Spatiotemporal Variability of Hydrologic Response: An Entropy-based Approach Using a Distributed Hydrologic Model

by

Aldrich Edra Castillo

B.S. Chem. E., University of the Philippines (2002)
M.S. Sys. E., University of Virginia (2008)

Submitted to the Department of Civil and Environmental Engineering
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in the Field of Hydrology

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2014

Signature redacted

Author .....
Department of Civil and Environmental Engineering
Signature redacted May 2, 2014

Certified by ....
Dara Entekhabi
Bacardi and Stockholm Water Foundations Professor of Civil and Environmental Engineering and Earth, Atmospheric, and Planetary Sciences
Thesis Supervisor

Accepted by ..... Signature redacted
Heidi M. Nepf
Chair, Departmental Committee for Graduate Students

Spatiotemporal Variability of Hydrologic Response: An Entropy-based Approach Using a Distributed Hydrologic Model

by

Aldrich Edra Castillo

Submitted to the Department of Civil and Environmental Engineering
on May 2, 2014, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in the Field of Hydrology

Abstract

Basin hydrologic response pertains to the partitioning of precipitation into streamflow, evapotranspiration, and change in storage. The ability to explain or predict the response has many applications e.g. flood forecasting, water budget studies, and design of hydrological observing systems. However, explaining the response is challenging because it is the combined manifestation of many complex and interrelated factors that naturally vary in space and time, and act over a variety of scales.

A possible key is better understanding of the space-time dynamics of the hydrologic state variable – the soil moisture field. This thesis uses the distributed hydrologic model MOBIDIC that uses a single soil layer with dual compartments: a capillary and a gravity reservoir composed of small, and large pores, respectively. Mass and energy fluxes are simultaneously solved using simple linear equations. These make the model computationally efficient. To improve soil moisture simulations, some model modifications were introduced. MOBIDICs ability to simulate the magnitude range and dynamics of soil moisture at the local scale is found comparable with a benchmark model that uses non-linear soil physics relations.

We derive an entropy-based dimensionless measure of hydrologic complexity $\mathcal{H}$ which measures the distance of a given soil moisture spatial probability distribution from two limiting cases. Using 8 test basins with area of 10^0–10^3 km^2 and representing semiarid, temperate, and humid climates, it is shown that $\mathcal{H}$ effectively tracks the evolution of soil moisture distribution, and captures the interplay between vertical and lateral fluxes. Furthermore, we investigate the relationship of $\mathcal{H}$ with observable basin attributes and traditional measures of hydrologic response. Clear and logical relationships emerge only after grouping basins based on similarity. For example, in the semiarid basins, $\mathcal{H}$ increases with catchment area, infiltration ratio and baseflow index. For basins of similar size, $\mathcal{H}$ is highest in temperate climate, consistent with soil moisture being double-bounded so its variability peaks at intermediate conditions. Finally, although not explicitly coded in MOBIDIC, hysteresis is evident in

the discharge-storage plots. It emerges from the use of a dual-pore soil structure that captures the threshold behavior of runoff. $\mathcal{H}$ helps in understanding the mechanisms involved.

Thesis Supervisor: Dara Entekhabi
Title: Bacardi and Stockholm Water Foundations Professor of Civil and Environmental Engineering and Earth, Atmospheric, and Planetary Sciences

Acknowledgments

First and foremost, I would like to thank my adviser, Professor Dara Entekhabi, for his wisdom, guidance, and patience throughout my stay here at MIT. I also would like to thank Professors Fabio Castelli for sharing his MOBIDIC model and more importantly his valuable insights and time. I am very grateful to Professor Dennis McLaughlin and Guido Salvucci for being part of my thesis committee and sharing their valuable time and wisdom.

This thesis would have not been possible without the financial support from MIT, the Center for Environmental Sensing and Modeling (CENSAM) of the Singapore-MIT Alliance for Research and Technology (SMART), the National Aeronautics and Space Administration Earth Science Technology Office (NASA ESTO), and JV Angeles Construction Corporation.

I would to thank my parents and brothers Mandrick, Lester, and Ramir; Mr. and Mrs. Angeles and their children Tricie, Jon, and Justin; the Beta Epsilon Fraternity Alumni Association in America esp. Ronne, Etao, Greg, and Pete; and the awesome people at Parsons esp. Siggi, Ben, Peter, Alex, Sheila, Vicki, Kris, and Jim. To my colleagues, friends, and family, here in the US, in the Philippines, in China, and in Singapore, thank you very much.

Lastly, I would like to thank Lijun — the classmate and colleague I met here in Parsons Lab during my first year, and who is now my lovely wife. Thanks for all the understanding, patience, and inspiration.

THIS PAGE INTENTIONALLY LEFT BLANK

I dedicate this thesis to

my parents

&

Lijun

THIS PAGE INTENTIONALLY LEFT BLANK

Contents

List of Figures **13**

List of Tables **19**

1 Introduction **21**

1.1 Motivation 21

1.2 Review of Related Literature 21

1.3 The Need for Organizing Principles 24

1.4 Thesis Outline 25

2 The Hydrologic Model MOBIDIC **29**

2.1 Introduction 29

2.2 Overview of MOBIDIC 31

2.3 Mass and Energy Balance 34

2.4 Model Modifications: Processes 37

2.4.1 Canopy Interception and Surface Storage 37

2.4.2 Evapotranspiration 39

2.4.3 Percolation and Lateral Subsurface Flow 40

2.4.4 Capillary Rise 43

2.5 Basin-scale Mass Balance 46

2.6 Model Modifications: Numerical Efficiency 46

2.7 Model Implementation 48

2.7.1 Collection and Preparation of Model Inputs 49

2.7.2 Calibration and Validation 50
2.8 Summary 52

3 Simulations of Local Scale Soil Moisture in a Catchment Hydrologic Model 55

3.1 Abstract 55
3.2 Introduction 56
3.3 The SHAW Model 60
3.4 Correspondence Between SHAW and MOBIDIC Variables 61
3.5 Test Sites 63
3.6 Calibration 65
3.7 Results and Discussion 65
3.7.1 Site 1 — Lucky Hills, Arizona 65
3.7.2 Site 2 — Mayday, Mississippi 68
3.8 Summary 72

4 An Entropy-based Measure of Hydrologic Complexity 75

4.1 Motivation 75
4.2 The Hydrologic Complexity Index of *Martina and Entekhabi* (2006) . 78
4.3 Development of a Revised Measure of Hydrologic Complexity 81
4.3.1 Shannon Entropy 81
4.3.2 Differential Entropy 84
4.3.3 Kullback-Liebler Divergence 85
4.3.4 Definition of the Revised Measure of Complexity 88
4.4 Application of the Measure 92
4.4.1 Example 1 93
4.4.2 Example 2 95
4.5 Summary 97

5 Understanding the Hydrology of a Tropical Rainforest Basin 99

5.1 Abstract 99

5.2 Introduction 100
5.3 Methods and Data 102
5.3.1 Study Area 102
5.3.2 Meteorological Forcings 104
5.3.3 Groundwater Interaction 106
5.3.4 Calibration 107
5.3.5 Measure of Hydrologic Complexity 107
5.4 Results and Discussion 109
5.4.1 Water Balance 109
5.4.2 Hydrologic Fluxes and States 110
5.4.3 Measure of Hydrologic Complexity 114
5.5 Summary 121

6 Spatiotemporal Variability of Hydrologic Response: an Entropy-based Approach 123

6.1 Abstract 123
6.2 Introduction 124
6.3 Methods 128
6.3.1 Measure of Hydrologic Complexity 128
6.3.2 Test Basins 129
6.3.3 Relationship of $\mathcal{H}$ to Other Hydrologic Variables 130
6.4 Results and Discussion 131
6.4.1 Basin-scale Partitioning of Precipitation 131
6.4.2 Hydrologic Complexity 132
6.4.3 Hysteresis and Threshold Behaviors 137
6.5 Summary and Conclusions 140

7 Summary, Major Contributions, and Recommendations 149

7.1 Summary 149
7.2 Major Contributions 153
7.3 Recommendations for Future Research 154

List of Acronyms 157

List of Symbols 159

List of Figures

2.1 The physical system described by MOBIDIC for a typical pixel; P is precipitation, T_a air temperature, U wind speed, RH relative humidity, S_t total solar radiation, T soil temperature, and θ soil moisture . . . 32

2.2 Schematic diagram of the mass balance for a typical pixel of MOBIDIC prior to model modifications 34

2.3 The surface energy balance of MOBIDIC at each computational unit 36

2.4 A quick reference of the mass balance of the modified MOBIDIC for a typical pixel 38

2.5 ET efficiency [-] vs. effective soil saturation S [-] using 'ET3 Model 1' 40

2.6 Comparison of the normalized percolation rate as a function of effective soil saturation computed using MOBIDIC's approximation and the analytic solution of *Eagleson* (1978), for 3 soil textures 41

2.7 Illustration of the cases when the position of the water table is inconsistent with the content of the soil gravity reservoir 42

2.8 Possible water table positions relative to the modeled soil layer 45

2.9 Overall water balance of a basin considering the system is a) the modeled soil layer; b) the modeled soil layer plus the groundwater aquifer 47

2.10 Procedure used to improve the numerical efficiency of MOBIDIC . . . 47

2.11 Data inputs of MOBIDIC 49

2.12 Example of how ArcGIS™Model Builder was used to automate the processing of topographic and geomorphologic data inputs 50

3.1 The physical system described by the 1–D SHAW model; P is precipitation, T_a is air temperature, u is wind speed, RH is relative humidity, S_t is total solar radiation, T is soil temperature, and θ is soil moisture 61

3.2 Observed vs. SHAW-simulated soil moisture at Site 1 66

3.3 a.) Observed precipitation [mm/day] and MOBIDIC-simulated ET [mm/day]; b.) observed equivalent depth [cm] of soil water stored in the top 50 cm vs. corresponding values simulated by SHAW and MOBIDIC for Site 1 66

3.4 The SHAW- and MOBIDIC-simulated equivalent depth [cm] of water stored in the soil a.) capillary reservoir; and b.) gravity reservoir, for the top 50 cm of Site 1 67

3.5 Temperature at diurnal damping depth of Site 1 68

3.6 Observed vs. SHAW-simulated soil moisture at Site 2 70

3.7 For Site 2: a.) observed precipitation [mm/day] and MOBIDIC-simulated ET [mm/day]; b.) observed, SHAW-, and MOBIDIC-simulated equivalent depth [mm] of soil water stored in the top 50 cm; and c.) MOBIDIC- and SHAW-simulated equivalent depth [mm] of water in the capillary reservoir of the top 50 cm of soil 71

3.8 Observed temperature at diurnal damping depth vs. corresponding values simulated by SHAW and MOBIDIC for Site 2 71

4.1 Conceptual diagram of the limiting cases: the soil water deficit is left) the same across the basin; and right) uniformly distributed from zero to a maximum value. Bottom plots show the PDF of $\forall_e$. 77

4.2 Dependence of left) the Shannon entropy H; and right) the index of hydrologic complexity $\mathcal{H}_{old}$, on discretization ($N = 1/\Delta x$). The inset shows the PDF of the relative available soil storage 81

4.3 The revised dimensionless hydrologic complexity index $\mathcal{H}$ is independent of numerical discretization (N = number of bins). The insets show the PDFs used. 90

4.4 Effect of using unreasonable number of bins N on $\mathcal{H}$ for peaky distributions. Inset shows the PDF of the 10^6 discrete values of X. 91

4.5 Time series of the hydrologic complexity of the basin, the capillary reservoir, and the gravity reservoir. Also shown are three snapshots showing the spatial PDF of relative soil water deficit in the dry, intermediate, and wet regimes. 94

4.6 The hydrologic complexity index of the soil capillary reservoirs, soil gravity reservoirs, and the (combined) soil reservoirs, vs. the relative soil water deficit. Inset shows hysteresis. Red and blue arrows indicate drying and wetting directions, respectively. 94

4.7 Time series of the hydrologic complexity of the basin, the capillary reservoir, and the gravity reservoir. Also shown are three snapshots showing the spatial PDF of relative soil water deficit in the dry, intermediate, and wet regimes. 96

4.8 The hydrologic complexity index of the soil capillary reservoirs, soil gravity reservoirs, and the (combined) soil reservoirs, vs. the relative soil water deficit. Inset shows hysteresis. Red and blue arrows indicate drying and wetting directions, respectively. 96

5.1 Maps of the Upper Tamugan River Basin showing a) the SRTM DEM, the basin boundary and river network derived from the DEM, and the location of the staff gauge and nearby rain gauges (in parenthesis are mean annual rainfall); b) its location (red dot) in southern Philippines; and c) its location relative to the sea and a weather station at the regional airport 103

5.2 The Tamugan basin is covered by 68% forest, 26% brushlands and woodlands, and 6% farmlands and human settlements 103

5.3 Distributed parameters: a) albedo [-]; b) turbulent heat exchange coefficient [-]; c) soil depth [mm]; d) soil saturated hydraulic conductivity [mm/hr]; e) capacity of soil capillary reservoir [m^3/m^3]; f) capacity of soil gravity reservoir [m^3/m^3]; g) capacity of plant canopy [mm]; and h) capacity of surface storage [mm] 103

5.4 a) Inter-annual variability of the TRMM rainfall; b) diurnal pattern of TRMM rainfall at 3-hour (data) and 1-hour (interpolated); c) annual mean rainfall fields at 90 m resolution 105

5.5 Net solar radiation generated by multiplying the analytic values for clear-sky with cloud factors based on qualitative description of conditions at the Davao City Airport, matches well with direct measurements on April 21–29, 2013. 107

5.6 Time series for WY 2009–2012 (left column focuses on Oct. & Nov. 2011) of daily-aggregated and basin-averaged a) rainfall, P [mm/d]; b) modeled vs. observed depth-equivalent streamflow [mm/d]; c) potential and actual ET, PET and ET [mm/d], and percolation, Q_{perc} [5 mm/d]; and d) depth-equivalent of water stored in the soil capillary reservoir, W_c [mm]; soil gravity reservoir, W_g [mm]; plant reservoir, W_p [0.02 mm]; and surface reservoir, W_s [0.02 mm], for 111

5.7 Groundwater table [m.a.s.l.] simulated for WY 2009–2012: a) time series of the spatial mean; b) net change [m] at each pixel; and c) temporal mean at each pixel 112

5.8 Maps and probability density function of temporal mean saturation [%] of the a) soil capillary reservoirs; b) soil gravity reservoirs; and c) plant/canopy reservoirs 113

5.9 Maps of simulated mean annual a) ET; b) net runoff generation; and c) percolation to deeper soil layers 114

5.10 Base flow and quick flow simulated for WY 2011–2012 115

5.11 Runoff (quick flow) depth vs. rainfall depth for each storm event. Trendline is the integral of an incomplete gamma function. 115

5.12 Time series of a) the percentiles of water deficit for WY 2011 (normal year); and b) the deviations from the mean of each percentile relative to WY 2011 115

5.13 Dynamics of the hydrologic complexity index $\mathcal{H}$ for WY 2010–2011. . 116

5.14 Measure of hydrologic complexity vs. basin-averaged water deficit . . 118

5.15 Macro-hysteresis in the evolution of the probability distribution of soil water deficit $\forall_e$ during a drying phase (top) and a wetting phase (bottom) 120

6.1 The location of the test basins: top) nested semiarid basins in California; middle) temperate basins in Oklahoma-Arkansas; and bottom) humid basin in the Philippines 142

6.2 Basin-scale precipitation partitioning with basins grouped by catchment area and arranged in increasing median slope 142

6.3 Map and PDF of characteristic soil water deficit, $x = \forall_e/\forall_{e,max}$ [-] of basins S1 (left) and T2 (right). Orange dots mark basin outlets. . . . 144

6.4 Time series of $\mathcal{H}$, $\mathcal{H}_c$, and $\mathcal{H}_g$ for basins S2 (top), T2 (middle), and H2 (bottom). On the right side are the 3-year median values. Days marked on basin H2 are representative of dry (1), intermediate (2), and wet (3) conditions 144

6.5 Example spatial PDF of relative soil water deficit at basin H2 in dry, intermediate, and wet conditions. 144

6.6 $\mathcal{H}$, $\mathcal{H}_c$, and $\mathcal{H}_g$ vs. basin-averaged total soil water deficit x for basins S2 (top), T2 (middle), and H2 (bottom). Red and blue arrows indicate drying and wetting directions for $\mathcal{H}$, respectively 144

6.7 From top: $\mathcal{H}$, $\mathcal{H}_c$, and $\mathcal{H}_g$, vs. aridity [-]. Points are temporal mean and whiskers are range; circle (humid); squares (temperate); triangles (semiarid). 144

6.8 Temporal mean of (left to right) $\mathcal{H}$, $\mathcal{H}_c$, and $\mathcal{H}_g$, vs. spatial scales: top) catchment area $\log_{10}A$ [km^2], and bottom) basin-scale topographic wetness index $\log_{10}(A/s_{50})$. Trendlines are for semiarid (red) and temperate (black) basins. 144

6.9 Temporal mean of (left to right) $\mathcal{H}$, $\mathcal{H}_c$, and $\mathcal{H}_g$, vs. relief ratio r_r [-]. Trendlines are for semiarid (red) and temperate (black) basins. 144

6.10 Temporal mean of (left to right) $\mathcal{H}$, $\mathcal{H}_c$, and $\mathcal{H}_g$, vs. infiltration ratio i_s/K_s [-]. Trendlines are for semiarid (red) and temperate (black) basins. 144

6.11 Temporal mean of (left to right) $\mathcal{H}$, $\mathcal{H}_c$, and $\mathcal{H}_g$, vs. (top to bottom) ET efficiency, runoff ratio, and BFI. Red and black trendlines are for semiarid and temperate basins, respectively. 145

6.12 Hysteresis in the evolution of the probability distribution of soil moisture deficit $\forall_e$ during a drying phase (top) and a wetting phase (bottom) at basin H2 146

6.13 Hysteresis in simulated daily discharge vs. mean relative soil water deficit $\overline{x}$ (left), and vs. $\mathcal{H}$ (right) for basins H2 and S2 in dry conditions, and T2 and S2 in wet conditions. Red and blue arrows indicate drying and wetting, respectively. 147

List of Tables

1.1 Comparison of previous studies that investigated the use of multiple predictors of hydrologic response through hydrograph analysis 23

2.1 Comparison of MOBIDIC with other hydrologic models 33

2.2 Lookup table of soil parameters for the capillary rise module of MOBIDIC. Source: *Rawls et al.* (1982) 44

3.1 Calibrated soil properties of the SHAW model of Site 1. b and ψ_e are the Campbell pore-size distribution index, and air-entry potential, respectively; D_1 and D_2, are the diffusion, and dispersion parameters, respectively 64

3.2 Performance of the SHAW and MOBIDIC models of Site 1 for the calibration period (year 2007 & 2008) and validation period (year 2009) 66

3.3 Performance of the SHAW and MOBIDIC models of Site 2 for the calibration period (water year 2006 & 2007) and validation period (water year 2008) 69

5.1 Properties of storm events derived from TRMM data for WY1999–2012 105

5.2 Multipliers of clear-sky solar radiation assigned to qualitative descriptions of weather condition at the Davao City Airport 106

6.1 Description of the test basins 142

6.2 Properties of the test basins 143

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 1

Introduction

1.1 Motivation

The hydrologic response (HR) of a basin pertains to how precipitation is partitioned into streamflow, evapotranspiration (ET), and change in storage. Streamflow can be further partitioned into quickflow and baseflow; ET into evaporation and plant transpiration; and storage into water in various reservoirs such as surface water, soil moisture, and groundwater. As one of the main goals of hydrology, the ability to explain the response or the spatiotemporal variability of the above mentioned hydrologic fluxes and states has many important applications from flood forecasting and risk management, to water budget studies, and to the design of efficient systems for observing or modeling hydrologic variables. However, it is difficult to explain much less predict the response because it is the combined manifestation of many complex and interrelated factors that naturally vary both in space and time, and act over a variety of scales.

1.2 Review of Related Literature

Many studies considered only the streamflow component of the hydrologic response and how it is influenced by a single factor or class of factors such as routing processes (*Dunne and Black*, 1970), catchment shape (*Viessmann et al.*, 1977), topographic

wetness index (*Beven and Kirby*, 1979, 1997), geomorphology (*K.Gupta and Mesa*, 1988; *Rodriguez-Iturbe and Rinaldo*, 1997; *Lazzaro*, 2008), and storm characteristics (*Vivoni et al.*, 2007). Other studies took a more holistic approach and investigated the combined effects of various basin attributes then estimated individual effects using techniques such as principal component analysis. Some of these studies are summarized and compared in Table 1.1.The majority used stream gauge measurements while *Berger and Entekhabi* (2001) used numerically-simulated streamflow, and (*Cerdan et al.*, 2004; *Buttle and McDonald*, 2002; *Buttle et al.*, 2004) used both observed and simulated streamflow. Over the years the approach generally progressed in terms of the number and coverage of basins used, the number and types of predictors and measures of hydrologic response considered, and the complexity of numerical models employed. However, the findings of these studies are inconsistent. For instance, the interbasin variability of runoff ratio (streamflow Q divided by precipitation P) was found to be controlled mainly by physiography(*Zecharias and Brutsaert*, 1988; *Sefton and Howarth*, 1998), climate (*Merz et al.*, 2006; *van Djik*, 2010), land use (*Cerdan et al.*, 2004), preferential flow paths and soil depth (*Buttle and McDonald*, 2002; *Buttle et al.*, 2004); or a combination of these attributes (*Berger and Entekhabi*, 2001; *Sankarasubramanian and Vogel*, 2002).

Some studies recognize that basin response and hydrologic fluxes are functions of distributed hydrologic states, most notably of soil moisture which controls the partitioning of rainfall into infiltration and runoff, and also controls land surface temperature through its effect on the partitioning of available energy into sensible and latent heat fluxes. Based on a rich literature, the spatial distribution of soil moisture is influenced mainly by the spatial variability of topography (*Yeh and Eltahir*, 1998; *Liu et al.*, 2012); soil properties (*Western and Grayson*, 2000; *Kim et al.*, 2002; *Famiglietti et al.*, 1999, 2008); macropore and preferential flow paths (*Brooks et al.*, 2009; *Beven and Germann*, 2013) groundwater-surface water interaction (*Levine and Salvucci*, 1999; *Kollet and Maxwell*, 2008); vegetation (*Eagleson*, 1978; *Ivanov et al.*, 2010); and meteorological forcings e.g. precipitation (*Entin et al.*, 2000; *Western et al.*, 2003; *Vivoni et al.*, 2010). Moreover, soil moisture is dynamic due to the intermittency,

Table 1.1: Comparison of previous studies that investigated the use of multiple predictors of hydrologic response through hydrograph analysis

Reference	*Zecharias & Brutsaert* (1988)	*Sefton & Howarth* (1998)	*Berger & Entekhabi* (2001)	*Sankarasubramanian and Vogel* (2002)
# of basins	19	60	10	1305
Area	8–180 km^2	9–900 km^2	10–325 km^2	> 50 km^2
Location	Appalachians, US	England and Wales	Continental US	Continental US
Predictors of HR	8 topography and geomorphology	29 physiography, climate and land cover	8 physiography & climate	4 physiography & climate
Measures of HR	baseflow Q_{bf}	streamflow Q	Q/P, ET/E_p, θ, extents of G.W. zones	runoff ratio Q/P
Streamflow data	observed	observed	modeled	observed
Hydrologic model	none	lumped and semi-physics-based	semi-distributed equilibrium model	none
Findings	98% of Q_{bf} variance = $f(A$, relief, total length of perennial streams)	63% of Q variance = f(elevation, TWI, stream freq., A	87% of Q/P variance = $f(P/E_p$, slope, stream density, infil. cap.	71% of Q/P variance = $f(P/E_p$, slope, stream density, infil. cap.
Reference	*Cerdan* (2004)	*Buttle et al.* (2002, 2004)	*Merz et al.* (2006)	*van Djik* (2010)
# of basins	$3 + 40$	1	337	183
Area	plot – 10 km^2	3.2 ha	80-10,000 km^2	51–1780 km^2
Location	France	Canadian Shield	Austria	Australia
Predictors of HR	physiography, land use	topography, soils, and plants	physiography, climate and land cover/use	physiography, climate and land cover
Measures of HR	streamflow Q	$Q/P, Q, Q_{bf}$, GW depth	streamflow Q	Q, Q_{bf}
Streamflow data	observed & modeled	observed & modeled	observed	observed
Hydrologic model	distributed conceptual model	semi-distributed physics-based model	none	simple linear-reservoir
Findings	R/P decreases as A and % of arable land increase	dominant controls are preferential flow paths and soil depth	Q/P is controlled by mean P and little by soil type or land cover	Q and Q_{bf} are controlled by E_p and monthly P

seasonality, and inter-annual variability of meteorological forcings (*Entekhabi et al.*, 1995; *Koster and Suarez*, 1999). Both the spatial mean and variance of soil moisture, i.e. its center of mass and distributional features, evolve over time.

1.3 The Need for Organizing Principles

The findings and conclusions of the various studies mentioned are inconsistent or confusing to say the least. *Beven* (2006a) and *Ebel and Loague* (2006) explain that because of uniqueness of place, the dominant processes and threshold levels can vary significantly from basin to basin, thus resulting to unique hydrologic signatures. For instance, the use of topographic indices has performed well in some cases but poorly in others (*Western et al.*, 1999). A possible key might come with better understanding and model representation of soil moisture. In this light, *Grayson et al.* (1997) argued that since soil is bi-modal with either local or nonlocal controls, realistic representation of hydrologic response cannot be achieved through the use of a single index. Using soil moisture measurements at three adjacent hillslopes on a steep alpine terrain in Italy, *Penna et al.* (2009) found that flow-related indices explain up to 42% of the spatial variation of soil moisture, while radiation-related indices are not as important.

Some studies propose the use of similarity analysis or classification schemes. For instance, based on catchment area, the hydrologic response of small basins is primarily controlled by topography and soils (*Robinson et al.*, 1995) or vegetation (*Bisht*, 2010); while that of large basins is primarily controlled by climatic properties (*Budyko*, 1974; *Vivoni et al.*, 2007; *Nicotina et al.*, 2008) or geomorphology (*Robinson et al.*, 1995). For soil moisture, the dominant controls depend on both spatial and temporal scales (*Grayson and Wetern*, 1998; *Sköien et al.*, 2003; *Wilson et al.*, 2004; *Vivoni et al.*, 2007; *Famiglietti et al.*, 2008). Adding another dimension, based on analysis of hillslope field studies in different parts of the world *Lawrence and Hornberger* (2007) found that a classification based on climate regime (e.g. dry, intermediate, and humid) explains the differences in soil moisture variability patterns. The magnitude of spatial variance of soil moisture is controlled by the wilting point at dry states, by

the hydraulic conductivity at intermediate states, and by the porosity at wet states. In semiarid regions, the spatial variance of soil moisture generally increases with the mean soil moisture (*Fernandez and Ceballos*, 2003). In humid regions, the variance generally decreases as the mean soil moisture increases (*Meyles et al.*, 2003; *Teuling and Troch*, 2005). In temperature regions, the variance peaks at intermediate soil moisture (*Famiglietti et al.*, 1999, 2008; *Lawrence and Hornberger*, 2007).

This thesis recognizes the complexity and multi-dimensionality of hydrologic response. Perhaps, instead of searching for universal relationships, the focus should be on developing a classification scheme or a multi-dimensional chart (e.g. analogous to the Periodic Table of Elements in Chemistry or the Linnaean Taxonomy in Biology as suggested by *Dooge* (1986) and *McDonnell et al.* (2007a)) that has multiple dimensions such as space, time, and climate. In particular, this thesis tests whether grouping basins according to climate and catchment area leads to a clearer understanding of the spatiotemporal variability of hydrologic response. Moreover, this thesis explores the use of the concept of entropy as an organizing principle, and a possible link between a basin's micro-states e.g. distributed soil moisture, to its macro-states e.g. streamflow. The test basins represent different spatial scales and climates, and where applicable, multiple basin attributes and measures of hydrologic response are used.

1.4 Thesis Outline

Chapter 2 describes the hydrologic model MOBIDIC which is the main tool used in this research. The model will be compared with a few other hydrologic models to point out its unique and key features. Next, the mass and energy balance of the model, and the model modifications made for this thesis in order to improve soil moisture simulations, will be discussed. Also discussed is the general procedure for development of basin-specific MOBIDIC models, from data collection to model setup and calibration.

Chapter 3 tests the ability of the improved MOBIDIC in simulating soil moisture

at the *local scale* e.g. at a plan element. MOBIDIC-simulated soil moisture will be compared with observations as well as with values simulated by a benchmark model that uses non-linear soil physics relations for flow in the unsaturated zone. The comparison is performed at two sites with contrasting climates.

Chapter 4 presents an entropy-based dimensionless index of hydrologic complexity $\mathcal{H}$ which measures the relative distance of a given distribution of soil moisture from two limiting distributions. $\mathcal{H}$ will be derived starting from the *Martina and Entekhabi* (2006) index of hydrologic complexity. By using the concepts of differential entropy and relative entropy, $\mathcal{H}$ is made discretization-invariant. The chapter concludes with example applications of $\mathcal{H}$ to demonstrate how it can be used to better understand the spatiotemporal variability of basin response and hydrologic fluxes in general, and to answer specific science questions.

In Chapter 5, MOBIDIC and $\mathcal{H}$ are used to understand the hydrology of a rain forest covered *river basin* in the Philippines. This chapter demonstrates in details how a MOBIDIC model is developed for a specific basin, and it also illustrates the kinds of hydrologic information and insights that the model can provide. Moreover, this chapter investigates the dynamics of $\mathcal{H}$ and how it relates to the evolution of the overall state of the basin and whether $\mathcal{H}$ captures the switching of dominant hydrologic processes.

Having demonstrated that MOBIDIC can successfully simulate the dynamics of soil moisture at the local scale (Chapter 2) and realistically capture the spatiotemporal variability of the hydrologic fluxes and states of a river basin (Chapter 5), Chapter 6 uses MOBIDIC to model *multiple basins* with area of 10^0–10^3 km^2 and representing semiarid, temperate and humid climates. $\mathcal{H}$ is computed on the simulated variables to answer the following science questions:

1. How do distributional features of soil water deficit evolve over time?
2. Is there hysteresis in the evolution?
3. Can the entropy-based measure of complexity $\mathcal{H}$ effectively track the evolution?
4. Can $\mathcal{H}$ be used to identify dominant hydrologic processes?

5. What factors (physiography, spatial scale, climate, etc.) affect the distribution and its evolution?

6. How are the distributional features related to basin response and hydrologic fluxes?

The final chapter summarizes the key findings and contributions of this thesis. Some future research directions are also suggested.

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 2

The Hydrologic Model MOBIDIC

This chapter describes the hydrologic model ***MO**dello **B**ilancio **I**drologico **DI**stributo e **C**ontinuo* (MOBIDIC) which is the main tool used in this research. The first section gives an overview and general description of the model. This is followed by a detailed description of the mass and energy balance. The next section discusses the key modifications made in order to improve the representation of some hydrologic processes and also significantly improve the numerical efficiency of the model. Finally, the general procedure for applying MOBIDIC on specific basins, from data collection to model validation, is outlined.

2.1 Introduction

More and more measurements of streamflow, soil moisture and temperature, piezometric head, latent heat flux, and other hydrologic variables, are being collected worldwide. These measurements are obtained through various methods such as monitoring stations, distributed networks of sensors, field campaigns, and remote sensing. The suitability of a certain method depends largely on the desired scale and resolution, both in space and in time. Unfortunately, as discussed in the introductory chapter, all observing methods or systems have shortcomings.

Numerical models can complement hydrologic observations. Incoming radiation and precipitation are used in conjunction with water and energy balance models to

simulate the evolution of state variables such as soil moisture and determine the water and energy fluxes across the landscape. *Harter and Hopmans* (2004) describes how hydrologic models have traditionally been used by two largely disconnected groups: the watershed hydrologists who deal with macro-processes; and the soil physicists who study soil properties and states at the laboratory or local scales. Watershed hydrologists typically use lumped or semi-distributed models such as TOPMODEL (*Beven and Kirby*, 1979) and SAC-SMA (*Burnash et al.*, 1973), which have been demonstrated as highly capable in simulating streamflow. The computational timestep is usually hourly, daily, or longer. The main shorthcomings of these models are that i) they do not adequately account the spatial variability in landuse, topography and other factors; and ii) they provide little or no information on the spatial variability of hydrologic states and fluxes within the basin. On the other hand, soil physicists who have detailed measurements of soil properties and states at the local to plot scales, model unsaturated flow by discretizing the hydrologically active soil column into several layers and using the nonlinear Richards equation (*Richards*, 1931),

$$\frac{\partial \theta}{\partial t} = -\frac{\partial}{\partial z}\left[K(\theta)\left(\frac{\partial \psi}{\partial z} + 1\right)\right] \tag{2.1}$$

where, K is the hydraulic conductivity, ψ pressure head, z elevation with respect to a datum, θ soil moisture, and t time (*Milly*, 1988). For numerical stability, the above non-linear differential equation is solved using sub-hourly time steps.

Over the years, the modeling efforts of the two disciplines have started to converge. The significant advances in computing, geographic information system (GIS), and the increasing availability of remotely-sensed hydrologic datasets, have made feasible the use of physics-based distributed hydrologic models (DHMs). These models discretize the landscape in computational elements that are $10^1 - 10^3$ m in the horizontal. Adopting the practice in soil physics, many DHMs employ Richards equation and discretize the hydrologically active soil layer into vertical layers that are $10^{-3} - 10^{-1}$ m thick. Some DHMs that use the Richards formulation are MIKE-SHE (*Refshaard and Storm*, 1995) and ParFlow (*Ashby and Falgout*, 1996) that use grids for hori-

zontal discretization; and PIHM (*Qu and Duffy*, 2007) and TRIBS (*Ivanov et al.*, 2004) that use triangulated irregular network (TIN) as horizontal elements, see Table 2.1. Some simpler models such as the SAC-SMA (*Burnash et al.*, 1973) represent the soil as having dual compartments — saturated and unsaturated. More DHMs are discussed by (*Smith et al.*, 2004, 2012) under the context of the Distributed Model Intercomparison Projects 1 and 2. There are studies that demonstrate the advantages of DHMs over lumped and semi-distributed model e.g. *Bartholomes and Todini* (2005); *Castelli et al.* (2009); *Smith et al.* (2004); *Vieux et al.* (2004). Although promising, the use of DHMs has its own challenges and criticisms including i) the need for a high number of inputs, often at fine spatiotemporal resolutions; ii) the use of many parameters (*Luzio and Arnold*, 2004; *Duan et al.*, 2006) which makes the calibration process tedious and raises the concern on equifinality *Beven* (2006a); and iii) the high computational requirements (*Smith et al.*, 2004, 2012). Hopefully, some of these challenges will be overcome as better DHMs are developed and as the use of DHMs mature, in general.

2.2 Overview of MOBIDIC

The main tool used in this research is the Modello Bilancio Idrologico DIstributo e Continuo (MOBIDIC). It is a physics-based and distributed catchment hydrologic model that simultaneously solves mass and energy balance. It was originally developed by Fabio Castelli at the University of Florence, Italy. Some modifications were introduced in this thesis in order to make the representation of some processes more realistic, especially to improve the ability of the model to simulate soil moisture. Changes were also made to improve the numerical efficiency of the model. Previous versions of MOBIDIC are described in *Castelli et al.* (2006, 2009).

MOBIDIC represents the hydrologic cycle as a system of reservoirs and the fluxes of mass and energy between these reservoirs, see Figure 2.1. In order to account for the different roles of gravity and capillary forces in moving and storing soil water, each pixel (i.e. raster cell) of soil is divided into a gravitational compartment composed of

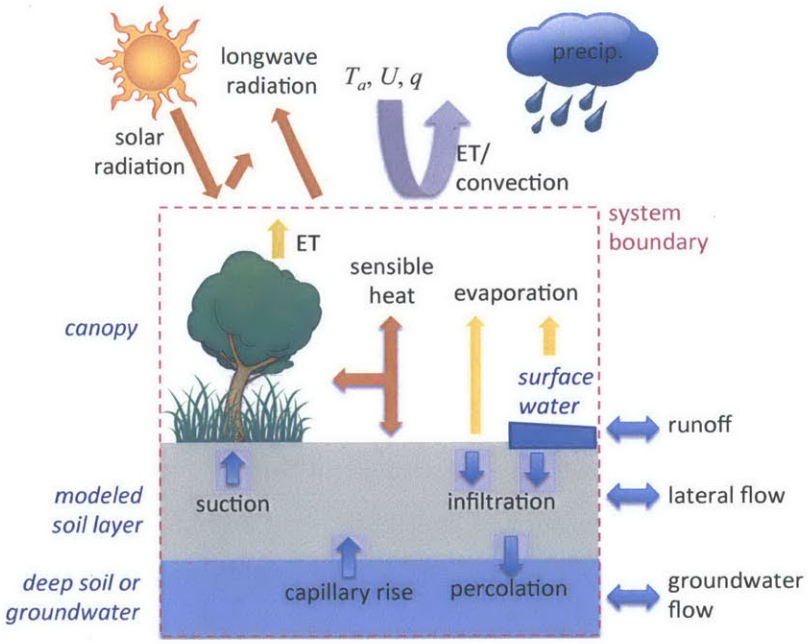


Figure 2.1: The physical system described by MOBIDIC for a typical pixel; P is precipitation, T_a air temperature, U wind speed, RH relative humidity, S_t total solar radiation, T soil temperature, and θ soil moisture

large pores, and a capillary compartment composed of small pores. For computational parsimony, the soil moisture flux in the unsaturated zone is solved empirically instead of using Richards equation.

MOBIDIC is written in Matlab with about a hundred subroutines. Topographic and geomorphologic model inputs are processed in ArcGIS™. Groundwater dynamics is modeled using either a linear reservoir model or a Laplace-type solver. Channel routing can be done by the lag, linear, Muskingum, or Muskingum-Cunge method. Surface reservoirs such as lakes and dams can be considered by specifying control rules. Table 2.1 compares MOBIDIC with other DHMs.

2.3 Mass and Energy Balance

A schematic diagram of MOBIDIC's mass balance at a typical pixel, prior to the modifications introduced by the author, is shown in Figure 2.2. This is simply water balance because although MOBIDIC has a subroutine for contaminant fate and

Table 2.1: Comparison of MOBIDIC with other hydrologic models

Name	TOPMODEL	**MOBIDIC**	MIKE-SHE	PIHM	TRIBS	ParFLow
Reference	*Beven & Kirby* (1979)	*Castelli et al.* (2009)	*Refshaard* (1995)	*Qu & Duffy* (2007)	*Ivanov et al.* (2004)	*Ashby & Falgout* (1996)
Distributed	semi	yes	yes	yes	yes	yes
Energy balance	no	yes	no	yes	yes	yes
Horizontal element	grid	grid	grid	TIN	TIN	grid
No. of soil layers	1	1	many	many	many	many
Dual-compartment	yes	yes	no	no	no	no
Unsaturated flow	analytic	analytic	1D Richards	1D Richards	1D Richards	3D Richards
Overland flow	steepest descent	steepest descent	2D St. Venant	1D St. Venant	steepest descent	Kinem. Wave
Channel routing	linear	linear, Dupuit or Muskingum	1D St. Venant	1D St. Venant	Kinem. Wave	Kinem. Wave
Groundwater	(boundary condition)	linear reservoir or MODFLOW	3D Boussinesq	3D Richards	(boundary condition)	3D Richards
Model complexity	low	medium	medium	medium	high	high
Computational need	low	medium	high	medium	high	high

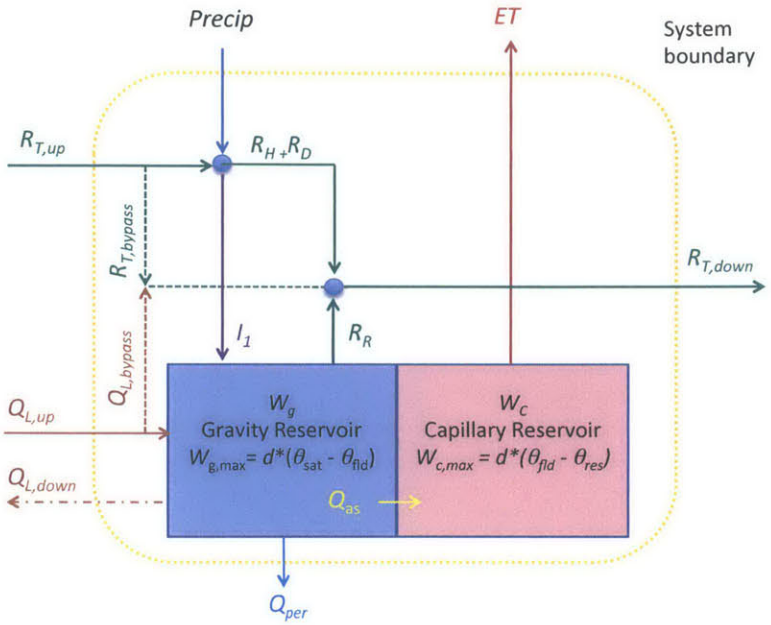


Figure 2.2: Schematic diagram of the mass balance for a typical pixel of MOBIDIC prior to model modifications

transport, it is not used in this thesis. The soil is modeled as a single layer with dual compartments, namely, the gravity reservoir composed of large pores that drain under gravity, and the capillary reservoir composed of small pores that do not drain under gravity. The water content states [L] of these reservoirs are W_g, and W_c, and the per unit area volume capacities [L] are $W_{g,max}$, and $W_{c,max}$, respectively, which are parameterized as,

$$W_{g,max} = d\,(\theta_{sat} - \theta_{fld}) \tag{2.2}$$

$$W_{c,max} = d\,(\theta_{fld} - \theta_{res}) \tag{2.3}$$

where, d is the thickness of the modeled soil layer, and θ_{sat}, θ_{fld}, and θ_{res}, are the volumetric soil moisture [-] at saturation, field capacity, and residual content, respectively. The values of θ_{sat}, θ_{fld}, and θ_{res} are initialized as function of soil texture type and using typical values reported by *Rawls et al.* (1982).

Within each computational timestep, dt [T], the hydrologic fluxes linking elements across the landscape include infiltration-excess runoff R_H, partial-area (saturation from below) runoff R_D, total runoff R_T, return flow R_R, and lateral subsurface flow

Q_L. These water fluxes can be limited by the available water to be transported, the transport capacity, or the available receiving storage. For each soil moisture storage unit the infiltration I, the absorption Q_{as} from W_g to W_c, the percolation Q_{per}, and the lateral subsurface flow Q_L, are formulated according to Equations 2.4 to 2.7,

$$I = min\left\{\phi_{ch}(P + R_{T,up}) + W_s, K_s dt, (W_{g,max} - W_g)\right\} \tag{2.4}$$

$$Q_{as} = min\left\{W_g, \kappa\,(1 - W_c/W_{c,max})\right\} \tag{2.5}$$

$$Q_{per} = \gamma\, W_g \tag{2.6}$$

$$Q_L = \beta\, W_g \tag{2.7}$$

where, P is precipitation [L], ϕ_{ch} the fraction of channelized flow, K_s the soil saturated hydraulic conductivity [L/T]; and κ, γ, and β, are dimensionless parameters with values from 0 to 1. The subscripts '*up*' and '*down*' denote incoming flow from upstream cell(s), and outgoing flow to downstream cell, respectively; and $Q_{L,bypass}$ and $R_{T,bypass}$ [L] are the portions of the lateral subsurface and total surface runoff from upstream cells that are routed directly to the downstream cell.

Infiltration fills the gravity storage at a rate limited by the saturated hydraulic conductivity. Absorption flux Q_{as} draws water from gravity storage into available capillary storage. The parameter κ is a linear rate coefficient. The water in gravity storage is lost to percolation or to lateral subsurface flow. Both are again characterized by linear rate coefficients γ and β.

The conceptualization of soil water storage as gravity and capillary storage and the flux relations shown in Equations 2.4 to 2.7, constitute the core of the simplified modeling system. Infiltration fills the larger pores with gravity storage. Water is moved from the gravity storage into the smaller capillary storage pores. Losses to the groundwater and lateral flow are only from gravity storage. Simple linear rate constants characterize the time scales of these exchanges. This simple representation is based on physical considerations and they result in a parsimonious and computationally efficient modeling approach.

Except during a precipitation event and the subsequent draining period, most of the fluxes are inactive. During dry conditions, the only significant flux is ET which occurs at the potential rate PET until W_c is consumed. PET is determined through a surface energy balance, as illustrated in Figure 2.3. From this, it follows that,

$$E = R_n - H - G \tag{2.8}$$

$$PET = E/\rho_w L_v \tag{2.9}$$

where, E is the latent heat, R_n the net incoming radiation, H the sensible heat, G the heat flux into the soil, ρ_w the density of water, and L_v the latent heat of vaporization.

