruijnzeel, L. A., and Mackensen, J. The hydrological and soil impacts of forestation in the tropics. In Bonell, M. and Bruijnzeel, L. A., editors, *Forests, water and people in the humid tropics*, pages 622–651. Cambridge University Press, Cambridge, UK, 2005.

Seibert, J. Estimation of parameter uncertainty in the HBV model. *Nordic Hydrology*, 28(4–5):247–262, 1997.

Seibert, J. Regionalisation of parameters for a conceptual rainfall-runoff model. *Agricultural and Forest Meteorology*, 98–99(1–4): 279–293, 1999.

Seibert, J. *HBV light, version 2 user's manual*. Stockholm University, Stockholm, Sweden, 2005. URL `http://people.su.se/~jseib/HBV/HBV_manual_2005.pdf`.

Seibert, J.; Bishop, K. H., and Nyberg, L. A test of TOPMODEL's ability to predict spatially distributed groundwater levels. *Hydrological Processes*, 11(9):1131–1144, 1997.

Sellers, P. J. Canopy reflectance, photosynthesis and transpiration. *International Journal of Remote Sensing*, 6(8):1335–1372, 1985.

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

Shao, J. Linear model selection by cross-validation. *Journal of the American Statistical Association*, 88(422):486–494, 1993.

Shepard, D. A two-dimensional interpolation function for irregularly-spaced data. In *ACM '68 Proceedings of the 1968 23rd ACM National Conference*, pages 517–524, 1968.

Shi, Z.; Chen, L.; Fanga, N.; Qinc, D., and Cai, C. Research on the SCS-CN initial abstraction ratio using rainfall-runoff event analysis in the Three Gorges Area, China. *CATENA*, 352(1): 1–7, 2009. doi: 10.1016/j.catena.2008.11.006.

Shiklomanov, A. I.; Lammers, R. B., and Vörösmarty, C. J. Widespread decline in hydrological monitoring threatens pan-arctic research. *EOS, Transactions American Geophysical Union*, 83(2):13, 2002.

Siles, P.; Vaast, P.; Dreyer, E., and Harmand, J. Rainfall partitioning into throughfall, stemflow and interception loss in a coffee (*Coffea arabica* L.) monoculture compared to an agroforestry system with *Inga densiflora*. *Journal of Hydrology*, 395(1–2): 39–48, 2010.

Silveira, L. and Alonso, J. Runoff modifications due to the conversion of natural grasslands to forests in a large basin in Uruguay. *Hydrological Processes*, 23(2):320–329, 2009.

Sivapalan, M.; Beven, K. J., and Wood, E. F. On hydrologic similarity 2. A scaled model of storm runoff production. *Water Resources Research*, 23(12):2266–2278, 1987.

Sivapalan, M.; Takeuchi, K.; Franks, S. W.; Gupta, V. K.; Karambiri, H.; Lakshmi, V.; Liang, X.; McDonnell, J. J.; Mendiondo, E. M.; O'Connell, P. E.; Oki, T.; Pomeroy, J. W.; Schertzer, D.; Uhlenbrook, S., and Zehe, E. IAHS decade on Predictions in Ungauged basins (PUB), 2003–2012: Shaping an exciting future for the hydrological sciences. *Hydrological Sciences Journal*, 48 (6):857–880, 2003.

Slater, A. G.; Bohn, T. J.; McCreight, J. L.; Serreze, M. C., and Lettenmaier, D. P. A multimodel simulation of pan-Arctic hydrology. *Journal of Geophysical Research: Biogeosciences (2005–2012)*, 112(G4), 2007. doi: 10.1029/2006JG000303.

Slayback, D. A.; Pinzon, J. E.; Los, S. O., and Tucker, C. J. Northern hemisphere photosynthetic trends 1982–99. *Global Change Biology*, 9(1):1–15, 2003. doi: 10.1046/j.1365-2486.2003.00507.x.

Smakhtin, V. U. Low flow hydrology: a review. *Journal of Hydrology*, 240(3–4):147–186, 2001.

