

**THE INFLUENCE OF CLIMATE ON WATERSHED TRANSIT TIMES, WITH  
IMPLICATIONS FOR NITRATE TRANSPORT IN THE CHESAPEAKE BAY  
WATERSHED**

by

Daniel C. Wilusz

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the  
degree of Doctor of Philosophy.

Baltimore, MD

July 2018

©Daniel C. Wilusz 2018

All rights reserved

## Abstract

The ability to accurately model the timing and quantity of contaminant transport from landscapes to surface waters under different climate conditions is vital to the development of climate-resilient watershed management tools and strategies. Although hydrologic transport cannot be directly measured at the full range of relevant scales, a measurable proxy at catchment scale is the integrated transit time distribution (TTD). The TTD is the time-varying, probabilistic distribution of water travel times or, equivalently, water ages in catchment outflow. This dissertation presents advances in hydrologic theory and catchment-scale modeling and uses them to learn about the influence of climate on TTD behavior at multiple sites. The specific contributions of this work include (1) the first benchmarking of the sensitivity of catchment transit times to present and projected climate conditions, which shows that climate change could significantly shift the phenology of stream age; (2) the introduction of a computationally efficient approach to calibrating integrated surface-subsurface hydrology models (ISSHMs) under realistic climate forcing to both discharge and stream age, and subsequent virtual experiments suggesting that the age of baseflow is significantly influenced by upper soil properties due to dynamic hydrologic partitioning, which is not captured in steady-state simulations; (3) a novel analysis using flowpath decomposition in an ISSHM to understand the influence of climate on catchment TTD dynamics, which reveals a complex relationship between flowpath and water age that belies suggestions in the literature of a one-to-one mapping; and (4) proof-of-concept of an enhancement to the popular Soil and Water Assessment Tool (SWAT) that allows users to calculate groundwater TTDs, calibrate TTDs to available data, and more realistically simulate groundwater nitrate transport with variable recharge rates. The dissertation concludes with a brief discussion of its implications on our understanding of groundwater nitrate transport in the Chesapeake Bay watershed under present and future climates.

**Dissertation Readers** (listed alphabetically):

Dr. William P. Ball, Johns Hopkins University (co-advisor)

Dr. Grace Brush, Johns Hopkins University

Dr. Ciaran J. Harman, Johns Hopkins University (co-advisor)

Dr. Benjamin Zaitchik, Johns Hopkins University

## Acknowledgments

Partial support for this work was generously provided by the NSF IGERT Water, Climate, and Health Fellowship (DGE1069213), the NSF CBET Water, Sustainability, and Climate Grant (CBET1360415), a JHU DoGEE Fellowship, and a Geological Society of America student research grant.

This thesis would have been impossible without the help and support of many colleagues, friends, and family. First and foremost, I thank my co-advisors Bill Ball and Ciaran Harman. Bill introduced me to the scenery and scientists of the Bay, reminded me to always respect the data, and helped me appreciate why DoGEE has been so special for so many years. Ciaran rekindled my interest in applied mathematics, ushered me into a global family of landscape hydrologists, and taught me the art of hydrologic modeling. They were generous with their time and advice, and pushed me to do more than I thought possible.

I am also indebted to members of my examination committees and other faculty for their insights and encouragement. Ben Zaitchik taught me remote sensing and mentored me through my first publication. Grace Brush encouraged me to think about the long-term ecological effect of catchment transit times. Ed Bouwer and Alan Stone helped me understand microbial engineering and aquatic chemistry. Darryn Waugh and Tom Haine introduced me to applications of transit time modeling outside of hydrology. Peter Wilcock and Kellogg Schwaab offered encouragement when it was needed, and Seth Guikema's lessons on data analytics were especially useful in Chapter Four. I benefited greatly from collaborations with outside faculty and researchers. Reed Maxwell and his research group helped me to use ParFlow-CLM and frame my questions. Tony Buda provided data and useful insight into the hydrology of the WE-38 watershed.

I am thankful to my classmates in the Ball and Harman research groups, especially to Qian Zhang in the Ball Group for his helpfulness and influential guidance, and to Shane Putnam, Minseok Kim, Cassie Cosans, Tianqi Liu, Sihan Zhao, Shuyu Chang, and Adam Wlostowski in the Harman Group for many hours of thoughtful discussion, constructive feedback, and spirited camaraderie. In

addition, I am thankful for exposure in my early years to a range of environmental engineering topics by members of the Ball-Bouwer group including Rebecca Murphy, Jin Yang, Chris Kelley, Laura MacDonald, Jessica Lawson, and Pavlo Bohutskyi.

I am also thankful to have worked with many other bright and dedicated students and professionals who added much to my educational experience: the IGERT program staff and students including Shahin Zand, Michael Reese, Luke MacDonald, Julie Shortridge, Gina Tonn, Jose Molina, Asha Jordan, Anna Scott, Steven Chow, Chris Brueck, Tania Banerji, Lesley Joseph, Ben Davis, Sara Rasmussen, Alexi Russell, Andrea Fraser, Margaret Fleming, Hanna Gray, Natalie Exum, Nicole DeLuca, Ramya Ambikapathi, and Kar'etta Venable; the Water, Sustainability, and Climate (WSC) research PIs including Damian Brady, Jeremy Testa, Michael Kemp, Lisa Wainger, Laura Murray, Ariel Ortiz-Bobea, Haoying Wang, Zach Easton, and Daniel Fuka; the DoGEE support staff including Adena Rojas, Jessica Elroy, Huan Luong, Michael Lester, Keith Ritchie, and Denise Nowlin; the Sudd Wetland research team including Martha Anderson, Chris Hain, Iva Mladenova, and Tugrul Yilmaz; the Chesapeake Bay Program modeling team including Gopal Bhatt, Gary Shenk, Lew Linker, Guido Yactayo, and Amy Dale; the Engineers Without Borders team including Emily Prosser and Hayley Normile; members of the Student Activities Committee including Mike Rose, Stephanie Lau, Diane Moeller, Eric Marble, Ali Barton, Zhufeng Fang, Liang Chen, Robin Hytowitz, Evangelia Spyrou, and Mitra Ebrahimi; my officemates including Venkat Prava, Fengwei Hung, Rui Shi, and Karim Rafeay; and various other scholars and student researchers including Lauren Foster, Nicholas Engdahl, Annalise Blum, Laura Sims, Temesgen Alemneh, Wes Zeller, Saleh Satti, Paolo Benettin, Ingo Heidbuchel, David Dralle, Mohammad Danesh, Lina Stein, Andres Peralta, Zhufeng Fang, Amanda Darrow, and Christina Cho.

Finally, I am thankful to all the friends and family who kept my spirits high. I am especially thankful to the Reidel family, the Dutta family, Ben Lum, Fred Weston, Lisa Aresco Smith, the Sanii family, and Norm Yavarkovsky for getting me out and keeping connected. To my entire family including Gramma, Cathy and Dave Wilusz, Sioux and Tom Wilusz, Dad and Mom, I owe a lifetime of gratitude for your love, encouragement and home cooked meals, and a special thanks to mom for welcoming me home at all hours for all these years. Finally, I owe a second lifetime of thanks to my fiancée Maria Tupac, for her patience, encouragement, wisdom, and wit.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Motivation . . . . .	2
1.1.1	Non-point source pollution . . . . .	2
1.1.2	Influence of climate on NPS pollution transport . . . . .	2
1.1.3	Transit-time models . . . . .	3
1.1.4	TTD model advancements . . . . .	5
1.1.4.1	rSAS-LPMs . . . . .	6
1.1.4.2	Particle tracking models . . . . .	8
1.1.5	Research objectives and dissertation outline . . . . .	8
<b>2</b>	<b>Influence of rainfall on transit times</b>	<b>10</b>
2.1	Highlights . . . . .	10
2.2	Summary . . . . .	10
2.3	Introduction . . . . .	11
2.3.1	Literature review and motivation . . . . .	11
2.3.2	Key metrics and notation . . . . .	15
2.3.3	Research questions . . . . .	16
2.4	Study site description . . . . .	17
2.4.1	Data description . . . . .	19
2.5	Modeling setup . . . . .	20
2.5.1	Rainfall-runoff model . . . . .	21
2.5.2	rSAS transit time model . . . . .	22
2.5.2.1	Previous rSAS calibration and the inverse storage effect . . . . .	23
2.5.2.2	Diagnosing an inverse storage effect . . . . .	24

2.5.3	Rainfall generator . . . . .	26
2.6	Sensitivity analysis methods . . . . .	27
2.6.1	Phase 1 methods: historic analysis . . . . .	27
2.6.2	Phase 2 methods: pattern analysis . . . . .	27
2.6.3	Phase 3 methods: climate change analysis . . . . .	29
2.7	Results . . . . .	30
2.7.1	Model calibration results . . . . .	30
2.7.1.1	Rainfall-runoff model calibration results . . . . .	30
2.7.1.2	rSAS model calibration results . . . . .	32
2.7.1.3	Rainfall generator calibration results . . . . .	33
2.7.2	Phase 1 results: historic analysis . . . . .	35
2.7.3	Phase 2 results: pattern analysis . . . . .	36
2.7.4	Phase 3 results: climate change analysis . . . . .	40
2.8	Discussion . . . . .	42
2.8.1	Diagnosis of the inverse storage effect and its implications . . . . .	42
2.8.2	Variation in young water fraction explained by mean rainfall . . . . .	43
2.8.3	Interplay between rainfall pattern and the inverse storage effect . . . . .	45
2.8.4	Seasonality and timescale of the inverse storage effect . . . . .	47
2.8.5	Evaluation of rainfall pattern using a young water index . . . . .	48
2.8.6	Water quality simulation under a changing climate . . . . .	49
2.8.7	Study limitations . . . . .	50
2.9	Conclusions . . . . .	51
2.10	Appendix A: Flux pattern independence in an idealized system . . . . .	52
<b>3</b>	<b>Calibration and sensitivity of transit times in an integrated model</b>	<b>54</b>
3.1	Highlights . . . . .	54
3.2	Summary . . . . .	54
3.3	Introduction . . . . .	55
3.4	Site Description . . . . .	59
3.4.1	Physiography . . . . .	59

3.4.2	Hydrographic and age tracer data . . . . .	60
3.5	ParFlow-CLM description . . . . .	61
3.6	Methods . . . . .	63
3.6.1	Model structure and initial parameters . . . . .	63
3.6.2	GLUE calibration . . . . .	65
3.6.2.1	Stage 1: Steady-state simulation . . . . .	66
3.6.2.2	Stage 2: Periodic steady-state simulation . . . . .	67
3.6.2.3	Stage 4: Fully transient simulation . . . . .	69
3.6.2.4	Model evaluation . . . . .	69
3.6.3	Sensitivity analysis . . . . .	70
3.6.3.1	Monte Carlo filtering . . . . .	70
3.6.3.2	Boosted regression trees . . . . .	71
3.7	Results . . . . .	72
3.7.1	GLUE calibration results . . . . .	72
3.7.2	Monte Carlo filtering results . . . . .	73
3.7.3	BRT sensitivity analysis - single parameter . . . . .	75
3.7.4	BRT sensitivity analysis - Parameter interactions . . . . .	77
3.7.5	2D versus 3D performance . . . . .	79
3.8	Discussion and Conclusions . . . . .	81
3.8.1	Implications for ISSHM calibration . . . . .	81
3.8.2	Controls on catchment discharge . . . . .	82
3.8.3	Controls on baseflow transit times . . . . .	83
3.8.4	Interactions between celerity and velocity . . . . .	84
3.8.5	Limitations and future work . . . . .	84
<b>4</b>	<b>Decomposition of transit times by flowpath</b>	<b>85</b>
4.1	Highlights . . . . .	85
4.2	Summary . . . . .	85
4.3	Introduction . . . . .	86
4.4	Background theory: age-based distributions . . . . .	91



4.4.1	Transit time distributions (TTDs)	91
4.4.2	Storage Selection (SAS) functions	92
4.4.3	Age-rank storage-flux (rSF) functions	93
4.5	Materials and methods	95
4.5.1	Study site	95
4.5.2	ParFlow-CLM Modeling	97
4.5.2.1	Model Structure	97
4.5.2.2	Model Parameters and Calibration	98
4.5.2.3	Particle tracking: $Q$ and $ET$	100
4.5.2.4	Particle tracking: constituent pathways	101
4.5.2.5	Age-based distributions calculation	103
4.5.2.6	Comparative analysis	104
4.6	Results	104
4.6.1	Model performance overview	104
4.6.2	Pathway partitioning	107
4.6.3	Partition flux	108
4.6.4	Age-based distribution analysis: $ET$	110
4.6.5	Age-based distribution analysis: $Q$	113
4.6.6	TTD Taxonomy	116
4.7	Discussion and conclusions	118
4.7.1	Drivers of shape and transience	118
4.7.1.1	$ET$	118
4.7.1.2	$Q$	119
4.7.2	Driver of Inverse Storage Effect	120
4.7.3	Implications of "age-equipfinality" in catchments	120
4.7.4	Implications of a "middle-aged depression" in TTDs	121
4.7.5	Implications for lumped parameter models	122
4.7.6	Implications for C-Q relationships	122
4.7.7	Limitations and future work	122

4.8	Appendix A - Estimating $S_T$ using forward TTDs . . . . .	123
<b>5</b>	<b>Simulating groundwater nitrate transit times in SWAT-SAS</b>	<b>124</b>
5.1	Highlights . . . . .	124
5.2	Summary . . . . .	124
5.3	Introduction . . . . .	125
5.4	SWAT-SAS theory . . . . .	128
5.4.1	Age and nitrate modeling in SWAT . . . . .	130
5.4.2	Proposed age and nitrate modeling in SWAT-SAS . . . . .	132
5.5	SWAT-SAS case study . . . . .	134
5.5.1	Study site and SWAT model setup . . . . .	134
5.5.2	SWAT-SAS implementation . . . . .	136
5.5.2.1	Default parameterization . . . . .	138
5.5.2.2	Uniform calibration . . . . .	139
5.5.2.3	Spatial calibration . . . . .	142
5.6	Discussion and conclusions . . . . .	144
5.6.1	Comparison with steady-state approximation . . . . .	144
5.6.2	Limitations and future work . . . . .	145
<b>6</b>	<b>Conclusion</b>	<b>146</b>
6.1	Summary of key findings . . . . .	146
6.1.1	Application to the Chesapeake Bay watershed . . . . .	148
6.1.1.1	Chesapeake Bay watershed overview . . . . .	148
6.1.1.2	Implications for nitrate lag times . . . . .	150
6.1.1.3	Implications for climate change . . . . .	151
<b>A</b>	<b>Supplemental Information: Chapter 2</b>	<b>153</b>
	<b>Bibliography</b>	<b>179</b>

## List of Figures

1.1	Three hypothetical TTDs that could describe the flow-weighted average TTD of base-flow using traditional TTD modeling methods. These TTDs are gamma distributions with identical means (10 years) but different median and mode values. The shape of the TTDs suggest that watershed A may have more fast overland flow paths, whereas watersheds B and C have more groundwater flow paths. . . . .	4
1.2	Illustrative example of rSAS-LPM operation. Panel A shows rSAS functions for two hydrologic fluxes (Q and ET) in a hypothetical watershed. The upper rSAS function shows the fraction of ET selected from different volumes of age-ranked storage. The median of the rSAS PDF represents the volume of water in storage from which the youngest half of water is sampled. The lower plot shows how the rSAS function might vary with time (i.e., t1 and t2) according to a state variable (e.g., total storage). Panel B shows how the rSAS functions act on storage ranked by age (shown as a conceptual storage column) for t1. The storage column represents the water in storage with the younger water stacked over the older water. The proportion of water ages selected in each outflow is represented by the size and color of the arrows, which (by definition) trace the shape of the rSAS functions. The composition of ages will vary with time due the dynamic fluxes in and out. Panel C illustrates that the model produces a different TTD at each time step. The TTD captures the erratic time history of precipitation inputs. (Conceptualization modified from Harman 2015.) . . . . .	7
2.1	Map showing the location of the Lower Hafren and Tanllwyth experimental catchments, with details about monitoring sites, topography, and the stream network. . . . .	18

2.2	<p>Rainfall scenarios used in the sensitivity analysis. Panel a shows the observed rainfall from 1999 to 2008 (top sub-panel) and the yearly average rainfall (bottom sub-panel). Panel b shows synthetic rainfall patterns with incrementally increasing variability. A representative 360 day period is shown. To show seasonal trends, the insets depict the average simulated rainfall for each day of the year, averaged over 300 years. The bottom sub-panel represents the full variability of the observed rainfall in our study sites. Panel c shows the rainfall change factors derived from the UKCP09 model (Murphy et al., 2007) downloaded at <a href="http://ukclimateprojections.metoffice.gov.uk">http://ukclimateprojections.metoffice.gov.uk</a> for each of the four climate change storylines described in the text. The gray shading shows the 90th percentile range of change factor values. . . . .</p>	28
2.3	<p>Calibration results for the rSAS model in both catchments. Panel a shows the histogram of parameter values and the most likely values (black triangles) given by DREAM. Panel b shows the parameterized shape of the rSAS function under 5th percentile (dry), 50th percentile (normal), and 95th percentile (wet) storage anomaly conditions. The lines bounded by color show the rSAS function based on the most likely parameter values, and the shaded areas show the 95th percentile confidence interval due to parameter uncertainty. The dotted lines illustrate the range of median <math>S_T</math> values in each watershed. . . . .</p>	31
2.4	<p>Values of key model inputs, outputs, and internal state variables for the coupled modeling framework for a representative 2-year period in the LH (left) and TW (right). Panels a-b, c-d, and e-f show <math>J(t)</math>, <math>ET(t)</math>, and <math>Q(t)</math>, respectively. Panels g-h show the storage anomaly <math>\Delta S(t)</math> derived from the rainfall runoff model. Panels i-j show the observed concentration of chloride tracer in rainfall <math>C_J(t)</math> and the stream <math>C_Q(t)</math>, overlaid with the stream concentration simulated by rSAS with the most likely parameter set. Panels k-l show the median <math>S_T</math> of <math>\Omega_Q(S_T, t)</math>. The shaded area captures the 95th percentile confidence interval due to parameter uncertainty. Panels m-n show the simulated fraction of water younger than 1 day, 30 days, 90 days, 1 year, and 5 years assuming the most likely parameter set. . . . .</p>	34

2.5	Relationship between the young water fraction anomaly $\Delta F_i(t)$ (defined in Section 2.7.2) and mean rainfall averaged over four timescales $i$ : daily (panel a), monthly (panel b), seasonally (panel c), and yearly (panel d). The dashed line shows a line of best fit. The error bars show the 90th percentile confidence interval. The right-most panels indicate the slope $s$ of the line of best fit and the median $F_{at}$ for all parameter sets. The gray triangular regions are described in Section 2.8.2. . . . .	36
2.6	Panel a reproduces from Figure 2.2a the observed rainfall record marked "full variability" (blue line) and the yearly mean rainfall marked "no within-yr variability" (black line). Panel b shows the relationship between mean annual precipitation $J_y(t)$ and modeled $F_y(t)$ in LH and TW with (1) full variability (closed circles) and (2) no within-year rainfall variability (open circles). Note that rainfall variability increases $F_y(t)$ , but changes in $J_y(t)$ have roughly the same marginal effect on $F_y(t)$ with or without within-year variability. The cases with no variability follow the steady-state solution (dashed line) described in Section 2.8.2. . . . .	37
2.7	The top row shows the effect of different modes of rainfall variability on the long-term mean FYW for all time (column a), all winters (panel b), and all summers (column c) in 100-year simulations. The error bars show the standard deviation of $F_y(t)$ . The bottom two rows (panels a1-a4) show the calculated storage-based and discharge-based young water indices versus estimates of $F_{at}$ taken from panel a. Each point represents one of the five scenarios, as labeled in panel a1. Increases in both indices are associated with increases in $F_{at}$ , especially in LH. The light gray lines show $F_{at}$ (right y-axis) for scenarios re-run with no inverse storage effect (i.e., $\lambda = 0$ ), as discussed in Sections 2.8.3 and 2.8.5. . . . .	38

- 2.8 The interplay between different storm patterns, model parameters, and  $F_{at}$  in the LH. The first seven days of repeating ten day rainfall patterns are shown (solid gray lines) along with simulated discharge (dashed line) split into young (pink) and old (blue) components. No rain fell during the last three days of the repeating patterns, which are not shown. The mean young water fraction and SBYW are given for each pattern. Note that discharge has a one-to-one non-linear relationship with  $\Delta S(t)$ , shown on the right y-axis. The panels show a tendency for  $F_{at}$  to increase when storms get more clustered (panel a), storms are negatively skewed (panel b-c), and storms are less than 3 days apart (panel d). Panels also show how both  $Q(t)$  and  $F(t)$  increase sharply when  $\Delta S(t) > 35mm$ . This value is both  $\eta_Q$  and  $\eta_\Omega$  as labeled in the lower right sub-panels and defined in Section 2.8.3. The last panels show how changes in the storage-discharge relationship (panel e) and inverse storage effect (panel f) affect  $F_{at}$ . 39
- 2.9 Results from the climate change simulations. Panels a-c show the simulated long term mean FYW averaged over three different time periods (i.e., all time  $F_{at}$ , all winter  $F_{aw}$ , and all summer  $F_{as}$ ) for each climate change storyline. Storyline acronyms are baseline (BL), median change in evapotranspiration (MCET), median change in rainfall (MCR), and extreme change in rainfall (ECR), as described in Section 2.6.3. The error bars show the 80th percentile range for  $F_y(t)$  over each 10 year modeling period. The bottom panels show the average monthly volume and age composition of water in discharge from the LH and TW for three of the four storylines. . . . . 41
- 2.10 Illustration of thresholds in the relationship between storage anomaly and discharge (solid line, left axis) and storage anomaly and the fraction of water in discharge from the youngest 350 mm of catchment storage (dashed line, right axis). Both rise more sharply at high storage, indicating mobilization of both large volumes of water and large fraction of the youngest age-ranked water in storage. The dashed lines indicate the approximate, overlapping location of  $\eta_Q$  and  $\eta_\Omega$  for LH and TW. . . . . 43

3.1	Map of the study site in the FD36 sub-catchment of the USDA Mahantango experimental catchment. (a) Location of the catchment in central PA, USA. (b) Satellite image of the catchment showing the rectangular region modeled in ParFlow and the FD36 catchment boundary. (c) 3D rendering of the rectangular area modeled by ParFlow, the catchment boundary, the location of the streambed and subsurface fragipan, the location of the stream outlet and CFC sampling well, and the transect A-A' used for the 2D calibration and shown in Figure 3.2. . . . .	59
3.2	Schematic of the 2D catchment transect modeled used in calibration, including the location of the four hydrogeologic units (soil, fragipan, streambed, and fractured rock) and the calibrated parameters (as defined in Section 3.6). . . . .	60
3.3	Flow chart showing steps in the model calibration process. . . . .	66
3.4	(a) The 2014 daily rainfall in the FD36 catchment and (b) the observed and simulated discharge. . . . .	72
3.5	(a) The observed and simulated flow duration curve and (b) comparison between ages using different estimation techniques. . . . .	74
3.6	Results of the global sensitivity analysis for the five most sensitive parameters. Each panel has four subpanels containing the Monte Carlo filtering results (top panel) and the single-parameter partial dependency plots (lower three panels). Each point in the blue cloud is the response from one of the 22886 parameter sets evaluated. . . . .	76
3.7	The two-parameter partial dependency plots for the porosity parameters (left columns) and the log of lateral hydrologic conductivity parameters (right columns). The response variables are high flows (upper row), low flows (middle row), and median baseflow age (bottom row). . . . .	78
3.8	The two-parameter partial dependency plots for the porosity parameters (left columns) and the log of lateral hydrologic conductivity parameters (right columns). The response variables are median baseflow travel times (upper row), median flow-weighted travel times (middle row), and median steady state travel times (bottom row) . . . . .	80
3.9	Comparison of 2D emulator and full 3D catchment simulation. . . . .	81

4.1	Map of the study site in the FD36 sub-catchment of the USDA Mahantango experimental catchment. (a) Location of the catchment in central PA, USA. (b) Satellite image of the catchment showing the rectangular region modeled in ParFlow and the FD36 catchment boundary. (c) 3D rendering of the rectangular area modeled by ParFlow, the catchment boundary, the location of the streambed and subsurface fragipan, the location of the stream outlet and CFC sampling well, and the transect A-A' used for the 2D calibration and shown in Figure 4.2. . . . .	96
4.2	Cross-section of the modeled domain across the transect A-A' (see Figure 4.1), including the location of the four geologic zones (soil, fragipan, streambed, and fractured rock), the calibrated model parameters, and the five modeled flow pathways into the stream: direct rainfall (DIR), overland flow (OVR), interflow (INT), shallow groundwater (SGW), and deep groundwater (DGW). Section 4.5.2.4 gives the operational definition for flow pathway classification. . . . .	99
4.3	Important ParFlow-CLM model inputs, outputs, and performance for the FD-36 catchment. Panel (a) shows total precipitation $J_{prec}$ and total infiltration to the land surface after accounting for above-surface processes including snow accumulation, snowmelt, and canopy interception and evaporation $J_{inf}$ . Panel (b) and (c) shows surface $ET$ and $Q$ simulated by the ParFlow-CLM model (black) and by particle tracking (gray). Panel (c) also shows observed daily discharge against time (red circles) and against simulations in log-log space (see inset). Panel (d) shows total water storage in the catchment over time. Panels (e) and (f) show the time-series of median transit time (MTT) for $ET$ and $Q$ (black lines) with the 10-90th percentile ages shaded gray. The red strip shows the tracer-derived age of baseflow in a nearby stream (McGuire, DeWalle, and Gburek, 2002). . . . .	106



4.4	The partitioning of rainfall into different flow pathways. Panel (a) shows the fraction of $J_{inf}$ exiting to different flow pathways for daily rainfall during the 1-year simulation. The white column breaks indicate periods with zero infiltration. Panel (b) shows the fraction of each flow pathway contributing to daily $Q$ . Panel (c) shows the median transit time for each flow pathways contributing to $Q$ and $ET$ . All flowpath acronyms are defined in the text. . . . .	108
4.5	Time-series of simulated $Q$ for each flowpath and comparisons to $S$ , $J_{prec}$ , and $Q$ . The first column shows total flux from DIR, OVR, INT, SQW, and DGW. Note the changing y-axis range. The second column plots simulated storage against discharge. The last column plots simulate discharge against $J_{prec}$ (panel a3) and total $Q$ (panel b3). For each scatter plot, lines of best fit are shown in red. . . . .	109
4.6	Analysis of age, storage, and outflow for $ET$ (top row) and its constituent flowpaths ETG and ETT (bottom two rows) during the 1-year simulation. The left column shows the median backward transit-time distribution $\overleftarrow{P}(T)$ at four different ranges of catchment storage: very dry (lower 10th percentile storage, lightest blue line), moderately dry (10-50th percentile, lighter blue line), moderately wet (50-90th percentile, darker blue line) and very wet (greater than 90th percentile, darkest line). The light red shading shows the 80th percentile daily range of $\overleftarrow{P}(T)$ . The middle column is similar but shows the SAS functions and the approximate analytic solutions derived in the text (red dashed lines, panels (e) and (h)). The last column shows the rSF functions. . . . .	111

4.7	Analysis of age, storage, and outflow for $Q$ (top row) and its constituent flowpaths DIR, OVR, INT, SGW, and DGW (bottom five rows) during the 1-year simulation. The left column shows the median backward transit-time distribution $\overleftarrow{P}(T)$ at four different ranges of catchment storage: very dry (lower 10th percentile storage, lightest blue line), moderately dry (10-50th percentile, lighter blue line), moderately wet (50-90th percentile, darker blue line) and very wet (>90th percentile, darkest line). The light red shading shows the 80th percentile daily range of $\overleftarrow{P}(T)$ . The middle column shows the SAS functions. The insets in panels (d) and (e) zoom into the youngest portion of the DIR flowpath. The last column shows the rSF functions. Note that DIR and OVL are intermittent (see activity in Table 4.2), and only days with rainfall are used to construct the distributions. . . . .	114
4.8	A taxonomy of transit time distributions and their constituent flow paths. Panel (a) shows the K-means clustering of daily $\overleftarrow{P}_Q(T)$ into four categorical shapes: type i, type ii, type iii, and type iv. Panel (b) shows the hydrograph, the occurrence of each type during the 1 year simulation (upper color bands), and the date of four archetypes (colored diamonds) explored in the four columns below. Each column contains the backward TTD $\overleftarrow{P}_Q(T)$ (upper row), the SAS function $\Omega(S_T)$ (middle row), and the rSF function (lower row). The distributions are filled in with colors to show the contribution from each flow pathway to the overall distribution. . . . .	117
5.1	Conceptualizing of SWAT and proposed changes. Panel (a) shows how nitrate transport is conceptualized in SWAT as a uniform selection from a control volume. Panel (b) shows the proposed approach using SAS functions to different proportions select age-ranked storage. The SAS functions are shown in the PDF form $\omega(S_T)$ , and the shape is indicative of uniform selection. . . . .	129
5.2	Map of the case study site and hydrology including the catchment site (Panel a), the distribution of HRUs (Panel b), and the amount and distribution of winter recharge (Panel c) and summer recharge (Panel d). . . . .	135

5.3	Time-series of inputs and outputs for a representative HRU using the default parameterization. The results demonstrate that SWAT-SAS can successfully reproduce SWAT simulations. . . . .	137
5.4	SWAT-SAS simulations of recharge (Panel a) and median transit time distribution (panel b) for a representative HRU using the default parameterization. . . . .	139
5.5	Map of spatially distributed groundwater ages including the reference age distribution taken from Sanford et al. (2012) (Panel a), along with the SWAT-SAS simulated ages using the default parameters (Panel b), the uniform calibration parameters (Panel c), and the spatially distributed calibration (Panel d). . . . .	140
5.6	Comparison of reference TTD with default parameters and spatially uniform parameterization (Panel a), along with the spatially distributed parameterization (Panel b). . . . .	142
6.1	Location of the Chesapeake Bay watershed and median groundwater age estimates. The physiographic provinces of the CBW are shown in shades of gray and the bay in white. The orange shading shows the median groundwater ages for sub-watersheds of the upper Potomac (western side) and eastern shore (eastern side), which were extracted from MODPATH modeling results provided by USGS. . . . .	149
A.1	Comparison between the discharge observed and the discharge simulated by the rainfall-runoff model at LH (top panels) and TW (lower panels). The Nash-Sutcliff efficiency (NSE) for the calibration period (1999-2008) and validation period (1993-1998) are written in the left panels. The right scatter plots show good agreement between observed and simulated data. . . . .	157
A.2	PDF of the seven rSAS model parameters at LH, including the three nuisance parameters $a_\sigma$ ("a coef"), $b_\sigma$ ("b coef") and $\phi$ ("autoreg") described in the text of this Supplemental Information. Values were determined from the last 3,200 parameter sets in the MCMC simulation performed by DREAM. . . . .	158
A.3	Same as Figure A.2 but for TW. . . . .	159

A.4	Scatter plot matrix for the four rSAS parameters shown in Figure A.2 for the LH watershed. Each dot represents one of the final 3,200 parameter sets in the MCMC simulation run by DREAM. Note, for example, that $S_0$ and $\lambda$ have an especially strong negative correlation. . . . .	160
A.5	Same as Figure A.4 but for TW. . . . .	161
A.6	Plot of observed stream chloride concentrations $C_Q$ (red dots) and the range of simulated values using the distribution of likely parameter sets identified by DREAM. At daily timesteps, each of the last 3,200 parameters sets in the MCMC simulation provided an estimate of $C_Q$ , with the 90th percentile range of these estimates shown in green. The gray region shows the total uncertainty taking into account the 90th percentile of the variance of the residuals from Equation (A.3). Note that most of the observations fall within the total uncertainty range. . . . .	162
A.7	Same as Figure A.6 but for TW. . . . .	163
A.8	Analysis of residuals of the most likely rSAS parameter set at LH, following the graphical analysis of Vrugt et al. (2009). The upper plot compares observations (blue line) with simulates (red dots) of stream chloride concentration. The lower left plot illustrates the presence of heteroskedasticity in the residuals. As the simulated chloride concentration (x-axis) increases, the spread in residuals also tends to increase (y-axis). The lower center plot shows the PDF of residuals (red dots) and the average of all residual PDFs that would be expected from each simulated data point given the heteroskedastic standard deviation described by Equation (A.3). The observed and expected distribution overlap reasonably well. The lower right panel shows the autocorrelation of the error $e_t(\mathbf{x})$ from the full residuals (light blue line) and the autocorrelation of the error $\eta_t$ from the partial residuals (dark blue) after accounting for a first-order autoregressive process. The partial residuals are better constrained within the natural scatter of the data, represented here as a 90th percentile interval (upper and lower red lines). . . . .	164
A.9	Same as Figure A.8 but for TW. . . . .	165

A.10	Plot illustrates the performance of the parameters in the rainfall generator model that control the mean, phase, and seasonal amplitude shift for storm duration (upper plot) and break duration (lower plot). The bars in the upper plot show the monthly average storm duration at Plynlimon from 1983-2008. The red lines shows the analytically derived monthly average storm duration after calibration of the mean $\delta_r$ , $\alpha_r$ , and $\tau_r$ . The bars in the bottom plot show the monthly average interval between storms at Plynlimon from 1983-2008. The red lines shows the analytically derived monthly average storm break after calibration of the mean $\delta_b$ , $\alpha_b$ , and $\tau_b$ . Plots follow analysis from Robinson and Sivapalan (1997). . . . .	166
A.11	Plot illustrates the performance of the parameters in the rainfall generator model that control the average storm intensity as a function of storm length (i.e., $a_1$ , $b_1$ , and their seasonal factors) and the coefficient of variability as a function of storm length (CV)(i.e., $a_2$ , $b_2$ ). The top (first) plot shows the expected storm intensity observed from 1983-2008 (black line) and analytically derived for the rainfall generator (red dashed line). The lower plots make the same comparison for the CV of storm intensity (second plot), the expected value of the product of storm intensity and storm duration (third plot), and the correlation coefficient for the intensity and duration (fourth plot). Plots and analytic calculations follow Robinson and Sivapalan (1997). . . . .	167
A.12	Plot illustrates the performance of the parameters in the rainfall generator that control within-storm variability of rainfall (i.e., parameters $\beta_1$ and $\beta_2$ , which are set equal). The 10th, 50th, and 90th percentile mass curves for all storms from 1983-2008 are given by the dashed lines. The mass curves from a representative simulation are shown by the solid lines. The overlap suggests the weather generator is able to simulate the observed within-storm variability. Plots follow approach of Robinson and Sivapalan (1997). . . . .	168

A.13	Results from the coupled-model evaluation at LH. The solid black lines show the CDF of observed rainfall (column a) and observed discharge (column b). The dashed black lines show the CDF of simulated discharge based on observed rainfall (column b) and simulated fraction of young water ( $F$ ) based on observed rainfall and discharge with the most likely rSAS parameter set (column c). The red area is the 95th percentile of 100 CDFs generated from 100 16-year simulations of rainfall (column a), from 100 16-year simulations of discharge based on the simulated rainfall (column b), and from 100 16-year simulations of the fraction of young water based on the simulated rainfall and discharge (column c). The black curves fall within the red regions, indicating the observed and simulated values have similar probability distributions for the all time (upper row), winter only (middle row), and summer only (lower row) cases. . . . .	169
A.14	Same as Figure S13 but for TW. . . . .	170
A.15	Same as Figure 5 in the manuscript, except the age fraction anomaly for water less than 1 day old is shown on the y axes instead of the age fraction anomaly of water less than 90 days old. The x axes show the mean rainfall averaged over four timescales $i$ : daily (panel a), monthly (panel b), seasonally (panel c), and yearly (panel d). The dashed line shows a line of best fit. The error bars show the 90th percentile confidence interval. The right-most panels indicate the slope $s$ of the line of best fit and the median $F_{at}$ for all parameter sets. . . . .	171
A.16	Same as Figure A.15 with the age fraction threshold for discharge set to water less than 3 days. . . . .	172
A.17	Same as Figure A.15 with the age fraction threshold for discharge set to water less than 30 days. . . . .	173
A.18	Same as Figure A.15 with the age fraction threshold for discharge set to water less than 60 days. . . . .	174
A.19	Same as Figure A.15 with the age fraction threshold for discharge set to water less than 90 days. Note these results are identical to the results shown in Figure 5 of the manuscript. . . . .	175

A.20 Same as Figure A.15 with the age fraction threshold for discharge set to water less than 120 days. . . . .	176
A.21 Same as Figure A.15 with the age fraction threshold for discharge set to water less than 365 days. . . . .	177
A.22 Same as Figure A.15 with the age fraction threshold for discharge set to water less than 1825 days. . . . .	178

## List of Tables

2.1	Notation used to represent the mean fraction of young water $F(t)$ and rainfall $J(t)$ at different timescales. Notation for $Q(t)$ and $ET(t)$ is similar to $J(t)$ . The last column lists the number $n$ of times the quantities were estimated in each of the two watersheds over the 10 year study period from 1999 to 2008. . . . .	15
2.2	Characteristics of the study sites. Soils are moorland (M), podzol (P), or gley (G). Catchment area, soil types, and modeled estimate of the mean transit time (MTT) are from Kirchner, Feng, and Neal (2001). . . . .	19
2.3	Results from the calibration of the rainfall-runoff model. . . . .	31
2.4	Results from the calibration of the rSAS model with 95th percentile confidence intervals in brackets. . . . .	32
2.5	Results from the calibration of the rainfall generator. Parameters with the $a$ superscript are adjusted an additional +19% to -51% depending on the season. Parameters are defined in Robinson and Sivapalan (1997) and in the SI. . . . .	33
3.1	List of fixed (i.e., not calibrated) model parameters. . . . .	64
3.2	List of parameters calibrated using GLUE, and the range of values that was explored. . . . .	65
3.3	List of GLUE calibration results including the 5th and 95th percentile of the prior and posterior distribution, and the median of the prior and posterior distribution. . . . .	74
4.1	List of key ParFlow-CLM model parameters. . . . .	98
4.2	Characteristics of pathway partitions for $ET$ and $Q$ . . . . .	115
4.3	Characteristics of the four types of transit time distributions (first four columns) and all types combined (last column). . . . .	118



## List of Abbreviations

<b>BL</b>	<b>BaseLine</b> scenario
<b>BRT</b>	<b>Boosting Regression Tree</b>
<b>bTTD</b>	<b>backward Transit Time Distribution</b>
<b>CDF</b>	<b>Cumulative Distribution Function</b>
<b>CLM</b>	<b>Common Land Model</b>
<b>DIR</b>	<b>DIRect</b> runoff flowpath
<b>DGW</b>	<b>Deep Groundwater</b> flowpath
<b>DREAM</b>	<b>Differential Evolution Adaptive Metropolis</b>
<b>ECR</b>	<b>Extreme Change in Rainfall</b> scenario
<b>ET</b>	<b>EvapoTranspiration</b>
<b>ETG</b>	<b>EvapoTranspiration: Ground</b> evaporation flowpath
<b>ETT</b>	<b>EvapoTranspiration: Transpiration</b> flowpath
<b>INT</b>	<b>INTerflow</b> flowpath
<b>ISSHM</b>	<b>Integrate Surface Subsurface Hydrology Model</b>
<b>fTTD</b>	<b>forward Transit Time Distribution</b>
<b>FYW</b>	<b>Fraction Young Water</b>
<b>GCM</b>	<b>General Circulation Model</b>
<b>HRU</b>	<b>Hydrologic Response Unit</b>
<b>GLUE</b>	<b>Generalized Likelihood Uncertainty Estimation</b>
<b>KGE</b>	<b>Kling-Gupta Efficiency</b>
<b>LH</b>	<b>Lower Hafren</b> catchment
<b>MC filtering</b>	<b>Monte Carlo</b> filtering
<b>MTT</b>	<b>Median Transit Time</b>
<b>MCMC</b>	<b>Markov Chain Monte Carlo</b>

<b>MCET</b>	<b>Median Change in EvapoTranspiration scenario</b>
<b>MCR</b>	<b>Median Change in Rainfall scenario</b>
<b>NSE</b>	<b>Nash-Sutcliffe Efficiency</b>
<b>OVR</b>	<b>OVeRland flow flowpath</b>
<b>ParFlow-CLM</b>	<b>Parallel Flow - Common Land Model</b>
<b>PDF</b>	<b>Probability Distribution Function</b>
<b>PTM</b>	<b>Particle Tracking Model</b>
<b>QBYW index</b>	<b>Q (Discharge) Based Young Water index</b>
<b>rSAS function</b>	<b>age-ranked StorAge Selection function</b>
<b>rSAS-LPM</b>	<b>age-ranked StorAge Selection - Lumped Parameter Model</b>
<b>rSF function</b>	<b>age-ranked Storage-Flux function</b>
<b>SAS function</b>	<b>StorAge Selection function</b>
<b>SBYW index</b>	<b>Storage Based Young Water index</b>
<b>SGW</b>	<b>Shallow Groundwater flowpath</b>
<b>SWAT</b>	<b>Soil and Water Assessment Tool</b>
<b>SWAT-SAS</b>	<b>Soil and Water Assessment Tool - StorAge Selection function coupled model</b>
<b>TT</b>	<b>Transit Time</b>
<b>TW</b>	<b>Tanllwyth catchment</b>
<b>TTD</b>	<b>Transit Time Distribution</b>
<b>USDA</b>	<b>United States Department of Agriculture</b>

## List of Symbols

$C_j$	concentration of solute in storage volume or outflux $j$	[M/L <sup>3</sup> ]
$DGW$	deep groundwater pathway flow	[L/T]
$DIR$	direct runoff pathway flow	[L/T]
$ET$	evapotranspiration	[L/T]
$ETG$	ground evaporation pathway flow	[L/T]
$ETT$	transpiration pathway flow	[L/T]
$F_{rock}$	e-folding depth for fractured rock $K$	[1/L]
$f_{T,j}$	age-rank storage-flux function of outflux $j$	[L/T]
$INT$	interflow pathway flow	[L/T]
$g$	catchment sensitivity function	[1/T]
$J$	rainfall	[L/T]
$J_{prec}$	rainfall reaching upper canopy	[L/T]
$J_{inf}$	rainfall infiltrating into land surface	[L/T]
$K$	saturated hydrologic conductivity	[L/T]
$K_{xy,j}$	lateral soil hydrologic conductivity multiplier in region $j$	[-]
$K_{z,j}$	vertical soil hydrologic conductivity multiplier in region $j$	[-]
$MTT_{bf}$	median transit time of baseflow based on periodic steady-state simulation	[T]
$MTT_{fw}$	flow-weighted median transit time based on periodic steady-state simulation	[T]
$MTT_j$	median transit time of outflux $j$	[T]
$MTT_{ss}$	median transit time of discharge based on steady-state simulation	[L/T]
$OVR$	overland pathway flow	[L/T]
$\overrightarrow{P}_{j,i}$	CDF of forward TTD of outflux $j$ and constituent flowpath $i$	[-]
$\overrightarrow{p}_{j,i}$	PDF of forward TTD of outflux $j$ and constituent flowpath $i$	[1/T]
$\overleftarrow{P}_{j,i}$	CDF of backward TTD of outflux $j$ and constituent flowpath $i$	[-]

$\overleftarrow{p}_{j,i}$	PDF of backward TTD of outflux $j$ and constituent flowpath $i$	[1/T]
$\overrightarrow{P}_s$	CDF of residence time distribution of catchment storage	[-]
$PET$	potential evapotranspiration	[L/T]
$Q$	discharge	[L/T]
$Q_T$	age-ranked discharge	[L/T]
$\overline{Q}_T$	age-ranked discharge complement	[L/T]
$Q_5$	discharge with 5% exceedance probability (peak flow)	[L/T]
$Q_{95}$	discharge with 95% exceedance probability (low flow)	[L/T]
$\overline{S}_T$	age-ranked storage complement	[L]
$S$	catchment storage	[L]
$S_{ref}$	reference storage	[L]
$S_T$	age-ranked storage	[L]
$SGW$	shallow groundwater pathway flow	[L/T]
$\overline{S}_T$	age-ranked storage complement	[L]
$T$	age	[T]
$t$	time	[T]
$\phi_j$	porosity in region $j$	[-]
$\Omega_{j,i}$	CDF of SAS function of outflux $j$ and constituent flowpath $i$	[-]
$\omega_{j,i}$	PDF of SAS function of outflux $j$ and constituent flowpath $i$	[1/L]

# Chapter 1

## Introduction

The ability to accurately model the timing and quantity of contaminant transport from landscapes to surface waters under different climate conditions is vital to the development of climate-resilient watershed management tools and strategies. Much progress has been made over the past decades in observing and modeling the relationships between climate and contaminant transport at the watershed scale. Less progress is evident in distinguishing the watershed processes underlying these relationships, which is important for developing robust predictive models, especially under a changing climate. A good approach to studying these watershed processes requires consideration of the watershed transit time distribution (TTD), though it has been historically difficult to measure the temporal variability of TTDs that result from realistic climatic conditions. There are, however, new advances in theory, computational power, and data availability for models that estimate time-varying TTDs at the watershed scale, encapsulated in (1) lumped parameter transit time models using rank StorAge Selection (rSAS) functions and (2) particle tracking models (PTMs) using discrete element numerical solvers. A central hypothesis of this dissertation is that these time-varying transit time modeling approaches can be used to improve our understanding of the influence of climate on catchment scale hydrologic transport, including transport of groundwater nitrate transport in the Chesapeake Bay watershed.

## 1.1 Motivation

### 1.1.1 Non-point source pollution

Non-point source (NPS) pollution is any contaminant that reaches surface waters from multiple, diffuse sources. NPS pollution most commonly originates from manure and synthetic fertilizer application and atmospheric deposition onto the landscape, and gets transported to surface waters by rainwater moving through complex catchment flow pathways. NPS has been described as “the nation’s largest water quality problem” (EPA, 1996) and “among the most difficult water-quality challenges” (Palanaippan et al., 2010). A major source of NPS pollution is excess nitrogen (N) and phosphorus (P) from agricultural and urban activities (Carpenter et al., 1998), and these excess nutrients are a leading cause of water impairment in the U.S. (EPA, 2012). For example, nitrate is one of the most pervasive NPS pollutants and a target of management action across the country. Nitrate that reaches surface waters is associated with human toxicity, harmful algal blooms, and eutrophication (e.g., Kemp et al., 2005). Human activity has more than doubled the rate of N inputs into the terrestrial N-cycle (Vitousek et al. 1997). A significant portion of N inputs take the form of nitrate that leaches into groundwater, where concentrations often rise to hazardous levels, especially in agricultural areas (Spalding and Exner, 1993).

### 1.1.2 Influence of climate on NPS pollution transport

A growing body of research shows that the quantity and timing of NPS loading to streams is influenced by the variability of climate, broadly defined here to include weather and its statistical properties. In addition, the overall influence of climate on NPS transport may be shifting due to global climate change (e.g., Najjar et al. 2010). Therefore, managing NPS pollution requires knowing, among other things, how climate affects NPS delivery to surface waters.

Climate can influence NPS transport by altering contaminant inputs, by altering hydrological processes, and by changing biogeochemical processes. While the influence of climate on these three pathways are often studied as a group, their individual influence on water quality is more difficult to understand. For example, researchers have used relatively simple input-output models to determine statistical relationships between historic climatic inputs and contaminant exports. These models

have shown, for example, that increases in precipitation are associated with higher fluxes of nitrogen from watersheds on both short (<1 year) and long time scales (e.g., Howarth et al. 2006). The statistical relationships determined by input-output models do not, however, allow one to determine which causal pathway is most important.

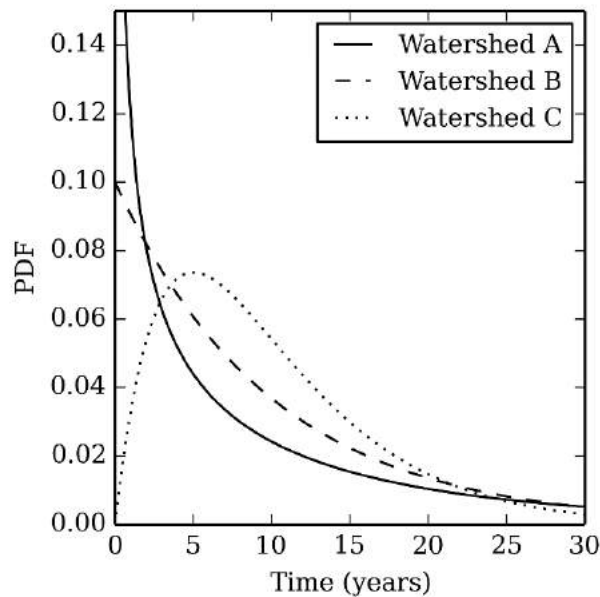
Other researchers have used process-based watershed models to study the link between climate and NPS transport. For example, Yang et al. (2015) calibrated a process-based model with data from 1901-2008 along eastern U.S. coastal watersheds. They found that changes in climate (temperature and precipitation) were the primary driver of interannual variability in nitrogen export to the Atlantic Ocean, with wetter periods seeing higher export. At a smaller scale, experiments run using the process-based Chesapeake Bay Community Watershed Model Phase 5.3.2 showed that increases in rainfall variability induced increases in the export of nitrate and other NPS pollution from the Patuxent river to the Chesapeake Bay (Wilusz and Ball, 2014). In principle, the internal structure and sensitivities of these process-based models could be scrutinized to understand how different causal pathways relate climate to transport. In practice, however, process based models often give uncertain representations of the interplay between climate and transport because of missing critical processes (Bernal and Hedin 2012). For example, many process-based models assume that water and the contaminants it carries are completely mixed within conceptual model storage zones (e.g., Neitsch et al., 2011), yet a growing body of research suggests that this assumption would be violated in many watersheds of interest (e.g., Benettin et al., 2015).

### 1.1.3 Transit-time models

Transit-time models are a promising alternative to input-output and overly simplistic process-based models for studying the link between climate and water quality. Transit-time modeling uses observed watershed behavior to simulate the watershed TTD, which is a probabilistic distribution of water travel times through a watershed. The TTD varies spatially and temporally according to local watershed characteristics including catchment geometry, topography, soil type, land use, and climatic conditions (McGuire and McDonnell, 2006). The shape of the TTD can be used to infer the dominant hydrologic processes within a watershed. For example, a higher proportion of younger water in watershed discharge is suggestive of more overland flow, and a higher proportion of older water is

suggestive of more groundwater flow. The shape of the TTD can be determined experimentally based on variations in the input and output of isotopes and other environmental tracers. Researchers have also shown that TTDs can be estimated from *a priori* knowledge of watershed flow pathways (McDonnell et al., 2010). Figure 1.1 gives illustrative examples of the kinds of TTDs that could be inferred in watersheds.

Climate is a very important factor in determining the shape and time-variability of the TTD. Consider, for example, a parcel of rain water that falls on watershed soils and flows towards surface waters. If climate conditions are wet due to high rainfall or low ET, the water parcel might travel quickly. If climate conditions are dry, the parcel might stagnate. Alternatively, wet conditions might open flow pathways to passive storage zones with long residence times, and dry conditions



**Figure 1.1:** Three hypothetical TTDs that could describe the flow-weighted average TTD of baseflow using traditional TTD modeling methods. These TTDs are gamma distributions with identical means (10 years) but different median and mode values. The shape of the TTDs suggest that watershed A may have more fast overland flow paths, whereas watersheds B and C have more groundwater flow paths.



might reduce soil permeability and increase fast-moving overland flow. Although simple, these cases illustrate the important but complex role of climate in determining the watershed TTD.

Model-based estimates of the catchment transit time distribution can be convolved with time-series of hydrochemical inputs such as nitrate to simulate transport to the watershed outlet:

$$C_{out}(t) = \int_0^\infty \overleftarrow{p}_{out}(\tau, t) C_{in}(T - \tau) e^{-\lambda\tau} d\tau \quad (1.1)$$

with an inflow concentration  $C_{in}(t)$ , catchment TTD  $\overleftarrow{p}_{out}(\tau, t)$ , first order rate of denitrification  $\lambda$ , time  $t$ , and a dummy variable of integration representing age  $\tau$  (Cook and Herczeg, 2012). Numerous papers have used the convolution integral to simulate nitrate transport at watershed scales (e.g., Morgenstern et al., 2015). To the extent that the catchment TTD depends on climate, Equation 1.1 defines a quantitative relationship between climate variability, shifting hydrologic pathways, and contaminant transport.

#### 1.1.4 TTD model advancements

In practice, a major barrier to using transit time models to study the link between climate and water quality has been that most transit time models have assumed that transit times are stationary. This may be reasonable over time scales much longer than the rate of major climatic variability (e.g., decadal). It can be shown, however, that actual TTDs vary significantly at times scales of interest to watershed managers (e.g., days to years). To date, relatively few studies have examined the time-variability of TTDs due to climate variability. For example, Rodhe, Nyberg, and Bishop (1996) showed how a discharge-weighted-time-invariant TTD could be used to estimate the time-varying TTD of a small hillslope in Sweden. Hrachowitz et al. (2010) used a long record of chloride tracer to parameterize steady-state gamma-shaped TTDs for three-year periods over roughly 20 years. Both studies found that the shape of the TTD varied systematically with precipitation, with more young water discharged during wet periods. Preliminary work for this dissertation looked at transit time studies in six locations around the world and compared their monthly estimates of stream age (mean or median) with remotely sensed estimates of watershed precipitation, temperature, and ET. The results showed a consistent negative correlation between stream age and rainfall volume, and a positive correlation between stream age and temperature and ET (Wilusz, 2015).

Two analytical methods have emerged that are capable of building on past efforts to estimate the time-variability of the TTD without *a priori* assumptions about discharge-weighted time invariance or smoothness. They are (a) the lumped parameter transit time models using a rank StorAge Selection (rSAS) functions and (b) particle tracking models (PTMs) for groundwater. Both are briefly described below.

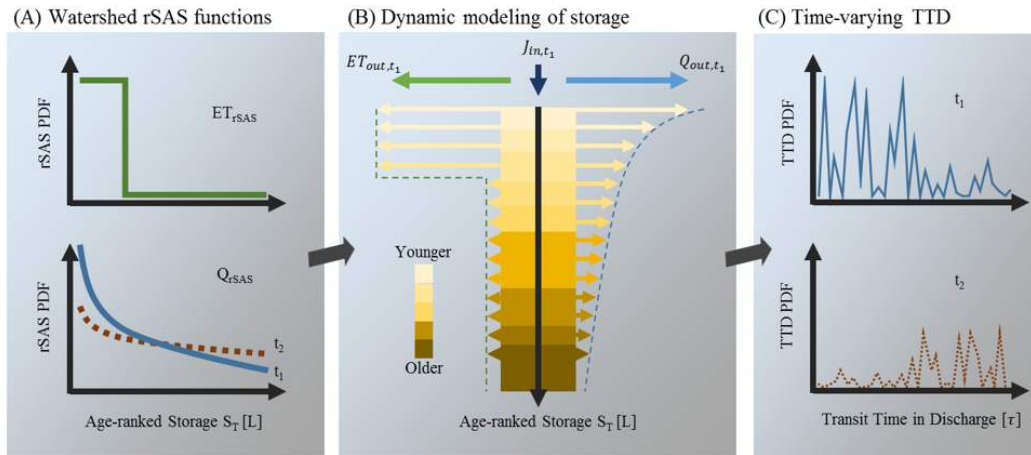
#### 1.1.4.1 rSAS-LPMs

The rSAS-LPM was introduced by Harman (2015) as an enhancement to the more general StorAge Selection modeling framework (Rinaldo et al. 2015). The model is able to (a) estimate the distribution of ages of water fluxes flowing out of a watershed (i.e., the time-varying TTD of  $Q$  and  $ET$ ) and (2) use the time-varying TTD to simulate solute transport from rainfall to discharge. In its most basic form, the rSAS-LPMs works by representing the watershed as a single control volume (see Figure 1.2). The input data includes fluxes coming in ( $P$ ) and out ( $Q$ ,  $ET$ ), which must be measured or estimated. Rainfall is assumed to be age zero. The distribution of ages of water "selected" from storage into each outflow is determined by a parameterized rSAS function, which is the probability distribution function (PDF) of storage ranked by age. If the rSAS functions for  $Q$  and  $ET$  are known, the time-varying TTD of  $Q$  and  $ET$  can be determined by solving a master differential equation expressing continuity of water age and mass (Botter, Bertuzzo, and Rinaldo, 2010). If input concentrations are known, they can be convolved with the time-varying TTD to estimate output concentrations (as in Equation 1.1). The convolution equation can be modified to account for both conservative and non-conservative transport. Importantly, an rSAS-LPM can have a different rSAS function for each outflow. For example, an application of the rSAS-LPM at Plynlimon, Wales suggested that the rSAS for  $ET$  flows should be a uniform distribution while the rSAS for  $Q$  should be a gamma distribution with a shape parameter that varies according to the total storage, as shown in Figure 1.2 (Harman, 2015).

The rSAS-LPMs have several advantages over direct parameterization of the TTD, which has been common in the literature (McGuire and McDonnell 2006). First, the rSAS function can be approximated by a simple and/or time-invariant CDF function while still representing the erratic time-variability of the TTD in a manner that conserves mass and age. Second, the parameters of the rSAS function may be more closely related to physical characteristics of the landscape compared to

parameters of the TTD, which are highly dependent on climate. Third, the rSAS-LPMs can directly simulate the effect of climatic inputs on the evolution of the time-varying TTD.

The principal challenge of rSAS modeling is the identification of the rSAS functions (shown in Figure 1.2(a)). rSAS functions describe the emergent tendency of all watershed processes to select outflow from different parts of storage. Identifying the rSAS function is closely related to identifying a watershed closure relation, a nonlinear hysteretic function that relates watershed inputs to outputs. Identification of the closure relationship is one of the most pressing problems in hydrology (Beven, 2006b). It is therefore promising that researchers have been able to use environmental tracers to parameterize rSAS functions as simple distributions (e.g., uniform, gamma distributions) that can



**Figure 1.2:** Illustrative example of rSAS-LPM operation. Panel A shows rSAS functions for two hydrologic fluxes ( $Q$  and  $ET$ ) in a hypothetical watershed. The upper rSAS function shows the fraction of  $ET$  selected from different volumes of age-ranked storage. The median of the rSAS PDF represents the volume of water in storage from which the youngest half of water is sampled. The lower plot shows how the rSAS function might vary with time (i.e.,  $t_1$  and  $t_2$ ) according to a state variable (e.g., total storage). Panel B shows how the rSAS functions act on storage ranked by age (shown as a conceptual storage column) for  $t_1$ . The storage column represents the water in storage with the younger water stacked over the older water. The proportion of water ages selected in each outflow is represented by the size and color of the arrows, which (by definition) trace the shape of the rSAS functions. The composition of ages will vary with time due the dynamic fluxes in and out. Panel C illustrates that the model produces a different TTD at each time step. The TTD captures the erratic time history of precipitation inputs. (Conceptualization modified from Harman 2015.)

vary with system state variables like storage (Harman, 2015; Benettin et al., 2015).

#### 1.1.4.2 Particle tracking models

An alternative approach to the empirically-based rSAS-LPMs for estimating the time-varying TTD is the physically-based approach adopted by particle tracking models (PTMs). The foundation of current-generation PTMs are process-based, finite element numerical models that solve fundamental mass-flux equations in a detailed 3D grid for surface and/or ground waters. Model parameters are calibrated against available field observations such as surface water maps, water depth, pressure measurements, and tracer-based age estimates. Once the numerical model is calibrated, a PTM is used to virtually introduce tracer particles into the groundwater recharge and track their movement through the subsurface until they are discharged into surface waters. By injecting and tracking particles across the entire landscape surface at different times, the time-varying TTD can be estimated. Although numerical models with PTMs have been in use since the 1960s (Pinder, 1973), their application has been limited by computational requirements to either steady-state, small-scale, and/or low resolution modeling. With modern computing capabilities, however, it is increasingly possible to model time-variability at relatively high resolution and scale. For example, Engdahl and Maxwell (2015) used the ParFlow PTM model to estimate the response of age distributions to changes in climate in different storage zones of a virtual watershed.

#### 1.1.5 Research objectives and dissertation outline

The overarching goal of this research was to make discoveries about the relationship between climate variability and catchment-scale transit times using a combination of (1) data-rich field sites, (2) recent developments in rSAS-LPM theory, and (3) innovative approaches to time-varying particle tracking in PTMs. The dissertation is split into four main chapters.