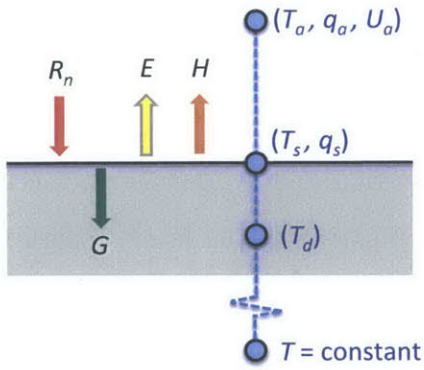


Figure 2.3: The surface energy balance of MOBIDIC at each computational unit

The turbulent fluxes are computed according to Equations 2.10 and 2.11 where, ρ_a is the density of air, C_a is the heat capacity of air, C_H is the turbulent heat exchange coefficient, and U is the wind speed; T_a and q_a are the temperature, and specific humidity of air, respectively; and T_s and q_s are the temperature, and specific humidity of the surface (combination of soil and vegetation), respectively.

$$H = \rho_a C_a C_H U(T_s - T_a) \tag{2.10}$$

$$E = \rho_a L_v C_H U(q_s - q_a) \tag{2.11}$$

The unknown surface temperature T_s and soil heat flux G are estimated using the

heat diffusion equation,

$$\rho_s C_s \frac{\partial T}{\partial t} = \frac{\partial}{\partial z}\left(k\frac{\partial T}{\partial z}\right) \tag{2.12}$$

where, ρ_s is the density, C_s the heat capacity, k is the thermal conductivity, and T the soil temperature. Equation 2.12 is integrated forward in time using a parsimonious 3-point vertical discretization:

$$T(z = 0) = T_s \tag{2.13}$$

$$T(z = z_d) = T_d \tag{2.14}$$

$$T(z = z_y) = T_{constant} \tag{2.15}$$

where, z_d and z_y are the damping depths of daily, and yearly heatwaves, respectively. The lower boundary condition is a constant temperature $T_{constant}$ roughly equal to the annual mean air temperature. The upper boundary condition is,

$$k\frac{\partial T_s}{\partial z} = -G \tag{2.16}$$

2.4 Model Modifications: Processes

Modifications were made on the mass balance of MOBIDIC in order to make the model more realistic, and applicable to diverse real-world settings. These include the addition of plant/canopy reservoir and surface reservoir, the addition of an alternative scheme for ET, and modifications of the conceptualization and formulation of lateral subsurface flow. Moreover, soil moisture simulations are improved by i) adding control rules for percolation; ii) adding a capillary rise module; and iii) introducing a modified form of the linear reservoir model for groundwater dynamics.

2.4.1 Canopy Interception and Surface Storage

Figure 2.4 is a quick reference guide showing the revised schematic diagram and formulations for the mass balance of MOBIDIC. This already incorporates the process modifications that are now herein discussed in detail. First, notice that two water

reservoirs have been added: i) the plant or canopy reservoir; and ii) the surface reservoir e.g. for depressions and ponds. The per unit area volume capacities [L] of these reservoirs are $W_{p,max}$ and $W_{s,max}$, and the water content states are W_p and W_s, respectively. Now, precipitation first encounters the vegetation canopy. Throughfall precipitation reaching the ground either infiltrates into the soil, gets stored on the surface, or becomes runoff.

2.4.2 Evapotranspiration

With the addition of the plant and surface reservoirs, evapotranspiration ET now has three components: ET_1 is from the plant reservoir, ET_2 is from the surface reservoir, and ET_3 is from the soil capillary reservoir. These fluxes are computed sequentially using simple formulations as shown in Equations 2.17 to 2.20. Note that the original version of MOBIDIC only has ET_3, which is limited by either W_c or PET (since $ET_1 = ET_2 = 0$).

$$ET = ET_1 + ET_2 + ET_3 \tag{2.17}$$

$$ET_1 = min\{W_p,\ PET\} \tag{2.18}$$

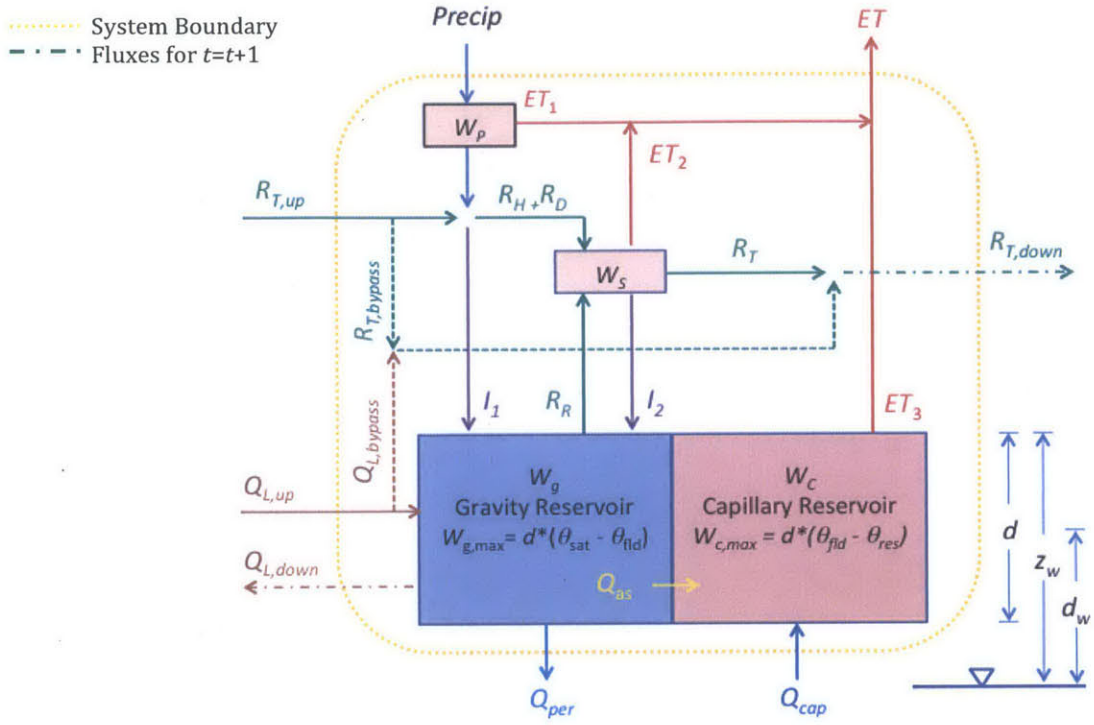
$$ET_2 = min\{W_s,\ PET - ET_1\} \tag{2.19}$$

$$ET_3 = min\{W_c,\ PET - ET_1 - ET_2\} \tag{2.20}$$

Let Equation 2.20 be denoted as 'ET3 Model 0'. ET_3 is shut-off only once W_c is empty i.e. when the soil is dried down to residual water content. The problem with this simple scheme is that it does not mimic the gradual drying behavior of soil. In reality, the ET efficiency ET/PET varies with soil moisture or plant-available soil moisture, see e.g. *Saxton et al.* (1974) and *Haan et al.* (1982).

$$\text{ET efficiency} = \frac{ET}{PET} = f(\theta) \tag{2.21}$$

So an alternative scheme, 'ET3 Model 1', is added. This scheme is formulated according to Equation 2.22. This S-curve type equation was chosen because it mimics



W_c	Capillary Reservoir	R_h	Hortonian Runoff	$\emptyset_{ch}$	channeled fraction
W_g	Gravity Reservoir	R_d	Dunne Runoff	d	soil depth
W_{Cl}	Canopy Reservoir	R_r	Return Flow	z_w	depth to water table
W_s	Surface Reservoir	R_T	Total Runoff	d_w	distance from center of unsaturated soil to water table

Infiltration $I = min\{\phi_{ch}(P + R_{T,up}) + W_s\,,\ K_s\Delta t,\ W_{g,max} - W_g\}$

Soil Absorption $Q_{as} = min\{W_g, \kappa(1 - W_c/W_{c,max})\}$

Evapotranspiration $ET = \sum_{i=1}^{3} ET_i;\ ET_1 = min\{W_p, PET\};\ ET_2 = min\{W_s, PET - ET_1\}$

$ET_3 = min\{W_c, (PET - ET_1 - ET_2)/(1 + exp(\xi - 10S))\}$

$S = (W_c + W_g)/(W_{c,max} + W_{g,max})$; ξ = constant

Percolation (or GW Seepage if negative)

$$Q_{per} = \begin{cases} min\left\{\gamma W_g, \left[W_g + \left(\frac{z_w}{d} - 1\right) W_{g,max}\right]/2\right\} & \text{if } z_w \geq 0 \\ min\{[W_{g,max} - W_g - z_w]/2, W_{g,max} - W_g\} & \text{if } z_w < 0 \end{cases}$$

Lateral Subsurface Flow $Q_L = \gamma W_g$

Capillary Rise $Q_{cap} = K_{sat} \frac{(d_w/\psi_1)^{-\eta} - (\psi/\psi_1)^{-\eta}}{1 + (\psi/\psi_1)^{-\eta} + (\eta - 1)(d_w/\psi_1)^{\eta}}$

Figure 2.4: A quick reference of the mass balance of the modified MOBIDIC for a typical pixel

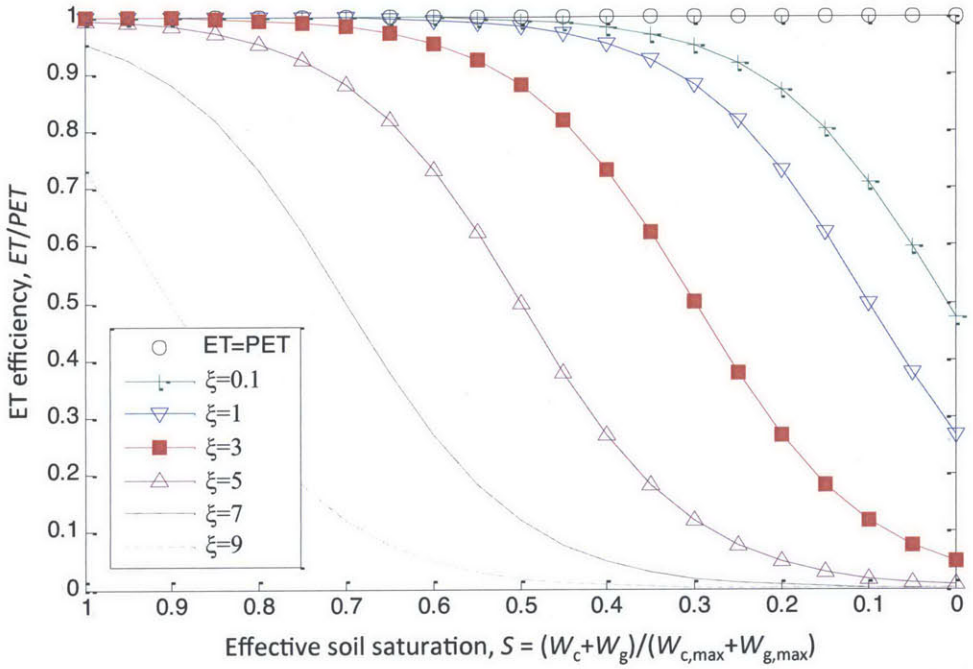


Figure 2.5: ET efficiency [-] vs. effective soil saturation S [-] using 'ET3 Model 1'

the nonlinear behavior of ET as function of the effective soil saturation S [-] and a single parameter ξ. The factor of 10 was set for convenience such that ξ takes on non-negative integer values (suggested value is 2 or 3). Figure 2.5 plots the resulting ET efficiency curves for several values of ξ, with the default value of 3 shown in red.

$$ET_3 = \frac{(PET - ET_1 - ET_2)}{1 + exp(\xi - 10S)} \tag{2.22}$$

$$S = (W_c + W_g)/(W_{c,max} + W_{g,max}) \tag{2.23}$$

2.4.3 Percolation and Lateral Subsurface Flow

Based on Equations 2.6 and 2.7, percolation Q_{per} and lateral subsurface flow Q_L are simply the products of the water content of the gravity reservoir with the parameters γ, and β, respectively. First, let us examine whether these simplistic formulations are realistic. *Eagleson* (1978) proposed the analytic solution for percolation,

$$Q_{per} = K_s S^{3+2m} \tag{2.24}$$

Figure 2.6 compares Q_{per} normalized by K_s computed using the analytic solution and MOBIDIC's formulation, for different values of S and three representative soil texture types. For MOBIDIC, γ was set equal to $K_s/W_{g,max}$. For coarse soil e.g. loamy sand, MOBIDIC gives higher Q_{per} than the analytic value, while for fine soil e.g. silty clay, it gives lower Q_{per}. For soils with intermediate texture e.g. sandy clay loam, Q_{per} computed by MOBIDIC is about the same as the analytic value.

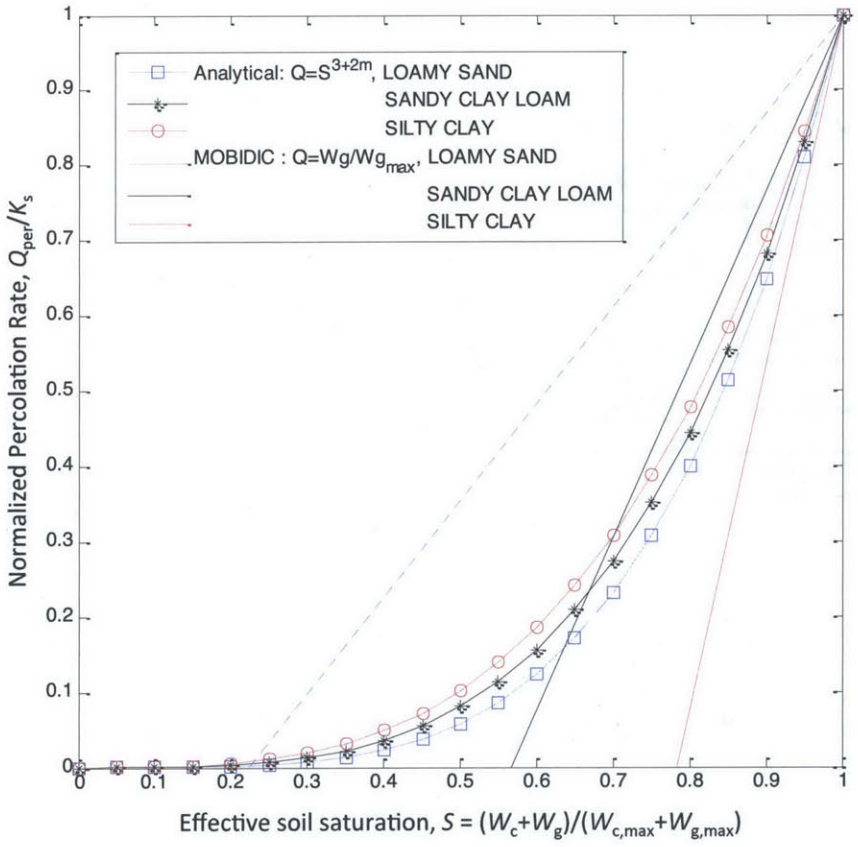


Figure 2.6: Comparison of the normalized percolation rate as a function of effective soil saturation computed using MOBIDIC's approximation and the analytic solution of *Eagleson* (1978), for 3 soil textures

Based on experience, γ is one of the important parameters of MOBIDIC. Thus, an additional benefit from Figure 2.6 is the realization that a good initialization of γ is $K_s/W_{g,max}$. If anisotropy is assumed, as is most often the practice, this value can

also be used to initialize the lateral subsurface parameter β. These imply a possible reduction of 2 degrees of freedom in MOBIDICs parameter space.

Still, with the current formulations, Q_{per} and Q_L occur if and only if the soil is wetter than field capacity. Moreover, an underlying assumption of Equation 2.6 is that the water table is deep and does not influence percolation. This is not the case in floodplains, riparian zones, or many humid environments. A solution is to use a piece-wise defined Q_{per}, depending on the position of the groundwater table z_w relative to the modeled soil layer. To enable MOBIDIC to dynamically reconcile z_w and W_g, Q_{per} is allowed to take negative values i.e. it is made bidirectional. $Q_{per} < 0$ means that water is moving from groundwater into the gravity reservoir.

Figure 2.7 illustrates the 3 pairs of possible scenarios, where z is the depth [L] below the surface (positive downward) and d is the thickness of the modeled soil. The scenarios are vertically exaggerated for clarity. Typically, z_w is below the bottom of the modeled soil layer. If $\gamma W_g < (z_w - d)W_{g,max}$, see Case A-1, then percolation proceeds until W_g is empty. Otherwise, see Case A-2, percolation stops once W_g and z_w are in hydrostatic equilibrium. For this pair of cases,

$$Q_{per} = min\{\gamma W_g, \left[W_g + (z_w - d)\left(\frac{W_{g,max}}{d}\right)\right]/2\} \qquad \text{if } z_w > d \tag{2.25}$$

The second case is when z_w is within the modeled soil layer. Either there is more water based on z_w than based on W_g (Case B-1), or vice versa. Water flows in the direction that makes these state variables consistent. The equation for this case is the same as that of the previous case except Q_{per} can be negative.

$$Q_{per} = min\{\gamma W_g, \left[W_g + (z_w - d)\left(\frac{W_{g,max}}{d}\right)\right]/2\} \qquad \text{if } 0 \leq z_w \leq d \tag{2.26}$$

In the third case, z_w is above the land surface. At the beginning of a time step, the gravity reservoir may or may not be full. Groundwater (or more correctly surface water) will fill the gravity reservoir until it is consistent with z_w (Case C-1) or until $W_g = W_{g_{max}}$ (Case C-2). This can be expressed by the following equation, with the

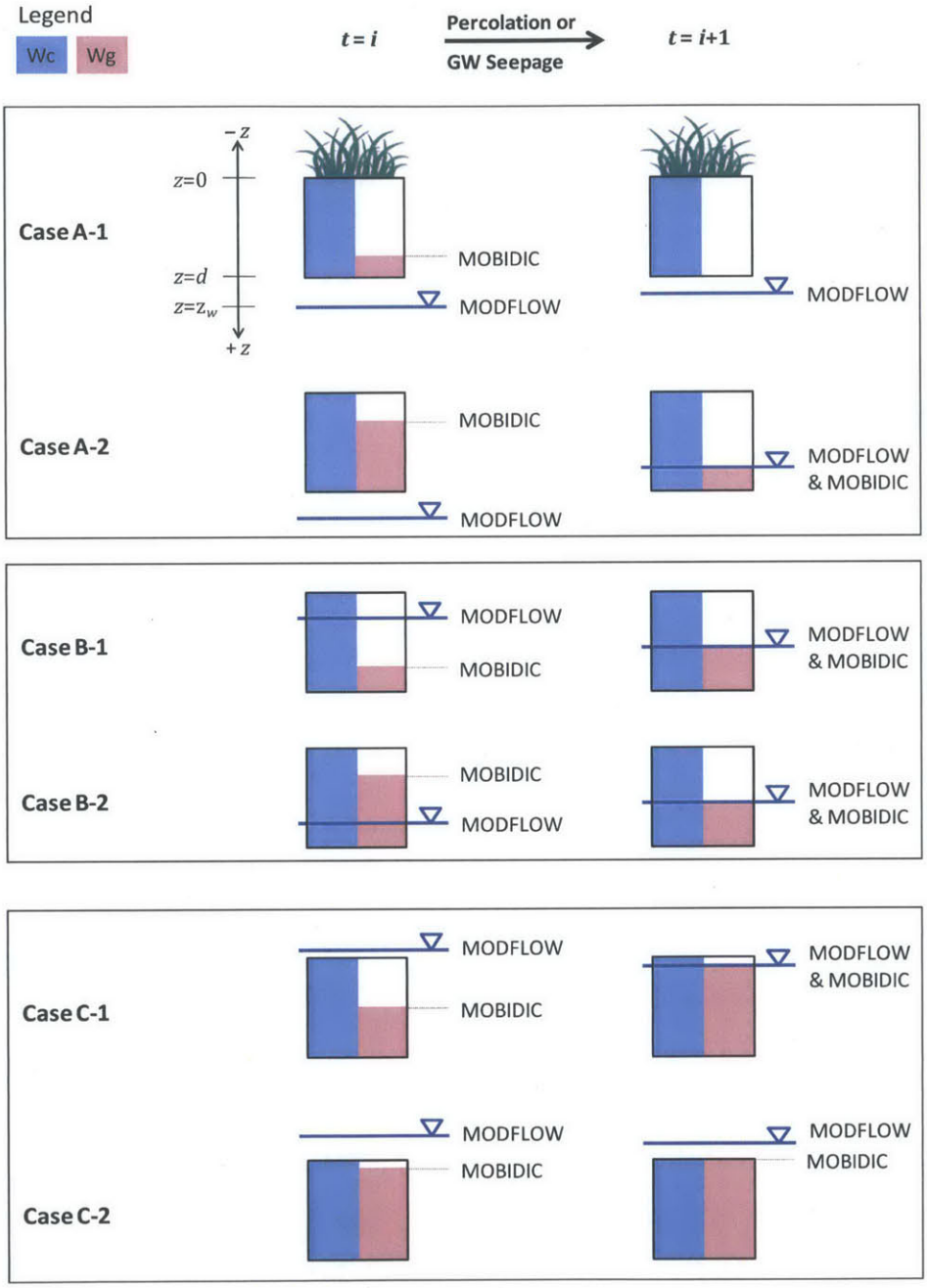


Figure 2.7: Illustration of the cases when the position of the water table is inconsistent with the content of the soil gravity reservoir

first part accounting for Case C-1 and the second for Case C-2,

$$Q_{per} = \begin{cases} \dfrac{[z_w - (W_{g,max} - W_g)]}{2} & \text{if } z_w < 0 \text{ and } (-z_w) < (W_{g,max} - W_g) \\ W_g - W_{g,max} & \text{if } z_w < 0 \text{ and } (-z_w) \geq (W_{g,max} - W_g) \end{cases} \tag{2.27}$$

which can be combined and written compactly as,

$$Q_{per} = min\{[(W_{g,max} - W_g) - z_w]/2,\ W_{g,max} - W_g\} \qquad \text{if} z_w < 0 \tag{2.28}$$

Combining Equations 2.25, 2.26, and 2.28, and simplifying, the final modified percolation model is:

$$Q_{per} = \begin{cases} min\{\gamma W_g,\ \left[W_g + (z_w - d)\left(\dfrac{W_{g,max}}{d}\right)\right]/2\} & \text{if} z_w \geq 0 \\ min\{[(W_{g,max} - W_g) - z_w]/2,\ W_{g,max} - W_g\} & \text{if} z_w < 0 \end{cases} \tag{2.29}$$

2.4.4 Capillary Rise

The soil capillary water storage unit can also receive water from capillary rise from shallow water table. There are a number of available capillary rise models e.g. *Gardner* (1958), *Eagleson* (1978), and *Bogaart et al.* (2008). They vary primarily based on their parameterization of K_s and the soil matric potential ψ [L] as function of soil moisture. The capillary rise model of *Salvucci* (1993) shown in Equation 2.30 was chosen because unlike other models, it allows direct calculation of capillary rise Q_{cap} [L] as function of ψ and the representative distance d_w [L] from the unsaturated soil layer to the water table.

$$Q_{cap} = \frac{[(d_w/\psi_1)^{-\eta} - (\psi/\psi_1)^{-\eta}]\, K_s dt}{1 + (\psi/\psi_1)^{-\eta} + (\eta - 1)(d_w/\psi_1)^{-\eta}} \tag{2.30}$$

where ψ_1 [L] is the bubbling pressure, and η [-] is the product of the Brooks-Corey pore-size distribution index, m [-], and the pore-size disconnectedness index, c [-]. *Brooks and Corey* (1964) is used to compute ψ,

$$\psi = \psi_1 S^{-1/m} \tag{2.31}$$

The required input of the capillary rise module is a map of soil texture type based on US Department of Agriculture (USDA) classification. Then the lookup table shown in Table 2.2 is used to extract soil parameters based on typical values reported by *Rawls et al.* (1982). The depth to water table d_w that is passed on as input to the capillary rise module is computed depending on z_w, see Figure 2.8. If the water table is above the land surface (case A), d_w is not computed and is automatically $Q_{cap} = 0$. If the water table is within the modeled soil layer (case B), $d_w = z_w/2$ since only the portion of the soil above z_w experiences capillary rise. If the water table is below the modeled soil layer, $d_w = z_w - d/2$. Tests conducted using all soil texture types and varying depth to water table and soil saturation, showed that the capillary rise model yields Q_{cap} with realistic values and correct dynamics.

Table 2.2: Lookup table of soil parameters for the capillary rise module of MOBIDIC. Source: *Rawls et al.* (1982)

Soil Texture	θ_{sat} %	θ_{fld} %	θ_{res} %	K_s mm/hr	ψ_1 mm	m [-]
clay (< 30% clay)	47.5	39.6	9.0	6	856	0.165
clay loam	46.4	31.8	7.5	23	564	0.242
loam	46.3	27.0	2.7	132	401	0.252
loamy sand	43.7	12.5	3.5	611	206	0.553
silt	46.0	34.0	3.4	30	700	0.250
silty loam	50.1	33.0	1.5	68	509	0.234
silty clay	47.9	38.7	5.6	9	765	0.150
silty clay loam	47.1	36.6	4.0	15	703	0.177
sand	43.7	9.1	2.0	2100	160	0.694
sandy clay	43.0	33.9	10.9	12	795	0.223
sandy clay loam	39.8	25.5	6.8	43	594	0.319
sandy loam	45.3	20.7	4.1	259	302	0.378

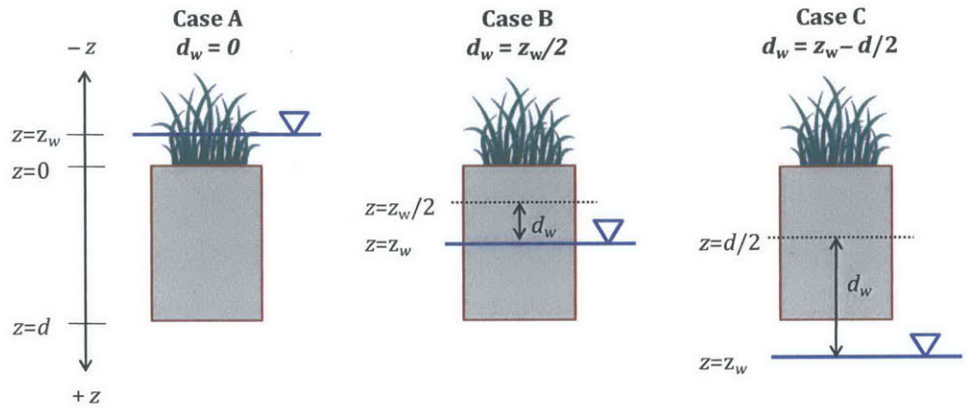


Figure 2.8: Possible water table positions relative to the modeled soil layer

2.5 Basin-scale Mass Balance

For long-term mass balance at the basin-scale, consider the system as composed of the 4 water reservoirs modeled by MOBIDIC. Figure 2.9(a) shows the water fluxes entering and leaving the system. If the basin boundary is a groundwater divide and there are no transbasin water imports or exports, the water balance is,

$$\dot{S} = (P + Q_{cap} + Q_{bf}) - (ET + Q_{per} + Q_{out}) \tag{2.32}$$

where, $\dot{S}$ is the net change in water storage, Q_{bf} the baseflow, Q_{out} the total streamflow at the outlet of the basin, and all other variables are as earlier defined and with units of [L]. Equivalently, $\dot{S}$ is the sum of the changes in water content of all reservoirs,

$$\dot{S} = (W_c + W_g + W_s + W_p)|_{initial}^{final} \tag{2.33}$$

For multi-year simulations, the change in storage is negligible. Another convenient system boundary for the mass balance includes the deep soil and groundwater aquifer, and neglects internal fluxes, see Figure 2.9(b). This yields a simplified equation for precipitation partitioning,

$$P = ET + Q_{out} \tag{2.34}$$

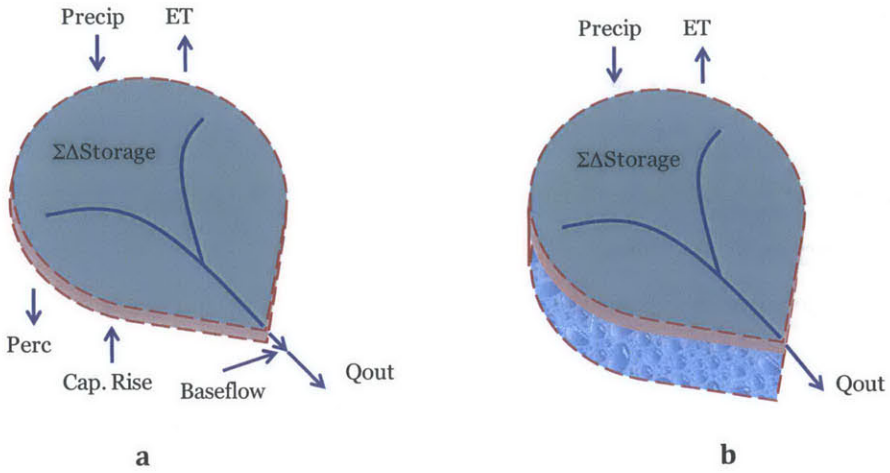


Figure 2.9: Overall water balance of a basin considering the system is a) the modeled soil layer; b) the modeled soil layer plus the groundwater aquifer

2.6 Model Modifications: Numerical Efficiency

MOBIDIC is written in MATLAB with several subroutines. Because it is a distributed model, it deals with large matrices, which makes it computationally intensive. This can be a limiting factor for applications across large domains and in ensemble mode. Thus, several changes were also made to the source codes in order to improve its numerical efficiency such that despite the increase in model complexity due to the modifications in hydrologic processes representation, the current version of MOBIDIC runs even faster than its previous version. In addition, code modifications were made to minimize hardware requirements e.g. random access memory and storage space.

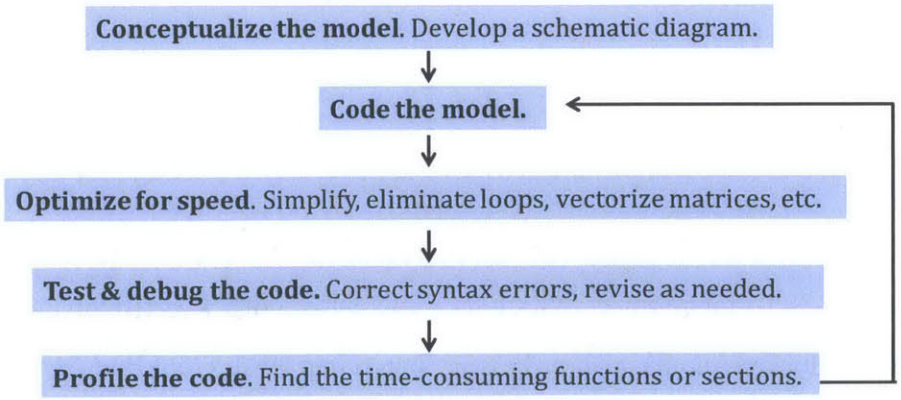


Figure 2.10: Procedure used to improve the numerical efficiency of MOBIDIC

The numerical efficiency of MOBIDIC was improved using the procedure shown in Figure 2.10. Before revising the codes, a conceptual diagram of the model must be developed. Fortunately, MOBIDIC is already an existing model so it does not need to be developed from scratch. Unfortunately however, since many subroutines have been added and changes made to the original version, the current version is not well organized and its performance is not optimized. Thus, considerable time was spent in the sub-loop of profiling and revising the code, indicated by the feedback arrow .

Profiling of the code was done using MATLAB Profiler, which allows tracking of parent-child relationship of each function/module, and automatically records the number of calls and the time used by each (sub-) sub-routine. The Profiler allows easy identification of time-consuming sections, which were then reviewed and revised as possible. Profiling and revising of the codes were iterated until the performance was acceptable. Some of the specific techniques used are:

- simplification and reorganization;
- elimination of loops and conditional statements;
- vectorization of matrices;
- use of common variables;
- use of consistent units;
- use of a basis of one time step (e.g. the unit for fluxes is [L]); and
- use of filters recognizing that often $P = 0$, and $W_g = 0$ for most pixels

The test basin used to assess the improvement in the speed of MOBIDIC is the Blue River Basin in Oklahoma. The basin was setup with $8.3x10^4$ pixels, and the model was run for one month at hourly time step on an IBM ThinkPad 64-bit 2.1GHz Intel™Core Duo processor. The total runtime was initially 1350 seconds, then reduced to 310 seconds or a 78 percent reduction.

2.7 Model Implementation

The main steps to develop a MOBIDIC model of a specific basin are the following:

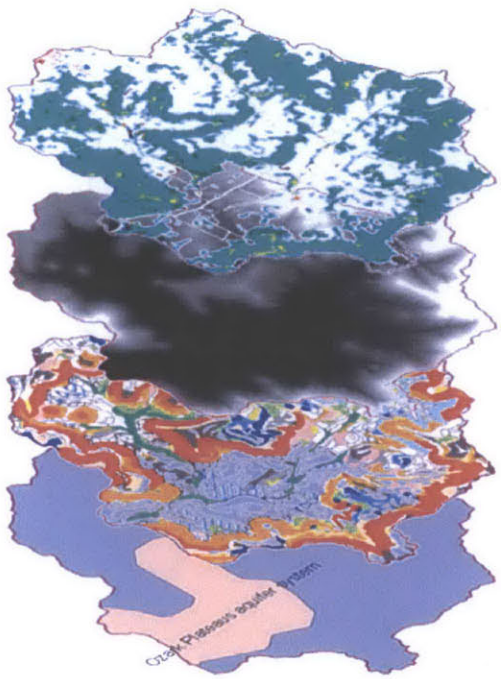
- Review of literature especially hydrologic investigation and similar studies
- Data collection
- Processing, quality check, and preliminary analysis of data
- If possible and as needed, field work, sampling and/or monitoring
- Preparation of model inputs including options and settings
- Model calibration and validation

2.7.1 Collection and Preparation of Model Inputs

Figure 2.11 illustrates the input data of MOBIDIC.The inputs can be categorized into four groups: GIS layers of basin properties; meteorological forcings; initial and boundary conditions; and model parameters and settings. In recent years, GIS data layers such as digital elevation model (DEM), aerial photos, and land cover and soil maps, have become more readily available. For basins in the US, most of these data layers can be downloaded for free from "The National Map" (http://viewer.nationalmap.gov/) of USGS. Additional data can be obtained from websites of research institutions and agencies at the state and local levels.

The collected GIS data are then imported into ArcGIS™for visualization and processing. Topographic and geomorphologic inputs are processed using the Hydrology Toolbox. The processing steps are listed below.

1. Define the projection system of the DEM
2. Fill the pits of DEM
3. Create the map of flow directions using the D8 method of *Tarbonton* (1997)
4. Create the flow accumulation map



Land cover (vegetation, surface storage, albedo, etc.)

DEM (elevation, flow directions, streams, etc.)

Soils (soil thickness, porosity, k_{sat}, etc.)

Aquifers – optional

+ Met. forcings (precip, air temp., solar radiation, wind speed, humidity)

+ Initial & boundary conditions

Figure 2.11: Data inputs of MOBIDIC

5. Delineate the river network and the basin boundary

6. Standardize the resolution, extent, and data format of each data layer

To automate these processes, a script was generated using the Model Builder of ArcGIS™, see Figure 2.12. The script was later edited in Python,

Other required model inputs are land cover and soil maps, which are in turn used to derive parameters such as albedo, turbulent heat exchange coefficient, canopy interception capacity, and soil hydraulic conductivity. On the other hand, MOBIDIC can output time series of streamflow at any point along the river network, and maps of hydrologic fluxes (e.g. infiltration, runoff, and ET) and states (e.g. soil temperature and water content of the soil capillary and gravity reservoirs).

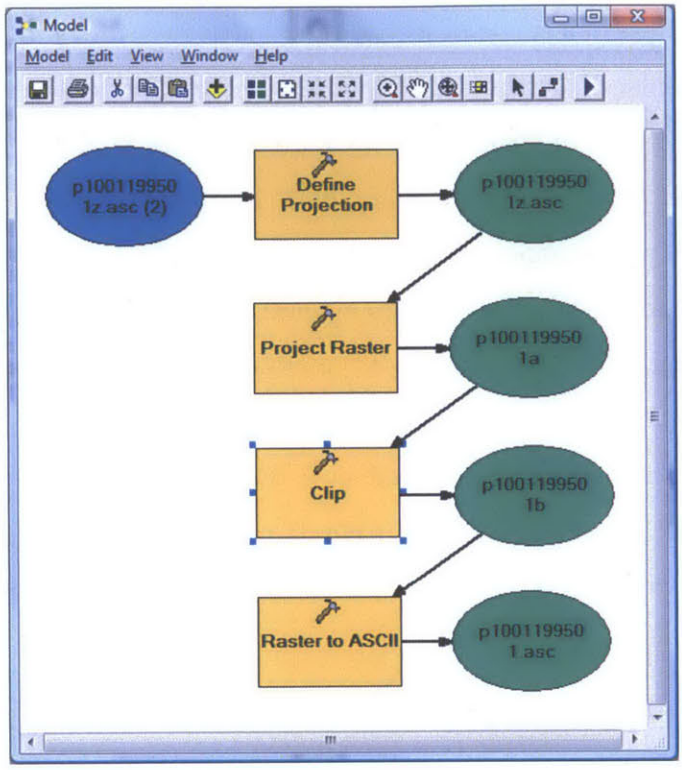


Figure 2.12: Example of how ArcGIS™Model Builder was used to automate the processing of topographic and geomorphologic data inputs

2.7.2 Calibration and Validation

The simulation period usually covers a few years and is composed of a warm-up period followed by a calibration period and a validation period. Having a warm-up period minimizes errors caused by incorrect initialization of the model. Alternatively, model spin-up can be performed until the simulated variables, esp. groundwater table and soil moisture, are practically yearly cyclical. Model parameters are adjusted during calibration. Once the model is satisfactory, the performance of the model is re-evaluated on the validation period.

Global parameters are initialized based on typical literature values, plus information on the specific basin being modeled, and previous experience with other basins. Calibration is initially done manually. Once the parameter set seems correct, then automatic calibration is performed to fine-tune the parameters. The entire calibration process is guided by both objective and qualitative checks.

The traditional use of a single-objective function as measure of model performance has slowly been replaced by the use of multi-objective functions or multiple-criteria measures that tests the models ability to match various aspects of the hydrologic response (*Gupta et al.*, 1998; *Vrugt et al.*, 2003; *Pokhrel et al.*, 2012). The objective measures of model performance are evaluated using observed and simulated streamflow and/or soil moisture. The statistical metrics used are i) the bias B; ii) the Pearson correlation coefficient R; iii) the modified index of agreement (*Willmott*, 1981); and iv) the Nash-Sutcliffe Efficiency index NSE (*Nash and Sutcliffe*, 1970). The latter 2 metrics are commonly used in hydrologic studies (*Krause et al.*, 2005; *Foglia et al.*, 2009; *Segui et al.*, 2009). For n pairs of observed x, and modeled y data, with mean values of $\overline{x}$ and $\overline{y}$, respectively, the metrics of model performance are defined as follows:

$$B = \frac{\sum_{i=1}^{n}(y_i - x_i)}{\sum_{i=1}^{n} x_i} \tag{2.35}$$

$$R = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2}} \tag{2.36}$$

$$IA = 1 - \frac{\sum_{i=1}^{n}(y_i - x_i)^2}{\sum_{i=1}^{n}(|y_i - \overline{x}| + |x_i - \overline{x}|)^2} \tag{2.37}$$

$$NSE = 1 - \frac{\sum_{i=1}^{n}(y_i - x_i)^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2} \tag{2.38}$$

NSE ranges from $-\infty$ to unity. $NSE < 0$ means that the observed mean is a better predictor than the model. *Foglia et al.* (2009) considers models with NSE less than 0.2 as insufficient, 0.2–0.4 as sufficient, 0.4–0.6 as good, 0.6–0.8 as very

good, and 0.8–1 as excellent. The overall objective function J used in calibration is expressed in Equation 2.39 where $\mathbf{w}$ is a 4x1 matrix of weights of the individual metrics, $\sum \mathbf{w} = 1$. A perfect fit has $J = 0$ which corresponds to $B = 0$, $R = 1$, $IA = 1$, and $NSE = 1$.

$$J = min\{[|B|, (1 - R), (1 - IA), (1 - NSE)]\,\mathbf{w}\} \tag{2.39}$$

However, the ultimate objective of the calibration process, is not just to minimize J. Emphasis is also given to the *realism* of the model. As mentioned, parameters are constrained according to literature values. Moreover, other simulated hydrologic fluxes and state variables are also plotted and qualitatively compared against observations and previously reported values.

2.8 Summary

This chapter describes the raster-distributed catchment hydrologic model MOBIDIC which is the main tool used in this thesis. Its mass and energy balance is described. Some model modifications were introduced to improve the representation of some hydrologic processes especially for better simulation of the heterogeneity and dynamics of soil moisture. These include i) the addition of plant and surface reservoirs; ii) the use of a single-parameter S-type model of ET as function of soil saturation; iii) the modifications of the conceptualization and formulation of lateral subsurface flow; iv) the addition of control rules for percolation; v) the addition of a capillary rise module; and vi) the use of a modified form of the linear reservoir model for groundwater dynamics. Moreover, profiling and modification of the MATLAB codes of MOBIDIC reduced its run time by about 78 percent. Finally, the general procedure for developing a MOBIDIC model of a certain site or basin is outlined.

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 3

Simulations of Local Scale Soil Moisture in a Catchment Hydrologic Model

3.1 Abstract

Distributed and continuous catchment models are used to simulate water and energy balance and fluxes across varied topography and landuse. The landscape is discretized into plan computational elements at resolutions of $10^1 - 10^3$ m, and soil moisture is the hydrologic state variable. Increasingly, catchment models are used in conjunction with replicates of forcings and parameters in order to generate an ensemble that characterizes the uncertainty in the inputs and models. At the local scale and within each of the spatial computational elements, the vertical soil moisture dynamics are important. In catchment models these local scale processes are modeled using one-dimensional soil columns. These hydrologically active soil columns are discretized into layers that are $10^{-3} - 10^{-1}$ m thick. This treatment of local processes creates a mismatch between the horizontal and vertical scales. Moreover, for applications across large domains and in ensemble mode, this treatment can be a limiting factor due to its computational demand. This chapter presents continuous multi-year simulations

of soil moisture at the local scale using i) the physically-based distributed hydrologic model MOBIDIC; and ii) a benchmark model — the legacy 1–D SHAW. SHAW was chosen as the benchmark because it solves both mass and energy balance and it uses non-linear soil physics relations for flow in unsaturated soils. Although MOBIDIC is intended for basin-scale catchment modeling, and despite its use of a single soil layer, simple partitioning of soil moisture into gravity (free) and capillary-bound water, and linear parameterization of infiltration and some other fluxes, this study shows that it can capture the more detailed dynamics implicitly included in the SHAW model. The comparison is performed at two sites with different climates (semi-arid and sub-humid).

3.2 Introduction

Soil moisture controls the partitioning of rainfall into infiltration and runoff, and it controls land surface temperature through its effect on the partitioning of available energy into sensible and latent heat fluxes. It is the hydrologic state variable in models of surface water and energy balance. Its dynamics are affected by hydrometeorological forcing of precipitation, radiation and atmospheric evaporative demand. Furthermore, topography, landuse, soil properties and lateral exchanges across the landscape affect its temporal evolution (*Western and Grayson*, 2000; *Lawrence and Hornberger*, 2007; *Vereecken et al.*, 2007; *Ivanov et al.*, 2010; *Liu et al.*, 2012; *Beven and Germann*, 2013).

There are various methods for measuring soil moisture, but unfortunately, as described in the previous chapter, all current observing systems have their shortcomings. Numerical models are used to complement hydrologic observations. Incoming radiation and precipitation are used in conjunction with water and energy balance models to simulate the evolution of soil moisture in the vadose zone and determine the water and energy fluxes across the landscape. *Harter and Hopmans* (2004) describes how hydrologic models have traditionally been used by two largely disconnected groups: the watershed hydrologists (and recently also climate modelers) who deal with macro-

processes; and the soil physicists who study soil properties and states at the laboratory or local to plot scales. Watershed hydrologists have traditionally used lumped or semi-distributed models that treat the vadose zone as a zero-dimensional black box. The computational timestep is usually hourly, daily, or even longer. Two examples of these models are the semi-distributed models TOPMODEL (*Beven and Kirby*, 1979) and SAC-SMA (*Burnash et al.*, 1973) which have both been demonstrated as highly capable in simulating streamflow. Unfortunately, these models do not account for the spatial heterogeneity in landuse, topography and other factors; and provide little or no information on the spatial variability of hydrologic states and fluxes within the basin. On the other hand, soil physicists who have detailed measurements of soil properties and states at the local to plot scales, model unsaturated flow by discretizing the hydrologically active soil column into several thin layers, and solving Richards equation, see Equation 2.1. For numerical stability, this nonlinear differential equation is solved using sub-hourly time steps.

Over the years, the modeling efforts of the two disciplines have started to converge. This convergence is probably best manifested by the emergence of physically-based distributed hydrologic models (DHMs). These models discretize the landscape in computational elements that are $10^1 - 10^3$ m along the horizontal. Adopting the practice in soil physics, many DHMs also employ Richards equation and discretize the hydrologically active soil layer into vertical layers that are $10^{-3} - 10^{-1}$ m thick. Some DHMs that use the Richards formulation include MIKE-SHE (*Refshaard and Storm*, 1995) and ParFlow (*Ashby and Falgout*, 1996) that use grids for horizontal discretization; and PIHM (*Qu and Duffy*, 2007) and TRIBS (*Ivanov et al.*, 2004) that use triangulated irregular network (TIN) as horizontal elements, see Table 2.1. Some simpler models such as the SAC-SMA (*Burnash et al.*, 1973) represent the soil as having dual compartments — saturated and unsaturated.

The hydrologically active soil mantle is but a thin layer draped over the landscape, and it serves as the intermediate water storage connecting the surface above and the groundwater or deeper soil layers below. Because of the horizontal-to-vertical scale disparity, flow dynamics in the soil are often treated in DHMs as one-dimensional i.e.

lateral subsurface flow is considered negligible. Exceptions include MIKE-SHE and ParFlow which solve the full 3–D Richards equation. This treatment is however very computationally intensive as demonstrated by *Kollet et al.* (2010) who had to utilize 16,384 processors of a supercomputer to achieve reasonable run time for ParFlow simulations of a basin on the order of 10^3 km^2 at fine spatial resolution ($10^0 - 10^1$ m in the horizontal and $10^{-2} - 10^{-1}$ m in the vertical).