Smith, E. A.; Asrar, G.; Furuhama, Y.; Ginati, A.; Kummerow, C., and others, . International Global Precipitation Measurement (GPM) program and mission: an overview. In *Measuring Precipitation From Space*, volume 28 of *Advances In Global Change Research*, pages 611–653. Springer, 2004.

Smith, E. A.; Asrar, G. R.; Furuhama, Y.; Ginati, G.; Kummerow, C.; Levizzani, V.; Mugnai, A.; Nakamura, K.; Adler, R.; Casse, V.; Cleave, M.; Debois, M.; John, J.; Entin, J.; Houser, P.; Iguchi, T.; Kakar, R.; Kaye, J.; Kojima, M.; Lettenmaier, D.; Luther, M.; Mehta, A.; Morel, P.; Nakazawa, T.; Neeck, S.; Okamoto, K.; Oki, R.; Raju, G.; Shepherd, M.; Stocker, E.; Testud, J., and Wood, E. *The International Global Precipitation Measurement (GPM) program and mission: An overview, in Measuring Precipitation From Space: EURAINSAT and the Future*. Springer, New York, 2006.

Smith, L. C. Satellite remote sensing of river inundation area, stage, and discharge: a review. *Hydrological Processes*, 11(10): 1427–1439, 1997.

Smith, R. M. and Abruña, A. Soil and water conservation research in Puerto Rico, 1938 to 1947. Bulletin no. 124, University of Puerto Rico Agricultural Experiment Station, Rio Piedras, Puerto Rico, 1955.

Sobrino, J. A.; Julien, Y., and Morales, L. Multitemporal analysis of PAL images for the study of land cover dynamics in South America. *Global and Planetary Change*, 51(3–4):172–180, 2006. doi: 10.1016/j.gloplacha.2006.01.006.

Song, Y.; Ma, M., and Veroustraete, F. Comparison and conversion of AVHRR GIMMS and SPOT VEGETATION NDVI data in China. *International Journal of Remote Sensing*, 31(9): 2377–2392, 2010. doi: 10.1080/01431160903002409.

Soulsby, C. and Tetzlaff, D. Towards simple approaches for mean residence time estimation in ungauged basins using tracers and soil distributions. *Journal of Hydrology*, 363(1–4):60–74, 2008.

Sperna Weiland, F. C.; Tisseuil, C.; Dürr, H. H.; Vrac, M., and van Beek, L. P. H. Selecting the optimal method to calculate daily global reference potential evaporation from CFSR reanalysis data. *Hydrology and Earth System Sciences*, 16(3):983–1000, 2011.

Staylor, W. F. Degradation rates of the AVHRR visible channel for the NOAA 6, 7, and 9 spacecraft. *Journal of Atmospheric and Oceanic Technology*, 7:411–423, 1990.

Steele-Dunne, S.; Lynch, P.; McGrath, R.; Semmler, T.; Wang, S.; Hanafin, J., and Nolan, P. The impacts of climate change on hydrology in Ireland. *Journal of Hydrology*, 356(1–2):28–45, 2008.

Steininger, M. K. Tropical secondary forest regrowth in the Amazon: age, area and change estimation with Thematic Mapper data. *International Journal of Remote Sensing*, 17(1):9–27, 1996. doi: 10.1080/01431169608948984.

Stellmes, M.; Udelhoven, T.; Röoder, A.; Sonnenschein, R., and Hill, J. Dryland observation at local and regional scale comparison of Landsat TM/ETM+ and NOAA AVHRR time series. *Remote Sensing of Environment*, 114(10):2111–2125, 2010. doi: 10.1016/j.rse.2010.04.016.

Stephens, G. L. and Kummerow, C. D. The remote sensing of clouds and precipitation from space: a review. *Journal of the Atmospheric Sciences*, 64(11):3742–3765, 2007.

Stöckli, R. and Vidale, P. L. European plant phenology and climate as seen in a 20-year AVHRR land-surface parameter dataset. *International Journal of Remote Sensing*, 25(17):3303–3330, 2004. doi: 10.1080/01431160310001618149.