In Chapter 2, I explore the relationship between rainfall variability and catchment transit time distributions at the Plynlimon experimental catchment in Wales, UK. The rSAS-LPM is coupled to a hydrologic model and rainfall generator to answer the following research questions. How do the amount and pattern of rainfall affect catchment transit times under historic conditions? How could shifts in rainfall and evapotranspiration affect catchment transit times under projected future

conditions? The contributions of the study include benchmarking the influence of rainfall variability on the young water fraction in streams, developing and describing a Bayesian approach to classifying catchment transport, and providing the first-ever estimate of the effect of climate change on catchment transit times.

Chapter 3 explores the relationship between catchment characteristics and the catchment transit time distribution under realistic climate conditions for a system simulated by an integrated surface-subsurface PTM. The model is configured and parameterized with data from the Mahantango experimental catchment in PA, USA to answer the following research questions: what is the sensitivity of catchment transit times to changes in catchment characteristics such as porosity and lateral hydrologic conductivity? Which hydrologic processes drive this sensitivity? Methodological contributions include the first-ever implementation of daily-scale particle tracking in an open source integrated hydrology model, the first-ever calibration of such a model to discharge and tracer-based age data, and a novel, non-parametric sensitivity analysis using boosted regression trees.

Chapter 4 introduces a new technique for understanding catchment transit time distribution behavior in time-varying systems by disaggregating outfluxes into constituent pathways. The approach analyzed results from the PTM model in Chapter 3 with the SAS function theory applied in Chapter 2 to answer the following questions: what climatic factors cause the time-variability of catchment transit times including the inverse storage effect? To what extent can the flowpaths of discharge and ET be inferred by the shape of its transit time distribution? Methodological contributions include a new particle tracking approach that accounts for losses to ET, a generalizable method to demarcating flowpaths in PTMs, and a novel statistical approach to classifying the shape of catchment transit times.

Chapter 5 is a short technical note demonstrating how the commonly-used Soil and Water Assessment Tool (SWAT) can be modified to enable the calculation and calibration catchment travel times. The capabilities of the proposed SWAT-SAS model were demonstrated at the Chesterville experimental catchment in the Eastern Shore, MD. The SWAT-SAS model is intended to facilitate future efforts to simulate the effects of climate on transit times and NPS transport.

Chapter 6 summarizes key findings and discusses their implication for groundwater nitrate transport in the Chesapeake Bay Watershed.

## Chapter 2

### Sensitivity of catchment transit times to rainfall variability under present and future climates <sup>1</sup>

#### 2.1 Highlights

- A data-driven rSAS model was used to benchmark effects of rainfall variability on fractions of young water (FYW) in two first-order streams
- Mean FYW was surprisingly sensitive to rainfall pattern as a result of "inverse storage effect" that may be common elsewhere
- Projected changes in rainfall due to climate change caused simulated FYW to significantly rise in winter and drop in summer

#### 2.2 Summary

Hydrologists have a relatively good understanding of how rainfall variability shapes the catchment hydrograph, a reflection of the celerity of hydraulic head propagation. Much less is known about the influence of rainfall variability on catchment transit times, a reflection of water velocities that control solute transport. This work uses catchment-scale lumped parameter models to decompose the

---

<sup>1</sup>This chapter was published as Wilusz, Daniel C., Ciaran J. Harman, and William P. Ball. "Sensitivity of catchment transit times to rainfall variability under present and future climates." *Water Resources Research* 53.12 (2017): 10231-10256. Data for this chapter was archived at Wilusz, Daniel C., Ciaran J. Harman, and William P. Ball. "Data and code associated with the publication 'Sensitivity of catchment transit times to rainfall variability under present and future climates'", Johns Hopkins University Data Archive, doi:10.7281/T1/OM3OMQ (2017). Ciaran Harman and Bill Ball were involved in hypothesis development, study design, results interpretation, and editing of draft text. Ciaran Harman provided the SAS model code. Other model code, the first draft of text, and all figures and tables were created by Dano Wilusz.

relationship between rainfall variability and an important metric of transit times, the time-varying fraction of young water (<90 days old) in streams (FYW). A coupled rainfall-runoff model and rank StorAge Selection (rSAS) transit-time model were calibrated to extensive hydrometric and environmental tracer data from neighboring headwater catchments in Plynlimon, Wales from 1999-2008. At both sites, the mean annual FYW increased more than 13 percentage points from the driest to the wettest year. Yearly mean rainfall explained most between-year variation, but certain signatures of rainfall pattern were also associated with higher FYW including: more clustered storms, more negatively skewed storms, and higher covariance between daily rainfall and discharge. We show that these signatures are symptomatic of an "inverse storage effect" that may be common among watersheds. Looking to the future, changes in rainfall due to projected climate change caused an up to 19 percentage point increase in simulated mean winter FYW and similarly large decreases in the mean summer FYW. Thus, climate change could seasonally alter the ages of water in streams at these sites, with concomitant impacts on water quality.

## 2.3 Introduction

The influence of rainfall on landscape hydrology is a function of both (1) the celerity with which hydraulic head perturbations are transmitted through the catchment, which controls the generation of streamflow and the hydrograph; and (2) the velocity with which rainwater moves through the catchment, which determines solute transport and controls the shape of the transit time distribution (TTD) (McDonnell and Beven, 2014). Much has been learned by studying the relationship between rainfall and the hydrograph, for example, in "rainfall-runoff" models. In contrast, relatively little is known about how rainfall drives the evolution of time-varying TTDs. In this paper we use recent advances in time-varying transit time modeling to understand the influence of rainfall variability on an important metric of transit times, the young water fraction, at two study sites in Plynlimon, Wales.

### 2.3.1 Literature review and motivation

TTDs are a fundamental descriptor of catchment behavior (McGuire and McDonnell, 2006). The TTD concept is based on the representation of catchments as a dynamic population of water parcels that

age as they move from inflow to outflow. In general terms, the TTD is the probability distribution of travel times of water parcels that move through a catchment, for example from rainfall to discharge. More specifically, TTDs can be defined in terms of tracking ages forward (i.e., time to exit) or backward (i.e., time since entrance) in time (Niemi, 1977; Botter, Bertuzzo, and Rinaldo, 2011). The backward TTD (bTTD) of discharge - defined as the distribution of water ages flowing through the catchment stream outlet at a certain time - is most relevant for considering catchment controls on the variability of stream chemistry (Harman, 2015), and is the focus of this work.

Rainfall is one of the primary factors determining the shape and transience of bTTDs (Botter, Bertuzzo, and Rinaldo, 2010). Conceptually, the rainfall history has two important consequences. First, it constrains the evolution of the bTTD through conservation of water mass by age. For example, none of the water in discharge can originate from a time period without rainfall. Second, rainfall influences the pathways that water travels through the watershed. Heidbüchel, Troch, and Lyon (2013) described this effect as a "hydrologic response cascade" in which the distribution and intensity of rainfall influences the way that water fills watershed storage, which in turn determines flow pathways and transit times. Kim et al. (2016) refined this observation by noting that rainfall can affect flow pathways in two ways: (1) by changing the flow rate by an equal proportion along all existing flow paths (flow path acceleration) or (2) by changing the arrangement of, and/or partitioning between flow paths (flow path reorganization or activation). The effect of rainfall on bTTDs is also mediated by other inherent properties of the watershed, including geology, topography, soils, and vegetation (Heidbüchel, Troch, and Lyon, 2013).

A major challenge to understanding the link between rainfall and bTTDs has been accurate estimation of the bTTD, which cannot be directly observed. Previous studies have inferred the bTTD in at least three different ways, providing three different perspectives on the importance of rainfall. First, isotope-based hydrograph separation and hydrochemical-based end-member mixing analysis (IHS/EMMA) use tracers to estimate a categorical bTTD by separating hydrographs into event (i.e., younger) and pre-event (i.e., older) contributions (Buttle, 1994; Hooper, Christophersen, and Peters, 1990). Second, models based on the assumption of a time-invariant bTTD use either inverse modeling and conservative tracer data (e.g., McGuire and McDonnell, 2006) or particle tracking in a steady-state distributed model (e.g., Pollock, 1988) to estimate an emergent steady-state bTTD



even when the true bTTD is known to be unsteady. Third, time-varying bTTD models use advanced approaches to inverse modeling (e.g., Rinaldo et al., 2015) and flux tracking (e.g., van der Velde et al., 2012; Davies et al., 2011) to make consecutive estimates of the time-varying bTTD averaged over shorter periods (i.e., hours to days) that integrate information from the history of time-variable inputs.

Studies using IHS/EMMA indicate that the intensity and pattern of rainfall have a complex and site-specific influence on the amount of event water in runoff. As summarized in the recent literature review by Klaus and McDonnell (2013), larger storms tend to yield higher fractions of event water in storm runoff. At the same time, they report that, in some case studies, storm runoff during wet periods tends to have a smaller fraction of event water than does runoff during dry periods. IHS/EMMA has also been combined with field observations to make inferences about dominant hydrologic processes. Studies at some sites find that intense storm events generate overland flow and activate other fast flow pathways that mobilize large volumes of event water into streams (e.g., Burns et al., 2001; Buttle, Vonk, and Taylor, 1995). Conversely, other studies show that intense storms trigger pressure waves (Torres et al., 1998), groundwater ridging (Sklash and Farvolden, 1979), and other phenomena that mobilize large volumes of pre-event water (e.g., McGlynn, McDonnell, and Brammer, 2002).

Of the many studies employing time-invariant transit time models (e.g., sine-wave model, convolution approach) (McGuire and McDonnell, 2006), only a few have explored the relationship between rainfall and the shape of the bTTD; they find that increased rainfall is sometimes (but not always) associated with faster transit times. For example, Hrachowitz et al. (2009) and Hrachowitz et al. (2010) compared estimates of the time-invariant bTTD between years and between sites at several humid Scottish watersheds and found a significant, but sometimes highly scattered, negative correlation between mean rainfall and mean transit time. Engdahl and Maxwell (2015) estimated the time-invariant bTTD in a series of virtual watershed experiments and found that higher effective precipitation (mean rainfall minus evapotranspiration) yielded faster transit times, especially through the vadose zone. Other studies spanning larger geographies, however, suggest that mean rainfall may not be a good predictor of travel time. Maxwell et al. (2016) tracked particles through a surface-subsurface flow simulation to estimate the bTTD of six major North American watersheds.

They reported no association between the mean transit time and rainfall, though they did find that the variance of the bTTD was greater in wetter watersheds. Jasechko et al. (2016) estimated a time-invariant fraction of young water in 254 catchments around the world and found no significant correlation with annual rainfall. A common limitation of many of these studies is the sparse sampling of isotope concentrations that underlie the analysis.

More recently, researchers have applied time-varying transit time models to study the relationship between rainfall and bTTDs, although careful sensitivity analyses have not been done. As might be expected, these studies show that stream age tends to be inversely related to rainfall intensity (e.g., Rodhe, Nyberg, and Bishop, 1996; Morgenstern, Stewart, and Stenger, 2010; Heidbüchel et al., 2012; Hrachowitz et al., 2013; Peters, Burns, and Aulenbach, 2014; Benettin et al., 2015). More unexpectedly, several studies have found the bTTD to be approximately time-invariant in discharge-weighted time (e.g., Rodhe, Nyberg, and Bishop, 1996). With regard to rainfall pattern, Heidbüchel, Troch, and Lyon (2013) found that rainfall clustering seemed to have an important effect on mean transit times at two zero-order, arid, mountainous catchments. Specifically, more clustered rainfall produced faster and younger runoff, presumably due to the exceedance of certain storage thresholds.

This brief literature review shows that fundamental questions about the relationship between rainfall variability and bTTDs remain unanswered, to the detriment of applied hydrology. A better understanding of what drives the evolution of bTTDs has the potential to (1) reveal dominant hydrological processes within the watershed (McGuire and McDonnell, 2006; McDonnell et al., 2010) and (2) improve the performance of catchment water quality models (Wellen, Kamran-Disfani, and Arhonditsis, 2015; Hrachowitz et al., 2016). Shifts in the bTTD due to variability in rainfall intensity and pattern, which could increase due to climate change (Lau, Wu, and Kim, 2013), might significantly alter the rate of chemical flushing through watersheds and modify the chemistry of water discharged to streams. From a watershed management perspective, this could call into question analyses that extrapolate estimates of transit times or stream chemistry from one time period or location to another (e.g., Sanford and Pope, 2013; Morgenstern et al., 2015). Further, watershed management models that incorrectly capture the relationship between rainfall and the bTTD might be missing important watershed processes (McDonnell and Beven, 2014), and thus be prone to giving

**Table 2.1:** Notation used to represent the mean fraction of young water  $F(t)$  and rainfall  $J(t)$  at different timescales. Notation for  $Q(t)$  and  $ET(t)$  is similar to  $J(t)$ . The last column lists the number  $n$  of times the quantities were estimated in each of the two watersheds over the 10 year study period from 1999 to 2008.

<b>Notation</b>	<b>Averaging period</b>	<b><math>n</math> (99-08)</b>
$F_i(t), J_i(t)$	some period $i$	varies
$F_d(t), J_d(t)$	daily	3653
$F_m(t), J_m(t)$	monthly	120
$F_q(t), J_q(t)$	quarterly	40
$F_y(t), J_y(t)$	yearly	10
$F_{at}, J_{at}$	all time	1
$F_{aw}, J_{aw}$	all winters	1
$F_{as}, J_{as}$	all summers	1

wrong results, or right results for wrong reasons (Kirchner, 2006).

### 2.3.2 Key metrics and notation

Although we are concerned with the behavior of the entire bTTD, the present analysis focuses on a summary statistic, the fraction of water in discharge younger than 90 days. We focus on this fraction of young water (FYW) for the following reasons. First, summary statistics facilitate the interpretation and visualization of changes in the bTTD, which varies across two dimensions (time and age). Previous transit time studies have also employed summary statistics such as the mean age, median age, and FYW (e.g., Rodhe, Nyberg, and Bishop, 1996; Hrachowitz et al., 2010; Peters, Burns, and Aulenbach, 2014; Benettin et al., 2013; Engdahl and Maxwell, 2015; Jasechko et al., 2016). Second, some evidence suggests that estimates of the FYW may be more certain than other commonly used summary statistics. Kirchner (2015b) and Kirchner (2015a) showed that tracer-based estimates of the mean and median water age may be more vulnerable than the FYW to aggregation error caused by catchment heterogeneities. Third, the FYW is an inherently interesting age fraction to study because young water is a significant portion of the global runoff budget that rapidly transmits soluble contaminants from landscapes to streams (Jasechko et al., 2016). The 90 day window also captures and isolates seasonal shifts in stream age that might affect the phenology of aquatic ecosystems. Finally, in the results and supplemental materials, we provide evidence that

some of our main conclusions are not sensitive to the choice of threshold within the range of 60 to 120 days.

The cumulative bTTD and FYW can be written in terms of age  $T$  and time  $t$  as  $\overleftarrow{P}_Q(T, t)$  and  $\overleftarrow{P}_Q(T = 90 \text{ days}, t)$ , respectively. The notation  $F(t)$  refers to a time-series of the FYW. Additional subscript notation describing  $F(t)$  over different timescales is shown in Table 2.1. Note all "mean" or "averaged" values of the FYW are flow-weighted using the equation presented in Section 2.5.2, in order to capture the FYW for all flux over the time period of interest.

### 2.3.3 Research questions

The overall goal of this study is to develop and demonstrate a method to analyze the sensitivity of the FYW to rainfall variability at catchment scale. We apply our methods at two neighboring experimental catchments in Plynlimon, Wales using a unique, publicly-available data record. We try to answer three questions.

1. How do changes in the *average intensity* of rainfall  $J_i(t)$  affect  $F_i(t)$  over different timescales  $i$ ?  
As discussed above, previous work has established that more rainfall generally yields more young water, but the functional form of the relationship (i.e., slope, linearity) and its dependence on the timescale of analysis are unclear.
2. How do changes in the *pattern of daily rainfall* affect  $F_i(t)$  over different timescales  $i$ ? It is evident that a change in rainfall pattern alters the FYW over short timescales (e.g., daily), but what about longer timescales (e.g., yearly)? What kinds of pattern have the biggest effect and why? Previous research studying the aggregate influence of rainfall pattern on the bTTD over long periods is scant and equivocal.
3. How might projected changes in rainfall due to *climate change* affect  $F_i(t)$  over different timescales  $i$ ? Although many studies have examined potential effects of climate change on stream discharge (Nohara et al., 2006) and water quality (Whitehead et al., 2009), we are unaware of prior efforts to isolate the potential influence of climate change on catchment bTTDs.

Answering these questions required addressing several common methodological challenges to catchment-scale transit time modeling (McGuire and McDonnell, 2006). First, to avoid uncertainties related to short data records, we identified study sites with an unusually long record of environmental tracer data. Second, to make our results more generalizable, we analyzed two distinct watersheds and used modeling approaches that could be applied at other sites. Third, to enable sensitivity analysis, we devised an appropriate means of forward modeling  $F(t)$  by coupling three models together: a rainfall generator (Robinson and Sivapalan, 1997), a rainfall-runoff model (Kirchner, 2009), and the rank StorAge-Selection (rSAS) transit time model (Harman, 2015). We forced the coupled models with a range of weather conditions, including projected climate change storylines. Finally, to address the challenge of parameter uncertainty in transit time modeling, we used the Differential Evolution Adaptive Metropolis (DREAM) algorithm (Vrugt, 2016) to identify a distribution of probable rSAS parameter values, using a likelihood function justified by analysis of model residuals. Previous studies using similar transit time models have either not considered parameter uncertainty (e.g., Harman, 2015; Danesh-Yazdi et al., 2016) or accounted for it based on unspecified assumptions about the underlying likelihood function (e.g., Benettin et al., 2017).

The remainder of the paper introduces the Plynlimon study sites, describes our coupled modeling framework, describes the three phases of our sensitivity analysis, and presents the key findings and conclusions.

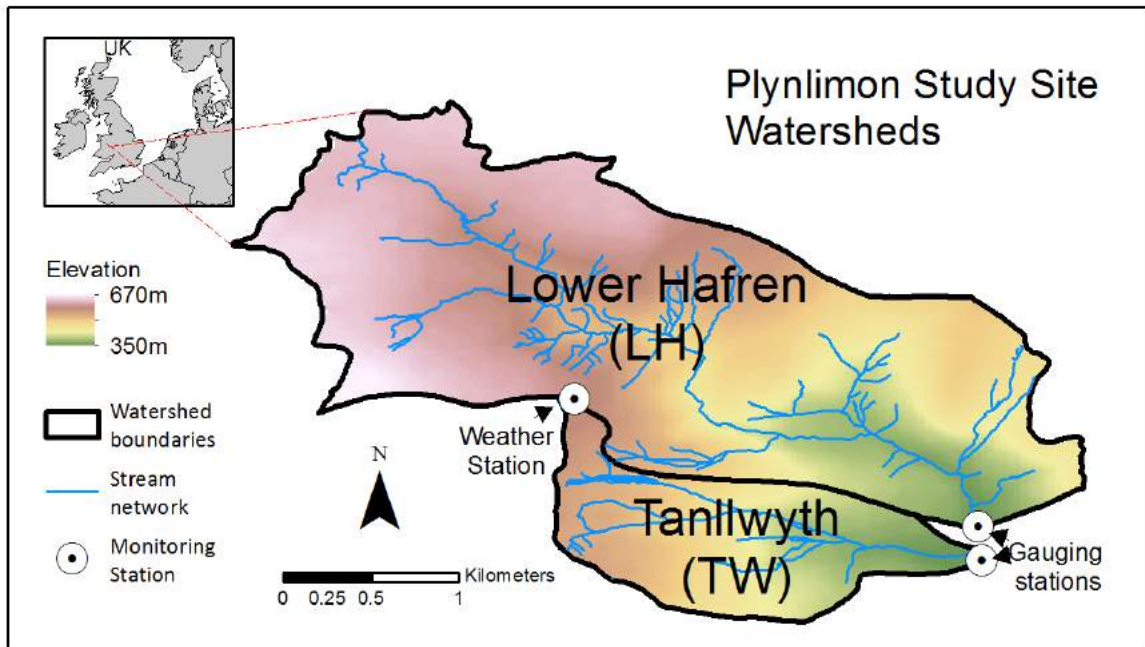
## 2.4 Study site description

Our work is based on data sets from the Lower Hafren (LH) and Tanllwyth (TW) headwater catchments of the Severn river basin in Plynlimon, Wales (see map in Figure 2.1 and catchment characteristics in Table 2.2). In addition to being extensively monitored and studied, the sites were for chosen for having similar climate but distinct physiography, hydrologic responsiveness, and transit time characteristics (one slower, one faster).

The catchments have been described elsewhere (Neal, 1997). In brief, the relatively large LH drains 3.5 km<sup>2</sup> across a 320 m elevation drop. The soils and vegetation of the upper LH are moorland, while the lower LH has a mixture of moorland, podzol, and gley soils with mostly Sitka Spruce vegetation. The TW site drains 0.5 km<sup>2</sup> across a 160 m elevation drop. Soils are primarily gley under

Sitka Spruce vegetation. Both sites are undeveloped but have a history of logging. No major felling occurred during the primary study period 1999-2008.

Over 30 years of monitoring data have been collected at the sites by the UK Center for Ecology, including rainfall, discharge, solar fluxes, temperature, and chloride concentration in the rainfall and discharge outlet (Neal, Kirchner, and Reynolds, 2013). The climate is wet with high annual rainfall (2857 mm) falling throughout the year and low evapotranspiration, yielding high runoff-ratios (0.78 in LH, 0.90 in TW). Besides atmospheric deposition and stream discharge, analyses of chloride hydrochemistry have not identified major sources or sinks for chloride at Plynlimon (Neal and Rosier, 1990) or in forested ecosystems generally (Svensson, Lovett, and Likens, 2012). Numerous studies have found that chloride acts as a conservative tracer at Plynlimon (e.g., Kirchner, Feng, and Neal, 2000; Benettin et al., 2015), which is the assumption made here. Input chloride concentrations were adjusted to account for unobserved occult deposition (as described in the next section). Using a steady-state transit time model calibrated with the chloride measurements, Kirchner, Feng, and Neal



**Figure 2.1:** Map showing the location of the Lower Hafren and Tanllwyth experimental catchments, with details about monitoring sites, topography, and the stream network.

**Table 2.2:** Characteristics of the study sites. Soils are moorland (M), podzol (P), or gley (G). Catchment area, soil types, and modeled estimate of the mean transit time (MTT) are from Kirchner, Feng, and Neal (2001).

<b>Watershed name</b>	<b>Area</b>	<b>Soil Types</b>	<b>Mean rainfall [mm/d]</b>	<b>Mean ET [mm/d]</b>	<b>Mean discharge [mm/d]</b>	<b>Runoff ratio</b>	<b>MTT [d]</b>
Lower Hafren (LH)	3.47	M/P/G	7.83	1.72	6.14	0.78	299.3
Tanllwyth (TW)	0.51	G	7.83	0.89	7.01	0.90	131.4

(2001) showed that the mean LH transit time is roughly two times larger than the mean TW transit time. This is possibly due to the relatively high permeability of LH podzolic soils, which can be associated with a larger volume of storage actively turning over and generating discharge.

### 2.4.1 Data description

Except where otherwise noted, daily time series of rainfall  $J(t)$ , discharge  $Q(t)$ , and actual evapotranspiration  $ET(t)$ , along with environmental tracer concentrations (chloride in rainwater  $C_J(t)$  and the stream  $C_Q(t)$ ) were gathered from a publicly-available data repository (Neal, Kirchner, and Reynolds, 2013) for the LH and TW catchments for the period 1993-2008 following methods described in Harman (2015) and summarized here. Daily bulk  $J(t)$  was measured from continuously open containers at Carren Wen near the edge of the LH watershed.  $Q(t)$  for both streams was recorded at 15-minute intervals and averaged daily at gauge locations shown in Figure 2.1.  $Q(t)$  data at TW was provided by the UK National River Flow Archive (O. Swain, personal communication). A small number of missing TW daily discharge measurements were filled by regressing a linear relationship with discharge at LH. Potential evapotranspiration ( $PET(t)$ ) was calculated from gap-filled, hourly weather station data at Carren Wen and Tanllwyth using the FAO Penman-Monteith equation for reference crops (Allen et al., 1998). To estimate  $ET(t)$ ,  $PET(t)$  was multiplied by a time-varying reduction factor  $\beta$ :

$$\beta(Q) = \min\left(1, \frac{PR_Q(Q)}{5}\right) \quad (2.1)$$

where  $PR_Q(Q)$  is the percentile rank of  $Q(t)$  relative to a reference period 1999-2008. The value of  $\beta(Q)$  begins to decline from 1 to 0 when  $Q$  drops below the 5th percentile, indicating the initiation

of water-limited conditions. The final equation for  $ET(t)$  was:

$$ET(t) = PET(t) * \beta * k_{ET} \tag{2.2}$$

where  $k_{ET}$  is an additional fitting parameter ( $k_{ET} = 1.74$  at LH,  $0.85$  at TW) set to obtain water mass balance over the reference period (i.e.,  $J_{at} = Q_{at} + ET_{at}$ ). More sophisticated approaches to  $ET(t)$  estimation under normal and dry conditions are possible (e.g., Monteith, 1965) but not explored here due in part to the limited role of  $ET(t)$  in the Plynlimon water budget.

Approximately weekly bulk measurements of  $C_J$  were collected from 1993-2008 (n=596). To facilitate modeling at daily timesteps, these were downscaled by assuming constant concentration during the bulk sampling period. Long data gaps were filled with the long-term average concentration for that month. To account for mist droplets and dry deposition missed by the collection, we assumed that occult chloride deposition from these sources was a fraction  $D$  of the total observed chloride in precipitation, where  $D$  was calibrated to achieve long-term chloride mass balance between inputs and outputs. Previous estimates of  $D$  have been on the order of 11-13% for moorland and 33-38% for forest (Neal and Kirchner, 2000; Wilkinson et al., 1997). Our calibrated value for LH, which is a combination of moorland and forest, was within the range of estimates for these two land-uses at 29%. The value for TW, which is mostly forest, was slightly above the range of previous estimates, at 43%. More complex accounting schemes for chloride inputs across space and time have been developed at these sites (Page et al., 2007) but are not necessary to obtain useful results (Harman, 2015). Grab samples of  $C_Q$  were also collected at approximately weekly intervals from the outlets of LH (n=805) and TW (n=797) over the same time period.

## 2.5 Modeling setup

Three different models were coupled, calibrated, and used to simulate the relationship between rainfall and the FYW: a rainfall-runoff model, a time-varying transit time model, and a rainfall generator. Each is described below.



### 2.5.1 Rainfall-runoff model

A catchment storage anomaly  $\Delta S(t)$  and discharge  $Q(t)$  were simulated using the catchment sensitivity function as described in Kirchner (2009). The method employs daily estimates of  $J(t)$ ,  $ET(t)$ , and a value of  $Q(t)$  at some initial time  $t_0$  to solve a reformulation of the water balance equation:

$$\frac{d(\ln Q(t))}{dt} = g(Q) \left( \frac{J(t) - ET(t)}{Q(t)} - 1 \right) \quad (2.3)$$

where  $g(Q)$  is the catchment sensitivity function and  $g(Q) = dQ/dS$ . The method assumes that  $S(t)$  has a one-to-one non-hysteric relationship with  $Q(t)$  (which is not generally true (Beven, 2006b) but was shown to be a reasonable approximation at Plynlimon (Kirchner, 2009)). Using the parameterization of  $g(Q)$  described below, Equation (2.3) was solved using the open source numerical integration solver "odeint" in SciPy v0.17.1 (Jones, Oliphant, and Peterson, 2014) to estimate  $Q(t)$ . The solver took  $PET(t)$  as input and converted it to  $ET(t)$  using equations (2.1) and (2.2) and the concurrent estimate of  $Q(t)$ .

The catchment sensitivity function  $g(Q)$  was parameterized using

$$\ln(g) = c_1 + c_2 * \ln(Q) + c_3 * (\ln(Q))^2 \quad (2.4)$$

with initial estimates of fitting parameters  $c_1$ ,  $c_2$ , and  $c_3$  determined from graphical techniques described in Kirchner (2009) for the period 1999-2008. These graphical estimates seeded an automated search algorithm (SciPy v0.17.1) that minimized the root mean square error (RMSE) between observations and simulations of  $Q$  derived from Equation (2.3). After obtaining the final parameters, an out-of-sample validation was performed for 1993-1998.

In both catchments, the calibrated value of  $g(Q)$  was used to calculate  $\Delta S(t)$ , the variation in catchment storage relative to an arbitrary datum (since the absolute storage is not known):

$$\Delta S(t) = \int_{Q(t_0)}^{Q(t)} \frac{1}{g(Q)} dQ - S_{ref} \quad (2.5)$$

where  $t_0$  is the initial condition for time and the parameter  $S_{ref}$  was chosen to ensure the mean of  $\Delta S(t)$  from 1999-2008 was zero.

### 2.5.2 rSAS transit time model

The rSAS model was used to estimate the time-varying bTTD and the FYW. The model was first described and applied by Harman (2015) in the LH watershed, and has since been used in a number of hydrologic systems (e.g., Kim et al., 2016; Harman, Ward, and Ball, 2016; Danesh-Yazdi et al., 2016). It is one in the family of StorAge Selection modeling frameworks (Botter, Bertuzzo, and Rinaldo, 2011; van der Velde et al., 2012; Rinaldo et al., 2015).

Briefly, the rSAS model assumes the catchment can be treated as a single control volume with a total storage  $S(t)$ , an influx of precipitation  $J(t)$  with a constant spatial distribution, and outfluxes  $ET(t)$  and  $Q(t)$ . The fluxes are inputs to the model derived from observations or other model simulations. At any time  $t$  the age of water in each outflux and storage can be represented by the cumulative transit time distributions  $P_Q(T, t)$  and  $P_{ET}(T, t)$  and the residence time distribution  $P_S(T, t)$ . The evolution of these age distributions is described by a reformulation of the "master equation" of Botter, Bertuzzo, and Rinaldo (2011) expressing continuity of water age and mass at the catchment scale (Harman, 2015):

$$\frac{\partial S_T(T, t)}{\partial t} = J(t) - Q(t)\overleftarrow{P}_Q(T, t) - ET(t)\overleftarrow{P}_{ET}(T, t) - \frac{\partial S_T(T, t)}{\partial T} \quad (2.6)$$

where  $S_T(T, t)$  is age-ranked storage:

$$S_T(T, t) = S(t)P_S(T, t) \quad (2.7)$$

with all precipitation assumed to have age zero. The terms on the right-hand side of Equation (2.6) capture the four ways the age of water in storage can change: water can enter as precipitation with age zero, water can leave as discharge or evapotranspiration with a distribution of ages, or water can age in place.

$\overleftarrow{P}_Q(T, t)$  and  $\overleftarrow{P}_{ET}(T, t)$  have a complex relationship with the history of fluxes that may be difficult to parameterize. To address this, it has been found convenient to re-express them as an equivalent CDF of age-ranked storage  $S_T(T, t)$  (Harman, 2015):

$$\Omega_Q(S_T, t) = \overleftarrow{P}_Q(T, t) \quad (2.8)$$

$$\Omega_{ET}(S_T, t) = \overleftarrow{P}_{ET}(T, t) \quad (2.9)$$

where  $\Omega_Q(S_T, t)$  and  $\Omega_{ET}(S_T, t)$  are the rSAS functions for the  $Q(t)$  and  $ET(t)$  fluxes, respectively. These rSAS functions describe the relationship between the age distribution of age-ranked water in storage and the age distribution of water in each outflux. This transformation between  $\Omega(S_T, t)$  and  $\overleftarrow{P}(T, t)$  is possible because of the one-to-one mapping between  $T$  and  $S_T$ .

Given time-series of  $J(t)$ ,  $Q(t)$ ,  $ET(t)$ ,  $\Omega_Q(S_T, t)$ ,  $\Omega_{ET}(S_T, t)$ , and the initial distribution of ages in storage  $S_T(T, t = 0)$  (determined here with a six-year warm-up period), the rSAS model uses Equations 2.6, 2.7, 2.8, and 2.9 to solve for the four unknowns:  $S_T(T, t)$ ,  $P_s(T, t)$ ,  $\overleftarrow{P}_Q(T, t)$ , and  $\overleftarrow{P}_{ET}(T, t)$ . The system of equations is solved numerically in Python 2.7 for daily timesteps following methods described by Harman (2015). The value of  $F(t)$  is determined as  $\overleftarrow{P}_Q(T = 90 \text{ days}, t)$ , and the flow-weighted mean over different timescales can be calculated using:

$$F_i(t) = \frac{\sum_{\forall t \in [t_i, t_{i+1}]} Q(t) \overleftarrow{P}_Q(T = 90 \text{ days}, t)}{\sum_{\forall t \in [t_i, t_{i+1}]} Q(t)} \quad (2.10)$$

where  $i$  is a period of averaging (e.g., yearly) within the interval  $t_i$  and  $t_{i+1}$ .

Finally, the rSAS model can be used to simulate movement of conservative tracers through the watershed. A spatially uniform concentration of conservative tracer in the rainfall  $C_J(t)$  can be convolved with the PDF of the bTTD ( $\overleftarrow{p}_Q(T, t)$ ) to estimate the output concentration in the stream (Barnes and Bonell, 1996; Rinaldo et al., 2011)

$$C_Q(t) = \int_0^\infty C_J(t - \tau) \overleftarrow{p}_Q(\tau, t) d\tau \quad (2.11)$$

with an additional adjustment made for the evapoconcentration of chloride (Harman, 2015).

### 2.5.2.1 Previous rSAS calibration and the inverse storage effect

Harman (2015) used conservative chloride tracer data from 1989-2008 to calibrate an rSAS model at the LH watershed. That work identified two cases of rSAS functions that did similarly well at reproducing the chloride stream concentrations. The first case exhibited an "inverse storage effect",

meaning that when storage was high the rSAS function for discharge shifted to increase the proportion of water coming from younger age-ranked storage. The second case had no inverse storage effect (i.e., the rSAS functions were time invariant) and performed just slightly worse than the first case. It was not clear if the difference between the two cases was significant or due to one or more potential sources of model uncertainty, including uncertainty in model inputs, the rSAS model structure, or the rSAS model parameterization.

This ambiguity in the rSAS calibration is problematic because the presence or absence of time-variability in the rSAS functions could significantly affect the relationship between rainfall pattern and the young water fraction. Kim et al. (2016) showed that, if the rSAS function  $\Omega_Q(S_T)$  is time-invariant, the bTTD should be approximately constant in discharge-weighted time. Under certain conditions, this implies that  $F(t)$  will be determined by the volume of discharge or rainfall over the prior 90 days, irrespective of its pattern. A proof of this (perhaps surprising) point is provided in Appendix A. By contrast, if the rSAS function is time-varying, the influence of rainfall pattern on the young water fraction could be more complex.

### 2.5.2.2 Diagnosing an inverse storage effect

To diagnose an inverse storage effect with more rigor, we parameterized the rSAS functions in a slightly different way from Harman (2015). The chosen form isolates the storage effect in a simple parameter  $\lambda$ , whose value distinguishes between the no storage ( $\lambda = 0$ ) and inverse storage ( $\lambda < 0$ ) cases.

With the approach applied here, the rSAS model requires estimation of four parameters in each watershed:  $\alpha$ ,  $\lambda$ ,  $S_0$ , and  $S_{ET}$ . The rSAS function for  $ET$  is defined as a time-invariant uniform distribution over a fixed volume  $S_{ET}$  of the youngest water in storage:

$$\Omega_{ET}(S_T) = \begin{cases} \frac{S_T}{S_{ET}}, & S_T \in [0, S_{ET}] \\ 1, & S_T > S_{ET} \end{cases} \quad (2.12)$$

The rSAS function for  $Q$  is defined as a storage-dependent gamma distribution with shape parameter  $\alpha$  and scale parameter  $S_Q(t)$ :

$$\Omega_Q(S_T, t) = \frac{\gamma(\alpha, \frac{S_T}{S_Q(t)})}{\Gamma(\alpha)} \quad (2.13)$$

$$S_Q(t) = \max(1 \text{ mm}, S_0 + \lambda * \Delta S(t)) \quad (2.14)$$

where  $\gamma(\alpha, \frac{S_T}{S_Q(t)})$  is the incomplete gamma function (Abramowitz and Stegun, 1964) and  $\Gamma(\alpha)$  is the gamma function. According to Equation (2.14),  $S_Q(t)$  varies with time depending on the catchment storage anomaly  $\Delta S(t)$  (derived in Section 2.5.1) and additional fitting parameters  $S_0$  and  $\lambda$ . If  $\lambda = 0$ , the scale parameter  $S_Q$  is constant and  $\Omega_Q(S_T, t)$  is time-invariant (i.e., no storage effect). When  $\lambda < 0$  then  $\Omega_Q(S_T, t)$  is time-varying and, as  $\Delta S(t)$  increases, the median and other quintiles of the distribution shift left to increase the fraction of discharge from smaller  $S_T$  (i.e., an inverse storage effect). The parameters of the gamma distribution must be positive, so  $S_Q(t)$  was not allowed to drop below 1 mm.

In order to test the hypothesis that  $\lambda < 0$  in each watershed, a probabilistic distribution of likely parameter combinations in each of the two watersheds was found using the Differential Evolution Adaptive Metropolis (DREAM) algorithm (Vrugt, 2016). DREAM applied a Markov chain Monte Carlo (MCMC) simulation to estimate the posterior probability of rSAS parameter combinations for the period 1999-2008, following a six year model warm-up. An iterative analysis of model residuals following Schoups and Vrugt (2010) was used to select a likelihood function. This approach has performed well for parameter identification in at least some cases (Schoups and Vrugt, 2010). It assumes, however, that model error is aleatory and does not account for epistemic uncertainty, which can result in overconfidence in parameter estimation (Beven, 2016; Vrugt, 2016).

The likelihood function used in the final analysis assumed residuals between simulated and observed  $C_Q(t)$  follow a heteroskedastic, Gaussian, first-order autoregressive process (see details in the supporting information (SI)). The standard deviation of errors was found to be a linear function of stream chloride concentration. The autoregressive process was needed to account for autocorrelation in the residuals, which was unsurprising given the dependence of  $C_Q(t)$  on the history of model inputs. In addition to generating a most likely parameter set, DREAM was used to generate a 95th percentile confidence interval of parameter sets, which provided a measure of parameter uncertainty

in our simulation results. The MCMC simulation was run in MATLAB on the Maryland Advanced Super Computing Center. All chains converged after 12,800 parameter trials, and an additional 3,200 chain links were run to reveal the distribution of likely parameter values.

### 2.5.3 Rainfall generator

To be able to study the FYW under a range of hypothetical rainfall scenarios, a rainfall generator was used to create daily rainfall time-series with statistical properties that matched or systematically differed from historic rainfall. Previous work has demonstrated the usefulness of rainfall generators for studying rainfall-response in the Plynlimon watershed (Cameron et al., 1999; Cameron et al., 2000). In this study, the rainfall patterns were used to force the coupled rainfall-runoff model and transit time model, as will be described in the methods section 2.6.2.

The rainfall generator algorithm described in Robinson and Sivapalan (1997) was chosen for its ability to isolate and adjust specific aspects of rainfall variability at both the seasonal and event scale. In brief, the model constructs a record of daily rainfall intensities in three steps. First, a series of alternating "storm" and "interstorm" durations are generated over the desired time period. The duration of each storm and interstorm is randomly sampled from a shifted exponential distribution with a seasonally varying mean. Second, the average intensity of each storm period is determined by sampling from a gamma distribution, with the expected value and coefficient of variation defined as power functions of the storm duration. Third, the normalized mass curve for each storm period – which determines the within-storm rainfall distribution – is randomly generated from repeat sampling of a beta PDF. In all three steps, the rainfall generator parameters were optimized using automated routines to match the statistical properties of daily rainfall observed at the study site from 1983-2008.

Unlike rain generators such as Cameron et al. (1999), the generator used here was not specifically designed to simulate extreme rainfall events, which could significantly bias young water estimates. To check the potential influence of extreme events on our simulations of water age, we generated an ensemble of 100 synthetic rainfall time-series with the same duration (16 years) as rainfall observed during the historic period 1993-2008 and used them to force the calibrated rainfall-runoff model and rSAS model.  $ET$  was kept constant at rates observed during the historic period. CDFs of key indicators (i.e., rainfall intensity, discharge, young water fraction) for different periods (i.e., all time,

all winters, all summers) were derived using data from both the historic period and the ensemble of synthetic rainfall records. The results were compared, with the expectation that CDFs from the historic period would lie within the 95th percentile interval of CDFs from the synthetic ensemble.

## 2.6 Sensitivity analysis methods

Using the data and models described above, we conducted three phases of sensitivity analysis and virtual experimentation at our study sites.

### 2.6.1 Phase 1 methods: historic analysis

The goal of phase 1 was to hindcast the FYW from 1999-2008 and explore the influence of mean rainfall intensity on the FYW over different timescales. The rainfall-runoff model was used to estimate daily values of  $\Delta S(t)$  with Equation (2.5) and the observed  $Q(t)$ . Daily observations of  $J(t)$ ,  $Q(t)$ , and  $ET(t)$  were fed into the rSAS transport model to simulate daily estimates of  $F(t)$ . The models were spun up for six years (1993-1998) prior to the period of analysis.  $F_i(t)$  was estimated with Equation (2.10) and compared with  $J_i(t)$  over daily, monthly, quarterly, and yearly timescales. Note that this phase did not use the rainfall generator, and the rainfall-runoff model was only used to estimate  $\Delta S(t)$  from the observed  $Q(t)$ ; it was not used to simulate  $Q(t)$ .

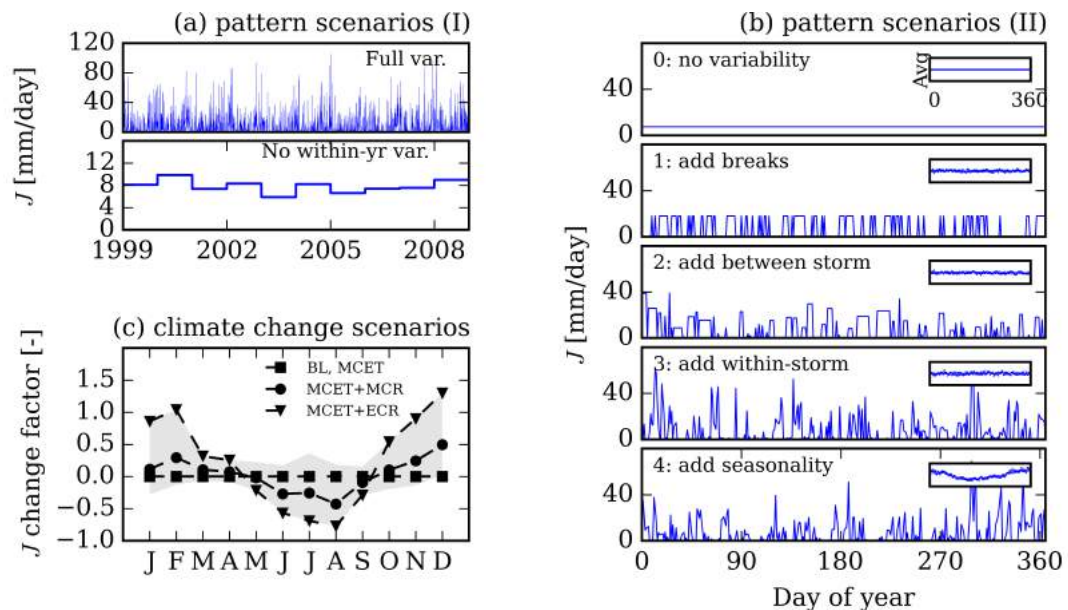
### 2.6.2 Phase 2 methods: pattern analysis

The goal of phase 2 was to test how changes in daily rainfall pattern influence the FYW over different timescales, when the long-term mean rainfall intensity does not change. The analysis was split into three virtual experiments, described below. To isolate the effect of rainfall pattern, only rainfall and discharge were simulated (i.e.,  $ET(t)$  was assumed to be zero).

The first experiment compared simulations of the FYW with and without intra-annual variability in the rainfall. The rainfall-runoff and rSAS models were forced with (1) the observed daily rainfall record from 1993-2008 and (2) the observed rainfall with daily rainfall replaced by the average yearly rainfall  $J_y(t)$  for that year (see Figure 2.2a). The effect of changes in  $J_y(t)$  on  $F_y(t)$  was compared

for the two cases during the years 1999-2008. Any differences could be attributed to intra-annual rainfall variability.

The second experiment compared simulations with identical mean rainfall intensity but increasing modes of rainfall variability, to isolate which kinds of variability have the greatest influence on the long-term mean FYW. To this end, we used the rainfall generator to create the set of 100-year rainfall patterns shown in Figure 2.2b. The base pattern has no variability; subsequent patterns add individual storm events, between-storm variability, within-storm variability, and finally seasonality. This final pattern represents the full range of rainfall variability at our study site. The synthetic rainfall patterns were used to force the runoff model to estimate daily  $Q(t)$  and  $\Delta S(t)$ , which were in



**Figure 2.2:** Rainfall scenarios used in the sensitivity analysis. Panel a shows the observed rainfall from 1999 to 2008 (top sub-panel) and the yearly average rainfall (bottom sub-panel). Panel b shows synthetic rainfall patterns with incrementally increasing variability. A representative 360 day period is shown. To show seasonal trends, the insets depict the average simulated rainfall for each day of the year, averaged over 300 years. The bottom sub-panel represents the full variability of the observed rainfall in our study sites. Panel c shows the rainfall change factors derived from the UKCP09 model (Murphy et al., 2007) downloaded at <http://ukclimateprojections.metoffice.gov.uk> for each of the four climate change storylines described in the text. The gray shading shows the 90th percentile range of change factor values.



turn used to force the rSAS model to estimate and compare the mean FYW for the entire period ( $F_{at}$ ), winter only ( $F_{aw}$ ), and summer only ( $F_{as}$ ).

The third experiment explored the effect of daily rainfall patterns on the long-term FYW in a more controlled manner. Several hypothetical series of repeating 10-day rainfall patterns were created to test the effect of changes in (1) storm breaks, (2) within-storm variation, (3) between-storm variation, and (4) seasonal variation on  $F_{at}$ . The patterns are presented in Section 2.7.3. Each repeating 10-day pattern was used to force the runoff model and rSAS model calibrated for the LH watershed. When  $F(t)$  reached periodic steady state,  $F_{at}$  was calculated and compared across runs. The rSAS model time step was reduced from 24 to 0.5 hours to capture sub-daily behavior.

### 2.6.3 Phase 3 methods: climate change analysis

Phase 3 explored how plausible changes in rainfall and evapotranspiration due to climate change might alter the long-term mean FYW during different times of year. The coupled modeling framework was run under a baseline storyline and three plausible future climatic storylines (shown in Figure 2.2c), developed as follows. The "baseline" (BL) storyline used  $J(t)$  and  $PET(t)$  from data observed in 1993-2008 and applied in the phase 1 analysis (see Section 2.6.1).

The three future storylines were developed using the "delta change" downscaling method described in Hay, Markstrom, and Ward-Garrison (2011). This method applies change factors to the data observed in 1993-2008 to produce a new climatic record representative of the period 2080-2095. Probability distributions of change factors for temperature, solar flux, relative humidity, and precipitation generated by the UKCP09 models (Murphy et al., 2007) were downloaded from the UK climate projections website (<http://ukclimateprojections.metoffice.gov.uk/>) for the overlying 1 km<sup>2</sup> grid cell for each month of the year. The probabilistic climate projections were based on a family of climate models from the Met Office Hadley Centre and other climate centers. The projections assume the high-emissions A1F1 development scenario with rapid, fossil fuel intensive economic growth (Nakicenovic et al., 2000). Figure 2.2c shows the precipitation change factors that were used to develop the three future storylines, described below.

The first future storyline simulated a median change in evapotranspiration (MCET), with no change in rainfall from baseline. The median projected monthly change factors for temperature, solar

radiation, and relative flux were applied to data from 1993-2008 to re-estimate  $PET(t)$  using the approach described in Section 2.4.1. Although projections for change in windspeed were not available, sensitivity analysis justified using the historic values (analysis not shown). Due to projected increases in temperature and other factors, the annual mean  $PET(t)$  increased 27%, with the largest increase (48%) in August.

The second future storyline simulated the same median change in evapotranspiration plus an additional median change in rainfall (MCET+MCR). To project  $J(t)$ , the median projected monthly rainfall change factors were applied to the baseline rainfall, with an up to 49% increase in monthly rainfall in winter and 43% decrease in monthly rainfall in summer.

The third future storyline simulated the median change in evapotranspiration plus an extreme change in rainfall (MCET+ECR). The extreme change in rainfall was calculated using the outer 5th percentile of change factor projections including an up to 130% increase and 77% decrease in monthly rainfall in winter and summer, respectively.

For each storyline, the projected values of  $J(t)$  and  $PET(t)$  were used to force the rainfall-runoff model and rSAS model with parameters calibrated from the historic period, yielding future projections for  $Q(t)$  and  $F(t)$ . These were used to estimate and compare the values of  $F_{at}$ ,  $F_{aw}$ , and  $F_{as}$ .

## 2.7 Results

### 2.7.1 Model calibration results

The calibration results for the rainfall-runoff model, the rSAS model, and the rainfall generator are discussed in turn.

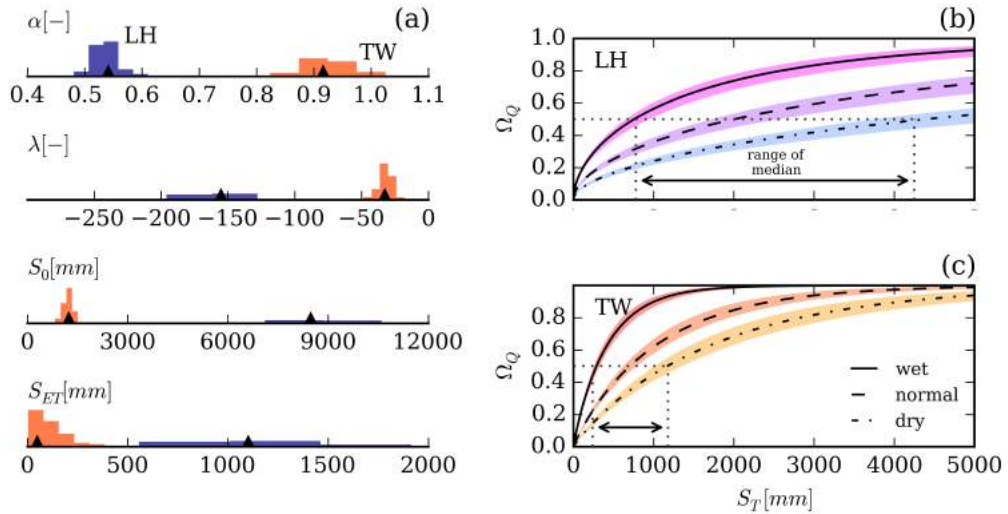
#### 2.7.1.1 Rainfall-runoff model calibration results

The calibrated rainfall-runoff model parameters and performance are given in Table 2.3. In both watersheds, the Nash-Sutcliffe efficiency (NSE) (Nash and Sutcliffe, 1970) for  $Q(t)$  and  $\log(Q(t))$  exceeded 0.82 during both calibration and validation time periods (see time-series comparison in Figure A1). The parameterized  $g(Q)$  was used to calculate the storage-discharge relationship in each watershed with Equation (2.5), which yielded results that were reasonable based on expectations

**Table 2.3:** Results from the calibration of the rainfall-runoff model.

	Optimal parameters			Calibration (1999-2008)		Validation (1999-2008)	
	$c_1$	$c_2$	$c_3$	NSE Q	NSE logQ	NSE Q	NSE logQ
LH	-4.29	2.04	-0.15	0.86	0.85	0.82	0.86
TW	-3.70	1.56	0.05	0.86	0.88	0.83	0.85

(see additional discussion in Section 2.8.3). During drier conditions ( $\Delta S < 0$ ) the discharge from LH tends to be higher than TW, which is consistent with LH having deeper soils contributing more to baseflow. Conversely, under wetter conditions ( $\Delta S > 0$ ) LH discharge tends to be lower than TW, which is consistent with TW having more impermeable soils that promote the formation of fast surficial flow pathways.



**Figure 2.3:** Calibration results for the rSAS model in both catchments. Panel a shows the histogram of parameter values and the most likely values (black triangles) given by DREAM. Panel b shows the parameterized shape of the rSAS function under 5th percentile (dry), 50th percentile (normal), and 95th percentile (wet) storage anomaly conditions. The lines bounded by color show the rSAS function based on the most likely parameter values, and the shaded areas show the 95th percentile confidence interval due to parameter uncertainty. The dotted lines illustrate the range of median  $S_T$  values in each watershed.

**Table 2.4:** Results from the calibration of the rSAS model with 95th percentile confidence intervals in brackets.

	rSAS parameters				Calibration (1999- 2008) NSE
	$\alpha[-]$	$\lambda[-]$	$S_0[\text{mm}]$	$S_{ET}[\text{mm}]$	
LH	0.55 [0.50, 0.59]	-155.3 [-228.0, -119.0]	8,472 [6,585, 12,202]	1,101 [664, 2,087]	0.69
TW	0.92 [0.85, 1.01]	-32.7 [-42.7,-23.5]	1,234 [953, 1,503]	48 [7, 326]	0.78

### 2.7.1.2 rSAS model calibration results

DREAM estimated the posterior distribution of rSAS parameters for the two watersheds (see Table 2.4, Figure 2.3a, and Figures A2-A5). The most likely parameter set in each watershed did well at predicting stream chloride concentrations over the ten year period (NSE = 0.69/0.78 in the LH/TW for the period 1999-2008, see Figures A6-A7). The model residuals for the most likely parameter set exhibited the characteristics (i.e., heteroskedastic, Gaussian, first-order autoregressive) that would be expected given our choice of likelihood function (see Figures A8-A9).

The results reject the hypothesis that either catchment has a time-invariant rSAS function. In both watersheds, the entire confidence interval of  $\lambda$  values is negative (see Table 2.4), indicating an inverse storage effect. Figure 2.3b shows the time-variability of the rSAS function at the 5th percentile storage anomaly (i.e., dry condition), the median, and the 95th percentile (i.e., wet condition).

The final parameters of  $\Omega_Q(S_T, t)$  show that discharge at LH generally comes from older age-rank storage than at TW, especially under dry conditions. At the median  $\Delta S$  (i.e., moderate wetness conditions) the median  $S_T$  for the  $\Omega_Q(S_T, t)$  distribution at LH is 2000 mm, meaning that 50% of discharge originates from the youngest 2000 mm of storage. By contrast, the median  $S_T$  at TW is less than half at 700 mm. The difference between the two watersheds is consistent with previous work showing longer transit times in LH compared to TW (Kirchner, Feng, and Neal, 2001). Under wet conditions, the median  $S_T$  is 760 mm and 280 mm for the LH and TW distributions, respectively, with a 480 mm gap. By contrast, the median  $S_T$  under dry conditions is 4400 mm and 1200 mm, with a much larger 3200 mm gap, suggesting that the LH is accessing a much larger storage volume when the watershed is dry.

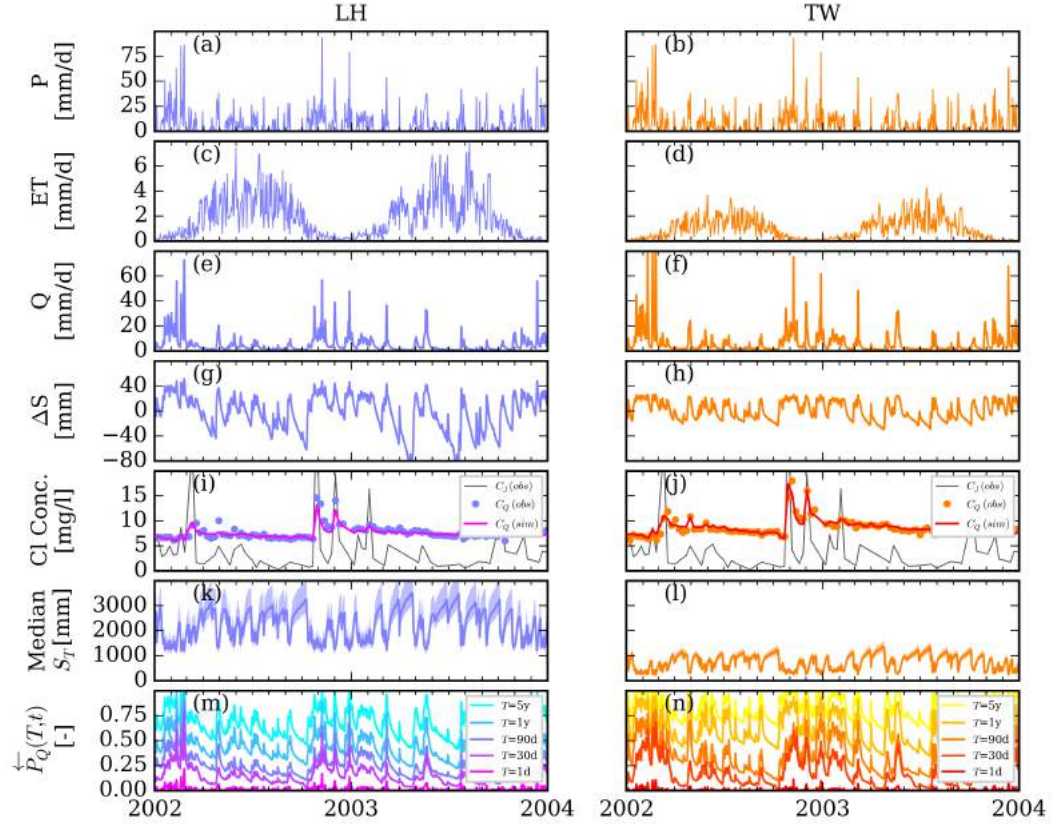
**Table 2.5:** Results from the calibration of the rainfall generator. Parameters with the *a* superscript are adjusted an additional +19% to -51% depending on the season. Parameters are defined in Robinson and Sivapalan (1997) and in the SI.

Parameter name	Value	Unit	Description
$\delta_r$	2.70	days	mean storm duration
$\alpha_r$	-0.74	days	amplitude of seasonal storm shift
$\tau_r$	74.10	days	phase of seasonal shift
$\delta_b$	3.92	days	mean break duration
$\alpha_b$	0.42	days	amplitude of seasonal break shift
$\tau_b$	43.80	days	phase of seasonal break shift
$a_1$	12.80 <sup>a</sup>	mm	coefficient 1 for expected storm intensity
$b_1$	0.23 <sup>a</sup>	-	coefficient 2 for expected storm intensity
$a_2$	0.63	-	coefficient 1 for expected storm variability
$b_2$	-0.83	-	coefficient 2 for expected storm variability
$\beta_1, \beta_2$	3.00	-	coefficient for within-storm variability
seasonality	1	-	with/without seasonality boolean
$p_{drizzle}$	0.31	-	probability of drizzle event
$\bar{J}$	7.83	mm/day	long-term mean rainfall

Finally, the parameter  $S_{ET}$ , which defines the uniform distribution  $\Omega_{ET}(S_T)$ , indicates that actual evapotranspiration is randomly sampled from the youngest 1,100 mm of storage at LH and 49 mm at TW. Although the 95% confidence intervals are relatively large, all results are lower than the median  $S_T$  in  $\Omega_Q(S_T, t)$  under moderate wetness conditions in each watershed. This indicates that evapotranspiration is sampled from generally younger age-rank storage than discharge.

### 2.7.1.3 Rainfall generator calibration results

The parameters of the rainfall generator model, shown in Table 2.5, were calibrated to reproduce observed rainfall variability (see calibration results in Figures A10-A12). These parameters were modified to produce the patterns shown in Figure 2.2b. The coupled model evaluation found that simulations using the ensemble of 100 synthetic rainfall patterns and a simulation using the observed rainfall gave comparable results in terms of the distribution of rainfall intensity, discharge, and young water fraction (see Figures A13-A14). These results demonstrate that the statistical properties of rainfall produced by the generator are reasonably similar to those of the historic period.