Models based on Richards formulation are useful when the vertical profile of soil moisture is desired especially when the soil column is significantly non-homogeneous. Information about the vertical soil structure is however often not available and highly uncertain where available.

As discussed, the scales mismatch between the vertical and horizontal discretization of DHMs (millimeters to centimeters in the vertical soil column versus tens to hundreds of meters in the horizontal) leads to two main problems: 1) solving the local scale vertical soil moisture dynamics based on Richards equation is computationally demanding; and 2) such fine vertical discretization increases the number of parameters to calibrate, and state variables to initialize.

Moreover, although Richards equation is probably an appropriate model for unsaturated flow at the local scale, it is questionable whether it is an appropriate physical model for watershed and regional scale applications (*Beven*, 1995; *Beven and Germann*, 2013; *Harter and Hopmans*, 2004). Using this equation for plan elements that are in the order of 10^1–10^3, makes the implicit assumptions that the vertical dynamics of soil moisture at the local scale is scale-invariant. In contrast, field measurements show that soil hydraulic conductivity and pore properties related to the soil retention curve (of ψ) vary significantly both in the horizontal and vertical (*Gelhar et al.*, 1992; *Rubin*, 2003; *Zhang et al.*, 2004). Furthermore, the review paper of *Beven and Germann* (2013) argues that the use of Richards equation to model field soil should not be considered physics-based but rather a convenient conceptual approximation. They explained that the Darcy and Richards equations have dominated soil physics in the last few decades because of the ready availability of numerical models based on these formulations, despite the convincing evidence that their underlying as-

sumptions, and carefully controlled experimental setups, are inappropriate for natural conditions. They highlighted the importance of macropores and suggested the use of at least a dual-pore soil structure. The catchment models TOPMODEL and SAC-SMA use this structure. Other models that use this dual-pore soil structure are the 1–D model of *Gerke and van Genuchten* (1993), the 1–D model MACRO (*Larsbo and Jarvis*, 2003; *Roulier et al.*, 2005), and the 1–D/2–D/3–D model HYDRUS (*Šimůnek and van Genuchten*, 2008). In these three models, the soil column is composed of a macropore and a matric compartment, and the water flow in the matric compartment is still solved using Richards equation.

Salvucci and Entekhabi (1994) demonstrated that a simple bucket-type equilibrium model can accurately reproduce the long-term mean evaporation, infiltration, and recharge, as well as the characteristic soil moisture profile. Using in situ measurements of soil moisture, *Walker et al.* (2004) showed that a simple bucket model captures the seasonality of the measurements although it overestimates the measurements during the dry-down period. They also showed that a Richards equation based model accurately reproduced the measurements at most of the instrumented soil depths. Finally, *Guswa et al.* (2002) found that for a woody African savanna, soil moisture predictions from a bucket model and a Richards equation based model are quite similar if plants can extract water from locally wet regions. However, similar studies that compare the performance of a dual-porosity model and a Richards equation based model in simulating the daily or seasonal dynamics of hydrologic states and fluxes are still not available.

This chapter tests the fidelity of a novel approach to modeling the soil moisture state in distributed hydrologic models. In particular, the distributed hydrologic model MOBIDIC which is the topic of the previous chapter, is used. Table 2.1 lists the features of MOBIDIC, and compares it with some of the hydrologic models that have been mentioned. A key feature of MOBIDIC is its use of a single layer of soil with dual compartments — one for free water and another for capillary-bound water. This representation accounts for both fast and slow processes. At the same time, it makes the model computationally efficient. Furthermore it reduces the number of state

variables in the overall dynamic modeling system. The fidelity of this approach to local processes is tested by comparing its soil moisture dynamics with that resulting from a numerical model that solves the vertical heat and moisture dynamics using detailed physics.

The goal of this chapter is to demonstrate that MOBIDIC which has a dual-pore structure, can simulate the dynamics of profile-averaged soil moisture as accurately as models using multiple soil layers and Richards equation. In addition, since most of the previous applications of MOBIDIC assessed its performance based mainly on stream-flow which is an area-integrated flux, this chapter also demonstrates that MOBIDIC is capable of correctly simulating the dynamics of soil moisture, soil temperature, and evapotranspiration.

The next section gives an overview of the selected benchmark model — the legacy 1–D SHAW. This is followed by sections describing the correspondence between SHAW and MOBIDIC variables, the measures of model performance, and the two study sites. Tests are performed at two sites representing semi-arid and humid contrasting conditions.

3.3 The SHAW Model

The Simultaneous Heat and Water (SHAW) models the transfer of heat, water, and solute within a 1–D vertical profile composed of multi-layered and multi-species plant cover, snow layer, dead plant residue layer, and multi-layered soil. It was first developed by *Flerchinger and Saxton* (1989) to simulate soil freezing and thawing, but has since undergone numerous modifications and extensions. It is available for free from the USDA Agricultural Research Service (ARS) Northwest Watershed Research Center (NWRC) website (ftp.nwrc.ars.usda.gov). It was chosen as the benchmark model for this study because i) it simultaneously solves mass and energy balance; ii) it solves Richards equation for soil moisture; and iii) it has detailed treatment of evapotranspiration (ET).

Figure 3.1 shows the physical system represented by the model. In SHAW, a soil

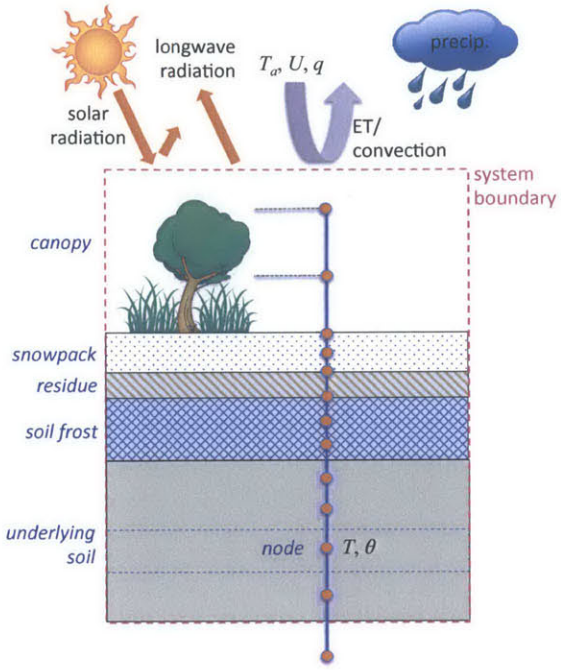


Figure 3.1: The physical system described by the 1–D SHAW model; P is precipitation, T_a is air temperature, u is wind speed, RH is relative humidity, S_t is total solar radiation, T is soil temperature, and θ is soil moisture

column is discretized into nodes. The fluxes between nodes are solved using implicit finite-difference. The required inputs include general site information (e.g. location, elevation, aspect); parameters for soil, snow, vegetation, and residue; meteorological forcings (precipitation, air temperature, total solar radiation, wind speed, and relative humidity); lower boundary conditions; and initial and final states for soil moisture and temperature. Optional inputs are time series of water sources or sinks, and time series of vegetation parameters. The latter, which includes canopy height, biomass, leaf diameter, leaf area index (LAI), and effective root depth, is specified in this study.

Over the years, SHAW has been successfully used in various environments and for different applications. For example, *Flerchinger and Pierson* (1997) used SHAW at the Reynolds Creek Experimental Watershed in Idaho, to model plant canopy effects on variability of soil moisture and temperature at three soil depths. For more details about SHAW, please see *Flerchinger* (2000a,b).

A similar conceptual diagram showing the hydrologic processes represented at

each computational unit of MOBIDIC is shown in Figure 2.1. In general, the two models have a lot in common. The main difference is that MOBIDIC uses a single soil layer and since it is a catchment model, it accounts for lateral movement of water both on the surface and in the subsurface, as opposed to only vertical fluxes in the 1–D SHAW.

3.4 Correspondence Between SHAW and MOBIDIC Variables

In order to compare the soil moisture dynamics between SHAW and MOBIDIC, the parameters used in both models were selected as consistently as possible. For example, the albedo, total soil depth and saturated water content are the same for both models.

SHAW and MOBIDIC output different state variables. SHAW gives the volumetric soil moisture θ_i [-] at each soil node i, while MOBIDIC gives the equivalent water depth W [L] stored as capillary and gravity water for its single soil layer. To allow comparison, the results of the two models were converted to depth-averaged soil moisture $\overline{\theta}$ [-]. Let the superscripts 'O', 'S' and 'M', denote observed, SHAW-simulated, and MOBIDIC-simulated variables, respectively. For SHAW, $\overline{\theta^S}$ (super-script S) is the depth-weighted average of the θ_i values,

$$\overline{\theta^S} = \frac{1}{D}\sum_{i=1}^{n} \theta_i^S d_i \tag{3.1}$$

where, D is the total soil depth or the sum of the thickness of each soil layer d_i [L],

$$d_i = z_{i+1/2} - z_{i-1/2} \qquad i = 1, 2, 3, \ldots, n \tag{3.2}$$

For MOBIDIC, $\overline{\theta^M}$ (super-script M) is the sum of the equivalent depth [L] of water stored in the capillary reservoir W_c^M, the gravity reservoir W_g^M, and the time-invariant

residual water content W_r^M, normalized by D,

$$\overline{\theta^M} = (W_c^M + W_g^M + W_r^M)/D \tag{3.3}$$

For a convenient comparison with the fluxes of precipitation and ET which have units of [L], the soil moisture can also be expressed as equivalent depth,

$$\overline{W^S} = D\,\overline{\theta^S} \tag{3.4}$$

$$\overline{W^M} = D\,\overline{\theta^M} \tag{3.5}$$

Moreover, in order to show that MOBIDIC's partitioning of soil moisture into gravity-water and capillary-bound water is realistic and has a correspondence in SHAW, the total water content simulated by SHAW for the ith soil layer is partitioned into gravity water $W_{g,i}^S$, and capillary water $W_{c,i}^S$. Water in excess of the field capacity is considered free water, while water between residual water content and field capacity, is considered capillary-bound.

$$W_{g,i}^S = \begin{cases} d_i\left(\theta_i^S - \theta_{fld,i}^S\right) & \text{if } \theta_i^S > \theta_{fld,i}^S \\ 0 & \text{if otherwise} \end{cases} \tag{3.6}$$

$$W_{c,i}^S = \begin{cases} d_i\left(\theta_{fld,i}^S - \theta_{res,i}^S\right) & \text{if } \theta_i^S > \theta_{fld,i}^S \\ d_i\left(\theta_i^S - \theta_{res,i}^S\right) & \text{if } \theta_{fld,i}^S \geq \theta_i^S > \theta_{res}^S \\ 0 & \text{if otherwise} \end{cases} \tag{3.7}$$

By summing over the soil column, the total water stored in the gravity and capillary reservoirs are obtained,

$$\overline{W_g^S} = \sum_{i=1}^{n} W_{g,i}^S \tag{3.8}$$

$$\overline{W_c^S} = \sum_{i=1}^{n} W_{c,i}^S \tag{3.9}$$

For a given site, the strategy is to first develop the SHAW model. Next, the depth-averaged saturated water content of the calibrated SHAW model, $\overline{\theta_{sat}^S}$, is carried over

to the MOBIDIC model as $\overline{\theta_{sat}^{M}}$. Since the same soil depth is used for both models, this fixes the sum of MOBIDIC's $W_{g,max}$ and $W_{c,max}$, but the partitioning between the two still needs to be calibrated.

$$W_{g,max} + W_{c,max} = \overline{W_g^S} + \overline{W_c^S} \tag{3.10}$$

Once the MOBIDIC model is calibrated, its $W_{c,max}$ is used to calculate SHAW's $\theta_{fld,i}^{S}$.

$$\theta_{fld,i}^{S} = W_{c,max} \,/\, D \tag{3.11}$$

3.5 Test Sites

The comparison was performed using two sites with contrasting climatic regimes. The first site is the 'Lucky Hills' catchment in Walnut Gulch Experimental Watershed, Arizona. The climate is semiarid with two-thirds of the annual precipitation occurring during the North American Monsoon from July to September (*Goodrich et al.*, 2008; *USDA-ARS*, 2007). The site has a mild topography with deep groundwater table. The vegetation is dominated by shrubs (creosote bush or Larrea tridentata) with sparse grass (*USDA-ARS*, 2007) The soil is sandy and gravelly loam. Meteorological data and measurements of soil moisture and temperature were downloaded from the USDA-ARS Southwest Watershed Research website (http://www.wcc.nrcs.usda.gov). Soil moisture is measured at seven depths (5, 15, 30, 50, 75, 100, and 200 cm). For consistency, the SHAW model was setup with nine soil nodes with the two extra nodes located at 0 and 300 cm. A subset of the calibrated soil parameters of the SHAW model for this site is shown in Table 3.1.

The second site is the USDA Soil Climate Analysis Network (SCAN) station 'Mayday' in Yazoo, west central Mississippi (32° 52" N, 90° 31" W, elevation 33 m.a.s.l). Located on the Mississippi Delta, this site is characterized by thick clayey alluvial soil, flat topography, shallow groundwater table, and agricultural land use. Its humid subtropical climate is significantly influenced by the warm and moist air often originating from the Gulf of Mexico. In contrast to Site 1, precipitation here

Table 3.1: Calibrated soil properties of the SHAW model of Site 1. b and ψ_e are the Campbell pore-size distribution index, and air-entry potential, respectively; D_1 and D_2, are the diffusion, and dispersion parameters, respectively

z	b	ψ_e	K_s	ρ	θ_{sat}	sand	silt	clay	OM	D_1	D_2
cm	-	cm	mm/hr	kg/m3	-	%	%	%	%	-	[-]
0	5.8	-100	11.0	1380	0.19	63	22	15	1.0	2.8	.005
5	6.1	-120	10.0	1380	0.20	63	22	15	0.6	2.8	.005
15	6.1	-150	6.00	1380	0.20	63	22	15	0.5	2.8	.005
30	6.1	-200	3.00	1380	0.20	62	22	16	0.4	2.8	.005
50	6.5	-220	0.50	1420	0.21	62	22	16	0.3	2.8	.005
75	9.0	-300	0.35	1450	0.21	54	21	25	0.2	2.8	.005
100	9.5	-300	0.30	1600	0.20	53	22	25	0.1	2.8	.005
200	10.0	-300	0.25	1600	0.19	52	22	26	0.0	2.8	.005
300	10.0	-300	0.25	1600	0.19	50	22	28	0.0	2.8	.005

is almost evenly distributed throughout the year. Hourly meteorological data and measurements of soil moisture and soil temperature were downloaded from the SCAN website (http://www.wcc.nrcs.usda.gov/scan). Soil moisture and temperature are measured at five depths (5, 10, 20, 50, and 100 cm). The SHAW model was setup with eight soil nodes with the three extra nodes located at 0, 75, and 150 cm.

3.6 Calibration

The period simulated for both sites covered four years with the first being the warm-up period, the second and third as the calibration period, and the last as the validation period. The use of a real-year warm-up period greatly reduced possible errors that can be caused by incorrect initialization of the model. To guide the manual calibration, several objective and qualitative checks were performed. For both sites, the objective part of the calibration process considered the equivalent depth of water stored in the top 50 cm of soil only. Later, the calibrated model was also evaluated for the top 30 cm of soil. As additional check, modeled soil temperature timeseries were also compared.

As objective measures of model performance, the percent bias PB and the Pearson correlation coefficient R, are used as statistical measures of goodness of fit. For

soil moisture comparisons, the Nash-Sutcliffe Efficiency index NSE was also evaluated. These statistical metrics are defined in Section 2.7.2. Moreover, the time series of SHAW-simulated soil moisture at various depths were also plotted and visually compared against observations. For MOBIDIC, the hourly time series and annual total of fluxes e.g. of ET, were qualitatively checked and compared against reported values.

3.7 Results and Discussion

3.7.1 Site 1 — Lucky Hills, Arizona

The soil moisture simulated by SHAW for the Lucky Hills site is plotted against observed values in Figure 3.2. The magnitude range and temporal dynamics of θ for all of the seven nodes are consistent. Particularly, SHAW is capable of reproducing the sharp difference between the drier and more dynamic top four soil nodes (z = 5, 15, 30, 50 cm) and the wetter and less dynamic bottom three nodes (z = 75, 100, 200 cm). Notice also that during precipitation events, the top four layers become wetter than the deeper layers, a process called 'profile inversion'. This particular phenomenon cannot be captured in single-layer models such as MOBIDIC.

Next, the modeled $\overline{W^S}$ and $\overline{W^M}$ for $z = 0 - 50$cm are plotted along-side observed values in Figure 3.3. Both SHAW ($PB = -1.8\%$, $R = 0.89$, $NSE = 0.79$) and MOBIDIC ($PB = -2.3\%$, $R = 0.88$, $NSE = 0.76$) accurately reproduced the observations for the 2-year calibration period. The performance in the validation period is comparable, see Table 3.2. Moreover, the results show that MOBIDIC captured the magnitude range but more importantly the temporal dynamics that are comparable to those with SHAW. Figure 3.3 also shows the time series of observed precipitation and the MOBIDIC-simulated ET. High ET occurs around Julian Day 200–300, with maximum of about 5 mm/day. For the rest of the year, ET rarely exceeds 0.5 mm/day.

The calibrated $\theta_{sat,i}$ ranges from 0.19 to 0.21. Although low, these values are

Table 3.2: Performance of the SHAW and MOBIDIC models of Site 1 for the calibration period (year 2007 & 2008) and validation period (year 2009)

Model	Depth	Calibration			Validation		
	cm	PB	R	NSE	PB	R	NSE
		Soil Moisture					
SHAW	0–50	-1.8	0.89	0.79	3.4	0.83	0.77
MOBIDIC	0–50	-2.3	0.88	0.76	1.6	0.84	0.70
SHAW	0 30	-5.9	0.86	0.71	0.0	0.95	0.90
MOBIDIC	0–30	-1.9	0.86	0.70	2.2	0.87	0.89
		Soil Temperature					
SHAW	z_d	1.7	0.98		2.3	0.98	
MOBIDIC	z_d	-7.4	0.93		-5.9	0.93	

realistic since the site is very gravelly and rocky. Using the $W_{c,max}$ of the calibrated MOBIDIC model, the derived values of $\theta_{fld,i}$ are 0.12–0.13. To illustrate the adequacy of the dual soil compartment of MOBIDIC, the MOBIDIC-simulated W_c and W_g are plotted against the corresponding values derived from the outputs of SHAW, see Figures 3.4 a and b. Two plots are used to highlight the difference in the dynamics of the capillary-bound and gravity water. The two models with contrasting level of detail are in general agreement indicating that the MOBIDIC W_c and W_g magnitude range and temporal dynamics have correspondence in SHAW. Gravity storage is comparable in magnitude to capillary storage in this semi-humid environment. Gravity storage is filled during rain storms and it is emptied rapidly. The capillary storage water has multi-day time scales in its dynamics. In contrast to gravity storage water the recession in its dynamics lasts for months.

As a further check, the soil temperature was compared. Since MOBIDIC outputs only the ground surface temperature and the temperature at diurnal damping depth, T_d^M, the latter was used. The corresponding values from the observed and SHAW-simulated soil temperature profiles were arbitrarily defined as the temperature at the minimum depth from the surface, where the change over the course of a given day is at most 0.25 degree Celsius . The diurnal damping depth is a dynamic variable that ranges from 0 to the depth of the modeled soil. Figure 3.5 shows the very good agreement between the observed soil temperature and the values modeled by SHAW ($R = 0.98$) and MOBIDIC ($R = 0.93$).

Using the SHAW and MOBIDIC models calibrated for the top 50 cm, the performance metrics were also evaluated for $z = 0 - 30$cm. Table 3.2 summarizes the results for Site 1. The degradation of model performance in the validation period is minimal. Actually, the performance even significantly improved for soil moisture in the validation period for the top 30 cm. According to the NSE categorization of *Foglia et al.* (2009), both models performed 'very good' to 'excellent'.

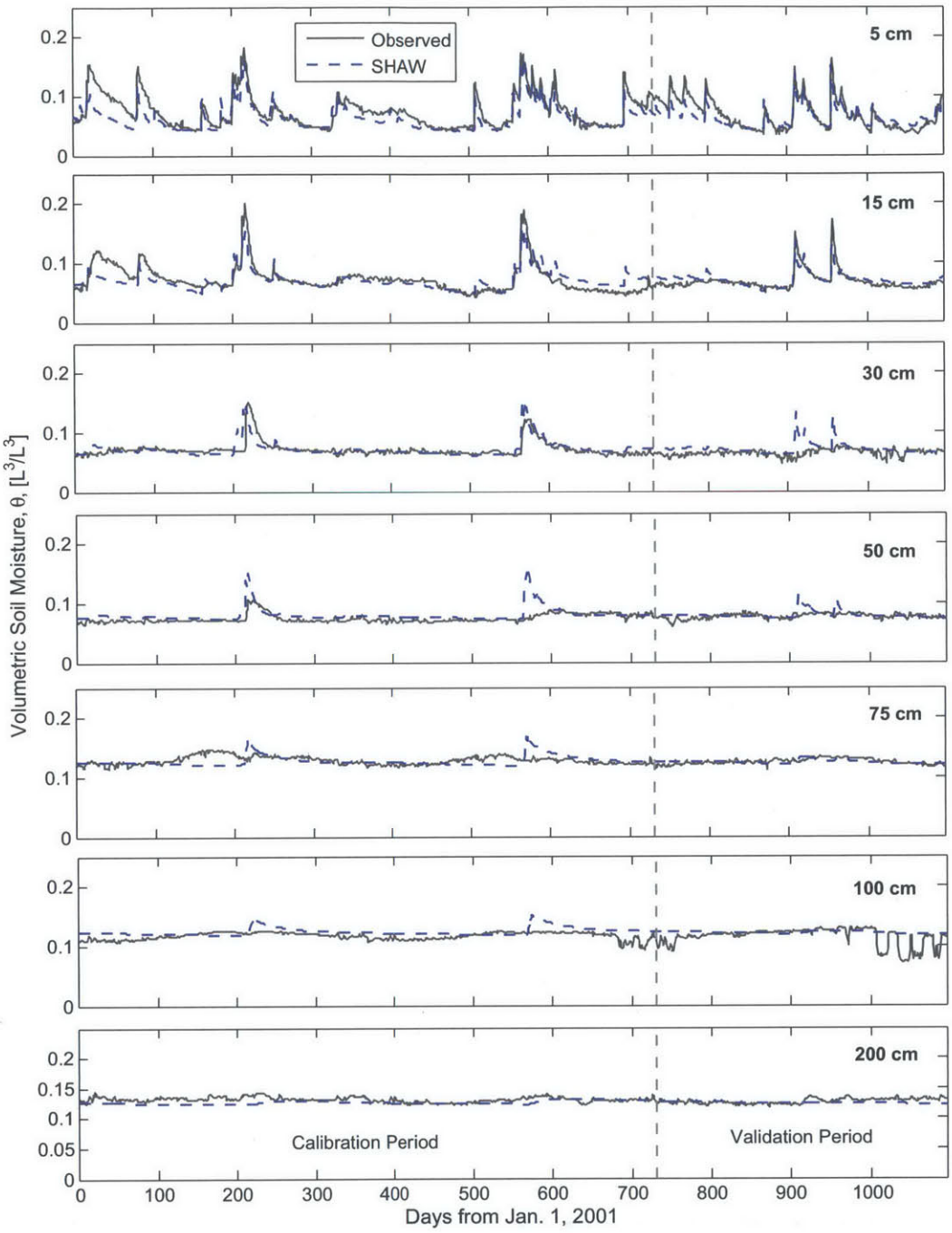


Figure 3.2: Observed vs. SHAW-simulated soil moisture at Site 1

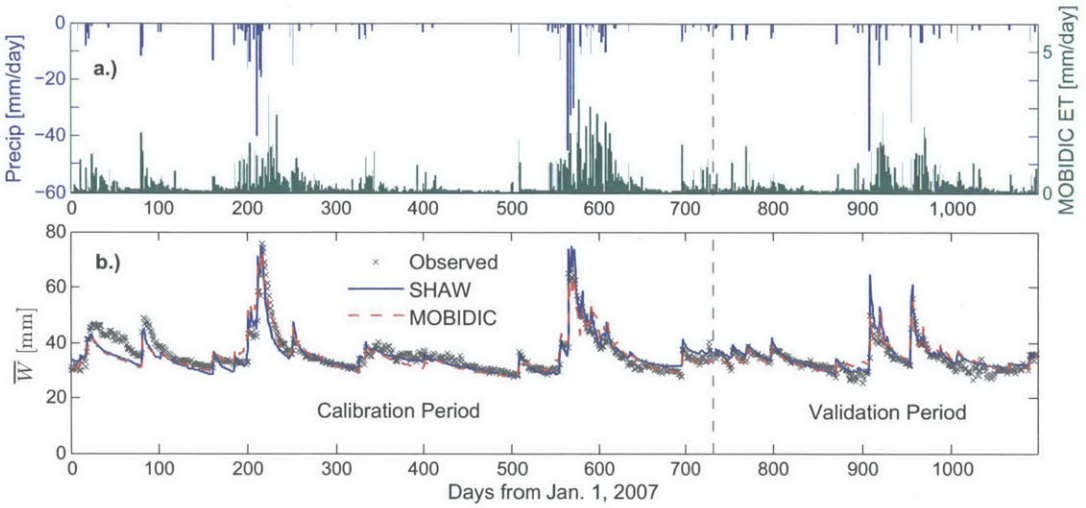


Figure 3.3: a.) Observed precipitation [mm/day] and MOBIDIC-simulated ET [mm/-day]; b.) observed equivalent depth [cm] of soil water stored in the top 50 cm vs. corresponding values simulated by SHAW and MOBIDIC for Site 1

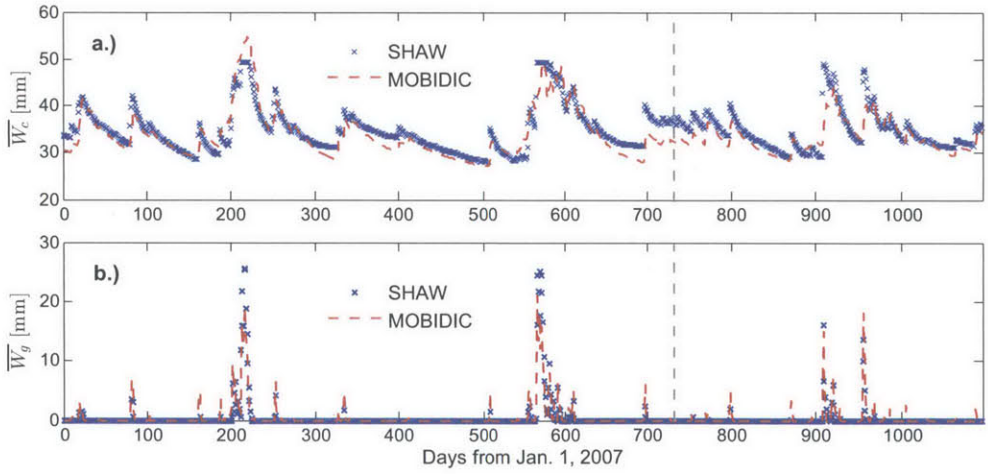


Figure 3.4: The SHAW- and MOBIDIC-simulated equivalent depth [cm] of water stored in the soil a.) capillary reservoir; and b.) gravity reservoir, for the top 50 cm of Site 1

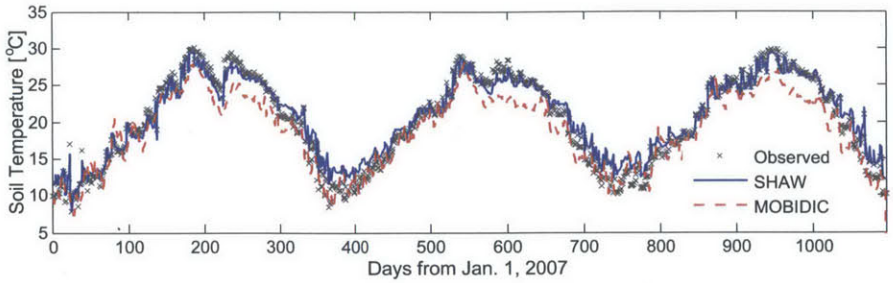


Figure 3.5: Temperature at diurnal damping depth of Site 1

3.7.2 Site 2 — Mayday, Mississippi

In contrast to the semi-arid Site 1, Site 2 is located in a humid climate region. For this site, the soil moisture simulated by SHAW (lines) are plotted alongside observations (points) in Figure 3.6. The soil moisture generally increases and becomes more stable with depth indicating the presence of a shallow water table. The soil node at $z = 50$ cm remained practically saturated during the entire simulated period. Overall, the SHAW-simulated θ_i^S time-series at the various depths track the field measurements in terms of magnitude range and temporal dynamics.

Figure 3.7a plots the time series of observed precipitation and the MOBIDIC-simulated ET. After precipitation wetting events, the evapotranspiration rate can be as high as about 12 mm/day. During the rest of the year, ET is normally 1–3 mm/day.

As mentioned earlier, the three objective measures of goodness-of-fit were evaluated using only the equivalent-depth of water stored in the top 50 cm of soil. For the 2-year calibration period, SHAW performed 'good' ($PB = -0.5\%$, $R = 0.78$, $NSE = 0.57$) while MOBIDIC performed 'very good' ($PB = 0.0\%$, $R = 0.86$, $NSE = 0.71$), see Figure 3.7b. For the validation period, both models significantly underestimated θ. As shown in Figure 3.7b, the soil column remained saturated during almost the entire validation period whereas SHAW and MOBIDIC naturally predicted the recession of θ due to ET and drainage. A possible reason for the dis-

crepancy is irrigation in upstream areas, which causes significant lateral subsurface flow and raises the groundwater table, and which is not properly accounted in the two models applied without upstream conditions.

Using the $W_{c,max}$ of MOBIDIC, the derived field capacity of the 8 SHAW nodes from top to bottom are 0.29, 0.30, 0.34, 0.37, 0.36, 0.31, 0.32, and 0.32, with a depth-averaged value of 0.33. As expected of a site with shallow groundwater table, clayey soil, and humid subtropical climate, the soil capillary reservoir remains full during non-drought years, i.e. the soil remains at or above field capacity. The fluctuation of the total soil moisture at this site is associated only with the soil gravity reservoir. Figure 3.7c shows that the MOBIDIC-simulated $\overline{W_g^S}$ and the equivalent values derived from SHAW, $\overline{W_c^S}$ track one another in both magnitude range and dynamics. Again, this indicates that MOBIDIC's dual soil compartment has correspondence in the Richards equation based SHAW model.

The measured and modeled temperature at the diurnal depth are compared in Figure 3.8. MOBIDIC has an excellent fit with $PB = 0.0$ and $R = 0.95$. SHAW has a positive bias of 23% but still a high R of 0.93.

The values of the performance metrics for soil moisture and temperature for Site 2 are summarized in Table 3.3. Similar to the findings in Site 1, the results here show that MOBIDICs simple dual pore storage model captures the essential local scale soil moisture dynamics that is comparable to those simulated with a solver like SHAW. Furthermore, the two models performed relatively better in Site 1 than in Site 2 because the former is well-represented by an independent vertical soil column, whereas in the latter, lateral subsurface fluxes and groundwater interactions are important.

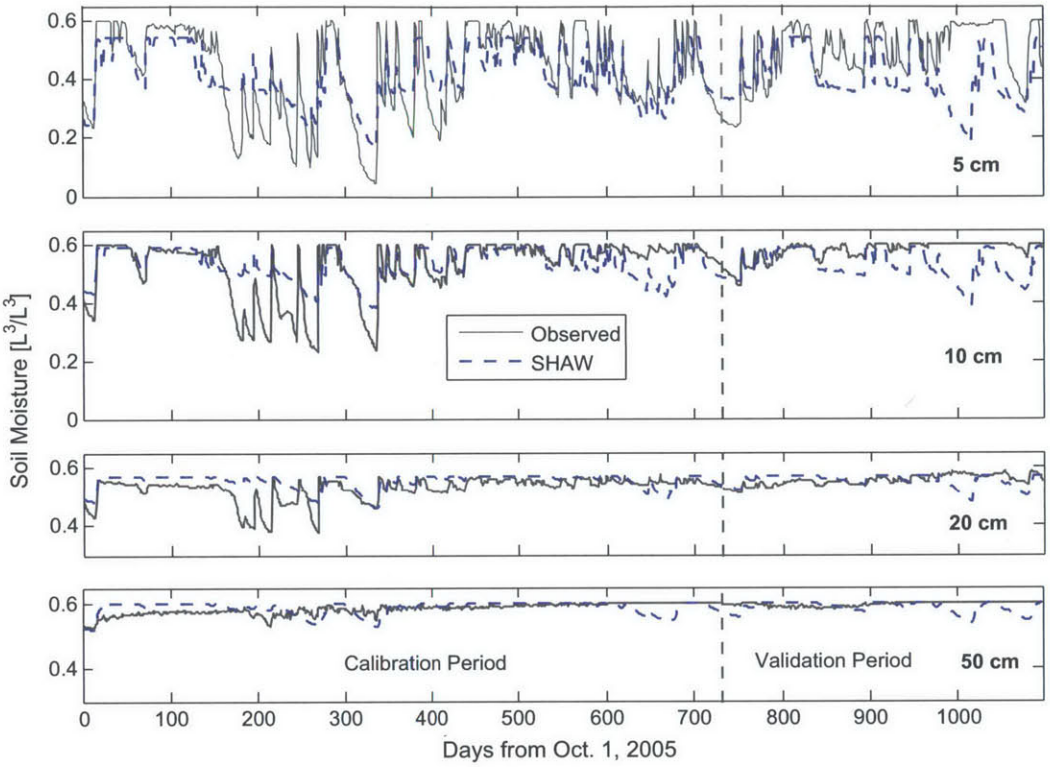


Figure 3.6: Observed vs. SHAW-simulated soil moisture at Site 2

Table 3.3: Performance of the SHAW and MOBIDIC models of Site 2 for the calibration period (water year 2006 & 2007) and validation period (water year 2008)

Model	Depth	Calibration			Validation		
	cm	PB	R	NSE	PB	R	NSE
		Soil Moisture					
SHAW	0–50	-0.5	0.78	0.57	-5.0	0.34	-0.36
MOBIDIC	0–50	0.0	0.86	0.71	-4.5	0.46	-0.34
SHAW	0–30	-0.7	0.79	0.57	-6.5	0.42	-0.17
MOBIDIC	0–30	4.9	0.82	0.46	-1.3	0.51	0.14
		Soil Temperature					
SHAW	z_d	23.2	0.93		21.9	0.92	
MOBIDIC	z_d	0.0	0.95		0.9	0.94	

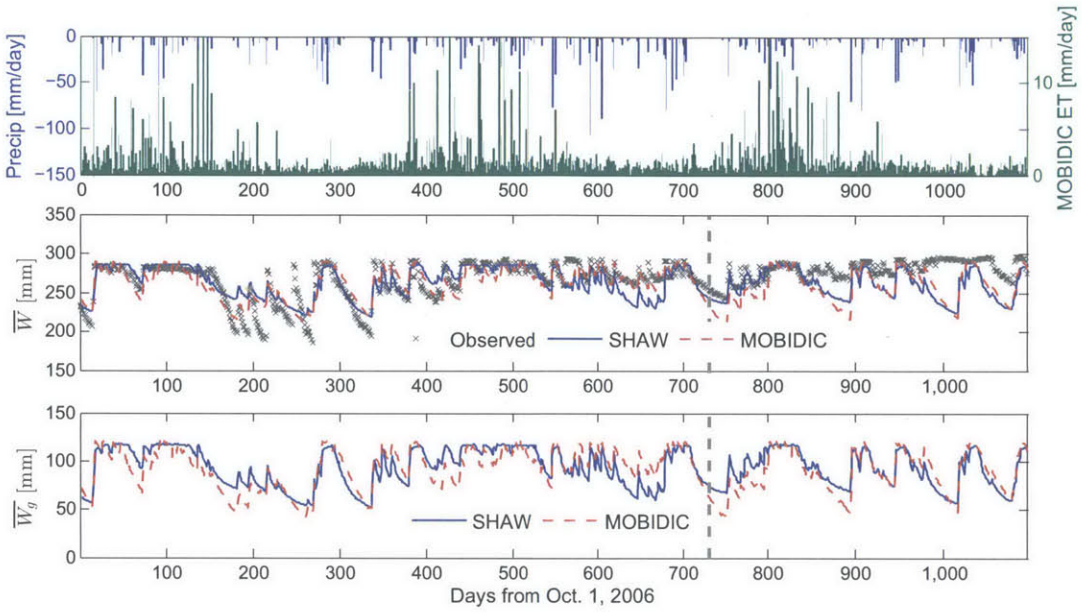


Figure 3.7: For Site 2: a.) observed precipitation [mm/day] and MOBIDIC-simulated ET [mm/day]; b.) observed, SHAW-, and MOBIDIC-simulated equivalent depth [mm] of soil water stored in the top 50 cm; and c.) MOBIDIC- and SHAW-simulated equivalent depth [mm] of water in the capillary reservoir of the top 50 cm of soil

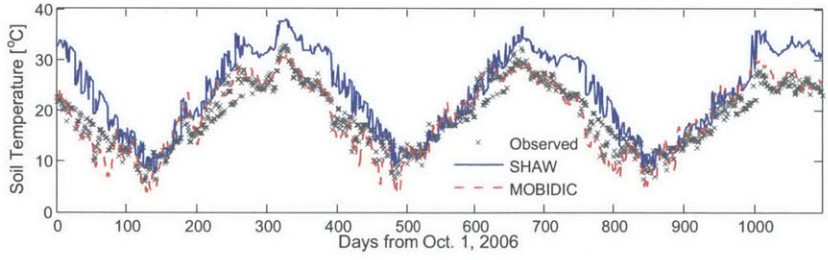


Figure 3.8: Observed temperature at diurnal damping depth vs. corresponding values simulated by SHAW and MOBIDIC for Site 2

3.8 Summary

The local scale (referring to vertical discretization of the soil column) in distributed hydrologic models is modeled using grids with millimeters to centimeters spacing. This is required for the stable and correct solution of vertical soil moisture dynamics based on Richards equation. This local scale treatment is embedded in distributed models with lateral gridding with tens to hundreds of meter scale. The distributed models are applied across entire basins. The desired applications to larger domains and in ensemble mode is limited by: 1) the computational demand of the detailed treatment of local scale processes, and 2) the number of model states that need to be initialized.

This study compared the effective performances of the two distinct approaches to the characterization of the local scale. In the detailed approach a numerical solver of the Richards equation for the vertical soil moisture dynamics (coupled to heat flow) is used. In the simpler and computationally efficient and parsimonious conceptual approach, a dual-pore characterization of a single soil unit is used. The various hydrologic fluxes act on the two reservoirs in different ways. Also an exchange flux links the two pore storages. This conceptual approach is based on physical reasoning and it is embedded in the MOBIDIC distributed hydrologic model.

The soil moisture state variables simulated by the two models are compared to field observations. The comparisons are made at two sites with contrasting climate (semiarid and sub-humid).

The parameters that can be linked between the two models are constrained to be comparable. The calibrated models are then compared with each other and the observations. At each of the two sites, the magnitude range and temporal dynamics of the gravity storage water and the capillary storage water are comparable. This result is the basis for using the simplified local scale characterization to large-domain and ensemble distributed hydrologic model applications.

Vertical structure in the soil column that is associated with horizons and parent geology cannot be captured in the dual-pore conceptual approach. The application

of models like MOBIDIC is applicable when the soil is homogeneous, or if there is limited or no information on the soil vertical stratification. Finally the role of roots and macropores cannot be captured or represented in both detailed and simplified conceptual approaches. Extensive field observations are required before an approach to these complications can be designed.

Chapter 4

An Entropy-based Measure of Hydrologic Complexity

4.1 Motivation

Basin response and hydrologic fluxes are functions of hydrologic states e.g. groundwater table and soil moisture. In this study, the focus is on soil moisture, which controls the partitioning of rainfall into infiltration and runoff, and also controls land surface temperature through its effect on the partitioning of available energy into sensible and latent heat fluxes. Based on a rich literature, the spatial distribution of soil moisture in a basin is influenced by the spatial variability of topography (*Yeh and Eltahir*, 1998; *Liu et al.*, 2012); soil properties (*Western and Grayson*, 2000; *Kim et al.*, 2002; *Famiglietti et al.*, 1999, 2008); macropore and preferential flow paths (*Brooks et al.*, 2009; *Beven and Germann*, 2013) groundwater-surface water interaction (*Levine and Salvucci*, 1999; *Kollet and Maxwell*, 2008); vegetation (*Eagleson*, 1978; *Ivanov et al.*, 2010); and meteorological forcings e.g. precipitation (*Entin et al.*, 2000; *Western et al.*, 2003; *Vivoni et al.*, 2010). Moreover, soil moisture is dynamic due to the intermittency, seasonality, and inter-annual variability of meteorological forcings (*Entekhabi et al.*, 1995; *Koster and Suarez*, 1999). Both the spatial mean and variance of soil moisture, i.e. its center of mass and distributional features, evolve over time.

Soil moisture can be expressed in several ways such as volumetric soil moisture [L^3/L^3], depth-aggregated volume per unit area [L^3/L^2], percent saturation, etc. Here the available soil water storage $\forall_e$ [L] which is particularly useful in water balance analysis, is used. Interchangeably, $\forall_e$ can be referred to as the soil water deficit. When the soil is saturated, $\forall_e = 0$, whereas when the soil is completely dry, $\forall_e = \forall_{e,max}$ which is simply the product of porosity and soil depth. Figure 4.1 shows the conceptual diagram of the two end members of the possible spatial distributions of $\forall_e$. The left is the simplest case wherein $\forall_e$ is the same across the basin. The probability density function $f(\forall_e)$ is a Dirac delta function as shown by the bottom plot. For this case, the basin can be represented by a single bucket with deterministic hydrologic response or spatially constant hydrologic fluxes. On the other extreme, the most complex case is when $f(\forall_e)$ is equally likely to take any value within its range i.e. it is uniformly distributed from 0 to $\forall_{e,max}$. For this case, the hydrologic fluxes are spatially variable and the hydrologic response is complicated. Naturally of course, a basin or a hillslope will have a spatial distribution of $\forall_e$ that is intermediate between these limiting cases. It is useful then to have a metric to quantify the distance of a given distribution of $\forall_e$ from these opposite extremes.

The marginal distribution of available soil water storage $f(\forall_e)$ evolves in time both in terms of its mean and distributional features. The specific science questions that this study aims to answer are the following:

1. How do distributional features evolve over time?
2. What factors (physiography, vegetation, climate, etc.) affect the distribution and its evolution?
3. Is there hysteresis in the evolution?

In this chapter, a dimensionless index of hydrologic complexity $\mathcal{H}$ is introduced. To develop the index, we first look at a similar index introduced by *Martina and Entekhabi* (2006) to quantify the complexity of event rainfall-runoff relationship. The theoretical basis of their index is reviewed and the sensitivity of their index to discretization is analyzed. Next, we go to improved measure of distance between two

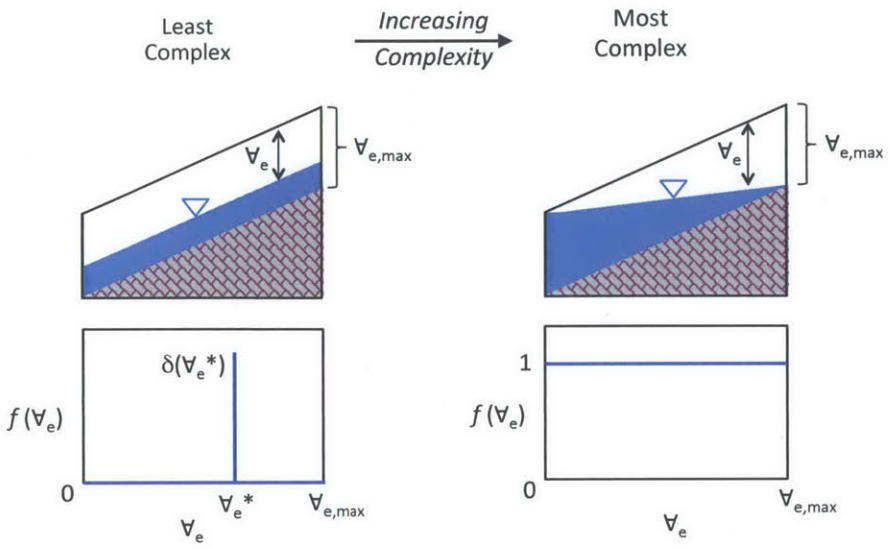


Figure 4.1: Conceptual diagram of the limiting cases: the soil water deficit is left) the same across the basin; and right) uniformly distributed from zero to a maximum value. Bottom plots show the PDF of $\forall_e$.

distributions where one is the reference distribution, which can be either of the two limiting cases previously described. However, a limitation of this approach is that both distributions must have the same support set. In hydrology, there are direct applications such as soil moisture which is naturally bounded. In this study, we use the relative soil water deficit with values ranging from 0 to 1. The effects of discretization and the universality of the new index is analyzed. This chapter concludes with an example application of the index on a real basin that demonstrates the use of the index to address the science questions.