Stokstad, E. Scarcity of rain, stream gages threatens forecasts. *Science*, 285(5431):1199–1200, 1999.

Stow, D.; Petersen, A.; Hope, A.; Engstrom, R., and Coulter, L. Greenness trends of Arctic tundra vegetation in the 1990s: comparison of two NDVI data sets from NOAA AVHRR systems. *International Journal of Remote Sensing*, 28(21):4807–4822, 2007. doi: 10.1080/01431160701264284.

Stowe, L. L.; David, P. A., and McClain, E. P. Scientific basis and initial evaluation of the CLAVR-1 global clear/cloud classification algorithm for the Advanced Very High Resolution Radiometer. *Journal of Atmospheric and Oceanic Technology*, 16(6):656–681, 1999.

Su, F.; Hong, Y., and Lettenmaier, D. P. Evaluation of TRMM Multisatellite Precipitation Analysis (TMPA) and its utility in hydrologic prediction in the La Plata Basin. *Journal of Hydrometeorology*, 9(4):622, 2008.

Su, F.; Gao, H.; Huffman, G. J., and Lettenmaier, D. P. Potential utility of the real-time TMPA-RT precipitation estimates in streamflow prediction. *Journal of Hydrometeorology*, 12(3): 444–455, 2011.

Sun, G.; Zhou, G.; Zhang, Z.; Wei, X.; McNulty, S. G., and Vose, J. M. Potential water yield reduction due to forestation across China. *Journal of Hydrology*, 328(3–4):548–558, 2006.

Sun, G.; Zuo, C.; Liu, S.; Liu, M.; McNulty, S. G., and Vose, J. M. Watershed evapotranspiration increased due to changes in vegetation composition and structure under a subtropical climate. *Journal of the American Water Resources Association*, 44(5): 1164–1175, 2008.

Szilagyi, J.; Harvey, F. E., and Ayers, J. F. Regional estimation of base recharge to ground water using water balance and a base-flow index. *Groundwater*, 41(4):504–513, 2003.

Tague, C. and Grant, G. E. A geological framework for interpreting the low-flow regimes of Cascade streams, Willamette River Basin, Oregon. *Water Resources Research*, 40(4):W04303, 2004.

Tallaksen, L. M. A review of baseflow recession analysis. *Journal of Hydrology*, 165:349–370, 1995.

Tang, Q.; Gao, H.; Lu, H., and Lettenmaier, D. P. Remote sensing: hydrology. *Progress in Physical Geography*, 33(4):490–509, 2009.

Tanré, D.; Holben, B. N., and Kaufman, Y. J. Atmospheric correction algorithm for NOAA-AVHRR products: theory and application. *IEEE Transactions on Geoscience and Remote Sensing*, 30(2):231–248, 1992.

Tarnavsky, E.; Garrigues, S., and Brown, M. E. Multiscale geostatistical analysis of AVHRR, SPOT-VGT, and MODIS global NDVI products. *Remote Sensing of Environment*, 112(2):535–549, 2008. doi: 10.1016/j.rse.2007.05.008.

Tateishi, R. and Ebata, M. Analysis of phenological change patterns using 1982-2000 Advanced Very High Resolution Radiometer (AVHRR) data. *International Journal of Remote Sensing*, 25(12):2287–2300, 2004. doi: 10.1080/01431160310001618455.

Te Linde, A. H.; Aerts, J. C. J. H.; Hurkmans, R. T. W. L., and Eberle, M. Comparing model performance of two rainfall-runoff models in the Rhine Basin using different atmospheric forcing data sets. *Hydrology and Earth System Sciences*, 12(3):943–957, 2008.

Tebaldi, C. and Knutti, R. The use of the multi-model ensemble in probabilistic climate projections. *Philosophical Transactions of The Royal Society of London, Series A*, 365(1857):2053–2075, 2007.