**Figure 2.4:** Values of key model inputs, outputs, and internal state variables for the coupled modeling framework for a representative 2-year period in the LH (left) and TW (right). Panels a-b, c-d, and e-f show  $J(t)$ ,  $ET(t)$ , and  $Q(t)$ , respectively. Panels g-h show the storage anomaly  $\Delta S(t)$  derived from the rainfall runoff model. Panels i-j show the observed concentration of chloride tracer in rainfall  $C_J(t)$  and the stream  $C_Q(t)$ , overlaid with the stream concentration simulated by rSAS with the most likely parameter set. Panels k-l show the median  $S_T$  of  $\Omega_Q(S_T, t)$ . The shaded area captures the 95th percentile confidence interval due to parameter uncertainty. Panels m-n show the simulated fraction of water younger than 1 day, 30 days, 90 days, 1 year, and 5 years assuming the most likely parameter set.

### 2.7.2 Phase 1 results: historic analysis

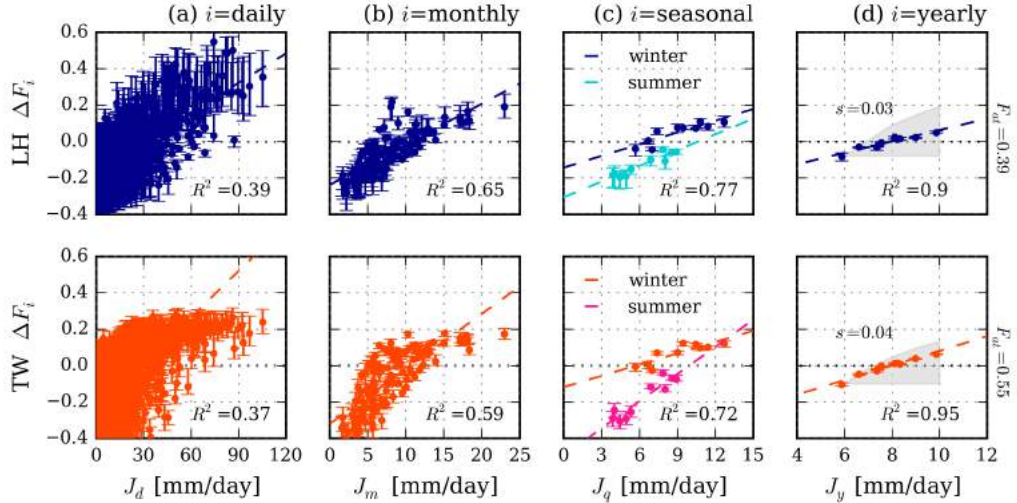
Figure 2.4 illustrates the coupled model inputs and outputs for two years (2002-2003) of the historic period. Precipitation is measured at the same station and therefore identical in both watersheds (Figure 2.4a-b). Rainfall is higher in winter and lower in summer. The  $ET$  peaks in summer and is higher in LH than TW due to differences in the fitting parameter  $k_{ET}$ , though the absolute values are still small compared to total rainfall (Figure 2.4c-d). The hydrograph and  $\Delta S(t)$  values appear slightly flashier in TW (Figure 2.4e-h), which is evidence of its steeper storage-discharge relationship. The chloride stream concentration at TW appears slightly more responsive to spikes in rainfall concentration (Figure 2.4i-j), which is evidence of having an rSAS function that selects more younger water from storage into discharge. As a result,  $F(t)$  is somewhat higher in TW than LH (see Figure 2.4m-n). The time-variability of  $F(t)$  is a consequence of both the changing composition of age-ranked storage and the changing shape of the rSAS functions. The  $F(t)$  seems highly correlated with other age fractions including the fraction of extremely young water (<1 day old) and fractions including much older water (e.g., < 5 years old) (see Figure 2.4m-n).

Based on a sampling of the last 3,200 chains of the DREAM simulation, the median  $F_{at}$  for LH and TW are 39% and 55%, respectively. We define a young water fraction anomaly  $\Delta F_i(t)$  for timescale  $i$  as the difference between  $F_i(t)$  and  $F_{at}$ . This metric can be directly compared between parameter sets (which each have a different  $F_{at}$ ) and used to construct 90% confidence intervals. At the daily timescale,  $\Delta F_d(t)$  tends to increase with  $J_d(t)$  (see Figure 2.5a), though there is substantial parameter uncertainty and scatter, presumably due to the influence of prior rainfall. At the monthly, quarterly, and yearly scales the correlation gets progressively stronger and less uncertain, culminating in a strong linear relationship ( $R^2 > 0.90$ ) between mean yearly rainfall  $J_y(t)$  and the yearly young water storage anomaly  $\Delta F_y(t)$  in both catchments (see Figure 2.5d). The value of  $\Delta F_y(t)$  increases by 13.0 and 16.8 percentage points from the driest to the wettest years in the LH and TW, respectively. Perhaps surprisingly, the marginal effect of higher rainfall on  $F_y(t)$  (i.e.,  $\partial F_y / \partial J_y$ ) is roughly constant and nearly the same in both watersheds: a 1 mm/day increase in  $J_y(t)$  results in a 3.1 and 4.0 percentage point increase in the  $F_y(t)$  in the LH and TW watershed, respectively.

### 2.7.3 Phase 2 results: pattern analysis

Our first analysis of rainfall pattern is shown in Figure 2.6, which compares the relationship between mean yearly rainfall  $J_y(t)$  and mean yearly FYW  $F_y(t)$  under conditions of (1) full rainfall variability and (2) no within-year rainfall variability. In the case of full rainfall variability in the LH watershed,  $J_y(t)$  varies from roughly 6 to 10 mm/day and  $F_y(t)$  varies from 32% – 45%. When within-year rainfall variability is removed, the value of  $F_y(t)$  drops by 4.4 – 8.6 percentage points in all years. At the same time, the marginal sensitivity of  $F_y(t)$  to changes in  $J_y(t)$  (i.e.,  $\partial F_y/\partial J_y$ ) remains roughly constant. This is evident by the fact that the "full variability" and "no within-year variability" curves run parallel in Figure 2.6. The same trends are evident in TW. These results suggest that within-year rainfall variability tends to increase  $F_y(t)$ , but that changes in  $F_y(t)$  from one year to the next are driven by differences in the mean intensity of rainfall, not the pattern.

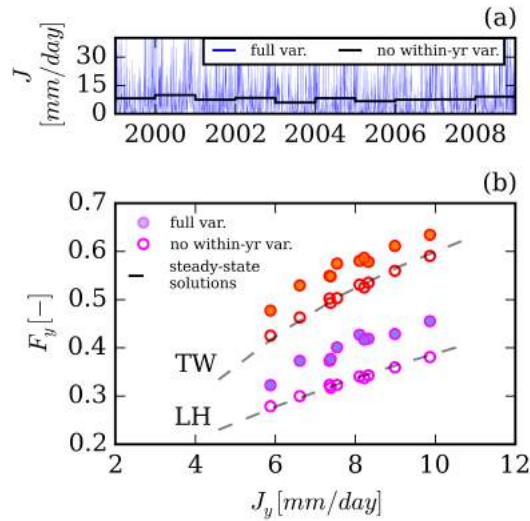
To better understand which kinds of rainfall pattern increase the long-term mean FYW, our second



**Figure 2.5:** Relationship between the young water fraction anomaly  $\Delta F_i(t)$  (defined in Section 2.7.2) and mean rainfall averaged over four timescales  $i$ : daily (panel a), monthly (panel b), seasonally (panel c), and yearly (panel d). The dashed line shows a line of best fit. The error bars show the 90th percentile confidence interval. The right-most panels indicate the slope  $s$  of the line of best fit and the median  $F_{at}$  for all parameter sets. The grey triangular regions are described in Section 2.8.2.

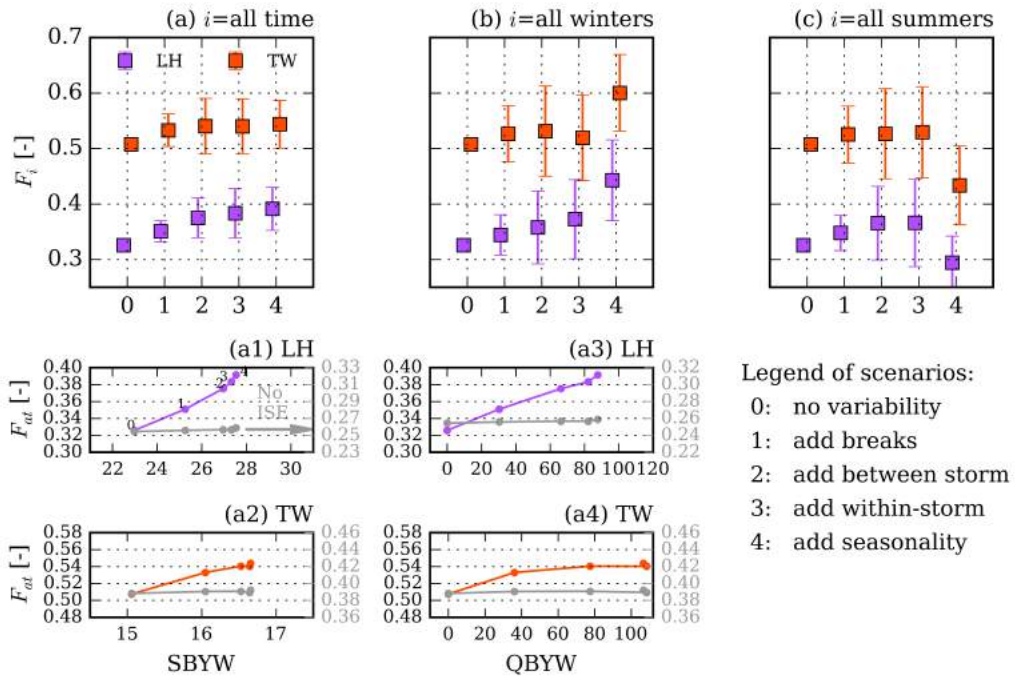


analysis forced the coupled models with 100-year synthetic rainfall records exhibiting increasing levels of rainfall variability (see results in Figure 2.7a-c). In both watersheds, rising variability in rainfall tended to increase both  $F_{at}$  and the variability of  $F_y(t)$  among years (indicated with the error bars). The largest increases occurred with the addition of storm breaks, followed by between- and within-storm variability. Although the addition of seasonality has a relatively small effect on  $F_{at}$ , it is the single most important factor affecting the long-term mean for winter  $F_{aw}$  and summer  $F_{as}$  (see Figures 2.7b-c). This is likely due to the redistribution of rainfall among the seasons, with higher amounts in winter and lower amounts in summer. Overall, the sensitivity to rainfall pattern was slightly greater in the LH.

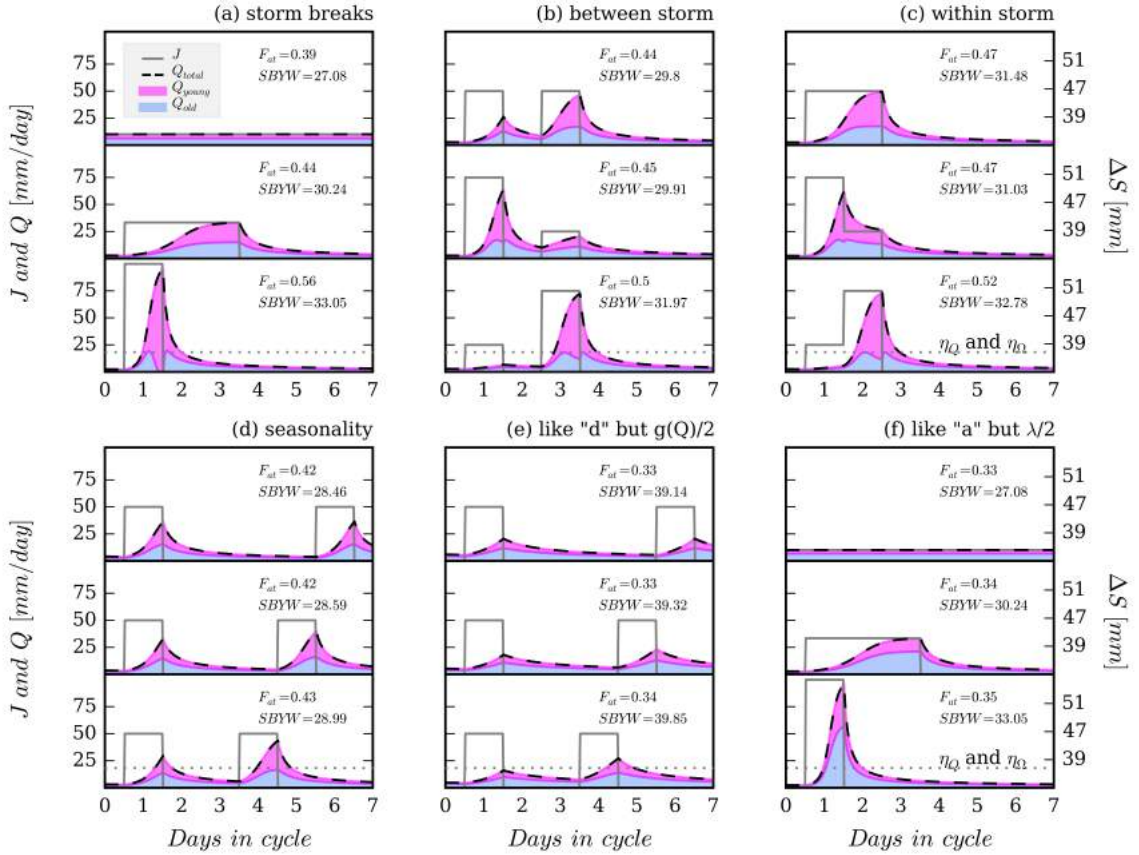


**Figure 2.6:** Panel a reproduces from Figure 2.2a the observed rainfall record marked "full variability" (blue line) and the yearly mean rainfall marked "no within-yr variability" (black line). Panel b shows the relationship between mean annual precipitation  $J_y(t)$  and modeled  $F_y(t)$  in LH and TW with (1) full variability (closed circles) and (2) no within-year rainfall variability (open circles). Note that rainfall variability increases  $F_y(t)$ , but changes in  $J_y(t)$  have roughly the same marginal effect on  $F_y(t)$  with or without within-year variability. The cases with no variability follow the steady-state solution (dashed line) described in Section 2.8.2.

Lastly, the influence of different rainfall patterns on the long-term mean FYW is demonstrated using repeating 10-day rainfall patterns shown in Figure 2.8. The series of three patterns in Figure 2.8a show that  $F_{at}$  increases significantly from 39% to 56% when the pattern transitions from having no storm variability to having storm breaks. The three patterns illustrate how large discharge events have a higher proportion of young water than smaller discharge events. This is evidence of the inverse storage effect, since discharge and storage (shown on the right y-axes in Figure 2.8) were assumed to be monotonically related. Similarly, panels 2.8b and 2.8c show how the  $F_{at}$  increases when between-storm and within-storm variability is introduced. Interestingly, the sequence of variability



**Figure 2.7:** The top row shows the effect of different modes of rainfall variability on the long-term mean FYW for all time (column a), all winters (panel b), and all summers (column c) in 100-year simulations. The error bars show the standard deviation of  $F_y(t)$ . The bottom two rows (panels a1-a4) show the calculated storage-based and discharge-based young water indices versus estimates of  $F_{at}$  taken from panel a. Each point represents one of the five scenarios, as labeled in panel a1. Increases in both indices are associated with increases in  $F_{at}$ , especially in LH. The light gray lines show  $F_{at}$  (right y-axis) for scenarios re-run with no inverse storage effect (i.e.,  $\lambda = 0$ ), as discussed in Sections 2.8.3 and 2.8.5.



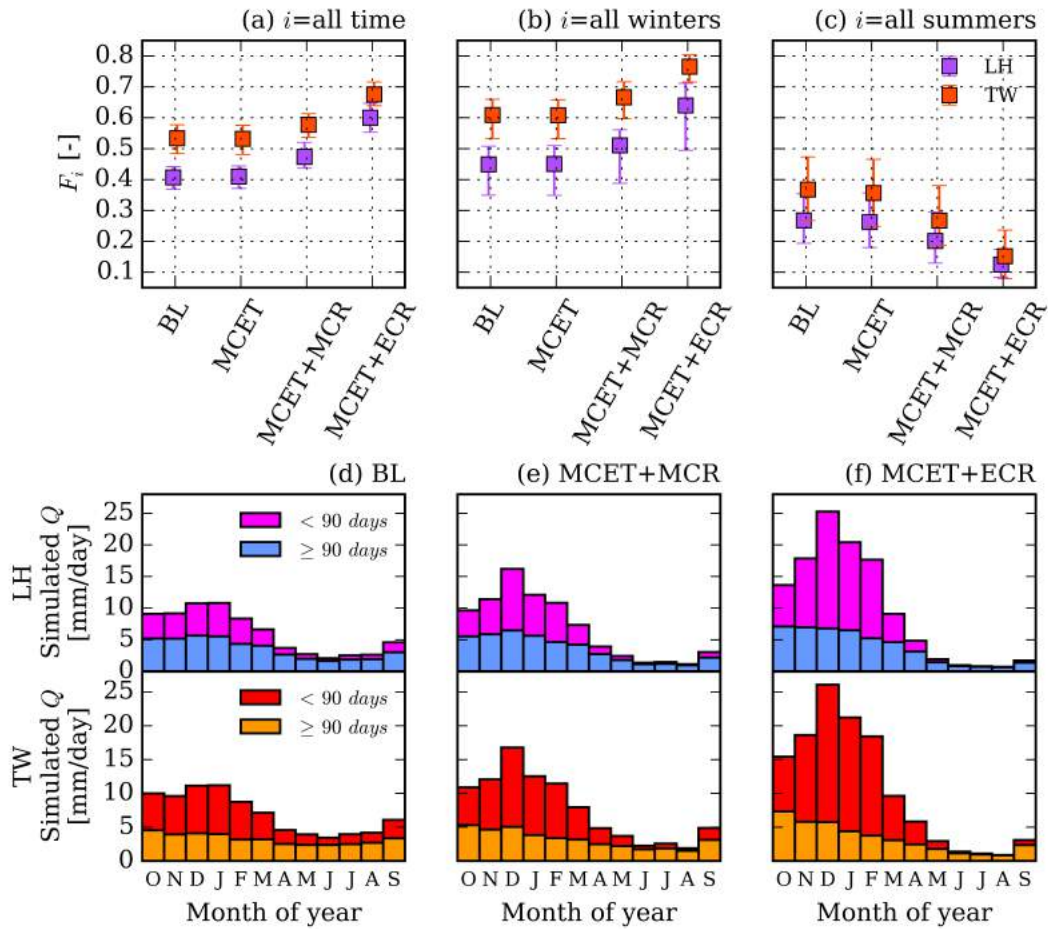
**Figure 2.8:** The interplay between different storm patterns, model parameters, and  $F_{at}$  in the LH. The first seven days of repeating ten day rainfall patterns are shown (solid gray lines) along with simulated discharge (dashed line) split into young (pink) and old (blue) components. No rain fell during the last three days of the repeating patterns, which are not shown. The mean young water fraction and SBYW are given for each pattern. Note that discharge has a one-to-one non-linear relationship with  $\Delta S(t)$ , shown on the right y-axis. The panels show a tendency for  $F_{at}$  to increase when storms get more clustered (panel a), storms are negatively skewed (panel b-c), and storms are less than 3 days apart (panel d). Panels also show how both  $Q(t)$  and  $F(t)$  increase sharply when  $\Delta S(t) > 35mm$ . This value is both  $\eta_Q$  and  $\eta_\Omega$  as labeled in the lower right sub-panels and defined in Section 2.8.3. The last panels show how changes in the storage-discharge relationship (panel e) and inverse storage effect (panel f) affect  $F_{at}$ .

is important. Small storms followed by large storms raise the  $F_{at}$  more than large storms followed by small storms, all else equal (see panel 2.8b). Similarly, negatively skewed storms – that is, those with the highest rainfall occurring at the end of the event – release more young water than positively skewed storms (see panel 2.8c). Finally, the influence of one storm on the next depends on the length of the break between them. Panels 2.8d shows that  $F_{at}$  begins to rise when storms are less than 3 days apart.

#### 2.7.4 Phase 3 results: climate change analysis

Figure 2.9a-c shows the influence of different climate change storylines on the long-term mean FYW. The BL and MCET storylines yielded almost the same  $F_{at}$ . The only difference in these scenarios is a projected increase in  $PET(t)$ , indicating that  $F_{at}$  is relatively insensitive to projected increases in  $PET(t)$  at these sites. In the other climate storylines featuring changes in rainfall, however, there is a significant increase in both  $F_{at}$  and  $F_{aw}$ . In the LH, for example, the  $F_{at}$  increases from 41% in the BL storyline to 60% under the MCET+ECR storyline. In contrast to the annual and winter trend, the summer shows a large drop in the  $F_{as}$  from 26(36)% in LH(TW) in the BL storyline to 12(15)% in the MCET+ECR storyline, presumably due to the large simulated decrease in summer precipitation. For both  $F_{aw}$  and  $F_{as}$ , the values at the two sites begin to overlap in the MCET+ECR storylines.

In addition to the young water fraction, we can also examine changes in the young water flux and total discharge. Figure 2.9d-f shows mean monthly discharge from the watersheds based on three projected storylines along with the mean monthly old water and young water fraction. The large increase in flux in the winter months in the MCET+MCR and MCET+ECR storylines is almost entirely young water. At the same time, the plot highlights a dramatic drop in young water during the summer months to near zero levels. The flux of water older than 90 days is relatively steady between the three storylines, suggesting that the old water fraction (presumably including baseflow) is more resilient to climatic shifts.



**Figure 2.9:** Results from the climate change simulations. Panels a-c show the simulated long term mean FYW averaged over three different time periods (i.e., all time  $F_{at}$ , all winter  $F_{aw}$ , and all summer  $F_{as}$ ) for each climate change storyline. Storyline acronyms are baseline (BL), median change in evapotranspiration (MCET), median change in rainfall (MCR), and extreme change in rainfall (ECR), as described in Section 2.6.3. The error bars show the 80th percentile range for  $F_y(t)$  over each 10 year modeling period. The bottom panels show the average monthly volume and age composition of water in discharge from the LH and TW for three of the four storylines.

## 2.8 Discussion

We discuss the relevance of our results to our research questions: how sensitive is the mean FYW to changes in the average intensity of rainfall over different timescales (subsections 2.8.1-2.8.2), how important is rainfall pattern (subsections 2.8.3-2.8.5), and what might be the effect of projected changes in rainfall due to climate change (subsection 2.8.6).

### 2.8.1 Diagnosis of the inverse storage effect and its implications

The positive diagnosis of an inverse storage effect at LH and TW provides useful insight into how these catchments respond to rainfall variability. First, it implies that flow through different flow pathways is rearranged and repartitioned during rain events and not simply accelerated (Kim et al., 2016), which informs our conceptual model of catchment processes. Second, as discussed in more detail below, it suggests that the time-averaged bTTD will be particularly sensitive to the pattern of rainfall. Third, it allows us to evaluate the performance of other models of these catchments based on their ability to simulate an inverse storage effect under different rainfall conditions. This may provide a stronger test of realism than solely fitting to the catchment hydrograph (McDonnell and Beven, 2014).

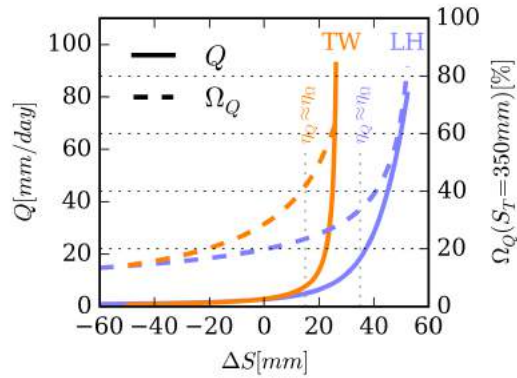
The inverse storage effect has been observed in a number of other hydrologic systems, but its global prevalence remains uncertain. Kim et al. (2016) used the rSAS model to identify an inverse storage effect in the bTTD of a soil lysimeter, and other groups have reported phenomena similar to an inverse storage effect in other catchments (Birkel et al., 2012; Soulsby et al., 2015; Hrachowitz et al., 2015; Benettin et al., 2017). The effect, however, does not seem to be universal. Several studies have measured catchment or hillslope bTTDs that are time-invariant in flow-weighted time (Roth et al., 1991; Rodhe, Nyberg, and Bishop, 1996), which is uncharacteristic of an inverse storage effect (Kim et al., 2016). Other studies have observed more event water discharged from storms following dry conditions (Klaus and McDonnell, 2013); if the catchment had an inverse storage effect, one would expect the opposite.

This work demonstrates a promising detection method that could be applied at other sites to test whether an inverse storage effect is the rule or the exception. A similar method was recently used by Benettin et al. (2017) to positively diagnose an inverse storage effect, albeit with a different

StorAge Selection function and parameterization. Although the rSAS functions at our study sites were best fit with gamma distributions, the approach of using Bayesian parameter identification to test whether  $\lambda = 0$  is more general and can be applied with any distribution having a parameter that scales the mean (e.g., exponential, weibull). An issue with Bayesian model identification, however, is the challenges raised in the ongoing debate on how to choose a likelihood function (Beven and Binley, 2014). Although our likelihood function is derived from first-order statistical principles based on a transparent analysis of model residuals, it cannot account for epistemic uncertainty and may be prone to over-conditioning of the likelihood surface (Vrugt, 2016; Beven, 2016), which could compromise conclusions about the value of  $\lambda$ . More sophisticated approaches to rSAS parameter estimation are being explored.

### 2.8.2 Variation in young water fraction explained by mean rainfall

From 1999-2008, the estimated mean FYW at LH and TW was very sensitive to changes in mean rainfall over various timescales. It was unsurprising to see this behavior over short timescales ( $< 1$  week), where young water discharge may be limited by supply. This sensitivity to mean rainfall



**Figure 2.10:** Illustration of thresholds in the relationship between storage anomaly and discharge (solid line, left axis) and storage anomaly and the fraction of water in discharge from the youngest 350 mm of catchment storage (dashed line, right axis). Both rise more sharply at high storage, indicating mobilization of both large volumes of water and large fraction of the youngest age-ranked water in storage. The dashed lines indicate the approximate, overlapping location of  $\eta_Q$  and  $\eta_\Omega$  for LH and TW.

persisted, however, over long time periods ( $\geq 1$  year), and explained much of the year-to-year variation in the mean FYW (see Figure 2.5). This does not imply that the FYW is insensitive to rainfall pattern, only that between-year variations in the total amount of rainfall had a larger effect on the FYW than between-year variations in rainfall pattern.

The catchments' sensitivity to mean rainfall should be considered when taking measurements of the young water fraction and, more generally, the bTTD. Many studies have used relatively short periods of observation ( $\leq 2$  years) to measure a "steady-state" bTTD (e.g., see review in McGuire and McDonnell, 2006). Our results suggest these measurements could have been substantially different if collected under other rainfall conditions. This source of variability is predictable but rarely acknowledged or quantified (though for exceptions see Hrachowitz et al. (2009) and Heimbüchel, Troch, and Lyon (2013)). This may partly explain the difficulty hydrologic scientists have had in finding catchment characteristics that predict observed transit times.

The highly linear relationship between annual rainfall and annual mean FYW (Figure 2.5) is a potentially useful hydrologic signature for model calibration and evaluation. The relationship can be plotted for any watershed with sufficient data, and is likely to lie between two conceptual end-members (shown as the area between the top and bottom of the gray triangles in Figure 2.5d). The bottom end-member (the horizontal base of the triangle) is the case where the bTTD is time-invariant and, for each year, all "base" rainfall (defined as the rainfall volume in the driest year) and "surplus" rainfall (defined as the rainfall volumes in excess of the base rainfall) is discharged with the same age fraction (chosen in Figure 2.5d as that of the driest year). The top end-member is the case where base rainfall still has the age fraction of the driest year, but all surplus rainfall is assumed to discharge as young water. The actual LH and TW curves lie between these two end-members, capturing the fact that surplus rainfall is more likely to discharge as young water but also continues to contribute to old water, even in the wettest years.

Finally, we observed that model runs without within-year rainfall variability had roughly the same sensitivity to changes in mean rainfall as runs with the full variability (see Figure 2.6). The results without within-year rainfall variability are comparable to steady-state solutions to Equation (2.6) (see equation 15 in Harman, 2015) assuming constant rainfall  $J_0$  and discharge  $Q_0$  (with  $J_0 = Q_0$ ) set equal to mean annual rainfall  $J_y$ . This is illustrated in Figure 2.6, where the set of quasi-steady-state



solutions for each year (derived from simple numerical integration) are shown to overlap our modeled solutions. These results point to a practical method for estimating the sensitivity of the mean annual young water fraction to changes in mean annual rainfall (i.e.,  $\partial F_y / \partial J_y$ ) based on the shape of the rSAS function alone. Conversely, if this sensitivity were determined experimentally, it could be used to constrain the shape of an unknown rSAS function.

To test the sensitivity of the results to the choice of a 90 day threshold used to define the FYW, we repeated the historic analyses shown in Figure 2.5 for the fraction of water less than 60 days and less than 120 days (see Figures A18 and A20). The overall trends were very similar, with a 1 mm/day increase in  $J_y(t)$  resulting in a 3-4 percentage point increase in the  $F_y(t)$  in both watersheds. A survey of other thresholds showed that, overall, the sensitivity of the yearly age fraction to changes in yearly rainfall seemed greatest for the middle age thresholds (30-120 days), and smallest for the very young (1 day) and very old (5 year) age thresholds (see Figures A15-A22).

### 2.8.3 Interplay between rainfall pattern and the inverse storage effect

Why do some rainfall patterns seem to produce more young water than others? In both catchments, rainfall pattern had a persistent influence on the mean fraction of young water in discharge, above and beyond the influence of rainfall amount. In particular, clustered rainfall produced more young water than steady rainfall (Figures 2.7 and 2.8a), and negatively skewed storms produced more young water than positively skewed storms (Figure 2.8c).

Sensitivity to rainfall pattern emerged from interactions between two catchment characteristics: (1) a dramatic rise in catchment discharge when storage anomaly exceeded some threshold  $\eta_Q$  and (2) a dramatic rise in the fraction of discharge from the youngest water in storage when storage anomaly exceeded some other threshold  $\eta_\Omega$ . These rise points are evident in Figure 2.8 and are further explored in Figure 2.10, which shows the storage-discharge relationship for each watershed as derived from Equation (2.5). These curves show a sharp increase in  $Q$  over a  $\Delta S$  threshold of  $\eta_Q \approx 35$  mm in LH and  $\eta_\Omega \approx 15$  mm in TW. The same figure shows the relationship between  $\Delta S$  and the fraction of discharge coming from the youngest 350 mm of storage (i.e.,  $\Omega_Q(S_T = 350\text{mm}, \Delta S)$ ), based on the most likely rSAS parameters listed in Table 2.4 and graphed in Figure 2.3. A  $S_T$  of 350 mm (equivalent to approximately 45 days of mean rainfall) was chosen to illustrate the behavior of

a volume of age-ranked storage that is likely to contain mostly young water. The positive slope of the  $\Delta S$  versus  $\Omega_Q$  curve is a manifestation of the inverse storage effect. In addition, the curve is concave up, with an acceleration in the fraction of water from the youngest 350 mm in storage above a threshold of  $\eta_\Omega \approx 35$  mm in LH and (less distinctly)  $\eta_\Omega \approx 15$  mm in TW. While there is no rigor to this analysis, it is curious that in both watersheds  $\eta_Q \approx \eta_\Omega$ . This may be considered unsurprising since the processes that control catchment celerity (and thus  $\eta_Q$ ) are inextricably linked to the those that control velocity (and thus  $\eta_\Omega$ ). Although an investigation of the hydrologic processes responsible for these non-linear dynamics is beyond the scope of this work, threshold behavior and storage effects have been widely observed (McDonnell, 2003) and explained by many mechanisms (e.g., Tromp-van Meerveld and McDonnell, 2006).

Due to the apparent threshold-like storage dependence of both discharge and the fraction of discharge from young storage, rainfall patterns with heavy rain during high storage periods induce "hot moments" with a disproportionately large and young volume of discharge at our study sites. Given steady rainfall (e.g., Figure 2.8a),  $\Delta S$  remains relatively low and the thresholds  $\eta_Q$  and  $\eta_\Omega$  are not exceeded. Given more clustered rainfall,  $\Delta S$  crosses the thresholds and large volumes of water from the youngest volumes of storage are discharged (see the thresholds marked with the dashed line in the bottom sub-panels of Figure 2.8a). When a rainfall cluster is negatively skewed (Figure 2.8c),  $\Delta S$  approaches the threshold in advance of the storm peak, so the storm peak more quickly exceeds it, yielding larger volumes of young water.

To test the sensitivity of our results to this threshold behavior, we tried adjusting the model parameters to reduce the non-linear dynamics. When the catchment celerity was decreased by halving the catchment sensitivity function (see Figure 2.8e compared with Figure 2.8d), the hydrograph flattened, reducing opportunities to exceed storage thresholds. Likewise, when the inverse storage effect was tempered by halving the  $\lambda$  parameter (see Figure 2.8f compared with Figure 2.8a), the sensitivity to pattern nearly disappeared, and the old water fraction represented a much more significant portion of the storm peak. Using the more realistic synthetic rainfall patterns, we eliminated the inverse storage effect by setting  $\lambda$  to zero. With no inverse storage effect, changes in rainfall variability had no clear effect on  $F_{at}$  (see flat gray lines in Figure 2.7a1-a4, discussed further in Section 2.8.5). This result is consistent with the fact that, assuming idealized conditions and a

time-invariant rSAS function,  $F(t)$  should only depend on the average intensity and not the pattern of rainfall over the past 90 days (see proof in Appendix A).

Our results suggest that time-variability in the rSAS function (referred to here as "internal variability" as defined in Kim et al. (2016)) is an important control on if and how rainfall pattern affects the FYW in catchment discharge. At our sites, the internal variability resembles a non-linear inverse storage effect. At other sites, the rSAS modeling framework could identify other kinds of internal variability that interact differently with rainfall pattern. For example, as mentioned previously, several studies have observed large young water fractions following dry conditions (Klaus and McDonnell, 2013), which could be characterized by another, as yet undiscovered form of internal variability.

#### 2.8.4 Seasonality and timescale of the inverse storage effect

The addition of seasonality into the rainfall pattern had little influence on the long-term mean young water fraction (Figure 2.7a). This was unexpected given that the addition of seasonality causes significant changes including making the mean duration of winter storms (3.4 days) longer than summer storms (2.0 days), the mean break between winter storms (3.5 days) shorter than summer storms (4.3 days), and the overall intensity of winter storms higher than summer storms (see Table 2.5). Given the importance of rainfall pattern discussed in the previous section, why would the long-term mean young water fraction be so insensitive to the seasonal distribution of the size and timing of storms?

The reason likely hinges on the transience of the inverse storage effect at our study sites. During storm events, catchment storage surpasses the thresholds  $\eta_Q$  and  $\eta_\Omega$  such that both discharge and the fraction of discharge from the youngest storage quickly increase, as discussed in the previous section. After storms, these "hot moments" persist until the storage anomaly drops below  $\eta_Q$  and  $\eta_\Omega$ , taking some timescale  $\psi$  to do so. Given that storage is monotonically related to discharge, the value of  $\psi$  may be on the order of the timescale of hydrograph recession. If we define  $\psi$  as the time for discharge to drop by a factor of  $e$  evaluated at the mean catchment discharge  $\bar{Q}$ , which is equivalent to  $g(\bar{Q})^{-1}$  (see derivation in SI), then  $\psi$  equals 2.9 days in LH and 1.7 days in TW.

It follows that seasonality had little effect on the long-term mean FYW because the timescale representing the persistence of hot moments after storms ( $\approx 1.7 - 2.9$  days) is shorter than the average break between storms in both winter and summer ( $3.5 - 4.3$  days). From the perspective of the internal variability of the rSAS function, each storm acts roughly independently of the last. Even with the larger and more frequent winter storms, they are still sufficiently far apart for the inverse storage effect to relax to a baseline state before the next storm. This is illustrated in Figure 2.8d, where reducing the break between storms from 5 days to 3 days has almost no effect on the long-term average FYW;  $F_{at}$  does not increase until the storm break drops below 3 days.

An understanding of the timescale of a catchment's internal variability can help explain and predict how the young water fraction may respond to different rainfall patterns. At Plynlimon the timescale is tied to a fast hydrograph recession, but in other watersheds the drivers of internal variability may be more persistent and/or decoupled from discharge (e.g., Danesh-Yazdi et al., 2016), in which case changes in rainfall pattern due to seasonality could take on added importance.

### 2.8.5 Evaluation of rainfall pattern using a young water index

Is there a practical way to predict if changes in the pattern of daily rainfall will yield a higher  $F_{at}$ , holding mean rainfall equal? In this section we propose a storage-based young water (SBYW) and discharge-based young water (QBYW) index that measure the tendency for a rainfall pattern to yield more or less young water.

As reasoned in Section 2.8.3, our modeling suggests that rainfall variability tends to increase  $F_{at}$  when heavy rainfall coincides with high catchment storage. This phenomenon can be quantified in a storage-based young water index as the average storage anomaly weighted by precipitation:

$$SBYW = \frac{\sum_{\forall t} \Delta S(t) J(t)}{\sum_{\forall t} J(t)} \quad (2.15)$$

This weighted average can be expanded into two terms:

$$SBYW = \frac{cov(\Delta S, J)}{E[J]} + E[\Delta S] \quad (2.16)$$

where  $E[\ ]$  is the expected value and  $cov(\ )$  is the covariance. The first term on the right-hand side

will be larger when periods of high rainfall coincide with high storage. The second term will be larger for rainfall patterns that raise the average storage anomaly. Due to the assumed relationship between  $\Delta S$  and  $Q$  in these watersheds, we can construct an analogous discharge-based young water index:

$$QBYW = \frac{cov(Q, J)}{E[J]} + E[Q] \quad (2.17)$$

If the mean rainfall and discharge are constant and we are only interested in relative changes, this can be simplified:

$$QBYW = cov(Q, J) \quad (2.18)$$

such that the QBYW index is simply the covariance between rainfall and discharge.

The SBYW and QBYW indices were calculated for each 100-year run described in Section 2.6.2 using the rainfall patterns shown in Figure 2.2b. Figure 2.7a1-a4 shows a linear relationship in the two watersheds between both indices and the long-term mean young water fraction  $F_{at}$ . The absolute value of the index, however, was watershed specific, suggesting that comparisons are only meaningful within the same watershed. When the inverse storage effect is removed (gray points in Figure 2.7a1-a4),  $F_{at}$  becomes independent of rainfall pattern and the indices have no meaning. The indices were also calculated for each of the periodic rainfall patterns in Figure 2.8 (values written in the plots). Again, higher index values were associated with higher average fraction of young water.

To the extent that other watersheds have an inverse storage effect, the SBYW index could be used to estimate the effect of a change in rainfall pattern on the age of water in discharge, even when specific information about the rSAS function is not available. To the extent the inverse storage effect is a function of discharge (as at our study sites), the QBYW may be a practical alternative when estimates of catchment storage are unavailable. For example, the SBYW or QBYW index could be calculated for rainfall patterns derived from different downscaling methods, in order to assess which (if any) is likely to elicit a significantly higher young water fraction.

### 2.8.6 Water quality simulation under a changing climate

The projected changes in the amount and fraction of young water under projected climate storylines (see Figure 2.9) could significantly shift water chemistry at the study sites. Kirchner (2015b) found

that stream concentrations of  $NO_3^-$ ,  $Ca^{2+}$ , and  $Al^{3+}$  had strong correlations with the young water fraction at both LH and TW. The association with nitrate may be especially important, since nitrogen is an essential and sometimes limited nutrient in stream ecosystems. Under the MCET+ECR climate storyline, for example, the simulated summer mean young water fraction  $F_{as}$  dropped from 26% to 12% in the LH. Based on the positive correlation between  $NO_3^-$  and the FYW presented in Figure 18 of Kirchner (2015a), this shift in young water would be associated with a 31% drop in mean stream nitrate concentration, from 0.28 mg-N/L to 0.19 mg-N/L. The ecological effect of summer declines in concentration could be exacerbated by the summer declines in discharge shown in Figure 2.9d-f. Drops in both discharge and concentration could significantly reduce the  $NO_3^-$  mass flux. Conversely, during winter the nitrate concentration and flow rate would be expected to increase, which might shift the ecological balance in the other direction.

The question could be raised: how sensitive are our results to the choice of climate downscaling technique? Although we used the relatively simple delta-change method, many methods are possible. The modeling framework presented here could be used to directly compare simulations of the young water fraction given different downscaling algorithms. Although such a comparison is beyond the scope of this work, our sensitivity analyses suggest that the long-term average young water fraction will be most sensitive to the mean intensity of downscaled rainfall and to the representation of storm breaks. Finer details such as the representation of within-storm variability may be of secondary importance, especially in the TW watershed.

### 2.8.7 Study limitations

While our results provide new insight into the relationship between rainfall variability and transit times, we highlight here some important limitations. For the Phase 1 analysis of historic trends, the rSAS model was calibrated to weekly chloride grab samples but used to simulate daily behavior. While this is a potential source of error, Harman (2015) showed that similar modeling in the LH did reasonably well at simulating daily chloride observations collected from 1994-1997. For the Phase 2 and 3 analyses, the coupled modeling framework was extrapolated to timescales (e.g., sub-daily) and climatic conditions (e.g., the extreme change in rainfall storyline) that were outside the calibration conditions and thus not directly tested. In this light, these simulation results should be viewed

as evidence-based hypotheses – developed using current theories describing observed behavior – which require further evaluation as new information becomes available. Finally, several simplifying assumptions were used to estimate catchment inputs (e.g., ET, chloride deposition) which yielded good fits to data but could nonetheless be important sources of modeling error.

## 2.9 Conclusions

Motivated by prior work and using novel approaches developed for the purpose, this study used a long record of hydrometric and environmental tracer data from neighboring headwater catchments in Plynlimon, Wales to answer three questions about the effect of rainfall variability on the FYW in discharge.

The first question was: what has been the historic relationship between rainfall and the FYW over different timescales? At the daily time scale, the correlation between mean rainfall intensity and the FYW was relatively weak, presumably due to the influence of antecedent rainfall. At longer time scales, however, rainfall intensity explained almost all of the variability in the FYW; a 1 mm/day increase in annual mean rainfall was associated with a 3.1 to 4.0 percentage point increase in the annual mean FYW. The sensitivity was roughly the same in both watersheds, despite the fact that the FYW was overall higher in TW (with less permeable soils and shallow flow paths) than LH (with more permeable soils and both shallow and deep flow paths).

The second question was: what is the influence of rainfall pattern on the FYW over long timescales? Although changes in average rainfall intensity seemed to dominate between-year differences in the mean annual FYW, the pattern of rainfall variability was also important. In particular, synthetic rainfall scenarios showed that precipitation patterns with heavy rainfall during periods of high storage (e.g., clustered rainfall events, negatively skewed rainfall events) resulted in a higher long-term mean FYW. The importance of different rainfall patterns was a consequence of non-linear interactions between catchment storage, discharge, and the fraction of discharge from the youngest part of storage. We showed that a storage-based or discharge-based young water index can be used to assess which rainfall patterns are likely to produce the highest long-term FYW. These indexes could be helpful in evaluating watershed models that simulate both celerity and velocity.

The third question was: how might changes in rainfall due to climate change affect the long-term FYW? Downscaled climate models were used to identify three plausible future storylines for rainfall, ranging from no change to a relatively extreme change. The most extreme rainfall storyline would see an increase in the long-term mean FYW of nearly 19 percentage points in the LH watershed and 15 percentage points in the TW watershed. In both watersheds, the increase would be especially pronounced in the winter season, whereas the mean young water fraction would decrease sharply in the summer period. The results from these catchments demonstrate how changes in rainfall due to climate change could significantly alter catchment velocities and stream water quality.

Looking forward, our methods can be adapted to other sites to understand the link between rainfall variability and transit times in a range of environments, using the young water fraction or whatever age fraction is most appropriate. The trends that emerge can be used to improve the development and evaluation of watershed models. At the same time, efforts to link these trends (e.g., the inverse storage effect) to physical processes in the catchment are ongoing (e.g., Kim et al., 2016; Pangle et al., 2017) and should continue. Finally, more attention should be paid to the effect of rainfall shifts on catchment velocities in climate change impact assessment and planning.

## 2.10 Appendix A: Flux pattern independence in an idealized system

We show here that under idealized conditions the  $F(t)$  of a catchment with a time-invariant rSAS function  $\Omega_Q(S_T)$  will only depend on the total volume of discharge over the past 90 days, irrespective of its pattern.

First, consider a catchment as a simple control volume with input rainfall  $J(t)$  equal to output discharge  $Q(t)$ , and a time-invariant  $\Omega_Q(S_T)$ . Kim et al. (2016) showed that such a system should exhibit steady-state flow pathways and storage. As such, the bTTD in flow-weighted time will be time-invariant (Niemi, 1977; Rodhe, Nyberg, and Bishop, 1996), and the real time  $\overleftarrow{P}_Q(T, t)$  will be equal to (Kim et al., 2016, equation 8):

$$\overleftarrow{P}_Q(t - t_i, t) = \overleftarrow{P}_Q^*(\tau_Q - \tau_J) \quad (2.19)$$



where  $t_i$  is injection time with  $T = t - t_i$ ,  $\overleftarrow{P}_Q^*$  is the time-invariant bTTD in flow-weighted time, and  $\tau_Q$  and  $\tau_J$  are the discharge-weighted and rainfall-weighted time, respectively, defined here with the same normalizing constant  $Q_{at}$  as:

$$\tau_Q(t) = \frac{1}{Q_{at}} \int_{t_0}^t Q(\tau) d\tau \quad (2.20)$$

$$\tau_J(t_i) = \frac{1}{Q_{at}} \int_{t_0}^{t_i} J(\tau) d\tau \quad (2.21)$$

Combining and rearranging Equations (2.19), (2.20), and (2.21) gives:

$$\overleftarrow{P}_Q^*(\tau_Q - \tau_J) = \overleftarrow{P}_Q^* \left( \frac{1}{Q_{at}} \int_{t_i}^t Q(\tau) d\tau + \frac{1}{Q_{at}} \int_{t_0}^{t_i} Q(\tau) d\tau - \frac{1}{Q_{at}} \int_{t_0}^{t_i} J(\tau) d\tau \right) \quad (2.22)$$

When  $J(t)$  and  $Q(t)$  are equal, the last two terms in Equation (2.22) cancel. We can rearrange Equations (2.19) and (2.22) to get:

$$\overleftarrow{P}_Q(T, t) = \overleftarrow{P}_Q^* \left( \frac{1}{Q_{at}} \int_{t-T}^t Q(\tau) d\tau \right) \quad (2.23)$$

With  $T = 90$  days, the left-hand-side is the definition of  $F(t)$ . From the integral on the right-hand-side, it is clear the time-variability of the young water fraction only depends on changes in the total volume of water discharged over the past 90 days.

Although the conditions described here are violated in real watersheds, the approximation of a time-invariant  $\overleftarrow{P}_Q^*$  has been effective at simulating tracer dynamics in a number of settings (Ali, Fiori, and Russo, 2014).

## Chapter 3

### Towards better understanding and calibration of baseflow transit times in an integrated surface-subsurface watershed model.<sup>1</sup>

#### 3.1 Highlights

- ParFlow-CLM catchment model calibrated to discharge and water age.
- Use of age-based calibration criteria reduced parameter equifinality.
- Hydrologic controls on baseflow transit times poorly represented by steady-state model.

#### 3.2 Summary

There is growing recognition that watershed models should be evaluated against both catchment celerities (i.e., the hydrograph) and catchment velocities (i.e., tracer-derived transit times). This is increasingly common with conceptual watershed models, but still rare with computationally expensive, spatially-distributed surface-groundwater models. The goal of this work was to (1) demonstrate a practical approach to calibrating a surface-groundwater model to both discharge and tracer data and (2) characterize the sensitivity of discharge and baseflow stream water age to changes in porosity, hydrologic conductivity, and other critical zone parameters. Virtual experiments were run with

---

<sup>1</sup>This chapter will be submitted for publication in a peer-reviewed journal. Ciaran Harman and Bill Ball were involved in hypothesis development, study design, and results interpretation. Reed Maxwell provided technical assistance with ParFlow-CLM and SLIM-FAST simulations. Anthony Buda provided data and interpretation of data from the study site. Dano Wilusz participated actively in all of the above and also produced the first draft of text and all figures and tables were created by Dano Wilusz.

the fully-distributed surface-groundwater model ParFlow-CLM with SLIM-FAST particle tracking. The ParFlow-CLM simulation was configured and forced with extensive field intelligence from the USDA's Mahantango Creek experimental catchment in PA, USA. We describe a novel calibration processing chain considering roughly 20,000 model realizations with plausible parameter sets in a 2D cross-section of the catchment. Model realizations were used to perform global sensitivity analyses (i.e., Monte Carlo filtering and boosted regression trees). The calibration identified behavioral parameter sets that reasonably simulated observed daily discharge (Kling-Gupta efficiency  $> 0.5$ ), as well as tracer-derived median baseflow age (240-300 days) and groundwater age ( $> 3$  years). The inclusion of age-related calibration criteria is shown to reduce parameter equifinality. Longer baseflow transit times were associated with (1) higher porosities and (2) a non-linear interaction between high lateral conductivity in the upper soil layer and low lateral conductivity in the underlying fractured bedrock. Overall, the results suggest that the age of baseflow may be significantly influenced by upper soil properties due to dynamic partitioning between recharge and overland flow, which is not captured in commonly used steady-state simulations. The findings are of significance for efforts to model water and non-point source pollution transport at catchment scale.

### 3.3 Introduction

Integrated surface/subsurface hydrologic models (ISSHMs) are evolving into a powerful and accessible tool for studying and simulating catchment-scale flow and transport processes (Paniconi and Putti, 2015). For purposes here, ISSHMs are defined as hydrology models that (1) are physically-based, continuum-based, and spatially-distributed (using terms defined in Hrachowitz and Clark (2017)) and (2) explicitly couple variably-saturated groundwater with surface water systems (e.g., ParFlow (Kollet and Maxwell, 2006), HydroGeoSphere (Brunner and Simmons, 2012), CATHY (Camporese et al., 2010)). Many ISSHMs simulate additional processes including land-surface energy and dynamic vegetation. An inter-comparison project evaluated eleven ISSHMs and found similar qualitative performance in benchmark experiments (Maxwell et al., 2014; Kollet et al., 2017), indicating progress towards technological maturity. However, although ISSHMs have the theoretical potential to make predictions about flow and transport without calibration data, actual applications have shown that calibration is generally required (Paniconi and Putti, 2015).

Although ISSHMs have many applications (see Table 1 in Maxwell et al. (2014)), this work focuses on the simulation of rainfall-runoff generation. The ability to accurately model runoff supports hydrologic decision making including water resources planning, flood protection, and contaminant transport (Beven, 2011). ISSHMs faithfully reproduce runoff related field processes including complex interactions between spatially-varying hydroclimatic forcing, topography, land cover, surface roughness, subsurface permeability and porosity, and other landscape features. Multiple interacting runoff mechanisms are simulated including Hortonian and Dunnian overland flow (Horton, 1933; Dunne and Black, 1970), subsurface stormflow, and groundwater flow.

A growing number of studies have used ISSHMs to simulate and describe runoff generation mechanisms. Model complexity ranges from idealized, single-event simulations in sloping slabs (e.g., Camporese et al., 2009; Meyerhoff and Maxwell, 2011; Delfs et al., 2013) to more realistic, multi-year simulations at catchment scale (e.g., Qu and Duffy, 2007; Li et al., 2008; Ebel et al., 2008; Heppner, Loague, and VanderKwaak, 2007; Mirus et al., 2009; Srivastava et al., 2014; Maxwell et al., 2015; Cornelissen, Diekkrüger, and Bogena, 2016; Bisht et al., 2017; Maier, Breuer, and Kraft, 2017). In addition to providing proof-of-concept of ISSHM capabilities, these studies use model forensics and local sensitivity analysis to explore factors affecting runoff generation. For example, early studies described the combined influence of Dunnian and Hortonian runoff on catchment discharge and the importance of antecedent soil moisture on the runoff response (VanderKwaak and Loague, 2001; Loague et al., 2005). Subsequent controlled studies in idealized domains showed how stochastic heterogeneity in hydrologic conductivity promotes localized overland flow and faster runoff response to rainfall events (Kollet and Maxwell, 2006; Maxwell and Kollet, 2008; Gilbert et al., 2016). Other research has begun to explore the relationship between runoff generation and various catchment characteristics, such as the soil-water retention curve (Morita and Yen, 2002), subsurface geologic representation (Gauthier et al., 2009), and the amount of evapotranspiration (ET) (Camporese et al., 2014), as well as issues related to model grid resolution (Maxwell and Kollet, 2008; Sulis, Paniconi, and Camporese, 2011) and spin-up (Seck, Welty, and Maxwell, 2015). The trajectory of studies suggest that ISSHMs will have an important role in the development of process-based solutions to the closure problem, the "holy grail" of scientific hydrology (Beven, 2006b).

An open question about the application of ISSHMs in runoff studies is when and how to perform

parameter calibration. Compared to conceptual models, ISSHMs have a large number of unknown or uncertain parameters. In a few highly-instrumented catchments, ISSHM parameters populated with commensurate field observations seem to reasonably reproduce runoff behavior, with no calibration needed (e.g., Mirus et al., 2009). In many cases, however, it remains desirable to calibrate ISSHMs against available field information. This may be a challenge for two reasons. First, ISSHMs are generally too computationally expensive to do a comprehensive search of the parameter space. As a result, most ISSHM calibrations are done manually, as is the case for many of the studies cited above. In this regard, however, computational speeds continue to improve and modelers are beginning to apply automated approaches (e.g., Haahti et al., 2015). Second, most ISSHM calibrations tune model parameters to fit catchment celerities (i.e., discharge at the outlet) (Sebben et al., 2013; Paniconi and Putti, 2015). Several studies have shown that a good model fit to outlet discharge does not guarantee a good fit to other model output including upstream discharge and distributed soil moisture (Ebel et al., 2008; Camporese et al., 2009).

It is reasonable to assume the calibration of ISSHMs can be improved by considering model fit to tracer-derived estimates of travel times of water through the watershed, hereafter referred to as catchment transit times. Previous studies using ISSHMs have not incorporated catchment transit times into the calibration objective function (though a few have used transit times to evaluate uncalibrated models (e.g., Maxwell et al., 2015)). As argued by McDonnell and Beven (2014), hydrologic models should be calibrated against both volumetric water discharge and catchment transit times because they are the observable consequences of fundamentally different integrated catchment properties (i.e., catchment celerities and flow velocities, respectively). Studies using more conceptual models have demonstrated the value of calibrating against both discharge and transit time data (Birkel and Soulsby, 2015). The same may not be true, however, for ISSHMs, for the following reason. In more conceptual models, the parameters that control catchment discharge (e.g., active storage, porosity) are often distinct from those that control travel times (e.g., passive storage, effective porosity) (e.g., Birkel and Soulsby, 2016). By contrast, in ISSHMs the parameters controlling discharge and transit times are unitary; they are the physical features of the catchment including topography, permeability, and porosity. Given this fundamental difference, more ISSHM-specific studies are needed to understand the extent to which travel time data helps constrains ISSHM

parameter values.

Efficient calibration of ISSHMs to discharge and travel times requires a good understanding of parameter sensitivity. Global sensitivity analysis investigates how changes in the plausible range of model parameters maps onto changes in model outputs. Sensitivity analysis helps reveal the inner working of complex models, the linear and non-linear interactions between parameters, and the most important parameters and associated hydrologic processes (Pianosi et al., 2016). In the first and perhaps only application of global sensitivity analysis in a catchment-scale ISSHM, Srivastava et al. (2014) analyzed the sensitivity of total streamflow, groundwater flow, and ET to 33 different parameters in simulations of a 3700km<sup>2</sup> catchment in Florida, USA. The analysis showed that peak streamflow was most sensitive to the hydrologic conductivity of the confining unit of the upper basin. The representation of runoff processes, however, was limited by the low lateral resolution of model grid cells (1500m). In a higher resolution study, Gilbert et al. (2016) explored the global sensitivity of runoff to variation in the subsurface permeability field, but in an idealized catchment that neglected the influence of topography, ET, and different geologic units on runoff. In both studies, transit times were not considered.

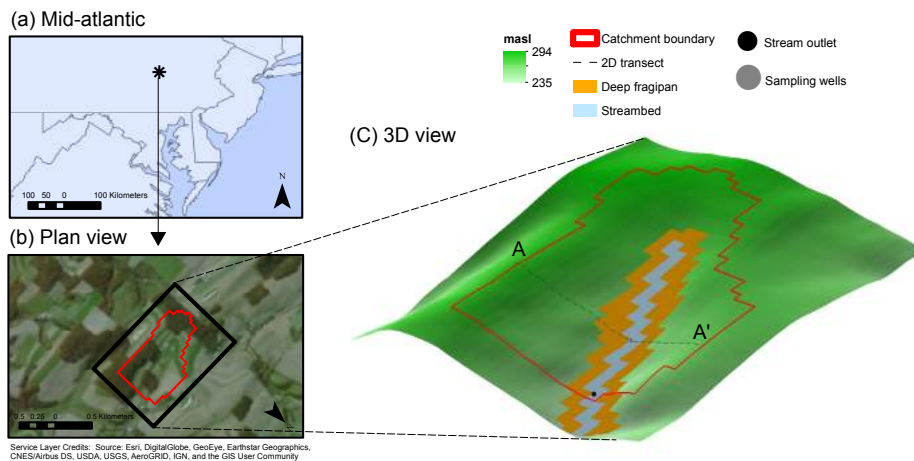
This paper has three objectives, all of which contribute toward the overall goal of developing methods for using ISSHMs to better understand and simulate runoff generation mechanisms at catchment scale. The first objective was to demonstrate a practical approach to computer-assisted calibration of a catchment-scale ISSHM to both discharge and transit time data. This requires novel, computationally efficient techniques, including the use of a 2D emulator and a periodic steady-state baseflow travel time estimator. The second objective was to quantify the benefit in terms of model performance of calibrating a catchment-scale ISSHM to both discharge and transit times. The results were intended to inform approaches to calibration in future studies. The third objective was to use a global sensitivity analyses (1) to understand how individual model parameters affect stream discharge and transit times and (2) to make inferences about dominant catchment processes affecting catchment celerities and velocities. Parameter sensitivity and process controls were assessed using Monte-Carlo filtering and boosted regression trees. We developed and applied the methods described here in the USDA Mahantango Experimental catchment using the ParFlow-CLM model. Both are described in the next two sections in more detail.

## 3.4 Site Description

### 3.4.1 Physiography

An ISSHM model was built and calibrated for the FD-36 subcatchment of the USDA Mahantango Creek experimental watershed near Harrisburg PA, USA (see Figure 3.1). FD-36 is a small ( $0.4 \text{ km}^2$ ) agricultural headwater basin in the Valley and Ridge physiographic province, underlain by shales, siltstones and sandstones (Urban, 1977). The basin drains to a single stream in the southwest quadrant of the intensively-studied,  $7.8 \text{ km}^2$  WE-38 experimental catchment, which is itself a subcatchment of the  $420 \text{ km}^2$  Mahantango Creek watershed and the larger Susquehanna River basin. Land use is approximately 50% cropland, 20% pasture, and 30% woodland (Needelman et al., 2004; Bryant et al., 2011). Previous research suggests that surface and subsurface flow leaves the catchment at the gauged outlet and does not cross other catchment boundaries (Gburek and Folmar, 1999).