4.2 The Hydrologic Complexity Index of *Martina and Entekhabi* (2006)

Recognizing the inadequacy of hydrographs to capture the spatial variability of processes within a basin, *Martina and Entekhabi* (2006) developed a novel method to infer the marginal probability distribution of the characteristic antecedent soil water deficit of a basin from time series of precipitation and streamflow which are often the

most readily available hydrologic data sets. The method is based on the non-linearity between storm precipitation depth and runoff volume. To quantify the complexity of the rainfall-runoff relationship, they introduced a dimensionless index of hydrologic complexity, denoted here as $\mathcal{H}_{old}$ to avoid later confusion, as:

$$\mathcal{H}_{old} = \frac{-\sum_{i=1}^{N} p_i \log(p_i)}{\log(N)} \tag{4.1}$$

where, $p = \{p_1, p_2, p_3, \ldots, p_i, \ldots, p_N\}$ are the discrete probabilities,

$$p_i = F(\forall_e)_{i+1/2} - F(\forall_e)_{i-1/2} \tag{4.2}$$

$F_{\forall_e}$ is the cumulative probability distribution, and

$$\sum_{i=1}^{N} p_i = 1 \qquad \text{for} \quad N \in \{1, 2, 3, \ldots\} \tag{4.3}$$

The numerator of Equation 4.1 is the Shannon entropy H used in the field of information theory. In other fields such as thermodynamics, H is referred to as discrete or extensive entropy (*Singh*, 2011; *Koutsoyiannis*, 2014).

$$H = -\sum_{i=1}^{N} p_i \log(p_i) \tag{4.4}$$

The intention for normalizing H by $\log(N)$ is to restrict the values of $\mathcal{H}_{old}$ to the interval [0,1] with $\mathcal{H}_{old} = 0$ for the simplest and $\mathcal{H}_{old} = 1$ for the most complex basin, respectively. This normalization was also thought to make $\mathcal{H}_{old}$ independent of numerical discretization. Although the definition was explained to be correct for the two limiting cases, no rigorous proof was given. The remainder of this section checks these claims, first for the two limiting cases then using a Beta distribution as example of a basin with intermediate complexity.

Dirac Delta

For the simplest case, as shown in Figure 4.1, the basin can be considered as a single bucket with a spatially homogeneous available soil water storage $\forall_e^*$. The corresponding probability density function (PDF) is the Dirac delta function

$$f(\forall_e) = \begin{cases} \infty & : \forall_e = \forall_e^* \\ 0 & : \forall_e \neq \forall_e^* \end{cases} \tag{4.5}$$

and the probability mass function (PMF) is

$$p(\forall_e) = \begin{cases} 1 & : \forall_e = \forall_e^* \\ 0 & : \forall_e \neq \forall_e^* \end{cases} \tag{4.6}$$

By convention, $0 \cdot \log(0) = 0$, which is consistent with

$$\lim_{p \to 0+} p \log(p) = 0 \tag{4.7}$$

Also, $\log(1) = 0$. Therefore, for this deterministic case, $\mathcal{H}_{old} = 0$ irrespective of the chosen N.

Uniform Distribution

For the most complex case wherein $\forall_e$ is uniformly distributed, the discrete probability is given by $\mathbf{p} = \{1/N, 1/N, \ldots, 1/N\}_{Nx1}$. Following the definition in Equation 4.4, the Shannon entropy for this distribution is,

$$\begin{aligned} H &= -\sum_{i=1}^{N} \left(\frac{1}{N}\right) \log\left(\frac{1}{N}\right) \\ &= -N \left(\frac{1}{N}\right) \log\left(\frac{1}{N}\right) \\ &= \log(N) \end{aligned} \tag{4.8}$$

This is the rationale of *Martina and Entekhabi* (2006) to normalize H by $\log(N)$ so that for this most complex case, $\mathcal{H}_{old} = 1$ for any $N \in \{1, 2, 3, \ldots\}$.

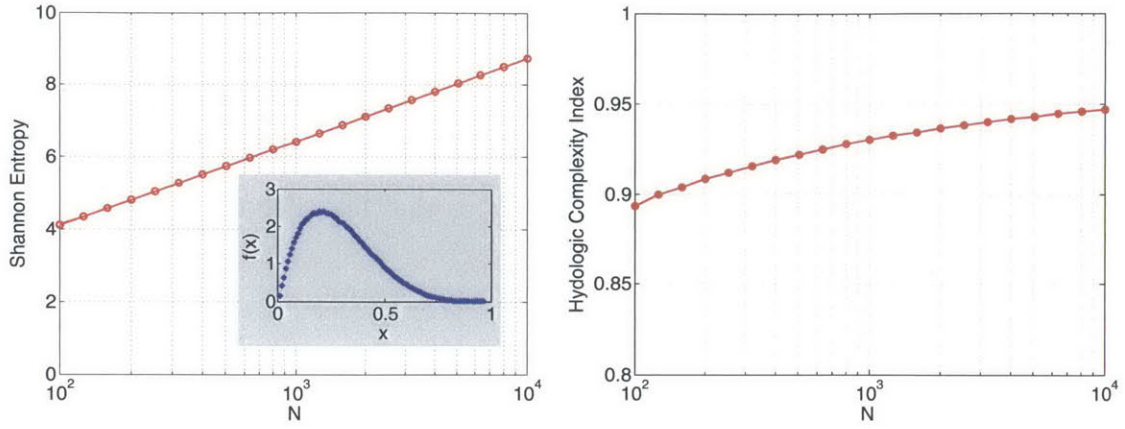


Figure 4.2: Dependence of left) the Shannon entropy H; and right) the index of hydrologic complexity $\mathcal{H}_{old}$, on discretization ($N = 1/\Delta x$). The inset shows the PDF of the relative available soil storage

Beta Distribution

Next, consider a random variable X with a beta distribution $X \sim \mathrm{B}(\alpha, \beta)$ where α and β are shape parameters that are both positive. For notation, X is the set of continuous values x or discrete values x_i. Since this distribution is defined only for $x \in [0, 1]$, let X be the relative available soil water storage [-]

$$X = \frac{\forall_e}{\forall_{e,max}} \tag{4.9}$$

The PDF and PMF of this distribution are given by Equations 4.10 and 4.11, respectively.

$$f(x) = \frac{x^{\alpha-1}\,(1-x)^{\beta-1}}{\mathrm{B}(\alpha, \beta)} \tag{4.10}$$

$$\begin{aligned} p_i &= p(x_i) \\ &= \int_{(i-1/2)\Delta x}^{(i+1/2)\Delta x} f(x)\,dx \\ &= f(x_i)\,\Delta x \end{aligned} \tag{4.11}$$

For illustration, consider $\alpha = 2$ and $\beta = 5$. Using this distribution, 10^6 discrete values were generated then binned into N equal intervals. The N values tried range from 10^2 to 10^4 which is equivalent to using $\Delta x = 10^{-4}$ to 10^{-2}. Figure 4.2 shows

that both the Shannon entropy and the index of hydrologic complexity based on the old definition, increase with N. The increase in H is intuitive because the entropy or uncertainty increases with the number of possible microstates or outcomes. What is less intuitive is that normalizing H by $\log(N)$ still does not make $\mathcal{H}_{old}$ invariant to numerical discretization.

4.3 Development of a Revised Measure of Hydrologic Complexity

4.3.1 Shannon Entropy

To fully understand the reason for the discretization problem just discovered and hopefully find a solution, the concept of entropy will be discussed in more details. For a discrete random variable X with potential states $\{x_1, x_2, \ldots, x_i, \ldots, x_N\}$ and PMF $p(x)$, the discrete or Shannon entropy as defined in Equation 4.4, can be expressed more generally as,

$$\begin{aligned} H(X) &= \sum_{i=1}^{N} p(x_i) \log \left[\frac{1}{p(x_i)}\right] \\ &= -\sum_{i=1}^{N} p(x_i) \log(p(x_i)) \end{aligned} \tag{4.12}$$

It can also be expressed as the expected value of the self information,

$$H(X) = E[-\log p(x)] \tag{4.13}$$

The base of the log function is arbitrary but e is adopted here, thus entropy has units of "nats". Since $0 \leq p(x) \leq 1$, it follows that the Shannon entropy is non-negative and it is equal to 0 only in the degenerate case that X has a singular value with probability of 1.

$$H(X) \geq 0 \tag{4.14}$$

Based on e.g. *Michalowicz et al.* (2014), some important properties of Shannon

entropy for arbitrary variables X and Y, and constant k, are:

1. It is transitive and symmetric.

$$H(p_1, p_2, \ldots) = H(p_2, p_1, \ldots) \tag{4.15}$$

2. It is invariant to translation.

$$H(X + k) = H(X) \tag{4.16}$$

3. Multiplication with a constant results to an additive constant.

$$H(kX) = H(X) + \log k \tag{4.17}$$

4. It is maximal if all possible states are equiprobable.

$$H(p_1, p_2, \ldots, P_N) \leq H(1/N, 1/N, \ldots, 1/N) \tag{4.18}$$

5. The joint entropy (or uncertainty) in X and Y is no more than the individual entropies, and are equal only if X and Y are statistically independent,

$$H(X, Y) \leq H(X) + H(Y) \tag{4.19}$$

6. The entropy (or uncertainty) in X cannot increase once the state (or outcome) or Y is known,

$$H(X|Y) \leq H(X) \tag{4.20}$$

4.3.2 Differential Entropy

Differential entropy (also known as continuous entropy) extends the concept of Shannon entropy to continuous probability distributions. For a continuous random variable X with PDF $f(x)$, the *differential entropy* is defined as (*Michalowicz et al.*, 2014),

$$h(X) = -\int_{\mathbb{X}} f(x) \log f(x) dx \tag{4.21}$$

where $\mathbb{X} = \{x | f(x) > 0\}$ is the support set of X.

Although Equation 4.21 seems to be a natural extension of Equation 4.13, the transition from the discrete to the continuous case or vice versa must be handled very carefully. To explain this, consider a random variable X with PDF $f(x)$, and let X^{Δ} be a discrete random variable where the probability of the outcome x_k^{Δ} is

$$p(x_k^{\Delta}) = \int_{k\Delta}^{(k+1)\Delta} f(x) dx = \Delta f(x_k) \tag{4.22}$$

The entropy of this discrete variable is

$$\begin{aligned} H(X^{\Delta}) &= -\sum_k p(x_k^{\Delta}) \log p(x_k^{\Delta}) \\ &= -\sum_k \Delta f(x_k) \log \left[\Delta f(x_k)\right] \\ &= -\Delta \sum_k f(x_k) \log f(x_k) - \left[\Delta \sum_k f(x_k)\right] \log \Delta \end{aligned} \tag{4.23}$$

Taking the limit as $\Delta \to 0$, the first term on the right is the differential entropy $h(X^{\Delta})$. Also, $\Delta \sum_k f(x_k) = 1$, thus

$$H(X^{\Delta}) = h(X^{\Delta}) - \log \Delta \tag{4.24}$$

Or rearranging,

$$h(X^{\Delta}) = H(X^{\Delta}) + \log \Delta \tag{4.25}$$

Therefore, differential entropy is NOT the same as discrete entropy. Unlike discrete

entropy, differential entropy can be negative. Particularly, as $\Delta \to 0$, $\log\Delta \to -\infty$, so as long as the discrete entropy is finite, the differential entropy is negative. Because of this, the interpretation of entropy as a measure of uncertainty in the discrete case cannot be used for differential entropy.

$$h(X) \in (-\infty, +\infty) \tag{4.26}$$

4.3.3 Kullback-Liebler Divergence

A modification of differential entropy that addresses its shortcomings is the Kullback-Liebler divergence (D_{KL}) also known as relative entropy, information divergence, or information gain (*Michalowicz et al.*, 2014).

For Continuous Random Variables

For a pair of continuous PDFs $f(x)$ and $g(x)$, D_{KL} is defined as,

$$D_{KL}\left(f||g\right) = \int_{S_f} f(x)\log\frac{f(x)}{g(x)}dx \tag{4.27}$$

where, S_f is the support set of $f(x)$. An important property of D_{KL} is that it is always non-negative with $D_{KL} = 0$ if and only if $f(x)$ and $g(x)$ have the same support set and$f(x) = g(x)$ everywhere in the support set. With the goal of obtaining a dimensionless index of hydrologic complexity, let the reference distribution $g(x)$ be uniformly distributed in the interval [a,b],

$$g(x) = \begin{cases} \dfrac{1}{b-a} & \text{for} \quad a \leq x \leq b \\ 0 & \text{otherwise} \end{cases} \tag{4.28}$$

By the definition in Equation 4.27,

$$D_{KL}(f||g) = \int_{S_f} f(x)\log\frac{f(x)}{\frac{1}{b-a}}\,dx \tag{4.29}$$

$$= \int_{S_f} f(x)\left[\log f(x) + \log(b-a)\right]\,dx \tag{4.30}$$

$$= \int_{S_f} f(x)\log f(x)dx + \log(b-a)\int_{S_f} f(x)\,dx \tag{4.31}$$

Since $\int f(x)dx = 1$,

$$D_{KL}(f||g) = \int_{S_f} f(x)\log\left(f(x)\right)dx \;+\log(b-a) \tag{4.32}$$

The integral term is the negative of the differential entropy of $f(x)$ as shown in Equation 4.21, thus,

$$D_{KL}(f||g) = -h(X) + \log(b-a) \tag{4.33}$$

If the support set $S_f = [0,1]$, i.e. $a = 0$ and $b = 1$, the previous equation simplifies to

$$D_{KL}(f||g) = -h(X) \qquad \text{for } x \in [0,1] \text{ and } g(x) \sim \mathrm{U}[0,1] \tag{4.34}$$

By definition, differential entropy is independent of Δx or N. Therefore, Equation 4.34 means that for a random variable X with PDF $f(x)$ such that $x \in [0,1]$ and $\int_0^1 f(x)dx = 1$, then $D_{KL}(f||g)$ *is also independent of* Δx or N. Moreover, because $D_{KL} \geq 0$, the negative of the differential entropy for this case is also non-negative i.e. $-h(X) \geq 0$.

For Discrete Random Variables

For completeness, consider discretizing the continuous PDFs $f(x)$ and $g(x)$ to generate their equivalent PMFs $p(x)$ and $q(x)$. The KL-divergence of $p(x)$ from the

reference $q(x)$ is,

$$D_{KL}{}^{[\Delta x,\delta x]}\left(p||q\right) = \sum_k p(x_k)\log\frac{p(x_k)}{q(x_k)} \tag{4.35}$$

Using the discretization Δx and δx, respectively, the probability of discrete outcomes can be expressed as,

$$p(x_k^{\Delta x}) = \int_{k\,\Delta x}^{(k+1)\Delta x} f(x)dx = \Delta x\, f(x_k) \tag{4.36}$$

$$q(x_k^{\delta x}) = \int_{k\,\delta x}^{(k+1)\delta x} g(x)dx \;= \delta x\, g(x_k) \tag{4.37}$$

Substituting Equations 4.36 and 4.37 into Equation 4.35,

$$D_{KL}{}^{[\Delta x,\delta x]}\left(p||q\right) = \sum_k \Delta x\, f(x_k)\log\frac{\Delta x\, f(x_k)}{\delta x\, g(x_k)} \tag{4.38}$$

If $\Delta x = \delta x$, then

$$D_{KL}\left(p||q\right) = \sum_k \Delta x\, f(x_k)\log\frac{f(x_k)}{g(x_k)} \quad \text{if } \Delta x = \delta x \tag{4.39}$$

As done for the continuous case, let $g(x)$ be uniformly distributed in the interval [a,b], recall Equation 4.28. Then,

$$D_{KL}\left(p||q\right) = \sum_k \Delta x\, f(x_k)\log\frac{f(x_k)}{\frac{1}{b-a}} \tag{4.40}$$

$$= \sum_k \Delta x\, f(x_k)\log f(x_k) + \log\left(b-a\right)\sum_k \Delta x\, f(x_k) \tag{4.41}$$

As $\Delta \to 0$, the first term is the negative of differential entropy, while for the second term, $\sum_k \Delta x\, f(x_k) = \sum_k p(x_k) = 1$. So,

$$D_{KL}\left(p||q\right) = \int f(x_k)\log f(x_k)\,dx + \log\left(b-a\right) \tag{4.42}$$

$$= -h(X) + \log\left(b-a\right) \tag{4.43}$$

Finally, if the support set of X is the interval [0,1], i.e. $a = 0$ and $b = 1$, then $\log(b - a) = 0$ and the result is Equation 4.44. Again, the KL-divergence is equal to the negative of the differential entropy which is independent of discretization.

$$D_{KL}(p||q) = -h(X) \qquad \text{for } x \in [0,1] \text{ and } q(x) \sim \mathrm{U}[0,1] \tag{4.44}$$

To summarize, given a dimensionless random variable X defined in the interval [0,1], the KL-divergence or relative entropy of either its continuous PDF $f(x)$ or its discrete PMF $p(x)$, with respect to a uniform distribution in the interval [0,1], is equal to the negative of the differential entropy of X and is independent of Δx or N.

As mentioned, since $D_{KL} \geq 0$ it follows that $-h(X) \geq 0$. However, the value of D_{KL} ranges from 0 to $+\infty$. By taking the exponential of $-D_{KL}$, the value can be constrained in the interval [0,1],

$$\begin{gathered} 0 \leq \exp\left[-D_{KL}(f||g)\right] \leq 1 \\ \text{for} \quad x \in [0,1], \quad \int_0^1 f(x)dx = 1, \quad \text{and} \quad g(x) \sim \mathrm{U}[0,1] \end{gathered} \tag{4.45}$$

Or equivalently for the discrete case,

$$\begin{gathered} 0 \leq \exp\left[-D_{KL}(p||q)\right] \leq 1 \\ \text{for} \quad x \in [0,1], \quad \sum_k x_k = 1, \quad \text{and} \quad q(x) \sim \mathrm{U}[0,1] \end{gathered} \tag{4.46}$$

4.3.4 Definition of the Revised Measure of Complexity

Now, we formally define $\mathcal{H}$ **as a dimensionless measure of complexity** of a dimensionless random variable X with distribution $f(x)$, for $x \in [0,1]$ and $\int_0^1 f(x)dx = 1$, as

$$\mathcal{H} = \exp\left[-\int f(x)\log f(x)\,dx\right] \tag{4.47}$$

Or equivalently, using Equations 4.21, 4.34, or 4.44,

$$\mathcal{H} = \exp\left[h(X)\right] \qquad \text{for } x \in [0,1] \tag{4.48}$$

$$= \exp\left[-D_{KL}\left(f||g\right)\right] \quad \text{for } x \in [0,1] \text{ and } g(x) \sim \mathrm{U}[0,1] \tag{4.49}$$

$$= \exp\left[-D_{KL}\left(p||q\right)\right] \quad \text{for } x \in [0,1] \text{ and } q(x) \sim \mathrm{U}[0,1] \tag{4.50}$$

In general, X can be any dimensionless random variable. In hydrology, the limitation that X be in the interval [0,1] is not a major drawback since many hydrologic variables are naturally bounded. For instance, soil moisture is bounded between zero[1] and saturated soil moisture content, and normalizing by the latter yields relative soil saturation or relative soil water deficit with values 0 to 1. In this thesis, unless otherwise specified $\mathcal{H}$ is computed using the relative available soil water storage, $\forall_e/\forall_{e,max}$ [-], in place of X.

$\mathcal{H}$ is dimensionless and takes on values from 0 to 1, with $\mathcal{H} = 0$ for the simplest, and $\mathcal{H} = 1$ for the most complex case, respectively. As check, if $f(x)$ is a Dirac delta (deterministic case or single-bucket), $D_{KL} = \infty$, so $\mathcal{H} = 0$. On the other hand, if $f(x)$ is a uniform distribution, $D_{KL} = 0$, so $\mathcal{H} = 1$. Figure 4.3 shows the value of $\mathcal{H}$ computed for these pair of limiting cases, plus two example distributions with intermediate complexity. As shown, the value of $\mathcal{H}$ is practically invariant to the number of bins used N, or Δx since $\Delta x = 1/N$. For direct comparison with $\mathcal{H}_{old}$, the third example (bottom right) uses the same distribution used in Figure 4.2.

As defined in Equation 4.47, $\mathcal{H}$ should be invariant to N or Δx. However, in actual numerical implementation, the continuous PDF $f(x)$ will be generated or derived from a finite set of discrete values e.g. values from an observational network, or values simulated for the nodes or pixels of a distributed hydrologic model. For instance, the distributions used in Figure 4.3 were generated from 10^6 discrete values. For these non-peaky distributions, it was observed that $\mathcal{H}$ is practically the same for $N = 10^1 - 10^5$. For peaky distributions such as shown in Figure 4.4, the choice of N

[1] In other applications, the wilting point or the residual soil moisture content can also be used as the lower bound, in which case the resulting dimensionless variable is the effective soil saturation.

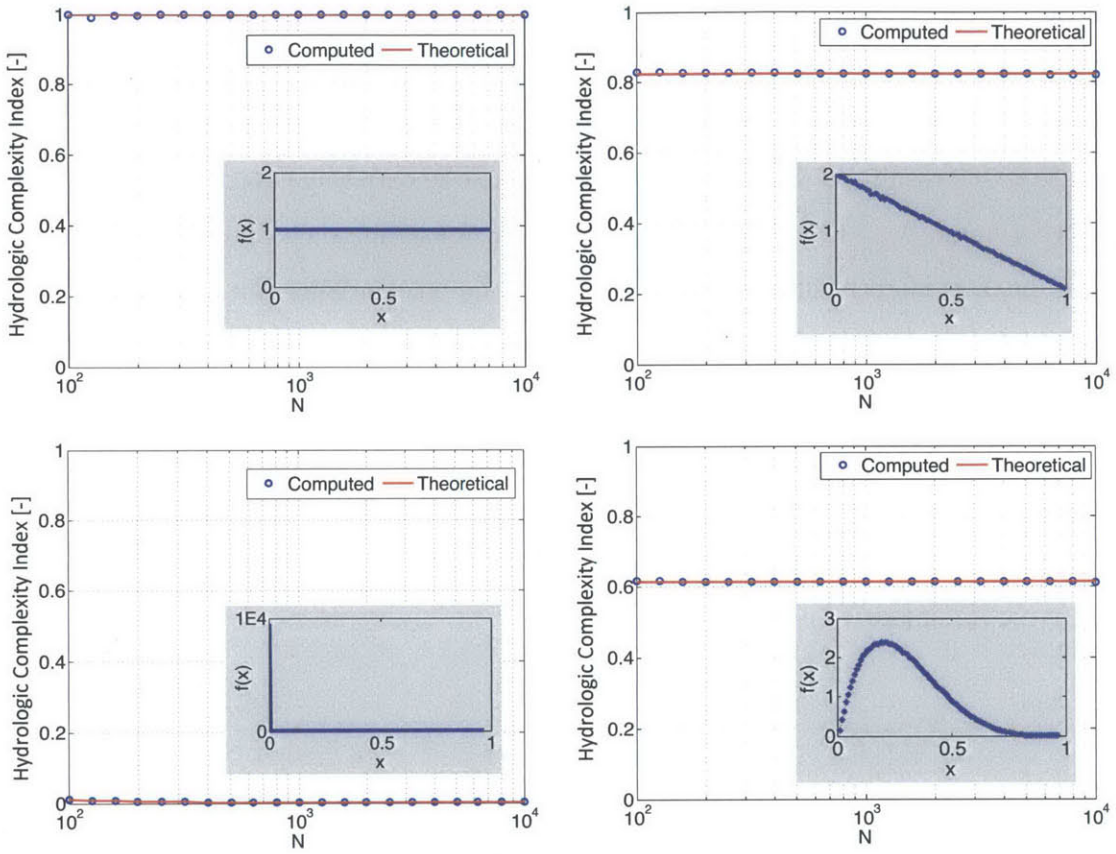


Figure 4.3: The revised dimensionless hydrologic complexity index $\mathcal{H}$ is independent of numerical discretization (N = number of bins). The insets show the PDFs used.

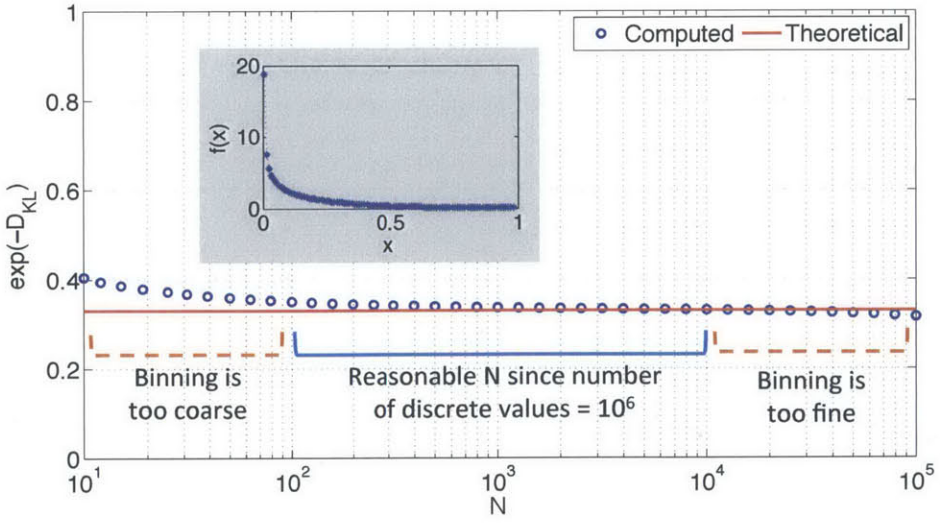


Figure 4.4: Effect of using unreasonable number of bins N on $\mathcal{H}$ for peaky distributions. Inset shows the PDF of the 10^6 discrete values of X.

can be critical. Nevertheless, as shown, as long as N is reasonable given the number of discrete values used, the computed $\mathcal{H}$ should be close to the theoretical value. For this example, the number of discrete points used is 10^6 so a reasonable N is 10^2 to 10^4. $N < 100$ results to a binning that is too coarse to correctly capture the peak. On the other hand, $N > 10^4$ results to a binning that is too fine such that the significance of the peak is slightly diminished. Moreover, the error arising from the use of too few bins is relatively more significant than using too many. However, this points a limitation of the use of $\mathcal{H}$ which is the need for a high number of discrete values e.g. $> 10^3$ so that a reasonable N e.g. $> 10^2$ can be used. This implies the need for long timeseries if the analysis is over time, or many points if the analysis is over space.

For reference, a sample MATLAB™code for computing $\mathcal{H}$ of a discrete data set 'Xdata' is provided.

```
%% Sample MATLAB code to compute hydrologic complexity index

N = 100                          ; % number of bins
dx = 1/N                         ; % width of a bin
xbin = 0+dx/2 : dx : 1-dx/2      ; % midpoints of bins

p = hist(Xdata,xbin)             ; % histogram
p(p==0) = []                     ; % ensures p*log(p)=0
p = p ./ sum(p)                  ; % PMF

f = p ./dx                       ; % Convert PMF to PDF

DKL = sum(f.*log(f))*dx          ; % KL-divergence

H = exp(-DKL)                    ; % hydrologic complexity index
```

4.4 Application of the Measure

In this section, for the purpose of demonstration only, $\mathcal{H}$ is applied to MOBIDIC-simulated timeseries of soil moisture fields for a 103 sq. km. mountainous tropical rain forest basin in southern Philippines. The modeled domain has 12,000 pixels and the simulated period is 4-year long, originally at hourly timestep then aggregated to

daily values. Detailed description of the basin and the model can be found in the previous chapter.

4.4.1 Example 1

$\mathcal{H}$ is evaluated using the relative soil water deficit, $x = \forall_e/\forall_{e,max}$. Since MOBIDIC uses a dual-compartmentalized soil, we can also look at the *individual complexity* of the soil capillary and gravity reservoirs, $\mathcal{H}_c$ and $\mathcal{H}_g$, respectively.

$$\begin{aligned} \mathcal{H} &= \mathcal{H}(x), & x &= \forall_e/\forall_{e,max} \\ \mathcal{H}_c &= \mathcal{H}_c(x_c), & x_c &= \forall_{e,c}/\forall_{e,c,max} \\ \mathcal{H}_g &= \mathcal{H}_g(x_g), & x_g &= \forall_{e,g}/\forall_{e,g,max} \end{aligned} \tag{4.51}$$

In this example, x_c and x_g are the relative water deficit of the soil capillary and gravity reservoirs, respectively. Note that $\forall_{e,max} = \forall_{e,c,max} + \forall_{e,g,max}$, and the range of x, x_c and x_g is the interval [0,1]. Later, a second example will use a different definition of x_c and x_g based on their contribution to the total soil water deficit.

The top panel of Figure 4.5 shows the hyetograph and the evolution of $\mathcal{H}$, $\mathcal{H}_c$, and $\mathcal{H}_g$. Also shown are the PDFs $f(x)$, $f(x_c)$, and $f(x_g)$, for three representative days. The timeseries only displays the results for WY2010 which was chosen for illustration as it includes an unusually dry period from February to June 2010. Several observations can be made from this plot. First, consider only $\mathcal{H}$ shown as red line. It has a median of 0.16 and mean of 0.19, which are closer to 0 than to 1, indicating that overall the basin has low hydrologic complexity. During a typical year, $\mathcal{H}$ only fluctuates a little around the median or characteristic value. In fact, the timeseries of $\mathcal{H}$ for the other 3 years is very similar to that shown for June to September.

Now consider also the individual complexity of the capillary and gravity reservoirs. When the basin is wet as is usual for this humid basin, see point 3, most of the capillary reservoirs are saturated so $f(x_c)$ is close to a Dirac delta at $x_c = 0$, and thus $\mathcal{H}_c \approx 0$.

A physical interpretation is that fluxes related to the state of the capillary reservoirs, such as ET is least complex, in this case ET occurs at the potential rate across the basin. Meanwhile, since this basin is well-drained, the gravity reservoirs never become completely saturated so x_g and x are spatially variable, and as a result $\mathcal{H}_g > 0$ and $\mathcal{H} > 0$. $\mathcal{H}$ is associated almost entirely on $\mathcal{H}_g$. However, because of the definition of x_g in Equation 4.51, the relative complexity $\mathcal{H}_g$ is not numerically equivalent to $\mathcal{H}$.

In dry conditions (see point 1), the gravity reservoirs are empty so $f(x_g)$ is close to a Dirac delta at $x_g = 1$, which yields $\mathcal{H}_g \approx 0$. A physical interpretation is that fluxes related to the state of the gravity reservoirs, such as percolation, and lateral subsurface flow are least complex, in this case there are no fluxes. Meanwhile, x and x_c are spatially variable so $\mathcal{H}$ and $\mathcal{H}_c$ are positive. While it appears that for this dry regime $\mathcal{H}_c$ has roughly the same magnitude as H, they are also not numerically equivalent. For reference the mean of $\forall_{e,c,max}$ is 262 mm and the mean of $\forall_{e,g,max}$ is 151 mm.

Comparing points 1 and 3, it can be seen that $f(x_c)$ in the dry regime has a wider spread than $f(x_g)$ in the wet regime, thus explaining why for this basin $\mathcal{H}$ is generally higher when the basin is unusually dry. Point 2 shows an intermediate condition wherein the total entropy of the system is attributed to the combined entropy of both soil reservoirs.

Figure 4.6 shows $\mathcal{H}$, $\mathcal{H}_c$ and $\mathcal{H}_g$ vs. the relative soil water deficit (all vs. total deficit x). As can be seen, $\mathcal{H}_c \approx 0$ when the basin is wet because the capillary reservoirs across the basin are saturated. $\mathcal{H}_c$ begins to increase at $x \approx 0.35$. $\mathcal{H}_g$ has opposite trend with $\mathcal{H}_g \approx 0$ when the basin is dry since the gravity reservoirs are mostly empty. $\mathcal{H}_g$ does not exactly go to 0 because the areas along drainage lines remain wet. $\mathcal{H}_g$ is high when the basin is wet since there is spatial variation in $\forall_{e,g}/\forall_{e,g,max}$. The total complexity $\mathcal{H}$ is the combined effect of $\mathcal{H}_c$ and $\mathcal{H}_g$, however as pointed out earlier, $\mathcal{H} \neq \mathcal{H}_c + \mathcal{H}_g$ Since the basin is usually wet, $\mathcal{H}$ is most of

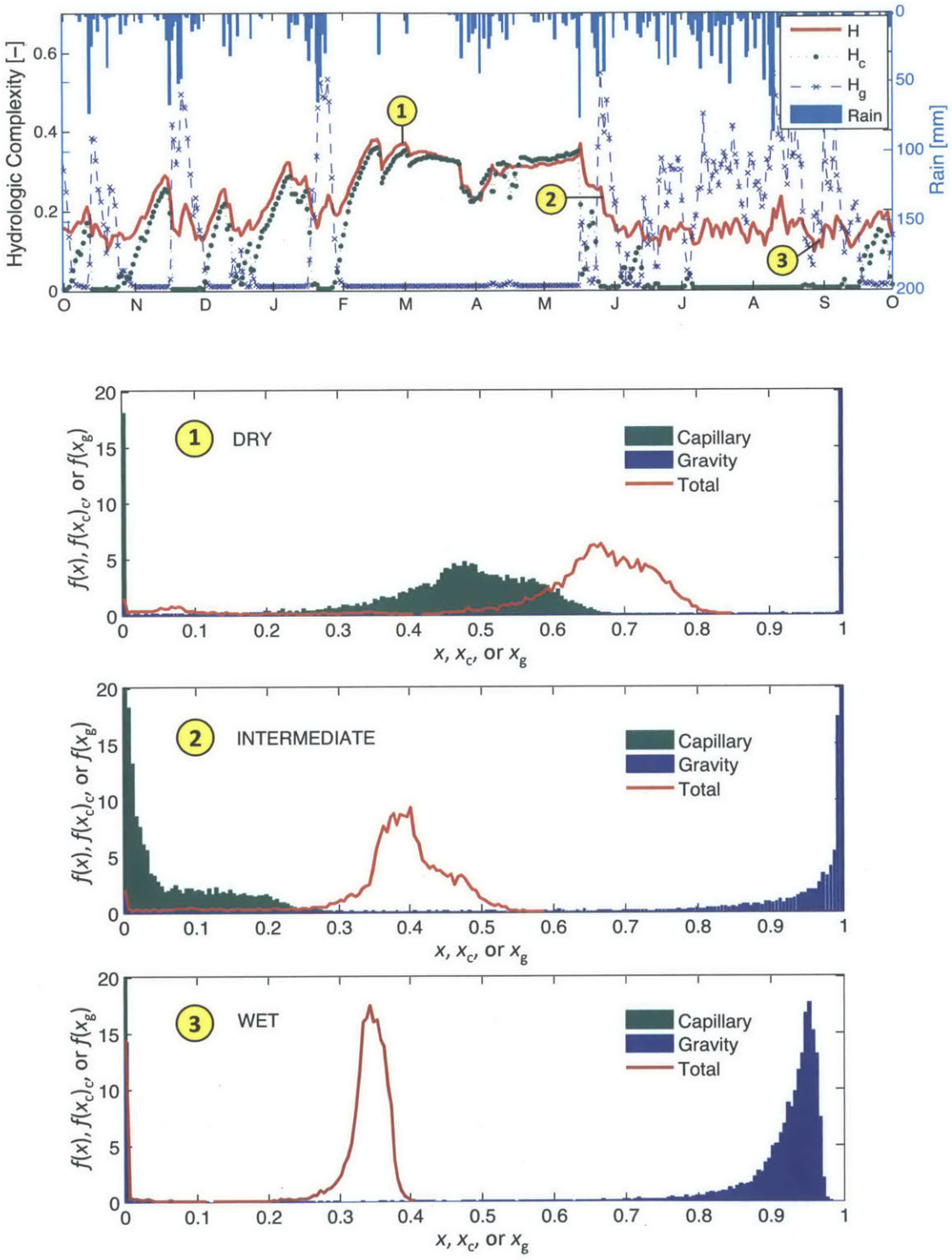


Figure 4.5: Time series of the hydrologic complexity of the basin, the capillary reservoir, and the gravity reservoir. Also shown are three snapshots showing the spatial PDF of relative soil water deficit in the dry, intermediate, and wet regimes.

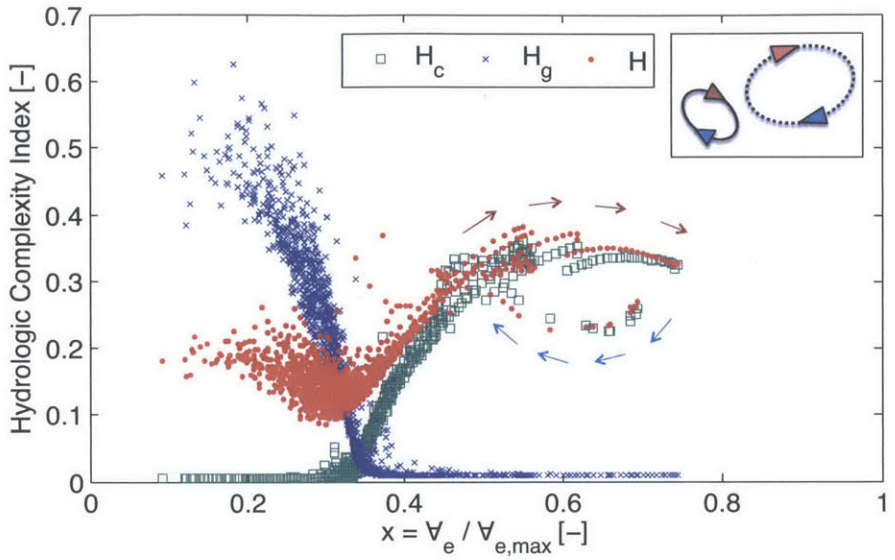


Figure 4.6: The hydrologic complexity index of the soil capillary reservoirs, soil gravity reservoirs, and the (combined) soil reservoirs, vs. the relative soil water deficit. Inset shows hysteresis. Red and blue arrows indicate drying and wetting directions, respectively.

the time attributed only to $\mathcal{H}_g$. Furthermore, the basin does not become completely saturated nor dry but since soil moisture is bounded, $\mathcal{H} = \mathcal{H}_c = \mathcal{H}_g = 0$ at both ends of the support. Finally, the looping behavior of $\mathcal{H}$ in the dry regime clearly indicates *hysteresis* at the basin scale. Although it is not as clear from the scatter plot, a looping behavior was also observed for the wet regime when the points were plotted sequentially.

4.4.2 Example 2

This example is the same as the first one except for different definitions of x_c and x_g,

$$\begin{aligned} \mathcal{H} &= \mathcal{H}(x)\,, \qquad x = \forall_e/\forall_{e,max} \\ \mathcal{H}_c &= \mathcal{H}_c(x_c)\,, \qquad x_c = \forall_{e,c}/\forall_{e,max} \\ \mathcal{H}_g &= \mathcal{H}_g(x_g)\,, \qquad x_g = \forall_{e,g}/\forall_{e,max} \end{aligned} \tag{4.52}$$

Notice the use of the same normalizer $\forall_{e,max}$. This definition, as will be seen shortly, is better for understanding the contribution of $\mathcal{H}_c$ and $\mathcal{H}_g$ to the total complexity. The support sets of x_c and x_g are still *contained in the interval* [0,1] so the definition of the measure is still valid, but the maximum values of x_c and x_g are now < 1,

$$\begin{aligned} 0 \leq \; & x \; && \leq 1 \\ 0 \leq \; & x_c \; && \leq \max\left(\forall_{e,c,max}/\forall_{e,max}\right) \; < 1 \\ 0 \leq \; & x_g \; && \leq \max\left(\forall_{e,g,max}/\forall_{e,max}\right) \; < 1 \end{aligned} \tag{4.53}$$

The new results are shown in Figures 4.7 and 4.8. Since the definition of x is unchanged, the overall complexity $\mathcal{H}$ and its interpretations are also unchanged. However, the magnitude, range, and interpretations of x_c, x_g, $\mathcal{H}_c$, and $\mathcal{H}_g$, are different. From the time series in Figure 4.7, it can be seen that the dynamics (trends) of $\mathcal{H}_c$, and $\mathcal{H}_g$ are the same as before but their magnitudes are reduced. This is because the support sets of x_c and x_g are now just subsets of the interval [0,1]. In other words, the maximum of $\mathcal{H}_c$ and $\mathcal{H}_g$ are < 1. However, one of the advantages of the new definitions of x_c and x_g is that now, $\mathcal{H}$ is roughly equal to $\mathcal{H}_c + \mathcal{H}_g$.

In wet conditions (e.g. point 3), the PDFs of x and x_c are almost the same so $\mathcal{H} \approx \mathcal{H}_g$. In dry conditions, when the gravity reservoirs are empty, the PDF of x_g is not a Dirac delta but equal to the PDF of $\forall_{e,g,max}/\left(\forall_{e,c,max} + \forall_{e,g,max}\right)$, which yields $\mathcal{H}_g = \mathcal{H}_g^* > 0$ as can be seen by the flat blue line from February to mid May. A physical interpretation of this is that when the gravity reservoirs are empty across the basin, the available gravity storage (unit of [L]) is not spatially uniform so this reservoir will have a spatially-variable effect on e.g. saturation-excess runoff. However, unlike in the previous example, the positive complexity does not mean that the other fluxes related to this reservoir such as percolation and lateral subsurface are still complex because they are in this case zero throughout the basin. For this basin, the capillary reservoir does not go near empty but if it does, $\mathcal{H}_c = \mathcal{H}_c^* > 0$. In this

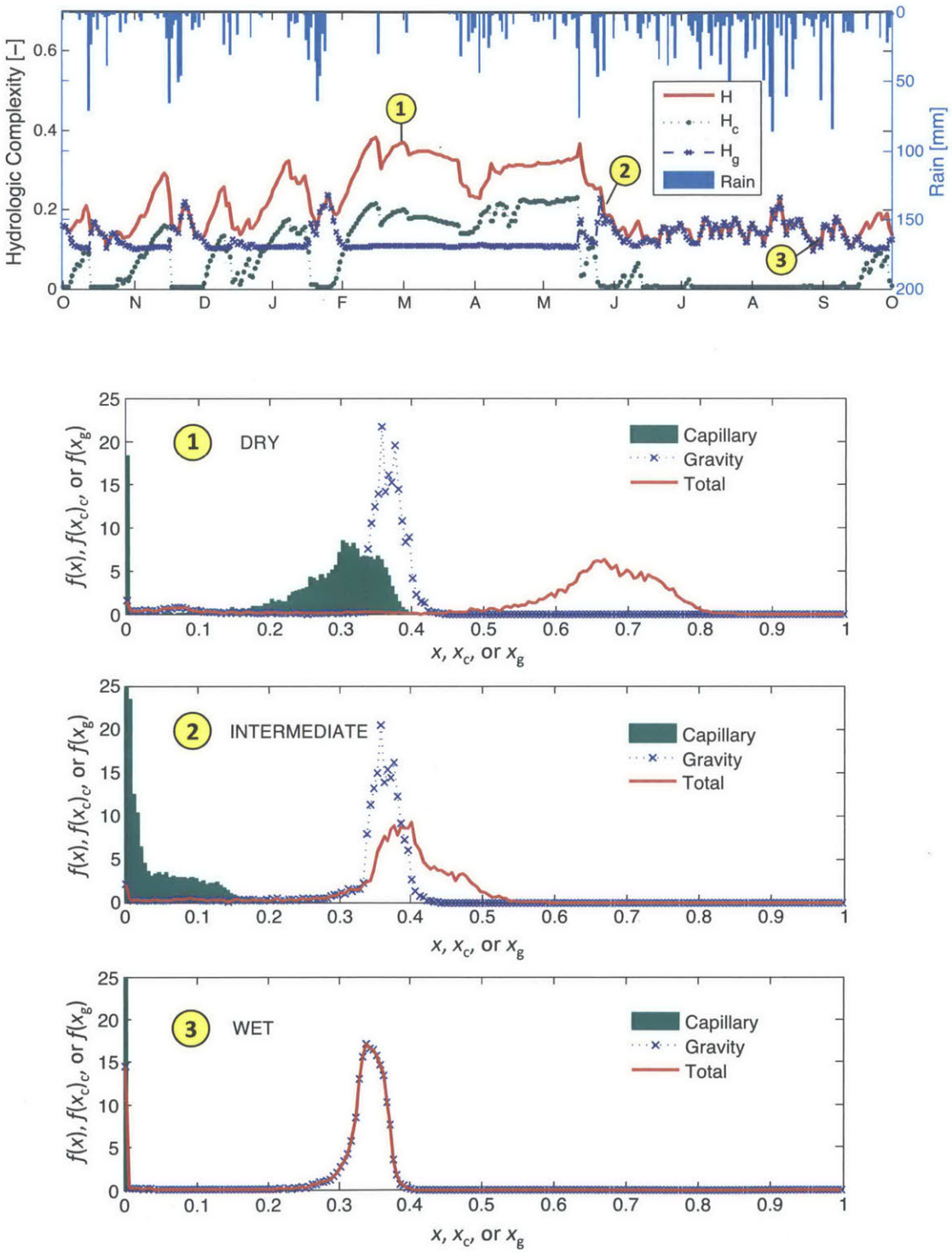


Figure 4.7: Time series of the hydrologic complexity of the basin, the capillary reservoir, and the gravity reservoir. Also shown are three snapshots showing the spatial PDF of relative soil water deficit in the dry, intermediate, and wet regimes.