Tetko, I. V.; Livingstone, D. J., and Luik, A. I. Neural network studies. 1. Comparison of overfitting and overtraining. *Journal of Chemical Information and Modeling*, 35(5):826–833, 1995.

Thomlinson, J. R.; Serrano, M. I.; del Mar López, T.; Aide, T. M., and Zimmerman, J. K. Land-use dynamics in a post-agricultural Puerto Rican landscape (1936–1988). *Biotropica*, 28(4):525–536, 1996.

Tomkins, K. M. Uncertainty in streamflow rating curves: methods, controls and consequences. *Hydrological Processes (in press)*, 2013. doi: 10.1002/hyp.9567.

Trajkovic, S. Hargreaves versus Penman-Monteith under humid conditions. *Journal of Irrigation and Drainage Engineering*, 133(1):38–42, 2007.

Trimble, S. W.; Weirich, F. H., and Hoag, B. L. Reforestation and the reduction of water yield on the Southern Piedmont since circa 1940. *Water Resources Research*, 23(3):425–437, 1987.

Tromp-van Meerveld, H. J. and McDonnell, J. J. Threshold relations in subsurface stormflow: 2. The fill and spill hypothesis. *Water Resources Research*, 42(2), 2006. doi: 10.1029/2004WR003800.

Tromp-van Meerveld, H. J.; Peters, N. E., and McDonnell, J. J. Effect of bedrock permeability on subsurface stormflow and the water balance of a trenched hillslope at the Panola Mountain Research Watershed, Georgia, USA. *Hydrological Processes*, 21

(6):750–769, 2007.

Tschinkel, H. M. Short-term fluctuation in streamflow as related to evaporation and transpiration. *Journal of Geophysical Research*, 68(24):6459–6469, 1963.

Tucker, C. J. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of the Environment*, 8(2):127–150, 1979.

Tucker, C. J.; Slayback, D. A.; Pinzon, J. E.; Los, S. O.; Myneni, R. B., and Taylor, M. G. Higher northern latitude Normalized Difference Vegetation Index and growing season trends from 1982 to 1999. *International Journal of Biometeorology*, 45(4): 184–190, 2001. doi: 10.1007/s00484-001-0109-8.

Tucker, C. J.; Pinzon, J. E.; Brown, M. E.; Slayback, D. A.; Pak, E. W.; Mahoney, R.; Vermote, E. F., and El Saleous, N. Z. An extended AVHRR 8-km NDVI dataset compatible with MODIS and SPOT vegetation NDVI data. *International Journal of Remote Sensing*, 26(20):4485–4498, 2005. doi: 10.1080/01431160500168686.

Tukey, J. W. Bias and confidence in not-quite large samples. *Annals of Mathematical Statistics*, 29(2):614, 1958.

Uchida, T.; Asano, Y.; Ohte, N., and Mizuyama, T. Seepage area and rate of bedrock groundwater discharge at a granitic unchanneled hillslope. *Water Resources Research*, 39(1), 2003. doi: 10.1029/2002WR001298.

UNEP, United Nations Environment Programme. World Atlas of Desertification. Technical report, United Nations Environment Programme (UNEP), Nairobi, 1992.

UNISDR, United Nations International Strategy for Disaster Reduction. Global assessment report on disaster risk reduction 2011. Technical report, United Nations International Strategy for Disaster Reduction, Geneva, Switzerland, 2011.

USDA, United States Department of Agriculture. *Urban hydrology for small watersheds*. Natural Resources Conservation Service, Conservation Engineering Division, 1986.

Van Beek, L. P. H. and Bierkens, M. F. P. The global hydrological model PCR-GLOBWB: conceptualization, parameterization and verification. Technical report, Utrecht University, 2009. URL `http://vanbeek.geo.uu.nl/suppinfo/vanbeekbierkens2009.pdf`.

Van der Molen, M. K. *Meteorological impacts of land use change in the maritime tropics*. PhD thesis, VU University Amsterdam, 2002.