Based on SSURGO soil surveys and other synoptic field measurements, the geology can be

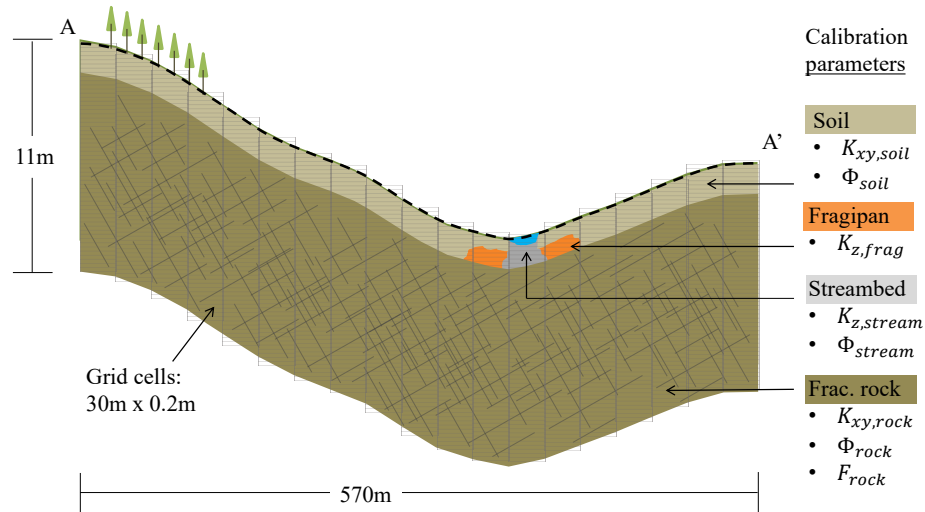


**Figure 3.1:** Map of the study site in the FD36 sub-catchment of the USDA Mahantango experimental catchment. (a) Location of the catchment in central PA, USA. (b) Satellite image of the catchment showing the rectangular region modeled in ParFlow and the FD36 catchment boundary. (c) 3D rendering of the rectangular area modeled by ParFlow, the catchment boundary, the location of the streambed and subsurface fragipan, the location of the stream outlet and CFC sampling well, and the transect A-A' used for the 2D calibration and shown in Figure 3.2.

classified into four hydrogeologic units (see Figures 3.1c and 3.2): well-drained soil, a poorly-drained fragipan horizon, streambed upper sediments, and fractured bedrock. Except for near the riparian area, the upper 1.4m consists of residual, fine-loamy, well-draining soils (Needelman et al., 2004). In the riparian zone, there is a lower layer (0.6m-1.4m) of poorly draining, colluvial fragipan with relatively high clay content (Needelman et al., 2004). The streambed is underlain by alluvial sediments. The location of the fragipan and streambed are shown in Figures 3.1 and 3.2. These layers are underlain by a fractured bedrock layer up to 60m thick, with relatively low porosity and a roughly exponential decline with depth in hydrologic conductivity (Gburek and Folmar, 1999).

### 3.4.2 Hydrographic and age tracer data

The hourly hydrology and meteorology of FD-36 and nearby catchments is well-studied and extensively monitored, which facilitated our modeling. Discharge was measured by a flume located at the catchment outlet as shown in Figure 3.1. Hourly rainfall was measured at a Fisher-Porter rain gauge designated RE-37 (Buda et al., 2011) and located approximately 0.8 km southeast of the



**Figure 3.2:** Schematic of the 2D catchment transect modeled used in calibration, including the location of the four hydrogeologic units (soil, fragipan, streambed, and fractured rock) and the calibrated parameters (as defined in Section 3.6).



catchment outlet. Hourly temperature, wind speed, wind direction, downward solar radiation, relative humidity, and saturated vapor pressure were measured at a monitoring station operated by the Natural Resource Conservation Service approximately 5 miles west. Downward longwave radiation and atmospheric pressure were obtained from the North American Land Data Assimilation System (NLDAS) for the overlaying 0.125 degree grid cell.

Several studies have used tracer data to estimate the age of water in catchment storage and discharge in the greater WE-38 catchment. Burton et al. (2002) used chlorofluorocarbon and tritium/helium-3 data to model groundwater age at a range of depths (3-20m) at a stream transect close to FD-36. The modeled mean age estimates were highly scattered but averaged around 10 years. The groundwater age estimates on the northern hillslope were somewhat older than the southern hillslope, which was attributed to anisotropic fracture geometry. In a separate study, McGuire, DeWalle, and Gburek (2002) used oxygen-18 isotopes to stream water travel times at another subwatershed of WE-38 in the south east corner, and found a mean age of stream baseflow of approximately 9.5 months.

### 3.5 ParFlow-CLM description

The FD-36 catchment was modeled using ParFlow-CLM (Ashby and Falgout, 1996; Jones and Woodward, 2001; Maxwell and Miller, 2005; Kollet and Maxwell, 2006; Maxwell, 2013) with SLIM-FAST particle tracking (Maxwell and Tompson, 2006). ParFlow is an open-source hydrology modeling simulation platform designed for parallel computing. It has been coupled to the Common Land Model (CLM) to simulate tree canopy processes (e.g., interception), atmospheric forcing (e.g., rain, snow) and land surface processes (e.g., latent heat flux, sensible heat flux). Model processes and equations related to rainfall runoff are described briefly. More details are available in the ParFlow User's Manual (Maxwell et al., 2016).

ParFlow-CLM uses a cell-centered finite difference scheme to solve the 3D Richard's Equation (see Equation 3.1) (Richards, 1931) for subsurface flows, and a coupled 2D kinematic wave equation for surface flows, which is only active when the ponded depth at the surface is greater than zero (Equation 3.2). To improve computational efficiency, ParFlow-CLM can represent topography using a terrain-following transform (Maxwell, 2013), where each grid cell indexed in the  $x$ ,  $y$ , and  $z$  directions

is re-indexed by  $i, j$ , and  $k$  and assigned a local slope  $\theta_{i,j}$ . The flow-depth discharge relationship is based on Manning's equation (Equation 3.3).

$$S_s S_w(h) \frac{\partial h}{\partial t} + \phi \frac{\partial S_w(h)}{\partial t} = \nabla \cdot -\mathbf{K}_s(\mathbf{x}) k_r(h) [\nabla(h+z) \cos(\theta) + \sin(\theta)] + q_s(\mathbf{x}) \quad (3.1)$$

$$(-\mathbf{k} \cdot \mathbf{K}_s(\mathbf{x}) k_r(h) \nabla(h+z)) = \frac{\partial ||h, 0||}{\partial t} - \nabla \cdot ||h, 0|| \mathbf{v}^{sw} - q_r(\mathbf{x}) \quad (3.2)$$

$$v_x^{sw} = \frac{\sqrt{\tan \theta_x}}{m} h^{2/3} \quad v_y^{sw} = \frac{\sqrt{\tan \theta_y}}{m} h^{2/3} \quad (3.3)$$

In these expressions,  $S_s$  is the specific storage [ $L^{-1}$ ];  $S_w$  is the relative saturation [-];  $h$  is the pressure head [L];  $\phi$  is the porosity [-];  $\mathbf{K}_s(\mathbf{x})$  is the saturated hydraulic conductivity tensor [ $LT^{-1}$ ];  $\mathbf{x}$  is the location vector [L];  $k_r$  is the relative permeability [-];  $z$  is the elevation with the  $z$ -axis upward [L];  $\theta$  is the local angle of slope [-];  $q_r$  represents infiltration to soil after accounting for processes including wet evaporation, throughfall, and snowmelt [ $LT^{-1}$ ];  $q_s$  is a general source/sink term [ $LT^{-1}$ ];  $\mathbf{k}$  is the vertical unit vector;  $\mathbf{v}^{sw}$  is the depth-averaged surface water velocity vector [ $LT^{-1}$ ]; and  $m$  is Manning's coefficient [ $TL^{-1/3}$ ]. The expression  $||h, 0||$  equals the greater value of the two quantities. The relationship between  $S_w$ ,  $k_r$ , and  $h$  is parameterized by the van Genuchten model (Genuchten, 1980) with parameters  $\alpha$  [ $L^{-1}$ ] and  $n$ .

The source/sink term  $q_s(\mathbf{x})$  in the subsurface mass balance equation (Equation 3.1) is defined as

$$q_s(\mathbf{x}) = E_{tr}(S_w, f_{root}(z), m_{rs}) + E_g \quad (3.4)$$

where  $E_{tr}$  is total plant transpiration from the root zone [ $LT^{-1}$ ] depending on the fraction of roots  $f_{root}$  at depth  $z$  and the slope of the conductance to photosynthesis relationship  $m_{rs}$ ; and  $E_g$  is the evaporative flux from the soil surface [ $LT^{-1}$ ]. In addition to the variables listed, the terms are functions of other model parameters and hourly atmospheric variables, as described in Dai (2003) and Maxwell and Tompson (2006).

Simulation results from ParFlow-CLM can be used with the Lagrangian particle tracking scheme SLIM-FAST (Maxwell and Tompson, 2006) to simulate flow paths and catchment travel times through

the catchment. For each timestep, ParFlow-CLM simulates both a subsurface and surface velocity flow field, which can be numerically integrated to define the pathlines that water particles follow through the catchment:

$$\mathbf{v} = \frac{d\mathbf{x}}{dt} \tag{3.5}$$

where  $\mathbf{v}$  is the instantaneous velocity vector,  $\mathbf{x}(t)$  is the pathline, and  $t$  is time.

## 3.6 Methods

### 3.6.1 Model structure and initial parameters

A ParFlow-CLM model was constructed for the FD-36 watershed in a rectangular surface domain, 1230m east-west by 870m north-south. The model domain extends at least 100m beyond the edges of FD-36, such that the catchment boundary and stream location emerge naturally from the flow simulation. Rock cores extracted in a nearby subcatchment of the WE-38 site showed extensive fracturing down to approximately 15m (Gburek and Urban, 1990), which was used as the bottom of the model domain. Selection of the model grid size required a trade-off between better model fidelity at high resolution, and faster solver convergence at low resolution. After initial tests, a model grid size of 30m in the x and y direction and 0.2m in the z direction yielded reasonable performance and run times, giving a grid cell count of 41 by 29 by 55, or a total of 65395 grid cells. The local slopes  $\theta_x$  and  $\theta_y$  at each grid cell was estimated by resampling 0.5m LIDAR data to 30m. Following procedures recommended in Maxwell et al. (2016) to accelerate solver convergence, the local slope  $\theta_x$  or  $\theta_y$  with the smallest absolute value was set to zero (i.e., diagonal lateral flow was neglected). Based on catchment characteristics discussed in Section 3.5, each grid cell was assigned one of the four hydrogeologic units: soil, fragipan, streambed, or fractured bedrock. The top 0.6m was either soil or streambed; from 0.6m to 1.4m was assigned soil, fragipan, or streambed; and from 1.4m to 15m was assigned fractured bedrock, as shown in Figures 3.1 and 3.2.

Several of the ParFlow-CLM parameters were fixed using field data and literature values for each hydrogeologic units (see Table 3.1). The Manning's roughness  $m$  was set to  $5.52 \times 10^{-6} \text{ hr m}^{-1/3}$  ( $0.20 \text{ s m}^{-1/3}$ ), a recommended value for bare cropland (Schall, Richardson, and Morris, 2008). The

**Table 3.1:** List of fixed (i.e., not calibrated) model parameters.

Param	Zone	Value	Units	Ref.	Description
$K$	a	0.062	$m/hr$	Troch et al. (1993)	saturated conductivity
$n$	a	1.29	-	Troch et al. (1993)	van Genuchten parameter
$\alpha$	a	2.32	$m^{-1}$	Troch et al. (1993)	van Genuchten parameter
$m$	a	$5.52 \times 10^{-6}$	$hr/m^{\frac{1}{3}}$	Maxwell et al. (2016)	mannings roughness
$r$	a	13	-	manually adjusted	slope for conductance to photosynthesis relationship

Zone abbreviations: 1: all zones

van Genuchten parameters  $n$  and  $\alpha$  for soil were set to 1.29 and  $2.32 \text{ m}^{-1}$ , respectively, based on typical values in soil sampling from the greater WE-38 catchment (Troch et al., 1993). In the absence of better data, the same values were assigned to the rest of the domain. The saturated hydraulic conductivity tensor  $\mathbf{K}_s(\mathbf{x})$  was assigned a base value of  $0.062 \text{ m/hr}$  for the entire domain, based on an areal average of soil sampling from the greater WE-38 catchment (Troch et al., 1993) and results from interval packer tests (Gburek, Folmar, and Urban, 1999).

Eight parameters - which were poorly constrained by the data and literature - were assigned a range of plausible values for subsequent calibration (see Table 3.2). Multiplicative factors were applied to  $\mathbf{K}_s(\mathbf{x})$  in the lateral  $x$  and  $y$  and the vertical  $z$  directions. Specifically, to account for potentially faster lateral flow due to macropores and other subsurface heterogeneity (Uchida, Tromp-Van Meerveld, and McDonnell, 2005; Lin, 2006; Fang et al., 2015), lateral multiplicative factors for soil ( $K_{xy,soil}$ ) and bedrock ( $K_{xy,rock}$ ) were set to the range [10,200] (i.e., from 1 to 200 inclusive) and [1,100], respectively. To account for the impermeable nature of the clay-rich fragipan layer (Buda et al., 2009), the vertical multiplicative factor for the fragipan ( $K_{z,fragipan}$ ) was set to  $[10^{-8}, 10^{-3}]$ . To account for the potential for high connectivity between the fractured bedrock and stream, the vertical multiplicative factor for the streambed was set to [1,100]. Results from packer tests in Gburek, Folmar, and Urban (1999) suggest that the hydrologic conductivity  $\mathbf{K}_s(\mathbf{x})$  in the fractured bedrock declines exponentially with depth. The value of  $\mathbf{K}_s(\mathbf{x})$  was assigned an  $e$ -folding depth  $F_{rock}$  [m] in the range [0.35,0.95]. The soil porosity was constrained to [0.27,0.54] based on the range of values reported in the field study by Lindeburg (2011). The fractured bedrock porosity was given the range [0.01,0.03] based on published tracer-aided MODFLOW modeling results (Gburek, Folmar, and Urban, 1999). Finally, the porosity of the streambed ( $\phi_{stream}$ ) was constrained to the

**Table 3.2:** List of parameters calibrated using GLUE, and the range of values that was explored.

Param	Zone	Range	Units	Ref.	Description
$K_{xy,soil}$	s	1-200	-	-	lateral $K$ multiplier of soil
$K_{xy,rock}$	r	1-100	-	-	lateral $K$ multiplier of rock
$K_{z,fragipan}$	f	$10^{-8}$ - $10^{-3}$	-	-	vertical $K$ multiplier of fragipan
$K_{z,streambed}$	b	1-100	-	-	vertical $K$ multiplier of streambed
$F_{rock}$	r	0.35-0.95	$m^{-1}$	4	length scale of exponential decline in $K$ of rock
$\phi_{soil}$	s,f	0.27-0.54	-	2	porosity of soil
$\phi_{streambed}$	b	0.01-0.54	-	1,3	porosity of streambed
$\phi_{rock}$	s,f	0.01-0.03	-	3	porosity of rock

Reference numbers: 1: Troch et al. 1993, 2: Lindeburg 2011,

3: Burton et al. 2002, 4: Gburek et al. 1999

Zone abbreviations: s: soil, f: fragipan, b: streambed, r: rock

full range of values of potential bedrock and soil porosity, [0.01, 0.54].

### 3.6.2 GLUE calibration

The 8 unknown parameters were calibrated using a variant of Generalized Likelihood Uncertainty Estimation (GLUE) (Beven and Binley, 1992; Beven and Binley, 2014) with an informal likelihood function. Figure 3.3 has a flow chart of the calibration process, with repeat Monte Carlo (MC) sampling of parameter sets, up to a target of at least 20000 sets. Each parameter set was run in parallel on one of 400 processors on the Maryland Super Computing Cluster (MARCC). To be computationally tractable, the calibration was run on a representative 2D slice of the full catchment (see Figures 3.1 and 3.2), with dimensions 570m long by 15m deep and a total of 1045 grid cells. Although this "2D emulator" lacks the complexity of the full 3D catchment, it approximates the first-order lateral interactions driven by topography, geology, and climate.

For each parameter set, the value of each parameter was randomly selected from the ranges described in Section 3.6.1 and shown in Table 3.2, according to either a uniform or - for parameters spanning several orders of magnitude - log-uniform probability distribution. The parameters were used to run the ParFlow-CLM model in three stages of increasing complexity: steady-state, periodic steady-state, and fully transient. The steady-state run was used to spin up the model and estimate the steady-state median transit time in the stream and groundwater. The periodic steady-state

was used to approximate the median baseflow transit time. The fully transient model was used to simulate one-year of hydrology. The results were used in post-processing to evaluate parameter sets according to the GLUE limits of acceptability approach (Beven, 2006a). These three stages and model evaluation are described below.

### 3.6.2.1 Stage 1: Steady-state simulation

A steady-state simulation was run to spin up the model and to estimate steady-state transit times. "Steady-state" refers to the fact the model forcing (i.e., rainfall) was set constant and the model ran until the catchment achieved equilibrium (i.e., rainfall inputs equal discharge outputs). The steady rainfall rate was set equal to the long term average observed discharge ( $\bar{Q}$ ). CLM was disabled such that  $ET$  and other land-atmosphere processes were not simulated. A model run time of  $2 \times 10^5$  hours (23 years) was found sufficient to achieve stasis. The final hydraulic pressure field, water table, and soil saturation levels were used as the initial condition for the periodic steady-state and fully transient stages.

For each parameter set, the steady-state result was used with SLIM-FAST particle tracking to estimate steady-state median transit time ( $MTT_{ss}$ ) of discharge. First,  $N_{ss}=1000$  particles were injected into the 2D surface and tracked forward. Each particle  $i$  was assigned a water volume  $V_{ss,i}$  equal to its share of infiltration  $\bar{Q}\Delta t/N_{ss}$  where  $\Delta t$  is the timestep of injection. The path and travel time of particles were tracked until they (a) left the domain by flowing out the outlet or (b) reached

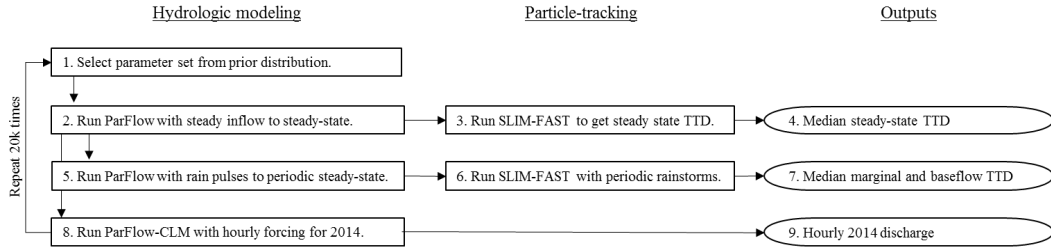


Figure 3.3: Flow chart showing steps in the model calibration process.

the surface (at which point they were assumed to flow out the outlet in a negligible amount of time). The particle transit times were used to estimate the forward transit time probability distribution (TTD) function  $\overrightarrow{P}_{Q,ss}(T)$ , defined as the distribution of transit times  $T$  taken by rainfall to travel to the outlet (Niemi, 1977; Rinaldo et al., 2011). In the steady-state case, this is equal to the backward travel time distribution  $\overleftarrow{P}_{Q,ss}(T)$ , defined as the distribution of transit times  $T$  of water in the stream since injection. The transit time of each particle was used to numerically estimate the forward and backward TTD in CDF form as:

$$\overleftarrow{P}_{Q,ss}(T) = \overrightarrow{P}_{Q,ss}(T) = \frac{\sum_{i=1}^{N_{ss}} V_{ss,i} \mathcal{H}(T - T_i)}{\sum_{i=1}^{N_{ss}} V_{ss,i}} \quad (3.6)$$

where  $\mathcal{H}$  is the Heaviside step function,  $T$  is the time since particle injection, and  $T_i$  is the total transit time from injection to exit for particle  $i$ . The median transit time  $MTT_{ss}$  satisfies the equation  $\overleftarrow{P}_{Q,ss}(MTT_{ss}) = 0.5$ .

### 3.6.2.2 Stage 2: Periodic steady-state simulation

A periodic steady-state simulation was run to estimate the median transit time of water in baseflow and groundwater. For each parameter set, the model was forced with a repeating rainfall pattern with 1 day (24 hours) of uniform precipitation and 13 days (312 hours) with no rainfall. The pattern was chosen to approximate a typical interval between large rainfall events at our study site. The mean simulated rainfall was set equal to long term average observed rainfall ( $\bar{J}$ ). CLM was disabled and  $ET$  was not simulated. Note that the overall flux through the periodic steady-state simulation ( $\bar{J}$ ) was higher than the steady-state simulation ( $\bar{Q}$ ) and can be seen as more typical of the winter period when  $ET$  is negligible. The model was assigned initial conditions from the steady-state simulation and run for a time sufficient to achieve an approximate periodic steady-state (33 cycles). Using the approach described below, the model results were used to estimate the median travel time of discharge on the last day of the periodic cycle (i.e., the 13th day of recession), which was used to approximate the median transit time of baseflow ( $MTT_{bf}$ ). Compared to the value of  $MTT_{ss}$ , the value of  $MTT_{bf}$  captures the effect of transient processes on the baseflow age and was the value adopted for subsequent model calibration.

A variation of the PERTH method (Harman and Kim, 2014) was used to estimate  $MTT_{bf}$  from the periodic steady-state simulation results. First, the hydraulic pressure field from the last 14 days of the periodic steady-state simulation was looped for 100 cycles (3.8 years) and used as input pressure field for SLIM-FAST particle tracking. For the first day in the first cycle, a swarm of  $N_{ps}=10000$  particles with volume  $V_{ps} = \bar{J}\Delta t/N_{ps}$  were introduced just below the land surface and tracked forward toward the catchment outlet. The transit time for each particle was used to calculate the cumulative forward travel time distribution  $\overrightarrow{P_{Q,ps}}(T, t = 1 \text{ day})$  in the manner of Equation 3.6. The value of  $\overrightarrow{P_{Q,ps}}(T, t)$  on the remaining 13 days of the cycle was undefined (due to zero rainfall inputs), and the value of  $\overrightarrow{P_{Q,ps}}(T, t)$  for all other times was determined by assuming periodic steady-state:

$$\overrightarrow{P_{Q,ps}}(T, t) = \overrightarrow{P_{Q,ps}}(T, t \pm \Delta n) \quad \text{for all positive integers } n \quad (3.7)$$

where  $\Delta$  is the period of the system (14 days). The PDF form of the backward TTD  $\overleftarrow{p_{Q,ps}}(T, t)$  was calculated with a rearranged version of Niemi's theorem for a system with a single inflow and outflow (Niemi, 1977):

$$\overleftarrow{p_{Q,ps}}(T, t) = \overrightarrow{p_{Q,ps}}(T, t - T) \frac{J(t - T)}{Q(t)} \quad (3.8)$$

where  $\overrightarrow{p_{Q,ps}}(T, t)$  is the PDF form of the forward TTD, and  $J(t)$  and  $Q(t)$  were obtained from the last cycle of the ParFlow simulation and extended using the assumption of periodic steady-state as in Equation 3.7. The  $MTT_{bf}$  was calculated from the CDF of the backward TTD (i.e.,  $\overleftarrow{P_{Q,ps}}(MTT_{bf}, 14 \text{ days}) = 0.5$ ). For comparison purposes, values of  $\overleftarrow{p_{Q,ps}}(T, t)$  and  $Q(t)$  were also used to calculate the median flow-weighted transit time  $MTT_{fw}$  (Peters, Burns, and Aulenbach, 2014).

Finally, the median age of groundwater  $MTT_{gw}$  was estimated by backward tracking a swarm of particles from a groundwater grid cell to the catchment surface, using the periodic steady-state pressure field. The grid cell was located 5m below the stream surface, where comparable data was available for calibration. Note that the maximum identifiable median transit time was the duration of the particle tracking simulation (3.8 years).



### 3.6.2.3 Stage 4: Fully transient simulation

A fully transient simulation was run to simulate catchment hydrology for the year 2014 including discharge at the catchment outlet. For each parameter set, ParFlow-CLM was run at hourly timesteps for one full year using observed hydroclimatic data as described in section 3.5. The hourly data was aggregated to produce daily estimates of  $Q$ . The steady-state simulation was used as the initial condition. Initial tests suggested that an additional 1 year transient spin up did not substantially affect simulated discharge, so no spin up was applied. Transit times were not simulated in the fully transient simulation, due to computational limitations.

### 3.6.2.4 Model evaluation

The simulation results from each parameter set were compared to observations and designated as "behavioral" or "non-behavioral" based on the limits of acceptability approach (Beven, 2006a). The parameter set was considered behavioral if it passed evaluation criteria related to (1) discharge, (2) baseflow transit time, and (3) groundwater transit time. For discharge, the criteria was a Kling-Gupta-Efficiency (KGE) performance metric greater than 0.5, with the KGE defined in Kling, Fuchs, and Paulin (2012) as:

$$KGE = 1 - \sqrt{(r - 1)^2 + \left(\frac{\mu_s}{\mu_o} - 1\right)^2 + \left(\frac{CV_s}{CV_o} - 1\right)^2} \quad (3.9)$$

where  $r [-]$  is the correlation coefficient between simulated and daily observed discharge,  $\mu [m/hr]$  is mean runoff,  $CV [-]$  is the coefficient of variation, and the subscripts  $s$  and  $o$  refer to simulated and observed data, respectively. Higher KGE values signal a better match between observations and simulations, and a KGE of 1 indicates perfect agreement. The KGE threshold of 0.5 was chosen because a similar threshold has been recommended for a closely related metric of watershed model performance, the Nash-Sutcliffe efficiency (Moriassi, Arnold, and Liew, 2007). For baseflow transit time, the criteria was an  $MTT_{bf}$  within the range of 240-300 days, which encompasses the value of 9.5 months estimated in the field (described in Section 3.4). For groundwater transit times, the criteria was a  $MTT_{gw}$  greater the maximum identifiable transit time of 3.8 years, which is at the lower end of tracer-based groundwater age estimates (also described in Section 3.4). The three criteria were used

to identify an ensemble of behavioral parameter sets. The ensemble was analyzed to investigate the plausible range of parameter values and to generate probabilistic model projections with uncertainty bounds, following the standard GLUE methodology (Beven, 1982; Beven, 2011).

### 3.6.3 Sensitivity analysis

The calibration produced simulations of discharge and transit times for a targeted 20,000 different parameter sets. The ensemble of parameter sets were subjected to global sensitivity analysis to understand how different model parameters affect model performance and to make inferences about the role of different hydrologic processes. The sensitivity analysis focused on three themes that were motivated in Section 3.3: (1) the sensitivity of the calibration results to the choice of calibration criteria, (2) the sensitivity of discharge to parameter selection, and (3) the sensitivity of transit times to parameter selection. The two main analytical tools for the sensitivity analysis were Monte Carlo filtering and boosted regression trees.

#### 3.6.3.1 Monte Carlo filtering

Monte Carlo (MC) filtering was done to identify which parameters were most important for model calibration, following methods first explored in Spear and Hornberger (1980). For each parameter, the marginal cumulative CDFs of the set of behavioral and non-behavioral parameter values were over-plotted. The greater the difference between the two CDFs, the more that particular parameter was constrained by the calibration, and the greater the suggested influence of the parameter on the calibration outcome. Note, however, that the absence of any difference in the CDFs is a necessary but not sufficient condition for screening out parameters (i.e., identifying parameters with negligible influence) due to the possibility of confounding parameter interactions (Pianosi et al., 2016).

The MC filtering analysis was repeated using different definitions for behavioral and non-behavioral. The greater the change in the CDFs when a calibration criterion was removed, the greater the suggested influence of that criterion on the calibration result. In this manner, MC filtering was used to evaluate the value of incorporating age-based criteria into the model calibration.

### 3.6.3.2 Boosted regression trees

Boosted regression trees (BRTs) were used to explore the relationship between model parameters and informative scalar representations of discharge and transit times extracted from the simulation results. The BRT model was chosen for its ability to simulate non-parametric interactions. The simulation quantities of interest (i.e., the response variables) were the high flow condition  $Q_5$  exceeded on 5% of days, the low flow condition  $Q_{95}$  exceeded on 95% of days, and the median transit time measured under steady-state  $MTT_{ss}$ , under baseflow conditions  $MTT_{bf}$ , and as the flow-weighted average  $MTT_{fw}$ . The values of  $Q_5$  and  $Q_{95}$  were derived from an empirical flow duration curve (FDC) constructed from each parameter set based on the Weibull plotting position, which provides an unbiased estimate of the exceedance probability (Blum, Vogel, and Archfield, 2016; Vogel and Fennessey, 1994):

$$FDC(q) = \frac{i_q}{N + 1} \quad (3.10)$$

where  $q$  is daily streamflow,  $i$  is the cardinal rank of  $q$  relative to all days, and  $N$  is the length of record (365 days).

The BRT model is an advanced form of regression described in Elith, Leathwick, and Hastie (2008) and briefly summarized here. For each response variable, the targeted 20,000 parameter runs were separated into a training group (80%) and a testing group (20%). A randomly selected subset of the training data was used to construct a decision tree. Each decision tree uses a series of thresholds to partition the parameter space into regions associated with a prediction, and thresholds are set to minimize residual errors between observations and predictions. Next, another randomly selected subset of data is used to construct another decision tree that predicts the residuals of the first, with special focus on observations that are hardest to predict. The second decision tree is shrunk according to a parameterized learning rate and added to the first. The process is repeated 100 times, generating 100 decision trees, each predicting the residual of the previous. The final BRT is a kind of regression model where each term is an individual tree. The final model performance is tested using the test data. The model was run using the Scikit-learn package (Pedregosa et al., 2011) in Python 2.7.

The BRT allows us to visualize the effect of varying model parameters on the response variables using one- and two-factor partial dependency plots (Friedman, 2001). Partial dependency plots illustrate how varying the value of one or two parameters affects a response variable after "averaging

out" the influence of all other model parameters. For more linear systems, the partial dependency plots should be lines or planes with slopes that mimic linear regression model parameters. For more complex models (such as non-linear hydrologic systems), the partial dependency plots can reveal non-linear and threshold behavior. Note, however, that partial dependency plots can mask strong parameter interactions and should be interpreted with care (as discussed further in Section 3.8).

### 3.7 Results

#### 3.7.1 GLUE calibration results

The calibration routine shown in Figure 3.3 was started and stopped shortly after surpassing 20000 realizations. A total of 380 out of 22886 parameter sets were designated behavioral based on their ability to simulate observed discharge and transit times, as described in Section 3.6.2.4.

The behavioral parameter sets did reasonably well at capturing key trends in the hydrograph. A time-series comparison of the simulated and observed hydrograph shows reasonably good agreement, with a KGE for the 50th percentile simulation of discharge and log discharge of 0.52 and 0.59, respectively (see Figure 3.4). The calibrated model reproduced the fast response and recession

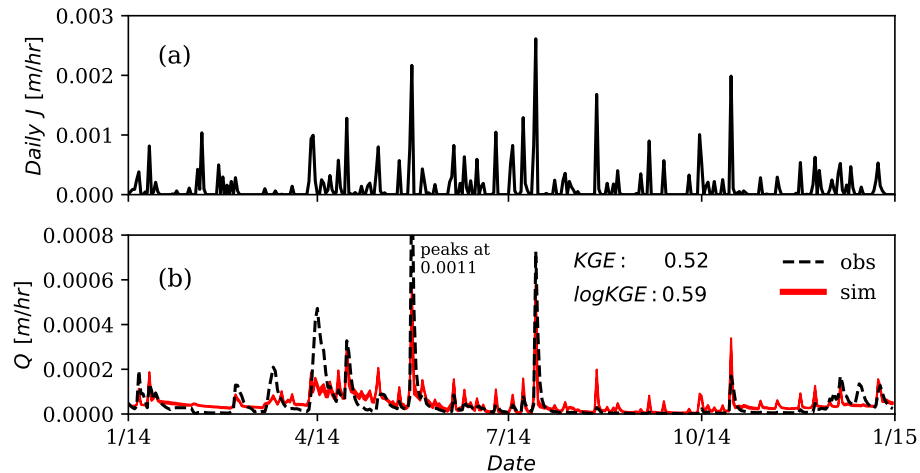


Figure 3.4: (a) The 2014 daily rainfall in the FD36 catchment and (b) the observed and simulated discharge.

following rainfall events, the suppressed rainfall-runoff response in January and February presumably due to snow accumulation, the higher baseflow in the winter, and the lower baseflow due to ET and low rainfall in the summer. Apparent deficiencies of the model include a tendency to overestimate the runoff response to small storms and underestimate the response to large storms. In addition, the model misses some large peaks in the hydrograph during the late winter and early spring, in a manner that suggests that the celerity of snowmelt is too slow.

The behavioral parameters also did well at capturing key trends in the flow duration curve including a steep rise in discharge at low exceedance probabilities and a steep drop in discharge at high exceedance probabilities (see Figure 3.5(a)). In general, behavioral simulations tended to have higher high flows and lower low flows than the non-behavioral models. Overall, however, none of the simulations were able to achieve the high peak flows and low baseflows observed in the data. Furthermore, in the middle ranges, the simulated FDC is flatter than the observed FDC. This suggests that simulated recessions are too steep, whereas the observed recessions are smoother over a range of discharge. Interestingly, the FDC curve does not vary much between parameter sets, suggesting that the basic shape of the FDC may be strongly influenced by other factors such as climate and topography.

The median baseflow transit time  $MTT_{bf}$  for the behavioral simulations was 251 days, which is within the calibration criteria of 240-300 days. The range of  $MTT_{bf}$  for non-behavioral simulation was much larger, from 100-550 days (see Figure 3.5(b)). Thus, although all parameter sets were drawn from evidence-based prior distributions, many could be excluded based solely on the transit time criteria. Figure 3.5(b) also shows that for any given parameter set the steady-state median transit time  $MTT_{ss}$  was significantly lower than  $MTT_{bf}$ , with almost all points falling below the 1:1 line. Almost all of the parameter sets simulated groundwater ages greater than 3 years.

### 3.7.2 Monte Carlo filtering results

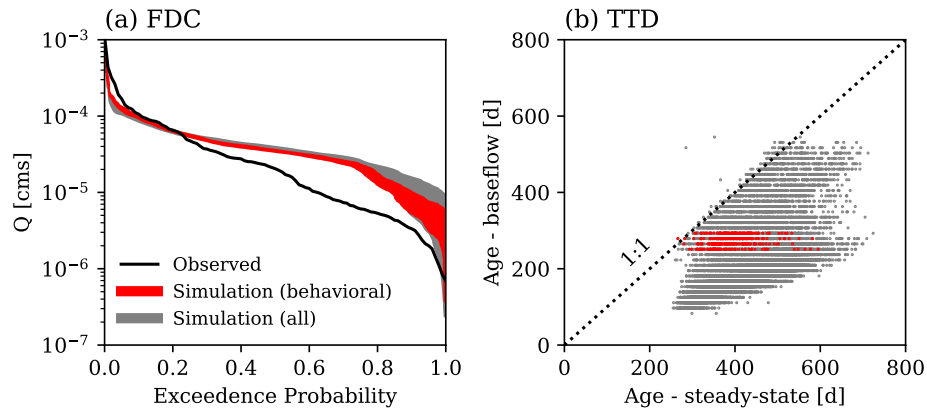
Monte Carlo filtering was used to help determine which parameters had the largest influence on variability in model performance, based on differences in the frequency of parameter values between behavioral and non-behavioral parameter sets. One of the biggest differences was seen in the upper soil layer  $\phi_{soil}$ . The top panel in Figure 3.6(a) shows that the marginal CDF for  $\phi_{soil}$  in non-behavioral

**Table 3.3:** List of GLUE calibration results including the 5th and 95th percentile of the prior and posterior distribution, and the median of the prior and posterior distribution.

Param	Prior range	Posterior range	Prior median	Posterior median	Units
$K_{xy,soil}$	1.3 - 153.5	40.5 - 192.7	44.6	104.0	-
$K_{xy,rock}$	1.3 - 79.4	1.0 - 2.31	10.0	1.4	-
$K_{z,fragipan}$	$1.5 \cdot 10^{-8}$ - $4.5 \cdot 10^{-4}$	$1.8 \cdot 10^{-8}$ - $5.6 \cdot 10^{-4}$	$2.6 \cdot 10^{-6}$	$2.4 \cdot 10^{-6}$	-
$K_{z,streambed}$	1.3 - 79.4	1.2 - 80.4	10.0	10.4	-
$F_{rock}$	0.38 - 0.92	0.41 - 0.93	0.65	0.71	$m^{-1}$
$\phi_{soil}$	0.28 - 0.53	0.27 - 0.38	0.40	0.30	-
$\phi_{streambed}$	0.03 - 0.51	0.03 - 0.51	0.28	0.28	-
$\phi_{rock}$	0.01 - 0.03	0.01 - 0.03	0.02	0.02	-

parameter sets is a straight line (i.e., the same as the prior distribution), whereas the behavioral parameter has a strong preference for lower porosity. This tendency for behavioral parameter sets to contain lower porosities is also reflected in Table 3.3, which shows that the median  $\phi_{soil}$  shifted from 0.40 before calibration to 0.30 after calibration.

Two other important parameters for achieving a behavioral calibration were the lateral hydrologic conductivity in soil  $K_{xy,soil}$  and fractured rock  $K_{xy,soil}$  (see Figure 3.6(c,d)). The behavioral parameter sets tended to have higher values of  $K_{xy,soil}$  (from a median of 44.6 before calibration to



**Figure 3.5:** (a) The observed and simulated flow duration curve and (b) comparison between ages using different estimation techniques.

104.0 after calibration) and lower values of  $K_{xy,rock}$  (from 10.0 to 1.4). The other five calibration parameters had nearly identical marginal CDFs and median values in the behavioral and non-behavioral sets. This was either because changes in these parameters had negligible influence on the calibration criteria, or because their influence was masked by interactions with other parameters.

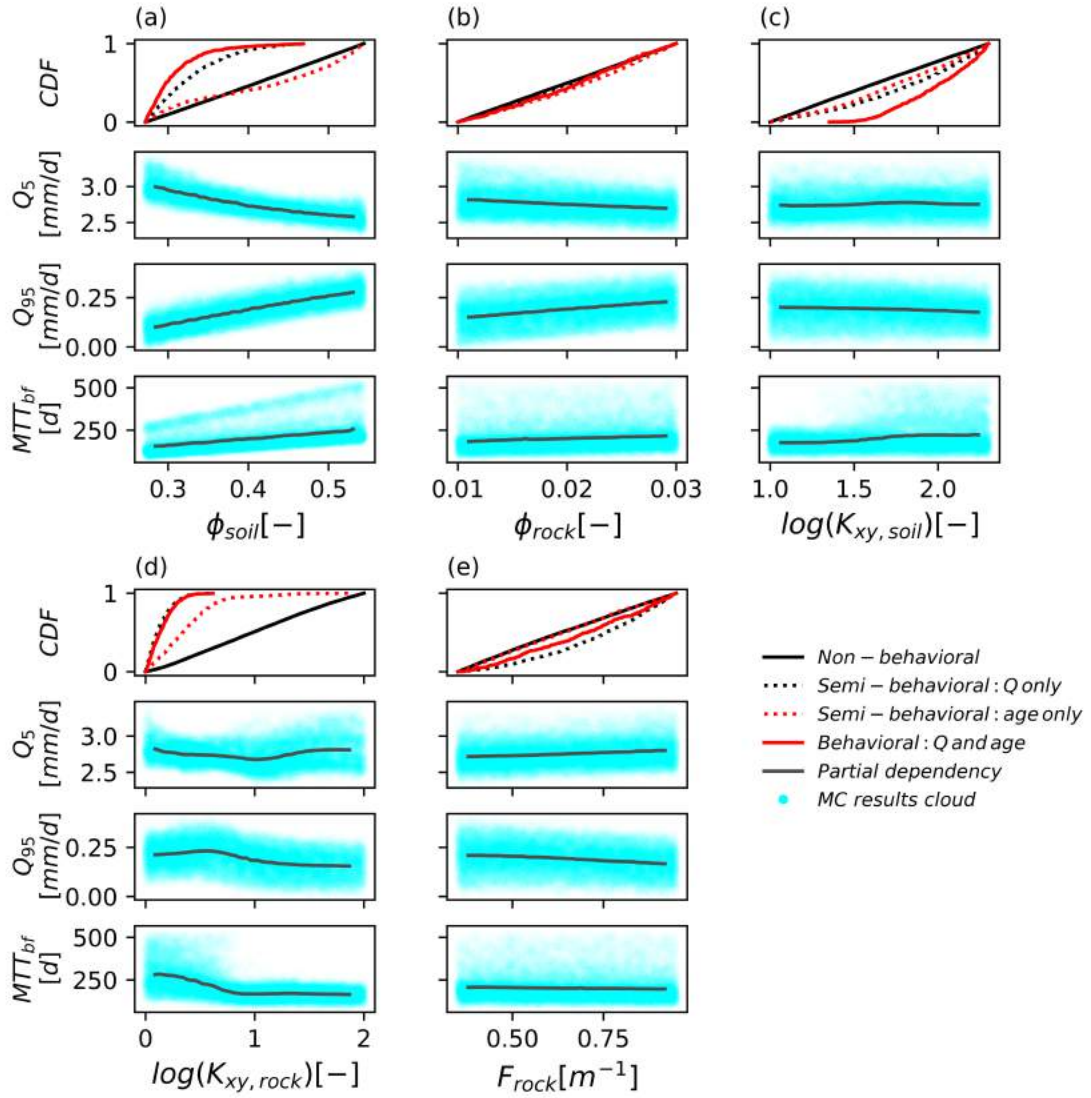
The Monte Carlo filtering analysis also showed that considering transit times in the calibration had a significant effect on the outcome. The top panel of Figure 3.6(c) shows that including the transit time criterion resulting in higher values of  $K_{xy,soil}$  in the behavioral parameter sets. For the other parameters, however, including the transit time criterion did not significantly alter the distribution of behavioral parameters. For example, for both  $\phi_{soil}$  and  $K_{xy,rock}$ , the distribution of behavioral parameters was roughly the same with or without considering transit times.

The Monte Carlo filtering suggests that the most important parameters were porosity and lateral hydrologic conductivity, such that these will be the focus of subsequent analysis.

### 3.7.3 BRT sensitivity analysis - single parameter

The BRT models were trained to predict response variables using 80% of the 22886 model runs, and then tested with the remaining runs. Model performance using the other 20% test data was satisfactory based on graphical analysis of scatter plots and performance statistics. For example, the Nash-Sutcliffe efficiency (Nash and Sutcliffe, 1970) ranged was 0.96 for the prediction of  $Q_5$ , 0.99 for the prediction of  $Q_{95}$ , 0.98 for prediction of  $MTT_{ss}$ , and 0.70 for the prediction of  $MTT_{bf}$ . The BRT model for  $MTT_{bf}$  performed the worst, consistent with the complex controls on transient transit time dynamics. The trained BRT models were used to generate the single parameter partial dependency plots for  $Q_5$ ,  $Q_{95}$ , and  $MTT_{bf}$ , shown in the lower three panels of Figure 3.6(a-e).

The partial dependency plots show that soil porosity  $\phi_{soil}$  has a large, linear influence on  $Q_5$ ,  $Q_{95}$ , and  $MTT_{bf}$  (Figure 3.6(a)). An increase in  $\phi_{soil}$  was associated with a steady decrease in peak flows  $Q_5$ , and a steady increase in low flows  $Q_{95}$  and baseflow transit times. An increase in the fractured rock porosity had a similar influence on discharge and transit times, though to a lesser degree (Figure 3.6(b)). These results make sense in that increases in porosity tend to increase catchment storage, which could in turn decrease the generation of overland flow that contributes to  $Q_5$ , increase the supply of baseflow for  $Q_{95}$ , and increase flushing time leading to higher  $MTT_{bf}$ .



**Figure 3.6:** Results of the global sensitivity analysis for the five most sensitive parameters. Each panel has four subpanels containing the Monte Carlo filtering results (top panel) and the single-parameter partial dependency plots (lower three panels). Each point in the blue cloud is the response from one of the 22886 parameter sets evaluated.



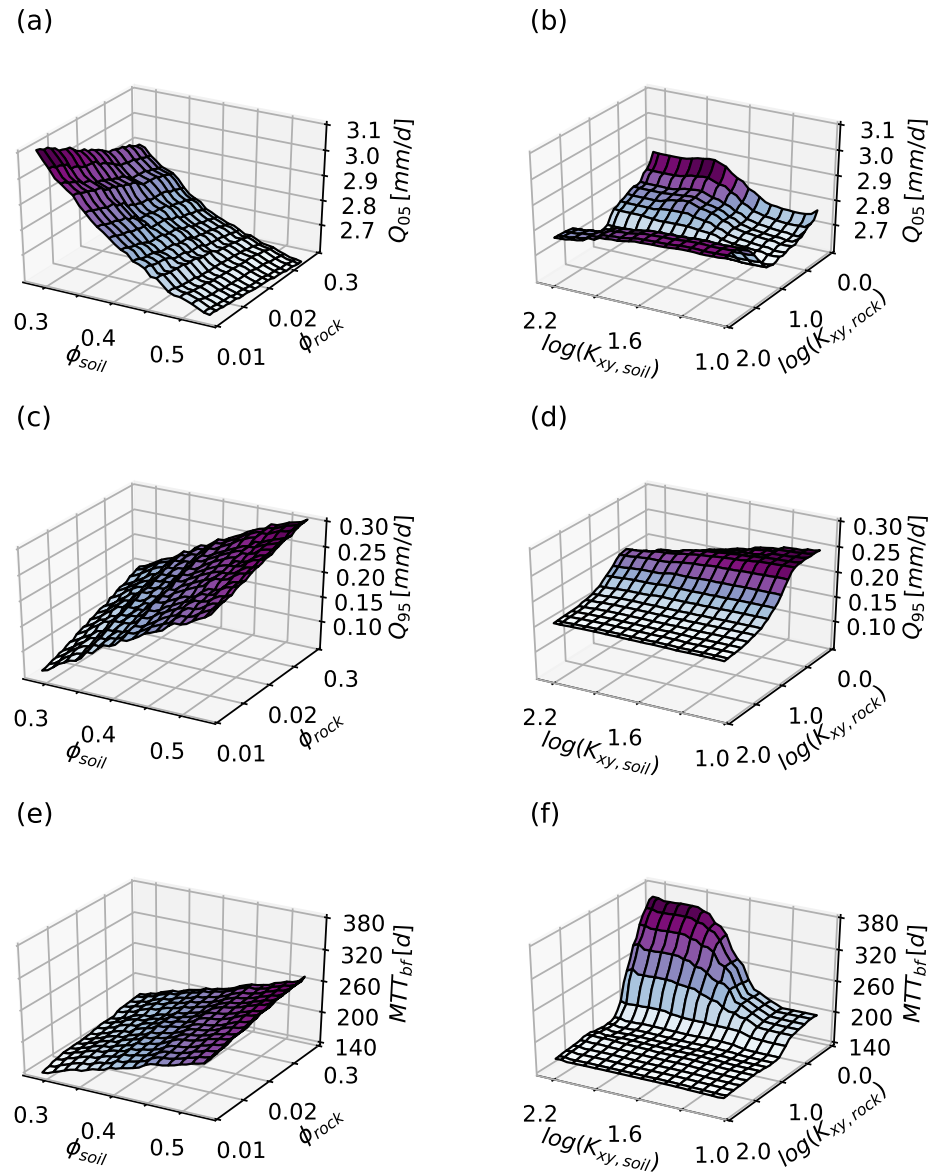
Compared to porosity, changes in the lateral soil hydrologic conductivity  $K_{xy,soil}$  over an order of magnitude had a relatively small global effect on the response variables (Figure 3.6(e)). A 20-fold increase in lateral hydrologic conductivity tended to neither increase high flows  $Q_5$ , decrease low flows  $Q_{95}$ , or lower median baseflow transit times  $MTT_{bf}$ . In fact, increases in  $K_{xy,soil}$  were associated with a small but counterintuitive increase in the median transit time for baseflow to reach the stream. Related to this, there was an abrupt transition when  $\log(K_{xy,soil})$  increased above 1.5 in the number of parameter sets with  $MTT_{bf}$  greater than 300 days (see blue point cloud in bottom panel of Figure 3.6(e)). This is indicative of possible threshold-like interactions with other parameters.

Finally, varying the lateral fractured rock hydrologic conductivity  $K_{xy,rock}$  induced the most complex and non-linear response in the model (Figure 3.6(d)). The peak flows  $Q_5$  had local maxima at both the extreme low end (i.e., slower draining fractured rock) and high end (i.e., faster draining fractured rock) of the assigned range of  $K_{xy,rock}$ . Thus, increases in  $K_{xy,rock}$  tended to both increase and decrease peak flows, depending on whether  $\log(K_{xy,rock})$  was above or below a threshold around 1.0. Low flows  $Q_{95}$  had a similar v-shaped partial dependency on  $K_{xy,rock}$ , though in the opposite direction. Finally, when  $\log(K_{xy,rock})$  was below a threshold of approximately 1.0, increases in  $K_{xy,rock}$  were associated with decreases in  $MTT_{bf}$ . Above this threshold, however, the value of  $MTT_{bf}$  had no partial dependency on  $K_{xy,rock}$  over a 10-fold increase.

### 3.7.4 BRT sensitivity analysis - Parameter interactions

The results from the single-parameter partial dependency plots indicate that potentially significant, non-linear parameter interaction may explain some variability in the response variables. BRT allows us to explore non-linear parameter interaction through two-parameter partial dependency plots. Several examples are shown in Figure 3.7 and 3.8.

Beginning with porosity in Figure 3.7(a,c,e), the results show very little interaction between  $\phi_{soil}$  and  $\phi_{rock}$  across all response variables. The surfaces are planer, indicating that (1) the partial dependency of the response variable on each parameter is linear (as also seen in the single-parameter plots) and (2) the influence of one parameter on the response variable is not modulated by the value of the other.



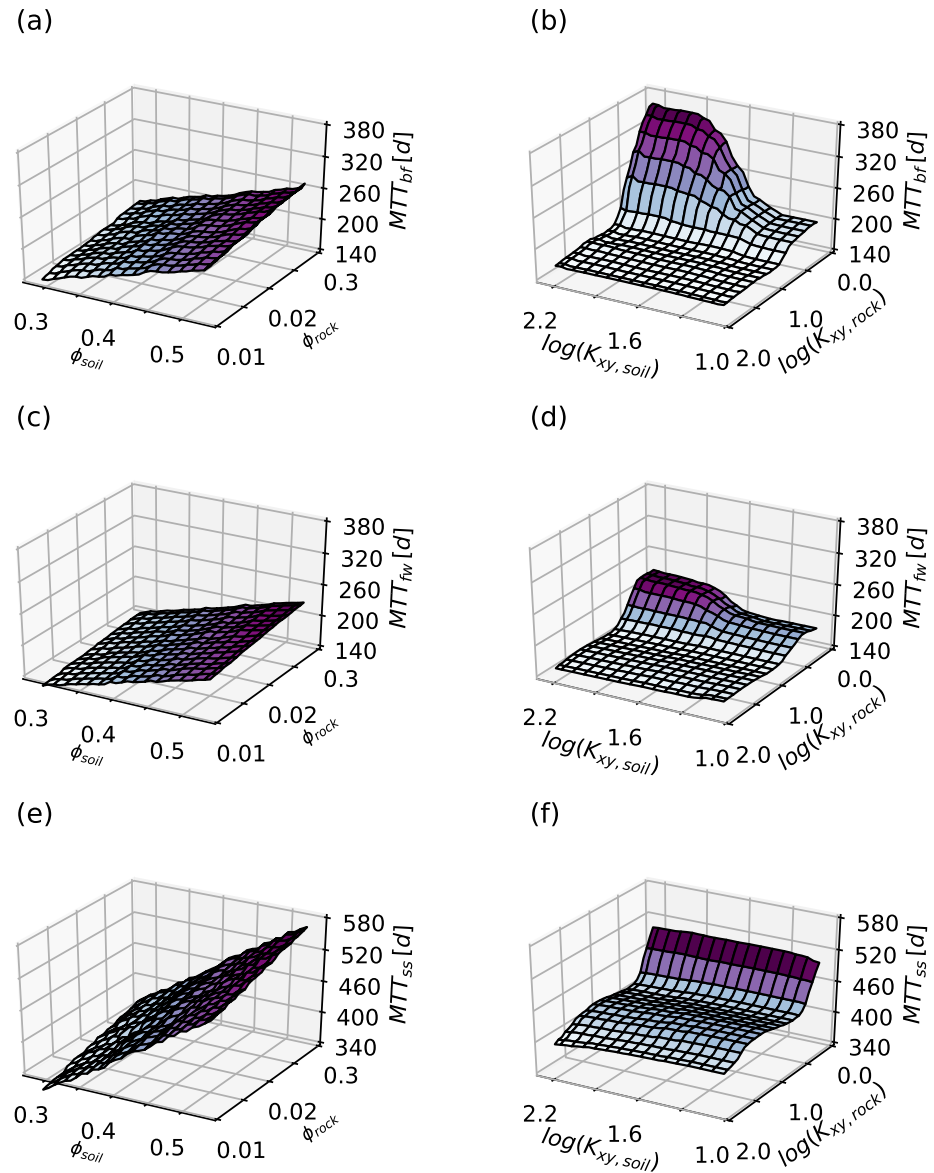
**Figure 3.7:** The two-parameter partial dependency plots for the porosity parameters (left columns) and the log of lateral hydrologic conductivity parameters (right columns). The response variables are high flows (upper row), low flows (middle row), and median baseflow age (bottom row).

By contrast, Figures 3.7(b,d,f) shows a strong interaction between the lateral hydrologic conductivity multiplier for upper soil layer ( $K_{xy,soil}$ ) and lower rock layer ( $K_{xy,rock}$ ). High flow  $Q_{05}$  is greatest, low flow  $Q_{95}$  is lowest, and median travel time  $MTT_{bf}$  is highest when  $K_{xy,soil}$  is at its upper limit and  $K_{xy,rock}$  at its lower limit. In addition, the response variables are most sensitive to changes in  $K_{xy,soil}$  when  $K_{xy,rock}$  is low. Some possible physical explanations for these interactions are discussed in Section 3.8.

The interaction between  $K_{xy,rock}$  and  $K_{xy,soil}$  had an especially dramatic influence on the median baseflow transit time  $MTT_{bf}$  (Figures 3.7(f)). Further investigations suggest that the strength of their interaction is at least partly related to (1) the focus on baseflow and (2) the transience of the system. Figure 3.8 compared the influence of  $K_{xy,rock}$  and  $K_{xy,soil}$  on three different representations of catchment transit time: the median baseflow transit time  $MTT_{bf}$ , the median flow-weighted transit time  $MTT_{fw}$ , and the median steady-state transit time  $MTT_{ss}$ . Compared to  $MTT_{bf}$ , the overall interaction was much smaller for  $MTT_{fw}$  (Figure 3.8d)), suggesting that that the interaction is less important in other stages of the hydrograph (e.g., peak flow). Furthermore, there was almost no interaction affecting  $MTT_{ss}$  (Figure 3.8(f)), suggesting that the conditions leading to the interaction were related to the intermittency of rainfall.

### 3.7.5 2D versus 3D performance

Recalling that a goal of the calibration was to parameterize the 3D catchment model, one of the behavioral parameter sets from our 2D emulator was used to simulate the full 3D catchment hydrology (shown in Figure 3.1). Overall, simulated daily discharge in the 3D catchment and 2D emulator were nearly identical with a KGE of 0.96 (Figure 3.9). The 2D model slightly overpredicted low flows. The simulated values of  $MTT_{bf}$  were more different but also relatively similar at 251 days for the 2D emulator and 321 days for the fully 3D model. The differences may be due to longer, indirect groundwater and interflow pathways in the 3D model that run both parallel and perpendicular to the stream channel, and larger contributing areas away from the stream due to convergence. Despite some differences, the results suggest that the 2D emulator was a reasonable approach to calibration under the given constraints. The 3D simulation is described in more detail in Chapter 4.



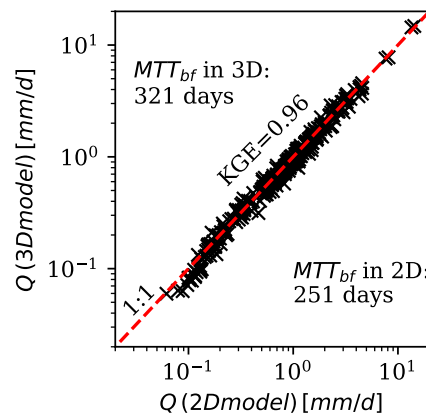
**Figure 3.8:** The two-parameter partial dependency plots for the porosity parameters (left columns) and the log of lateral hydrologic conductivity parameters (right columns). The response variables are median baseflow travel times (upper row), median flow-weighted travel times (middle row), and median steady state travel times (bottom row)

## 3.8 Discussion and Conclusions

The results are discussed in the context of the three main research objectives: to demonstrate an efficient approach to calibrating an ISSHM to discharge and age data, to assess the advantages of including transit times in the calibration criteria, and to investigate the global sensitivity of discharge and transit times to model parameters and what it suggests about catchment processes.

### 3.8.1 Implications for ISSHM calibration

This work demonstrates a computationally efficient technique to calibrate a 3D ISSHM model using several novel techniques. First, the full 3D catchment was represented with a 2D emulator. A post-calibration analysis showed that the 2D emulator faithfully reproduced the hydrograph of the 3D model, but moderately underestimated the 3D median baseflow age. The approach may prove appropriate for other catchments with similar topography. Second, a periodic steady-state simulation was shown to make efficient estimates of the median transit time distribution of baseflow for over 20,000 model realizations. A simple approach to simulating baseflow transit times under transient conditions may be especially useful for modeling field data, because studies of stream transit times have traditionally focused on baseflow measurements. Third, the study illustrates how



**Figure 3.9:** Comparison of 2D emulator and full 3D catchment simulation.

GLUE calibration can be applied to a highly parameterized ISSHM to fit both discharge and transit time data. Overall model performance for discharge was consistent with the small number of other studies using ISSHMs over multiple seasons at catchment scale. Further, this is the only study to our knowledge to simulate the contributions of deep fractured rock to streamflow in a non-snowmelt dominated system. No direct comparison can be made related to the performance of the model to simulate observed catchment transit times, which has not considered in other catchment-scale, transient ISSHM simulations. Finally, we used for the first time Monte Carlo filtering and boosted regression trees with a large ensemble of ISSHM model realizations to understand the global sensitivity of model parameters, which is discussed in the next sections. Overall, our approach to calibration and sensitivity analysis can be generalized and applied to other ISSHMs and in other settings.

### 3.8.2 Controls on catchment discharge

For our modeling conditions and among the parameter ranges evaluated, catchment discharge was most sensitive to porosity in the upper soil and lateral hydrologic conductivity. The influence of these parameters on low flows and high flow and potential mechanistic explanations are discussed below. These results give insight into catchment processes and can help inform the manual calibration and interpretation of other ISSHM models.

Higher peak flows were associated with lower porosities, presumably due to faster pore velocities and more frequent overland flow. The relationship between peak flow and the lateral fractured hydrologic conductivity of bedrock  $K_{xy,rock}$  was highly non-linear. Peak flows were relatively high when  $K_{xy,rock}$  was high, presumably due to higher pore velocities. However, peak flows were also high when  $K_{xy,rock}$  was low, potentially due to limited infiltration causing saturation excess overland flow. When  $K_{xy,rock}$  was between these extreme, peak flow was lowest, potentially because lateral flow through bedrock was too fast to promote surface saturation but too slow to concentrate stormflow at the outlet. Finally, when  $K_{xy,rock}$  was high, peak flow was relatively insensitive to changes in the lateral hydrologic conductivity of the upper soil  $K_{xy,soil}$ , presumably because most flow through the upper soil layer was vertical infiltration. When  $K_{xy,rock}$  was low and infiltration was minimal, however, increases in  $K_{xy,soil}$  were associated with increased in peak flow, perhaps due to increased perching and faster lateral flow through the upper soils. In general, combinations of parameters

that promoted higher peak flows also promoted lower baseflow, and vice versa, according to the mechanisms already described (but from the perspective of low flows). Note, however, that the change in the absolute value of  $Q_{95}$  was much smaller.

Neither the discharge nor the transit times (discussed below) were very sensitive to the other calibrated parameters including: the porosity of the fractured bedrock, the vertical hydrologic conductivity of the streambed and fragipan, and the e-folding depth of the fractured bedrock hydrologic conductivity. It is important to note, however, that the parameters included and the ranges considered were based on an *a priori* literature review and assessment, which was itself subject to considerable uncertainty. Also, the sensitivity analyses were global and did not consider interactions between three parameters or more. Therefore, the actual local sensitivity may vary at any specific parameter set.