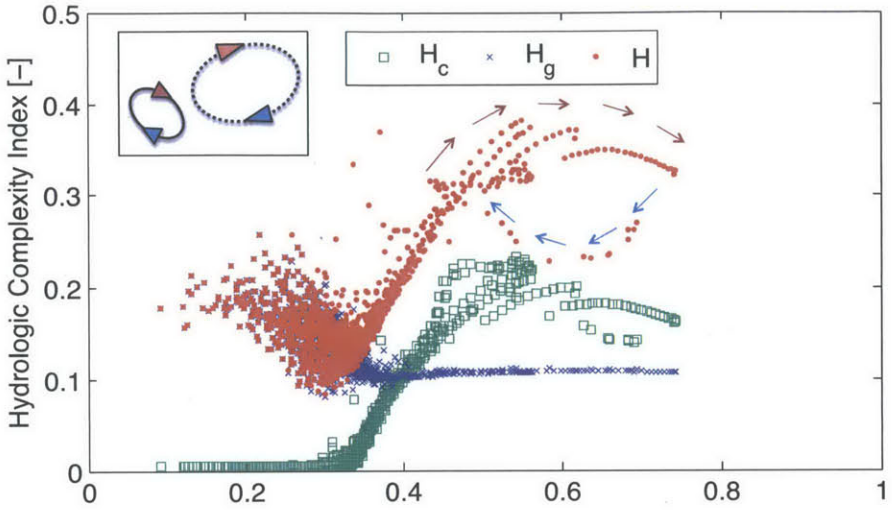


Figure 4.8: The hydrologic complexity index of the soil capillary reservoirs, soil gravity reservoirs, and the (combined) soil reservoirs, vs. the relative soil water deficit. Inset shows hysteresis. Red and blue arrows indicate drying and wetting directions, respectively.

limit, the capillary reservoir will cause complexity or variability in runoff response but it can not be used to mean that there is still complexity in other fluxes related to this reservoir such as ET.

In both examples, the definition of x is the same, so the values and interpretations of the total complexity $\mathcal{H}$ are unchanged. The measure of complexity introduced here was developed for any dimensionless random variable in the interval [0,1] so it is valid in both examples. However, as shown, care must be taken in interpreting the values of this metric computed using different variables, and relating the values with each other. In a review on the use of entropy theory in hydrology, *Singh* (2011) noted that an entropy-based analysis becomes complicated when dealing with multiple subsystems or variables, and suggested that more work needs to be done. With the pair of examples used, we showed that the physical interpretation, i.e. in relation to various hydrologic fluxes and states, of the measure of hydrologic complexity, in particular of $\mathcal{H}_c$ and $\mathcal{H}_g$, should consider the definition of the variables used.

4.5 Summary

Basin response and hydrologic fluxes are functions of hydrologic states e.g. soil moisture. However, soil moisture is itself heterogeneous in space and dynamic in time. To characterize the spatial distribution of soil moisture and understand its evolution, we introduced a modified version of the *Martina and Entekhabi* (2006) dimensionless index of hydrologic complexity $\mathcal{H}$ which measures the distance of a given distribution from two limiting distributions: Dirac delta (simplest case) and uniform distribution (most complex case). The modifications were done to make $\mathcal{H}$ discretization-invariant. The key features of the index are: i) it is computed for random variables with values strictly in the interval 0 to 1; ii) it is computed based on differential entropy instead of Shannon entropy; iii) it uses the Kullback-Leibler divergence to ensure non-negativity; and iv) it uses an exponential transformation so that its value ranges from 0 to 1, with $\mathcal{H} = 0$ for the simplest, and $\mathcal{H} = 1$ for the most complex case, respectively.

For demonstration purposes only, we applied the index on MOBIDIC-simulated timeseries of soil moisture fields of a real basin. $\mathcal{H}$ was computed from the spatial distribution of total relative soil water deficit. Since a key feature of MOBIDIC is the partitioning of each soil moisture storage unit into a gravity reservoir and a capillary reservoir, we also investigated the complexity of these sub-systems, $\mathcal{H}_c$ and $\mathcal{H}_g$, respectively. It was shown that $\mathcal{H}$, $\mathcal{H}_c$ and $\mathcal{H}_g$, can track the evolution of the distributional features of soil moisture and can provide insights on the switching of dominant hydrologic processes. Care must be taken in interpreting the values of the index computed using different variables, and relating the values with each other and to various hydrologic fluxes and states. Physical interpretations should consider the definition of the variables used. Finally, although hysteresis is not explicitly coded in MOBIDIC, this phenomenon is exhibited by the plots of $\mathcal{H}$ and $\mathcal{H}_c$ versus soil moisture. These findings are just just preliminary with the main purpose of demonstrating how $\mathcal{H}$ can be applied to address the science questions. More tests

and analysis will be performed in the next chapters. Particularly, the analysis of the factors affecting the characteristic value of $\mathcal{H}$ will be investigated using multiple basins representing different spatial scales and climatic regimes.

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 5

Understanding the Hydrology of a Tropical Rainforest Basin

5.1 Abstract

In this chapter, MOBIDIC is used to investigate a 103 sq. km. mountainous tropical rain forest (TRF) basin in southern Philippines. Aside from being the first application of MOBIDIC on a TRF basin, this chapter also demonstrates how a sophisticated hydrologic model can be developed using freely-available remotely-sensed data, plus only minimal field observations and measurements. In order to understand how the basin responds to interannual variability of rainfall, a 4-year period that includes a dry year and a wet year was simulated. Based on the simulations, the average annual rainfall of 3877mm was partitioned into 22% quick flow, 38% base flow, 37% evapotranspiration, and 3% groundwater recharge. Canopy interception is about 12% of the total rainfall. This study also investigates the dynamics of the revised dimensionless measure of hydrologic complexity $\mathcal{H}$ which was found to behave differently in the wet and dry regimes. The transition occurs when the dominant processes in the basin switch between vertical and lateral fluxes. Moreover, although

hysteresis is not coded in MOBIDIC, the simulated soil moisture fields clearly exhibit this phenomenon. This hysteretic behavior emerges as a result of MOBIDIC's use of a dual-compartmentalized soil which captures the different roles of capillary and gravity-driven processes, as well as the threshold-dependent nature of lateral subsurface runoff.

5.2 Introduction

Tropical rain forests (TRF) are important ecosystems that serve many environmental and economic functions. Although much progress have been made in recent years, the role and importance of TRFs in providing reliable water supply, minimizing flood hazard, and controlling erosion, are still not well understood (*Bruijnzeel*, 2004; *Douglas et al.*, 2005; *Lamb*, 2011). For water supply in particular, more study is needed to understand how a humid basin behaves during an unusually dry period. As TRFs are increasingly threatened by deforestation, land conversion, and climate change, there is an urgent need for more scientifically-based hydrologic assessments of these ecosystems in order to better articulate their values and better design policies, infrastructures, and management strategies.

The dramatic advances in computing and remote sensing during the last few decades, along with the improved understanding of hydrologic processes, have encouraged the use of physically-based distributed hydrologic models (DHMs). Unlike lumped-models that treat a basin as a single bucket, DHMs resolve hydrologic fluxes and states within the basin, use physically-meaningful parameters, and do not require long time series e.g. of streamflow measurements for model calibration. Chapter 2 describes some specific DHMs, and discusses the merits and criticisms of DHMs in general. An example application of DHM in TRF environment is *Molicova et al.* (1997) who used TOPMODEL to study the streamflow dynamics of a 1-hectare catchment

in French Guinea. Also notable is *Lamb* (2011) who assessed the impacts of land cover changes on runoff for all major tropical forested river basins in the world using a coarse-scale (0.5° grid) process-based water balance model. However, although the use of DHMs for TRF basins seems promising, the applications are limited probably because of the low technical capability in the region with respect to using DHMs, and the lack of hydrologic data needed as inputs to these models.

In this study, MOBIDIC is used to investigate a 103 sq. km. mountainous TRF basin in southern Philippines. This is the first application of MOBIDIC on a TRF basin. Moreover, this chapter demonstrates how a sophisticated hydrologic model in general can be developed using freely-available remotely-sensed global data sets namely: digital elevation model (DEM) from the SRTM satellite, aerial photo from GoogleEarth™, rainfall measurements from the TRMM satellite, weather data from a typical weather website, plus minimal field observations and flow measurements.

Furthermore, this study extends the works of *Martina and Entekhabi* (2006) who, upon recognizing the inadequacy of hydrographs to capture the spatial variability of processes within a basin, developed a method to infer the marginal probability distribution of the characteristic antecedent soil water deficit across a basin using only time series of precipitation and streamflow which are often the most readily available hydrologic data sets. The method is based on the nonlinearity between storm depth and runoff volume. They also introduced an entropy-based dimensionless measure of hydrologic complexity which ranges in value from zero for the simplest, to one for the most complex basin, respectively. The current study uses the revised measure of hydrologic complexity $\mathcal{H}$ derived in the previous chapter.

The method employed here to develop the model of the basin, as well as the method to infer sub-basin variability using only streamflow and rainfall time series, both directly address the problem of hydrologic data scarcity in TRF environment.

5.3 Methods and Data

5.3.1 Study Area

The study area is the 103 sq. km. Upper Tamugan River Basin ($125°14' - 125°22' E$ and $7°5' - 7°12' N$) located in the Province of Davao, in the southern island of Mindanao, the Philippines. Figures 5.1(a-c) shows the location map, the 90-m resolution DEM from the Shuttle Radar Topography Mission (SRTM) satellite, and the basin boundary and river network that were derived from this DEM. The basin has a funnel shape and is bounded in the southwest by Mt. Tipolog. The basin generally drains north-east, with the two main tributaries being the Upper Tamugan River that drains the northern and western parts of the basin, and the Panigan River that drains the southern and eastern parts. The topography is mountainous with elevation of 400–1905 m.a.s.l. and mean slope of 35%. The geology is characterized by thin clayey soil underlain by porous volcanic tuff formation.

The landcover map was created by patching several tiles of aerial photos from GoogleEarth™, which were then geo-referenced and converted into a polygon shapefile in ArcGIS™. As shown in Figure 5.2, 68% of the basin is covered by closed-canopy dipterocarp forest and 26% by brushlands and woodlands. The remaining 6% is used as farms (mostly subsistence upland cultivation) and human settlements.

The landcover map was then used to derive distributed basin properties such as albedo, canopy interception capacity and turbulent heat transfer parameter. Other distributed properties were set using simple linear functions. For example, the soil depth h_{soil} [L] was assumed to increase from 0.3 m on steep hillslopes to 1.2 m on the valleys,

$$h_{soil} = 1.2 - 0.9\frac{s}{s_{max}} \tag{5.1}$$

where s is the local slope and s_{max} is the maximum slope in the basin. Each distributed basin property was calibrated using a corresponding global multiplier. Values

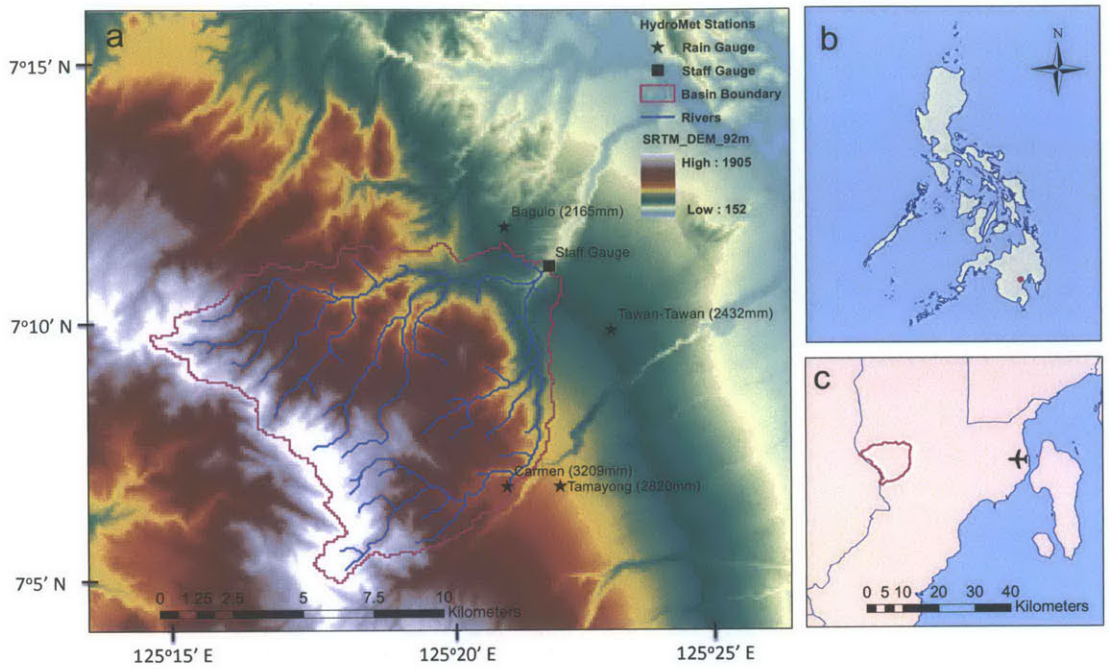


Figure 5.1: Maps of the Upper Tamugan River Basin showing a) the SRTM DEM, the basin boundary and river network derived from the DEM, and the location of the staff gauge and nearby rain gauges (in parenthesis are mean annual rainfall); b) its location (red dot) in southern Philippines; and c) its location relative to the sea and a weather station at the regional airport

were constrained such that they are consistent with field observations and literature values. The calibrated basin properties are shown in Figure 5.3.

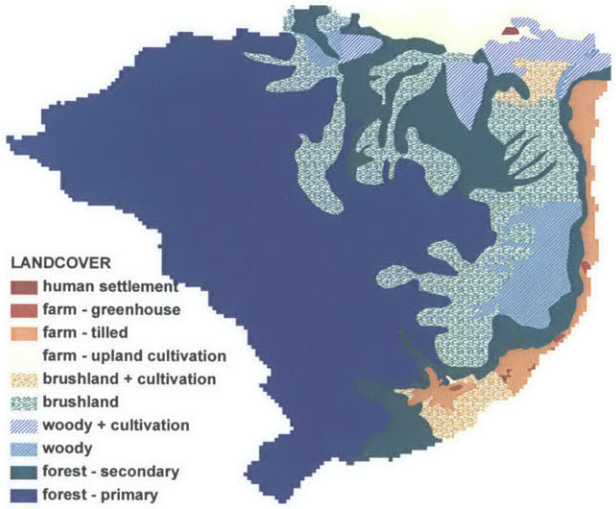


Figure 5.2: The Tamugan basin is covered by 68% forest, 26% brushlands and woodlands, and 6% farmlands and human settlements

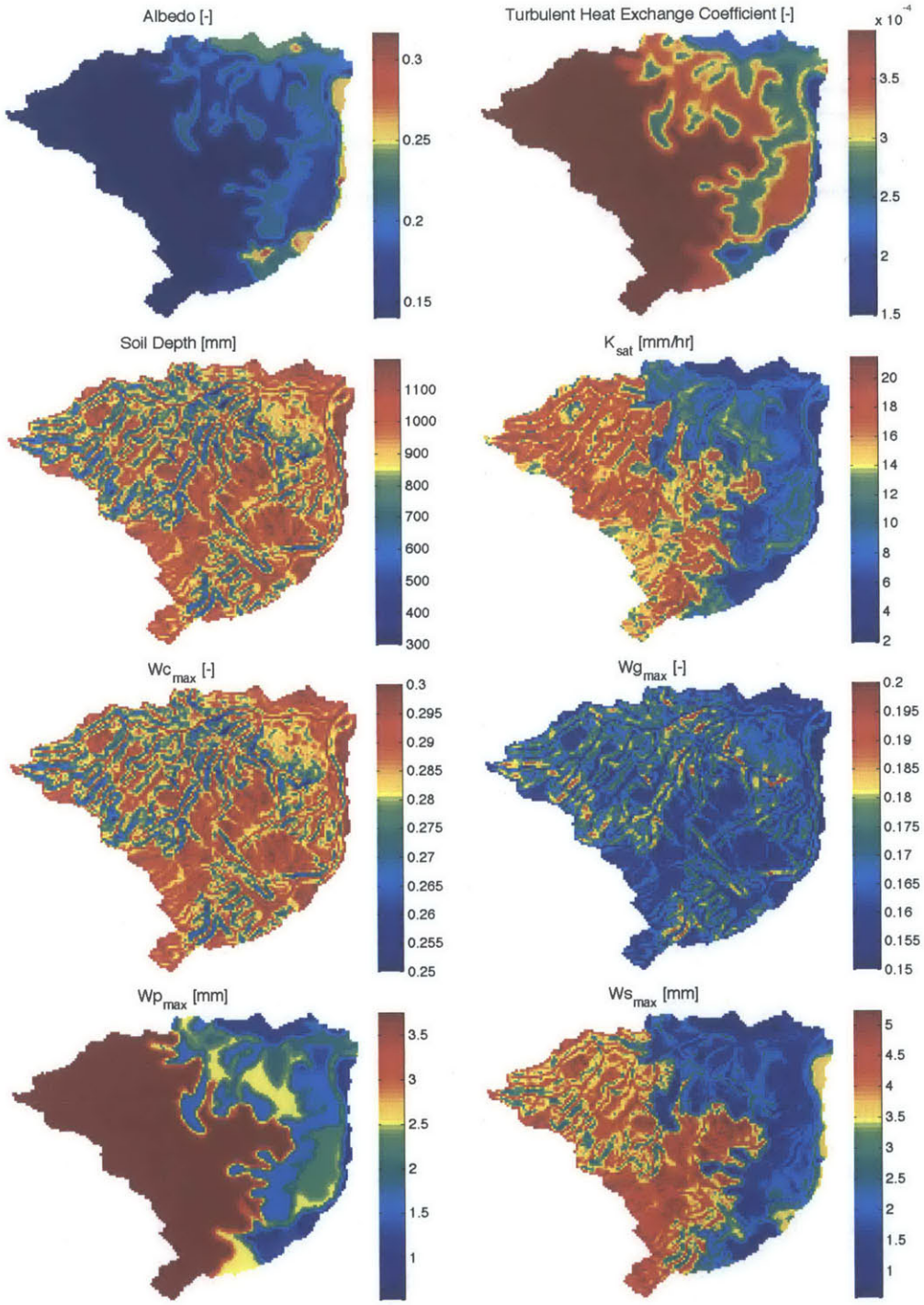


Figure 5.3: Distributed parameters: a) albedo [-]; b) turbulent heat exchange coefficient [-]; c) soil depth [mm]; d) soil saturated hydraulic conductivity [mm/hr]; e) capacity of soil capillary reservoir [m^3/m^3]; f) capacity of soil gravity reservoir [m^3/m^3]; g) capacity of plant canopy [mm]; and h) capacity of surface storage [mm]

5.3.2 Meteorological Forcings

Like many tropical forested river basins particularly in South East Asia, the Upper Tamugan Basin is not well-instrumented. The nearest weather station is at the Davao City Airport located about 30 km to the east, see Figure 5.1(c). To address this, data sets from various alternative sources were combined. Particular emphasis was given to generating good rainfall and solar radiation forcings.

Historical rainfall record was obtained from the Tropical Rainfall Measuring Mission (TRMM) satellite for a 14-year period from Oct. 1, 1998 to Sept. 30, 2012 (water years, WY, 1999–2012). To offset the coarse spatiotemporal resolution of TRMM, both temporal and spatial downscaling were performed. The temporal resolution of TRMM is 3 hours but in order to better represent the diurnal dynamics, the MOBIDIC model was setup with an hourly time step. The temporal downscaling was accomplished by first extracting the diurnal pattern of rainfall from the entire data set, and then linearly-interpolating hourly mean values. Figure 5.5(b), shows that the interpolated hourly values produce a smoother curve with a greater diurnal amplitude.

The spatial resolution of TRMM is 0.25 ° so a single pixel (7 − 7.25 °N, 125.25 − 125.5 °E) practically covers the entire basin. Spatial downscaling was performed by using the elevation and mean annual rainfall at 4 community-based rain gauges near the foothills of the basin, as mapped in Figure 5.1(a). The mean annual rainfall P [mm] appears to be a linear function of elevation z [m.a.s.l.] according to Equation (5.2) with $R^2 = 0.97$.

$$P = 1603 + 1.69z \tag{5.2}$$

Using ArcGIS™, 55 synthetic rain gauges were created and manually distributed over the basin. Since the spatial pattern of rainfall is influenced by the landscape but not as much by the micro-topography, no synthetic gauge was placed near rivers.

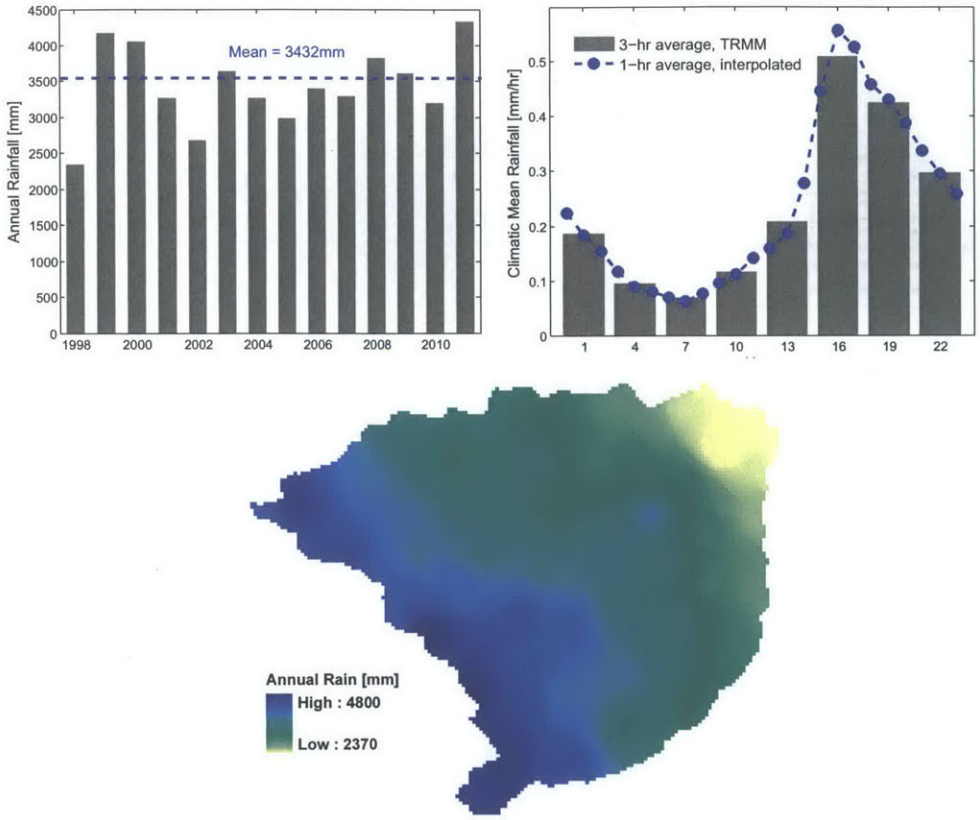


Figure 5.4: a) Inter-annual variability of the TRMM rainfall; b) diurnal pattern of TRMM rainfall at 3-hour (data) and 1-hour (interpolated); c) annual mean rainfall fields at 90 m resolution

The hourly times series of rainfall at each of these locations were computed using Equation (5.2). Finally, MOBIDIC has a built-in inverse-distance interpolation subroutine that generates the continuous rainfall fields. The inter-annual variability of the TRMM-derived rainfall over the basin is shown in Figure 5(a). The basin-averaged annual rainfall ranged between 2300 and 4300 mm, with a mean of about 3430 mm. Figure 5(c) shows the spatial distribution of the mean annual rainfall, varying between 2370 mm at the mouth of the basin to about 4800 mm at the ridges. In addition, storm events were extracted from the hourly rainfall time series, see summary in Table 5.1. The mean storm event pours 23 mm over 11 hours, followed by 48 hours of interstorm period.

Table 5.1: Properties of storm events derived from TRMM data for WY1999–2012

Property	Unit	Min	Median	Mean	Max
Storm depth	mm	0.16	12	23	337
Storm duration	hour	1	6	11	159
Storm interarrival time	day	0.58	2.2	2.0	24

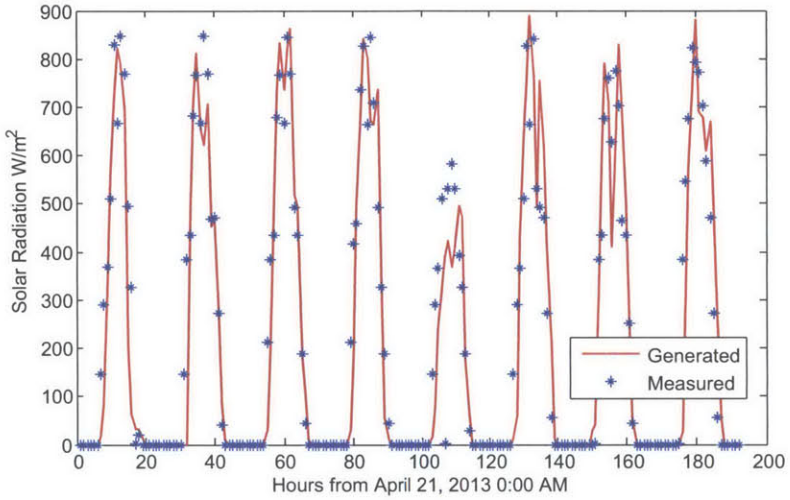


Figure 5.5: Net solar radiation generated by multiplying the analytic values for clear-sky with cloud factors based on qualitative description of conditions at the Davao City Airport, matches well with direct measurements on April 21–29, 2013.

Hourly data on air temperature, relative humidity, and wind speed, for the regional airport, were obtained from a typical weather website (www.wunderground.com). The airport does not measure solar radiation but it gives a qualitative description of the conditions e.g. scattered clouds, heavy rain showers, etc. So, the solar radiation was computed as the product of the analytic value for solar irradiance at the top of the troposhere as a function of latitude, longitude, and time of the day, see *Bras* (1990), and a multiplier based on the reported weather condition. The multipliers shown in Table 5.2 were first arbitrarily assigned then calibrated using direct measurements of solar radiation at the Davao University (*The Manila Observatory*, 2013) for an 8-day period from April 21 to April 29, 2013. The result of the calibration is shown in Figure 5.5.

Table 5.2: Multipliers of clear-sky solar radiation assigned to qualitative descriptions of weather condition at the Davao City Airport

Qualitative Sky Condition	Multiplier
Clear	0.68
Scattered Clouds	0.65
Partly Cloudy	0.60
Dust Whirls	0.55
Smoke	0.50
Light Sandstorm	0.48
Mostly Cloudy	0.45
Sandstorm	0.42
Light Rain Showers	0.40
Rain Showers	0.38
Haze	0.35
Overcast	0.32
Heavy Haze	0.30
Heavy Rain Showers	0.25
Light Rain	0.22
Rain	0.20
Heavy Rain	0.18
Light Thunderstorm	0.17
Light Thunderstorms and Rain	0.15
Thunderstorm	0.14
Thunderstorms with Hail	0.13
Thunderstorms and Rain	0.12
Heavy Thunderstorms with Hail	0.11
Heavy Thunderstorms and Rain	0.10
Unknown or No Data	N/A

5.3.3 Groundwater Interaction

The soil layer modeled by MOBIDIC interacts with the groundwater through the vertical fluxes of percolation and capillary rise. In addition, the bidirectional groundwater-river interaction is modeled using a modified linear reservoir formulation (*Domenico and Schwartz*, 1998). Unlike in typical applications where the outlet of a basin is used as the single reference elevation, here, the reference assigned to each point is the stage of its nearest river segment. The fluxes to and from the groundwater aquifer strongly depend on the position of the groundwater table which was initialized as equal to

the height above nearest drainage (*HAND*) introduced by *Nobre et al.* (2011). To minimize errors associated with initial conditions, each MOBIDIC simulation was spun-up until the groundwater table was practically in steady-state.

5.3.4 Calibration

Scalar parameters and global multipliers of distributed parameters were calibrated using several objective and qualitative checks. First, the goodness-of-fit between the observed and simulated streamflow at the outlet of the basin was objectively evaluated using the absolute bias and exceedance probability index. Other popular metrics such as the Nash-Sutcliffe Efficiency Index (*Nash and Sutcliffe*, 1970) were not used because the TRMM rainfall does not closely match the timing of the measured streamflow. However, it should be noted that the aim was not simply to get the best value of the objective metrics, but to develop a *realistic* model. To reiterate, parameters were constrained within typical literature values. Also the hourly time series and annual total of fluxes e.g. of ET were qualitatively checked. Another criterion is that the change in groundwater table elevation at each pixel between the start and end of the 4-year simulation is at most 20 cm.

5.3.5 Measure of Hydrologic Complexity

The fraction of precipitation that becomes runoff depends on the fraction of the basin that is saturated (*Dunne and Black*, 1970). This is particularly true for forested basins which have high infiltration capacity. To quantify the complexity of the rainfall-runoff relationship, *Martina and Entekhabi* (2006) introduced an entropy-based dimensionless index of hydrologic complexity $\mathcal{H}_{old}$ which ranges in value from zero to one. $\mathcal{H}_{old} = 0$ for the simplest basin which behaves like a single bucket with a spatially-homogeneous (i.e. constant) available soil water storage $\forall_e$ [L], whereas $\mathcal{H}_{old} = 1$ for a basin with $\forall_e$ uniformly distributed from 0 to a maximum value $\forall_{e,max}$.

This thesis uses the revised dimensionless index of hydrologic complexity $\mathcal{H}$ derived in the previous chapter, recall Equation 4.47. $\mathcal{H}$ is a function of the distribution of the relative soil water deficit x [-], $x \in [0, 1]$.

$$\mathcal{H} = \mathcal{H}(x), \qquad x \in [0, 1] \tag{5.3}$$

In this 'direct method', x is equal to the MOBIDIC-simulated soil water deficit $\forall_e$ normalized by the total soil water storage capacity $\forall_{e,max}$ at each pixel, $\forall_{e,max} = W_{c,max} + W_{g,max}$. In addition, the characteristic or time-invariant $\mathcal{H}$ is also computed using the 'indirect method' developed by *Martina and Entekhabi* (2006). In this method, $\mathcal{H}$ is derived using only time series of precipitation and discharge which are often the most readily available hydrologic data sets. x is extracted from the nonlinear relationship between the event rainfall depth H_s and the resulting quick flow depth H_r, see Equation 5.5. Compared to *Martina and Entekhabi* (2006) who used a Poisson-distributed rainfall and the smooth-minima technique (*Institute of Hydrology*, 1980) for base flow separation, the indirect method employed here is more straightforward because historical rainfall time series was used and the base flow and quick flow components of the hydrograph were directly simulated by MOBIDIC. The maximum storm depth $H_{s,max}$ is used as normalizer .

$$\text{Direct Method}: \quad x = \forall_e / \forall_{e,max} \tag{5.4}$$

$$\text{Indirect Method}: \quad x = H_s / H_{s,max} \tag{5.5}$$

5.4 Results and Discussion

5.4.1 Water Balance

For the simulated 4-year period, the average annual rainfall of 3877 mm was partitioned into 60% streamflow, 37% ET, and 3% recharge to the regional groundwater system. The streamflow contribution can be further partitioned into 22% quick flow and 38% base flow. The vegetation intercepted about 12% of the total rainfall. It was also found that the soil layer over the entire basin was at or above field capacity for about 90% of the time. The values obtained for ET and canopy interception are consistent with literature values reported for TRFs in the region e.g. in *Dykes* (1997).

The basin-averaged and daily-aggregated values of the TRMM-derived rainfall and the MOBIDIC-simulated hydrologic fluxes and states for WY 2009–2012, are shown in Figure 5.6. Notice from the top two columns that there are some mismatches between the TRMM rainfall and the observed streamflow. For instance, a 150 mm rain on Day 752 did not produce a proportional peak flow. Nonetheless, the dynamics (i.e. shape of the rising and falling limbs) and amplitudes of the observed and modeled streamflow are quite consistent (bias of 0.03 and exceedance probability index of 0.05). The fit is very good especially for the base flow regime. The third row shows that the potential and actual evapotranspiration, PET and ET [mm/d], as well as the percolation, Q_{perc} [5mm/d]. As expected for humid basins, ET is almost always equal to PET. It can also be noticed that ET and Q_{perc} are negatively correlated. The bottom plot shows the equivalent water depth stored in the soil capillary reservoir W_c [mm], soil gravity reservoir W_g [mm], plant reservoir W_p [0.02 mm], and surface reservoir W_s [0.02 mm]. Except for W_c, the water stored in these reservoirs are peaky and quickly respond to individual rain events. Furthermore, W_c is almost always saturated except during the abnormally dry period around Day 500 which caused a reduction in base flow and ET.

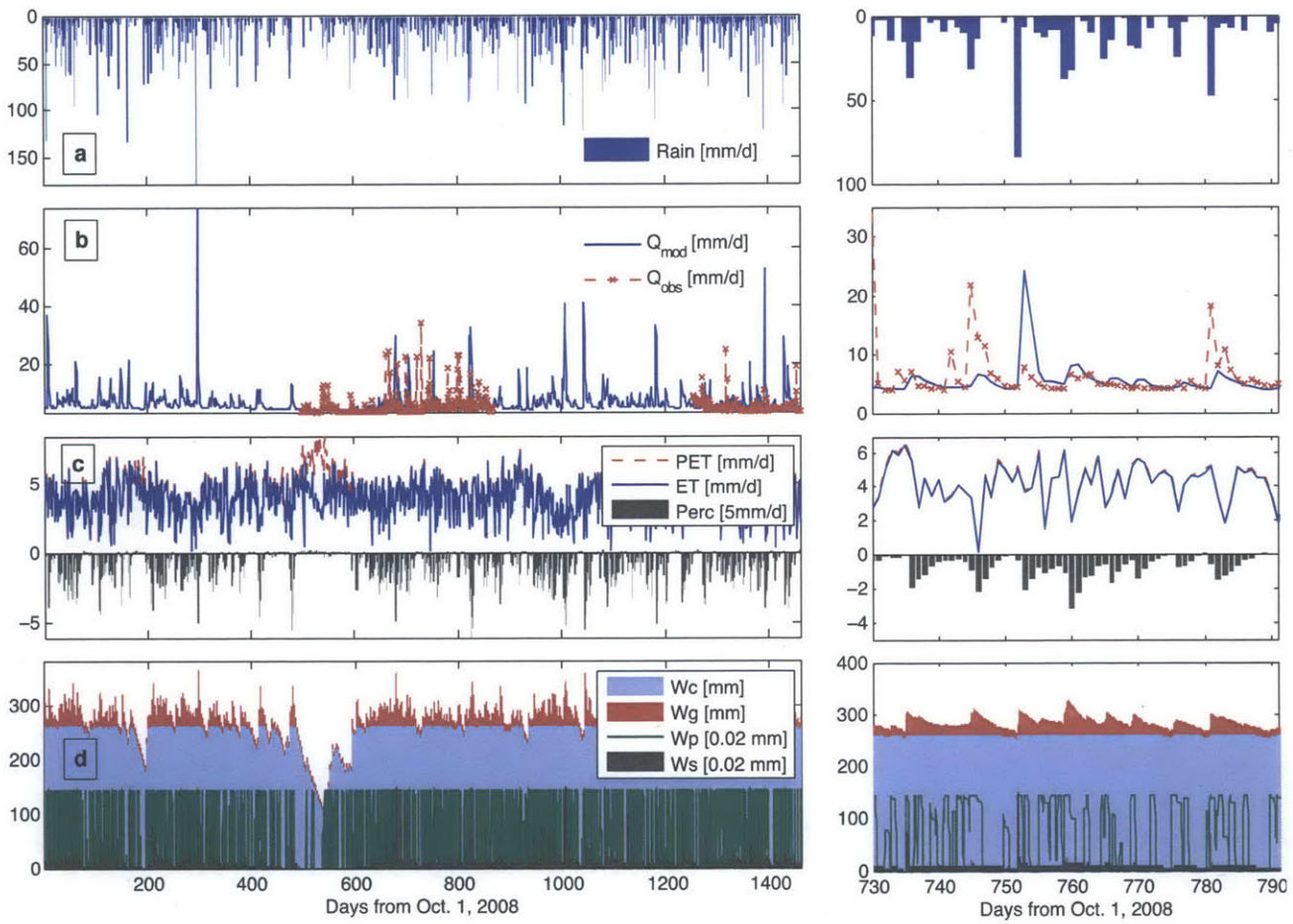


Figure 5.6: Time series for WY 2009–2012 (left column focuses on Oct. & Nov. 2011) of daily-aggregated and basin-averaged a) rainfall, P [mm/d]; b) modeled vs. observed depth-equivalent streamflow [mm/d]; c) potential and actual ET, PET and ET [mm/d], and percolation, Q_{perc} [5 mm/d]; and d) depth-equivalent of water stored in the soil capillary reservoir, W_c [mm]; soil gravity reservoir, W_g [mm]; plant reservoir, W_p [0.02 mm]; and surface reservoir, W_s [0.02 mm], for

5.4.2 Hydrologic Fluxes and States

The time series of the basin-averaged groundwater table elevation [m.a.s.l.] is shown in Figure 5.7(a). The range is about 0.8 m. Notice that the values at the start and end of the 4-year period are almost the same. As a further check, Figure 5.7(b) shows that the net movement of the groundwater table is less than 20 cm for all points in the basin. Figure 5.7(c) shows the temporal mean groundwater table. The dynamics of the groundwater table indicates how the basin is affected by the inter-annual variability of rainfall. For instance, the unusually dry period from December 2009 to May 2010 caused a significant dip of the groundwater table, see the red box in Figure 5.7(a). In contrast, the relatively wet WY 2012 did not raise the groundwater table above its usual position. Thus, this TRF basin is more sensitive to drought than to increased rainfall.

Figure 5.8(a–c) show the spatial distribution of the mean saturation [%] of the soil capillary reservoirs $W_c/W_{c,max}$, soil gravity reservoirs $W_g/W_{g,max}$, and plant reservoirs $W_p/W_{p,max}$. First, in Figure 5.8(a), the map shows that the temporal mean of $W_c/W_{c,max}$ is 90–100% and the spatial pattern is strongly influenced by both soil properties and topography. The probability density function (PDF) of this map shows a Gaussian curve centered around 95% and an exponentially-shaped curve around 100%. The latter curve represents the areas on or along drainage lines that are always at or above field capacity, while the former curve represents the rest of the basin. Next, Figure 5.8(b) shows that the spatial pattern of $W_g/W_{g,max}$ is controlled mainly by topography with the areas along drainage lines remaining saturated while the rest of the basin is on average only about 5–25% full. This is realistic because of the high hydraulic conductivity of the volcanic soil and the steep topography. Lastly, Figure 5.8(c) shows that the spatial pattern of $W_p/W_{p,max}$ manifests the combined effects of vegetation and rainfall spatial variability. The plant reservoirs are on average about 35–40% full.

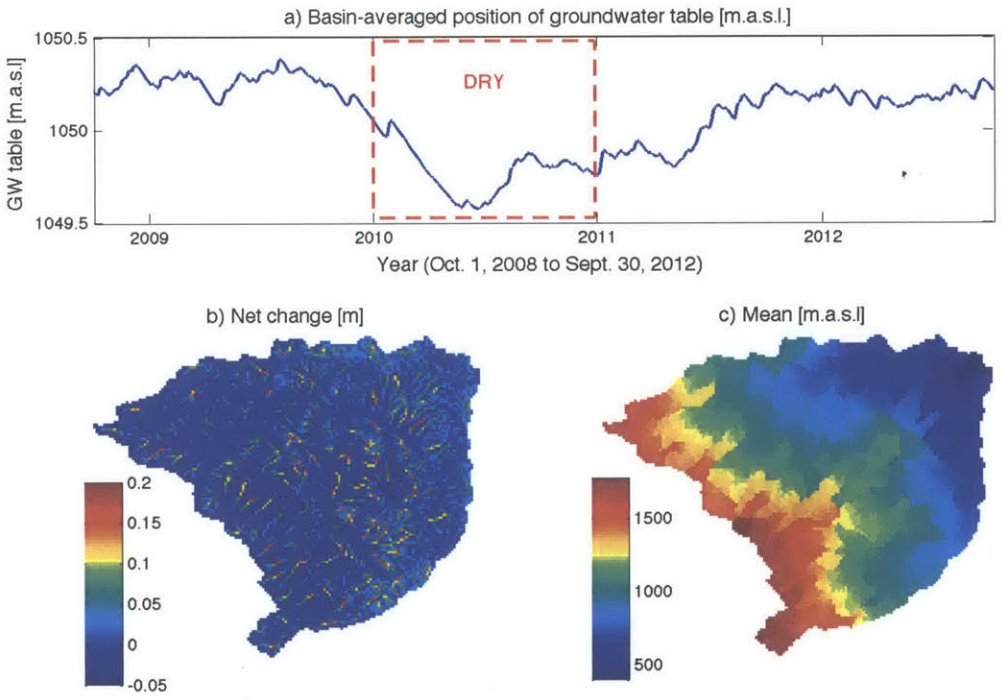
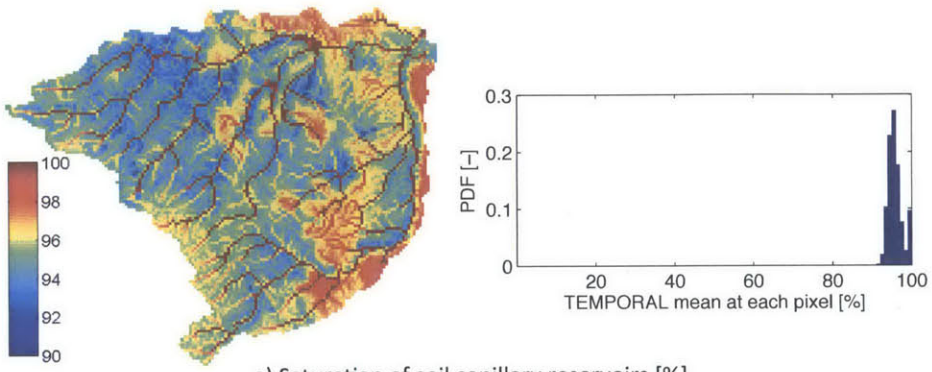
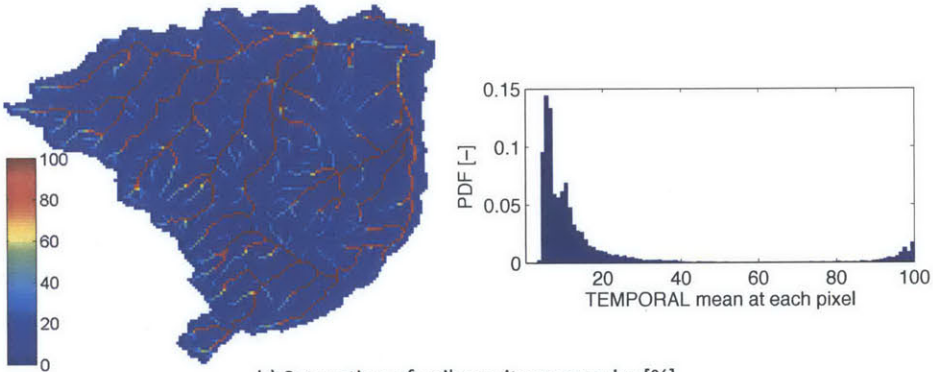


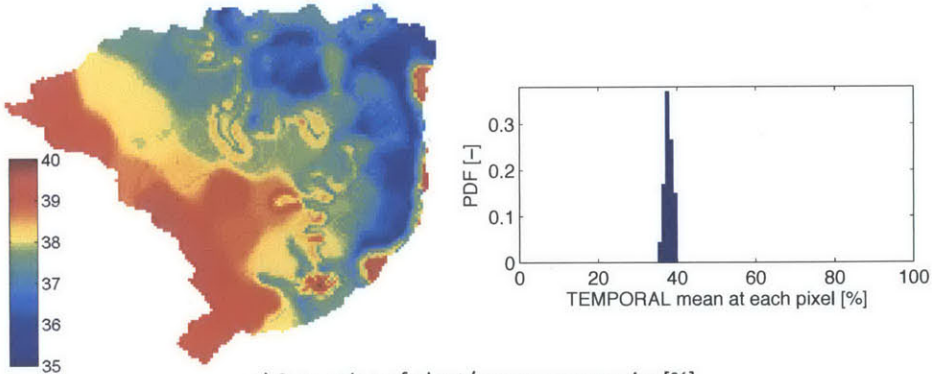
Figure 5.7: Groundwater table [m.a.s.l.] simulated for WY 2009–2012: a) time series of the spatial mean; b) net change [m] at each pixel; and c) temporal mean at each pixel



a) Saturation of soil capillary reservoirs [%]



b) Saturation of soil gravity reservoirs [%]



c) Saturation of plant/canopy reservoirs [%]

Figure 5.8: Maps and probability density function of temporal mean saturation [%] of the a) soil capillary reservoirs; b) soil gravity reservoirs; and c) plant/canopy reservoirs

Better understanding of the spatial distribution and improved estimates of mean annual fluxes of ET, net runoff generation, and groundwater recharge, are useful for water budget and environmental studies. As shown in Figure 5.9(a), the annual ET ranges between 750 mm in the lower parts of the basin to about 1700 mm in the upland riparian zones. This illustrates how ET in this basin is primarily controlled by vegetation, and secondarily by water availability. Although the basin receives a lot of rainfall, the ET can also be water-limited such as during the dry months and occasional droughts.

The net runoff (surface plus sub-surface) generated per unit area $(Q_{run})_{net}$ is defined here as,

$$(Q_{run})_{net} = (Q_{run} + Q_{lat})_{down} - (Q_{run} + Q_{lat})_{up} \tag{5.6}$$

where Q_{run} is surface runoff, Q_{lat} is lateral subsurface flow, and the subscripts *up* and *down* refer to inflows from upstream and outflows to downstream, respectively. All fluxes have consistent units such as mm/year.