Van der Molen, M. K.; Dolman, A. J.; Waterloo, M. J., and Bruijnzeel, L. A. Climate is affected more by maritime than by continental land use change: a multiple scale analysis. *Global and Planetary Change*, 54(1–2):128–149, 2006.

Van der Weert, R. Hydrological conditions in Indonesia. Delft Hydraulics, Delft, The Netherlands, 1994.

Van Dijk, A. I. J. M. Climate and terrain factors explaining streamflow response and recession in Australian catchments. *Hydrology and Earth System Sciences*, 14(1):159–169, 2010.

Van Dijk, A. I. J. M. and Renzullo, L. J. Water resource monitoring systems and the role of satellite observations. *Hydrology and Earth System Sciences*, 15:39–55, 2011. doi: 10.5194/hess-15-39-2011.

Van Dijk, A. I. J. M.; van Noordwijk, M.; Calder, I. R.; Bruijnzeel, L. A.; Schellekens, J.; Chappell, N. A.; Smyle, J., and Aylward, B. Forest-flood relation still tenuous – comment on 'Global evidence that deforestation amplifies flood risk and severity in the developing world' by C. J. A. Bradshaw, N. S. Sodi, K. S.-H. Peh and B. W. Brook. *Global Change Biology*, 15(1):110–115, 2009.

Van Dijk, A. I. J. M.; Peña-Arancibia, J. L., and Bruijnzeel, L. A. Top-down analysis of collated streamflow data from heterogeneous catchments leads to underestimation of land cover influence. *Hydrology and Earth System Sciences*, 16(9):3461–3473, 2012.

van Zyl, J. J. The Shuttle Radar Topography Mission (SRTM): a breakthrough in remote sensing of topography. *Acta Astronautica*, 48(5–12):559–565, 2001.

Vázquez, R. Water requirements of sugarcane under irrigation in Lajas Valley, Puerto Rico. Bulletin no. 224, University of Puerto Rico Agricultural Experiment, Río Piedras, Puerto Rico, 1970.

Venturini, V.; Bisht, G.; Islam, S., and Jiang, L. Comparison of evaporative fractions estimated from AVHRR and MODIS sensors over South Florida. *Remote Sensing of Environment*, 93(1–2):77–86, 2004. doi: 10.1016/j.rse.2004.06.020.

Verbyla, D. The greening and browning of Alaska based on 1982-2003 satellite data. *Global Ecology and Biogeography*, 17(4): 547–555, 2008. doi: 10.1111/j.1466-8238.2008.00396.x.

Vermote, E. F. and Kaufman, Y. J. Absolute calibration of AVHRR visible and near-infrared channels using ocean and cloud views. *International Journal of Remote Sensing*, 16(13): 2317–2340, 1995.

Vermote, E. F.; El Saleous, N. Z.; Kaufman, Y. J., and Dutton, E. Data pre-processing: Stratospheric aerosol perturbing effect on the remote sensing of vegetation: Correction method for the composite NDVI after the Pinatubo eruption. *Remote Sensing Reviews*, 15(11):7–21, 1997.

Vermote, E. F.; El Saleous, N. Z., and Justice, C. O. Atmospheric correction of MODIS data in the visible to middle infrared: first results. *Remote Sensing of Environment*, 83(1–2):97–111, 2002. doi: 10.1016/S0034-4257(02)00089-5.

Verstraete, M. M.; Pinty, B., and Myneni, R. B. Potential and limitations of information extraction on the terrestrial biosphere from satellite remote sensing. *Remote Sensing of Environment*, 58(2):201–214, 1996. doi: 10.1016/S0034-4257(96)00069-7.

Voepel, H.; Ruddell, B.; Schumer, R.; Troch, P. A.; Brooks, P. D.; Neal, A.; Durcik, M., and Sivapalan, M. Quantifying the role of climate and landscape characteristics on hydrologic partitioning and vegetation response. *Water Resources Research*, 47: W00J09, 2011.

Vogel, R. M. and Kroll, C. N. Regional geohydrologic-geomorphic relationships for the estimation of low-flow statistics. *Water Resources Research*, 28(9):2451–2458, 1992.