### 3.8.3 Controls on baseflow transit times

High median baseflow transit times were associated with the combination of relatively high lateral hydrologic conductivity in the upper soil  $K_{xy,soil}$  and low hydrologic conductivity in the lower fractured bedrock  $K_{xy,rock}$ . This may be driven by three phenomena. First, low values of  $K_{xy,rock}$  promote perched aquifers that move water laterally through the upper soil later. Second, high values of  $K_{xy,soil}$  promote rapid lateral flows in the upper soil that drain too quickly to contribute to baseflow. Third, the relatively small amount of water that does infiltrate takes a long to reach the outlet due to the low  $K_{xy,rock}$ . The results of our transient simulations suggest that the hydrologic properties of the upper soil can have a large influence of the age of baseflow. By contrast, our analysis using the steady-state median transit times shows that the hydrologic conductivity of the upper soil has almost no effect on the median transit time. Although steady-state transit time modeling is common practice, this apparent tendency of the steady-state assumption to misrepresent the importance of upper soil hydrologic conductivity is not well appreciated in the relevant literature reviews (e.g., McGuire and McDonnell, 2006; McDonnell et al., 2010; Hrachowitz et al., 2016).

### 3.8.4 Interactions between celerity and velocity

Recent papers have called for the combined evaluation of watershed models against catchment celerities (i.e., the hydrograph based on volumetric water discharge) and velocities (e.g., the travel time distribution based on tracer movement) (McDonnell and Beven, 2014). The idea that watershed models can simulate hydrology correctly but travel times poorly is largely based on virtual experimentation using lumped parameter models. This work provides insight into the importance of evaluating both celerities and velocities using process-based ISSHMs. Specifically, we found that the inclusion of travel times in the calibration criteria significantly changed the composition of behavioral parameter sets. Although a larger range of  $K_{xy,soil}$  values could reproduce the discharge data, only relatively high values of  $K_{xy,soil}$  were able to generate the transit times inferred from the data. It is worth noting, however, that no parameter emerged as especially sensitive "knobs" to turn for isolated tuning of either discharge or travel times: rather, the most sensitive parameters were associated with changes in both.

### 3.8.5 Limitations and future work

The scope of our analysis was necessarily limited by the complexity of conducting integrated surface subsurface hydrologic modeling. Limitations that might be addressed in future work include a higher modeling spatial resolution especially in the upper soil layer and near the streambed, consideration of a wider range of parameter values and additional parameters, inclusion of heterogeneity in the hydraulic field, evaluation of the full transit time distribution, and calibration of other model state variables against data including water table level and soil moisture. Notwithstanding the potential for further improvements, the calibrated model shows promise for simulation of non-point source pollution transport at catchment scale.



## Chapter 4

Using flowpath decomposition to understand time-varying transit time distributions and storage selection with an integrated surface-groundwater model of a small watershed. <sup>1</sup>

### 4.1 Highlights

- Shape of catchment transit time distributions explained by examining constituent flowpaths in integrated hydrology model.
- Catchment "inverse storage effect" driven by increases in overland flow, which itself showed direct storage effect.
- Stream TTs exhibited middle-aged depression not captured by typical fitting distributions.

### 4.2 Summary

Catchment studies make inferences about flowpaths using tracer-derived estimates of transit time distributions (TTDs) and, more recently, using StorAge Selection (SAS) functions at the largest (most integrated) scale of study. These inferences may be highly uncertain, because few studies have rigorously examined the mapping between flowpaths and age-based distributions in a realistic testbed. Here we use the physically-based, fully coupled surface-groundwater model ParFlow-CLM

---

<sup>1</sup>This chapter will be submitted for publication in a peer-reviewed journal. Ciaran Harman and Bill Ball were involved in hypothesis development, study design, and results interpretation. Reed Maxwell provided technical assistance with ParFlow-CLM and SLIM-FAST simulations. Anthony Buda provided data and interpretation of data from the study site. Dano Wilusz participated in all of the above and also produced the first draft of text and all figures.

with a modified SLIM-FAST particle tracking scheme to study how different flowpaths contribute to the time-varying catchment TTD and SAS functions of discharge ( $Q$ ) and evapotranspiration ( $ET$ ). The model was parameterized and forced with one year of field data from the USDA's Mahantango Creek experimental catchment in PA, USA. Model evaluation showed reasonable consistency with observed daily discharge ( $KGE = 0.62$ ) and available tracer-derived water ages. We find that the time-variability of the TTD for  $ET$  results from seasonally varying contributions of ground evaporation and transpiration drawn from constant distributions of age-ranked storage, which can be analytically derived. The time-variability of the TTD for  $Q$  is more complex and evolves from shifting contributions of five flowpaths: direct rainfall, overland flow, interflow, and shallow and deep groundwater. The transit times for different flowpaths overlap substantially, suggesting that age alone cannot uniquely identify flowpath. As observed elsewhere,  $Q$  contains more of the youngest water in storage when catchment storage is high (an "inverse storage effect"). This occurs due to a non-linear increase in relatively young overland flow when the catchment is wet, and despite the fact that overland flow and other pathways tend to become older when storage is high (a "direct storage effect"). The TTD and SAS function exhibit a bimodal distribution of younger and older water, resulting in a "middle-aged depression" that cannot be captured with common unimodal fitting distributions (e.g., gamma, beta). We develop a taxonomy of TTD shapes that occur during a typical recession, each with a distinct composition of flowpaths. Finally, we discuss the implications of our results on the conceptualization of lumped parameter models, the physical origins of SAS functions and their variability, and the origins of different C-Q relationships.

### 4.3 Introduction

Despite ongoing progress in the hydrologic sciences (e.g., Hrachowitz and Savenije, 2013), we still have a poor understanding of the pathways and velocities of water moving through catchments (McDonnell and Beven, 2014), to the detriment of effective water resources management (Hrachowitz et al., 2016). The catchment transit time distribution (TTD) has proven to be a useful, spatially-integrated measure of catchment velocities that gives insight into catchment flowpaths (McGuire and McDonnell, 2006; McDonnell et al., 2010). The TTD is the time-varying distribution of water ages at a catchment outflux, where water age is defined as the time since entry into the catchment. Research to describe

and predict catchment TTDs has been advancing in two parallel tracks related to (a) integrated surface/subsurface hydrology modeling (ISSHMs) and (b) StorAge Selection (SAS) function theory.

One research track focuses on the simulation of TTDs using ISSHMs, which are physically-based, spatially-distributed watershed models with growing capability to realistically simulate catchment processes at catchment scale (Maxwell et al., 2014; Kollet et al., 2017). ISSHMs are numerical, mesh or grid-based models designed to simulate major hydrologic processes including groundwater flow, overland flow, and their interaction (i.e., infiltration and exfiltration). ISSHMs can be coupled with land-atmosphere models to simulate snowfall and snowmelt, canopy process, and spatially-varying evapotranspiration. ISSHMs including ParFlow (Kollet and Maxwell, 2006), CATHY (Camporese et al., 2010), and HydroGeoSphere (Brunner and Simmons, 2012) have been used in a host of applications (Paniconi and Putti, 2015; Fatichi et al., 2016), and can be integrated with particle tracking algorithms to map flow paths and velocities through the critical zone and calculate TTDs. To date, however, relatively few studies have used ISSHMs to simulate catchment TTDs. A review of previous studies and key findings is in Chapter 2. Among these studies, most are in idealized and/or steady-state domains, and catchment-scale studies with time-varying forcing are extremely rare. Limitations of ISSHMs include large data requirements, high computational expense, potentially significant model structure uncertainties, and a steep learning curve for users (Clark et al., 2017).

The other TTD-related research track focuses on simulation of TTDs using SAS functions in a lumped modeling framework, where SAS function parameters are inferred from tracer data and catchment characteristics (Rinaldo et al., 2015). Broadly speaking, a SAS function describes the relative proportion of younger and older storage that contributes to the outflux of a control volume in a lumped modeling framework. Because the age of storage in each control volume can be tracked over time, the SAS function can be used to estimate the TTD of outflux. Notwithstanding the growing number of catchment-scale studies using SAS functions to simulate catchment TTDs (e.g., van der Velde et al., 2014; Harman, 2015; Benettin et al., 2017; Wilusz, Harman, and Ball, 2017; Rodriguez, McGuire, and Klaus, 2018), the physical origins of SAS functions are poorly understood, and most studies rely on inverse modeling with tracer data to infer their shape. Although SAS function behavior can be explained in laboratory experiments (Kim et al., 2016; Pangle et al., 2017), more work is needed to identify factors controlling the shape and time-variability at catchment

scale (Rinaldo et al., 2015). While previous research has tried with limited success to isolate factors affecting TTD functions (e.g. McGuire and McDonnell, 2006), SAS functions may be less sensitive to "external variability" from unsteady flows (Kim et al., 2016) and thus easier to parameterize. To the extent that the time-varying shape of SAS functions can be explained and predicted, they may be an efficient way to simulate catchment TTDs in data-scarce regions.

Although these two research tracks have a common purpose, they are not in competition. Rather, previous work has recognized a deep synergy between the development of ISSHMs and SAS function theory. On one hand, ISSHMs can be used to interrogate the relationship between catchment characteristics and SAS functional form in controlled modeling experiments. For example, Kim et al. (2016) used the ISSHM CATHY to describe the physical origins of SAS functions estimated in a lab-scale lysimeter tracer experiment. Danesh-Yazdi et al. (2018) used the ISSHM ParFlow to show how that heterogeneity in hydrologic conductivity added significant variability and bias to the shape of the SAS function. On the other hand, SAS functions can facilitate synthesis and visualization of potentially copious ISSHM output (e.g., Yang et al., 2018). In addition, SAS functions derived from available data could be a useful signature for ISSHM model calibration and evaluation, just as flow duration curves are a useful signature for calibrating the catchment hydrograph (e.g., Jothityangkoon, Sivapalan, and Farmer, 2001). Following the arc of previous work, this paper presents complementary advances related to both ISSHMs and SAS function theory.

The first goal of this study was to use an ISSHM to describe event-scale catchment TTDs in a hydrologically complex watershed with year-round precipitation, significant snowfall, high summer evapotranspiration (ET), and a heterogeneous geology including a shallow riparian fragipan and deep fractured bedrock. Our study site is the extensively studied FD-36 subcatchment of the U.S. Department of Agriculture (USDA) Mahantango Creek experimental catchment near Harrisburg, PA, USA (Bryant et al., 2011). The catchment hydrology was simulated for calendar year 2014 using the open-source ISSHM ParFlow-CLM with SLIM-FAST particle tracking. The particle code was enhanced to support sub-daily tracking of both discharge  $Q$  and  $ET$ , as well as flowpath characterization (discussed below). The model was parameterized using field data and calibration to daily discharge and available tracer-derived water ages, which is detailed in Chapter 3. To interpret our results, model outputs were visualized as TTDs, SAS functions, and the recently proposed

age-rank storage-discharge relationship [Harman, *submitted*]. Our modeling approach has similarities with two recent publications (Yang et al., 2018; Remondi et al., 2018), but in a significantly different hydrological setting and using an open-source ISSHM that is well-documented with a large user base (Maxwell et al., 2016).

To better understand the simulated TTDs and SAS functions, we applied a novel approach to deconstructing the  $Q$  and  $ET$  outfluxes and analyzing their constituent sub-components (hereafter labeled "flowpaths"). This approach is premised on the idea that each outflux (e.g.,  $Q$ ) is comprised of several constituent flowpaths (e.g., overland flow, interflow, groundwater flow), and the TTD and SAS function of each flowpath may have a unique shape, time-variability, and relationship with catchment characteristics. While the behavior of the aggregated outflux TTD and SAS function may seem chaotic, the behavior of individual flowpaths may be easier to interpret. Compared to the analysis of complete flowlines through the catchment - which is also feasible using ISSHMs - the analysis of flowpaths is simpler to operationalize and (looking forward) potentially more comparable across watersheds. Another advantage of analyzing flowpaths is that modeled TTD and SAS functions may be usefully commensurate with certain field measurements, including estimates of soil water age (e.g., McGuire, DeWalle, and Gburek, 2002) and groundwater age (e.g., Lindsey et al., 2003). Although this is the first flowpath-based analysis using ISSHMs, the approach is inspired by the common practice of partitioning discharge into flowpaths in end-member mixing analysis (e.g., Buttle, 1994; Hooper, Christophersen, and Peters, 1990) and lumped parameter models (e.g., Dunn, McDonnell, and Vaché, 2007; Birkel, Soulsby, and Tetzlaff, 2011; Hrachowitz et al., 2013; Benettin et al., 2015), which has been shown to improve model performance (e.g., McMillan et al., 2012).

The second goal of this study was to use our ISSHM modeling framework to explore two particular questions about the physical origins of SAS functions. For the first question, we explore why some catchments seem to route more of the youngest water in storage to discharge when the catchment is wet, a phenomenon dubbed the "inverse storage effect" (Harman, 2015). The inverse storage effect is associated with a particular kind of SAS function, which has been observed in many catchments (e.g., Harman, 2015; Wilusz, Harman, and Ball, 2017; Rodriguez, McGuire, and Klaus, 2018). Studies have attributed the inverse storage effect to the activation of fast flowpaths when storage is high (e.g., Kim et al., 2016; Soulsby et al., 2015), but the precise mechanism has not been modeled or

observed. We use our ISSHM to determine whether the SAS function in our study site indicates an inverse storage effect and, if so, how it arises from the flowpaths of discharge. For the second question, we examine an assumption that is sometimes implicit in the interpretation of TTD and SAS functions: that there is a one-to-one relationship between the apparent mean age of water in outflux and the flowpath that produced it. Perceptual models (e.g., Sanford and Pope, 2013) and steady-state 2D analytical solutions to the catchment flow field (e.g., Małozewski and Zuber, 1982) tend to show a very clear, one-to-one relationship between the age of water and how it contributes to stream discharge (i.e., younger water arrives from overland flow, middle-aged water from interflow, and older water from groundwater). Here we use the ISSHM modeling framework to describe and benchmark the nature of this mapping between water age and flowpath in a 3D model with complex topographies, unsteady flows, and interacting processes. In addition to informing the interpretation of TTD and SAS functions in other catchments, the analysis is a step towards relating water age to flow pathways, which could inform transport modeling based on source exposure and contact time (e.g., Benettin et al., 2015).

Based on these two goals, the objectives of the study can be summarized as (a) to use a physically-based hydrology model to calculate and explain the TTD of  $Q$  and  $ET$  and their respective flowpaths in a small experimental catchment, (b) to test whether analysis of flowpaths can elucidate the physical origins of age-based distributions for  $Q$  and  $ET$ , including any storage effect, and (c) to examine the null hypothesis that there is a one-to-one mapping between the age of outfluxes and the flowpath that produced it.

The remainder of the paper is organized as follows. Section 4.4 provides background theory. Section 4.5 introduces the methods including the study site, the model setup and calibration, the particle tracking routines, and the TTD and SAS function calculation. Section 4.6 describes the results, and Section 4.7 has discussion and conclusions including the implications of our results on the hydrologic behavior of our study site, the interpretation of TTD and SAS functions at other sites, the conceptualization of lumped parameter models, and the origins of different concentration-discharge relationships.

## 4.4 Background theory: age-based distributions

This section gives a brief comparative description of the TTD, the SAS function, and the age-ranked storage discharge function relationship, collectively called "age-based distributions". In addition to providing essential background, the section shows how the distributions can be deconstructed into their constituent pathways. Then, the remainder of the paper applies this deconstruction to answer questions about the nature of age-based distributions at our study site.

### 4.4.1 Transit time distributions (TTDs)

We begin by conceptualizing water moving through the watershed as a collection of discrete water parcels or particles. Each particle represents a mass and volume of water moving from injection (e.g., as rainfall) to some outflux (e.g., as  $Q$  or  $ET$ ). The flowpath and velocity of the particles can be described by two types of transit time distributions. The forward transit time distributions (TTDs)  $\overrightarrow{P}_Q(T, t_i)$  and  $\overrightarrow{P}_{ET}(T, t_i)$  are the cumulative probability distribution functions (CDFs) of transit times (or, equivalently, exit ages) of all particles entering the catchment at time  $t_i$  and exiting after time  $T$  as  $Q$  and  $ET$ , respectively. The backward TTDs  $\overleftarrow{P}_Q(T, t)$  and  $\overleftarrow{P}_{ET}(T, t)$  are the CDFs of transit times  $T$  of particles exiting the catchment at time  $t$ .

TTDs are described as a master variable linking catchment hydrology and contaminant transport (Hrachowitz et al., 2016), which can be estimated from tracer experiments and numerical models. As summarized in review papers including McGuire and McDonnell (2006) and McDonnell et al. (2010), the TTD provides an integrated window into transport through the subsurface, which cannot be observed directly. In addition, the backward TTD can be convolved with hydrochemical inputs to simulate contaminant transport (Barnes and Bonell, 1996). In natural systems, however, the backward TTD is highly dynamic and sensitive to the history of inputs (e.g., Wilusz, Harman, and Ball, 2017), making it more difficult to characterize and predict.

To better understand the shape and time-variability of the TTD, we propose its disaggregation into component parts. Consider the case where total discharge  $Q$  is composed of  $n$  different flowpaths labeled  $i$ :

$$Q(t) = \sum_{i=1}^n Q_i(t) \quad (4.1)$$

We can define the quantity age-ranked discharge  $Q_T = Q_T(T, t)$  as the flux in discharge with age less than or equal to  $T$ . It follows from the definition of the backward TTD that:

$$Q_T(T, t) = Q(t) \overleftarrow{P}_Q(T, t) \quad (4.2)$$

The conservation of age and mass requires that the flux of total discharge younger than age  $T$  must equal the total contribution younger than age  $T$  from each pathway:

$$Q_T(T, t) = \sum_{i=1}^n Q_{T,i}(T, t) \quad (4.3)$$

Combining Equations 4.1-4.3 and rearranging to solve for  $\overleftarrow{P}_Q(T, t)$  gives:

$$\overleftarrow{P}_Q(T, t) = \frac{\sum_{i=1}^n Q_i(t) \overleftarrow{P}_{Q,i}(T, t)}{Q(t)} \quad (4.4)$$

Equation 4.4 shows how information about fluxes and backward TTDs of individual pathways can be used to understand the emergent behavior of the backward TTD of total discharge. Specifically, the backward TTD for total discharge is simply the flow-weighted average of the backward TTD of its component pathways. The derivation can be generalized to any outflux including  $ET$ .

#### 4.4.2 Storage Selection (SAS) functions

The aforementioned complexity and time-variability of catchment backward TTDs may make them difficult to analyze and model. The SAS function is a transformed representation of the backward TTD that reflects the dynamic ages of water in storage, and may be easier to interpret and model. SAS function theory was developed in a series of papers (Botter, Bertuzzo, and Rinaldo, 2011; van der Velde et al., 2012; Harman, 2015), are synthesized in Rinaldo et al. (2015), and reviewed here briefly.

We define the age-rank storage  $S_T(T, t)$  as the amount of water in catchment storage with age younger than or equal to  $T$  at time  $t$ :

$$S_T(T, t) = S(t) P_S(T, t) \quad (4.5)$$



where  $S(t)$  is total catchment storage and  $P_S(T, t)$  is the cumulative residence time distribution, defined as the cumulative fraction of water in catchment storage with age younger than  $T$  at time  $t$ . Note that, in the case study presented,  $S(t)$  is determined by our modeling. Typically, however,  $S(t)$  is not known and must be set relative to some reference state. As long as discontinuities are appropriately addressed (e.g., Harman, 2015), then the relationship between  $S_T$  and  $T$  is invertible and one maps onto the other. As a result, the backwards TTD for  $Q$  can be re-expressed as a distribution called the SAS function:

$$\Omega_Q(S_T, t) = \overleftarrow{P}_Q(T, t) \quad (4.6)$$

where the SAS function  $\Omega_Q(S_T, t)$  is the cumulative distribution of age-ranked storage contributing to discharge  $Q$ . As with the backward TTD, the SAS function is the flow-weighted average of the SAS functions of its constituent pathways:

$$\Omega_Q(S_T, t) = \frac{\sum_{i=1}^n Q_i(t) \Omega_{Q,i}(S_T, t)}{\sum_{i=1}^n Q_i(t)} \quad (4.7)$$

which follows by substituting 4.6 into 4.4. The SAS function for  $ET$  denoted  $\Omega_{ET}(S_T, t)$  can be similarly defined.

### 4.4.3 Age-rank storage-flux (rSF) functions

Harman [*submitted*] introduced a new approach for describing and analyzing the time-varying transit time dynamics of catchments, referred to here as age-rank Storage-Flux (rSF) functions. For any flux exiting a catchment (e.g., discharge,  $ET$ ), the rSF function describes which volume of age-rank storage contributes to the youngest flux, the oldest flux, and all ages of flux in between. The rSF function is similar to the SAS function, but it is not a probability distribution (i.e., there is no normalization to total flux) and the axes are transformed, such that the rSF function can be plotted alongside (and indeed encompasses) the canonical storage-discharge relationship. The plotting of the rSF function may be useful for several reasons, which are elaborated upon below.

We begin by describing the rSF function for total discharge  $Q$ , which is easily generalized to other fluxes including  $ET$ . The rSF function  $f_{T,Q}$  is the age-rank discharge complement  $\overline{Q}_T$  as a function

of the age-rank storage complement  $\bar{S}_T$ :

$$f_{T,Q}(\bar{S}_T, t) = \bar{Q}_T(\bar{S}_T, t) \quad (4.8)$$

where:

$$\bar{S}_T(T, t) = S(t) - S_T(T, t) \quad (4.9)$$

$$\bar{Q}_T(T, t) = Q(t) - Q_T(T, t) \quad (4.10)$$

and  $S_T$  and  $Q_T$  are defined in Equations 4.2 and 4.5. The age-ranked storage complement  $\bar{S}_T(T, t)$  has units of [L] and represents the amount of water in storage that is older than age  $T$  at time  $t$ .  $\bar{Q}_T(T, t)$  has units of [L/T] and represents the flux in discharge that is older than age  $T$  at time  $t$ . Both  $\bar{S}_T$  and  $\bar{Q}_T$  are mathematical complements of  $S_T$  and  $Q_T$  in the sense that they sum to total storage  $S$  and total discharge  $Q$ , respectively. Equations 4.2, 4.6, 4.8, and 4.9 can be combined to give:

$$f_{T,Q}(\bar{S}_T, t) = Q(t)(1 - \Omega_Q(S(t) - \bar{S}_T, t)) \quad (4.11)$$

which shows that the rSF function is the product of discharge and the cumulative exceedance probability of discharge expressed in terms of age-ranked storage complement. To give an example, a rSF function value of  $f_{T,Q}(\bar{S}_T = 300 \text{ mm}, t = 100 \text{ days}) = 1.0 \text{ mm/d}$  indicates that the oldest 300 mm of storage contributed to the oldest 1.0 mm/d of discharge on day 100.

The rSF function for each pathway  $i$  of  $Q$  can be derived in a similar fashion and denoted  $f_{T,Q,i}(\bar{S}_T, t)$ . Following the logic used to derive Equation 4.3, the rSF function for discharge  $Q$  is simply the sum of the rSF function for each of its constituent pathways:

$$f_{T,Q}(\bar{S}_T, t) = \sum_{i=1}^n f_{T,Q,i}(\bar{S}_T, t) \quad (4.12)$$

The rSF function for  $ET$  denoted  $f_{T,ET}(\bar{S}_T, t)$  and its constituent pathways  $f_{T,ET,i}(\bar{S}_T, t)$  can be derived in an analogous manner.

The rSF function has several advantages over SAS functions for visualizing and interpreting age-related catchment dynamics. First, unlike the TTD and SAS function, the rSF function for the total flux is the simple arithmetic sum of its constituent pathways. Thus, simple visual inspection can

give insight into the relative flux and age contributions of each pathway to the total fluxes. Second, the use of age-ranked storage complement  $\bar{S}_T$  means that increases in  $S(t)$  due to influxes of new water will not necessarily shift the plotting position of older  $\bar{S}_T$  values on the x-axis. As a result, the rSF is a better tool for analyzing the relationship between the age of water in outfluxes and the oldest volumes of water in storage. Finally, the rSF function explicitly shows the storage-discharge relationship, which is a reflection of catchment celerities, plotted alongside a representation of the SAS function, which is a reflection of catchment velocities. As a result, the rSF functions are a promising approach to describing and parameterizing catchment celerity and velocity in a unified, lumped modeling framework.

## 4.5 Materials and methods

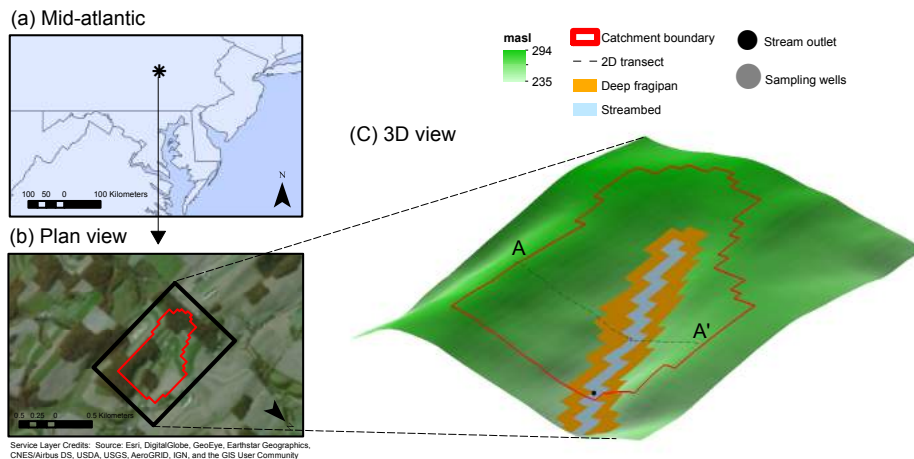
### 4.5.1 Study site

Our study site was the FD-36 subcatchment of the USDA Mahantango Creek experimental catchment near in central PA, USA (see Figure 4.1). The 0.4km<sup>2</sup> agricultural headwater basin is situated within the larger 7.8km<sup>2</sup> WE-38 experimental catchment, which drains to the Susquehanna River and discharges into the Chesapeake Bay estuary. Regional runoff contributes a significant load of nutrients and sediment to the bay, which impairs coastal water quality and has been targeted for management action (Linker et al., 2013a). The most pertinent catchment details are summarized below, and described in more detail in Chapter 3 and the references therein.

FD-36 is situated in the Valley and Ridge physiographic province, underlain by shale, siltstones, and sandstones (Urban, 1977). The climate is humid and temperate with mean annual rainfall of 1080mm, of which roughly 10% falls as snow and 50% leaves as evapotranspiration (Bryant et al., 2011). Landuse is 70% cropland and pasture and 30% wooded (Needelman et al., 2004; Bryant et al., 2011). Previous research suggests that catchment boundaries are well-defined by surface topography (i.e., interbasin groundwater exchange is minimal) (Gburek, 1977). A combination of SSURGO soil surveys and synoptic field measurements (Needelman et al., 2004) suggest the geology can be classified into three major units (see Figure 4.1 and 4.2). Except for near the riparian area, the upper 1-2 meters is well-draining, residual, fine-loamy soil. The lower 1m of soil in the riparian zone

is a confining layer of poorly draining, clay-rich, colluvial fragipan. Finally, beneath the soil and fragipan is up to 60m of fractured bedrock, with low porosity and exponential decline in hydrologic conductivity (Gburek and Folmar, 1999).

The site is well-instrumented and extensively studied, which facilitated our 1-year simulation of 2014 conditions. Unless otherwise noted, data was provided by USDA staff [*Anthony Buda, personal communication*]. Discharge was measured by a flume at the outlet (see location in Figure 4.1) and hourly rainfall was measured at a rain gauge approximately 0.8km southeast of the catchment outlet (Buda et al., 2011). Hourly temperature, wind velocity, downward solar radiation, relative humidity, and pressure were measured by the Natural Resource Conservation Service approximately 8km west of the study site. Several studies have used tracer data to estimate catchment transit times. McGuire, DeWalle, and Gburek (2002) analyzed  $^{18}\text{O}$  variations in rain and streamflow and obtained a mean residence time of 9.5 months for a nearby, hydrologically similar catchment. Burton et al. (2002) used various tracers to model groundwater age at a stream transect less than 1km from FD-36, and



**Figure 4.1:** Map of the study site in the FD36 sub-catchment of the USDA Mahantango experimental catchment. (a) Location of the catchment in central PA, USA. (b) Satellite image of the catchment showing the rectangular region modeled in ParFlow and the FD36 catchment boundary. (c) 3D rendering of the rectangular area modeled by ParFlow, the catchment boundary, the location of the streambed and subsurface fragipan, the location of the stream outlet and CFC sampling well, and the transect A-A' used for the 2D calibration and shown in Figure 4.2.

obtained highly scattered estimates of mean transit time ranging averaging near 10 years.

## 4.5.2 ParFlow-CLM Modeling

FD-36 critical zone hydrology was simulated using ParFlow-CLM (Ashby and Falgout, 1996; Jones and Woodward, 2001; Maxwell and Miller, 2005; Kollet and Maxwell, 2006; Maxwell, 2013) with SLIM-FAST particle tracking (Maxwell and Tompson, 2006). A detailed model description appears in the references listed and the ParFlow User’s Guide (Maxwell et al., 2016), with the most pertinent details summarized here. ParFlow-CLM (Parallel Flow - Common Land Model) is an open-source, fully-distributed simulation platform designed for parallel computing. ParFlow simultaneously solves the 3D Richards equation for subsurface flow (Richards, 1931) and - in surfaces with a ponded depth greater than zero - the 2D kinematic wave equation for surface flows using a cell-centered finite difference scheme and terrain-following transform (Maxwell, 2013). ParFlow is fully coupled with CLM, which simulates land-atmosphere processes including atmospheric forcing (e.g., rain, snow), tree canopy processes (e.g., interception), and land surface processes (e.g., latent and sensible heat flux). The model simulated the FD-36 watershed at hourly timesteps using hydrometeorological forcing from 2014. Prior to the model simulation, the model was spun up to a steady-state equilibrium with constant inflow of average precipitation to get the water table close to accurate, and then run for one full year.

### 4.5.2.1 Model Structure

The model structure, parameterization, and calibration is described in a Chapter 3 and summarized here. The ParFlow-CLM model was constructed for the FD-36 watershed using a rectangular domain, 1230m east-west by 870m north-south by 15m top to bottom. The domain surface extends at least 100m beyond the edges of FD-36, allowing catchment boundaries to emerge naturally from the flow simulation. Initial tests suggested that a grid size of 30m in the  $x$  and  $y$  direction and 0.2m in the  $z$  direction yielded a reasonable balance between performance and run times, resulting in a cell count of 41 by 29 by 55, or 65395. High resolution LIDAR data was used to assign a local slope to each cell in the cardinal direction of steepest decent (i.e., a D4 flow algorithm). Each cell was parameterized as one of three different geology types described earlier: soil, fragipan, and fractured bedrock. In

**Table 4.1:** List of key ParFlow-CLM model parameters.

Param	Zone	Value	Units	Ref.	Description
$K$	a	0.062	$m/hr$	1	saturated conductivity
$K_{xy,soil}$	s	35.2	-	GC	lateral $K$ multiplier of soil
$K_{xy,rock}$	r	1.0	-	2,GC	lateral $K$ multiplier of rock
$K_{z,fragipan}$	f	$2.2e - 5$	-	GC	vertical $K$ multiplier of fragipan
$K_{z,streambed}$	b	37.4	-	GC	vertical $K$ multiplier of stream
$F_{rock}$	r	0.79	$m^{-1}$	2,GC	length scale of exponential decline in $K$ of rock
$\Phi_{soil}$	s,f	0.29	-	3,GC	porosity of soil
$\Phi_{streambed}$	b	0.18	-	1,4,GC	porosity of streambed
$\Phi_{rock}$	s,f	0.02	-	4,GC	porosity of rock
$n$	a	1.29	-	1	van Genuchten parameter
$\alpha$	a	2.32	$m^{-1}$	1	van Genuchten parameter
$m$	a	$5.52e-6$	$hr/m^{\frac{1}{3}}$	5	mannings roughness
$r$	a	13	-	MC	slope for conductance to photosynthesis relationship

Zone legend: 1: all zones; s: soil; f: fragipan; b: streambed; r: rock

Reference legend: 1: Troch et al. 1993; 2: Gburek et al. 1999; 3: Lindeburg 2011;

4: Burton et al. 2002; 6: Maxwell et al. 2016; GC: GLUE calibration;

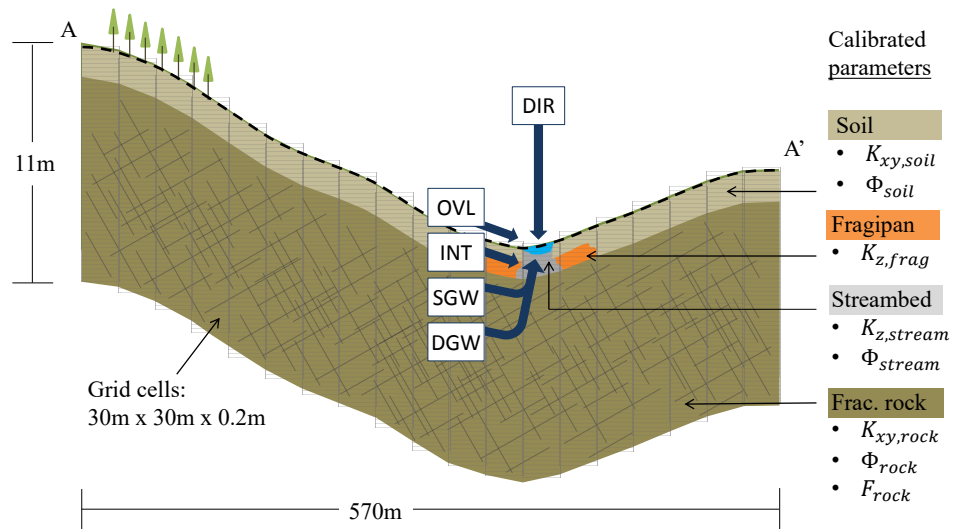
MC: Manual calibration to achieve water balance

addition, the streambed and underlying soil was parameterized separately to account for physical changes due to alluvial sediment transport and hyporheic processes. The location of the 4 geologic units are shown in plane view (Figure 4.1) and cross-section (Figure 4.2).

#### 4.5.2.2 Model Parameters and Calibration

Model parameters were set using a combination of field data, literature values, and automated calibration. Table 4.1 lists key model parameters and their source data. The mean annual ET was manually adjusted to get annual water balance (e.g.,  $\overline{ET} = \overline{J} - \overline{Q}$ ) using a fitting parameter  $r$  that describes the efficiency of photosynthesis. The soil water retention curve was parameterized using the van Genuchten model (Genuchten, 1980) with values reported in the literature for the watershed (Troch et al., 1993). A baseline hydrologic conductivity value was assumed to be the areal average of soil samples for the region (Troch et al., 1993). Model results were relatively insensitive to changes in Mannings  $n$  (results not shown), so it was assigned the default value (Maxwell et al., 2016). A total of eight other parameters were calibrated, including anisotropy in the hydrologic

conductivity, the rate of exponential decline of hydrologic conductivity in the fractured bedrock, and the porosity of upper layers and the lower bedrock. The calibration process worked as follows. Each parameter was assigned a uniform prior distribution of values informed from literature. A total of 22886 simulations were run for the 2D cross-section shown in Figure 4.2 using parameter combinations randomly selected from prior distributions obtained from the literature. Parameters sets that met the following performance criteria were considered "behavioral": Kling-Gupta Efficiency (Gupta et al., 2009) between simulated and observed daily discharge greater than 0.5, simulated baseflow water age between 8 and 10 months, and simulated groundwater age older than 3 years. Out of 342 behavioral parameter sets, one representative set was selected for the full 3D model run. Due to long simulation times for the full 3D simulation, it was not feasible to use the entire ensemble of behavioral parameters for uncertainty analysis.



**Figure 4.2:** Cross-section of the modeled domain across the transect A-A' (see Figure 4.1), including the location of the four geologic zones (soil, fragipan, streambed, and fractured rock), the calibrated model parameters, and the five modeled flow pathways into the stream: direct rainfall (DIR), overland flow (OVR), interflow (INT), shallow groundwater (SGW), and deep groundwater (DGW). Section 4.5.2.4 gives the operational definition for flow pathway classification.

#### 4.5.2.3 Particle tracking: $Q$ and $ET$

A modified version of the Lagrangian particle tracking module SLIM-FAST (Maxwell and Tompson, 2006) was used to estimate the time-varying TTD, SAS function, and rSF function for  $ET$ ,  $Q$ , and their seven constituent pathways. At the start of each hourly timestep, for each surface grid cell inside the catchment receiving infiltration, a fixed number of particles ( $n=10$ ) were injected into the top 1 mm of the grid cell. There were 492 surface grid cells inside the catchment, meaning up to 4920 particles were injected for each of the 912 hours with non-zero soil infiltration. Each particle was assigned an initial mass equivalent to an equal share ( $1/10^{th}$ ) of the mass of infiltrating water. The density of water was assumed constant throughout the domain. The particles were moved at hourly timesteps through the domain according to the advective flux driven by pressure differences between cells (i.e., dispersion was assumed to be negligible) using the approach of Pollock (1988).

$ET$  was accounted for in particle tracking by removing mass from particles as they traversed through grid cells from which  $ET$  was being removed. Specifically,  $ET$  was assumed to be uniformly selected from each grid cell:

$$\Delta m_j = \frac{m_j}{m_k} ET_k \quad (4.13)$$

where  $\Delta m_j$  was the mass removed from particle  $j$ ,  $m_j$  was the mass of particle  $j$  at the start of the timestep,  $m_k$  was the mass of water in grid cell  $k$  at the start of the timestep (determined from grid cell saturation and porosity), and  $ET_k$  was the total flux removed from grid cell  $k$  during the timestep.

The age of each particle (i.e., the time since infiltration) was tracked, such that the age of all water mass routed to  $ET$  and  $Q$  could be tabulated. This accounting was used to calculate the forward TTDs  $\overrightarrow{P}_Q(T, t)$  and  $\overrightarrow{P}_{ET}(T, t)$  that describe the distribution of transit times  $T$  that water infiltrating at time  $t$  takes to exit the catchment as  $Q$  and  $ET$ , respectively. The forward TTDs were upscaled from hourly to daily. Note that, while the mass of each particle was needed to weight contributions to the age distributions, the mass had no influence on the pathline of motion. The mass of particles contributing to  $ET$  and  $Q$  was compared to the mass of  $ET$  and  $Q$  simulated by ParFlow-CLM to ensure consistency.

Each particle was tracked for 3 years into the future (i.e.,  $\max(T)=1095$  days) by repeating the 1 year flow field simulation 3 times. Watershed conditions were assumed to be a loop of the 2014



model simulation such that:

$$u(t) = u(t \pm \Delta n) \quad \text{for all positive integers } n \quad (4.14)$$

where  $u(t)$  represents any model variable (e.g., rainfall, soil infiltration discharge, grid cell pressure) and  $\Delta$  is the period of the system (1 year). Equation 4.14 equates any model variable at any time to its value at the same time of the simulation year 2014. The assumption of periodic state is useful for estimating the first-order characteristics of catchment TTDs including the effect of seasonality, though longer simulations would be needed to better understand between-year variability. Note that 3 years was found to be sufficient to recover greater than  $> 90\%$  of particles from injection to discharge.

#### 4.5.2.4 Particle tracking: constituent pathways

The particle tracking results were analyzed to estimate the forward TTDs  $\overrightarrow{P_{Q,i}}(T, t)$  and  $\overrightarrow{P_{ET,i}}(T, t)$  for each constituent flowpath  $i$ . Beginning with  $ET$ , the total flux was disaggregated into two flowpaths: ground evaporation ( $ETG$ ) and transpiration ( $ETT$ ). In ParFlow-CLM code, the  $ETG$  and  $ETT$  are calculated and extracted separately from each grid cell in the model domain, as described in Dai (2003).  $ETG$  represents evaporation from the land surface and depends on radiative forcing, wind, humidity, and surface moisture; it is only extracted from the top model cell.  $ETT$  represents water transpired by plants and depends on plant physiology, atmospheric conditions, and soil moisture; it is extracted from the five model grid cells in the top 1m of soil, according to a parameterized distribution of plant roots. Root fraction and other parameters controlling  $ET$  processes were assigned the default ParFlow-CLM values for cropland. Because the different  $ET$  fluxes were explicitly calculated by the model, the mass of  $ET$  removed from each grid cell at each timestep could be directly attributed to either  $ETG$  or  $ETT$  and used to make separate estimates of  $\overrightarrow{P_{ET,ETG}}(T, t)$  and  $\overrightarrow{P_{ET,ETT}}(T, t)$ .

The time-series of  $\overrightarrow{P_{Q,i}}(T, t)$  were calculated in a post-processing analysis of the path that each particle traversed through the watershed. Unlike the flowpaths of  $ET$ , the flowpaths of  $Q$  were not individually modeled by ParFlow-CLM. Instead, a simple classification scheme was developed and applied to each particle based on its residence time in different parts of the catchment. The following

is a description of each flowpath and classification criterion, along with comments about the potential for misclassification.

- Direct runoff (DIR): rainfall falling directly into the stream. Particles directly injected into grid cells designated as streambed were classified as DIR. This approach is likely to overestimate direct runoff due to the low spatial resolution of the grid cell (30m) compared to the actual width of streams in the catchment (1-2m). Also, the method assumes the location of the streambed is fixed (e.g., not ephemeral), which was a reasonable assumption at the FD-36 study site.
- Overland flow (OVR): infiltration and saturation excess overland flow into the stream. Particles that traveled as overland flow before reaching the streambed were classified as OVR. Particles reaching the catchment surface were assumed to flow to the catchment outlet in less than 1 day.
- Shallow Groundwater (SGW): groundwater flow from the upper part of the fractured bedrock. Particle that travel through the upper 2m of the bedrock (but no deeper) including the area just beneath the streambed were classified as SGW.
- Deep Groundwater (DGW): groundwater flow through the lower fractured bedrock. Particles that meet the requirements of SGW but traverse more than 2m deep into the fractured bedrock were classified as DGW.
- Interflow (INT): flow that enters from the streambed laterally, including through the perched aquifer just above the fragipan. Particles that do not satisfy any other classification were classified as INT.

These criteria were used to classify the pathway taken by each particle. Then the forward TTD for each flowpath could be calculated using the particles assigned to that flowpath, yielding values of  $\overrightarrow{P_{Q,DIR}}(T, t)$ ,  $\overrightarrow{P_{Q,OVR}}(T, t)$ ,  $\overrightarrow{P_{Q,INT}}(T, t)$ ,  $\overrightarrow{P_{Q,SGW}}(T, t)$ , and  $\overrightarrow{P_{Q,DGW}}(T, t)$ . A note on notation: because the subscripts in  $\overrightarrow{P_{Q,DIR}}(T, t)$  are redundant, they will be shortened to  $\overrightarrow{P_{DIR}}(T, t)$ , and likewise for the other pathways.

#### 4.5.2.5 Age-based distributions calculation

The simulation results and estimates of the forward TTDs were used to calculate the backward TTD, the age-rank SAS functions, and the age-rank storage-flux (rSF) functions.

To estimate the backward TTDs for  $Q$  and  $ET$ , Niemi's theorem (Niemi, 1977) was rearranged and applied with the periodic steady state assumption (Equation 4.14) to solve for the PDF form of the backward TTD for each day of 2014 for ages up to 3 years:

$$\overleftarrow{p}_Q(T, t) = \overrightarrow{p}_Q(T, t - T) \frac{J_{inf}(t - T)\theta_Q(t - T)}{Q(t)} \quad (4.15)$$

$$\overleftarrow{p}_{ET}(T, t) = \overrightarrow{p}_{ET}(T, t - T) \frac{J_{inf}(t - T)\theta_{ET}(t - T)}{ET(t)} \quad (4.16)$$

where  $\overleftarrow{p}$  and  $\overrightarrow{p}$  are the PDF form of the backward and forward TTD,  $\theta(t - T)$  is the fraction of total infiltration at time  $t - T$  that exited from pathway  $i$  over the 3 year tracking period, and the subscripts  $Q$  and  $ET$  refer to the relevant flux. The PDFs were converted to CDFs for subsequent analysis. The values of  $Q_i(t)$ ,  $ET_i(t)$ , and the partition functions  $\theta_Q(t)$  and  $\theta_{ET}(t)$  were determined from the particle tracking partitioning described in the previous section. The partition functions  $\theta_Q(t - T)$  and  $\theta_{ET}(t - T)$  are the fraction of infiltration  $J_{inf}(t - T)$  that eventually exits the catchment as  $Q$  and  $ET$ , respectively (Botter, Bertuzzo, and Rinaldo, 2010). As states earlier, greater than 90% of the particles took less than the three year tracking period to exit the catchment. Similar equations were used to estimate the backward transit time of each constituent flowpath:

$$\overleftarrow{p}_{Q,i}(T, t) = \overrightarrow{p}_{Q,i}(T, t - T) \frac{J_{inf}(t - T)\theta_{Q,i}(t - T)}{Q_i(t)} \quad (4.17)$$

$$\overleftarrow{p}_{ET,i}(T, t) = \overrightarrow{p}_{ET,i}(T, t - T) \frac{J_{inf}(t - T)\theta_{ET,i}(t - T)}{ET_i(t)} \quad (4.18)$$

where the partition function  $\theta_{Q,i}(t - T)$  refers to the fraction of infiltration at time  $J_{inf}(t - T)$  exiting as discharge that exits as flowpath  $i$ , and likewise for  $\theta_{ET,i}(t - T)$ .

Next, the backward TTDs and results from the particle tracking were used to calculate the SAS functions for  $Q$  and  $ET$ , which was defined in Section 4.5. A novel approach was used to estimate the age-rank storage  $S_T(T, t)$  using the forward TTD without direct estimation of the residence time

distribution. Given that our 3 years of forward particle tracking accounted for nearly all contributions to  $Q$  and  $ET$ , we could approximate  $S_T(T, t)$  as follows (see derivation in Appendix 4.8, adapted from (Rinaldo et al., 2011)):

$$S_T(T, t) = \sum_{i=0}^{i=N} J_{inf}(t-i\Delta t)\Delta t \left( 1 - \overrightarrow{P}_Q(\tau, t-i\Delta t)\theta_Q(t-i\Delta t) - \overrightarrow{P}_{ET}(\tau, t-i\Delta t)\theta_{ET}(t-i\Delta t) \right) \quad (4.19)$$

where  $N$  is the number of timesteps ( $N=1095$ ) and  $\Delta t$  is the timestep duration (1 day). For  $Q$ , the value of  $S_T$  from Equation 4.19 and  $\overleftarrow{P}_{Q,i}(T, t)$  from Equation 4.15 were used to calculate the SAS function  $\Omega_Q(S_T, t)$  from Equation 4.6. The SAS functions for  $ET$  and the constituent pathways were estimated in a similar manner.

Lastly, we used our modeling results to calculate the rSF functions presented in Section 4.4.3. The rSF function for discharge  $f_{T,Q}(\overline{S}_T, t)$  was calculated using Equation 4.11 with the value of  $S_T$  given by Equation 4.19, the value of  $\overline{S}_T$  given by Equation 4.9, the value of  $Q(t)$  estimated from the particle tracking, and total storage  $S(t)$  extracted from the ParFlow-CLM simulation. A similar method was used to estimate the rSF functions for  $ET$  and the constituent flowpaths.

#### 4.5.2.6 Comparative analysis

The model results were compared and decomposed in quantitative and qualitative ways to answer the main research questions described in the introduction (Section 4.3). The analysis focused on the extent to which the emergent shape and time-variability of the age-based distributions of  $Q$  and  $ET$  could be explained by the behavior of their constituent pathways.

## 4.6 Results

### 4.6.1 Model performance overview

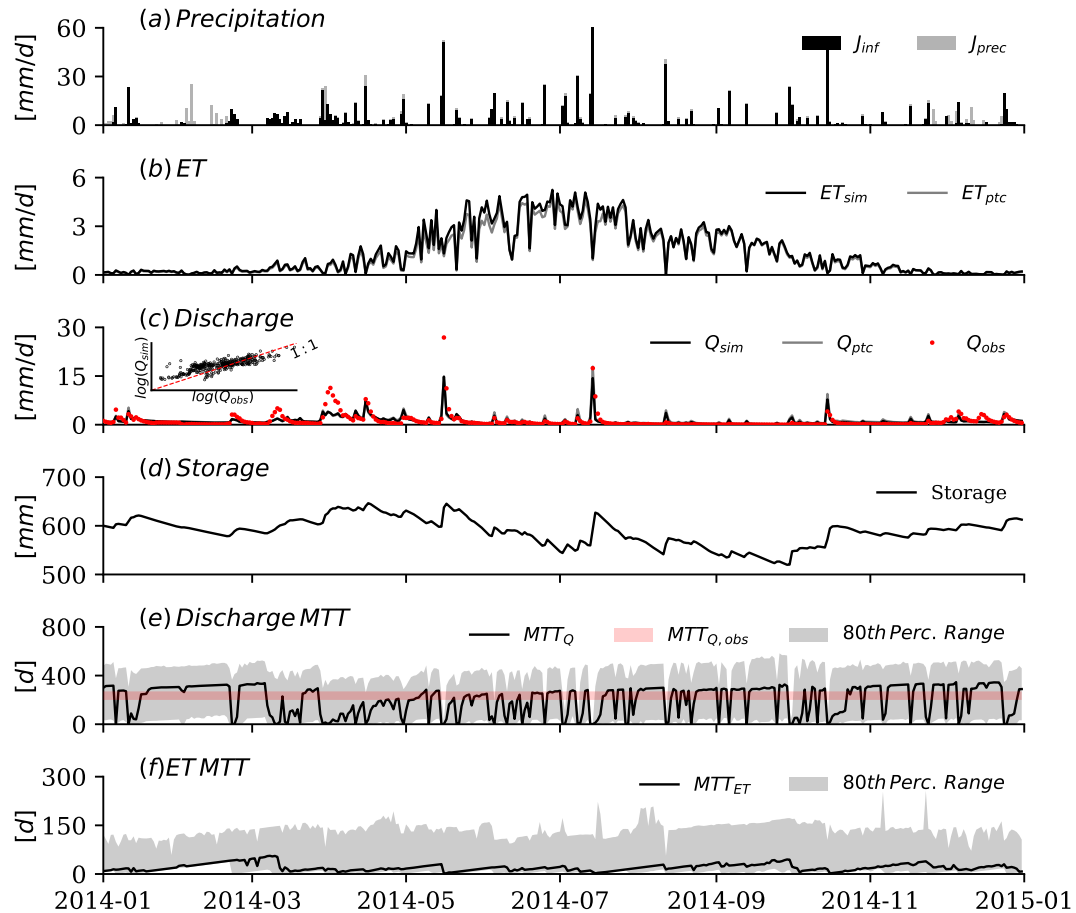
The model simulated major aspects of the catchment hydrology and its seasonal behavior. Figure 4.3(a) shows the observed input rainfall to the catchment  $J_{prec}$  and - after simulating surface-atmosphere processes including canopy interception, snowpack, and snowmelt - the infiltration to the catchment land surface  $J_{inf}$ . Total rainfall  $J_{prec}$  was 1078 mm with several large events in

the spring and summer time, of which 983 mm infiltrated as  $J_{inf}$ . Although  $J_{prec}$  and  $J_{inf}$  were closely synchronized ( $r^2=0.86$ ), they differ most in winter when  $J_{prec}$  falls as snow and does not produce  $J_{inf}$  until snowmelt. This behavior is also seen in the streamflow data, in that some large precipitation events in the winter elicit a negligible increase in discharge (e.g., the 30mm rainfall events on 2/1/2014). Total annual catchment  $ET$  was 522 mm with high values peaking in the summer and near zero values in the winter (Figure 4.3(b)). Total catchment storage varied between 519 and 646 mm with a peak during spring showers and snowmelt, and a trough following a long period of high  $ET$  in the summer.

The calibrated model captured key aspects of observed discharge and water ages. The fit between daily observed and simulated  $Q$  and  $\log(Q)$  had a KGE of 0.60 and 0.65, respectively (see Figure 4.3(c) including inset). The simulation did well at capturing the general timing and size of storm events as well as the large drop in baseflow during the summer. The model showed less skill at capturing the peak flows during a snowmelt event in late March, possibly because the model did not simulate the influence of freezing temperatures on soil hydrology.

The median backward transit time for  $Q$  ( $MTT_Q$ ) shown in Figure 4.3(d) dips significantly during rain events, but quickly relaxes to baseflow values that coincide with and sometimes exceed the tracer-derived calibration target of 240 – 300 days. The simulated median backward transit time of groundwater at 5 m depth in the subsurface was found to be greater than 3 years (results not shown), which is consistent with previous groundwater age-dating studies (Burton et al., 2002).

The pattern of  $MTT_Q$  and  $MTT_{ET}$  (Figure 4.3(e-f)) have a complex relationship with catchment storage and hydrological characteristics. The time-average  $MTT_Q$  is 264 days with a large interdecile range of 16-465 days (see Table 4.2). Although catchment storage has been shown to be an important control on transit times in other catchments (van der Velde et al., 2012; Harman, 2015; Wilusz, Harman, and Ball, 2017), the correlation between  $S(t)$  and  $MTT_Q$  is very weak ( $r^2 = 0.05$ ). The time-average  $MTT_{ET}$  is much lower than discharge, at 18 days with an interdecile range of 5-137 that skews heavily toward younger ages. The  $MTT_{ET}$  drops after large influxes of  $J_{inf}$  and slowly rises between large events.



**Figure 4.3:** Important ParFlow-CLM model inputs, outputs, and performance for the FD-36 catchment. Panel (a) shows total precipitation  $J_{prec}$  and total infiltration to the land surface after accounting for above-surface processes including snow accumulation, snowmelt, and canopy interception and evaporation  $J_{inf}$ . Panel (b) and (c) shows surface  $ET$  and  $Q$  simulated by the ParFlow-CLM model (black) and by particle tracking (gray). Panel (c) also shows observed daily discharge against time (red circles) and against simulations in log-log space (see inset). Panel (d) shows total water storage in the catchment over time. Panels (e) and (f) show the time-series of median transit time (MTT) for  $ET$  and  $Q$  (black lines) with the 10-90th percentile ages shaded gray. The red strip shows the tracer-derived age of baseflow in a nearby stream (McGuire, DeWalle, and Gburek, 2002).

## 4.6.2 Pathway partitioning

The particle tracking methodology was used to determine the partition functions  $\theta$  for the two pathways contributing to  $ET$  (i.e.,  $ETG$ ,  $ETT$ ) and five pathways contributing to  $Q$  (i.e.,  $DIR$ ,  $OVR$ ,  $INT$ ,  $SGW$ ,  $DGW$ ). Figure 4.4(a) shows how the influx from each infiltration event is partitioned by the fraction  $\theta_i$  to each flowpath  $i$ . The largest fraction of  $J_{inf}$  is partitioned to  $ETT$  (49%), predominantly in the summer months. The second largest is  $OVR$  (16%), followed by  $INT$  (12%). Notwithstanding the significant day-to-day variability in infiltrating rainfall, the changes in  $\theta_i$  occur over longer, seasonal timescales. In other words, variability in the partitioning of influxes seems to be driven by the season, and not by the size of any particular influx event. The average daily sum of  $\theta_i$  for all pathways is close to unity (0.97), reflecting the fact that nearly all influx was discharged from the catchment over the 3-year particle tracking period.

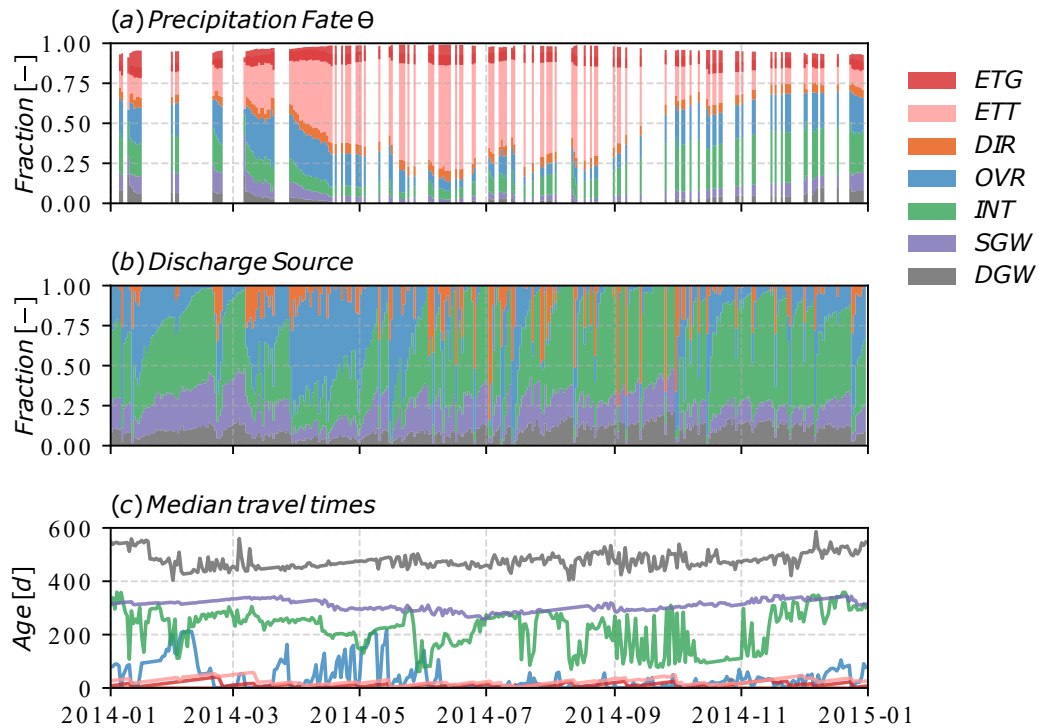
While the destination pathways of infiltrating rainfall changed on the seasonal timescale, the source pathways of discharge showed much more day-to-day variation. Figure 4.4(b) shows the fraction of each pathway contributing to  $Q$  for each day of 2014. Table 4.3 shows the time-average median contribution from each: 48% from  $INT$ , 18% from  $OVR$ , and 16% from  $SGW$ , and less than 10% from  $DGW$  and  $DIR$ . The drivers of variability in outflux composition is explored in the next section.

An initial comparison of the  $MTT_i$  of each flow pathway  $i$  shows a wide spread both between pathways and between days of the year. As shown in Figure 4.4(c) and described in Table 4.2,  $DGW$  is significantly older than other pathways with a  $MTT$  of 474 days that remains steady throughout the year. At the other extreme, the  $MTTs$  of  $ETG$ ,  $ETT$ , and  $DIR$  are the youngest pathways with  $MTTs$  below 30 days that also remain relatively steady during the year. Between these end members are pathways with the most dynamically intermingled ages:  $OVR$  ( $MTT_{OVR} = 22$  days),  $INT$  (216 days), and  $SGW$  (251 days). In the spring, the  $MTT$  of  $OVR$  and  $INT$  intersect, and, later in the year, the  $MTT$  of  $INT$  and  $SGW$  also intersect. These results suggest a complex many-to-many mapping between age and pathway, depending on the pathway type and time of year. The following sections try to unpack this complexity by analyzing the flux and age-based distributions of each pathway in more detail.