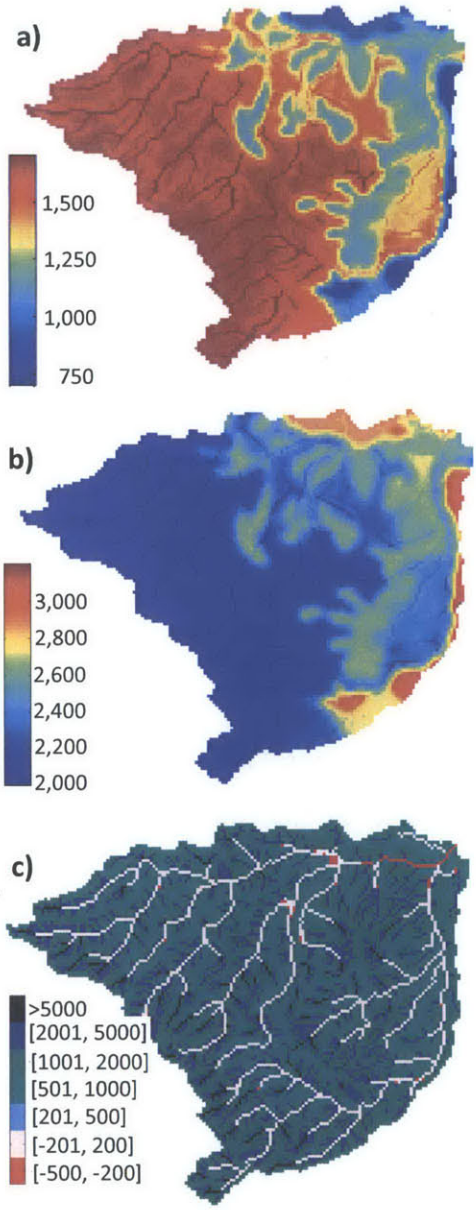


Figure 5.9: Maps of simulated mean annual a) ET; b) net runoff generation; and c) percolation to deeper soil layers

Figure 5.9(b) shows that although the lowlands esp. the east and north-east boundary of the basin receive less rainfall, they generate the most runoff per unit area. This suggests that best management practices for these areas such as the rehabilitation of riparian buffer zones might have significant impact on streamflow

and water quality.

The mean annual percolation from the soil layer modeled by MOBIDIC to the deeper soil layer or groundwater aquifer, is shown in Figure 5.9 (c). This generally follows the decreasing soil hydraulic conductivity from the forested uplands to the valleys. For most of the basin, Q_{perc} = 500–2000 mm/year. On some forested hillslopes, Q_{perc} is as high as 2000–9000 mm/year, while on the river network and riparian zones, Q_{perc} is low or even negative i.e. there is net exfiltration or wetting from below.

5.4.3 Measure of Hydrologic Complexity

Indirect Method

Figure 5.10 shows the partitioning of the total streamflow into base flow and quick flow. Base flow accounted for 63% of the streamflow and it remained high throughout the 4-year period. From the quick flow time series, storm event runoff depth H_r [mm] was extracted. Figure 5.11 plots H_r vs. the corresponding storm event rainfall depth H_s [mm]. Values were grouped into 10, then fitted with the integral of an incomplete gamma function

$$f(H_s; k, \theta) = \frac{1}{\theta^k} \frac{1}{\Gamma(k)} H_s^{k-1} e^{-\frac{H_s}{\theta}} \tag{5.7}$$

where, k and θ are the shape and scale parameters, respectively. The trendline ($k = 106$, $\theta = 0.46$), was then used to generate the PDF of H_s which in this method is equivalent to $\forall_e$. Evaluation of Equation 4.47 yields $\mathcal{H} = 0.17$.

Direct Method

Figure 13(a) shows time series of soil water deficit $\forall_e$ for WY 2011 which is a normal year. The lines show the 5, 10, 50, 90, and 95 percentiles. As shown, about 5% of the basin stays saturated year round, and the big jump in $\forall_e$ is within the 5–10$^{\text{th}}$ percentile. This means that although the basin receives a lot of rain, most of the basin

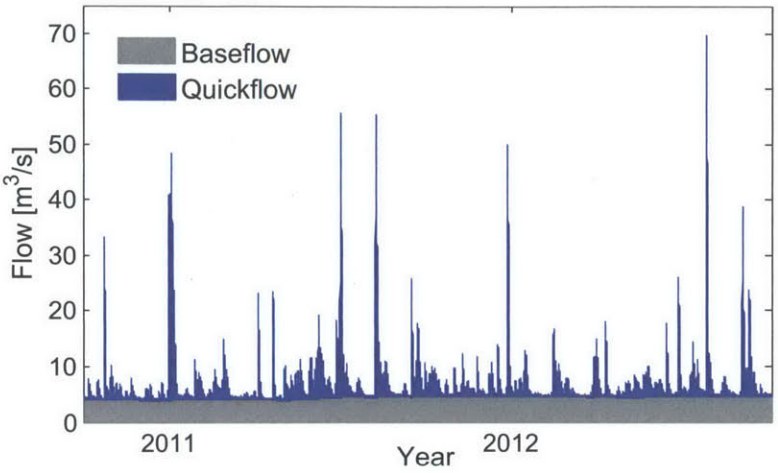


Figure 5.10: Base flow and quick flow simulated for WY 2011–2012

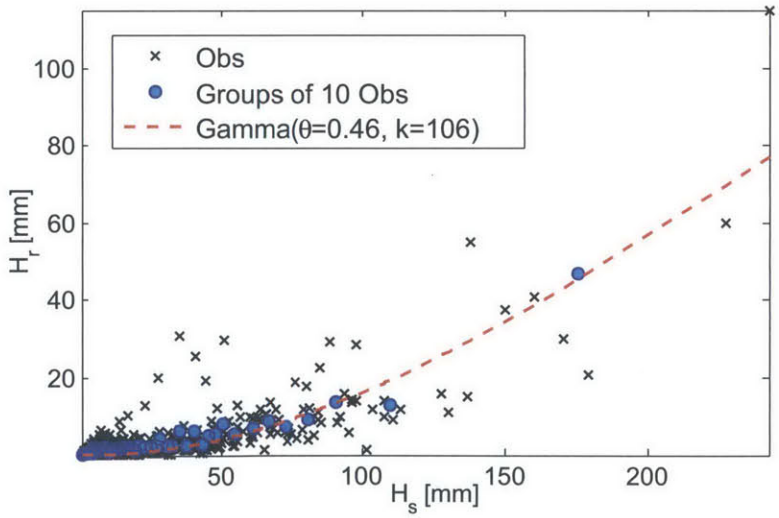


Figure 5.11: Runoff (quick flow) depth vs. rainfall depth for each storm event. Trendline is the integral of an incomplete gamma function.

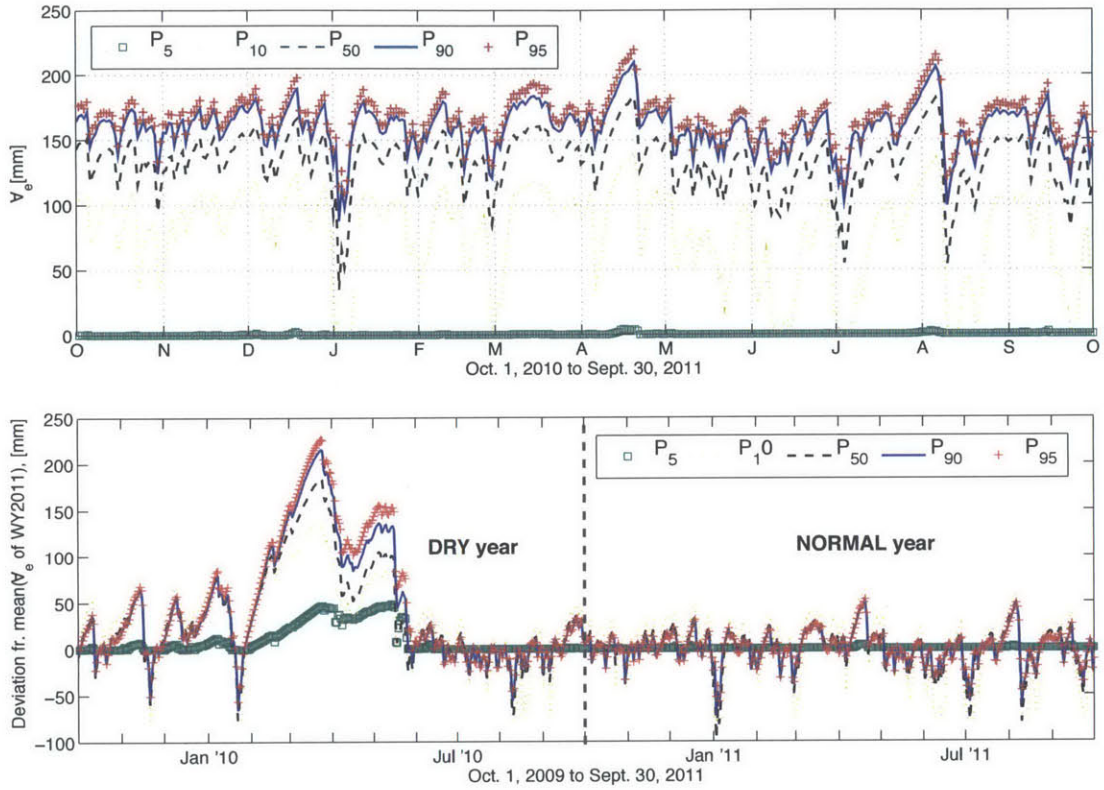


Figure 5.12: Time series of a) the percentiles of water deficit for WY 2011 (normal year); and b) the deviations from the mean of each percentile relative to WY 2011

always have high $\forall_e$ with median of 140 mm. Even during the intense storm in early January 2011, 50% of the basin has $\forall_e > 50$ mm. Figure 13(b) shows the deviation of percentiles of $\forall_e$ relative to the mean values of WY 2011. This plot illustrates the significant difference between a normal year (WY 2010) and an unusually dry year (WY 2011). The plot for WY 2009 and 2012 are not shown as they are similar to that of WY 2011.

Figure 5.13 plots the time series of $\mathcal{H}$ and P for WY 2010–2011. The characteristic $\mathcal{H}$ is 0.19 which is in good agreement with the value of 0.17 obtained through the indirect method, thus demonstrating the success of the indirect method in providing insights about the spatial variability of $\forall_e$.

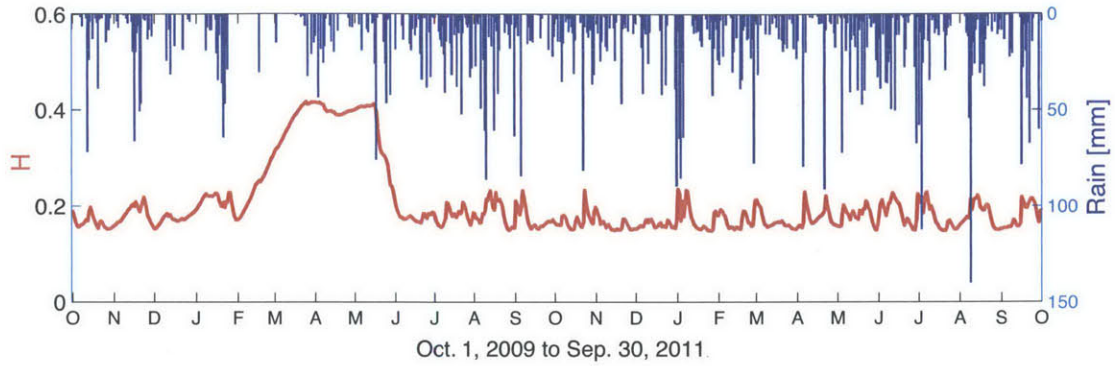


Figure 5.13: Dynamics of the hydrologic complexity index $\mathcal{H}$ for WY 2010–2011.

To investigate further, Figure 15 plots $\mathcal{H}$ vs. the basin-averaged soil water deficit $\overline{\forall_e}$. The basin has a preferred state with an $\mathcal{H}$ of about 0.18 which corresponds to $\overline{\forall_e}$ of about 140 mm. Interestingly, in this preferred state, most of the basin is at field capacity. More interestingly, *$\mathcal{H}$ behaves differently in the wet and dry regimes, and exhibits hysteresis.* In the wet regime, while rain causes a decrease in $\forall_e$ across the basin, the result is an initial increase in $\mathcal{H}$ because the differences in infiltration capacity generally increase the overall variability of $\forall_e$. Once drainage and redistribution of water through surface and subsurface runoff become the dominant processes, $\overline{\forall_e}$ increases while $\mathcal{H}$ decreases. The decrease in $\mathcal{H}$ means a decrease in spatial variability of $\forall_e$ or a decrease in the basin's complexity. This post-rainfall decrease in spatial variability can be attributed to the homogenizing effect of vegetation (*Ivanov et al.*, 2010) and the influence of topography that leads to organization of wet areas along drainage lines (*Grayson et al.*, 1997). These factors result to a bi-modal distribution of $\forall_e$ which has lower $\mathcal{H}$ compared to e.g. a wide uni-modal Gaussian distribution. Meanwhile, for the dry regime, further drying increases $\forall_e$ across the basin which leads to an increase in $\mathcal{H}$. From a dry state, rain initially decreases $\forall_e$ across the basin with little or no lateral flows. This results to only a slight decrease in $\mathcal{H}$. As additional rain further wets the basin, a threshold (which for this basin is close to field capacity) will be reached, after which $\mathcal{H}$ steeply decreases. This

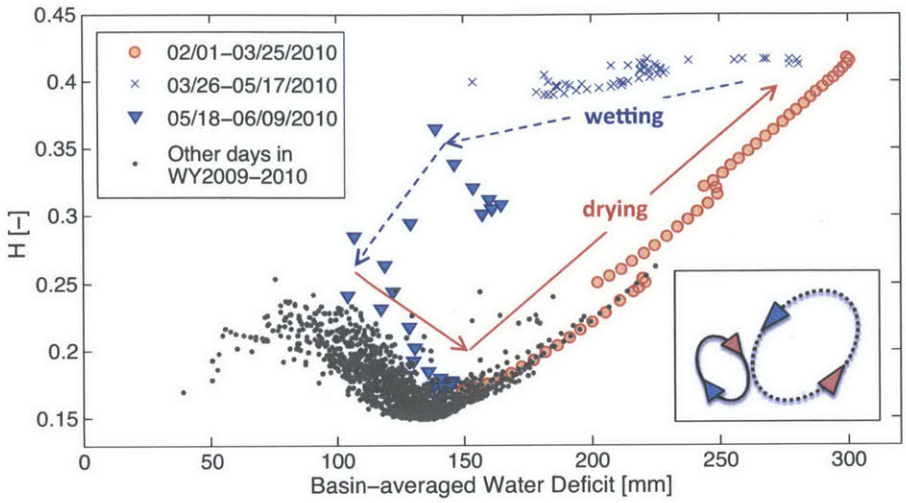


Figure 5.14: Measure of hydrologic complexity vs. basin-averaged water deficit

threshold behavior has been reported in previous studies e.g. *Grayson et al.* (1997) and *Tromp van Meerveld and McDonnell* (2006a), who explained this as a switch in dominant processes from vertical to lateral fluxes.

The looping behavior of $\mathcal{H}$ in the dry regime indicates *hysteresis* of soil moisture at the basin scale. Although it is not as clear from Figure 15, a looping behavior was also observed for the wet regime when the points were plotted sequentially. The inset (bottom right) shows that the hysteresis loops in the dry and wet regimes have opposite direction. The red arrow indicates drying while the blue arrow indicates wetting. Hysteresis refers to the dependence of a system on not only its current state but also on its past state (*Joekar-Niasar et al.*, 2013; *Camporose et al.*, 2014). This phenomenon has long been observed since at least the pioneering works of *Haines* (1930) in the field of soil physics. However, hysteresis is difficult to quantify or parameterize esp. at the basin scale or even at the hillslope scale, so it is often ignored (*Bras*, 1990; *Jaynes*, 1990; *Beven*, 2006b). Recognizing this challenge, *O'Kane* (2005) introduced a simple way to insert rate-independent hysteresis in hydrologic models. Thus, the hysteresis observed in this current study is an unexpected result since MOBIDIC does

not explicitly account for this phenomenon. It is hypothesized that this hysteretic behavior emerged as a result of MOBIDIC's use of a dual-compartmentalized soil which captures the different roles of capillary and gravity-driven processes, as well as the threshold-dependent nature of subsurface runoff as previously discussed.

The behavior of $\mathcal{H}$ in the dry regime can be further illustrated by plotting the evolution of the probability distribution of $\forall_e$. During drying, as the mean $\forall_e$ increases, the distribution also becomes wider which corresponds to an increase in variability. During initial wetting, the mean $\forall_e$ decreases but the distribution simply translates to the left with little decrease in variability. Around field capacity, further wetting causes a sharp decrease in spatial variability, as lateral subsurface flow become the dominant process in the basin.

A few recent modeling studies also reported hysteresis in the plots of spatial variability and mean of depth-averaged soil moisture e.g. *Ivanov et al.* (2010) for a zero-order semiarid basin, and *Vivoni et al.* (2010) for a small arid river basin. They explained that hysteresis, particularly macro-hysteresis, emerges from the interplay between different basin properties (soil, climate, and vegetation) and is related to the switch in dominant hydrologic processes i.e. vertical fluxes in dry conditions to lateral flows in wet conditions. Their conclusions and explanations are similar to those made in this study. Moreover, this research adds to the relatively small literature on macro-hysteresis especially in TRF environment.

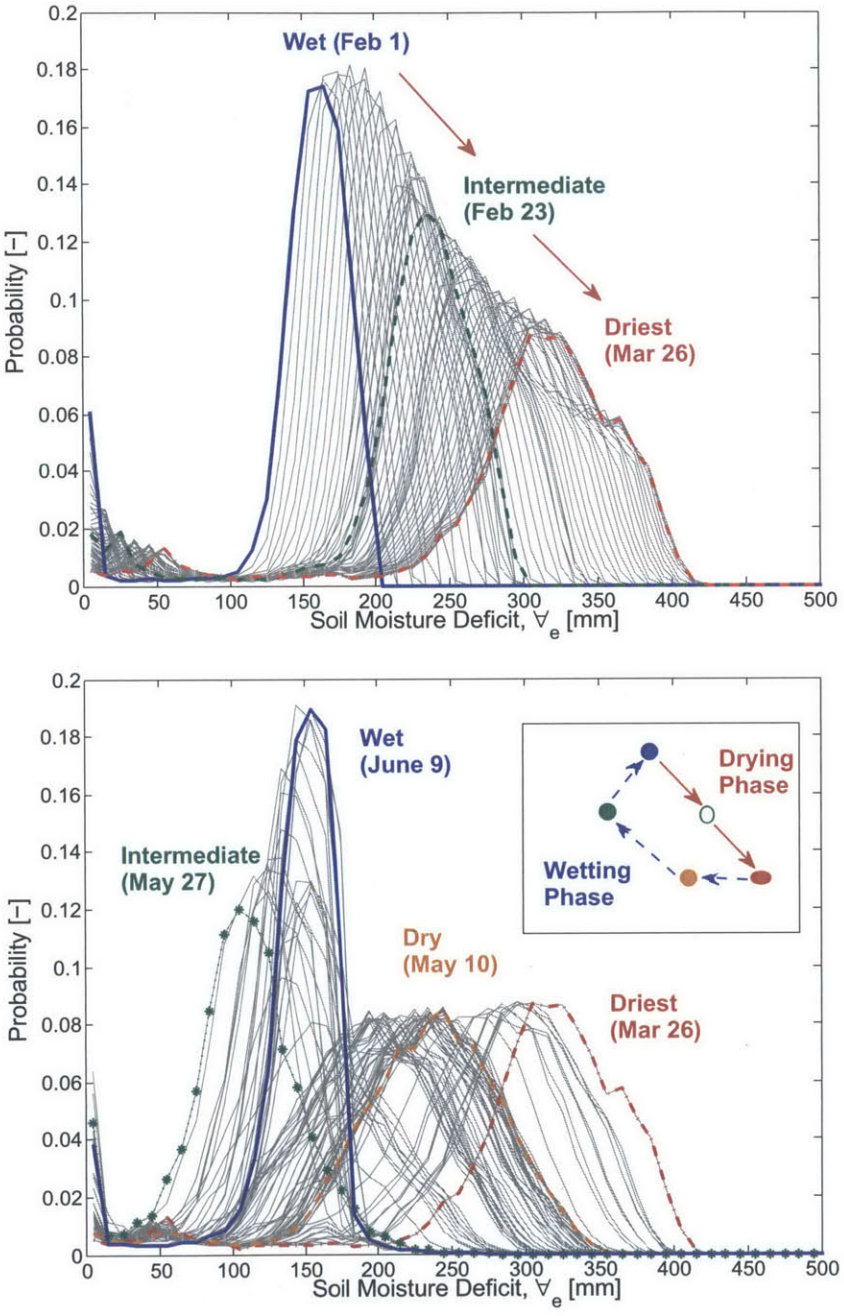


Figure 5.15: Macro-hysteresis in the evolution of the probability distribution of soil water deficit $\forall_e$ during a drying phase (top) and a wetting phase (bottom)

5.5 Summary

The distributed hydrologic model MOBIDIC was used to understand the hydrology of a mountainous tropical rain forest river basin in the Philippines. This chapter demonstrated how a sophisticated hydrologic model can be developed for data-scarce environment using freely- and readily-available SRTM DEM, GoogleEarth™aerial photos, TRMM rainfall, typical online weather data, and only minimal field observations and flow measurements. For the simulated 4-year period, the average annual rainfall of 3877 mm was partitioned into 22% quick flow, 38% base flow, 37% ET and 3% groundwater recharge. The forest cover intercepted about 12% of the total rainfall, and the soil layer was at least at field capacity for about 90% of the time.

This study also applies the revised dimensionless measure of hydrologic complexity $\mathcal{H}$. $\mathcal{H}$ was computed using two methods: first indirectly using only timeseries of precipitation and streamflow, then directly using simulated soil moisture fields. The close agreement of the $\mathcal{H}$ obtained from the two methods (0.17 and 0.19) demonstrates the utility of the indirect method to provide insights about the spatial variability of soil water deficit within the basin. Moreover, it was found that $\mathcal{H}$ behaves differently in the wet and dry regimes. In the wet regime, an increase in soil moisture increases $\mathcal{H}$ while drainage and drying causes the opposite. In the dry regime, further drying increases $\mathcal{H}$ while rain causes the opposite albeit at much slower rate. Once the basin has been wet to field capacity, additional rain causes a sharp decrease in $\mathcal{H}$ due to the activation of lateral subsurface flows which lead to organization of wet areas along drainage lines. Furthermore, this study shows that although hysteresis is not coded in MOBIDIC, the simulated soil moisture fields clearly demonstrate this phenomenon. The hysteretic behavior emerges as a result of the interplay between vertical and lateral fluxes and the threshold-dependent nature of lateral subsurface flow. MOBIDIC's use of a dual-compartmentalized soil appears to be able to realistically represent these complex processes.

Chapter 6

Spatiotemporal Variability of Hydrologic Response: an Entropy-based Approach

6.1 Abstract

Basin response and hydrologic fluxes are functions of hydrologic states, most notably of soil moisture. However, characterization of soil moisture is challenging since it is both heterogeneous in space and dynamic in time. This chapter applies the revised dimensionless index of hydrologic complexity $\mathcal{H}$ on MOBIDIC-simulated timeseries of soil moisture fields for eight test basins with area of 10^0–10^3 km^2 and representing semiarid, temperate and humid climates. Since a key feature of MOBIDIC is the partitioning of each soil unit into a gravity reservoir and a capillary reservoir, the complexity of these sub-systems, $\mathcal{H}_c$ and $\mathcal{H}_g$, respectively, were also investigated. It was shown that $\mathcal{H}$ can effectively track the evolution of the spatial distribution of soil moisture, and together with $\mathcal{H}_c$ and $\mathcal{H}_g$, can capture the switching between vertical and lateral fluxes as the dominant hydrologic processes. This chapter also explores

what and how basin attributes affect $\mathcal{H}$, and how $\mathcal{H}$ can be used to explain inter-basin variability of hydrologic response. Clear and logical relationships are found only by grouping basins with similar climate or size. For basins of similar size, $\mathcal{H}$ and $\mathcal{H}_g$ are highest in temperate climate, consistent with soil moisture being lower- and upper-bounded so its variability peaks at intermediate conditions. $\mathcal{H}_c$ generally increases with aridity since the capillary reservoirs are often saturated in the temperate and humid basins i.e. they are less hydrologically active. For the semiarid basins, $\mathcal{H}$, $\mathcal{H}_c$ and $\mathcal{H}_g$, are positively correlated with catchment area and infiltration ratio, but negatively correlated with relief ratio; while $\mathcal{H}$ and $\mathcal{H}_g$ are positively correlated with baseflow index. Moreover, although not explicitly coded in MOBIDIC, hysteresis is exhibited in the discharge-storage and discharge-$\mathcal{H}$ plots, in both low- and high-flow conditions. Hysteresis emerges from MOBIDIC's novel use of a dual-pore soil structure which not only captures the different roles of capillary and gravity-driven processes, but also mimics the threshold-like behavior of runoff generation. The use of entropy-based measure of complexity helps understand the link between internal patterns and states to lumped or macro processes and behaviors.

6.2 Introduction

The hydrologic response (HR) of a basin pertains to how precipitation is partitioned into streamflow, evapotranspiration (ET), and change in storage. These fluxes can be further partitioned e.g. streamflow into quickflow and baseflow. As one of the main goals of hydrology, the ability to explain or predict the response or the spatiotemporal variability of the above mentioned hydrologic fluxes and states has many important applications from flood forecasting and risk management, to water budget studies, and to the design of efficient systems for observing or modeling hydrologic variables. However, as has been mentioned throughout this thesis, it is very difficult

to explain the response because it is the combined manifestation of many complex and interrelated factors that naturally vary both in space and time, and act over a variety of scales.

Basin response and hydrologic fluxes are functions of distributed hydrologic states, most notably of soil moisture. (*Grayson et al.*, 1997) argues that a possible key to collapsing the variability and complexity of basin response might come with better understanding and model representation of soil moisture which controls the partitioning of rainfall into infiltration and runoff, and controls land surface temperature through its effect on the partitioning of available energy into sensible and latent heat fluxes. However, characterization of soil moisture is itself challenging since it is both heterogeneous in space and dynamic in time. Based on a rich literature, the heterogeneity of soil moisture is mainly caused by the spatial variability of topography (*Yeh and Eltahir*, 1998; *Liu et al.*, 2012); soil properties (*Western and Grayson*, 2000; *Kim et al.*, 2002; *Famiglietti et al.*, 1999, 2008); macropore and preferential flow paths (*Brooks et al.*, 2009; *Beven and Germann*, 2013) groundwater-surface water interaction (*Levine and Salvucci*, 1999; *Kollet and Maxwell*, 2008); vegetation (*Eagleson*, 1978; *Ivanov et al.*, 2010); and meteorological forcings e.g. precipitation (*Entin et al.*, 2000; *Western et al.*, 2003; *Vivoni et al.*, 2010). Moreover, soil moisture is dynamic due to the intermittency, seasonality, and inter-annual variability of meteorological forcings (*Entekhabi et al.*, 1995; *Koster and Suarez*, 1999). Both its spatial mean and variance, i.e. its center of mass and distributional features, evolve over time.

This study believes that a possible key to unraveling the complexity of hydrologic response is better understanding and model representation of the space-time dynamics of the hydrologic state variable – the soil moisture field. The approach taken in this thesis has the following unique features:

- focus on distributed soil moisture fields;
- use of a process-based distributed hydrologic model (MOBIDIC) which has a

novel representation of soil moisture;

- use of an entropy-based dimensionless index of hydrologic complexity $\mathcal{H}$ which measure the spatial distribution of soil moisture within a basin;
- use of test basins representing different spatial scales and climatic regimes; and
- grouping of test basins based on similarity e.g. similarity in climatic regime or similarity in catchment area.

As mentioned, this thesis also recognizes the utility of better model representation of soil moisture. To recall, Chapter 2 describes the advantages of process-based distributed hydrologic models (DHMs) over lumped or semi-distributed models. Chapter 3 discussed the horizontal-vertical scale disparity in catchment DHMs and demonstrated that the single layer dual-pore soil structure of MOBIDIC can simulate the magnitude range and dynamics of soil moisture at the local scale, with comparable performance as that of a benchmark model that uses non-linear physics relation. The computational efficiency of the former is advantageous for applications on large or multiple domains such as in this study. *Tromp van Meerveld and McDonnell* (2006a) adds that good DHMs should also be able to capture non-linear behaviors, particularly hysteresis.

Hysteresis refers to the dependence of a system on not only its current state but also on its past state (*Joekar-Niasar et al.*, 2013; *Camporose et al.*, 2014). This phenomenon has long been observed since at least the pioneering works of *Haines* (1930) in the field of soil physics. *O'Kane and Flynn* (2007) gives a review of hysteresis in hydrology at different scales and for different process dynamics. Several mechanisms have been proposed to explain hysteresis such as the flipping of the contact angle of the air-water interface attached to soil particles (*de Gennes et al.*, 2003), the branching of a system near a threshold or a "critical" state (*Phillips*, 2003), and the switching between vertical and lateral fluxes as dominant processes (*Vivoni et al.*, 2010; *Ivanov*

et al., 2010). In soil physics, hysteresis is commonly demonstrated by a looping pattern in the relationship of soil matric potential and soil saturation. The Haverkamp's GRIZZLY database contains probably the largest database of hysteretic soil moisture measurements in the world (*O'Kane*, 2005; *Beven*, 2006a). In catchment hydrology, hysteresis is typically demonstrated by a looping pattern in the discharge-storage plot, where 'storage' can be groundwater table, soil moisture, etc. For instance, *Camporose et al.* (2014) observed clockwise loop in the plot of groundwater table as a function of discharge, where clockwise means that for a given discharge value the groundwater table is lower in the rising than in the falling limb. In terms of the same variables, *Myrabø* (1997) and *McDonnell et al.* (2007b) observed clockwise loop for near-stream areas and counter-clockwise loop for hillslopes.

While observations of hysteresis abound, this phenomenon is difficult to quantify or parameterize in hydrologic models esp. at the basin scale or even at the hillslope scale, thus it is often ignored (*Bras*, 1990; *Jaynes*, 1990; *Beven*, 2006b). Recognizing this challenge, *O'Kane* (2005) presents a simple way to insert hysteresis in hydrologic models using a binary switch. *O'Kane and Flynn* (2007) presents more general mathematical mechanisms of hysteresis using switches, thresholds, and branches. A few modeling studies that explicitly attempted to model hysteresis are described in *Camporose et al.* (2014).

However, a few recent studies show that hysteresis need not be explicitly coded esp. in process-based DHMs. For instance, using a Richards equation based DHM, *Vivoni et al.* (2010) observed hysteresis in the plots of spatial variability vs. mean of depth-averaged soil moisture for a small arid river basin. The macro-hysteresis was argued to be caused by the switching of controls on surface soil moisture variability, from precipitation during wetting stage, to landscape during drying stage. Similarly, *Ivanov et al.* (2010) observed similar hysteresis for a zero-order semiarid basin and explained that the observed macro-hysteresis is related to the switch in

dominant hydrologic processes i.e. vertical fluxes in dry conditions to lateral flows in wet conditions. The evolution of soil moisture variability is said to depend on whether topography-induced subsurface stormflow is triggered. Furthermore, the interplay between different basin properties (soil, climate, and vegetation) modulates the strength of hysteresis. Hysteresis is not explicitly coded in MOBIDIC so it is interesting to see whether this non-linear behavior will be evident in the simulated variables.

6.3 Methods

6.3.1 Measure of Hydrologic Complexity

This chapter uses the revised dimensionless measure of hydrologic complexity $\mathcal{H}$ derived in Chapter 4, as an analysis tool and an organizing principle to better understand the spatiotemporal variability of hydrologic response and fluxes. For clarity, $\mathcal{H}$ is defined as,

$$\mathcal{H} = \exp\left[-\int f(x)\log f(x)\,dx\right] \tag{6.1}$$

where x is the relative soil water deficit, with spatial distribution $f(x)$ for $x \in [0,1]$, and $\int_0^1 f(x)dx = 1$. Since MOBIDIC uses a dual-compartmentalized soil, the complexity of the soil capillary and gravity reservoirs, $\mathcal{H}_c$ and $\mathcal{H}_g$, respectively, are also investigated.

$$\begin{aligned} \mathcal{H} &= \mathcal{H}(x), & x &= \forall_e/\forall_{e,max} \\ \mathcal{H}_c &= \mathcal{H}_c(x_c), & x_c &= \forall_{e,c}/\forall_{e,max} \\ \mathcal{H}_g &= \mathcal{H}_g(x_g), & x_g &= \forall_{e,g}/\forall_{e,max} \end{aligned} \tag{6.2}$$

The definitions of x_c and x_g are the same as those used in Example 2 in Chapter 4 (Section 4.4.2). Note the use of the same normalizer $\forall_{e,max}$. The range of x is the interval [0,1] but since $\forall_{e,max} = W_{c,max} + W_{g,max}$, the range of x_c and x_g are only

subsets of this interval,

$$x_c \in [0, W_{c,max}/\forall_{e,max}] \tag{6.3}$$

$$x_g \in [0, W_{g,max}/\forall_{e,max}] \tag{6.4}$$

$$x = x_c + x_g, \quad x \in [0, 1] \tag{6.5}$$

As shown in Section 4.4.2, these definitions of x_c and x_g are good for understanding the contribution of $\mathcal{H}_c$ and $\mathcal{H}_g$ to the total complexity $\mathcal{H}$. In this study, $\mathcal{H}$, $\mathcal{H}_c$, and $\mathcal{H}_g$ are applied on a number of test basins to answer the following science questions:

- How do distributional features of soil water deficit evolve over time?
- Is there hysteresis in the evolution?
- What factors (physiography, vegetation, climate, etc.) affect the distribution and its evolution?
- How are the distributional features related to basin response and hydrologic fluxes?

6.3.2 Test Basins

Table 1 lists the eight test basins used for this study. The first column shows the basin ID with the letter denoting the climate ('S' for semiarid, 'T' for temperate, and 'H' for humid) and the number denoting the size of the basin ('0', '1', '2', and '3', means an area of about 10^0, 10^1, 10^2, and 10^3 sq. km., respectively). The last column shows the dominant land cover and land use. The location maps of these basins are shown in Figure 6.1. Basins S0, S1, S2, S3 and T0 cover field sites of the Soil moisture Sensing Controller and oPtimal Estimator (SoilSCAPE) Project, see *Moghaddam et al.* (2013). T2 and T3 are tests basins in the Distributed Model Intercomparison Project (DMIP), see *Reed et al.* (2004). Basin H2 is the study basin

used in Chapter 5. Additional properties of the test basins are provided in Table 2. Calibrated MOBIDIC models of each of the test basin was developed using $5.3\text{x}10^3$ to $3.5\text{x}10^4$ pixels, and 3-year simulation period at hourly time steps. Simulated hourly variables were later aggregated to daily values.

6.3.3 Relationship of $\mathcal{H}$ to Other Hydrologic Variables

Basin Attributes as Predictors of $\mathcal{H}$

The characteristic distribution of soil moisture might be related to observable basin attributes. To check this, the temporal mean of $\mathcal{H}$, $\mathcal{H}_c$, and $\mathcal{H}g$, were related to the following basin attributes:

- *Catchment area* A [km^2] – a basic spatial scaling predictor.
- *Median slope* s_{50} [-] – the gradient along hillslopes, which influences both surface and subsurface runoff. The median was chosen instead of the mean to avoid bias from a few very steep points.
- $\log_{10}(A/s_{50})$ [km] – the basin-scale equivalent of the topographic wetness index (TWI) of *Beven and Kirby* (1979).
- *Relief ratio* r_r [-] – the range in elevation, $z_{max} - z_{min}$, normalized by the length of the mainstream L. As opposed to s_{50}, this gives an estimate of the gradient along channels.
- *Drainage density* D_d [km^{-1}] – the ratio of total stream length to A. High D_d implies high channelization of precipitation and shorter residence time. This is also indicates the extent of riparian zones.
- *Mean soil moisture capacity* $\overline{\forall_{e,max}}$ [m]. – the spatial mean of the product of the saturated soil moisture content θ_{sat} [-] and the height d [m] of the modeled soil layer.

- *Aridity* $\overline{E_p}/\overline{P}$ [-] – the ratio between the annual mean potential evaporation (computed based on *Penman* (1948)) divided by the annual mean precipitation.

- *Infiltration ratio* i_s/K_s [-] – the mean precipitation intensity during a storm event, i_s, divided by K_s. Typically, $i_s/K_s < 1$ because i_s is averaged over the entire storm duration which includes short periods without rain.

$\mathcal{H}$ as Predictor of Hydrologic Response

The utility of $\mathcal{H}$, $\mathcal{H}_c$, and $\mathcal{H}g$ to explain the basin response is also investigated by relating them with the traditional measures of hydrologic response listed below. The overbar symbol indicates temporal average.

- ET efficiency $\overline{ET}/\overline{E_p}$ – ratio of basin-averaged ET to the potential rate E_p

- Runoff ratio $\overline{Q}/\overline{P}$ – fraction of precipitation that becomes runoff

- Baseflow index BFI $(\overline{Q_{bf}}/\overline{Q})$ – fraction of streamflow contributed by baseflow

6.4 Results and Discussion

6.4.1 Basin-scale Partitioning of Precipitation

Calibrated MOBIDIC models were developed for each of the test basins. The basin-scale partitioning of precipitation is shown in Figure 6.2. Looking at the test basins altogether, there is no clear pattern. However, by grouping the basins based on catchment area, clear and logical patterns emerge. As shown, for basins with similar catchment area, the basin with higher median slope has higher runoff ratio and lower ET efficiency. Basin T0 has the lowest runoff ratio because in addition to having the lowest median slope, it has anthropogenic features that increase its infiltration capacity and allow impoundment of runoff. The net groundwater recharge constitutes

only a small fraction of the annual precipitation in all the test basins especially in the large basins where the spatial mean of percolation and channel leakage are balanced by capillary rise and channel seepage.

The basins can also be grouped according to climate. Compare the four semiarid basins by picturing columns 2, 3, 4, and 8 of Figure 6.2 together. S0 and S2 have relatively higher ET efficiency because they are mostly covered by woody savanna whereas S1 and S3 are mostly covered by grasslands.

6.4.2 Hydrologic Complexity

Figure 6.3 shows the maps and PDFs of the characteristic relative soil water deficit x for basin S1 which is the least complex with $\mathcal{H} = 0.05$, and basin H2 which is the most complex with $\mathcal{H} = 0.57$. Clearly the map for T2 is much more complex than that of S1. For the former, x spans almost the entire interval [0,1] and its $f(x)$ is closer to a uniform distribution than a Dirac delta. In contrast, the entire basin S1 is often very dry so its map appears only in shades of blue and its $f(x)$ is concentrated near $x = 0$, making it more similar to a Dirac delta than a uniform distribution. These mean that qualitatively, $\mathcal{H}$ is a correct measure of complexity.

Figure 6.4 shows the hyetograph and time series of $\mathcal{H}$, $\mathcal{H}_c$, and $\mathcal{H}_g$ for basins S2, T2, and H2. First for the semiarid basin (top plot), there is marked seasonality with most of the year being dry and the months of November to February receiving almost all of the annual precipitation. The soil moisture distribution lags precipitation by about a month. In dry conditions, most of the gravity reservoirs are completely empty so $\mathcal{H}_g$ is low and its value corresponds to the complexity of the time-invariant maximum capacity of the gravity reservoirs, $\mathcal{H}_g^*$,

$$\mathcal{H}_g^* = \mathcal{H}_g \left(\forall_{e,g,max} / \forall_{e,max} \right) \tag{6.6}$$

Meanwhile, there is more spatial variability in x_c so $\mathcal{H}_c$ is higher and it contributes most of the total complexity $\mathcal{H}$. In wet conditions, the trend is reversed. Since the capillary reservoirs are saturated, $x_c \approx 0$ across the basin so $\mathcal{H}_c \to 0$. Meanwhile the gravity reservoirs are variably saturated so $\mathcal{H}_g$ is high and it accounts for most of the total complexity. Similar trends are seen in the timeseries of the humid basin (bottom plot). The other two years simulated for this basin are similar to the wet months of June to September. Although this basin receives a lot of rain, it has mountainous topography and well-drained soil so the gravity reservoirs never become completely saturated, hence x and x_g are spatially variable, and as a result $\mathcal{H} > 0$ and $\mathcal{H}_g > 0$. In the temperate basin (middle plot), the soil has intermediate total wetness year-round. Most of the capillary reservoirs are completely or partially saturated so $\mathcal{H}_c$ is low; while most of the gravity reservoirs are either empty or only partially saturated so $\mathcal{H}_g$ is high and it accounts for most of the total complexity of the basin.

The right side of Figure 6.4 shows the 3-year median values. The median of $\mathcal{H}$ is highest for T2 and roughly the same for S2 and H2. Moreover, as the climatic wetness increases, the median of $\mathcal{H}_c$ decreases. In terms of magnitude, the median of $\mathcal{H}_g$ is highest for T2. However, the relative contribution of the gravity reservoir to the total basin complexity, $\mathcal{H}_g/(\mathcal{H}_c + \mathcal{H}_g)$, increases with climatic wetness. The opposite relationship is observed, for the relative contribution of the capillary reservoir, $\mathcal{H}_c/(\mathcal{H}_c + \mathcal{H}_g)$. These indicate that the dominant controls on the spatial distribution of hydrologic states and fluxes vary with climate.

In the timeseries of H2, the markers 1- 3 are representative days with dry, intermediate, and wet conditions, respectively. The PDFs of x, x_c, and x_g for these representative days are illustrated in Figure 6.5. In dry conditions (point 1), most gravity reservoirs are empty and most capillary reservoirs are variably saturated, so the center of mass of $f(x_g)$ and $f(x_c)$, and consequently also of $f(x)$, are far from 0 or 1, and the complexity indices are all positive. A physical interpretation of this is that the

heterogeneity of available capillary and gravity storage both have spatially-variable effects on e.g. saturation-excess runoff. $\mathcal{H}_c > 0$ also indicates spatial variability of fluxes related to the state of the capillary reservoirs, e.g. ET. As can be seen, $f(x_c)$ has a wider spread than $f(x_g)$ so $\mathcal{H}_c > \mathcal{H}_g$. In intermediate conditions (point 2), $f(x_c)$ has shifted towards $x_c = 0$ and is now less spread resulting to a decrease in $\mathcal{H}_c$. On the other hand, $f(x_g)$ is practically unchanged since precipitation has so far only wet the capillary reservoirs. Finally in wet conditions (point 3), $f(x_c)$ is roughly a Dirac delta at $x_c = 0$ thus $\mathcal{H}_c \approx 0$. Some physical interpretations are i) ET is spatially deterministic, in this case it occurs at the potential rate across the basin; and ii) the states of the capillary reservoirs do not cause variability in e.g. runoff. Since $f(x)$ and x_g are the same, the total complexity of the basin is only due to the complexity of available gravity storage.

Based on the definition of the hydrologic complexity index and the definition of x, x_c, and x_g used here, as the entire basin becomes completely saturated,

$$\lim_{x \to 0} \mathcal{H} = \lim_{x \to 0} \mathcal{H}_c = \lim_{x \to 0} \mathcal{H}_g = 0 \tag{6.7}$$

On the other extreme, if the entire basin becomes completely dry,

$$\lim_{x \to 1} \mathcal{H} = \mathcal{H}\left(\forall_{e,max}/\forall_{e,max}\right) = 0 \tag{6.8}$$

$$\lim_{x_c \to 1} \mathcal{H}_c = \mathcal{H}_c\left(W_{c,max}/\forall_{e,max}\right) = \mathcal{H}_c^* \tag{6.9}$$

$$\lim_{x_g \to 1} \mathcal{H}_g = \mathcal{H}_g\left(W_{g,max}/\forall_{e,max}\right) = \mathcal{H}_g^* \tag{6.10}$$

Since $\mathcal{H}_c^*$ and $\mathcal{H}_g^*$ are typically > 0, the sum of the complexity of the dual reservoirs are usually higher than $\mathcal{H}$ in very dry conditions. For these conditions, the relative contribution of the capillary and gravity reservoirs to the total basin complexity can be evaluated by simply comparing the relative magnitude of $\mathcal{H}_c$ and $\mathcal{H}_g$.

In order to see the behavior of $\mathcal{H}$, $\mathcal{H}_c$, and $\mathcal{H}_g$ between the two extreme conditions and see the effects of wetting and drying, Figure 6.6 shows scatter plots of daily $\mathcal{H}$, $\mathcal{H}_c$, and $\mathcal{H}_g$, vs. x. As expected, $\mathcal{H}_c$ is close to 0 in wet conditions, then it increases as the basin dries. Further drying, as can be seen in basin S2 and T2, eventually causes a decrease in $\mathcal{H}_c$. On the other hand, $\mathcal{H}_g$ is generally high when the basin is wet (but not saturated) then decreases as the basin dries, and eventually approaches $\mathcal{H}_g^*$ at very dry conditions.

Figures 6.7–6.10 show the temporal mean of $\mathcal{H}$, $\mathcal{H}_c$, and $\mathcal{H}_g$ of the 8 test basins and their relationships with some of the tested basin attributes. The mean is used instead of the median because it is more representative of the dynamic values. As pointed in the analysis of basin-scale precipitation partitioning, it is difficult to see relationships if all basins are analyzed together. Clearer and more logical relationships are found by comparing only basins with similar climate or size. For instance, Figure 6.7 shows how the complexity indices vary with aridity for basins of similar spatial scale. Overall, the total complexity $\mathcal{H}$ (and also $\mathcal{H}_g$) is highest in temperate basins. This is consistent with the understanding that soil moisture is lower- and upper-bounded, so its variability peaks at intermediate conditions, see e.g. *Laurence and Hornberger* (2007). Independently, $\mathcal{H}_c$ of S2 and S3 are higher than temperate and humid basins of similar size because the capillary reservoirs are often saturated in the temperate and humid basins i.e. they are less hydrologically active. Figure 6.8 shows that for the semiarid and temperate test basins, $\mathcal{H}$, $\mathcal{H}_c$, and $\mathcal{H}_g$ all increase with catchment area $\log_{10}A$ (top row). This is logical because across smaller area, there is naturally less variability in e.g. soil properties and land cover. The relationships with the basin-scale topographic wetness index $\log_{10}(A/s_{50})$ (bottom row) are practically the same as with $\log_{10}A$, with marginal improvement for the semiarid basins and marginal degradation for the temperate basins. Furthermore, Figures 6.9 and 6.10 show that $\mathcal{H}$, $\mathcal{H}_c$, and $\mathcal{H}_g$ of the semiarid and temperate basins are negatively cor-

related to relief ratio, and positively correlated to infiltration ratio. Weaker or no significant relationships were found for the other basin attributes considered.