Vogel, R. M. and Kroll, C. N. Estimation of baseflow recession constants. *Water Resources Management*, 10(4):303–320, 1996.

Vörösmarty, C. J.; Fekete, B., and Tucker, B. A. River discharge database, Version 1.0 (RivDIS v1.0). A contribution to IHP-V Theme 1. Technical Documents in Hydrology Series 0–6, UNESCO, Paris, France, 1996. URL `http://www.sage.wisc.edu/riverdata/`.

Vörösmarty, C. J.; Green, P.; Salisbury, J., and Lammers, R. B. Global water resources: vulnerability from climate change and population growth. *Science*, 289(5477):284–288, 2000. doi: 10.1126/science.289.5477.284.

Wagener, T. and Montanari, A. Convergence of approaches toward reducing uncertainty in predictions in ungauged basins. *Water Resources Research*, 47(6), 2011. doi: 10.1029/2010WR009469.

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

Wagner, W.; Lemoine, G., and Rott, H. A method for estimating soil moisture from ERS scatterometer and soil data. *Remote Sensing of Environment*, 70(2):191–207, 1999.

Wandishin, M. S.; Mullen, S. L.; Stensrud, D. J., and Brooks, H. E. Evaluation of a short-range multimodel ensemble system. *Monthly Weather Review*, 129(4):729, 2001.

Wang, Q.; Adiku, S.; Tenhunen, J., and Granier, A. On the relationship of NDVI with leaf area index in a deciduous forest site. *Remote Sensing of Environment*, 94(2):244–255, 2005. doi: 10.1016/j.rse.2004.10.006.

Wanner, W.; Li, X., and Strahler, A. H. On the derivation of kernels for kernel-driven models of bidirectional reflectance. *Journal of Geophysical Research*, 100(D10):21077–21089, 1995. doi: 10.1029/95JD02371.

Ward, E.; Buytaert, W.; Peaver, L., and Wheater, H. Evaluation of precipitation products over complex mountainous terrain: a water resources perspective. *Advances in Water Resources*, 34 (10):1222–1231, 2011.

Waterloo, M. J.; Schellekens, J.; Bruijnzeel, L. A., and Rawaqa, T. T. Changes in catchment runoff after harvesting and burning of a *pinus caribaea* plantation in Viti Levu, Fiji. *Forest Ecology and Management*, 251(1–2):31–44, 2007.

Weaver, P. L.; Birdsey, R. A., and Lugo, A. E. Soil organic matter in secondary forests of Puerto Rico. *Biotropica*, 19(1):17–23,

1987.

Wesseling, C. G.; Karssenberg, D. J. A. N.; Burrough, P. A., and Deursen, W. Integrating dynamic environmental models in GIS: the development of a dynamic modelling language. *Transactions in GIS*, 1(1):40–48, 1996.

Wessels, K. J.; van den Bergh, F., and Scholes, R. J. Limits to detectability of land degradation by trend analysis of vegetation index data. *Remote Sensing of Environment*, 125:10–22, 2012.

Widén-Nilsson, E.; Halldin, S., and Xua, C. Global water-balance modelling with WASMOD-M: parameter estimation and regionalisation. *Journal of Hydrology*, 340(1–2):105–118, 2007.

Wilcox, B. P. and Huang, Y. Woody plant encroachment paradox: Rivers rebound as degraded grasslands convert to woodlands. *Geophysical Research Letters*, 37(7):L07402, 2010. doi: 10.1029/2009GL041929.

Wilk, J.; Andersson, L., and Plermkamon, V. Hydrological impacts of forest conversion to agriculture in a large river basin in northeast Thailand. *Hydrological Processes*, 15(14):2729–2748, 2001.

Wittenberg, H. Baseflow recession and recharge as nonlinear storage processes. *Hydrological Processes*, 13(5):715–726, 1999.