### 4.6.3 Partition flux

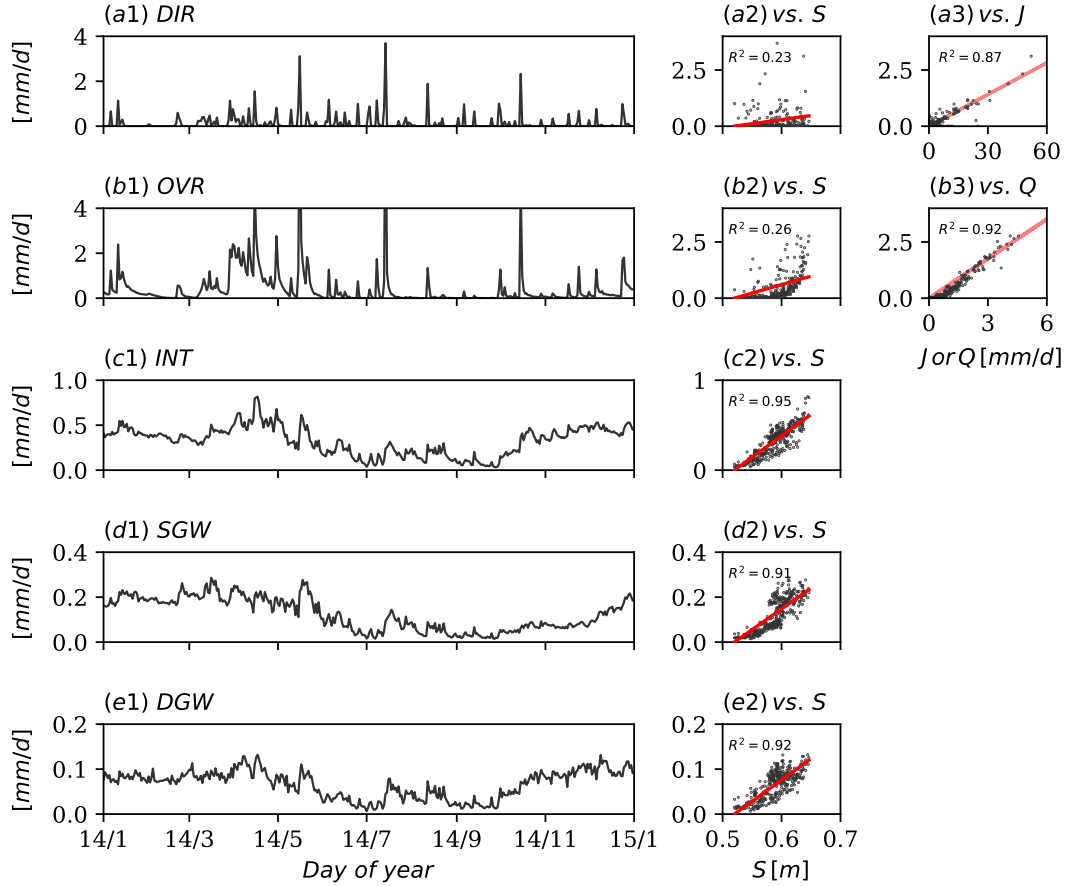
Section 4.4 showed that the age distributions of  $Q$  and  $ET$  are driven by (1) the total flux of each constituent pathway and (2) the age-based distribution of each constituent pathway. We investigate the total flux in this section, and the age distribution in the next. For  $ET$ , the flux of  $ETG$  is relatively small and roughly constant throughout the year. By contract, the flux of  $ETT$  is highly seasonal and contributes the majority  $ET$  in the summer (results not shown). The average flux of  $ETT$  is 6.5 times greater than  $ETG$  (see Table 4.2).

For  $Q$ , the contribution from different pathways is shown in Table 4.2.  $INT$  had the highest



**Figure 4.4:** The partitioning of rainfall into different flow pathways. Panel (a) shows the fraction of  $J_{inf}$  exiting to different flow pathways for daily rainfall during the 1-year simulation. The white column breaks indicate periods with zero infiltration. Panel (b) shows the fraction of each flow pathway contributing to daily  $Q$ . Panel (c) shows the median transit time for each flow pathways contributing to  $Q$  and  $ET$ . All flowpath acronyms are defined in the text.





**Figure 4.5:** Time-series of simulated  $Q$  for each flowpath and comparisons to  $S$ ,  $J_{prec}$ , and  $Q$ . The first column shows total flux from DIR, OVR, INT, SGW, and DGW. Note the changing y-axis range. The second column plots simulated storage against discharge. The last column plots simulated discharge against  $J_{prec}$  (panel a3) and total  $Q$  (panel b3). For each scatter plot, lines of best fit are shown in red.

median flux ( $0.36 \text{ mm/d}$ ), suggesting that the TTD and SAS function of  $INT$  has the largest influence on the emergent TTD and SAS function of  $Q$ . By contrast,  $DGW$  contributed the least ( $0.07 \text{ mm/d}$ ), suggesting that  $DGW$  has a relatively small influence.

The time-variability of fluxes in each pathway contributing to  $Q$  had two distinct patterns. The fluxes  $DIR$  and  $OVR$  shared a relatively flashy pattern with a correlation of  $r^2=0.67$ . The fluxes  $INT$ ,  $SGW$ , and  $DGW$  shared a smoother and more seasonal pattern that dropped low in summer, with pair-wise correlations all greater than  $r^2=0.50$ . The fact that the pair-wise correlation between

the two groupings is relatively low ( $r^2 < 0.10$ ) suggests that variability in the two groups is driven by different hydrologic processes.

To help understand what drives the time-variability of pathway fluxes, Figure 4.5(a2-e2) shows the relationship between total catchment storage  $S$  and pathway flux. The subsurface pathways  $INT$ ,  $SGW$ , and  $DGW$  all have a strong linear relationship with  $S$  ( $r^2 = 0.91-0.95$ ). By contrast, the pathway  $OVR$  has a relatively non-linear ( $r^2 = 0.23$ ) but monotonically increasing (spearman's  $\rho^2 = 0.62$ ) relationship with  $S$  that resembles a typical non-linear storage discharge relationship, and Figure 4.5(b3) shows that  $OVR$  is highly correlated with total  $Q$  ( $r^2 = 0.92$ ). Finally, the flow pathway  $DIR$  has the weakest relationship with  $S$  ( $r^2 = 0.23$ ,  $\rho^2 = 0.02$ ), but is strongly correlated with  $J_{obs}$  ( $r^2 = 0.87$ ).

#### 4.6.4 Age-based distribution analysis: $ET$

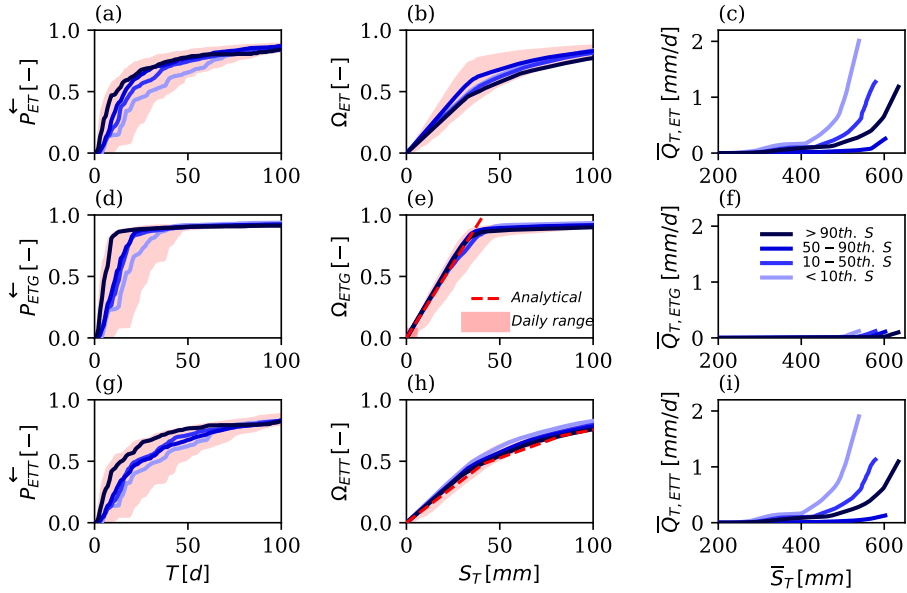
We begin by analyzing the TTD, SAS function, and rSF function of  $ET$ , to understand the time-variability of the ages of  $ET$  and its two constituent pathways. The case of  $ET$  is relatively simple and provides a good introduction to the analysis before applying it to  $Q$ , which has five flowpaths and greater conceptual complexity.

The TTD for  $ET$  and its flowpaths show significant time-variability, with the distribution of ages younger when catchment storage is high. The median of  $MTT_{ET}$  over the course of the year (and the interdecile age) was 18 (5-137) days.  $ETT$  was slightly older at 23 (5-145) days and  $ETG$  younger at 11 (4-45) days (see Figure 4.6(abg)). All three distributions are skewed left but with a small fraction of relatively old water. In general, the TTD for  $ET$  is more like the TTD for  $ETT$  than  $ETG$ .

While the TTDs vary with time and storage, the SAS functions for  $ETG$  and  $ETT$  tend to collapse into a single distribution with no clear storage relationship and small day-to-day variability (see Figure 4.6(eh)). For example, for the  $ETG$  pathway, the coefficient of variation (CV) drops from 0.62 for the median age  $T$  of the TTD distribution to 0.24 for the median age-ranked storage  $S_T$  of the SAS distribution. The results show that the time-variability in the SAS function for  $ET$  results from the supposition  $ETT$  and  $ETG$  with roughly time-invariant SAS functions. It is interesting to note that, even though the TTDs for  $ET$  tend to get younger at higher storage, the SAS functions show no inverse storage effect.

The SAS functions for  $ETG$  and  $ETT$  are very close to the distribution we would expect based on the ParFlow-CLM model assumptions. In ParFlow-CLM,  $ETG$  is uniformly selected from the top model grid cell which had an average storage volume of  $S_0=47$  mm (determined as the product of relative saturation, porosity, and cell depth averaged across the top catchment cells for the entire model simulation). This is equivalent to a well mixed system with storage of 47 mm, which has a uniform SAS function (Harman, 2015):

$$\Omega_{ETG}(S_T) = \begin{cases} \frac{S_T}{47 \text{ mm}}, & S_T \in [0, 47 \text{ mm}] \\ 1, & S_T > 47 \text{ mm} \end{cases} \quad (4.20)$$



**Figure 4.6:** Analysis of age, storage, and outflow for  $ET$  (top row) and its constituent flowpaths  $ETG$  and  $ETT$  (bottom two rows) during the 1-year simulation. The left column shows the median backward transit-time distribution  $\overleftarrow{P}(T)$  at four different ranges of catchment storage: very dry (lower 10th percentile storage, lightest blue line), moderately dry (10-50th percentile, lighter blue line), moderately wet (50-90th percentile, darker blue line) and very wet (greater than 90th percentile, darkest line). The light red shading shows the 80th percentile daily range of  $\overleftarrow{P}(T)$ . The middle column is similar but shows the SAS functions and the approximate analytic solutions derived in the text (red dashed lines, panels (e) and (h)). The last column shows the rSF functions.

Similarly, an approximate analytical solution can be derived for  $ETT$ , for which a fraction  $\eta_k$  is uniformly selected from the top  $k=5$  cells, where  $\eta_k$  is the parameterized root fraction distribution. By consideration of the value of  $\eta_k$  and the average saturation, porosity, and cell depth, a 1D solution was derived for  $ETT$ .

The analytically-derived 1D SAS functions for  $ETG$  and  $ETT$  show very close agreement to the SAS functions derived from the particle tracking (see Figure 4.6), but also a telling area of disagreement. The good agreement indicates that the 1D approximation is a good representation of mixing in the upper cells, and that lateral transfers are minimal. The major difference is a small but significant volume of older age-rank storage contributing to  $ET$  not anticipated by the 1D theory. This fraction is likely driven by the upwelling and evaporation of old water in the upper cells of the streambed.

The rSF functions (Figure 4.6(cfi)) show how different parts of the age-rank storage complement  $\bar{S}_T$  contribute to the  $ET$  and its pathways. For example, the  $f_{T,ET}(\bar{S}_T, t)$  function for the lowest storage scenarios (<10th percentile) shown in Figure 4.6(c) indicates that the maximum  $ET$  was approximately 2 mm/d and the maximum total storage  $S$  was 525 mm. The sharp initial drop in this curve (from right to left) indicates that 1 mm/d of  $ET$  comes from the oldest 500 to 525 mm mm of storage, which is equivalent to the youngest 25 mm of storage. The remaining 1 mm/d of  $ET$  comes from the oldest 300 mm to 500 mm of storage. Almost none of the  $ET$  is selected from the oldest 300 mm of storage. Other curves in Figure 4.6(cfi) can be interpreted in the same manner. Given that  $f_{T,ET}$  is the linear supposition of  $f_{T,ETG}$  and  $f_{T,ETT}$ , the curves in Figure 4.6(c) are roughly equivalent to the sum of Figure 4.6(f) and (i).

Inspection of the  $f_{T,ET}$  functions gives several insights into what drives the distribution of ages in  $ET$ . First, comparison of Figures 4.6(b) and (c) show that the flux of  $ETT$  can be an order of magnitude larger than the flux of  $ETG$ , and during these times it will dominate the shape of the  $ET$  age-based distributions. At times when the fluxes are nearly the same, they will both influence of the shape of  $ET$  age-based distributions. Second, almost none of the  $ET$  is selected from the oldest 300mm of storage. Third, there is no clear correlation between catchment storage and  $ET$ . The period with highest storage has a very moderate  $ET$ . This likely reflects the fact that  $ETT$  is very sensitive to the pattern of solar radiation throughout the year.

### 4.6.5 Age-based distribution analysis: $Q$

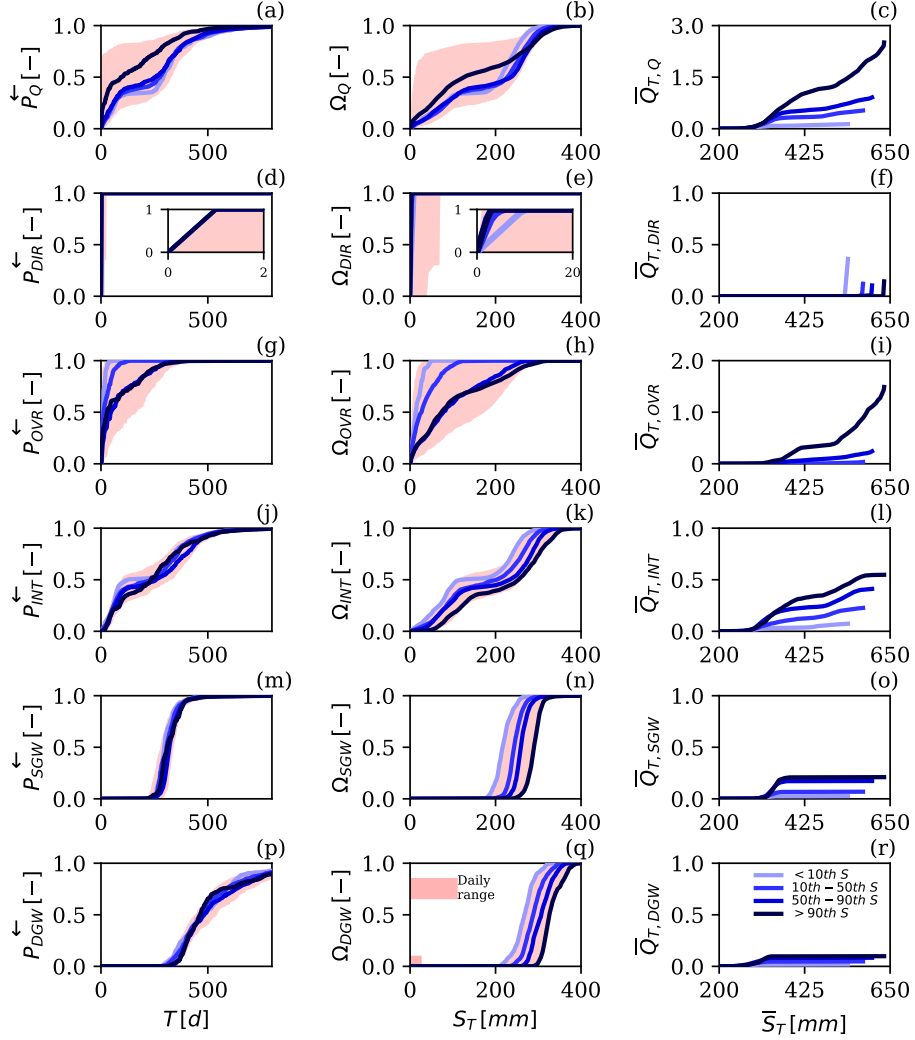
The TTD of  $Q$  has time-variable behavior and a distinct bimodal shape (see Figure 4.7(a)). The median transit time of discharge  $MTT_Q$  varies widely from  $<1$  to 346 days. The period with the highest decile of storage has a somewhat younger distribution of water compared with the other storage scenarios. In these lower storage scenarios, the TTD shows a relatively large fraction of relatively young water (0-50 days) and older water (200-500 days) and a "middle-aged depression" in between with no contribution aged 50-200 days (marked by the middle plateau in the CDF).

The time-variability and shape of  $Q$  can be explained in part by inspection of its constituent pathways (see Figure 4.7(d-p)). Nearly all of the time-variability seems to be explained by the *OVR* pathway as well as the intermittent *DIR* pathway. The other pathways *INT*, *SGW*, and *DGW* are relatively time-invariant. While  $MTT_{INT}$  showed large oscillations in our initial analysis in Figure 4.4(c), this seems to be at least in part an artifact of the bimodal shape of the distribution, where small shifts cause the median to toggle from one mode to the other. The portion of younger water in the distribution is from *DIR* and *OVR*, while the older water is from *SGW* and *DGW*. *INT* is the only flowpath to contribute both older and younger water. Although  $Q$  tends to get younger when storage is very high, this is not true for any of the constituent flowpaths.

The shape and time-variability of the SAS function of  $Q$  in Figure 4.7(b) is similar to that of the TTD. The day-to-day median  $S_T$  varies from 2 to 258 mm with a similar bi-modal distribution. When catchment storage is in the highest decile, the median  $S_T$  drops to 126 mm (compared to 244 mm at other storage times). This is indicative of the inverse storage effect.

Unlike the SAS functions for *ET* pathways, the SAS functions for  $Q$  pathways exhibit significant time-variability (shown in Figure 4.5e,h,k,n,q). As with the TTD function, the greatest time-variability is in the *OVR* pathway and the intermittent *DIR* pathway. Surprisingly, the *INT*, *SGW*, and *DGW* SAS functions also vary in time, and in all three pathways the shift is strongly correlated with storage. For example, the correlation coefficients between median  $S_T$  and catchment storage in *INT*, *SGW*, and *DGW* are 0.31, 0.82, and 0.89 respectively. Interestingly, although the  $Q$  SAS function has an inverse storage effect, each individual pathway has the opposite relationship with storage: a direct storage effect. When storage increases, so does the volume of  $S_T$  selected into storage.

This apparent paradox - that many pathways with a direct storage effect produce an emergent  $Q$



**Figure 4.7:** Analysis of age, storage, and outflow for  $Q$  (top row) and its constituent flowpaths DIR, OVR, INT, SGW, and DGW (bottom five rows) during the 1-year simulation. The left column shows the median backward transit-time distribution  $\overleftarrow{P}(T)$  at four different ranges of catchment storage: very dry (lower 10th percentile storage, lightest blue line), moderately dry (10-50th percentile, lighter blue line), moderately wet (50-90th percentile, darker blue line) and very wet (>90th percentile, darkest line). The light red shading shows the 80th percentile daily range of  $\overleftarrow{P}(T)$ . The middle column shows the SAS functions. The insets in panels (d) and (e) zoom into the youngest portion of the DIR flowpath. The last column shows the rSF functions. Note that DIR and OVL are intermittent (see activity in Table 4.2), and only days with rainfall are used to construct the distributions.

**Table 4.2:** Characteristics of pathway partitions for *ET* and *Q*.

<b>Flux</b>	<b>Partition</b>	<b>Active</b> %	<b>Flow</b> [ <i>mm/d</i> ]	<b>TTD</b> [ <i>d</i> ]	<b>rSAS</b> [ <i>mm</i> ]
<i>ET</i>	<i>TOT</i>	100	0.83 ( 0.1 - 3.94)	18 ( 5 - 137)	33 ( 6 - 142)
	<i>ETG</i>	100	0.11 ( 0.04 - 0.22)	11 ( 4 - 45)	19 ( 3 - 60)
	<i>ETT</i>	100	0.7 ( 0.04 - 3.71)	23 ( 5 - 145)	39 ( 7 - 149)
<i>Q</i>	<i>TOT</i>	100	0.76 ( 0.17 - 2.33)	264 ( 16 - 465)	220 ( 34 - 293)
	<i>DIR</i>	46	0.13 ( 0.0 - 0.79)	0 ( 0 - 0)	1 ( 0 - 3)
	<i>OVR</i>	96	0.14 ( 0.01 - 1.27)	22 ( 3 - 111)	44 ( 3 - 121)
	<i>INT</i>	100	0.36 ( 0.1 - 0.51)	243 ( 34 - 464)	216 ( 53 - 294)
	<i>SGW</i>	100	0.11 ( 0.03 - 0.22)	312 ( 278 - 371)	251 ( 233 - 273)
	<i>DGW</i>	100	0.07 ( 0.02 - 0.1)	474 ( 362 - 777)	297 ( 270 - 330)

Active: percent time flow path is non-zero;

TTD: Transit time distribution showing median age and 80th percentile range;

rSAS: rank StorAge Selection function showing median age-rank storage and 80th percentile range.

with an inverse storage effect - can be resolved by analyzing the  $f_T$  functions in the last column of Figure 4.7. Under high storage conditions, *OVR* contributes significantly more flux than any other pathway. Even though *OVR* is selected from older  $S_T$  when storage is high, it is still younger than all other flowpaths except *DIR*. Thus, even though individual flowpaths get older when storage is high, their combined distribution is younger because of disproportionate growth of the youngest flowpaths.

The  $f_T$  functions also elucidate the different nature of the direct storage effect in each flowpath. For example, Figure 4.7(n) shows that the *SGW* pathway is selected from a relatively wide 130 mm interval of  $S_T$ , from approximately the youngest 190 mm of storage to the youngest 320 mm of storage. On the other hand, Figure 4.7(o) shows that that *SGW* is selected from a relatively narrow 25 mm interval of  $\bar{S}_T$ , from the oldest 375 mm of storage to the oldest 400 mm. Thus, *SGW* is selected from a narrow range of the oldest water in the catchment, irrespective of how much younger

water is present.

#### 4.6.6 TTD Taxonomy

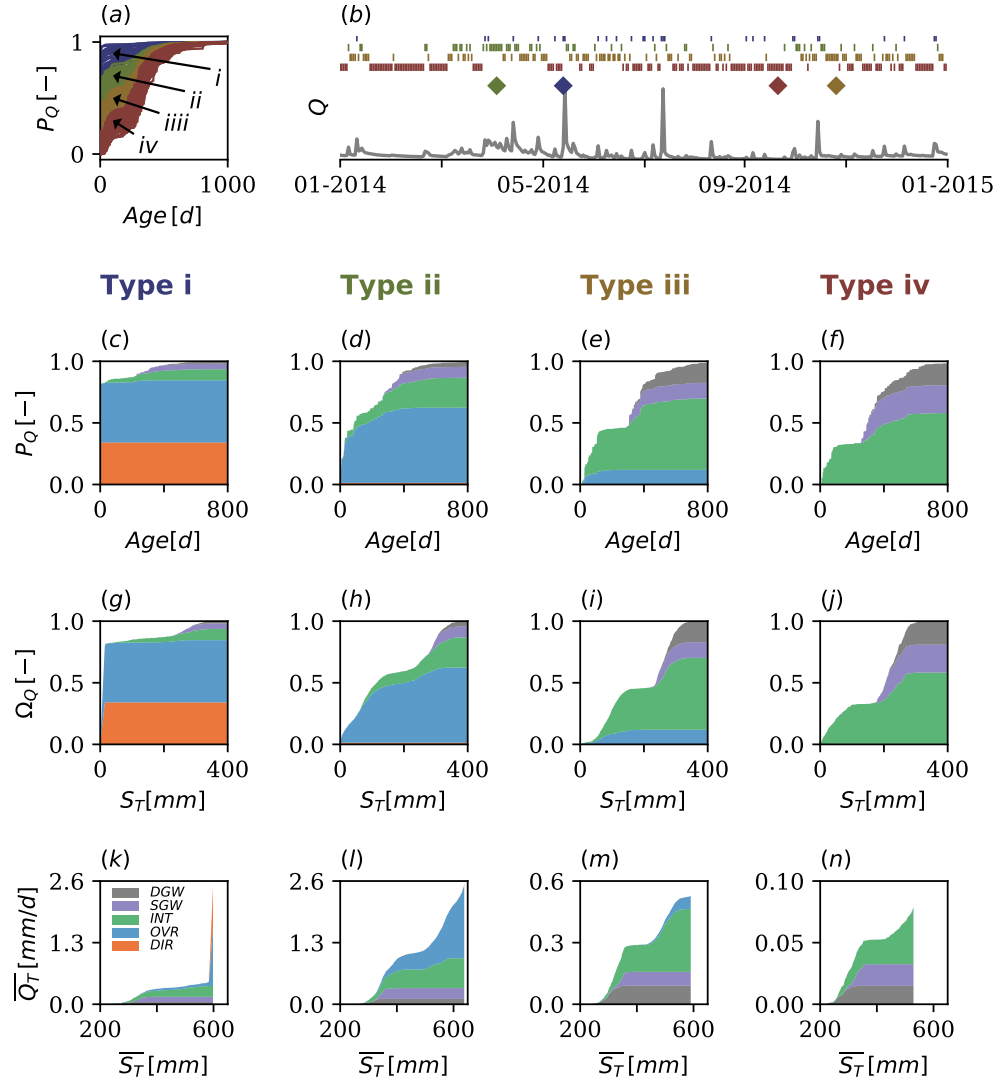
The previous analysis shows there is a many-to-many mapping between the source age or age-rank storage of discharge and its flowpath. For example, throughout the course of the simulation, the range of *OVR* ages overlaps with the ages of *INT* (see Table 4.2). In this section, we broaden the question to investigate the mapping between the shape of the age distribution and its constituent flowpaths. In other words, to what extent can the shape of the age-based distribution at a point in time be associated with a particular set of pathways?

The daily TTDs were analyzed to create a taxonomy of four characteristic TTD types. To do this, a K-mean clustering algorithm was applied to the quintiles of the simulated daily TTD distributions. Figure 4.8(a) shows the four types, ranging from the youngest distributions (type i) with a median MTT of 1 day to oldest distribution (iv) with a median MTT of 299 days (see Table 4.3). Figure 4.8(b) shows that type i occurs in isolated episodes on the fewest number of days in the yearlong simulation ( $n=30$ ), type ii is more common ( $n=59$ ), and type iii and iv are the most common ( $n=107, 169$ ). The shape of TTD often, but not always, cascades in time from type i to type iv in series.

Figure 4.8(c-n) shows archetype age-based distributions for each of the types, and Table 4.3 shows various characteristics of each. In general, the transition from type i to type iv is characterized by decreasing fractions of DIR and OVR and increasing fractions of INT, SGW, and DGW. The type i age distributions tend to coincide with heavy rainfall and higher discharge, though the mapping between type and discharge is not one-to-one. For example, Figure 4.8(k) and (j) show two  $f_T$  curves with similar total  $Q$  but very different profiles. Interestingly, there is no clear relationship between total storage  $S$  and type.

Given information about the shape of the age-based distributions, the source pathway can be easier to infer, especially for younger parts of the distribution. When the distribution is type i, younger water is mostly likely a mix of DIR and OVR pathways. In type ii distributions, younger water is likely to come from OVR. In type iii and iv, younger water is likely to come from INT.





**Figure 4.8:** A taxonomy of transit time distributions and their constituent flow paths. Panel (a) shows the K-means clustering of daily  $\hat{P}_Q(T)$  into four categorical shapes: type i, type ii, type iii, and type iv. Panel (b) shows the hydrograph, the occurrence of each type during the 1 year simulation (upper color bands), and the date of four archetypes (colored diamonds) explored in the four columns below. Each column contains the backward TTD  $\hat{P}_Q(T)$  (upper row), the SAS function  $\Omega(S_T)$  (middle row), and the rSF function (lower row). The distributions are filled in with colors to show the contribution from each flow pathway to the overall distribution.

**Table 4.3:** Characteristics of the four types of transit time distributions (first four columns) and all types combined (last column).

	Type i	Type ii	Type iii	Type iv	All
$n$	30	59	107	169	365
$\overleftarrow{P}_Q(0.5)$	1	37	205	299	265
	(1 – 1)	(1 – 114)	(78 – 297)	(268 – 329)	(2 – 323)
$\Omega(0.5)$	11	68	210	240	220
	(8 – 23)	(9 – 150)	(105 – 237)	(215 – 252)	(21 – 247)
$J$	17.6	4.0	0.0	0.0	0.0
	(10.1 – 38.4)	(0.1 – 8.2)	(0.0 – 3.3)	(0.0 – 0.1)	(0.0 – 7.7)
$Q$	2.6	1.7	0.8	0.6	0.8
	(1.2 – 7.6)	(0.7 – 3.3)	(0.4 – 1.5)	(0.1 – 0.9)	(0.2 – 2.3)
$S$	593	608	596	584	591
	(541 – 627)	(561 – 638)	(567 – 627)	(541 – 606)	(548 – 626)
$DIR$	33	14	0	0	0
	(18 – 67)	(0 – 34)	(0 – 19)	(0 – 0)	(0 – 27)
$OVR$	51	46	23	6	18
	(20 – 65)	(16 – 63)	(8 – 46)	(0 – 25)	(2 – 55)
$INT$	10	24	44	56	48
	(4 – 19)	(16 – 36)	(30 – 64)	(43 – 66)	(18 – 64)
$SGW$	4	9	14	21	16
	(1 – 6)	(6 – 14)	(10 – 21)	(14 – 28)	(7 – 26)
$DGW$	2	5	8	12	9
	(1 – 4)	(3 – 8)	(6 – 14)	(9 – 16)	(4 – 15)

The top row  $n$  shows the number of days for each type of TTD during the 1-year simulation. Subsequent rows show the median value and the 80th percentile range in parentheses.

## 4.7 Discussion and conclusions

### 4.7.1 Drivers of shape and transience

The age-distributions for  $ET$  and  $Q$  had very different shapes and time dependence. Both are described in turn.

#### 4.7.1.1 $ET$

The shape and temporal variability of the TTD for  $ET$  were well explained by examining its constituent flowpaths, ETG and ETT, and estimating the TTD and SAS function for each. While the TTD and SAS function for all  $ET$  were time-varying, the SAS functions for ETG and ETT were nearly time-invariant, and their shape was derived using the model governing equations, the

physically-based parameters, and a 1D approximation. The results conclusively show that the time-variability in the TTD results from a combination of variability in soil infiltration (i.e., external variability causing flowpath acceleration as defined in (Kim et al., 2016)) and shifting contributions from the ETG and ETT pathways from winter to summer (i.e., internal variability causing flowpath activation). The analytical approach could be used in ParFlow-CLM and similar models to estimate *ET* age distributions using model parameters instead of computationally-intensive particle tracking.

The simulation results have at least two implications for our understanding of *ET* age distributions. First, the good agreement between model results and the 1D analytical solution suggest that the lateral movement of water can be neglected when estimating *ET* ages, with the exception of the riparian zone where upwelling groundwater is evapotranspired. The relative unimportance of lateral flow probably reflects the fact that transport in the vadose zone is mostly vertical. Second, the finding that the MTT of *ET* (18 days) was significantly younger than *Q* (264 days) informs the current debate on the "two water worlds" hypothesis, which posits that the water transpired by plants is generally older than water delivered to streams (McDonnell, 2014). In fact, our results suggest the water transpired is actually *younger* than water delivered to streams. The representation of *ET*, however, is relatively simple in ParFlow-CLM and important ecohydrologic processes causing the two water worlds phenomena may be missing.

#### 4.7.1.2 *Q*

The behavior of the TTD for *Q* can be explained in part by the activation and deactivation of the five flowpaths considered, which had distinct but overlapping age distributions. The younger water was routed by DIR and OVR, the older water by SGW and DGW, and a combination of old and young routed by INT. Different combinations of flowpaths occurred throughout the year, which could be statistically clustered into four different TTD types. The overall shape was driven by the largest fluxes, with DIR more likely to dominate during large rain events, OVR during high storage periods with no rainfall, and INT otherwise. Our results are consistent with the small number of other studies looking at the relationship between discharge flowpaths and age distributions. For example, Yang et al. (2018) analyzed results from an ISSHM and found that shallow flowpaths contribute young water and deeper groundwater contributes older water.

Unlike the case for  $ET$ , the behavior of the TTD for  $Q$  was also driven by the time-variability of the SAS functions for individual pathways. The SAS function for  $OVR$  was particularly sensitive and changed shape to select more older storage when storage increased; this is likely due to the exfiltration of deeper groundwater flowpaths contributing to saturation excess overland flow. The SAS functions for INT, SGW, and DGW were also time-dependent and shifted with total storage in the catchment.

### 4.7.2 Driver of Inverse Storage Effect

In our study catchment, for  $ET$  there was no systematic relationship between catchment storage and the preference of the SAS function for young age-rank storage. In other words,  $ET$  did not exhibit the inverse storage effect. While the distribution of  $ET$  ages does get younger during high storage conditions, the distribution of age-rank storage  $S_T$  remains the same.

For  $Q$ , our model results show that the inverse storage effect can be explained by (1) the dramatic increase of overland flow when storage is high and (2) the fact that overland flow is selected from relatively young age-rank storage. This is, to our knowledge, the first demonstration of the effect in a fully coupled surface groundwater model at catchment scale. It is interesting to note, however, that overland flow and other flowpaths do not themselves demonstrate and ISE. That is, under high storage conditions, all the flowpaths except direct runoff select from *older* age-rank storage, and thus exhibit a direct storage event.

### 4.7.3 Implications of "age-equifinality" in catchments

Our analysis rejects the null hypothesis of a time-invariant, one-to-one mapping between the age of a parcel of water exiting a catchment and the flowpath that it traversed. To the contrary, the mapping seemed to be dynamic and many-to-many. It is instructive to see where and how the mapping failed. For  $ET$ , the ground evaporation was almost entirely young (<45 days) water. By contrast, the transpiration included young and older water (5-145 days). Therefore, only older  $ET$  could be mapped to a single pathway. Without additional information, the source pathway of the younger  $ET$  is ambiguous. The mapping between the age (or age-rank storage) and source pathway was more complex for the five pathways contributing to  $Q$ . The oldest water in discharge (>400 days

old) could be definitely mapped to deep groundwater. No other one-to-one mapping was possible; moderately old water ( 300-400) could be either shallow groundwater, interflow, or overland flow, and the youngest water ( <100 days old) could be either interflow, overland flow, or direct runoff.

Equifinality is the principle that an end state can be reached by multiple possible means. In hydrology, the principle is most commonly associated with model selection and parameterization (Beven, 2006a). This work shows that another kind of equifinality exists where water of a certain age or age-ranked storage in the outflux can originate from multiple possible flowpaths. This "age-equifinality" may complicate the interpretation of TTDs and SAS functions. Taxonomy schemes such as the one presented show potential for mitigating age-equifinality and strengthening the ability to make inferences about flowpaths from stream ages. For example, in general the youngest discharge can be either direct runoff or overland flow, but if the overall shape of transit time distribution is type ii, it is more likely to be overland flow.

#### 4.7.4 Implications of a "middle-aged depression" in TTDs

The age-based distributions for  $Q$  exhibited a bimodal distribution with a selection of the youngest water and oldest water, and a "middle aged depression" in between with no contribution. This depression occurs around 150-250 days in the TTDs, around 150-200 mm of the youngest storage in the SAS functions, and around 425-450 mm of the oldest storage in the  $f_T$  functions. While a full investigation of its origin is outside the present scope, the depression is an interesting phenomena for at least two reasons. The first reason is that previous age measurements using calibrated, lumped parameter models have suggested that age distributions have a bimodal distribution (e.g., Benettin et al., 2015), with an old contribution from groundwater that contains no young water (e.g., Georgek et al., 2017). Our results demonstrate, for the first time, that this bimodal distribution could emerge from the processes simulates in a fully-distributed, time-varying catchment model. The second reason is that studies typically fit age-based distributions to unimodal distributions (e.g., gamma, beta) (van der Velde et al., 2014). Future work may try to fit SAS functions to bimodal distributions, as has been done using binary exponential piston flow models in Morgenstern et al. (2015).

### 4.7.5 Implications for lumped parameter models

The present analysis offers insight into the consistency of two complementary approaches to hydrologic modeling: fully distributed and lumped parameter. While this analysis is based on fully distributed modeling, the outfluxes were divided into flowpaths that are commonly used in lumped parameter modeling. This has two useful aspects. First, it suggests that the very complex fully-distributed model might be reasonably emulated by a lumped parameter model that captures the basic trends presented here. Second, the features of our model can be compared with those of a lumped parameter model. The presence of inconsistencies may indicate model structural errors to be addressed in future work. For example, our results suggest that ages from groundwater flowpaths should be relatively constant, and that the overland flow path should deliver water with a relatively broad range of catchment transit times, especially under high storage conditions.

### 4.7.6 Implications for C-Q relationships

The relationship between concentration of solutes ( $C$ ) and discharge  $Q$  varies significantly between river systems (e.g. Zhang, 2018). As  $Q$  increases,  $C$  can either increase (i.e., concentration), decrease (i.e., dilution), or stay the same (i.e., chemostasis). In some watersheds there is a strong, consistent C-Q relationships; in others it may be highly scattered and time-varying. As discussed in Section 4.3, the concentration of hydrochemicals in water reaching the stream is sometimes believed to be linked to its transit times and pathway, which is in turn is often assumed to be closely related to storage and discharge level. Yet, our results show how storage and discharge are imperfect proxies for transit times of discharge and its component flowpaths. For example, Figure 4.7 (k) and (l) have roughly equal discharge but different age distributions and constituent flowpaths. The fact that similar discharge can have different component flowpaths could be a significant source of variability in the C-Q relationship, which could be modeled and explained using the approaches described here.

### 4.7.7 Limitations and future work

The ParFlow-CLM modeling had a number of limitations that were discussed in the conclusion of Chapter 3. In addition, a limitation of this work was the inability to evaluate the simulated age-based distributions against significant field data, especially at the level of individual flow pathways.

Notwithstanding the limited evaluation data, the hypotheses generated here can be tested in the field and used in turn to improve future watershed models, along the lines of the iterative dialogue between modeler and field experimentalist proposed by Weiler and McDonnell (2004). In addition, the methods can be repeated in other types of watersheds (e.g., semi-arid, high relief) to better understand the influence of catchment characteristics on watershed transit times.

## 4.8 Appendix A - Estimating $S_T$ using forward TTDs

This appendix derives an equation for calculating the age-rank storage  $S_T$  from the forward TTD of  $Q$  and  $ET$ . The derivation assumes the system is a simple control volume with influx  $J$  (with age zero) and outfluxes  $Q$  and  $ET$ . As shown by Rinaldo et al. (2011) using somewhat different notation, total catchment storage can be expressed as:

$$S(t) = \int_{T=0}^{T=\infty} J(t-T)P_{EX}(T, t-T)dT \quad (4.21)$$

where the exceedance probability  $P_{EX}(T, t-T)$  is the probability that water entering the catchment at time  $t-T$  exits with age greater than  $T$  (and is thus still in storage at time  $t$ ). The upper limit of integration of Equation 4.21 can be truncated at  $T$  to get the volume in storage with age less than or equal to  $T$ , which is the age-rank storage:

$$S_T(T, t) = \int_{\tau=0}^{\tau=T} J(t-\tau)P_{EX}(\tau, t-\tau)d\tau \quad (4.22)$$

where  $\tau$  is a dummy variable of integration. The exceedance probability can be written as:

$$P_{EX}(T, t-T) = 1 - \overrightarrow{P}_Q(T, t-T)\theta_Q(t-T) - \overrightarrow{P}_{ET}(T, t-T)\theta_{ET}(t-T) \quad (4.23)$$

which equates the probability of a parcel of water that enters at time  $t-T$  leaving after time  $t$  (the LHS) with the complement of the probability of a parcel of water that enters at time  $t-T$  exiting as either  $Q$  or  $ET$  before time  $t$  (the RHS). Combining Equations 4.22 and 4.23 gives Equation 4.19 in the main body of the text.

# Chapter 5

## Improved simulation of groundwater transit times in SWAT using StorAge Selection functions.<sup>1</sup>

### 5.1 Highlights

- The SWAT model was enhanced using StorAge Selection (SAS) functions to better simulate groundwater transit times (TTs).
- The prototype SWAT-SAS model was able to estimate and calibrate TTs to reference data at a small study catchment.
- The approach could improve representation of nitrate lag times in many watershed management models.

### 5.2 Summary

Intensive agriculture and atmospheric deposition add nitrogen to the landscape, which leaches as nitrate into groundwater aquifers. Depending in part on local groundwater transit times, nitrate can be stored and released over years to decades, which must be anticipated in watershed management plans. A commonly-used, open-source model for nitrate transport is the Soil and Water Assessment Tool (SWAT). The ability of SWAT to simulate nitrate transport is limited because (1) SWAT

---

<sup>1</sup>This chapter will be submitted for publication in a peer-reviewed journal. Ciaran Harman was involved in hypothesis development, study design, and results interpretation. Christina Cho and Shuyu Shang provided background research on the SWAT model. Daniel Fuka initialized and ran the SWAT model. Portions of the introduction were adapted from a report written by Dano Wilusz and submitted in partial fulfillment of the requirements for course EN.570.647 and used with permission of the instructor. The SWAT-SAS code and runs, the first draft of text, and all figures and tables were created by Dano Wilusz.



assumes that groundwater is well-mixed, which could be a poor assumption in many watersheds, and (2) SWAT does not explicitly calculate transit times, which precludes calibration against available data. To overcome these limitations, we demonstrate proof-of-concept of a modified version of SWAT that (1) allows the user to relax the well-mixed groundwater assumption using the recently developed theory of StorAge Selection (SAS) functions, and (2) calculates and outputs groundwater travel times to facilitate interpretation and evaluation. The resultant SWAT-SAS concept is described and applied at a small watershed in the Eastern Shore, MD, USA. Compared with original SWAT, we show that SWAT-SAS can simulate transit times that are much more consistent with available calibration data. We conclude that SAS functions show promise for improving simulation of groundwater transit times in SWAT and could be useful in other lumped-parameter watershed management models.

### 5.3 Introduction

Since the 18th century, the rate of anthropogenic nitrogen (N) loading to the surface of the earth has increased 10-fold to 150 Tg N per year (Gruber and Galloway, 2008). Sources of the increase include N fertilization, fossil fuel combustion, biological fixation by crops, and deforestation and other losses of long term N storage pools (Breuer et al., 2008). The increase has upset the delicate balance of N-availability in many watershed systems and caused environmental and human health problems including biodiversity loss, increases in nitrophile species, eutrophication, and drinking water contamination (Galloway and Cowling, 2002). As a result many jurisdictions in the U.S. and elsewhere are implementing watershed management action plans with the goal of reducing the delivery of excess N to sensitive surface waters (e.g., Linker et al., 2013b).

A significant portion of surface water N loading originates from one particular species and flowpath: nitrate-N ( $NO_3^-$ -N) in groundwater recharge to streams. Unlike other relatively reactive nitrogen species (e.g., ammonia-N), nitrate-N is relatively stable in groundwater with potentially long residence times, especially in carbon-depleted and/or aerobic environments. Under these conditions, a significant portion of nitrogen transport through the landscape can occur as dissolved nitrate. In the Chesapeake Bay Watershed, for example, an estimated 50% of the annual nitrogen load to the bay derives from nitrate-N in groundwater recharge (STAC, 2013).

The movement of nitrate through watersheds can exhibit large "lag times" that should be anticipated in nitrate control strategies (Meals, Dressing, and Davenport, 2010). For purposes here, we define the nitrate lag time as the time between (1) a change in nitrate inputs to the landscape and (2) a corresponding change in nitrate outputs to surface waters. Large nitrate lag times can be associated with the accumulation of large stores of "legacy nitrate" in watersheds as nitrate slowly travels from injection to discharge. Recent work suggests that large lag times and legacy nitrate stores are a feature of agricultural catchments around the world (Meter et al., 2016). The presence of nitrate lag times can delay the watershed response to best management practices and, if not considered in planning, jeopardize the achievement of watershed restoration goals (Van Meter, Van Cappellen, and Basu, 2018). Many factors can contribute to the emergent lag time of nitrate including nitrogen cycling and - to the extent that water acts as the mobile phase for dissolved nitrate transport - the transit times of water through different watershed flowpaths.

To the extent that water acts as the mobile phase for dissolved nitrate, nitrate lag times can be heavily influenced by groundwater transit times through shallow aquifers (Sanford and Pope, 2013). Over the past 20 years, our understanding of groundwater transit time distributions (TTDs) has significantly improved due to a combination of inverse modeling with tracer data and particle tracking through physically-based models. For this work, we define the groundwater TTD  $\overleftarrow{P}_{GW}(T, t)$  as the distribution of ages  $T$  of baseflow discharged from the groundwater aquifer to the stream at time  $t$ . The age  $T$  is equivalent to the transit time of water from injection to baseflow, since all water is assumed to be injected at age zero. Studies using time-series of environmental tracer inputs and outputs typically infer TTDs with medians on the order of days to decades.

The Soil and Water Assessment Tool (SWAT) is a widely used watershed model developed to help simulate the fate and transport of water and its solutes, including nitrate (Neitsch et al., 2011). SWAT is a daily time-step, semi-distributed, process-based river basin model built on algorithms first developed by the U.S. Department of Agriculture in the 1970s (Krysanova and Arnold, 2008). SWAT is designed for modeling mesoscale catchments, defined by Breuer et al. (2008) as ranging in size from 10 km<sup>2</sup> to 10,000 km<sup>2</sup>. Extensive documentation and a GIS-based user interface makes SWAT relatively user friendly. Perhaps as a result, SWAT is one of the most commonly applied watershed model by researchers and practitioners worldwide. At the time of the 2014 International

SWAT Conference, over 1700 peer-reviewed articles in at least six different languages were published and cataloged into an online database. Approximately 10% of these were related to the fate and transport of nitrate.

Despite the popularity of SWAT, the current version has at least two major limitations for simulation and calibration of realistic groundwater transit times, which can be critical for modeling overall nitrate transport and lags. First, SWAT does not explicitly simulate groundwater transit times. The model does not report groundwater travel times, which precludes the possibility of directly calibrating SWAT modeling results to increasingly-common groundwater transit time data. Second, SWAT imposes that all discharge from the shallow groundwater aquifer be uniformly selected from total aquifer storage. This "uniform selection" assumption is computationally efficient, but not supported by significant field data.

StorAge Selection (SAS) functions are a promising new approach for explicitly simulating and calibrating dynamic TTDs in lumped parameter models. The SAS function of flow exiting a control volume describes the distribution of storage, from youngest to oldest, that contributes to that outflow. SAS functions have been used to describe the transit time dynamics in different hydrologic systems including experimental lysimeters (Kim et al., 2016), the hyperheic zones of streams (Harman, Ward, and Ball, 2016), small experimental catchments (Harman, 2015), and large regional river basins (Danesh-Yazdi et al., 2016). SAS functions are applied to an entire control volume and therefore compatible with lumped or semi-lumped (i.e., semi-distributed) modeling frameworks. Unlike TTDs, the shape of SAS functions may be roughly independent of hydrometeorological inputs (Kim et al., 2016), and work is ongoing to link its shape to catchment characteristics.

The goal of this short technical note is to (1) describe how SWAT and SAS functions could be coupled in a new modeling framework called SWAT-SAS to better simulate and calibrate groundwater TDDs in mesoscale catchments and (2) demonstrate proof-of-concept for SWAT-SAS in a case study site in the Eastern Shore, MD, USA. The remainder of the note is organized as follows. Section 5.4 describes how SWAT currently models groundwater transit times, and how SWAT and SAS could be coupled to improve on current capabilities. Section 5.5 described how SWAT-SAS was developed and tested at a realistic modeling testbed. Finally, Section 5.6 discuss the implications, limitations, next steps, and opportunities to extend the approach to other lumped-parameter, semi-distributed

watershed models.

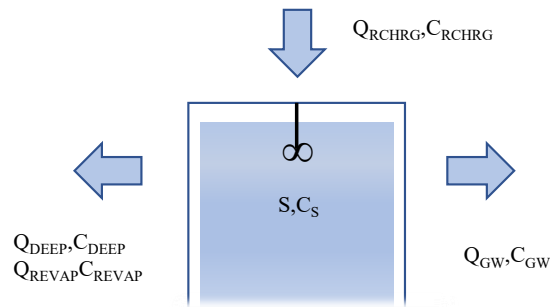
## 5.4 SWAT-SAS theory

The SWAT model is based on a network of conceptual storage volumes that receive and discharge water and hydrochemicals according to a multitude of parameterized governing equations representing different hydrologic and biogeochemical processes. The aim of SWAT-SAS is to improve the representation of groundwater transit times through the shallow aquifer control volume, which has an important control on the transport of nitrate. Due to the compartmentalized model structure, the changes required to implement SWAT-SAS are isolated to equations governing the shallow aquifer control volume and a few connecting flowpaths. We begin with an overview of how SWAT models nitrate transport, and then focus on processes in the shallow aquifer.

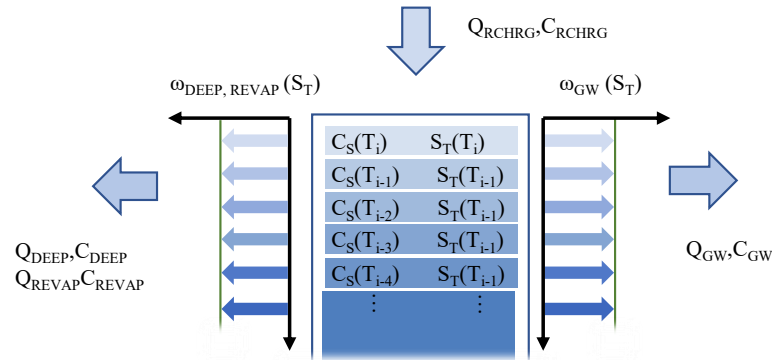
SWAT simulates nitrate transport to the edge of stream in three basic steps, as detailed in Neitsch et al. (2011). First, the watershed is delineated into Hydrological Response Units (HRUs). HRUs are land areas that are most alike based on average surface slope, soil type, and land use. HRUs are not necessarily conterminous and can be distributed throughout the landscape. Each HRU has a unique set of model parameters, state variables, and flux contributions to the edge of stream. Second, the daily hydrology is simulated through each HRU. In general, input requirements include rainfall and other hydrometeorological data. Rainfall onto the catchment is partitioned into either ET, surface water runoff, or infiltration. Infiltration moves through the soil zone and a portion is uptaken by plants. The remainder flows laterally to surface waters or leaches down into the vadose zone, where it percolates into the shallow aquifer. From there, governing equations determine the amount and timing of flow to each of three destinations: the deep aquifer as deep recharge (where it does not contribute to the stream), back up to the vadose zone as "revap", and out to the stream as groundwater discharge. Third, the fate and transport of nitrate is simulated through the catchment domain. The sources of N inputs into the model include atmospheric deposition, fertilizer application, nitrogen fixation, and initial stores in the soil. The N species move through the upper soil undergoing while biogeochemical processes including decomposition, immobilization, mineralization, nitrification, and denitrification. Nitrogen is transported out of the HRU as either sediment-bound organic N or dissolved nitrate. Some of the dissolved nitrate seeps as recharge into the shallow aquifer, where

additional denitrification may occur. The next section describes how nitrate moves through the shallow aquifer in SWAT and the proposed SWAT-SAS.

(a) Conventional Conceptualization for SWAT Shallow Aquifer (after Arnold 2009).



(b) Proposed conceptualization with StorAge Selection (SAS) Function theory (after Harman 2015).



**Figure 5.1:** Conceptualizing of SWAT and proposed changes. Panel (a) shows how nitrate transport is conceptualized in SWAT as a uniform selection from a control volume. Panel (b) shows the proposed approach using SAS functions to different proportions select age-ranked storage. The SAS functions are shown in the PDF form  $\omega(S_T)$ , and the shape is indicative of uniform selection.

### 5.4.1 Age and nitrate modeling in SWAT

The current conceptualization of the shallow aquifer nitrate transport module in SWAT is shown in Figure 5.1. The flux of water through the shallow aquifer in SWAT is constrained by a numerical solution to the water balance equation:

$$\frac{dS}{dt} = J_{rchrg} - Q_{gw} - Q_{deep} - Q_{revap} \quad (5.1)$$

where  $S$  is the storage in mm of water in the shallow aquifer,  $J_{rchrg}$  is the flux in mm/day of recharge entering the shallow aquifer,  $Q_{gw}$  is the flux in mm/day into the main channel (i.e., groundwater baseflow),  $Q_{deep}$  is the flux in mm/day of water entering the deep aquifer (where it leaves the simulation), and  $Q_{revap}$  is the flux in mm/day of water moving back up into the soil zone in response to water deficiencies.

The concentration of nitrate in fluxes leaving the shallow aquifer is determined by assuming that the concentrations of outflows are equal to the concentration in storage (see conceptualization in Figure 5.1). Thus:

$$C_S = C_{gw} = C_{deep} = C_{revap} \quad (5.2)$$

where  $C_S$ ,  $C_{gw}$ ,  $C_{deep}$ , and  $C_{revap}$  are the concentration of nitrate in mg/l in  $S$ ,  $Q_{gw}$ ,  $Q_{deep}$ , and  $Q_{revap}$ , respectively. After the hydrology is simulated, the concentrations of outfluxes can be determined by combining Equation 5.2 with the nitrate mass balance equation:

$$\frac{d(C_S S)}{dt} = C_{rchrg} J_{rchrg} - C_{gw} Q_{gw} - C_{deep} Q_{deep} - C_{revap} Q_{revap} \quad (5.3)$$

Where  $C_{rchrg}$  is the concentration of  $J_{rchrg}$  in mm/day determined by the SWAT model for the upper soil zone. An initial value of  $S$  and  $C_S$  must be assumed. SWAT also models groundwater pumping and first order nitrate decay due to denitrification, which are relatively simple to include by extension.

It is important to note SWAT does not make any explicit calculation of transit times through the shallow aquifer. However, Equation 5.3 is derived from two implicit assumptions: (1) that fluxes out of the shallow aquifer are uniformly sampled from shallow aquifer storage and (2) that nitrate moves

with the water. Given these assumptions, the TTD for groundwater in PDF form ( $\overleftarrow{p}_{gw}(T, t)$ ) and other outfluxes under steady-state has the well-known solution of an exponential distribution with mean age  $T = S_0/Q_0$  (Małozzewski and Zuber, 1982):

$$\overleftarrow{p}_{gw}(T, t) = \frac{Q_0}{S_0} e^{-\frac{TQ_0}{S_0}} \quad (5.4)$$

where  $Q_0$  is the steady-state flow through the shallow aquifer (which may be partitioned between the multiple outfluxes) and  $S_0$  is the steady-state storage. The solution to the TTD shown in Equation 5.4, however, is not valid for unsteady flows simulated in SWAT. While Equation 5.3 does not have simple analytical solution for the non-steady case, numerical solutions can be derived (e.g., Ali, Fiori, and Russo, 2014).

This uniform-sampling approach to nitrate transport modeling has several attractive features for lumped parameter watershed modeling. First, the approach is computationally efficient and only requires state variable values for the current and previous timestep. Second, the equations ensure conservation of the mass and age of water moving through the system. Third, the approach seems to capture important first-order behaviors, including faster transit times when fluxes through the system are high or when total storage is low, and a general spreading of ages that might occur as water moves through multiple pathways. Finally, the approach is consistent with physically-based (but highly idealized) groundwater transport theory, which also suggests that sampling should be uniform with an exponential distribution (Haitjema and Mitchell-Bruker, 2005).

In many situations, however, the uniform-sampling approach may be overly simplistic. First, processors are becoming more powerful, so computational efficiency is less important. Second, water age estimates derived from tracers (e.g., McGuire and McDonnell, 2006) and particle tracking (e.g., Rooij, Graham, and Maxwell, 2013) is increasingly common, so tools are needed to calculate and report model water ages for calibration and validation. Third, recent studies are showing a range of groundwater age distributions that do not seem consistent with uniform sampling, including distributions with a bimodal peak (e.g., Morgenstern et al., 2015) or significant time-variability (Peters, Burns, and Aulenbach, 2014). The exact nature of the groundwater age distribution is still poorly understood, which seems to call for a flexible approach to defining distributions.

### 5.4.2 Proposed age and nitrate modeling in SWAT-SAS

In this section we propose a more transparent and flexible approach to simulating age distributions and nitrate fluxes in the SWAT shallow aquifer. The approach replaces the uniform sampling assumption with a StorAge Selection (SAS) function, which can be parameterized to represent whatever sampling is most appropriate. The theory of SAS functions has been developed in a series of papers including Botter, Bertuzzo, and Rinaldo (2011), van der Velde et al. (2012), and Harman (2015) and is summarized in a review paper by (Rinaldo et al., 2015). There are several different formulations of SAS functions in the literature. This work most closely follows the formulation presented in Harman (2015).

The proposed conceptualization is shown in Figure 5.1(b). In the shallow aquifer of SWAT-SAS, the distribution of ages of water in storage is explicitly tracked using a new state variable called age-ranked storage  $S_T(T, t)$  which varies as a function of age  $T$  and time  $t$ :

$$S_T(T, t) = P_S(T, t)S(t) \quad (5.5)$$

where  $P_S(T, t)$  is the cumulative residence time distribution of ages  $T$  in storage at time  $t$ . The age-rank storage  $S_T(T, t)$  has units of length (i.e., mm) and is the volume of water in storage with age less than or equal to  $T$  at time  $t$ . Water entering the storage volume as  $J_{rchrq}$  is assumed to have age  $T=0$ . For example, a value of  $S_T(T = 3 \text{ days}, t = 50 \text{ days}) = 8 \text{ mm}$  indicates that the volume of storage in the shallow aquifer with an age of 3 days or less on the day 50 of the simulation is equal to 8 mm.

In addition to calculating  $S_T$ , SWAT-SAS also defines a new probability distribution for each outflow called the SAS function  $\Omega(S_T, t)$ . As illustrated in Figure 5.1(b), the SAS function defines the probability of selecting water from different age-ranked storages. For example, the value  $\Omega_{gw}(S_T = 8 \text{ mm}, t = 50 \text{ days}) = 0.10$  indicates that the fraction of  $Q_{gw}$  coming from the youngest 8 mm of storage on the 50th day is equal to 10%. The distribution of age-ranked storage that contributes to an outflow can be used to directly calculate the TTD of outflows. Building on the previous example, if the youngest 8mm of rain in storage originated from a recharge event 3 days prior, then the youngest 10% of  $Q_{gw}$  must be 3 days old. If, on the other hand, there has been a



drought and the youngest 8mm of water in storage is 100 days old, then the same SAS function would indicate that the youngest 10% of  $Q_{gw}$  is 100 days. In this manner, the invertible relationship between  $S_T$  and  $T$  can be exploited to transform values of  $\Omega(S_T, t)$  to values of  $\overleftarrow{P}(T, t)$ . Further, time-invariant SAS functions can yield time-varying transit times that respond to changing inputs.

SAS functions can exactly reproduce the original SWAT conceptualization of uniform sampling, but can also represent other kinds of sampling. The SAS functions in Figure 5.1(b) (in PDF form) are a constant value that uniformly sample from all age-ranked water in storage. These SAS functions embody the implicit assumption made in SWAT that water flowing out of the shallow aquifers is uniformly sampled from the water in storage. Alternatively, SAS functions could be constructed to represent other forms of selection, including the ideas that (1)  $Q_{revap}$  originates from the upper layer of the aquifer and is therefore more likely to be sampled from younger age-ranked storage and (2)  $Q_{deep}$  originates from the deepest part of the aquifer and is more likely to be sampled from older age-ranked storage.

The SWAT-SAS framework is based on an expression of conservation of mass for each age of water. A "master equation" (Botter, Bertuzzo, and Rinaldo, 2011; Harman, 2015) describes the evolution of age-ranked storage and TTDs in time:

$$\frac{\partial S_T(T, t)}{\partial t} = J_{rchrg}(t) - \left( Q_{gw}(t) \overleftarrow{P}_{gw}(T, t) + Q_{deep}(t) \overleftarrow{P}_{deep}(T, t) + Q_{revap}(t) \overleftarrow{P}_{revap}(T, t) \right) - \frac{\partial S_T}{\partial T} \quad (5.6)$$

Equation 5.6 expresses that the rate of change of shallow aquifer storage younger than age  $T$  (the LHS) is equal to the rate of recharge (assumed to have age zero), minus the rate of all discharge leaving the storage with age less than  $T$ , minus the rate of water in storage becoming older than  $T$  (the RHS). As described above, the TTD for each flux, which is expressed as a distribution of  $T$ , can be transformed into a distribution of  $S_T$ :

$$\overleftarrow{P}_i(T, t) = \Omega_i(S_T, t) \quad \text{for } i \in gw, deep, revap \quad (5.7)$$

Given the time-series of each flux, the SAS functions, and an initial condition for  $S_T(T, t = 0)$ , Equations 5.6 and 5.7 can be solved numerically to determine  $S_T(T, t)$  and  $\overleftarrow{P}(T, t)$  for each outflow.