The slope of the trendlines for $\mathcal{H}$ and $\mathcal{H}_g$ in Figure 6.8–6.10 are generally the same for the semiarid and temperate basins. The linear relationships would have been even better if $\mathcal{H}_g$ (and thus $\mathcal{H}$) of basin T2 is somewhat lower. T2 has many watering holes and infiltration ponds which although altogether covers only about one percent of the basin, has significant impacts on the hydrology of the basin. It is hypothesized that these structures artificially increase $\mathcal{H}_g$ since they make some areas wetter than they naturally should be while at the same time minimizing surface and subsurface runoff which otherwise can have a homogenizing effect on soil moisture of downstream areas. For future work, the calibrated models of the test basins can be re-run without the anthropogenic factors. It is emphasized that the current study is exploratory and more basins with various settings, scales, and climates, are needed to verify the relationships found.

Figure 6.5 shows the relationship between the temporal mean of the complexity indices with three traditional measures of hydrologic response. The top row shows that the ET efficiency is practically independent of the complexity indices. In fact and as expected, ET efficiency is mainly a function of aridity. The middle row shows that the runoff ratio generally increases with $\mathcal{H}$, $\mathcal{H}_c$, or $\mathcal{H}_g$, but the relationships are weak. The bottom plot shows a stronger positive linear correlation between BFI and the three complexity indices. Notice that the relationships with runoff ratio or BFI are stronger with $\mathcal{H}$ or $\mathcal{H}_g$, than with $\mathcal{H}_c$, because the state of the capillary reservoirs have less influence on runoff. Again, more basins are needed to verify these relationships or the lack thereof.

6.4.3 Hysteresis and Threshold Behaviors

The scatter plots of $\mathcal{H}$ in Figure 6.6 especially in the dry regime of basins H2 and T2, show looping patterns which indicate hysteresis of soil moisture at the basin scale. Although not as clear from these plots, looping patterns are also seen for the dry regime of basin S2, and also for the wet regime of these basins when the points are plotted sequentially one at a time. Hysteresis is not explicitly coded in MOBIDIC so the observation of this non-linear behavior in simulated variables is an encouraging result. The observed hysteretic behaviors emerge as a result of MOBIDIC's use of a dual-pore soil structure which not only captures the different roles of capillary and gravity-driven processes, but also mimics the threshold-like behavior of runoff generation.

To better understand the nature of the observed hysteresis in the relationship of $\mathcal{H}$ and storage, Figure 6.12 shows the evolution of the probability distribution of $\forall_e$ for basin H2 in dry conditions. As the mean $\forall_e$ increases, the distribution widens, thereby increasing $\mathcal{H}$. During initial wetting, the mean $\forall_e$ decreases but the distribution simply translates to the left with little change in the shape of $f(\forall_e)$. Further wetting causes significant narrowing of the PDF which corresponds to a sharp decrease in spatial variability. The threshold or critical state is when most soil elements are at field capacity, above which gravity reservoirs start to fill-up which triggers lateral subsurface flow leading to greater hydraulic connectivity of the landscape or homogenization of the soil moisture fields.

Figure 6.5 shows hysteresis in the MOBIDIC-simulated discharge Q vs. the basin-averaged relative soil water deficit $\overline{x}$ and Q vs. $\mathcal{H}$ for basin H2 in dry conditions i.e. low-flow regime, basin S2 in dry conditions, basin T2 in wet conditions i.e. high-flow regime, and basin S2 in wet conditions, respectively. The red and blue arrows indicate drying and wetting directions, respectively. The direction is easier to follow on the left plots since increase in $\overline{x}$ means drying, i.e. arrow pointing to the right, while

wetting is vice versa.

Focus first on the Q vs. $\overline{x}$ plots on the *left*. These plots can be flipped horizontally to obtain the more common plots of discharge-storage, since relative saturation is equal to one minus relative deficit. $\overline{x}$ is used here for consistency with the rest of the Chapter and to allow easy cross-referencing with scatter plots of $\mathcal{H}$ vs. $\overline{x}$ in Figure 6.6. The red arrows point to the right or towards increasing in $\overline{x}$, whereas the blue arrows point to the left. Except for basin T2 in wet conditions, the loops are clockwise. However the evolution is not unidirectional as the points can go back-and-forth along a trajectory e.g. see the bottom left area of the plot for H2 in dry conditions which show alternating wetting and drainage.

In addition to the horizontal components, the arrows also have vertical components. The evolution of Q vs. $\overline{x}$ has two limiting cases: i) a change in $\overline{x}$ with little change in Q (mostly horizontal movement); or ii) a a change in Q with little change in $\overline{x}$ (mostly vertical movement). The first case is when the dominant hydrologic fluxes are vertical fluxes e.g. ET or wetting of the capillary reservoirs. The other case is when lateral surface or subsurface runoff are the dominant processes.

In the left plots, a given $\overline{x}$ corresponds to non-unique values of Q. A possible explanation is that a given $\overline{x}$ can have non-unique values of $\mathcal{H}$ as shown in Figure 6.6. This is similar to the notion that a mean can have different associated variance. Difference in $\mathcal{H}$ means difference in spatial distribution of relative soil water deficit, which implies i) difference in the extent of runoff contributing area (either fully or partly wet); and ii) difference in the heterogeneity or connectivity of wet areas. These differences in spatial patterns result to differences in discharge for the same $\overline{x}$.

The right plots of Figure 6.5 show Q vs. $\mathcal{H}$. Although the patterns are not necessarily the same as those on the left, these plots also show hysteresis. As shown in Figure 6.6, a given $\mathcal{H}$ can correspond to different values of $\overline{x}$, and thus different values of Q as shown here.

For illustration, consider basin H2 in dry conditions. From the dry initial state ($\overline{x} \approx 0.36$, $Q \approx 5,200L/s$), further drying decreases Q and increases $\mathcal{H}$. Initial rain typically decreases $\overline{x}$ with little change in Q as most of the rain are absorbed by the capillary reservoirs. Occasionally if the rain is intense, some infiltration-limited (i.e. Hortonian) runoff will be produced, and these are shown in the plots by a sharp increase in Q during wetting, followed by a sharp decrease in Q during drainage. Additional rain further decreases $\overline{x}$ and once a threshold is reached, $\mathcal{H}$ decreases to low values while Q sharply increases. Again, the threshold is when most of the basin is at or above field capacity, $\overline{x} \approx 0.4$. This threshold behavior of runoff generation has been well analyzed in hydrological literature e.g. *Zehe et al.* (2005); *Tromp van Meerveld and McDonnell* (2006a,b); *Penna et al.* (2013).

The Q vs. $\overline{x}$ plots for the high flow regimes also show hysteresis but the mechanisms are more complicated. As mentioned, counter-clockwise loop is observed in basin T2 while a clockwise loop is observed in basin S2. Moreover, wetting can correspond to either increase or decrease in Q. The differences and complications are hypothesized to be related to differences in magnitude range and spatial distribution of infiltration capacity, and differences in concentration time for both surface and sub-surface runoff.

Hysteresis and threshold-like behavior are manifestations of non-linear hydrologic response. According to *Tromp van Meerveld and McDonnell* (2006b), the detection of these non-linear behaviors can be used as a benchmark for models to capture and perhaps a way to classify behaviors of different hillslopes and basins. This study shows that MOBIDIC with its use of a dual-pore soil structure is able to mimic these non-linear behaviors, and more importantly, the mechanisms that cause these behaviors to emerge are hydrologically meaningful. Moreover, the use of entropy-based measures of hydrologic complexity helps understand the link between spatial patterns and processes, or the micro-states of the basin to the macro-state (e.g. Q

and BFI) and the macro-behaviors (hysteresis and threshold behavior).

6.5 Summary and Conclusions

Basin response and hydrologic fluxes are functions of hydrologic states. In this study the focus is on soil moisture. To characterize the spatial distribution of soil moisture and understand its evolution in time, a revised formulation of the *Martina and Entekhabi* (2006) dimensionless index of hydrologic complexity $\mathcal{H}$ which measures the distance of a given distribution from two limiting distributions (Dirac delta and uniform distribution), is used. The modifications make $\mathcal{H}$ discretization-invariant. The revised $\mathcal{H}$ is based on differential entropy instead of Shannon entropy, and is computed for a dimensionless random variable that is strictly in the interval [0,1]. Here the relative soil water deficit is used.

For demonstration, $\mathcal{H}$ is applied on 8 test basins with area ranging from 10^0–10^3 km^2 and representing semiarid, temperate and humid climates. Calibrated models of these basins were developed using the distributed hydrologic model MOBIDIC. Since a key feature of MOBIDIC is the partitioning of each pixel of soil into a gravity reservoir and a capillary reservoir, the complexity of these sub-systems, $\mathcal{H}_c$ and $\mathcal{H}_g$, respectively, were also investigated . It was shown that $\mathcal{H}$, $\mathcal{H}_c$ and $\mathcal{H}_g$, can track the evolution of the distributional features of soil moisture and can provide insights on the switching of dominant hydrologic processes.

Furthermore, this study explored what and how basin attributes affect the characteristic value of $\mathcal{H}$, $\mathcal{H}_c$ and $\mathcal{H}_g$, and how these indices can be used to explain inter-basin variability in hydrologic response. Clear and meaningful relationships were found only by grouping basins with similar climate or size. For basins of similar size, $\mathcal{H}$ and $\mathcal{H}_g$ are highest in temperate climate, consistent with soil moisture being lower- and upper-bounded so its variability peaks at intermediate values. $\mathcal{H}_c$ gen-

erally increases with aridity since the capillary reservoirs are often saturated in the temperate and humid basins i.e. they are less hydrologically active. For the semiarid and temperate basins, $\mathcal{H}$, $\mathcal{H}_c$ and $\mathcal{H}_g$, are positively correlated with catchment area, basin-scale topographic wetness index, and infiltration ratio, but negatively correlated with relief ratio; while $\mathcal{H}$ and $\mathcal{H}_g$ are positively correlated with baseflow index. No significant relationships were found for the other basin attributes considered, as well as with ET efficiency and runoff ratio. The current study is exploratory and it is emphasized that more basins with various settings, scales, and climates, are needed to verify the relationships found.

Hysteresis and threshold-like behavior are manifestations of non-linear hydrologic response. The detection of these non-linear behaviors is a potential benchmark for hydrologic models to capture and perhaps a way to classify behaviors of different hillslopes and basins. This study shows that although hysteresis is not explicitly coded in MOBIDIC, the results clearly demonstrate this phenomenon. More importantly, the mechanisms that cause these behaviors to emerge in the simulated variables are found to be hydrologically meaningful. Hysteresis emerges as a result of MOBIDIC's use of a dual-pore soil structure which not only captures the different roles of capillary and gravity-driven processes, but also mimics the threshold behavior of runoff generation. Moreover, investigation of these non-linear behaviors is aided by the use of entropy-based measures of hydrologic complexity which help understand the link between spatial patterns and processes or the micro-states of the basin (i.e internal soil moisture distribution), to the macro-states e.g. Q and the macro-behaviors e.g. hysteresis and threshold behavior.

Table 6.1: Description of the test basins

ID	Name and Location	Climate	Land cover/use
S0	Tonzi Ranch, CA	semiarid	woody savanna
S1	Vaira Ranch, CA	semiarid	grassland, woody savanna, shrubland
S2	Willow Creek Basin, CA	semiarid	woody savanna, grassland, shrubland
S3	San Joaquin R.B., CA	semiarid	grassland, farmland, woody savanna
T0	SoilSCAPE site, Canton, OK	temperate	pasture with some trees
T2	Baron Fork, Dutchmills, AR	temperate	pasture, farmland, forest
T3	Blue River Basin, OK	temperate	grassland, farmland, woodlands
H2	Tamugan Basin., Philippines	humid	upland tropical rain forest

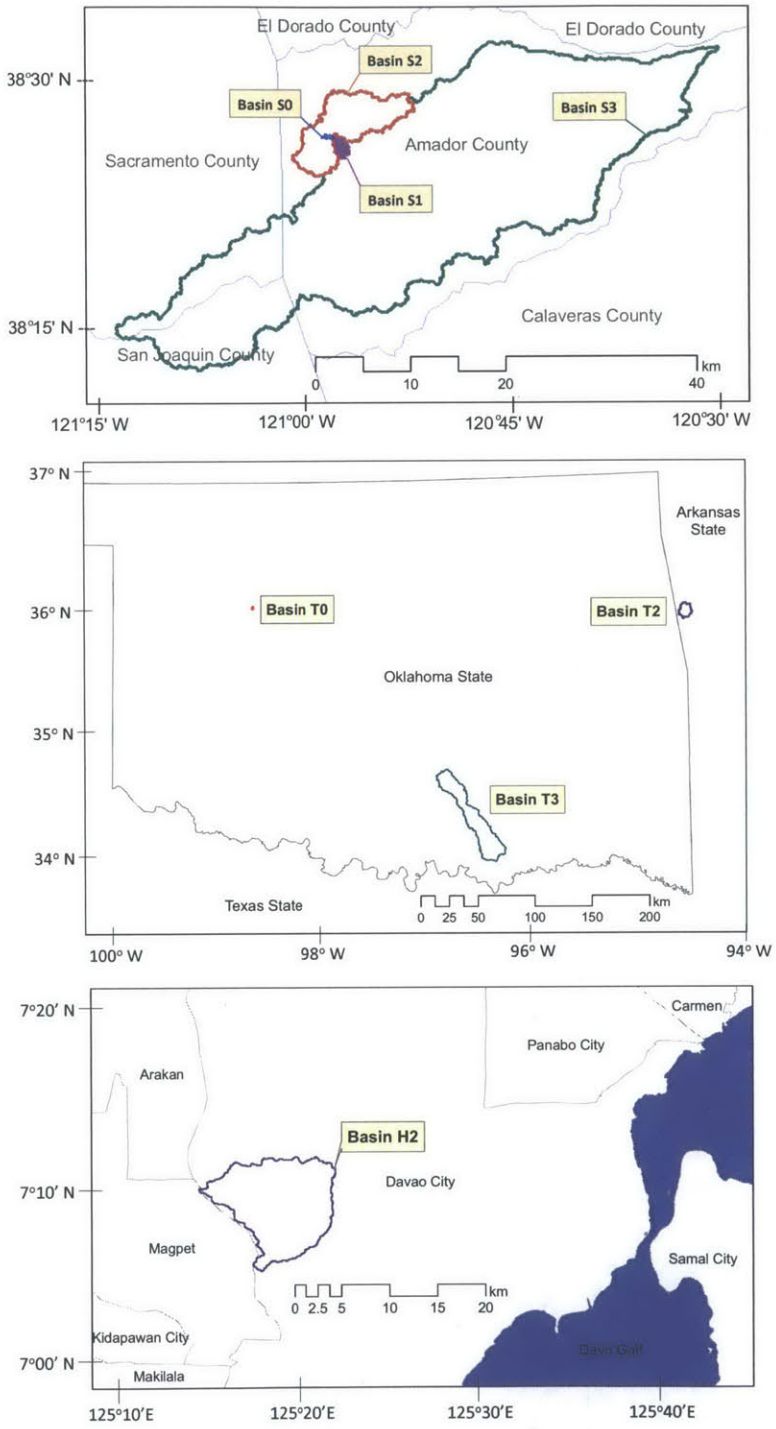


Figure 6.1: The location of the test basins: top) nested semiarid basins in California; middle) temperate basins in Oklahoma-Arkansas; and bottom) humid basin in the Philippines

Table 6.2: Properties of the test basins

		Topography		Geomorph.		- - - Soil Properties - - -				Climate		Storm Prop.		Mixed
ID	Area	Relief	s_{50}	D_d	L	$\overline{K_s}$	$\overline{d}$	$\overline{\theta_{fld}}$	$\overline{\theta_{sat}}$	$\overline{P}$	$\overline{E_p}/\overline{P}$	i_s	t_s	i_s/K_s
	km^2	m	[-]	km^{-1}	km	mm/hr	mm	[-]	[-]	mm/yr	[-]	mm	hr	[-]
S0	0.45	34	0.04	2.67	0.98	28	368	0.15	0.24	569	3.71	11.1	14.4	0.028
S1	3.09	76	0.14	2.42	2.92	45	353	0.18	0.30	579	3.73	11.1	14.4	0.014
S2	54	316	0.08	2.17	19.2	17	842	0.24	0.43	576	3.50	11.1	14.4	0.046
S3	838	316	0.18	1.17	91.2	17	613	0.14	0.31	580	3.48	11.1	14.4	0.046
T0	0.45	18	0.03	0.73	0.16	38	1113	0.15	0.29	1210	1.64	14.4	9.8	0.039
T2	108	252	0.11	1.95	18.9	11	594	0.16	0.30	1027	1.61	9.0	12.5	0.066
T3	1259	242	0.03	0.36	143.	7	1483	0.15	0.41	941	1.92	10.1	13.9	0.108
H2	103	1594	0.36	083	17.0	13	916	0.29	0.45	3877	0.56	23.0	12.0	0.153

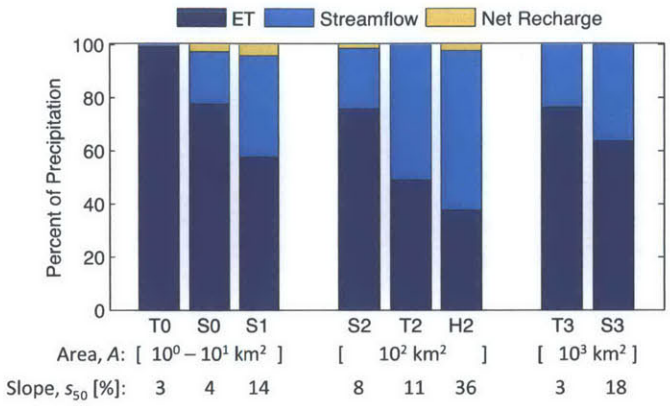


Figure 6.2: Basin-scale precipitation partitioning with basins grouped by catchment area and arranged in increasing median slope

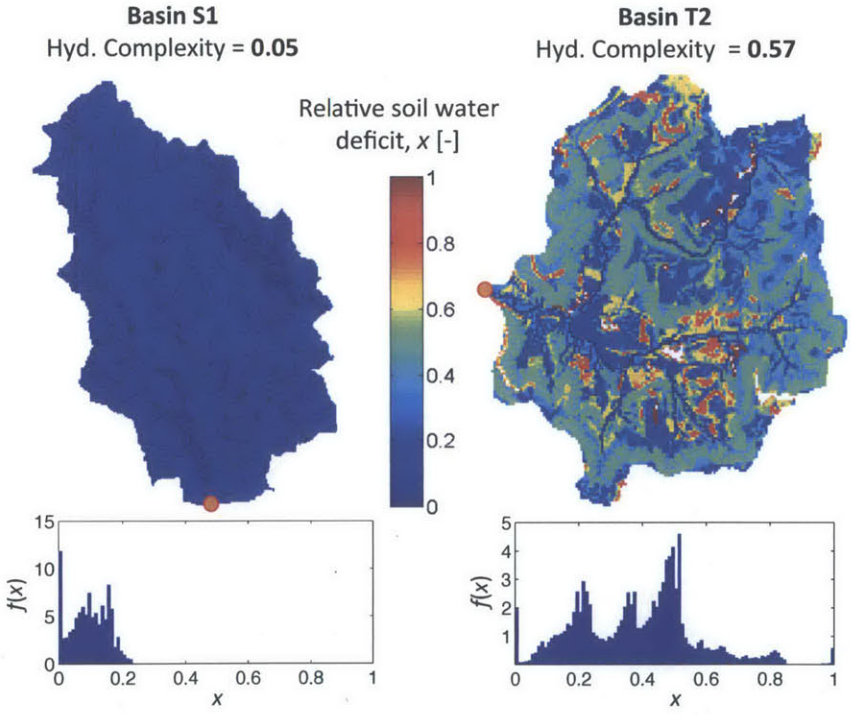


Figure 6.3: Map and PDF of characteristic soil water deficit, $x = \forall_e / \forall_{e,max}$ [-] of basins S1 (left) and T2 (right). Orange dots mark basin outlets.

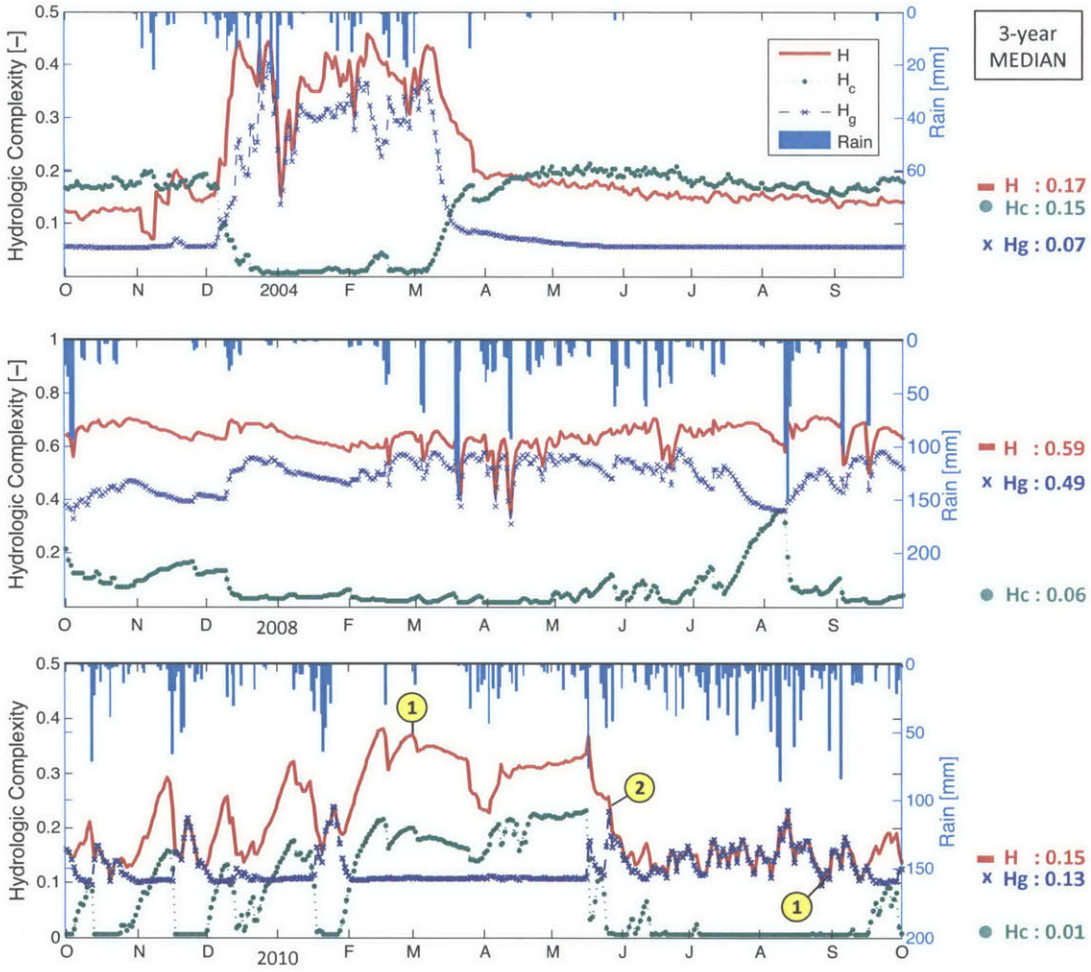


Figure 6.4: Time series of $\mathcal{H}$, $\mathcal{H}_c$, and $\mathcal{H}_g$ for basins S2 (top), T2 (middle), and H2 (bottom). On the right side are the 3-year median values. Days marked on basin H2 are representative of dry (1), intermediate (2), and wet (3) conditions

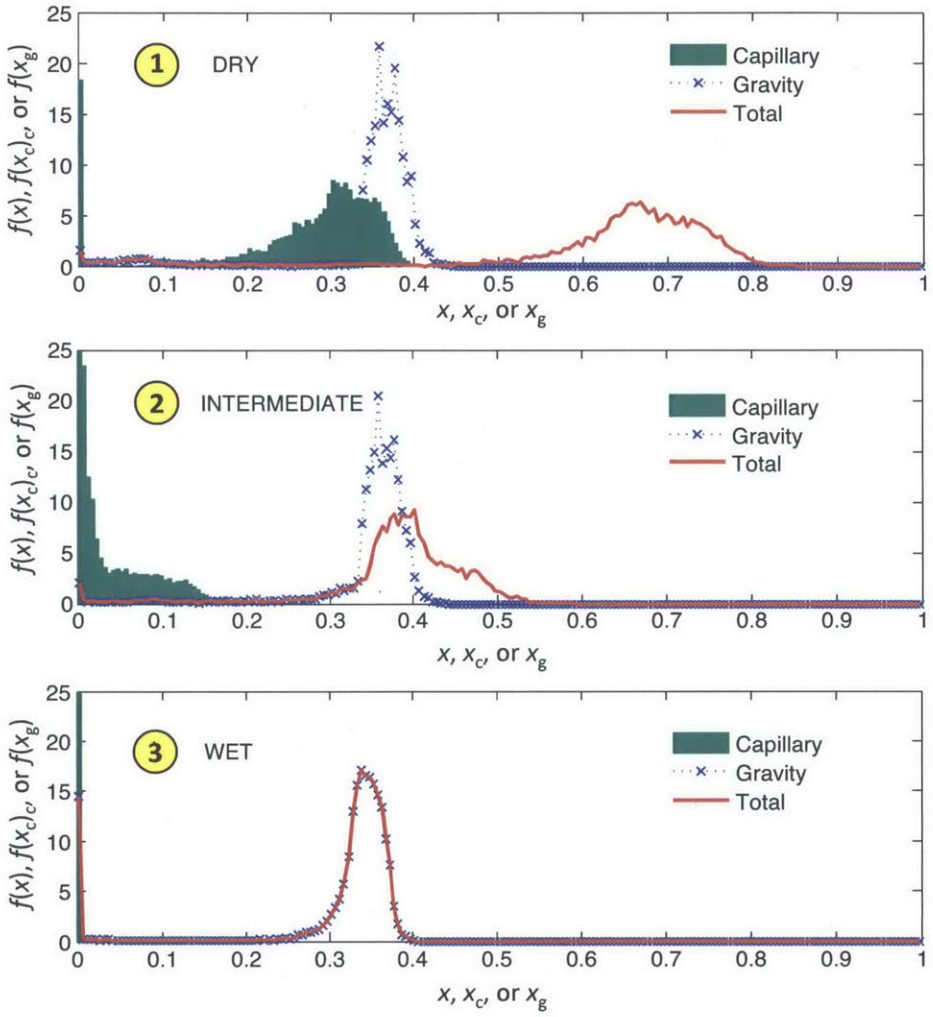


Figure 6.5: Example spatial PDF of relative soil water deficit at basin H2 in dry, intermediate, and wet conditions.

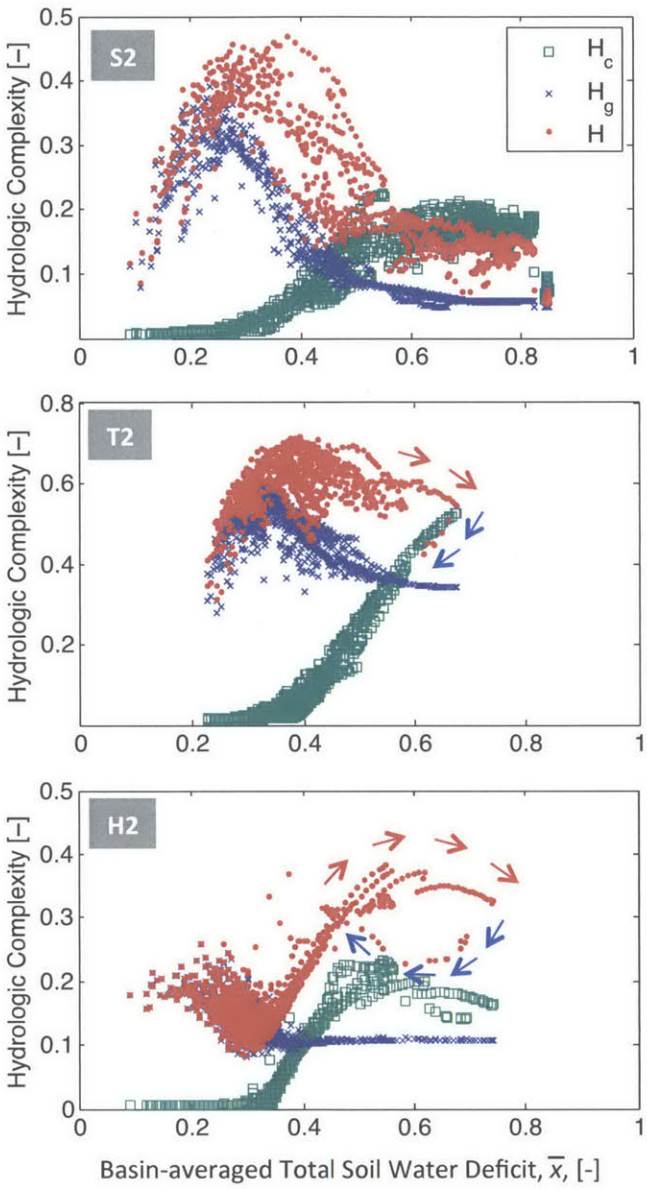


Figure 6.6: $\mathcal{H}$, $\mathcal{H}_c$, and $\mathcal{H}_g$ vs. basin-averaged total soil water deficit x for basins S2 (top), T2 (middle), and H2 (bottom). Red and blue arrows indicate drying and wetting directions for $\mathcal{H}$, respectively

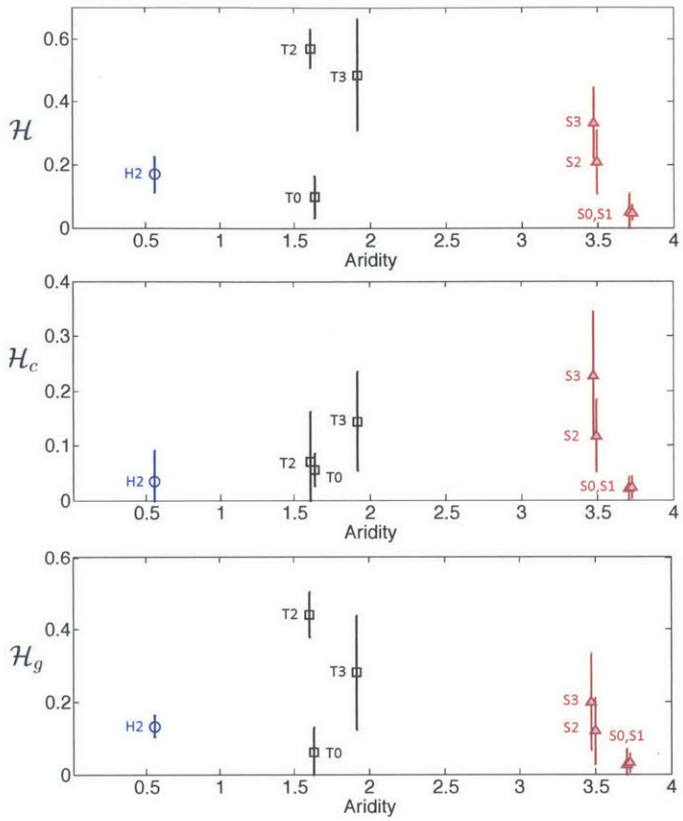


Figure 6.7: From top: $\mathcal{H}$, $\mathcal{H}_c$, and $\mathcal{H}_g$, vs. aridity [-]. Points are temporal mean and whiskers are range; circle (humid); squares (temperate); triangles (semiarid).

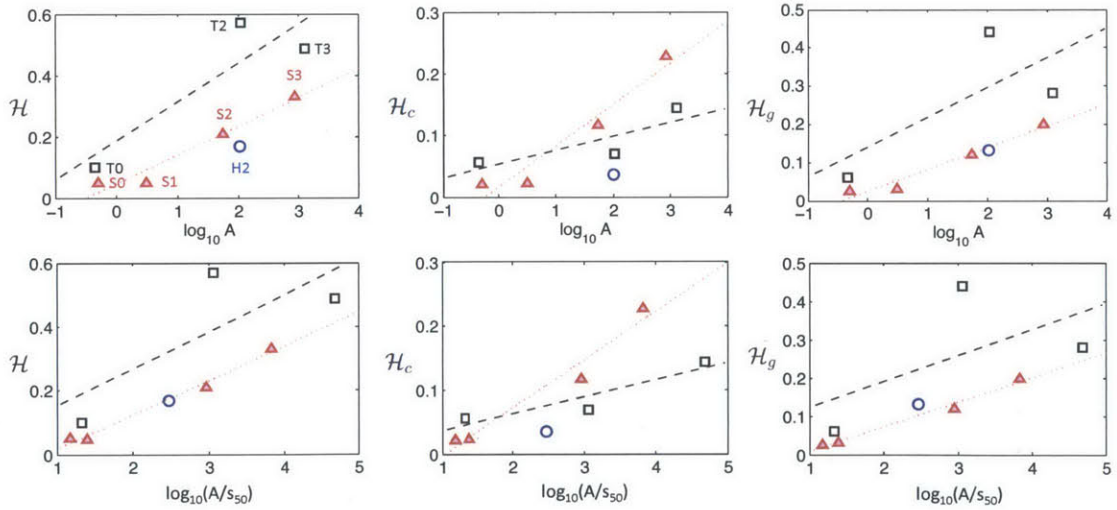


Figure 6.8: Temporal mean of (left to right) $\mathcal{H}$, $\mathcal{H}_c$, and $\mathcal{H}_g$, vs. spatial scales: top) catchment area $\log_{10}A$ [km^2], and bottom) basin-scale topographic wetness index $\log_{10}(A/s_{50})$. Trendlines are for semiarid (red) and temperate (black) basins.

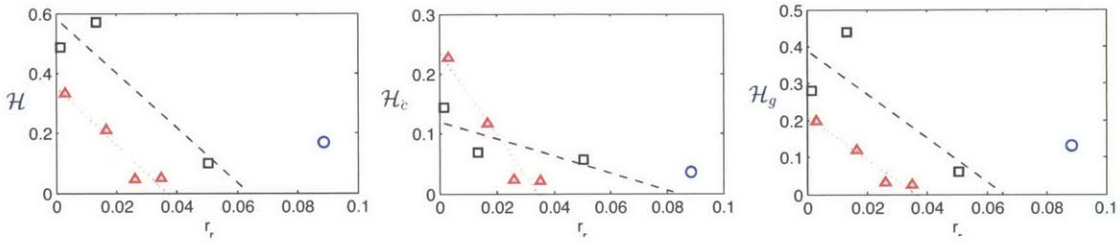


Figure 6.9: Temporal mean of (left to right) $\mathcal{H}$, $\mathcal{H}_c$, and $\mathcal{H}_g$, vs. relief ratio r_r [-]. Trendlines are for semiarid (red) and temperate (black) basins.

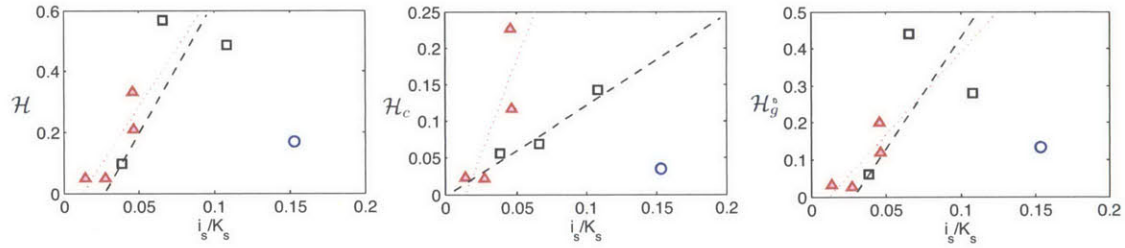


Figure 6.10: Temporal mean of (left to right) $\mathcal{H}$, $\mathcal{H}_c$, and $\mathcal{H}_g$, vs. infiltration ratio i_s/K_s [-]. Trendlines are for semiarid (red) and temperate (black) basins.

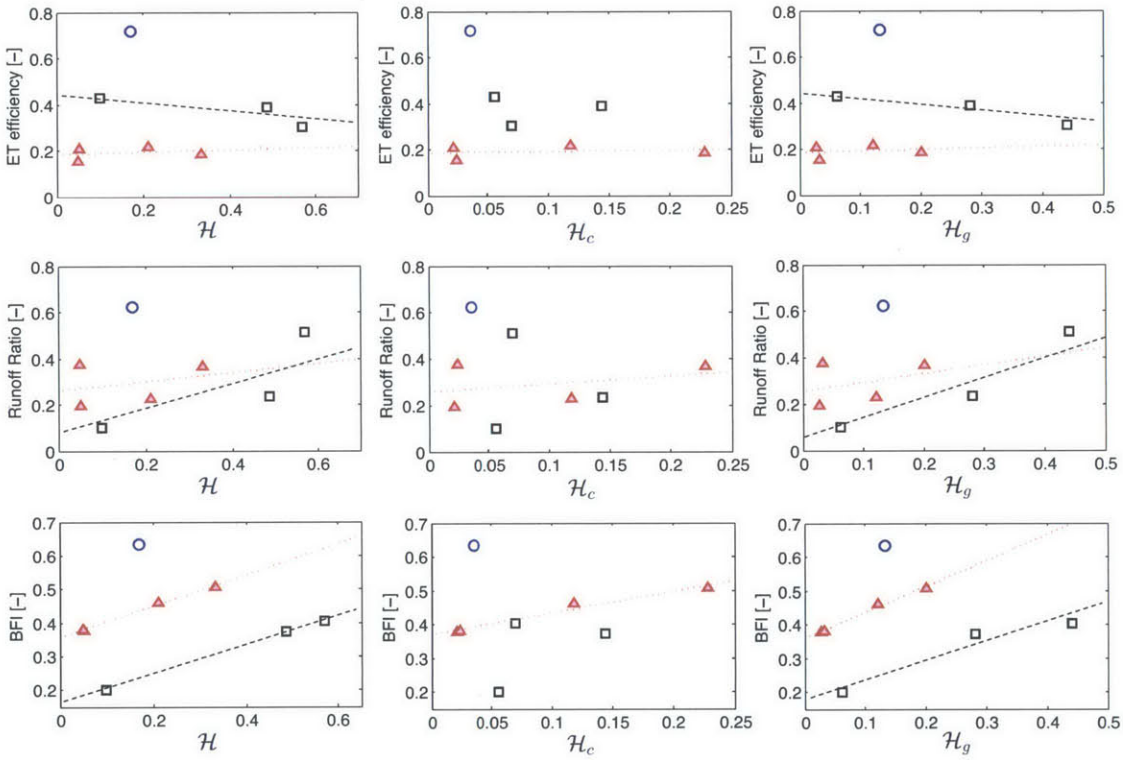


Figure 6.11: Temporal mean of (left to right) $\mathcal{H}$, $\mathcal{H}_c$, and $\mathcal{H}_g$, vs. (top to bottom) ET efficiency, runoff ratio, and BFI. Red and black trendlines are for semiarid and temperate basins, respectively.

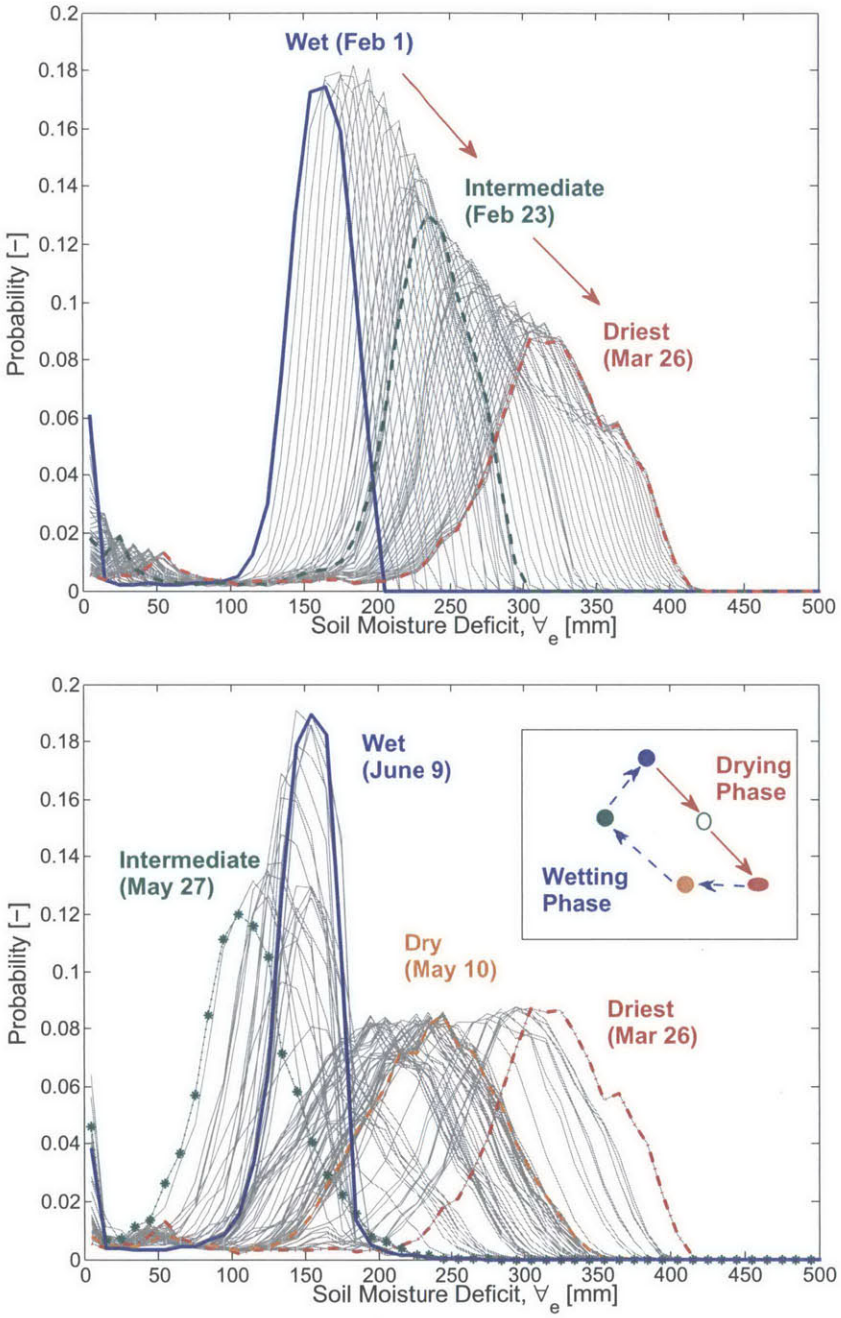


Figure 6.12: Hysteresis in the evolution of the probability distribution of soil moisture deficit $\forall_e$ during a drying phase (top) and a wetting phase (bottom) at basin H2

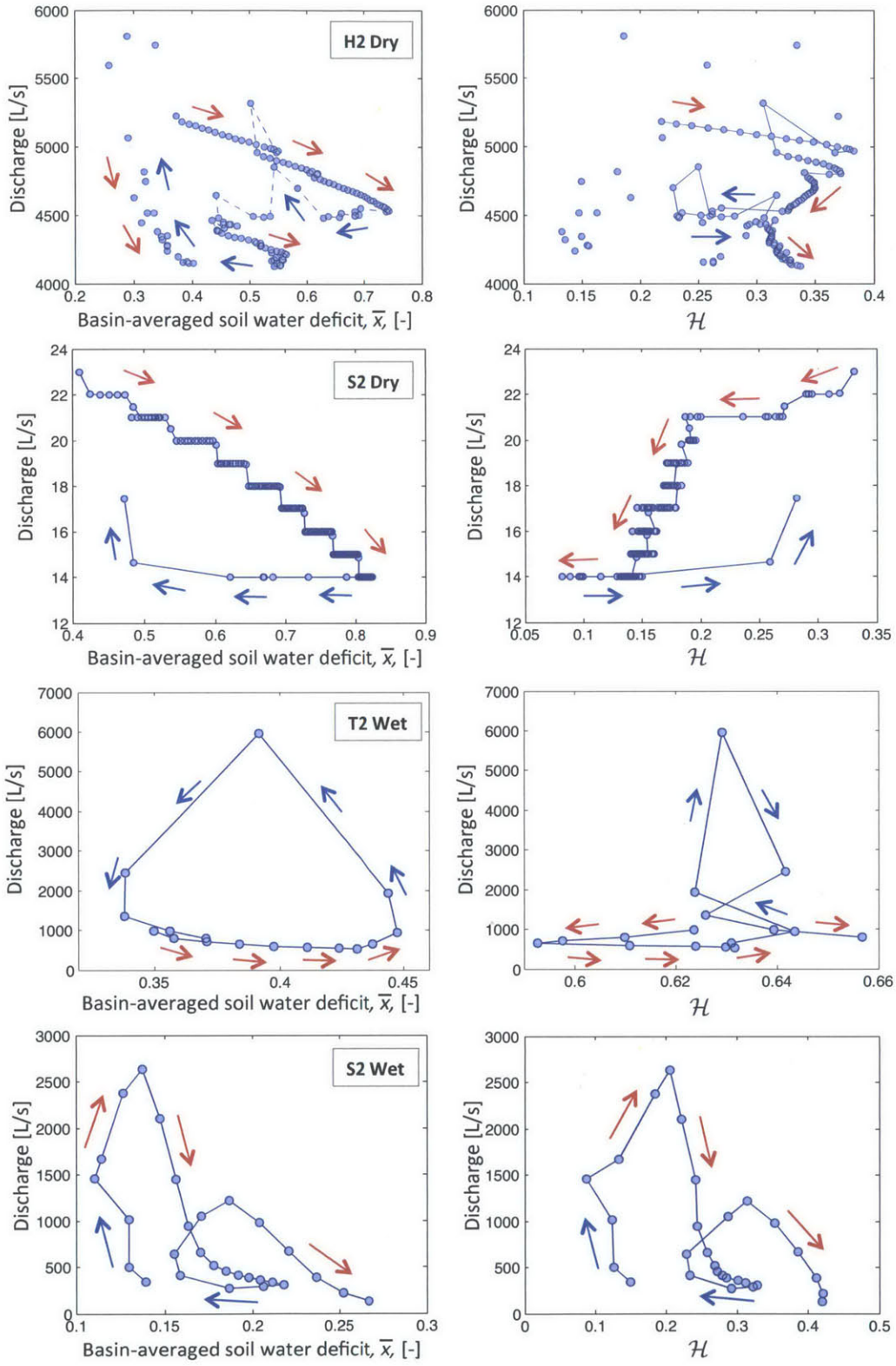


Figure 6.13: Hysteresis in simulated daily discharge vs. mean relative soil water deficit $\overline{x}$ (left), and vs. $\mathcal{H}$ (right) for basins H2 and S2 in dry conditions, and T2 and S2 in wet conditions. Red and blue arrows indicate drying and wetting, respectively.