Wohl, E.; Barros, A.; Brunsell, N.; Chappell, N. A.; Coe, M.; Giambelluca, T.; Goldsmith, S.; Harmon, R.; Hendrickx, J. M. H.; Juvik, J.; McDonnell, J., and Ogden, F. The hydrology of the humid tropics. *Nature Climate Change*, 2:655–662, 2012. doi: 10.1038/nclimate1556.

Wolock, D. M. and McCabe, G. J. Explaining spatial variability in mean annual runoff in the conterminous United States. *Climate Research*, 11:149–159, 1999.

Wolock, D. M.; Hornberger, G. M.; Beven, K. J., and Campbell, W. G. The relationship of catchment topography and soil hydraulic characteristics to lake alkalinity in the northeastern United States. *Water Resources Research*, 25(5):829–837, 1989.

Wolock, D. M.; Winter, T. C., and McMohan, G. Delineation and evaluation of hydrologic-landscape regions in the United States using geographic information system tools and multivariate statistical analyses. *Environmental Management*, 34(1): 71–88, 2004.

Wood, E. F.; Lettenmaier, D.; Liang, X.; Nijssen, B., and Wetzel, S. W. Hydrological modeling of continental-scale basins. *Annual Review of Earth and Planetary Sciences*, 25:279–300, 1997. doi: 10.1146/annurev.earth.25.1.279.

Wood, E. F.; Roundy, J. K.; Troy, T. J.; van Beek, L. P. H.; Bierkens, M. F. P.; Blyth, E.; de Roo, A.; Döll, P.; Ek, M.; Famiglietti, J.; Gochis, D.; van de Giesen, N.; Houser, P.; Jaffé, P. R.; Kollet, S.; Lehner, B.; Lettenmaier, D. P.; Peters-Lidard, C.; Sivapalan, M.; Sheffield, J.; Wade, A., and Whitehead, P. Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water. *Water Resources Research*, 47(5), 2011. doi: 10.1029/2010WR010090.

Woodward, D. E.; Hawkins, R. H.; Jiang, R.; Hjelmfelt, A. T.; van Mullem, J. A., and Quan, Q. D. Runoff curve number method: Examination of the initial abstraction ratio. In *World Water and Environmental Resources Congress 2003 and Related Symposia*, 2003.

Wright, S. J. The future of tropical forests. *Annals of the New York Academy of Sciences*, 1195:1–27, 2010. doi: 10.1111/j.1749-6632.2010.05455.x.

Wu, W.; Hall, C. A. S.; Scatena, F. N., and Quackenbush, L. J. Spatial modelling of evapotranspiration in the Luquillo experimental forest of Puerto Rico using remotely-sensed data. *Journal of Hydrology*, 328(3–4):733–752, 2006.

Wu, W.; Hall, C. A. S., and Scatena, F. N. Modelling the impact of recent land-cover changes on the stream flows in northeastern Puerto Rico. *Hydrological Processes*, 21(21):2944–2956, 2007.

Xiao, J. and Moody, A. Geographical distribution of global greening trends and their climatic correlates: 1982–1998. *International Journal of Remote Sensing*, 26(11):2371–2390, 2005. doi: 10.1080/01431160500033682.

Xie, P.; Chen, M.; Yang, S.; Yatagai, A.; Hayasaka, T.; Fukushima, Y., and Liu, C. A gauge-based analysis of daily precipitation over East Asia. *Journal of Hydrometeorology*, 8(3):607–626, 2007.

Xie, Y.; Sha, Z., and Yu, M. Remote sensing imagery in vegetation mapping: a review. *Journal of Plant Ecology*, 1(1):9–23, 2008.

Xu, J.; Fox, J.; Lu, X.; Podger, N.; Leisz, S., and Ai, X. Effects of swidden cultivation, state policies, and customary institutions on land cover in a Hani Village, Yunnan, China. *Mountain Research and Development*, 19(2):123–132, 1999.

Yadav, M; Wagener, T., and Gupta, H. Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins. *Advances in Water Resources*, 30(8):1756–1774, 2007.