Finally, the time-series of nitrate inputs  $C_{rchrg}(t)$  and the groundwater TTD  $\overleftarrow{P}_{gw}(T, t)$  can be

used to estimate the groundwater nitrate concentration  $C_{gw}(t)$ . Conceptually, this is done by linking each age of water in storage with its injection concentration, as illustrated in Figure 5.1. In this way, knowing the distribution of ages in  $Q_{gw}$  is sufficient to calculate the groundwater concentration. This is calculated as the convolution of the input concentration with the pdf form of the TTD (Barnes and Bonell, 1996):

$$C_{gw}(t) = \int_0^{\infty} \overleftarrow{p}_{gw}(\tau, t) C_{rechg}(t - \tau) e^{-\lambda\tau} d\tau \quad (5.8)$$

where  $\lambda$  could represent a first order removal of nitrate due to denitrification. Similar equations can be derived for  $Q_{deep}$  and  $Q_{revap}$ .

## 5.5 SWAT-SAS case study

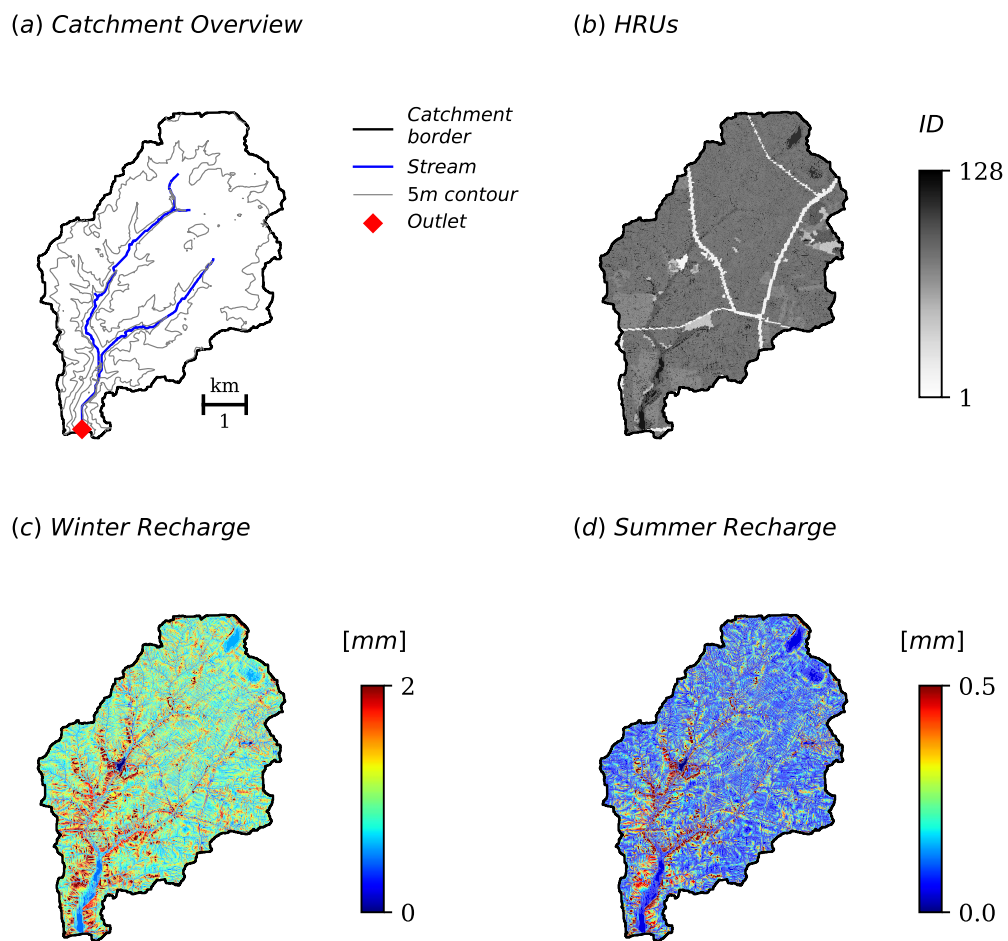
A simple SWAT model was built, modified, and tested to demonstrate proof-of-concept for the SWAT-SAS approach to estimating groundwater transit times. Although the model set-up used observation-based meteorological forcing and catchment characteristics, it was not calibrated and was not designed to accurately simulate most aspects of hydrology. Rather, the case study model was used here as a testbed to develop the SWAT-SAS software architecture and demonstrate its capabilities in a catchment-scale SWAT model with realistic features. Nitrate concentrations were not examined directly, but the effect of changes in TTDs on nitrate transport is discussed in light of Equation 5.8.

### 5.5.1 Study site and SWAT model setup

The SWAT-SAS testbed was built for the Chesterville watershed on the Delmarva Peninsula in Maryland, MD (see map in Figure 5.2(a)). The catchment feeds into the Chester River, which drains into the Chesapeake Bay. The landscape is highly agricultural, which contributes to high legacy nitrates in groundwater storage and high nitrate concentrations in streams (e.g., Denver, 1995). A deep, sandy unconfined aquifer supplies stream baseflow, which accounts for roughly 80% of discharge and 30% of total nitrate load. The watershed has been studied extensively by the U.S. Geological Survey (USGS) (Hyer et al., 2016), which maintains a stream gauge at the outlet and regularly collects hydrochemical samples. Although the present study is focused more on model development

than site-specific discovery, the Chesterville Branch was chosen because it is data-rich and has high nitrate concentrations with potentially long lag times, making it a good candidate for future work.

The ArcSWAT application (Winchell et al., 2010) was used to delineate the watershed into 128 HRUs using soil type, land use, slope, and topographic index. Soil type was obtained from a 30m resolution STATSGO soil map (Sheshukov et al., 2011), and land-use / land-cover was obtained from



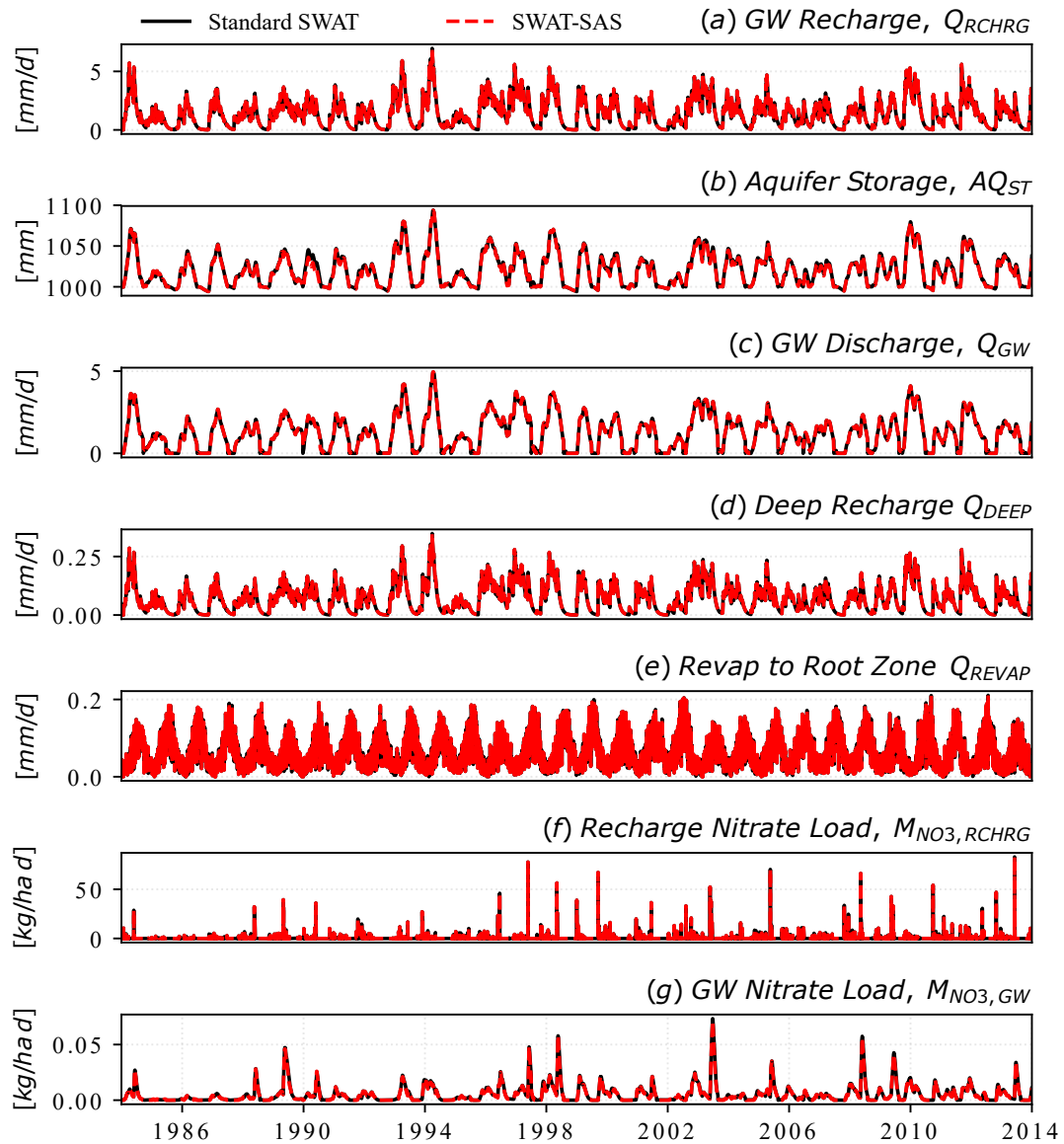
**Figure 5.2:** Map of the case study site and hydrology including the catchment site (Panel a), the distribution of HRUs (Panel b), and the amount and distribution of winter recharge (Panel c) and summer recharge (Panel d).

the 2011 National Land Cover Database (Homer et al., 2015) with 10m resolution. A 10m resolution USGS DEM was used to calculate slope and topographic index. Based on these four characteristics, the watershed was delineated into 128 HRUs with similar characteristics, shown in Figure 5.2(b). Note that topographic index is not typically used in SWAT to delineate HRUs, but was applied here using the "TopoSWAT" modifications detailed in Fuka et al. (2016) and Easton et al. (2008). The use of the topographic index helps to correctly demarcate and parameterize HRUs that are most likely to contribute saturation excess overland flow. The model was forced with hourly meteorological data (e.g., rainfall, solar radiation, relative humidity) from the SWAT weather generator, which draws from locally available observations. The parameters for each HRU were assigned the default values, which were set based on the range of reasonable values in the literature.

The model was run for the 30-year period 1984-2014 and state variables relevant to the shallow aquifer were extracted from the output:  $Q_{rchrg}$ ,  $S$ ,  $Q_{gw}$ ,  $Q_{deep}$ ,  $Q_{revap}$ , and the recharge nitrite load  $M_{rchrg} = Q_{rchrg}C_{rchrg}$  and  $M_{gw} = Q_{rchrg}C_{rchrg}$ . The results for a representative HRU are shown in Figure 5.3. The values of the inflows, storage, and outflows are dynamic with different magnitudes and seasonal patterns that must be accounted for in any transit time modeling framework. For example,  $Q_{rchrg}$  has high recharge in winter and low recharge in summer, which drives a similar seasonal pattern in  $S$ . Of the fluxes leaving the shallow aquifer,  $Q_{gw}$  is roughly an order of magnitude larger than  $Q_{deep}$  and  $Q_{revap}$ . The seasonal recharge seems to drive  $S$ ,  $Q_{gw}$ , and  $Q_{deep}$  to peak in the winter. The  $Q_{revap}$ , however, has the opposite pattern and peaks in the summer, presumably due to ET-driven soil moisture deficits in the upper soil layers. The nitrate load entering in recharge  $M_{rchrg}$  is relatively flashy, while the load exiting in  $M_{gw}$  is more dispersed. As shown in Equation 5.8, the translation of nitrate inputs to groundwater outputs are a function of the groundwater TTD  $\overleftarrow{P}_{gw}(T, t)$ , even though the TTD is not explicitly calculated. The SWAT simulation also shows significant spatial variability across HRUs. The maps in Figure 5.2(c-d) show that, regardless of season, recharge is generally higher in the riparian areas and generally lower further from the streams.

### 5.5.2 SWAT-SAS implementation

The results from the SWAT model were reanalyzed using the SWAT-SAS conceptualization described in Section 5.4. The SWAT output were imported into Python scripts (v2.7) that run SAS using



**Figure 5.3:** Time-series of inputs and outputs for a representative HRU using the default parameterization. The results demonstrate that SWAT-SAS can successfully reproduce SWAT simulations.

Python 2.7 This "soft-coupling" between SWAT and SAS was simpler to develop than a more integrated "hard-coupling". The capabilities of the SWAT-SAS model were demonstrated in three phases of increasing complexity. First, the groundwater TTDs were estimated and visualized using the default parameters. Second, the groundwater TTDs were calibrated to a reference groundwater TTD using identical SAS parameters in all HRUs (i.e., ignoring spatial heterogeneity). Third, the groundwater TTD was calibrated separately to different regions of the catchment (i.e., incorporating spatial heterogeneity).

### 5.5.2.1 Default parameterization

SWAT-SAS was used to replicate the nitrate transport in regular SWAT by using a uniform SAS function for each outflow. The SAS function for  $Q_{gw}$  was described by:

$$\Omega_{gw}(S_T, t) = \begin{cases} \frac{S_T}{S(t)}, & S_T \in [0, S(t)] \\ 1, & S_T > S(t) \end{cases} \quad (5.9)$$

with similar definitions for  $Q_{deep}$  and  $Q_{revap}$ . Note that combining Equations 5.5, 5.7, and 5.9 shows that, for the uniform sampling case, the residence time distribution  $P_S(T, t)$  is equal to the backward TTD  $\overleftarrow{P}_{gw}(T, t)$ .

The results from SWAT-SAS and regular SWAT using default parameters show very good agreement (see Figure 5.3). The models used different numerical schemes to solve the governing hydrology and transport equations, but the results are nearly identical (Figure 5.3(a-e)). Similarly, the models used different approaches to simulating uniform selection but estimate the same nitrate concentration in the outflux (see Figure 5.3(g)).

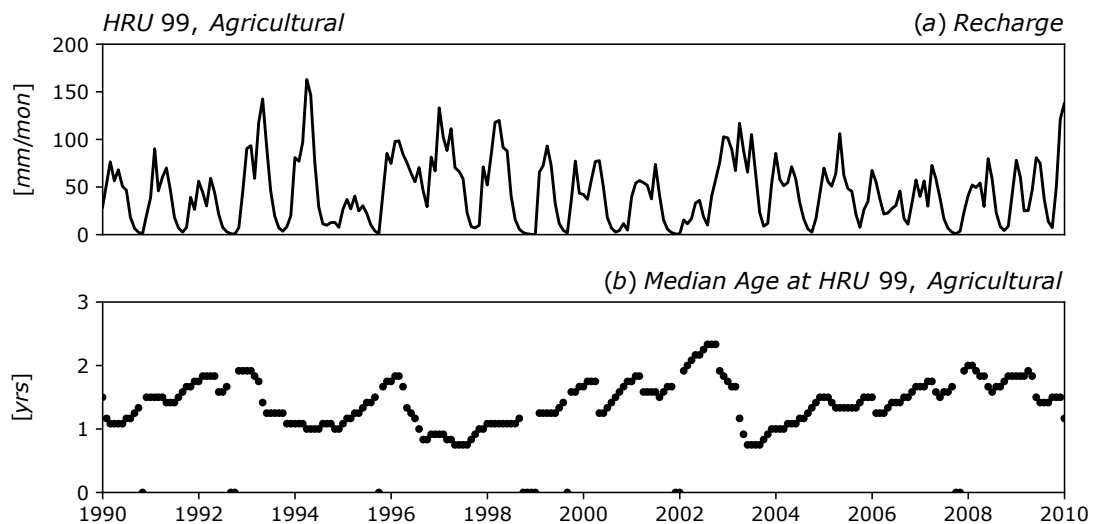
SWAT-SAS was used to estimate the groundwater TTD for each HRU in the model. The overall median age of  $Q_{gw}$  was 2.3 years, with an interquartile range of 0.9 years to 5.1 years with high spatial and temporal variability. Figure 5.4 shows an illustrative example for the agricultural HRU number 99. In this particular HRU, the median daily age is 1.4 years with an interquartile range of 1.1-1.7 years. The year 1994 illustrates how wetter periods result in with faster flushing and younger outflows. Conversely, the year 2003 shows how dryer periods result in older outflows.

The spatial distribution of the median groundwater age across all HRUs is shown in Figure 5.5(b). Because the default parameters in the shallow aquifer are identical across HRUs, the spatial variability is entirely driven by differences in  $Q_{rechg}$ , which are shown in Figure 5.2(c-d).

### 5.5.2.2 Uniform calibration

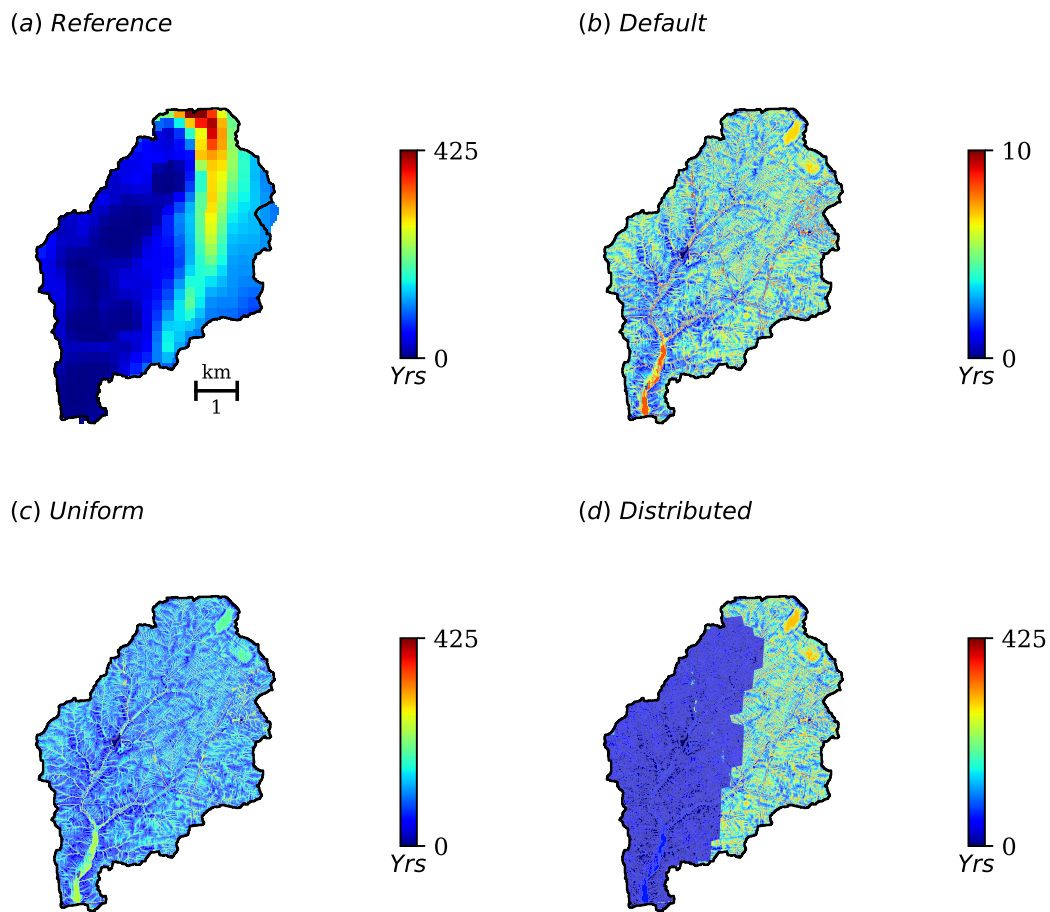
SWAT-SAS revealed that the median groundwater TTD using the default parameterization was 2.3 years. Is this a reasonable value, and if not, how could it be calibrated to data? Recent studies show that - even in models that fit the hydrograph - the simulation of transit times can be poorly constrained because of the difference between catchment celerity and velocity (McDonnell and Beven, 2014). In fact, in the Chesterville catchment, evidence suggests that baseflow ages should be much older than values obtained from the default parameters. Due to its flat terrain and deep, sandy aquifers, groundwater ages in this region may be decades or even centuries (Sanford et al., 2012).

The SWAT-SAS model was calibrated to match a set of reference ages extracted from a recent USGS report (Sanford et al., 2012). In the report, a MODFLOW groundwater model was built using local hydrogeologic data and used to track the transit time of particles through the shallow aquifer.



**Figure 5.4:** SWAT-SAS simulations of recharge (Panel a) and median transit time distribution (panel b) for a representative HRU using the default parameterization.

The results from the report were digitized, georeferenced, and used to construct the empirical CDF shown in Figure 5.5(a). The reference ages provide an example of the kind of age data to which the SWAT-SAS data might be calibrated. The shape of the backward TTD is similar to the shape of a CDF of an exponential distribution, with a median age of 69.2 years and an interquartile distribution of 28.7 - 138.4 years. The long transit times required the SAS modeling to be run for 300 years,



**Figure 5.5:** Map of spatially distributed groundwater ages including the reference age distribution taken from Sanford et al. (2012) (Panel a), along with the SWAT-SAS simulated ages using the default parameters (Panel b), the uniform calibration parameters (Panel c), and the spatially distributed calibration (Panel d).



which necessitated (1) extending the SWAT hydrology output by looping the 30-year simulation 10 times and (2) increasing the timestep from 1 day to 1 month to accelerate runtimes.

The SAS functions were assumed to select a uniform distribution of age-ranked storage, as they were for the default parameterization:

$$\Omega_{gw}(S_T, t) = \begin{cases} \frac{S_T}{S(t) + S_p}, & S_T \in [0, S(t) + S_p] \\ 1, & S_T > S(t) + S_p \end{cases} \quad (5.10)$$

with similar definitions for  $Q_{deep}$  and  $Q_{revap}$ . The key difference with the default parameterization is that the range of age-ranked storage that contributed to discharge was increased to  $S(t) + S_p$  where  $S(t)$  can be called "active storage" and  $S_p$  the "passive storage" (with the caveat noted below). In the hydrologic literature on lumped parameter models, "active storage" is the storage that controls the magnitude of discharge; by contrast, passive storage is additional storage that influences the velocity of water traveling through the control volume. For example, active storage is sometimes conceptualized as the water flowing through fractured rock, and passive storage as water captured between the fractures. The passive storage exchanges water with the active storage, but does not contribute to the connectivity of discharge.

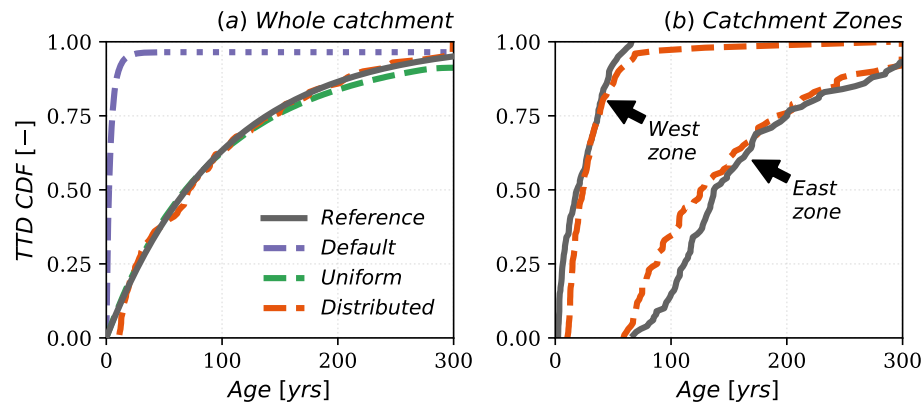
The SWAT-SAS parameter  $S_p$  was manually adjusted to visually fit the shape of the reference TTD. Figure 5.6(a) shows that the uniformly calibrated CDF matches the reference distribution. The spatial distribution of the median TTD is shown in Figure 5.5(c). The spatial variability is driven by differences in recharge and does not reproduce the spatial variability of the reference pattern shown in Figure 5.5(a).

We note that original SWAT can also be parameterized to uniformly select a combination of active and passive storage. In each HRU, the parameters  $aq_{shth,gw}$  and  $aq_{shth,revap}$  are the minimum amount of storage needed for  $Q_{gw}$  and  $Q_{revap}$  to occur, respectively. A uniform increase in these two parameters causes  $S(t)$  to increase by the same amount and creates an equal amount of new passive storage. These thresholds can be adjusted to incorporate a certain amount of passive storage into the value of  $S(t)$  simulated by SWAT-SAS. This implicit passive storage will be equal to the minimum of  $aq_{shth,gw}$  and  $aq_{shth,revap}$ . To ensure that  $S_p$  represents the true passive storage, at least one of these parameters should be set to zero.

Our results demonstrate how SWAT-SAS can be used to calibrate the parameters of all HRUs against a reference TTD distribution. A major limitation is that the spatial variability of the reference TTD is poorly represented. Another limitation is that the uniform SAS function can only simulate TTDs with a long-term average exponential shape; other types of SAS functions are needed for other distributions. We will address both these limitations in the next phase of calibration.

### 5.5.2.3 Spatial calibration

This section demonstrates how SAS functions for different HRUs can be given unique parameterizations to represent spatial variability. We observed that the reference distribution of groundwater ages is much older in the eastern zone than the western zone (Figure 5.5(a)). This may be important to capture in the SWAT-SAS model, especially when exploring the influence of beneficial management practices (BMPs) in different regions of the watershed. The HRUs, however, were not assigned on the basis on groundwater TTDs, such that HRUs can be located in regions with both old and young transit times (Figure 5.2(b)). To overcome this, the reference map was used to split the region into zones with the lowest 50% of transit times (the "west zone") and the highest 50% of transit times (the "east zone"). An empirical probability distribution was plotted for each zone (Figure 5.6(d))



**Figure 5.6:** Comparison of reference TTD with default parameters and spatially uniform parameterization (Panel a), along with the spatially distributed parameterization (Panel b).

with a median transit time of 18.7 y and 136.9 y for the east and west zone, respectively. The zones were used as an additional criteria for HRU selection, which nearly doubled their number.

The SAS functions were parameterized using a different distribution for each zone. For the younger west zone, the distribution looks approximately exponential (see Figure 5.6(b)) so the uniform SAS function was applied at each western HRU:

$$\Omega_{gw,W}(S_T, t) = \begin{cases} \frac{S_T}{S(t) + S_{p,W}}, & S_T \in [0, S(t) + S_{p,W}] \\ 1, & S_T > S(t) + S_{p,W} \end{cases} \quad (5.11)$$

with similar definitions for  $Q_{deep}$  and  $Q_{revap}$ , where the subscript  $W$  is used for the western zone. Although the form of the SAS function is identical to Equation 5.10, the passive storage parameter  $S_{p,W}$  should be significantly different to reflect the younger transit times of the system. For the older east zone, the youngest water reaching the stream is nearly 75 years old. This suggests that the youngest age-rank storage should not contribute to discharge. To account for this, a "piston-uniform" SAS function was applied at each eastern HRU:

$$\Omega_{gw,E}(S_T, t) = \begin{cases} 0, & S_T \leq S_{min} \\ \frac{S_T - S_{min,E}}{S(t) + S_{p,E} - S_{min,E}}, & S_T \in [0, S(t) + S_{p,E} + S_{min,E}] \\ 1, & S_T > S(t) + S_{p,E} + S_{min,E} \end{cases} \quad (5.12)$$

with similar definitions for  $Q_{deep}$  and  $Q_{revap}$ , where  $S_{min,E}$  is the youngest age-rank storage that contributes to discharge in the east zone. Thus, only the age-ranked storage between the volumes  $S_{min,E}$  and  $S_{min,E} + S(t) + S_{p,E}$  is uniformly selected into discharge.

The parameters  $S_{p,W}$ ,  $S_{p,E}$ , and  $S_{min,E}$  were manually tuned to fit the reference groundwater TTDs to the SWAT-SAS simulations (see figure 5.6(b)). For the west zone, a value of  $S_{p,W}$  of 7000 mm gave overall good agreement. For the east zone, values for  $S_{min,E}$  and  $S_{p,E}$  of 20000 mm and 34000 mm respectively fit the data reasonably well. Note that the parameter  $S_{min,E}$  was tuned to fit the minimum age of the discharge (i.e., 75 years) while  $S_{p,E}$  was tuned to fit the shape of the distribution. The spatial pattern of median groundwater age from each HRU is shown in Figure 5.5(d). The new SWAT-SAS configuration and calibration captures the disparity between the east

and west sides. There are other features, however, that are still not captured from the reference map, such as the north-south divide in the eastern zone. This could have been handled, for example, by using three zones to re-allocate the HRUs instead of two.

## 5.6 Discussion and conclusions

This paper has demonstrated a novel approach to simulating time-varying groundwater TTDs using the SWAT-SAS framework. An illustrative case study used SWAT-SAS to visualize and calibrate the simulated TTDs without and with spatial fidelity. The proposed SWAT-SAS has two main advantages over the current SWAT model. First, it permits direct calculation of the time-varying TTD in each HRU and their total TTD reaching the edge of stream. The modeled TTDs can be directly evaluated against available data. Second, the approach allows the selection of water for each outflux to be parameterized according to the data.

SWAT-SAS is intended to support more realistic simulation of groundwater TTDs and nitrate lag times for historic analysis as well as "what-if" scenario development for decision support. Although not used here, the baseline SWAT model simulates a range of beneficial management processes (BMP). The relationship between groundwater TTDs and lag times for each BMP might be different. For example, the lag time of BMPs that reduce nitrate near the point of entry into the aquifer (e.g., cover crops) might be heavily influenced by groundwater TTDs, whereas the lag time of BMPs that reduce nitrate at the point of egress (e.g., riparian buffers) might be independent of groundwater TTDs. These differences could be explored and quantified using SWAT-SAS.

### 5.6.1 Comparison with steady-state approximation

We compare SWAT-SAS with other methods to simulate groundwater TTDs and nitrate lag times in the SWAT framework. A status quo approach is to use the baseline SWAT and rely on steady-state approximations to estimate groundwater TTDs. As suggested in Equation 5.4, the mean residence time through the shallow aquifer can be approximated as the ratio of the mean storage to the mean recharge. In our case study, the steady-state approximation and SWAT-SAS simulation are likely to give reasonably similar values for the spatially-averaged TTD. The steady-state approximation does not, however, provide information about the time-variability of the TTD (e.g., Figure 5.3) and

is not valid when selection is not uniform (e.g., the piston-uniform selection used in the east zone). Notwithstanding the limitations, the steady-state approximation is a practical approach that might be useful in many settings where more comprehensive methods are not feasible.

### 5.6.2 Limitations and future work

A major limitation is data availability for calibrating groundwater TTDs. Even where data is available, is it often incomplete in time and space relative to the simulations of SWAT-SAS. In our case study, for example, the reference data did not provide any information about the time-variability. As data becomes more available, we believe the ability to estimate the TTD distributions in space and time will be increasingly important. In the interim, SWAT-SAS can still be useful for order-of-magnitude corrections. For example, even if our reference groundwater TTD were highly uncertain, we can have relative confidence that the TTDs simulated by the default parameters were much younger than actual field conditions. A second limitation is computational efficiency. The approach described here required additional HRUs to simulate the spatially distributed HRUs. In addition, the retention of the age-ranked storage variable in each HRU is memory intensive and could increase simulation times significantly. Future work should streamline the code to reduce computational burden.

# Chapter 6

## Conclusion <sup>1</sup>

This dissertation has explored the central hypothesis that recent advances in time-varying transit time modeling approaches can improve our understanding of the hydrologic transport at watershed scale. The first part of this final chapter is a synthesis of the key dissertation findings. The second part is an illustrative discussion of the implications of key findings on an important environmental issue: groundwater nitrate transport in the Chesapeake Bay Watershed. The short case study highlights conceptual links between this work and ongoing efforts to control nitrate transport as well as potential work areas for future exploration.

### 6.1 Summary of key findings

The combined scientific contribution of the four studies presented in Chapters 2-5 fall into three categories: applied, methodological, and fundamental. The applied contribution includes two site-specific analyses using recent advances in time-varying transit time modeling to show how hydrometeorological forcing influences catchment transit times, and how this influence might exert non-stationarity in catchment transit times under a changing climate. The methodological contribution is the articulation and demonstration of advanced approaches to modeling catchment transit times and quantifying their sensitivity to climate and catchment characteristics, with a conscious focus on open source modeling tools. The fundamental contribution is better understanding of how catchment-scale transport (as represented by the catchment transit time distribution) is driven by various hydrologic processes including natural rainfall variability and surface - groundwater interactions.

---

<sup>1</sup>A portion of the background information on the Chesapeake Bay watershed was adapted from a report written by Dano Wilusz and submitted in partial fulfillment of the requirements for course AS.270.641 and used with permission of the instructor.

The first study described in Chapter 2 uses catchment-scale lumped parameter models to decompose the relationship between rainfall variability and the time-varying fraction of young water (<90 days old) in streams. A coupled rainfall-runoff model and rank StorAge Selection transit-time model were calibrated to hydrometric and environmental tracer data from neighboring headwater catchments in Plynlimon, Wales from 1999-2008. At both sites, the mean annual young water fraction increased significantly from the driest to the wettest year. Yearly mean rainfall explained most between-year variation, but certain signatures of rainfall pattern were also associated with higher young water fraction. These signatures were symptomatic of an "inverse storage effect" that may be common among watersheds. The last phase of the study found that projected changes in climate due to global warming could seasonally alter the ages of water in streams at these sites, with concomitant impacts on water quality.

The second study described in Chapter 3 introduced a novel approach to calibrating a surface-groundwater model to both discharge and tracer data, and used the calibrated model to characterize the sensitivity of discharge and baseflow stream water age to changes in porosity, hydrologic conductivity, and other critical zone parameters under transient forcing conditions. The model used the fully-distributed surface-groundwater model ParFlow-CLM with SLIM-FAST particle tracking. The ParFlow-CLM simulation was configured and forced with extensive field data from the USDA's Mahantango Creek experimental catchment in PA, USA. The inclusion of age-related calibration criteria was shown to reduce parameter equifinality. Longer baseflow transit times were associated with higher porosities and a non-linear interaction between high lateral conductivity in the upper soil layer and low lateral conductivity in the underlying fractured bedrock. Overall, the results suggest that the age of baseflow may be significantly influenced by upper soil properties due to dynamic partitioning between recharge and overland flow, which is not captured in commonly used steady-state simulations.

The fourth study (Chapter 4) used the model calibrated in Chapter 3 to study how different flowpaths through the watershed contribute to the time-varying catchment transport. We found that the shape and time-variability of transit time distribution for evapotranspiration was readily explained by existing theory, though this may be a consequence of its simplified representation in the model. The shape and time-variability of discharge was more complex and evolved from

shifting contributions of five flowpaths: direct rainfall, overland flow, interflow, and shallow and deep groundwater. The transit times for different flowpaths overlapped substantially, suggesting that age alone cannot uniquely identify flowpath. The modeling demonstrated a mechanistic explanation for an inverse storage effect which was observed in Chapter 2 and many other studies. The chapter developed a taxonomy of TTD shapes that occur during a typical recession, each with a distinct composition of flowpaths. The results have implications for the conceptualization of lumped parameter models, the physical origins of SAS functions and their variability, and the origins of different C-Q relationships.

The final study (Chapter 5) is a technical note demonstrating proof-of-concept for a modified version of the commonly-used Soil and Water Assessment Tool (SWAT) that (1) allows the user to relax the well-mixed groundwater assumption using the recently developed theory of StorAge Selection (SAS) functions, and (2) calculates and outputs groundwater travel times to facilitate interpretation and evaluation. The resultant SWAT-SAS concept was applied at a small watershed in the Eastern Shore, MD, USA. Compared with original SWAT, the results show that SWAT-SAS can simulate transit times that are much more consistent with available calibration data. The study concludes that SAS functions show promise for improving simulation of groundwater transit times in SWAT and could be useful in other lumped-parameter watershed management models.

### **6.1.1 Application to the Chesapeake Bay watershed**

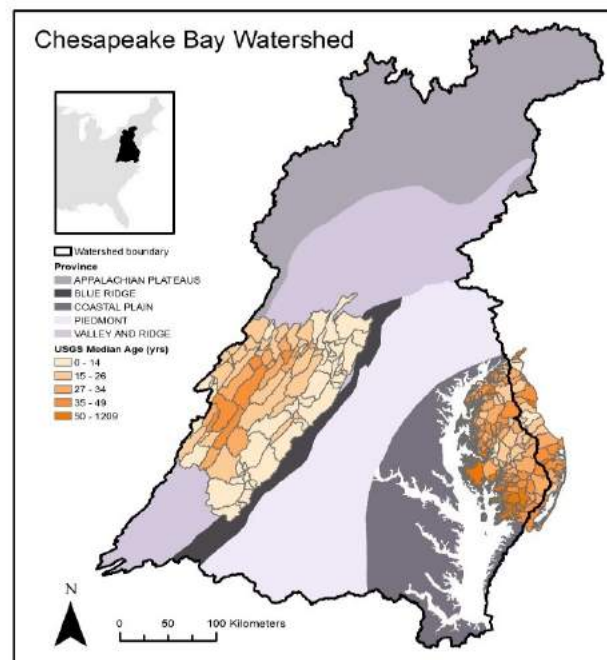
Like watersheds, this dissertation was shaped by its environment. It was written in an office overlooking the waters of the Chesapeake Bay, in the confines of the Chesapeake Bay watershed on the mid-Atlantic coast, USA (see Figure 6.1). Although the ideas explored here have a broad scientific audience, they were inspired by regional environmental imperatives. This final section briefly discusses how the dissertation findings might inform efforts to maintain and improve local water quality, with a focus on nitrate groundwater transport in the Chesapeake Bay watershed, along with potential avenues for future research.

#### **6.1.1.1 Chesapeake Bay watershed overview**

The watershed of the Chesapeake Bay is one of the most economically important and ecologically diverse natural systems in the eastern United States (Committee on the Evaluation of Chesapeake



Bay Program Implementation for Nutrient Reduction to Improve Water Quality, 2011). The vast basin covers 166000 km<sup>2</sup> including all or part of six states (New York, Pennsylvania, Delaware, West Virginia, Maryland, Virginia) and the District of Columbia (see Figure 1). Approximately 20 million people live, work, and farm in the Chesapeake Bay Watershed, which provided an estimated \$3.4 billion per year in ecosystem goods and services to Delaware alone (Kauffman et al., 2011). The watershed's development has contributed to several major regional water quality challenges, including the runoff of nutrients that increase eutrophication of the Chesapeake Bay, the nation's largest estuary (Kemp et al., 2005). In response, the seven local governments along with the U.S. Environmental Protection Agency (EPA) and other federal partners founded the Chesapeake Bay Program (CBP) to restore and conserve the Chesapeake water system. An important part of this effort has been the establishment of county-level Watershed Implementation Plans that reduce pollution diets to levels deemed necessary to meet bay restoration goals by 2025 (Batiuk, Linker, and Cerco, 2013).



**Figure 6.1:** Location of the Chesapeake Bay watershed and median groundwater age estimates. The physiographic provinces of the CBW are shown in shades of gray and the bay in white. The orange shading shows the median groundwater ages for sub-watersheds of the upper Potomac (western side) and eastern shore (eastern side), which were extracted from MODPATH modeling results provided by USGS.

Nitrate pollution due to groundwater leaching from surficial fertilizer application and atmospheric deposition is an important water quality concern. Municipal and industrial wastewater account for only 19% of nitrogen loading to the estuary (Committee on the Evaluation of Chesapeake Bay Program Implementation for Nutrient Reduction to Improve Water Quality, 2011). The remainder is transported as non-point pollution from landscape sources (e.g. fertilizer, atmospheric deposition) to receiving surface. Many streams and portions of the estuary within the Chesapeake Bay water system do not meet federal water quality standards due in part to the high nitrate levels, and an estimated 50% of all nitrate delivered to the bay is leached from groundwater (STAC, 2013).

### 6.1.1.2 Implications for nitrate lag times

In the Chesapeake Bay watershed, nitrate lag times are a poorly understood but significant impediment to bay restoration efforts. As defined in Chapter 5, nitrate lag times are the delay between the implementation of beneficial management practices and the achievement of intended nitrate reductions in surface water quality. One of the important drivers of groundwater lag times in the Chesapeake Bay watershed is groundwater transit times. A USGS report estimated median groundwater transit times of 7-11 years across the physiographic provinces of the watershed (Lindsey et al., 2003), though more localized estimates have ranged from weeks to centuries (see for example the data in Figure 6.1). The Chesapeake Bay Program has identified poor understanding of transit times as an important impediment to their ability to monitor progress and maintain political will for restoration efforts (STAC, 2013). The influence of transit times and climate on nitrate transport is likely, however, to vary across the watershed, which spans five different physiographic provinces.

This dissertation makes advances related to modeling and prediction of groundwater transit times driving lag times in the landscape. Chapter 2 introduces a novel, parsimonious lumped parameter model that simulates and predicts groundwater transit times under different climatic conditions. Chapter 5 shows how the commonly used Soil and Water Assessment Tool (SWAT) can be enhanced to calculate and more flexibly simulate groundwater transit times. While these models require *a priori* information about catchment transit times, Chapter 4 and 5 demonstrate new physically-based modeling approaches that do not, in principle, require prior transit time data. All three chapters provide guidance for modelers in the Chesapeake Bay watershed on when and how to consider groundwater transit times in model formulation. In addition, a sensitivity analysis in

Chapter 4 points to how watershed characteristics (e.g., porosity, lateral hydrologic conductivity) might eventually be used to make first-order predictions about transit times in other parts of the Chesapeake Bay watershed. The sensitivity analysis also showed that groundwater transit times are heavily influenced by the time-variability of rainfall inputs. This suggests that Chesapeake Bay groundwater transit time studies, which typically make steady-state assumptions (e.g., Gburek and Folmar, 1999), should also consider transient conditions and surface groundwater interactions. Finally, these modeling improvements provide opportunities for "what if" scenario development that account for lag times, which can help stakeholders understand the benefits of different management actions (e.g., Sanford and Pope, 2013).

#### **6.1.1.3 Implications for climate change**

The climate of the Chesapeake Bay watershed is likely to become less stationary under global change. General circulation models (GCMs) show with high certainty that increasing CO<sub>2</sub> concentrations in the atmosphere will change weather patterns in the northeast U.S., resulting in higher temperatures, which could increase evapotranspiration and potentially slow hydrologic transport. GCMs also predict (with lower certainty) wetter winters and more extreme precipitation events throughout the year, which could accelerate transport (Walsh et al. 2014). Changes in climate could have a detrimental effect on efforts to restore Chesapeake Bay water quality (Najjar et al., 2010), and multiple technical working groups and research programs have been established to assess the ways in which climate change (e.g., regional warming, sea level rise) could affect the economic and environmental health of the region (e.g., Pyke et al., 2008). Although studies are limited, climate variability across the Chesapeake Bay watershed could have significant effects on transit time distributions and associated transport. For example, heavy rainfall from tropical storms have occasionally made landfall with disproportionately large impacts on water quality (Schubel and Hirschberg, 1978).

This dissertation offers new tools for modeling transit times under a changing climate and benchmarks, for the first time, the potential effect of climate change on catchment transit times. As just described above, the transit time models used and improved here were specifically formulated to handle the kind of changing climate conditions projected in the Chesapeake Bay watershed. Their development was partly motivated by a desire to more accurately capture the effect of climate change on the phenology of nutrient loading to the bay, which can alter the amount and timing of

bay eutrophication (Kemp et al., 2005). There is local demand for this modeling capability. For example, this author worked with collaborators to incorporate many of the modifications made in SWAT in Chapter 5 into the watershed model used investigate the effect of climate change on the achievement of regional water quality goals. In addition, the modeling results from Chapter 2 showed that projected climate change would cause a significant, seasonal shift in the fraction of young water in a small seasonal stream in Wales, UK. The approach from that study could be used in the Chesapeake Bay watershed to help understand the effect of climate change on regional groundwater transit times. This would be a potentially valuable paradigm shift in the practice of climate change impact assessment for water resources, which tends to look at how climate change will affect many individual hydrochemicals (e.g., particulate phosphorus, dissolved nitrate) without understanding the underlying effect on hydrologic transit times.

# Appendix A

## Supplemental Information: Chapter 2

This supporting information provides additional detail on the setup and calibration of the coupled modeling framework used to simulate the influence of rainfall variability on the fraction of young water (FYW) at two neighboring catchments in Plynlimon, Wales (sections S1 and S2). There is also a summary of tests run to check the sensitivity of our results to the choice of the FYW as our summary statistic of the full TTD (section S3). The last section derives the formula used to estimate the timescale of hydrograph recession (section S4).

### S1. Bayesian parameter estimation

This section provides more detail on the Bayesian parameter estimation applied to the rSAS transit time model. The parameter estimation was done with a MATLAB toolbox of the Differential Evolution Adaptive Metropolis (DREAM) algorithm developed by Vrugt et al. (Vrugt, 2016). The DREAM run files associated with this manuscript are available upon request.

In order to perform the required Markov chain Monte Carlo (MCMC) simulation, a likelihood function was selected that quantifies the similarity between model simulations and observations. The likelihood function assumes that the error between simulations and observations (i.e., residuals) are Gaussian, heteroskedastic, and first-order autoregressive. It is described by the equation (Vrugt, 2016):

$$\mathcal{L}(\mathbf{x}|\tilde{\mathbf{Y}}, \phi, \hat{\boldsymbol{\sigma}}^2) = -\frac{n}{2}\log(2\pi) + \frac{1}{2}\log(1-\phi^2) - \frac{1}{2}(1-\phi^2)\hat{\sigma}_1^{-2}e_1(\mathbf{x})^2 - \sum_{t=2}^n \log(\hat{\sigma}_t) - \frac{1}{2} \sum_{t=2}^n \left( \frac{e_t(\mathbf{x}) - \phi e_{t-1}(\mathbf{x})}{\hat{\sigma}_t} \right)^2 \quad (\text{A.1})$$

where  $\mathcal{L}$  is the log likelihood,  $\mathbf{x}$  is a vector of model parameters,  $\tilde{\mathbf{Y}}$  is the vector of observed data (i.e., stream chloride  $C_Q(t)$ ),  $n$  is the number of observed data points,  $e_t(\mathbf{x})$  is the vector of error

residuals,  $\phi$  is the first-order autoregressive coefficient, and  $\hat{\boldsymbol{\sigma}} = \{\hat{\sigma}_1, \dots, \hat{\sigma}_n\}$  is the vector of the standard deviation of error residuals. The error residuals are assumed to take on an AR-1 process:

$$e_t(\mathbf{x}) = \phi * e_{t-1}(\mathbf{x}) + \eta_t \tag{A.2}$$

$$Var[e_t(\mathbf{x})] = \frac{\hat{\sigma}_t^2}{1 - \phi^2} \tag{A.3}$$

with  $\eta_t \stackrel{D}{\sim} \mathcal{N}(0, \hat{\sigma}_t^2)$ . The heteroskedastic standard deviation can be expressed as a linear function of model predictions  $\mathbf{Y}$ :

$$\hat{\boldsymbol{\sigma}} = a_\sigma + b_\sigma * \mathbf{Y} \tag{A.4}$$

with fitting parameters  $a_\sigma$  and  $b_\sigma$ . Thus, in addition to the four parameters of the rSAS model, the DREAM algorithm also estimates the value of the three nuisance parameters  $\phi$ ,  $a_\sigma$ , and  $b_\sigma$ . The distribution of likely values for all seven parameters are shown in Figure A.2 for LH and A.3 for TW. Note that all parameters were assumed to have uninformative priors.

An analysis of the model residuals using the most likely parameter set was performed to help justify the choice of likelihood function. Figures A.8 and A.9 show that the assumptions of heteroskedastic residuals (lower left panel), normally distributed residuals (lower center panel), and first-order auto-correlated residuals (lower right panel) are qualitatively consistent with the actual residuals produced by the model.

## S2. Weather generator calibration

This section describes the calibration of the weather generator in more detail. First, all of the parameters listed in Table 5 of the main paper, except the ones noted below, were calibrated following the methods in Robinson and Sivapalan (1997). Second, the value of four seasonal factors for parameter  $a_1$  (one for each season, final values = 1.02 in winter, 0.94 in summer, 1.05 in spring, 1.12 in summer) and four seasonal factors for  $b_1$  (1.19, 0.49, 0.57, 0.89) were optimized to get a better match between observed and simulated monthly average rainfall. These factors were multiplied by the values of  $a_1$  and  $b_1$  based on the time of year. This was needed because the original weather generator algorithm could not account for the highly seasonal rainfall at Plynlimon based on seasonal variations in storm length alone. Finally, a set of parameters were calibrated that determine the probability and size of small rainfall events (<4mm) on each day between storms based on the 1983-2008 record.

The parameters were: a probability  $p_{drizzle}$  of having a small one-day rainfall event on days between storms (found to be 0.31), a maximum small rainfall event size (assumed to be 4mm), and a random scaling factor between 0 and 1 (found by drawing from a beta distribution with shape factors 1.5 and 0.7) that scaled the maximum small rainfall event size to get the actual small rainfall event size. The final rainfall pattern was scaled to have the same mean rainfall recorded during the 1999-2008 time period.

The performance of all the final parameters was tested following methods in Robinson and Sivapalan (1997), and the results are shown in Figures A.10, A.11, and A.12. The performance was also tested in the coupled model evaluation shown in Figures A.13 and A.14.

### S3. Sensitivity to choice of age fraction

The manuscript focuses on the influence of rainfall variability on the sensitivity of the FYW (i.e., fraction of water less than 90 days in discharge). To see if a subset of our results were sensitivity to the 90 day threshold, we reproduced the calculations used to produce Figure 5 using a threshold of 1, 3, 30, 60, 90, 120, 365, and 1825 days (see Figure A.15-A.22).

### S4. Derivation of the timescale of hydrograph recession

We derive here an approximate solution to the time needed for discharge  $Q(t)$  to recess from the mean discharge  $\bar{Q}(t)$  by a factor of  $e$ . The derivation assumes that the catchment sensitivity function  $g(Q)$  (Kirchner, 2009) is known, and can be defined as:

$$g(Q) = \frac{dQ}{dS} = \frac{dQ}{dt} \frac{dt}{dS} \quad (\text{A.5})$$

which can be rearranged to give:

$$g(Q) \frac{dS}{dt} = \frac{dQ}{dt} \quad (\text{A.6})$$

If we assume  $P$  and  $ET$  are negligible, the water balance is:

$$\frac{dS}{dt} = -Q \quad (\text{A.7})$$

Substituting equation A.7 into equation A.6 and rearranging yields:

$$-\frac{dQ}{Q} = g(\bar{Q})dt \quad (\text{A.8})$$

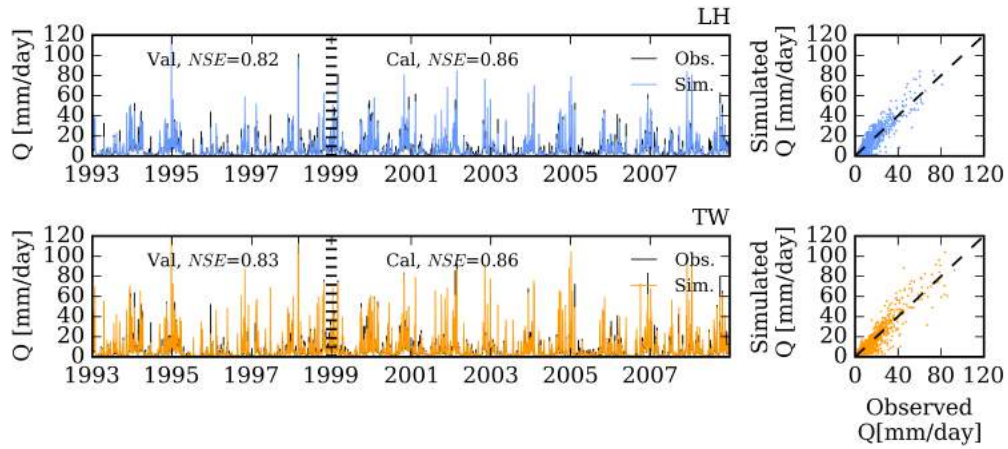
where  $g(Q)$  is evaluated at the mean discharge  $\bar{Q}$ . Integrating both sides of equation A.8 from  $t = 0$  to  $t$  gives:

$$t = -\frac{1}{g(\bar{Q})} \ln\left(\frac{Q(t)}{Q(0)}\right) \quad (\text{A.9})$$

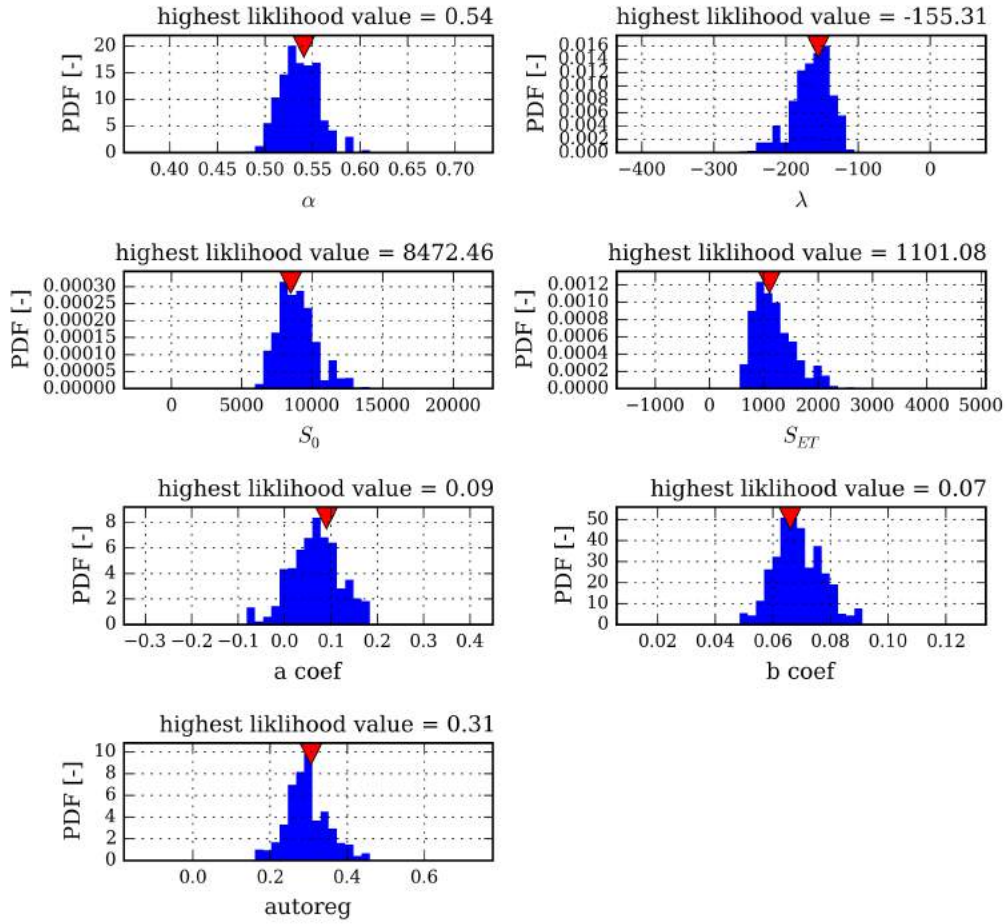
If the initial discharge recesses by a factor of  $e$ , equation A.9 simplifies to:

$$t = \frac{1}{g(\bar{Q})} \quad (\text{A.10})$$





**Figure A.1:** Comparison between the discharge observed and the discharge simulated by the rainfall-runoff model at LH (top panels) and TW (lower panels). The Nash-Sutcliffe efficiency (NSE) for the calibration period (1999-2008) and validation period (1993-1998) are written in the left panels. The right scatter plots show good agreement between observed and simulated data.



**Figure A.2:** PDF of the seven rSAS model parameters at LH, including the three nuisance parameters  $a_\sigma$  ("a coef"),  $b_\sigma$  ("b coef") and  $\phi$  ("autoreg") described in the text of this Supplemental Information. Values were determined from the last 3,200 parameter sets in the MCMC simulation performed by DREAM.

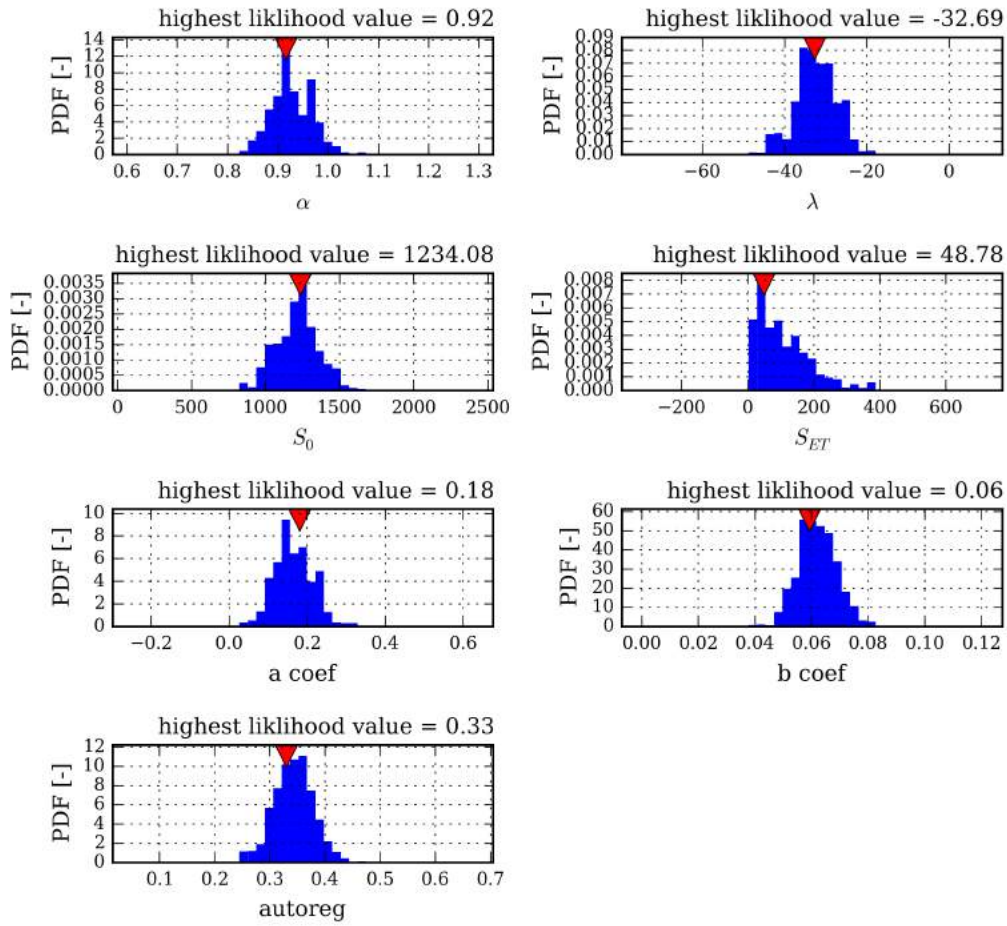
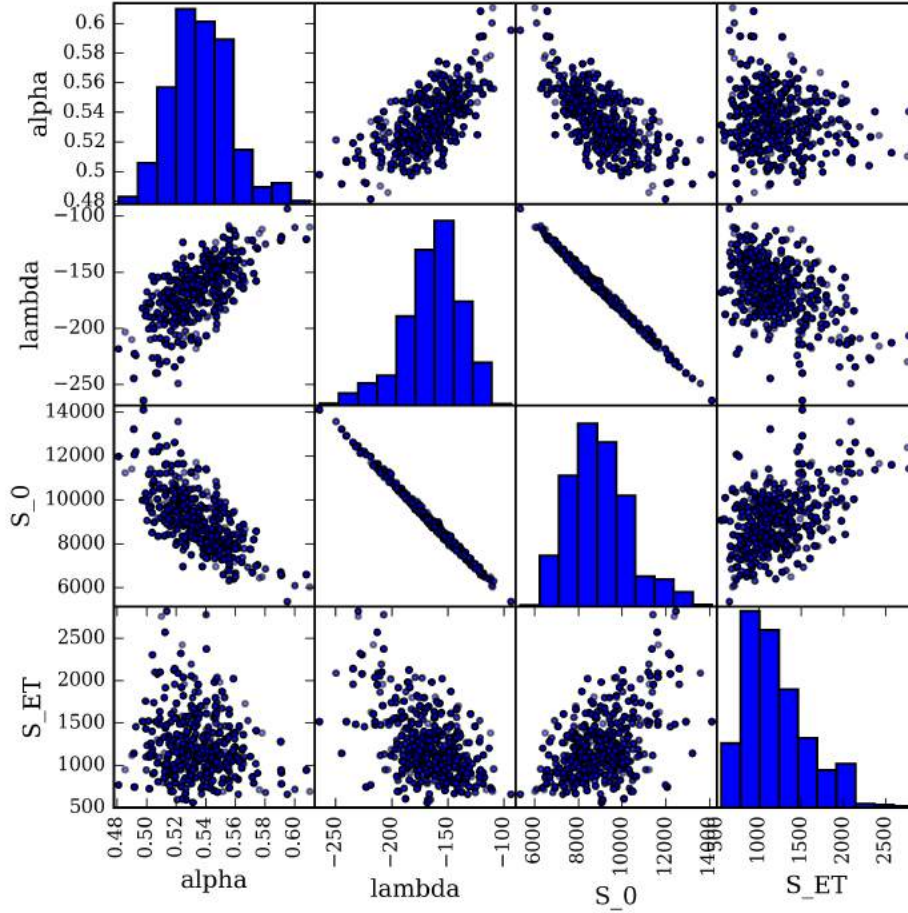


Figure A.3: Same as Figure A.2 but for TW.