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 7

Summary, Major Contributions, and Recommendations

7.1 Summary

The ability to explain or predict the hydrologic response of a basin has many applications. Unfortunately, although hydrologists have long sought for general organizing principles to relate the hydrologic response of a basin to its features, the ability to regionalize of extrapolate observations from one place to another or across different scales, has not been realized. This is because the hydrologic response is the combined manifestation of many complex and interrelated factors that naturally vary both in space and time, and act over a variety of scales. Perhaps, a possible key is better understanding and model representation of soil moisture which strongly controls basin response and hydrologic fluxes. In this light, this thesis aims to improve our understanding of the complexity and spatiotemporal variability of basin response and hydrologic fluxes by i) using a distributed hydrologic model with a novel approach to modeling soil moisture; and ii) using an entropy-based measure of hydrologic complexity as an analysis tool and an organizing principle.

The main tool used in this research is the distributed catchment hydrologic model MOBIDIC which simultaneously solves mass and energy balance. A key feature of the model is its use of a single layer of soil with dual compartments: a gravity reservoir composed of large pores that drain under gravity, and a capillary reservoir composed of small pores that do not drain under gravity. This representation accounts for the different roles of gravity and capillary in storing and moving soil water, and at the same time makes the model computationally efficient. Model modifications were introduced to improve the representation of some hydrologic processes especially for better simulation of soil moisture. These include i) the addition of plant and surface reservoirs; ii) the introduction of a single-parameter S-type model of ET as function of soil saturation; iii) the modifications of the conceptualization and formulation of lateral subsurface flow; iv) the addition of control rules for percolation; v) the addition of a capillary rise module; and vi) the use of a modified form of the linear reservoir model for groundwater dynamics. Modifications were also done on the MATLAB codes of MOBIDIC reducing run time by about 80 percent.

The ability of the improved MOBIDIC to correctly simulate the magnitude range and temporal dynamics of soil moisture at the *local scale* e.g. at a plan element was tested by comparing its simulated soil moisture against both observations and values simulated by a benchmark model. The 1–D SHAW model was chosen as the benchmark because it solves both mass and energy balance and it uses the nonlinear Richards Equation to simulate flow in unsaturated soil. Using two sites with contrasting climates, it was shown that although MOBIDIC is intended for basin-scale catchment modeling and despite its use of a single soil layer with dual compartments, and linear parameterization of infiltration and some other fluxes, that it can capture the more detailed dynamics implicitly included in the SHAW model. The comparison was performed at two sites with contrasting climates.

This thesis also introduces a revised formulation of the *Martina and Entekhabi*

(2006) dimensionless measure of hydrologic complexity $\mathcal{H}$ which measures the relative distance of a given distribution of soil moisture from two limiting distributions. The modifications make $\mathcal{H}$ discretization-invariant. The key features of $\mathcal{H}$ are: i) it is computed for random variables with values strictly in the interval 0 to 1, here the relative soil water deficit is used; ii) it is computed based on differential entropy instead of Shannon entropy; iii) it uses the Kullback-Leibler divergence to ensure non-negativity; and iv) it uses an exponential transformation so that its value ranges from 0 to 1, with $\mathcal{H} = 0$ for the simplest, and $\mathcal{H} = 1$ for the most complex case, respectively.

In Chapter 5, the improved MOBIDIC and the revised $\mathcal{H}$ are used to conduct a detailed investigation of the hydrology of a tropical rain forest (TRF) *river basin* in the Philippines. This demonstrates the details of how a MOBIDIC model is developed for a specific basin, and it also illustrates the hydrologic information and insights that the model can provide. Moreover, aside from being the first application of MOBIDIC on a TRF basin, this study demonstrates how a sophisticated hydrologic model can be developed using freely-available remotely-sensed data plus only minimal field observations and measurements. This is particularly useful in data-scarce environments. For this basin, $\mathcal{H}$ behaves differently in the wet and dry regimes. The transition occurs when the dominant processes in the basin switch between vertical and lateral fluxes. The spatial distribution of soil moisture also exhibited hysteresis and threshold-like behaviors.

The final chapter applies $\mathcal{H}$ on MOBIDIC-simulated hydrologic variables for eight test basins with area of $10^0 - 10^3$ km^2 and representing semiarid, temperate and humid climates. The results show that for basins with roughly the same catchment area, higher median slope leads to higher runoff ratio and lower ET efficiency. This finding, though rather simple, makes perfect hydrological sense (which is what we want) but is not clear if all test basins are analyzed together. This highlights the need for a

systematic and multi-dimensional analysis framework or basin classification scheme that will help collapse and unravel the complexity of hydrologic response.

$\mathcal{H}$ was evaluated using the simulated relative soil water deficit for each of the test basins. Since a key feature of MOBIDIC is the partitioning of each pixel of soil into a **g**ravity reservoir and a **c**apillary reservoir, the complexity of these sub-systems, $\mathcal{H}_c$ and $\mathcal{H}_g$, respectively, were also investigated. It was shown that $\mathcal{H}$ can effectively track the evolution of the distributional features of soil moisture, and together with $\mathcal{H}_c$ and $\mathcal{H}_g$ can capture the switching of between vertical and lateral fluxes as the dominant hydrologic processes. This study also explored what and how basin attributes affect the measures of hydrologic complexity and how these measures can be used to explain inter-basin variability in hydrologic response. For basins of similar size, $\mathcal{H}$ and $\mathcal{H}_g$ are highest in temperate climate, consistent with soil moisture being double-bounded so its variability is maximal at intermediate values. $\mathcal{H}_c$ generally increases with aridity since the capillary reservoirs are often saturated in the temperate and humid basins i.e. they are less hydrologically active. For the semiarid basins, $\mathcal{H}$, $\mathcal{H}_c$ and $\mathcal{H}_g$, are positively correlated with catchment area, basin-scale topographic wetness index, and infiltration ratio, but negatively correlated with relief ratio; while $\mathcal{H}$ and $\mathcal{H}_g$ are positively correlated with baseflow index. The trends (slope direction) are generally the same for the temperate basins.

Hysteresis and threshold-like behavior are manifestations of non-linear hydrologic response. The detection of these non-linear behaviors is a potential benchmark for hydrologic models to capture and perhaps a way to classify behaviors of different hillslopes and basins. This study shows that although hysteresis is not explicitly coded in MOBIDIC, it is exhibited in the plots of discharge vs. mean soil water deficit, and discharge vs. $\mathcal{H}$, both in low- and high-flow conditions. In addition, the mechanisms that cause these behaviors to emerge in the simulated variables are hydrologically meaningful. Hysteresis emerges as a result of MOBIDIC's use of a

dual-pore soil structure which not only captures the different roles of capillary and gravity-driven processes, but also mimics the threshold behavior of runoff generation. The use of entropy-based measure of complexity helps understand the link between internal patterns and states to lumped or macro processes and behaviors.

7.2 Major Contributions

The major contributions of this thesis are the following:

1. A more computationally efficient version of MOBIDIC with more realistic and flexible representation of several hydrologic processes, was developed.

2. This thesis demonstrates that a simple and computationally parsimonious conceptual approach using a dual-pore characterization of a single soil unit

   - can perform as good as a numerical solver of Richards equation, in simulating the magnitude range and temporal dynamics of depth-averaged soil moisture at the local scale; and
   - exhibits hysteresis and threshold behaviors.

3. A discretization-invariant dimensionless measure of hydrologic complexity $\mathcal{H}$ was developed. It was shown how $\mathcal{H}$ can be used as a tool

   - to understand the evolution of the spatial distribution of soil moisture;
   - to investigate the link between spatial patterns or internal states, to lumped basin processes or behaviors;
   - to understand the interplay between various hydrologic processes; and
   - to explain interbasin variability in hydrologic response.

4. This thesis demonstrates how a sophisticated hydrologic model can be developed using freely-available remotely-sensed data, plus only minimal field observations and measurements.

5. This thesis presents the first application of MOBIDIC on a tropical rain forest basin, and adds to the hydrological literature on these important and fragile ecosystems.

6. This thesis emphasizes the multi-dimensionality of hydrologic response. It is shown that clearer and meaningful relationships between measures of hydrologic response and observable basin attributes can only be obtained if basins are classified based on e.g. spatial scale or climatic regime.

7.3 Recommendations for Future Research

The following are some thoughts and recommendations for future research:

1. Chapter 6 is of exploratory nature and as mentioned, the relationship found between the measures of hydrologic complexity ($\mathcal{H}$, $\mathcal{H}_c$, and $\mathcal{H}_g$) with basin attributes and traditional measures of hydrologic response, should be verified using more test basins of various settings, spatial scales, and climates.

    - Specifically, it would be nice to have other humid test basins with different spatial scales.
    - For climate sensitivity, having (additional) test basins across different locations is recommended.
    - For spatial scale sensitivity, well-studied nested river basins can be used. Other basins larger than 10^3 sq. km. should also be added.

2. Apply the measures of hydrologic complexity on actual observations of soil moisture fields such as from extensive networks of in-situ soil moisture sensors or from remotely-sensed soil moisture products such as from SoilSCAPE (http://soilscape.usc.edu), AirMOSS (http://airmoss.jpl.nasa.gov), and SMAP (http://smap.jpl.nasa.gov).

3. Check the sensitivity of $\mathcal{H}$ to space-time resolution

   - This thesis used daily-aggregated soil moisture fields, computed from a model using hourly time-step. The effect of using different temporal resolutions can be investigated.
   - The simulated soil moisture fields can also be aggregated to coarser spatial resolution.

4. The relationships between the space-time statistics (mean, median, variance, kurtosis, etc.) of the measure and the statistics of distributed basin attributes and forcings can also be investigated.

5. Synthetic experiments using spatially uniform precipitation, vegetation, and soil types can be performed.

6. The utility of $\mathcal{H}$ as a measure of complexity or uncertainty of any random variable in the interval $[0, 1]$, not just hydrologic variables, should be explored.

7. The strategy of grouping basins based on similarity of attributes can be applied to previous or future studies involving multiple basins.

THIS PAGE INTENTIONALLY LEFT BLANK

List of Acronyms

BFI	base flow index
DEM	digital elevation model
DHM	distributed hydrologic model
DMIP	Distributed Model Intercomparison Project
ET	evapotranspiration
GW	groundwater
HR	hydrologic response
MOBIDIC	Modello Bilancio Idrologico DIstributo e Continuo
PET	potential evapotranspiration
RH	relative humidity
SHAW	Simultaneous Heat And Water
SoilSCAPE	Soil moisture Sensing Controller and oPtimal Estimator
TRMM	Tropical Rainfall Measuring Mission
TWI	topographic wetness index
WY	water year

THIS PAGE INTENTIONALLY LEFT BLANK

List of Symbols

Note: Fluxes are based on 1 time step

Symbol	Definition	Units
α	surface runoff parameter	[-]
β	lateral subsurface runoff parameter	[-]
η	product of Brooks-Corey parameters m and c	[-]
γ	soil percolation parameter	[-]
κ	soil absorption parameter	[-]
ξ	parameter for ET; $\xi \in 1, 2, 3, 4, 5$	[-]
ϕ_{ch}	fraction of flow that is channelized	[-]
ψ	matric potential	[L]
ψ_1	bubbling pressure	[L]
ρ_a	density of air	[M/L^3]
ρ_s	density of soil	[M/L^3]
ρ_w	density of water	[M/L^3]
θ	soil moisture (state variable)	[-]
$\overline{\theta}^{mod}$	modeled basin-averaged θ	[-]
$\overline{\theta}^{obs}$	observed basin-averaged θ	[-]
θ_{fld}	soil moisture at field capacity	[-]
θ_{sat}	soil moisture at saturation	[-]
θ_{sat}	soil moisture at saturation	[-]
$\forall_e$	total soil water deficit	[L]
$\forall_{e,c}$	water deficit of the soil capillary reservoir	[L]
$\forall_{e,g}$	water deficit of the soil gravity reservoir	[L]
$\forall_{e,max}$	total water storage capacity of the soil	[L]

Symbol	Definition	Units
A	catchment area	[km^2]
B	bias	[-]
c	Brooks-Corey pore-size disconnectedness index	[-]
C_a	heat capacity of air	[$ML^2T^{-2}\Theta^{-1}$]
C_s	heat capacity of soil	[$ML^2T^{-2}\Theta^{-1}$]
C_H	turbulent heat exchange coefficient	[-]
d	thickness of the modeled soil	[L]
D	total soil depth of SHAW	[L]
D_d	drainage density	[-]
d_i	thickness of ith soil layer of SHAW	[L]
dt	computational time step	[T]
d_w	representative distance from the unsaturated soil layer to the water table	[L]
E	Shannon entropy	[-]
E	latent heat	[$L^2T^{-2}\Theta^{-1}$]
ET	evapotranspiration	[L]
ET_1	ET from the plant canopy	[L]
ET_2	ET from the surface water	[L]
ET_3	ET from the soil	[L]
$f(x)$	probability density function of x	[-]
G	heat flux into the soil	[$L^2T^{-2}\Theta^{-1}$]
H	sensible heat	[$L^2T^{-2}\Theta^{-1}$]
$\mathcal{H}$	dimensionless measure of hydrologic complexity	[-]
i_s	mean storm intensity	[L/T]
I	infiltration	[L]
IA	index of agreement	[-]

Symbol	Definition	Units
J	overall objective function for calibration	[-]
k	soil thermal conductivity	$[\mathrm{MLT^{-3}\Theta^{-1}}]$
K_s	soil saturated hydraulic conductivity	[L/T]
L_v	latent heat of vaporization	[J/M]
m	Brooks-Corey pore-size distribution index	[-]
NSE	Nash-Sutcliffe Efficiency index	[-]
P	precipitation	[L]
PB	percent bias	[%]
PET	potential evapotranspiration	[L]
$p(x)$	probability mass function of x	[-]
q_a	specific humidity of air	[-]
q_s	specific humidity of the soil-vegetation continuum	[-]
Q	basin discharge	[L]
Q_{as}	absorption from W_g to W_c	[L]
Q_{bf}	baseflow	[L]
Q_{cap}	capillary rise	[L]
Q_L	lateral subsurface flow	[L]
$Q_{L,bypass}$	Q_L that is directly routed downstream	[L]
Q_{out}	total streamflow at the basin outlet, same as Q	[L]
Q_{per}	percolation	[L]
R	Pearson correlation coefficient	[-]
R_D	Dunne runoff	[L]
R_H	Hortonian runoff	[L]
R_n	net incoming radiation	$[\mathrm{L^2T^{-2}\Theta^{-1}}]$
R_r	relief ratio	[-]
R_R	return flow	[L]
R_T	total runoff	[L]

Symbol	Definition	Units
$R_{T,bypass}$	R_T that is directly routed downstream	[L]
s	local slope	[-]
s_{50}	median slope	[-]
s_{max}	maximum slope in the basin	[-]
S	effective soil saturation	[-]
$\overline{S}$	basin-averaged soil saturation	[-]
$\dot{S}$	net change in water storage of a basin	[L]
t	time	[T]
T	soil temperature	[Θ]
T_a	air temperature	[Θ]
$T_{constant}$	lower boundary condition for T roughly equal to the annual mean T_a	[Θ]
T_d^M	temperature at diurnal damping depth	[Θ]
T_s	temperature of soil and vegetation continuum	[Θ]
U	wind speed	[L/T]
x	relative soil water deficit	[-]
x_c	relative water deficit of the capillary reservoir	[-]
x_g	relative water deficit of the gravity reservoir	[-]
$\mathbf{w}$	4 by 1 matrix of weights for PB, R, IA, and NSE	[-]
W_c	current content of the soil capillary reservoir	[L]
$W_{c,max}$	capacity of the soil capillary reservoir	[L]
W_g	current content of the soil gravity reservoir	[L]
$W_{g,max}$	capacity of the soil gravity reservoir	[L]
W_p	current content of the plant/canopy reservoir	[L]
$W_{p,max}$	capacity of the plant/canopy reservoir	[L]
W_s	current content of the surface reservoir	[L]

Symbol	Definition	Units
$W_{s,max}$	capacity of the surface reservoir	[L]
z	depth below surface (positive downward)	[L]
z_d	damping depths of daily heatwaves	[L]
z_w	depth to groundwater table	[L]
z_y	damping depths of yearly heatwaves	[L]

Bibliography

Institute of Hydrology (1980), Low flow studies, *Tech. rep.*, Walingford, UK.

The Manila Observatory (2013), Daily meteorological data for Davao, http://www.observatory.ph/Metdata/MOD, accessed: 04/29/2013.

Ashby, S. F., and R. D. Falgout (1996), A parallel multigrid preconditioned conjugate gradient algorithm for groundwater flow simulations, *Nuclear Science and Engineering, 124*, 145–159.

Bartholomes, J., and E. Todini (2005), Coupling meteorological and hydrological model for flood forecasting, *Hydrol. and Earth Sys. Sci., 9(4)*, 333–346.

Berger, K. P., and D. Entekhabi (2001), Basin hydrologic response relations to distributed physiographic descriptors and climate, *Journal of Hydrology, 247*, 169–182.

Beven, K. J. (1995), Linking parameters across scales: subgrid parameterizations and scale dependent hydrological models, *Hydrological Processes, 9(5/6)*, 507–525.

Beven, K. J. (2006a), A manifesto for the equifinality thesis, *Journal of Hydrology, 320*, 18–36, doi:10.1016/j.jhydrol.2005.07.007.

Beven, K. J. (2006b), Searching for the holy grail of scientific hydrology, *Hydrol. Earth Syst. Sci. Discuss., 3*, 769–792.

Beven, K. J., and P. Germann (2013), Macropores and water flow in soils revisited, *Water Resour. Res., 49*, 3071 3092, doi:10.1002/wrcr.20156.

Beven, K. J., and M. J. Kirby (1979), A physically based, variable contributing area model of basin hydrology, *Hydrol. Sci. Bull., 24*(1), 43–69.

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

Bisht, G. (2010), Satellite-based estimates of net radiation and modeling the role of topography and vegetation on inter-annual hydroclimatology, PhD dissertation, MIT, Department of Civil and Environmental Engineering.

Bogaart, P. A. Teuling, and P. Troch (2008), A state-dependent parameterization of saturated-unsaturated zone interaction, *Water Resour. Res.*, W11423, doi:10.1029/2007WR006487.

Bras, R. (1990), *Hydrology: An Introduction to Hydrologic Science*, Addison-Wesley, Massachusetts.

Brooks, J. R., H. R. Barnard, R. Coulombe, and J. J. McDonnell (2009), Ecohydrologic separation of water between trees and streams in a mediterranean climate, *Nature Geoscience*, *3*, 100–104, doi:10.1038/NGE0722.

Brooks, R. H., and A. T. Corey (1964), Hydraulic properties of porous media, *Tech. rep.*, Agricultural Research Service, Soil and Water Conservation Research Division, and the Agricultural Engineering Dept., Colorado State Univ., Fort Collins, Colorado.

Bruijnzeel, L. A. (2004), Hydrological impacts of tropical forest conversion, *Nature & Resources*, *27*(2), 36–46.

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

Burnash, R. J., R. L. Ferral, and R. A. McGuire (1973), A generalized streamflow simulation system: conceptual models for digital computers, *Tech. rep.*, Joint Federal-State River Forecast Center, Sacramento, CA.

Buttle, J. M., and D. J. McDonald (2002), Coupled vertical and lateral preferential flow on a forested slope, *Water Resour. Res.*, *38*, doi:10.1029/2001WR000773.

Buttle, J. M., P. J. Dillon, and G. R. Eerkes (2004), Hydrologic coupling of slopes, riparian zones and streams: an example form the Canadian Shield, *Journal of Hydrology*, *287*, 161–177, doi:10.1016/j.jhydrol.2003.09.022.

Camporose, M., D. Penna, M. Borga, and C. Paniconi (2014), A field and modeling study of nonlinear storage-discharge dynamics for an Alpine headwater catchment, *Water Resour. Res.*, *50*, 2693–2712, doi:10.1002/2013WR013604.

Castelli, F., G. Menduni, and B. Mazzanti (2006), Use of multi-platform, multi-temporal remote-sensing data for calibration of a distributed hydrologic model: an application in the Arno basin, Italy, *Hydrological Processes*, *20*, 2693–2712, doi:10.1002/hyp.6061.

Castelli, F., G. Menduni, and B. Mazzanti (2009), A distributed package for sustainable water management: a case study in the Arno basin, *IAHS Publ.*, *327*.

Cerdan, O., Y. L. Bissonnais, G. Govers, V. Lecomte, K. van Oost, A. Couturier, C. King, and N. Dubreuli (2004), Scale effect on runoff from experimental plots to catchments in agricultural areas in Normandy, *Journal of Hydrology*, *1-2*, 4–14, doi:10.1016/j.jhydrol.2004.02.017.

de Gennes, P. G., F. Brochard Wyard, and D. Quere (2003), *Capillarity and Wetting Phenomena: Drops, Bubbles, Pearls, Waves*, Springer.

Domenico, P. A., and F. W. Schwartz (1998), *Physical and Chemical Hydrogeology*, John Wiley & Sons, Inc., New Jersey.

Dooge, J. C. I. (1986), Looking for hydrologic laws, *Water Resour. Res.*, *22*(9), 46–58.

Douglas, E. M., K. Sebastian, C. J. Vörösmarty, S. Wood, and K. M. Chomitz (2005), *The Role of Tropical Forests in Supporting Biodiversity and Hydrological Integrity*, World Bank.

Duan, Q., J. Schaake, V. Andreassian, S. Franks, G. Goteti, and H. V. G. *et al.* (2006), Model Parameter Estimation Experiment (MOPEX): An overview of science strategy and major results from the second and third workshops, *Journal of Hydrology*, *320*, 3–17, doi:10.1016/j.jhydrol.2005.05.031.

Dunne, T., and R. D. Black (1970), Partial area contributions to storm runoff in a small New England watershed, *Water Resour. Res.*, *6*, 1296–1311.

Dykes, A. P. (1997), Rainfall interception from a lowland tropical rainforest in Brunei, *Journal of Hydrology*, *200*, 260–279, doi:10.1016/S0022-1694(97)00023-1.

Eagleson, P. S. (1978), Climate, soil, and vegetation 1: Introduction to water balance dynamics, *Water Resour. Res.*, *14*(5).

Ebel, B., and K. Loague (2006), Physics-based hydrologic-response simulation: seeing through the fog of equifinality, *Hydrol. Processes*, *20*, 2887–2900.

Entekhabi, D., I. Rodriguez-Iturbe, and F. Castelli (1995), Mutual interaction of soil moisture state and atmospheric processes, *Adv. Water Res.*, *184*, 3–17.

Entin, J. K., A. Robock, K. Y. Vinnikov, S. E. Hollinger, S. Liu, and A. Namkai (2000), Temporal and spatial scales of observed soil moisture variations in the extratropics, *J. Geophys. Res.*, *105*, 865–877, 2000JD900051.

Famiglietti, J., D. Ryu, A. A. Berg, M. Rodell, and T. J. Jackson (2008), Field observations of soil moisture variability across scales, *Water Resour. Res.*, *44*, W01423, doi:10.1029/2006WR005804.

Famiglietti, J. S., J. A. Devereaux, and C. A. L. *et al.* (1999), Ground-based investigation of soil moisture variability within remote sensing footprints during the Southern Great Plains 1997 (SGP97) Hydrology Experiment, *Water Resour. Res.*, *35*, 1839–1851, doi:10.1029/1999WR900047.

Fernandez, J. M., and A. Ceballos (2003), Temporal stability of soil moisture in a large-field experiment in Spain, *Soil Sci. Soc. Am. Journal*, *67*, 1647–1656.

Flerchinger, G. N. (2000a), The simultaneous Heat And Water (SHAW) model: Technical documentation, *Technical Report NWRC 9*, USDA Agricultural Research Service, Boise, Idaho.

Flerchinger, G. N. (2000b), The simultaneous Heat And Water (SHAW) model: User manual, *Technical Report NWRC 10*, USDA Agricultural Research Service, Boise, Idaho.

Flerchinger, G. N., and P. B. Pierson (1997), Modeling plant capnopy effects on variability of soil temperature and water: model calibration and validation, *J. of Arid Environments*, *35*, 641–653.

Flerchinger, G. N., and K. E. Saxton (1989), Simultaneous Heat And Water model of a freezing snow-residue-soil system i. theory and development, *Transactions of the ASAE*, *32*(2), 565–571.

Foglia, L., M. C. Hill, S. W. Mehl, and P. Burlando (2009), Sensitivity analysis, calibration, and testing of a distributed hydrologic model using error-based weighting and one objective function, *Water Resour. Res.*, *45*, doi:10.1029/2008WR007255.

Gardner, W. R. (1958), Some steady-state solutions of the unsaturated moisture flow equation with application to evaporation from a water table, *Hydrol. and Earth Sys. Sci.*, *85*(4), 228–232.

Gelhar, L. W., C. Welty, and K. R. Wehfeldt (1992), A critical review of data on field-scale dispersion in aquifers, *Water Resour. Res.*, *28*(7), 1955–1974.

Gerke, H. H., and M. T. van Genuchten (1993), A dual-porosity model for simulating the preferential movement of water and solutes in structured porous media, *Water Resour. Res.*, *29*(2), 305–319.

Goodrich, D. C., T. O. Keefer, C. L. Unkrich, M. H. N. nad H. B. Osborn, J. J. Stone, and J. R. Smith (2008), Long-term precipitation database, Walnut Gulch Experimental Watershed, Arizona, United States, *Water Resour. Res.*, *44*, W05S04, doi:10.1029/2006WR005782.

Grayson, R. B., and A. W. Wetern (1998), Towards areal estimation of soil water content from point measurements: Time and space stability of mean response, *Journal of Hydrology*, *207*(1-2), 68–82.

Grayson, R. B., A. W. Wetern, F. H. Chiew, and G. Blöschl (1997), Preferred states in spatial soil moisture patterns: local and nonlocal controls, *Water Resour. Res.*, *33*(12), 2897–2908.

Gupta, H. V., S. Sorooshian, and P. O. Yapo (1998), Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information, *Water Resour. Res.*, *34*(4), 751763.

Guswa, A. J., M. A. Celia, and I. Rodriguez-Iturbe (2002), Models of soil moisture dynamics in ecohydrology: a comparative study, *Water Resour. Res.*, *38*(9), doi:10.1029/2001WR000826.

Haan, C. T., H. P. Johnson, and D. L. Brakensiek (1982), Hydrologic modeling of small watersheds, *Am. Soc. Agr. Eng., Monograph no. 5.*

Haines, W. B. (1930), Studies in the physical properties of soil: V. The hysteresis effect in capillary properties, and the modes of moisture associated therewith, *Am. Soc. Agr. Eng., Monograph no. 5.*

Harter, T., and J. W. Hopmans (2004), *Unsaturated-zone Modeling: Progress, Challenges and Applications*, chap. Role of vadose-zone flow processes in regional-scale hydrology: review, opportunities and challenges, pp. 179–208, Kluwer Academic Publishers, the Netherlands.

Ivanov, V. Y., E. R. Vivoni, R. L. Bras, and D. Entekhabi (2004), Catchment hydrologic response with a fully distributed irregular network model, *Water Resour. Res., 40*, W11102, doi:10.1029/2004WR003218.

Ivanov, V. Y., S. Fatichi, G. D. Jenerette, J. F. Espeleta, P. A. Troch, and T. E. Huxman (2010), Hysteresis of soil moisture spatial heterogeneity and the homogenizing effect of vegetation, *Water Resour. Res., 46*, W09521, doi:10.1029/2009WR00861.

Jaynes, D. B. (1990), *Process studies in hillslope hydrology*, chap. Soil water hysteresis: models and implications, pp. 93–126, Wiley, iSBN 0-471-92714-7.

Joekar-Niasar, V., F. Doster, R. T. Armstrong, D. Wildenschild, and M. A. Celia (2013), Trapping and hysteresis in two-phase flow in porous media: A pore-network study, *Water Resour. Res., 49*, 4244–4256, doi:10.1002/wrcr.20313.

K.Gupta, V., and O. J. Mesa (1988), Runoff generation and hydrologic response via channel network geomorphology - recent progress and open problems, *Journal of Hydrology, 102*(1-4), 3–28.

Kim, G., J. Chung, and J. Kim (2002), Spatial characterization of soil moisture estimates from the Southern Great Plain (SGP 97) hydrology experiment, *KSCE J. Civil Engg., 6*(2), 177–184.

Kollet, S. J., and R. M. Maxwell (2008), Capturing the influence of groundwater dynamics on land surface processes using an integrated distributed watershed model, *Water Resour. Res., 44*, doi:10.1029/2007WR006004.

Kollet, S. J., R. M. Maxwell, and C. S. W. *et al.* (2010), Proof of concept of regional scale hydrologic simulations at hydrologic resolution utilizing massively parallel computer resources, *Water Resour. Res., 46*, W04201, doi:10.1029/2009WR008730.

Koster, R. D., and M. J. Suarez (1999), A simple framework for examining the interannual variability of land surface moisture fluxes, *J. Climate, 12*, 1911–1917.

Koutsoyiannis, D. (2014), Uncertainty, entropy, scaling adn hydrological processes. 1. marginal distributional properties of hydrologic processes and state scaling, *Hydrological Sciences Journal, 50(3)*, 381–404.

Krause, P., D. P. Boyle, and F. Base (2005), Comparison of different efficiency criteria for hydrological model assessment, *Adv. in Geosciences*, *5*, 89–97.

Lamb, D. (2011), *Regreening the bare hills: tropical forest restoration in the Asia-Pacific region*, Springer.

Larsbo, M., and N. Jarvis (2003), *MACRO 5.0. A model of water flow and solute transport in macroporous soil.*, Swedish University of Agricultural Sciences, ISBN91-576-6592-3.

Lawrence, J. E., and G. M. Hornberger (2007), Soil moisture variability across climate zones, *Geophys. Res. Lett.*, *34*, L20402, doi:10.1029/2007GL031382.

Lazzaro, M. D. (2008), Correlation between channel and hillslope lengths and its effects on the hydrologic response, *Journal of Hydrology*, *362*, 260–273, doi:10.1016/j.jhydrol.2008.08.022.

Levine, J. B., and G. D. Salvucci (1999), Equilibrium analysis of groundwater-vadose zone interactions and the resulting spatial distribution of hydrologic fluxes across a canadian prairie, *Water Resour. Res.*, *35*(5), 1369–1383.

Liu, M., A. Bardossy, J. Li, and Y. Jiang (2012), Physically-based modeling of topographic effects on spatial evapotranspiration and soil moisture patterns through radiation and wind, *Hydrol. and Earth Sys. Sci.*, *16*, 357–373, doi:10.5194/hess-16-357-2012.

Luzio, M. D., and J. G. Arnold (2004), Formulation of a hybrid calibration approach for a physically based distributed model with NEXRAD data input, *Journal of Hydrology*, *298*, 136–154, doi.10.1016/j.jhydrol.2004.03.034.

Martina, M. L., and D. Entekhabi (2006), Identification of runoff generation spatial distribution using conventional hydrologic gauge time series, *Water Resour. Res.*, *42*, W08431, doi:10.1029/2005WR004783.

McDonnell, J. J., M. Sivapalan, K. Vache, and S. D. *et al.* (2007a), Moving beyond heterogeneity and process complexity: A new vision for watershed hydrology, *Water Resour. Res.*, *43*(7), doi:10.1029/2006WR005467.

McDonnell, J. J., B. McGlynn, K. Kendall, J. Shanley, and C. Kendall (2007b), The role of near-stream riparian zones in the hydrology of steep upland catchments, *Water Resour. Res.*, *43*(7).

Merz, R., G. Blöschl, and J. Parajka (2006), Spatio-temporal variability of event runoff coefficients, *Journal of Hydrology*, *331*, 591–604, doi:10.1016/j.jhydrol.2006.06.008.

Meyles, E., A. Williams, L. Ternan, and J. Dowd (2003), Runoff generation in relation to soil moisture patterns in a small dartmoor catchment, Southwest England, *Hydrol. Processes*, *17*, 251–264, doi:10.1002/hyp.1122.

Michalowicz, J. V., J. M. Nichols, and F. Bucholtz (2014), *Handbook of Differential Entropy*, CRC Press, iSBN 13:978-1-4665-8317-7.

Milly, P. C. (1988), Advances in modeling of water in the unsaturated zone, *Transport in porous media, 3(5)*, 491–514.

Moghaddam, M., A. Silva, R. Akbar, D. Clewley, M. Burgin, A. Castillo, and D. Entekhabi (2013), in *SoilSCaPE in situ network for multi-scale validation of SMAP data products*, IEEE, Melbourne, Australia, international Geoscience and Remote Sensing Symposium (IGARSS).

Molicova, H., M. Grimaldi, M. Bonell, and P. Hubert (1997), Using TOPMODEL towards identifying and modelling the hydrological patterns within a headwater humid, tropical catchment, *Hydrological Processes, 11*.

Myrabø, S. (1997), Temporal and spatial scale of response area and groundwater variation in a till, *Hydrol. Processes, 11*, 1861–1880.

Nash, J. E., and J. V. Sutcliffe (1970), River flow forecasting through conceptual models Part I - a discussion of principles, *Journal of Hydrology, 10*, 282–290.

Nicotina, L., E. A. Celegon, A. Rinaldo, and M. Marani (2008), On the impact of rainfall patterns on the hydrologic response, *Water Resour. Res., 44*, W12401, doi:10.1029/2007WR006654.

Nobre, A. D., L. A. Cuartas, M. Hodnett, C. D. Renno, G. Rodrigues, A. Silveira, M. Waterloo, and S. Saleska (2011), Height Above the Nearest Drainage - a hydrologically relevant new terrain model, *Journal of Hydrology, 404*, 13–29, doi:10.1016/j.jhydrol.2011.03.051.

O'Kane, J. P. (2005), Hysteresis in hydrology, *Acta Geophysica Polonica, 53*, 373–383.

O'Kane, J. P., and D. Flynn (2007), Thresholds, switches and hysteresis in hydrology from the pedon to the catchment scale: A non-linear systems theory, *Hydrol. and Earth Sys. Sci., 11(1)*, 443–459.

Penman, H. L. (1948), Natural evaporation from open water, bare soil and grass, *Proc. Roy. Soc. London, A(194)*, S. 120–145.

Penna, D., M. Borga, D. Norbiato, and G. Dalla (2009), Hillslope scale soil moisture variability in a steep alpine terrain, *Journal of Hydrology, 364*, doi:10.1016/j.jhydrol.2008.11.009.

Penna, D., N. Mantese, L. Hopp, G. D. Fontana, and M. Borga (2013), Spatio-tempral variability of piezometric response on two alpine hillslopes, *Hydrological Processes*, doi:10.1002/hyp.10140.

Phillips, J. D. (2003), Sources of nonlinearity and complexity in geomorphic systems, *Prog. Phys. Georg., 27*(1), 1–23.

Pokhrel, P., K. K. Yilmaz, and H. Gupta (2012), Multiple-criteria calibration of a distributed watershed model using spatial regularization sand response signatures, *Journal of Hydrology, 418-419,*, 49–60, doi:10.1016/j.jhydrol.2008.12.004.

Qu, Y., and C. J. Duffy (2007), A semidiscrete finite volume formulation for multiprocess watershed simulation, *Water Resour. Res.*, *43*, W08,419, doi:10.1029/2006WR005,752.

Rawls, W. J., D. L. Brakensiek, and K. E. Saxton (1982), Estimation of soil water properties, *Transaction of ASCE*, pp. 1316–1321.

Reed, S., V. Koren, M. Smith, and Z. Z. *et al.* (2004), Overall DMIP results, *Journal of Hydrology, 298*, 27–60, doi:10.1016/j.jhydrol.2004.03.031.

Refshaard, J. C., and B. Storm (1995), *Computer models of watershed hydrology*, pp. 809–846.

Richards, L. A. (1931), Capillary conduction of liquids through porous mediums, *Physics, 1(5).*

Robinson, J. S., M. Sivapalan, and J. D. Snell (1995), On the relative roles of hillslopes processes, channel routing, and network geomorphology in the hydrologic response of natural catchments, *Water Resour. Res.*, *31(12)*, 3089–3101.

Rodriguez-Iturbe, I., and A. Rinaldo (1997), *Fractal River Basins: Chance and Self-Organization*, Cambridge University Press.

Roulier, M. L. S., F. Stenemo, R. Kasteel, and N. Jarvis (2005), An improved dual-permeability model of water flow and solute transport in the vadose zone, *Vadose Zone Journal, 4*, 398–406, doi:10.2136/vzj2004.0137.

Rubin, Y. (2003), *Applied stochastic hydrogeology*, Oxford University Press, Oxford, UK.

Salvucci, G. D. (1993), An approximate solution for steady vertical flux of moisture through an unsaturated homogenous soil, *Water Resour. Res.*, *29*(11), 3749–3753.

Salvucci, G. D., and D. Entekhabi (1994), Comparion of the eagleson statistical-dynamical water balance model with numerical simulations, *Water Resour. Res.*, *30*(10), 2751–2757.

Sankarasubramanian, A., and R. M. Vogel (2002), Comment on the paper: basin hydrologic response relations to distributed physiographic descriptors and climate by Karen Plaut Berger, Dara Entekhabi, 2001. *Journal of Hydrology* 247, 169-182, *Journal of Hydrology, 263*, 257–261.

Saxton, K. E., H. P. Johnson, and R. H. Shaw (1974), Modeling evapotranspiration and soil moisture, *Trans. Am. Soc. Agr. Eng.*, *7(4)*, 673–677.

Sefton, C. E., and S. M. Howarth (1998), Relationships between dynamic response characteristics and physical descriptors of catchments in England and Wales, *Journal of Hydrology, 211*, 1–6.

Segui, P. Q., E. Martin, F. Habets, and J. Noilhan (2009), Improvement, calibration and validation of a distributed hydrologic model over france, *Hydrol. and Earth Sys. Sci., 13*, 163–181.

Šimůnek, J., and M. T. van Genuchten (2008), Modeling non-equilibrium flow and transport processes using HYDRUS, *Vadose Zone Journal, 7*, 782–797, doi:10.2136/vzj2007.0074.

Singh, V. P. (2011), Hydrologic synthesis using entropy theory: Review, *Journal of Hydrologic Engineering, 16(5)*, 421–433, iSSN 1084-099/2011/5.

Sköien, J. O., G. Blöschl, and A. W. Western (2003), Characteristic space scales and timescales in hydrology, *Water Resour. Res., 39*(10), 1304, doi:10.1029/2002WR001736.

Smith, S., D. J. Seo, V. I. Koren, S. M. Reed, Z. Zhang, Q. Duan, F. Moreda, and S. Cong (2004), The Distributed Model Intercomparison Project (DMIP): motivation and experiment design, *Journal of Hydrology, 298*(1-4), 4–26, doi:10.1016/j.jhydrol.2004.03.031.

Smith, S., et al. (2012), The Distributed Model Intercomparison Project - Phase 2: Motivation and design of the Oklahoma experiments, *Journal of Hydrology, 418-419*, 3–16, doi:10.1016/j.jhydrol.2011.08.056.

Tarbonton, D. (1997), A new method for the determination of flow directions and upslope areas in grid DEM, *Water Resour. Res., 33*, 309–319.

Teuling, A., and P. Troch (2005), Improved understanding of soil moisture variability dynamics, *Geophys. Res. Lett., 32*, L05404, doi:10.1029/2004GL021935.

Tromp van Meerveld, I. J., and J. J. McDonnell (2006a), Threshold relations in subsurface stormflow: 1. A 147-storm analysis of the panola hillslope, *Water Resour. Res., 42*, doi:10.1029/WR003778.

Tromp van Meerveld, I. J., and J. J. McDonnell (2006b), Threshold relations in subsurface stormflow: 2. The fill and spill hypothesis, *Water Resour. Res., 42*, doi:10.1029/WR003800.

USDA-ARS (2007), Southwest Watershed Research Center & Walnut Gulch Experimental Watershed, *Tech. rep.*, US Department of Agriculture – Agricultural Research Service, Tucson, Arizona.

van Djik, A. I. J. M. (2010), Climate and terrain factors explaining streamflow response and recession in Australian catchments, *Hydrol. and Earth Sys. Sci., 14*, 159–169, iSSN:1027-5606.

Vereecken, H., T. Kamai, T. Harter, R. Kasteel, J. Hopmans, and J. Vanderborght (2007), Explaining soil moisture variability as a function of mean soil moisture measurements in vadose zone hydrology: a stochastic unsaturated flow perspective, *Geophys. Res. Lett.*, *34*, L22,402.

Viessmann, W., J. W. Knapp, and G. L. Lewis (1977), *Introduction to hydrology*, second ed., Harper and Row, New York, New York.

Vieux, B. E., Z. Cui, and A. Gaur (2004), Evaluation of a physics-based distributed hydrologic model for flood forecasting, *Journal of Hydrology*, *298*, 155–177, doi:10.1016/j.jhydrol.2004.03.035.

Vivoni, E. R., D. Entekhabi, R. L. Bras, and V. Y. Ivanov (2007), Controls on runoff generation and scale-dependence in a distributed hydrologic model, *Hydrol. and Earth Sys. Sci.*, *4*, 1–47.

Vivoni, E. R., J. C. Rodriguez, and C. J. Watts (2010), On the spatiotemporal variability of soil moisture and evapotranspiration in a mountainous basin within the North American monsoon region, *Water Resour. Res.*, *46*, W09521, doi:10.1029/2009WR008611.

Vrugt, J. A., H. V. Gupta, L. A. Bastidas, W. Bouten, and S. Sorooshian (2003), Effective and efficient algorithm for multiobjective optimization of hydrologic models, *Water Resour. Res.*, *39(8)*, doi:10.1029/2002WR001746.

Walker, J. P., G. R. Willgoose, and J. D. Kalma (2004), In situ measurement of soil moisture: a comparison of techniques, *Journal of Hydrology*, *293*, 85–99, doi:10.1016/j.jhydrol.2004.01.008.

Western, A. W., and R. B. Grayson (2000), *Spatial Patterns in Hydrological Processes: Observations and Modelling*, chap. Soil moisture and runoff processes at Tarrawarra, pp. 209–246, Cambridge University Press.

Western, A. W., R. B. Grayson, G. Blöschl, G. R. Willgoose, and T. A. McMahon (1999), Observed spatial organization of soil moisture and its relation to terrain indices, *35*(3), 797–810, doi:10.1029/1998WR900065.

Western, A. W., R. B. Grayson, G. Blöschl, and D. J. Wilson (2003), *Scaling methods in soil physics*, chap. Spatial variability of soil moisture and its implications for scaling, pp. 119–142, CRC Press, Boca Raton, Fla.

Willmott, C. J. (1981), On the validation of models, *Physical Geography*, *2*, 184194.

Wilson, D. J., A. W. Western, and R. Z. B. Grayson (2004), Identifying and quantifying sources of variability in temporal and spatial soil moisture observations, *Water Resour. Res.*, *40*, doi:10.1029/2003WR002306.

Yeh, P., and E. Eltahir (1998), Stochastic analysis of the relationship between topography and the spatial distribution of soil moisture, *Water Resour. Res.*, *34(5)*, 1251–1263, doi:10.1029/98WR00093.

Zecharias, Y. B., and W. Brutsaert (1988), The influence of basin morphology on groundwater outflow, *Water Resour. Res.*, *24*(10), 1645–1650, WR03083.

Zehe, E., R. Becker, A. Bardossy, and E. Plate (2005), Uncertainty of simulated catchment runoff response in the presence of threshold processes: role of initial soil moisture and precip, *Journal of Hydrology*, *315*, 183–202, doi:10.1016/j.jhydrol.2005.03.038.

Zhang, L., K. Hickel, W. R. Dawes, F. H. Chiew, A. W. Western, and P. R. Briggs (2004), A rational function approach for estimating mean annual evapotranspiration, *Water Resour. Res.*, *40*, doi:10.1029/2003WR002710.