Yong, B.; Hong, Y.; Ren, L.-L.; Gourley, J. J.; Huffman, G. J.; Chen, X.; Wang, W., and Khan, S. I. Assessment of evolving TRMM-based multisatellite real-time precipitation estimation methods and their impacts on hydrologic prediction in a high latitude basin. *Journal of Geophysical Research: Atmospheres (1984–2012)*, 117(D9), 2012. doi: 10.1029/2011JD017069.

Yu, P-S; Yang, T-C, and Chen, S-J. Comparison of uncertainty analysis methods for a distributed rainfall-runoff model. *Journal of Hydrology*, 244(1–2):43–59, 2001.

Zaitchik, B. F.; Rodell, M., and Olivera, F. Evaluation of the global land data assimilation system using global river discharge data and a source-to-sink routing scheme. *Water Resources Research*, 46(6), 2010. doi: 10.1029/2009WR007811.

Zecharias, Y. B. and Brutsaert, W. The influence of basin morphology on groundwater outflow. *Water Resources Research*, 24(10):1645–1650, 1988.

Zhang, B.; Yang, Y., and Zepp, H. Effect of vegetation restoration on soil and water erosion and nutrient losses of a severely eroded clayey Plinthudult in southeastern China. *Catena*, 57(1):77–90, 2004.

Zhang, L.; Dawes, W. R., and Walker, G. R. Response of mean annual evapotranspiration to vegetation changes at catchment scale. *Water Resources Research*, 37(3):701–708, 2001. doi: 10.1029/2000WR900325.

Zhang, L.; Potter, N.; Hickel, K.; Zhang, Y., and Shao, Q. Water balance modeling over variable time scales based on the Budyko framework - model development and testing. *Journal of Hydrology*, 360(1–4):117–131, 2008a.

Zhang, Z.; Wagener, T.; Reed, P., and Bhushan, R. Reducing uncertainty in predictions in ungauged basins by combining hydrologic indices regionalization and multiobjective optimization. *Water Resources Research*, 44(12), 2008b. doi: 10.1029/2008WR006833.

Zhou, G.; Wei, X.; Luo, Y.; Zhang, M.; Li, Y.; Qiao, Y.; Lui, H., and Wang, C. Forest recovery and river discharge at the regional scale of Guangdong Province, China. *Water Resources Research*, 46(9), 2010. doi: 10.1029/2009WR008829.

Zhou, G. Y.; Morris, J. D.; Yan, J. H.; Yu, Z. Y., and Peng, S. L. Hydrological impacts of reafforestation with eucalypts and indigenous species: a case study in Southern China. *Forest Ecology and Management*, 167(1–3):209–222, 2002.

Zhou, X.; Zhang, Y.; Wang, Y.; Zhang, H.; Vaze, J.; Zhang, L.; Yang, Y., and Zhou, Y. Benchmarking global land surface models against the observed mean annual runoff from 150 large basins. *Journal of Hydrology*, 470–471:269–279, 2012.

Ziegler, A. D.; Giambelluca, T. W.; Tran, L. T.; Vana, T. T.; Nullet, M. A.; Fox, J.; Duc Vien, T.; Pinthong, J.; Maxwell, J. F., and Evett, S. Hydrological consequences of landscape fragmentation in mountainous northern Vietnam: evidence of accelerated overland flow generation. *Journal of Hydrology*, 287 (1–4):124–146, 2004.

Zimmermann, B.; Elsenbeer, H., and De Moraes, J. M. The influence of land-use changes on soil hydraulic properties: Implications for runoff generation. *Forest Ecology and Management*, 222(1–3):29–38, 2006.

Zimmermann, B.; Papritz, A., and Elsenbeer, H. Asymmetric response to disturbance and recovery: Changes of soil permeability under forest-pasture-forest transitions. *Geoderma*, 159 (1–2):209–215, 2010.

Zou, X. and Gonzalez, G. Changes in earthworm density and community structure during secondary succession in abandoned tropical pastures. *Soil Biology Biochemistry*, 29(3–4):627–629, 1997.