**Figure A.4:** Scatter plot matrix for the four rSAS parameters shown in Figure A.2 for the LH watershed. Each dot represents one of the final 3,200 parameter sets in the MCMC simulation run by DREAM. Note, for example, that  $S_0$  and  $\lambda$  have an especially strong negative correlation.

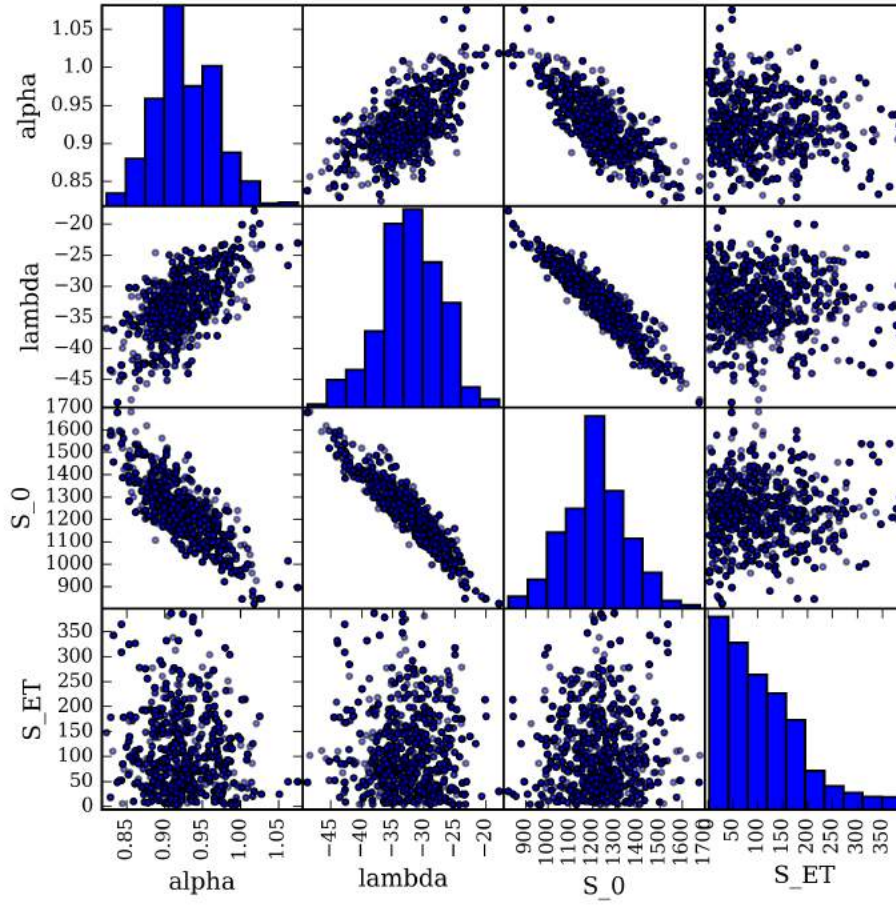
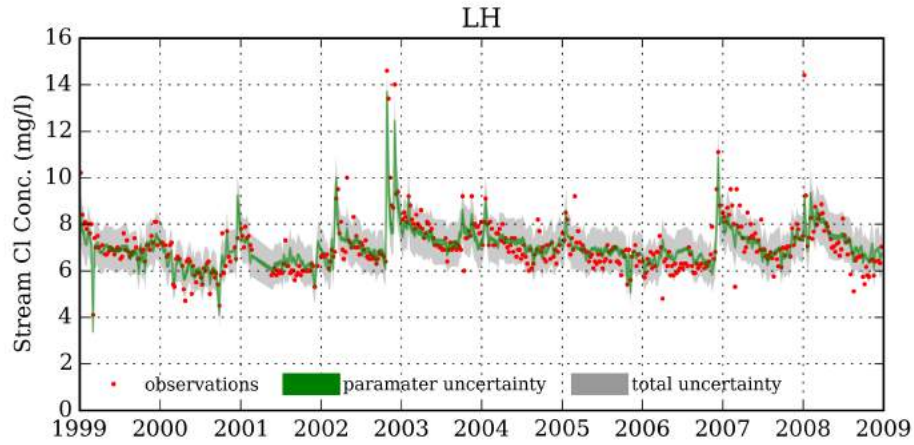
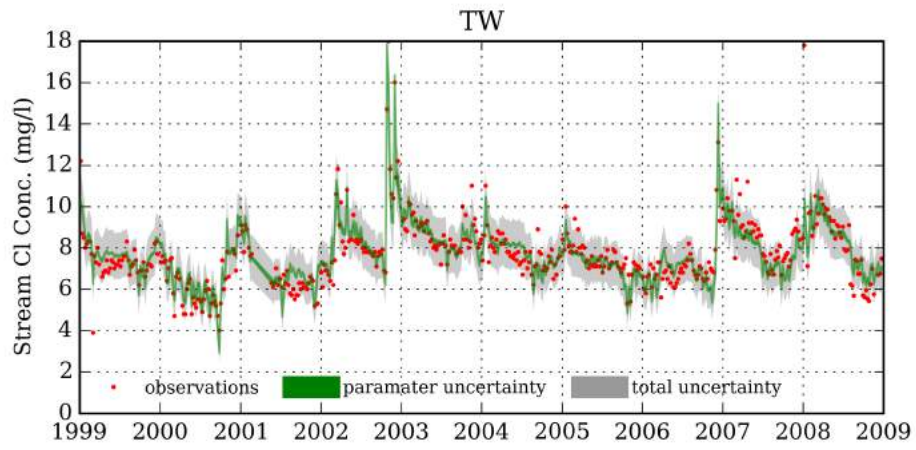


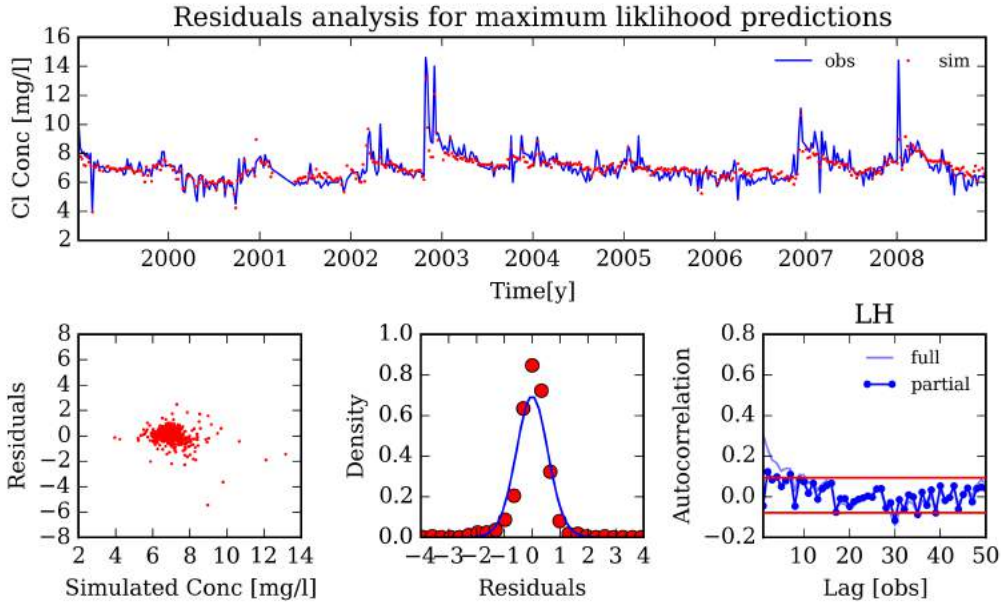
Figure A.5: Same as Figure A.4 but for TW.



**Figure A.6:** Plot of observed stream chloride concentrations  $C_Q$  (red dots) and the range of simulated values using the distribution of likely parameter sets identified by DREAM. At daily timesteps, each of the last 3,200 parameters sets in the MCMC simulation provided an estimate of  $C_Q$ , with the 90th percentile range of these estimates shown in green. The gray region shows the total uncertainty taking into account the 90th percentile of the variance of the residuals from Equation (A.3). Note that most of the observations fall within the total uncertainty range.



**Figure A.7:** Same as Figure A.6 but for TW.



**Figure A.8:** Analysis of residuals of the most likely rSAS parameter set at LH, following the graphical analysis of Vrugt et al. (2009). The upper plot compares observations (blue line) with simulates (red dots) of stream chloride concentration. The lower left plot illustrates the presence of heteroskedasticity in the residuals. As the simulated chloride concentration (x-axis) increases, the spread in residuals also tends to increase (y-axis). The lower center plot shows the PDF of residuals (red dots) and the average of all residual PDFs that would be expected from each simulated data point given the heteroskedastic standard deviation described by Equation (A.3). The observed and expected distribution overlap reasonably well. The lower right panel shows the autocorrelation of the error  $e_t(\mathbf{x})$  from the full residuals (light blue line) and the autocorrelation of the error  $\eta_t$  from the partial residuals (dark blue) after accounting for a first-order autoregressive process. The partial residuals are better constrained within the natural scatter of the data, represented here as a 90th percentile interval (upper and lower red lines).



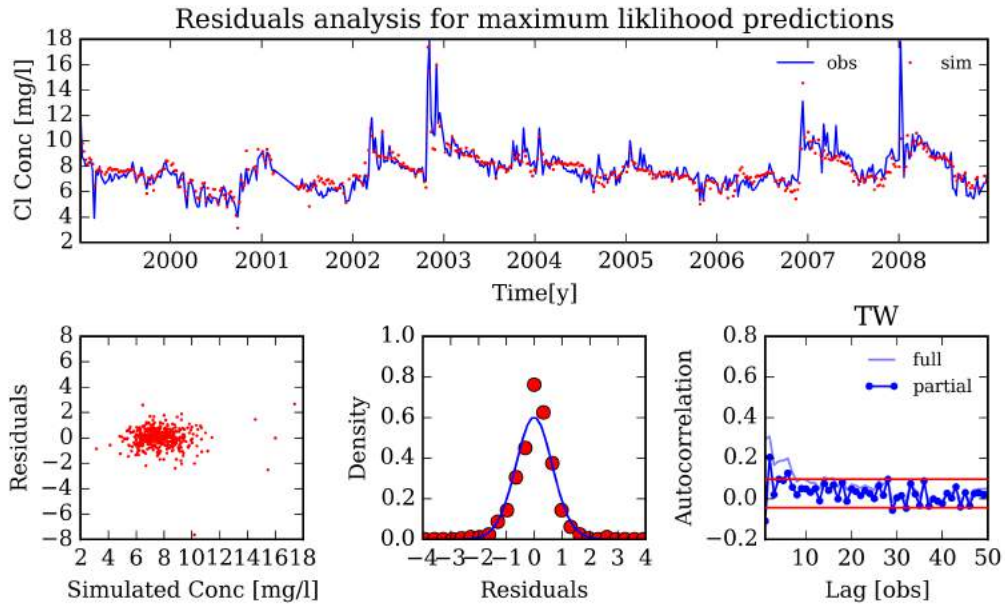
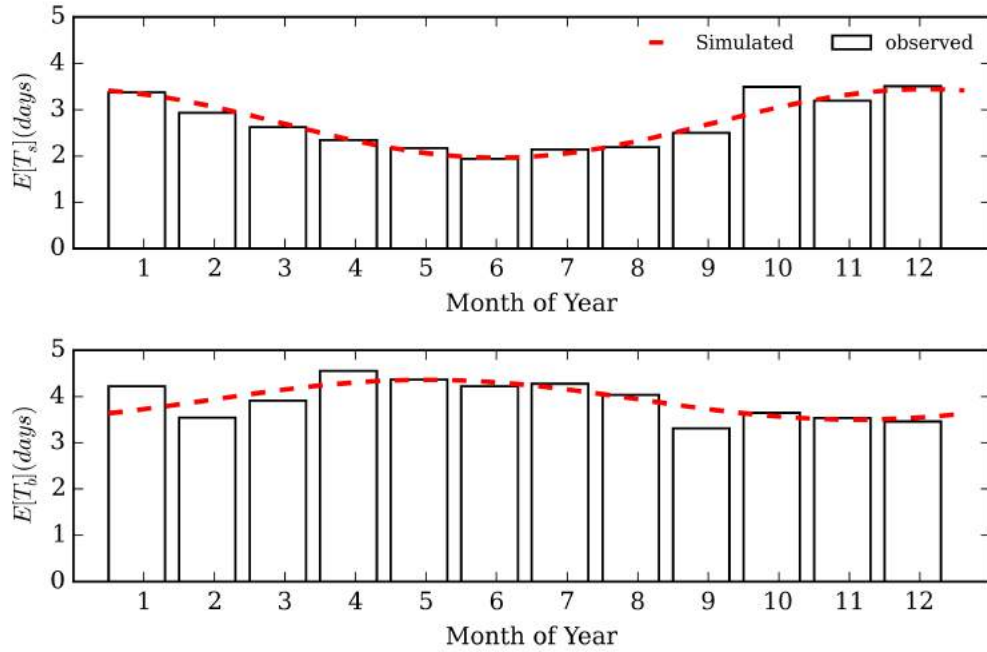
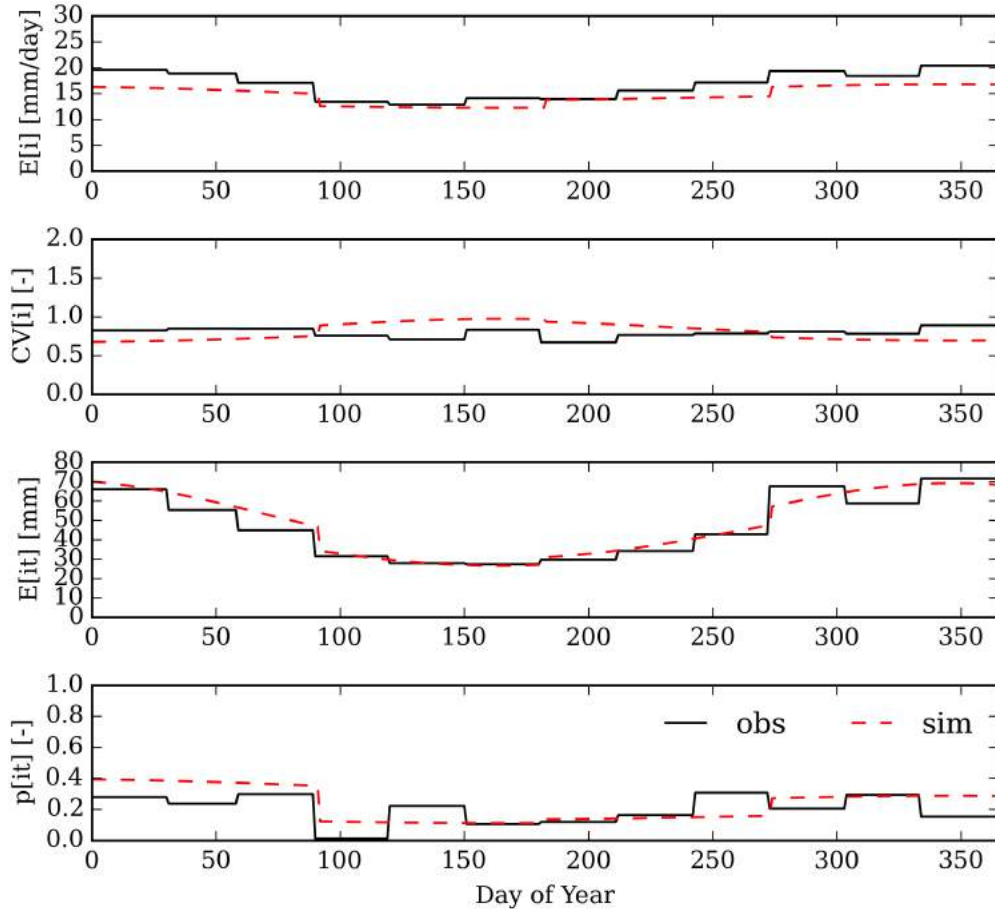


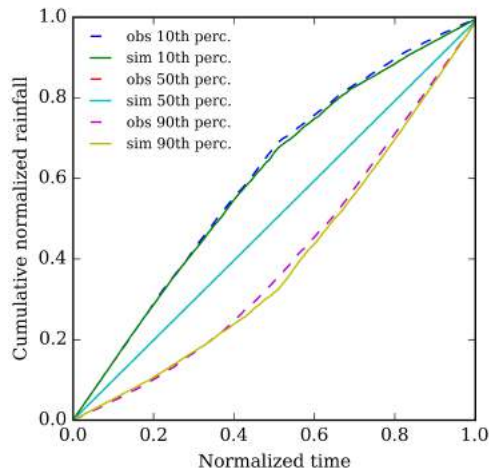
Figure A.9: Same as Figure A.8 but for TW.



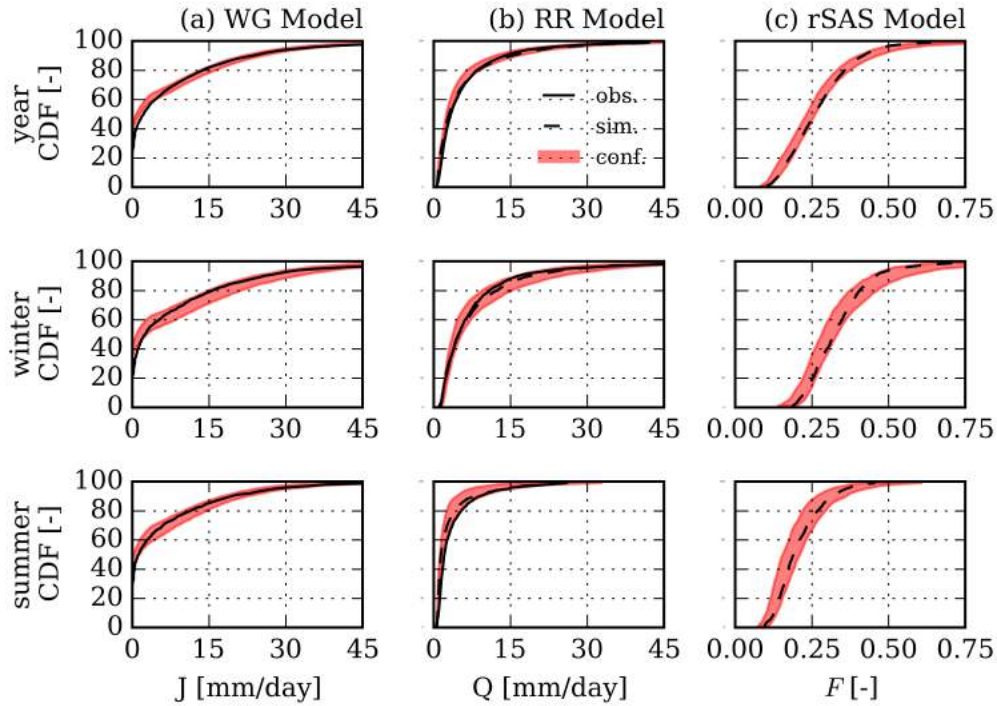
**Figure A.10:** Plot illustrates the performance of the parameters in the rainfall generator model that control the mean, phase, and seasonal amplitude shift for storm duration (upper plot) and break duration (lower plot). The bars in the upper plot show the monthly average storm duration at Plynilimon from 1983-2008. The red lines shows the analytically derived monthly average storm duration after calibration of the mean  $\delta_r$ ,  $\alpha_r$ , and  $\tau_r$ . The bars in the bottom plot show the monthly average interval between storms at Plynilimon from 1983-2008. The red lines shows the analytically derived monthly average storm break after calibration of the mean  $\delta_b$ ,  $\alpha_b$ , and  $\tau_b$ . Plots follow analysis from Robinson and Sivapalan (1997).



**Figure A.11:** Plot illustrates the performance of the parameters in the rainfall generator model that control the average storm intensity as a function of storm length (i.e.,  $a_1$ ,  $b_1$ , and their seasonal factors) and the coefficient of variability as a function of storm length (CV)(i.e.,  $a_2$ ,  $b_2$ ). The top (first) plot shows the expected storm intensity observed from 1983-2008 (black line) and analytically derived for the rainfall generator (red dashed line). The lower plots make the same comparison for the CV of storm intensity (second plot), the expected value of the product of storm intensity and storm duration (third plot), and the correlation coefficient for the intensity and duration (fourth plot). Plots and analytic calculations follow Robinson and Sivapalan (1997).



**Figure A.12:** Plot illustrates the performance of the parameters in the rainfall generator that control within-storm variability of rainfall (i.e., parameters  $\beta_1$  and  $\beta_2$ , which are set equal). The 10th, 50th, and 90th percentile mass curves for all storms from 1983-2008 are given by the dashed lines. The mass curves from a representative simulation are shown by the solid lines. The overlap suggests the weather generator is able to simulate the observed within-storm variability. Plots follow approach of Robinson and Sivapalan (1997).



**Figure A.13:** Results from the coupled-model evaluation at LH. The solid black lines show the CDF of observed rainfall (column a) and observed discharge (column b). The dashed black lines show the CDF of simulated discharge based on observed rainfall (column b) and simulated fraction of young water ( $F$ ) based on observed rainfall and discharge with the most likely rSAS parameter set (column c). The red area is the 95th percentile of 100 CDFs generated from 100 16-year simulations of rainfall (column a), from 100 16-year simulations of discharge based on the simulated rainfall (column b), and from 100 16-year simulations of the fraction of young water based on the simulated rainfall and discharge (column c). The black curves fall within the red regions, indicating the observed and simulated values have similar probability distributions for the all time (upper row), winter only (middle row), and summer only (lower row) cases.

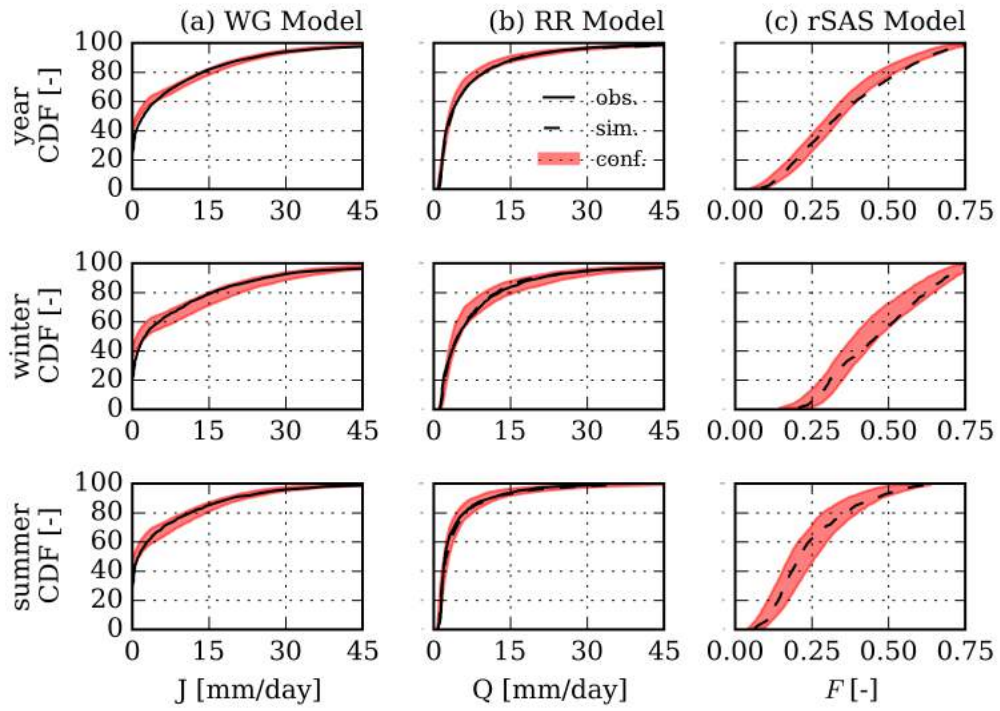
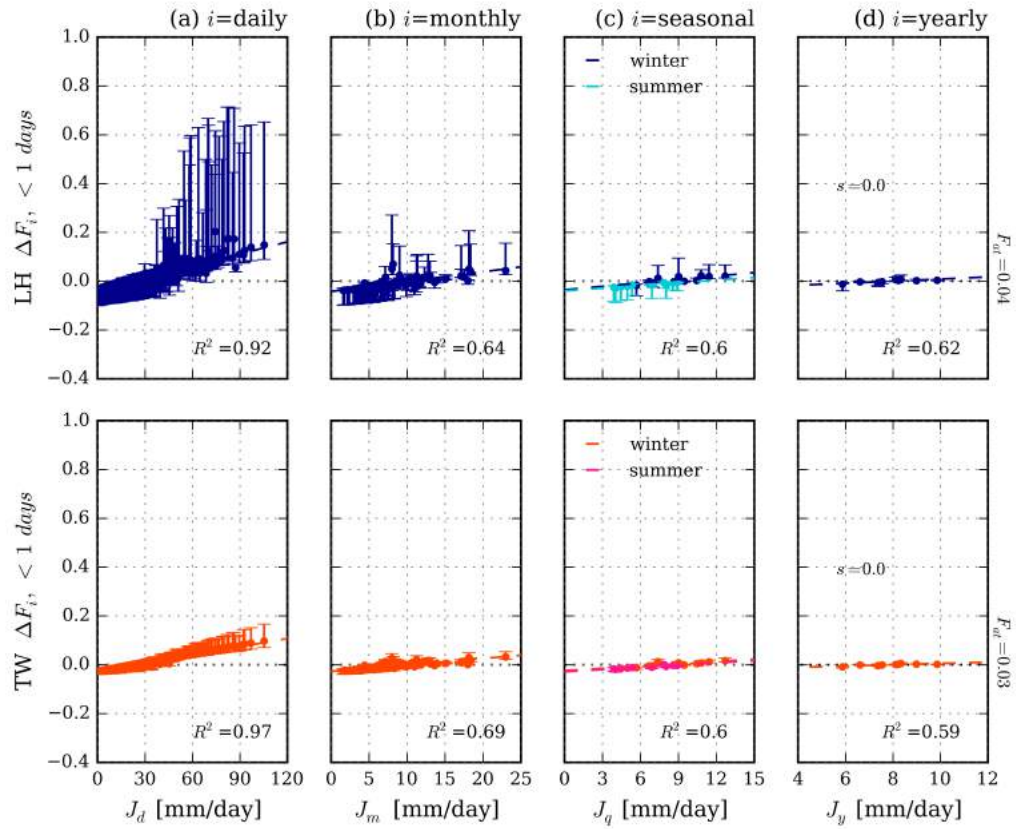


Figure A.14: Same as Figure S13 but for TW.



**Figure A.15:** Same as Figure 5 in the manuscript, except the age fraction anomaly for water less than 1 day old is shown on the y axes instead of the age fraction anomaly of water less than 90 days old. The x axes show the mean rainfall averaged over four timescales  $i$ : daily (panel a), monthly (panel b), seasonally (panel c), and yearly (panel d). The dashed line shows a line of best fit. The error bars show the 90th percentile confidence interval. The right-most panels indicate the slope  $s$  of the line of best fit and the median  $F_{at}$  for all parameter sets.

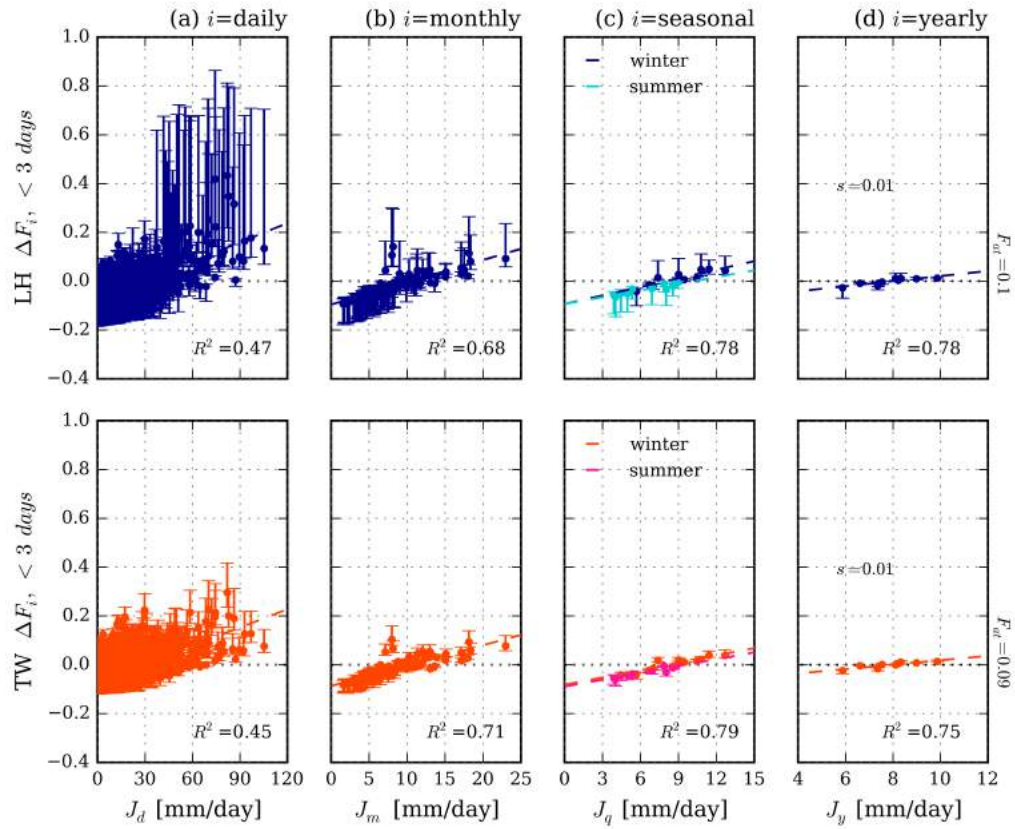


Figure A.16: Same as Figure A.15 with the age fraction threshold for discharge set to water less than 3 days.



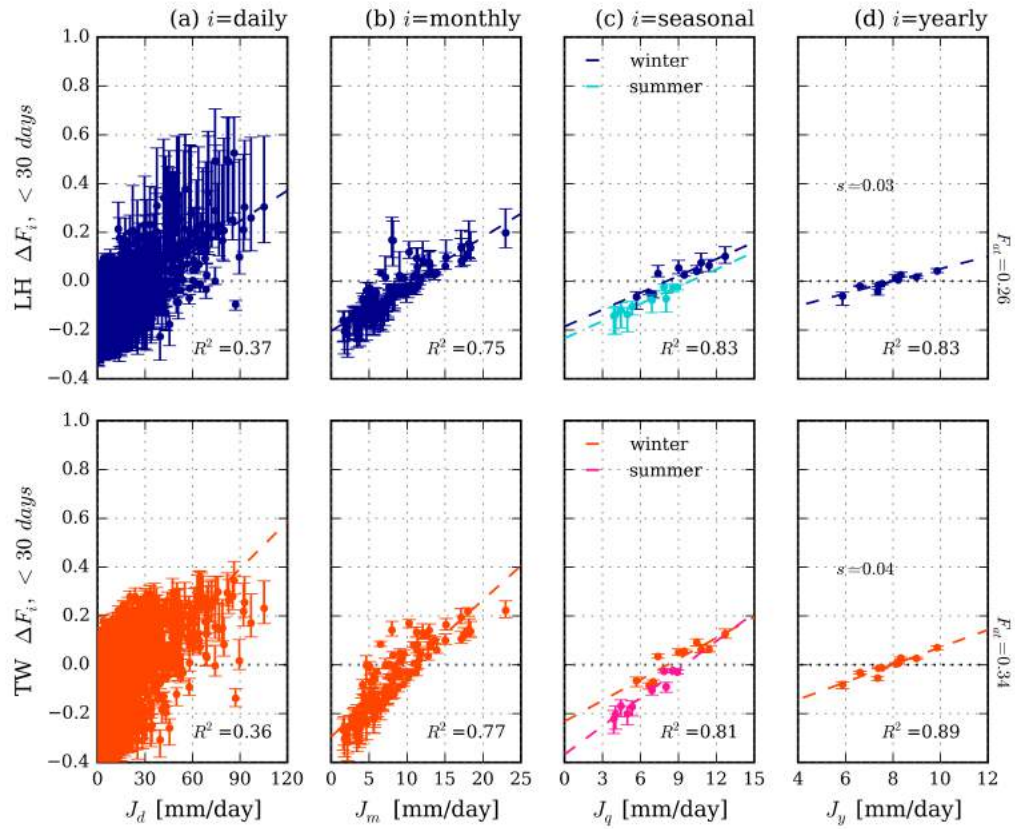


Figure A.17: Same as Figure A.15 with the age fraction threshold for discharge set to water less than 30 days.

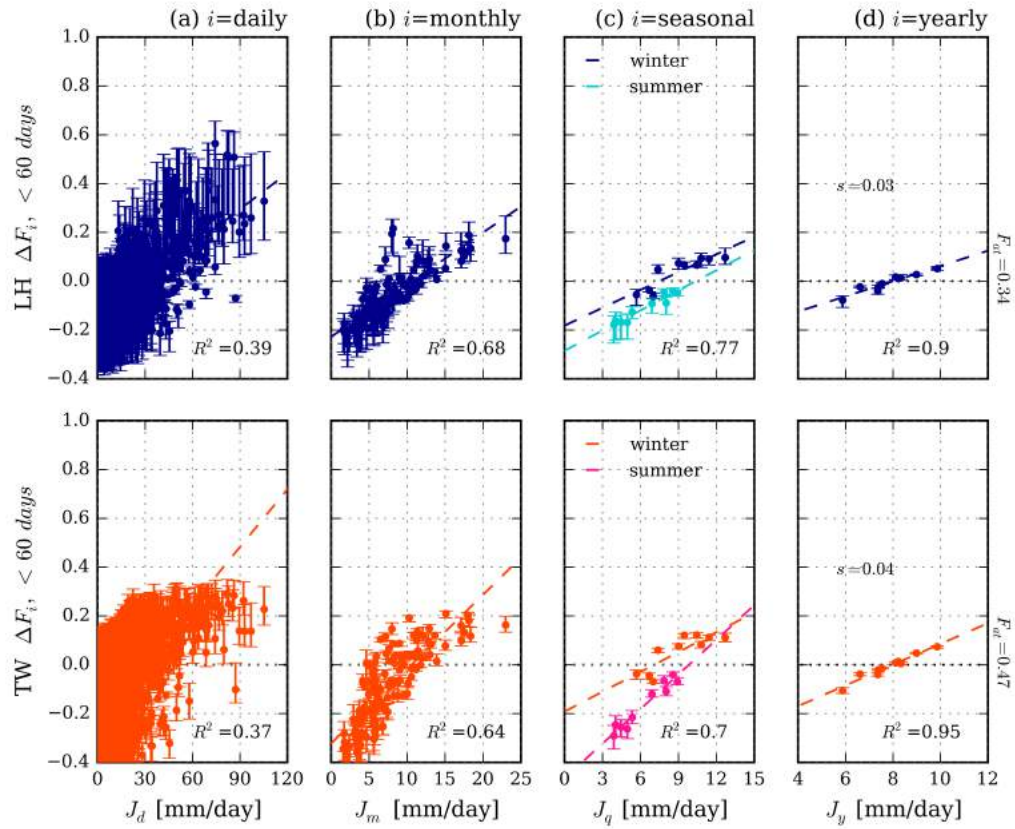
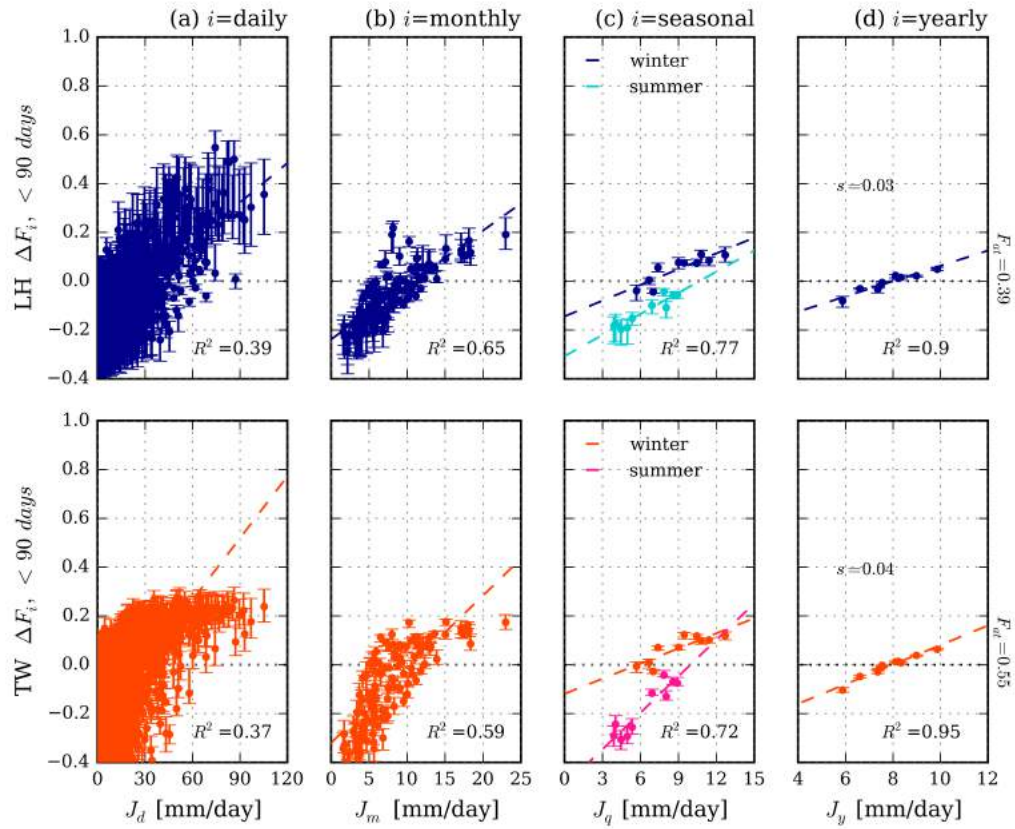


Figure A.18: Same as Figure A.15 with the age fraction threshold for discharge set to water less than 60 days.



**Figure A.19:** Same as Figure A.15 with the age fraction threshold for discharge set to water less than 90 days. Note these results are identical to the results shown in Figure 5 of the manuscript.

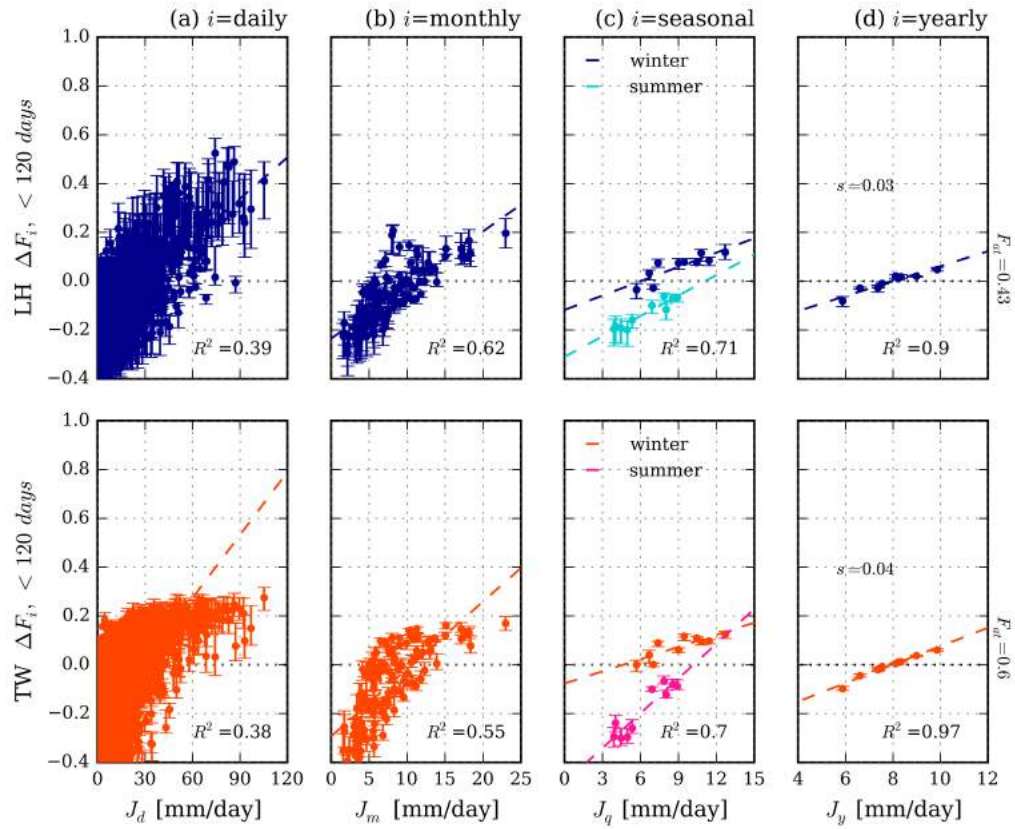


Figure A.20: Same as Figure A.15 with the age fraction threshold for discharge set to water less than 120 days.

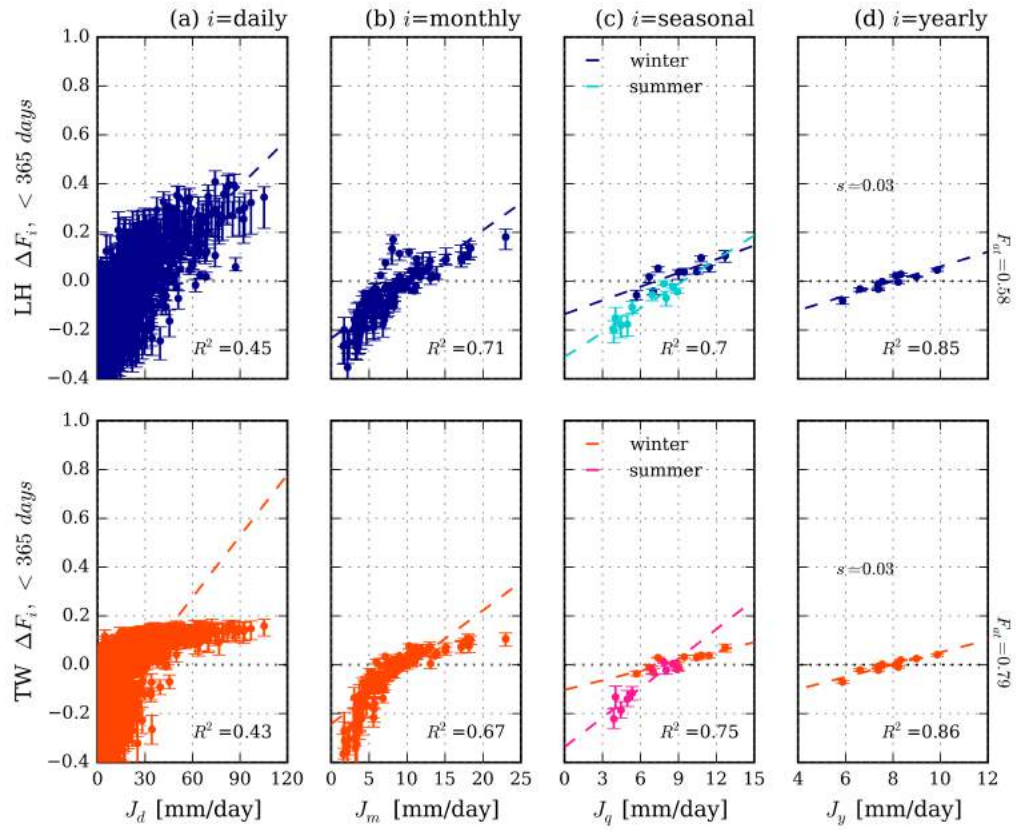


Figure A.21: Same as Figure A.15 with the age fraction threshold for discharge set to water less than 365 days.

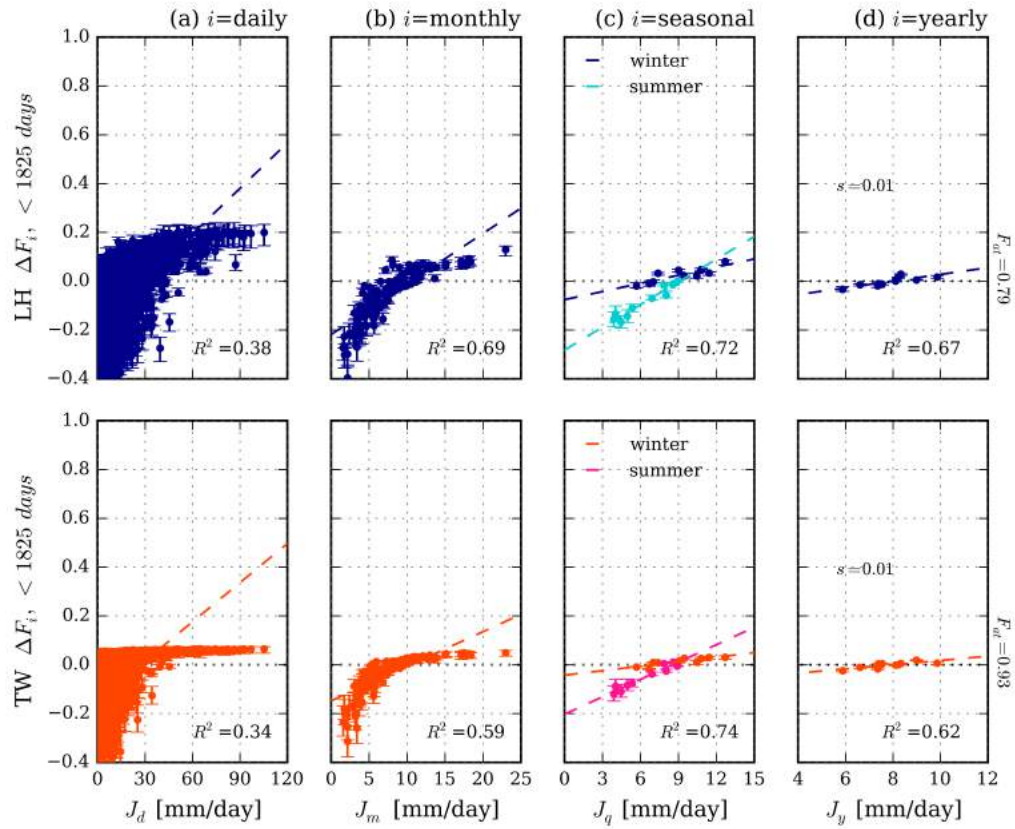


Figure A.22: Same as Figure A.15 with the age fraction threshold for discharge set to water less than 1825 days.

## Bibliography

- Abramowitz, M. and I. A. Stegun (1964). *Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables*. Vol. 55. DoverPublications. com.
- Ali, M., A. Fiori, and D. Russo (2014). “A comparison of travel-time based catchment transport models, with application to numerical experiments”. In: *Journal of Hydrology* 511, pp. 605–618.
- Allen, R. G. et al. (1998). *Crop evapotranspiration - Guidelines for computing crop water requirements - FAO Irrigation and drainage paper 56*. Tech. rep. Rome: Food and Agriculture Organization of the United Nations, pp. 1–15.
- Ashby, S. F. and R. D. Falgout (1996). “A parallel multigrid preconditioned conjugate gradient algorithm for groundwater flow simulations”. In: *Nuclear Science and Engineering* 124.1, pp. 145–159.
- Barnes, C. J. and A. M. Bonell (1996). “Application of Unit Hydrograph Techniques To Solute Transport in Catchments”. In: *Hydrological Processes* 10, pp. 793–802.
- Batiuk, R. A., L. C. Linker, and C. F. Cerco (2013). “Featured Collection Introduction: Chesapeake Bay Total Maximum Daily Load Development and Application”. In: *JAWRA Journal of the American Water Resources Association* 39180, pp. 1–5.
- Benettin, P. et al. (2013). “Chloride circulation in a lowland catchment and the formulation of transport by travel time distributions”. In: *Water Resources Research* 49.8, pp. 4619–4632.
- Benettin, P. et al. (2015). “Modeling chloride transport using travel time distributions at Plynlimon, Wales”. In: *Water Resources Research* 51, pp. 3259–3276.
- Benettin, P. et al. (2017). “Using SAS functions and high-resolution isotope data to unravel travel time distributions in headwater catchments”. In: *Water Resource Research* 53, pp. 5375–5377.
- Beven, K. (1982). “On subsurface stormflow: an analysis of response times”. In: *Hydrological Sciences Journal* 27.4, pp. 505–521.
- (2006a). “A manifesto for the equifinality thesis”. In: *Journal of Hydrology* 320.1, pp. 18–36.

- Beven, K (2006b). “Searching for the Holy Grail of scientific hydrology:  $Q=H(S, R, t)A$  as closure”. In: *Hydrology and Earth System Sciences* 10.5, pp. 609–618.
- Beven, K. (2011). *Rainfall-runoff modelling: the primer*. John Wiley and Sons, Ltd, p. 452.
- (2016). “Facets of uncertainty : epistemic uncertainty , non- stationarity , likelihood , hypothesis testing , and communication”. In: *Hydrological Sciences Journal* 61.9, pp. 1652–1665.
- Beven, K. and A. Binley (1992). “The future of distributed models: model calibration and uncertainty prediction.” In: *Hydrological Processes* 6.May 1991, pp. 279–298.
- (2014). “GLUE: 20 years on”. In: *Hydrological Processes* 28.24, pp. 5897–5918.
- Birkel, C. and C. Soulsby (2015). “Advancing tracer-aided rainfall-runoff modelling: A review of progress, problems and unrealised potential”. In: *Hydrological Processes* 29.25, pp. 5227–5240.
- (2016). “Linking tracers, water age and conceptual models to identify dominant runoff processes in a sparsely monitored humid tropical catchment”. In: *Hydrological Processes* 30.24, pp. 4477–4493.
- Birkel, C., C. Soulsby, and D. Tetzlaff (2011). “Modelling catchment-scale water storage dynamics: Reconciling dynamic storage with tracer-inferred passive storage”. In: *Hydrological Processes* 25.25, pp. 3924–3936.
- Birkel, C. et al. (2012). “High-frequency storm event isotope sampling reveals time-variant transit time distributions and influence of diurnal cycles”. In: *Hydrological Processes* 26.July 2011, pp. 308–316.
- Bisht, G. et al. (2017). “Coupling a three-dimensional subsurface flow and transport model with a land surface model to simulate stream-aquifer-land interactions (PFLOTRAN\_CLM v1.0)”. In: *Geoscientific Model Development Discussions* February, pp. 1–45.
- Blum, A. G., R. M. Vogel, and S. A. Archfield (2016). “The probability distribution of daily streamflow in the conterminous United States”. In: *Hydrology and Earth System Sciences Discussions*, pp. 1–33.
- Botter, G., E. Bertuzzo, and A. Rinaldo (2010). “Transport in the hydrologic response: Travel time distributions, soil moisture dynamics, and the old water paradox”. In: *Water Resources Research* 46.3.
- (2011). “Catchment residence and travel time distributions: The master equation”. In: *Geophysical Research Letters* 38.11.



- Breuer, L. et al. (2008). “Current concepts in nitrogen dynamics for mesoscale catchments”. In: *Hydrological Sciences Journal* 53.5, pp. 1059–1074.
- Brunner, P. and C. T. Simmons (2012). “A Fully Integrated Physically Based Hydrogeological Model”. In: *Ground Water* 50.2, pp. 170–176.
- Bryant, R. B. et al. (2011). “U . S . Department of Agriculture Agricultural Research Service Mahantango Creek Watershed , Pennsylvania , United States : Physiography and history”. In: 47, pp. 1–5.
- Buda, A et al. (2009). “Factors influencing surface runoff generation from two agricultural hillslopes in central Pennsylvania”. In: *Hydrologic Processes* 23.November 2008, pp. 1295–1312.
- Buda, A. R. et al. (2011). “U.S. Department of Agriculture Agricultural Research Service Mahantango Creek Watershed , Pennsylvania , United States : Long-term precipitation database”. In: 47, pp. 1–5.
- Burns, D. a. et al. (2001). “Quantifying contributions to storm runoff through end-member mixing analysis and hydrologic measurements at the Panola Mountain research watershed (Georgia, USA)”. In: *Hydrological Processes* 15.10, pp. 1903–1924.
- Burton, W. C. et al. (2002). “Influence of fracture anisotropy on ground water ages and chemistry, Valley and Ridge province, Pennsylvania”. In: *Ground Water* 40.3, pp. 242–257.
- Buttle, J. M. (1994). “Isotope hydrograph separations and rapid delivery of pre-event water from drainage basins”. In: *Progress in Physical Geography* 18.1, pp. 16–41.
- Buttle, J. M., A. Vonk, and C. H. Taylor (1995). “Applicability of Isotopic Hydrograph Separation in a Suburban Basin During Snowmelt”. In: *Hydrological Processes* 9, pp. 197–211.
- Cameron, D. et al. (2000). “Flood frequency estimation by continuous simulation (with likelihood based uncertainty estimation)”. In: *Hydrology and Earth System Sciences* 4.1, pp. 23–34.
- Cameron, D. S. et al. (1999). “Flood frequency estimation by continuous simulation for a gauged upland catchment (with uncertainty)”. In: *Journal of Hydrology* 219, pp. 169–187.
- Camporese, M. et al. (2010). “Surface-subsurface flow modeling with path-based runoff routing, boundary condition-based coupling, and assimilation of multisource observation data”. In: *Water Resources Research* 46.2.

- Camporese, M. et al. (2009). “Ensemble Kalman filter data assimilation for a process-based catchment scale model of surface and subsurface flow”. In: *Water Resources Research* 45.10, pp. 1–14.
- Camporese, M. et al. (2014). “Simplified modeling of catchment-scale evapotranspiration via boundary condition switching”. In: *Advances in Water Resources* 69, pp. 95–105.
- Carpenter, S. et al. (1998). “Nonpoint pollution of surface waters with phosphorus and nitrogen”. In: *Ecological Applications* 8.3, pp. 559–568.
- Clark, M. P. et al. (2017). “The evolution of process-based hydrologic models: Historical challenges and the collective quest for physical realism”. In: *Hydrology and Earth System Sciences* 21.7, pp. 3427–3440.
- Committee on the Evaluation of Chesapeake Bay Program Implementation for Nutrient Reduction to Improve Water Quality (2011). *Achieving nutrient and sediment reduction goals in the Chesapeake Bay: An evaluation of program strategies and implementation*. Tech. rep. National Research Council.
- Cook, P. and A. Herczeg (2012). *Environmental tracers in subsurface hydrology*. Springer Science & Business Media.
- Cornelissen, T., B. Diekkrüger, and H. Bogen (2016). “Using High-Resolution Data to Test Parameter Sensitivity of the Distributed Hydrological Model HydroGeoSphere”. In: *Water* 8.5, p. 202.
- Dai, Y. (2003). “The common land model”. In: *American Meteorological Society* August, pp. 1013–1023.
- Danesh-Yazdi, M et al. (2016). “Inferring changes in water cycle dynamics of intensively managed landscapes via the theory of time-variant travel time distributions”. In: *Water Resource Research*, pp. 613–615.
- Danesh-Yazdi, M. et al. (2018). “Bridging the gap between numerical solutions of travel time distributions and analytical storage selection functions”. In: *Hydrological Processes* 32.8, pp. 1063–1076.
- Davies, J. et al. (2011). “A discrete particle representation of hillslope hydrology : hypothesis testing in reproducing a tracer experiment at Gardsjon, Sweden”. In: *Hydrological Processes* 25, pp. 3602–3612.

- Delfs, J. O. et al. (2013). “A coupled surface/subsurface flow model accounting for air entrapment and air pressure counterflow”. In: *Environmental Earth Sciences* 69.2, pp. 395–414.
- Denver, J. M. (1995). “Combined use of groundwater dating , chemical , and isotopic analyses to resolve the history and fate of nitrate contamination in two agricultural watersheds , Atlantic coastal plain , Maryland of recharging discharging”. In: 31.9, pp. 2319–2339.
- Dunn, S. M., J. J. McDonnell, and K. B. Vaché (2007). “Factors influencing the residence time of catchment waters: A virtual experiment approach”. In: *Water Resources Research* 43.6, pp. 1–14.
- Dunne, T. and R. D. Black (1970). “Partial area contributions to storm runoff in a small New England watershed”. In: *Water Resources Research* 6.5, pp. 1296–1311.
- Easton, Z. M. et al. (2008). “Re-conceptualizing the soil and water assessment tool (SWAT) model to predict runoff from variable source areas”. In: *Remote Sensing of Environment* 348.3-4, pp. 279–291.
- Ebel, B. A. et al. (2008). “Physics-based continuous simulation of long-term near-surface hydrologic response for the Coos Bay experimental catchment”. In: *Water Resources Research* 44.7, pp. 1–23.
- Elith, J., J. R. Leathwick, and T. Hastie (2008). “A working guide to boosted regression trees”. In: *Journal of Animal Ecology* 77.4, pp. 802–813.
- Engdahl, N. B. and R. M. Maxwell (2015). “Quantifying changes in age distributions and the hydrologic balance of a high-mountain watershed from climate induced variations in recharge”. In: *Journal of Hydrology* 522, pp. 152–162.
- EPA (1996). *Nonpoint Source Pollution: The Nation’s Largest Water Quality Problem*. Tech. rep.
- (2012). *The Facts about Nutrient Pollution*. Tech. rep.
- Fang, Z. et al. (2015). “Spatio-temporal validation of long-term 3D hydrological simulations of a forested catchment using empirical orthogonal functions and wavelet coherence analysis”. In: *Journal of Hydrology* 529, pp. 1754–1767.
- Fatichi, S. et al. (2016). “An overview of current applications, challenges, and future trends in distributed process-based models in hydrology”. In: *Journal of Hydrology* 537, pp. 45–60.
- Friedman, J. (2001). “Greedy function approximation: a gradient boosting machine”. In: *The Annals of Statistics* 29.5, pp. 1189–1232.

- Fuka, D. R. et al. (2016). "Improving the spatial representation of soil properties and hydrology using topographically derived initialization processes in the SWAT model". In: *Hydrological Processes* 30.24, pp. 4633–4643.
- Galloway, J. N. and E. B. Cowling (2002). "Reactive nitrogen and the world: 200 years of change." In: *Ambio* 31.2, pp. 64–71.
- Gauthier, M. J. et al. (2009). "A modeling study of heterogeneity and surface water-groundwater interactions in the Thomas Brook catchment, Annapolis Valley (Nova Scotia, Canada)". In: *Hydrology and Earth System Sciences* 13.9, pp. 1583–1596.
- Gburek, W. J. and G. J. Folmar (1999). "Flow and chemical contributions to streamflow in an upland watershed : a baseflow survey". In: *Journal of Hydrology* 217, pp. 1–18.
- Gburek, W. J. and J. Urban (1990). "The Shallow Weathered Fracture Layer in the Near Stream Zone". In: *Ground Water* 28.6.
- Gburek, W. J., G. J. Folmar, and J. Urban (1999). "Field Data and Ground Water Modeling in a Layered Fractured Aquifer". In: *Groundwater* 37.2, pp. 175–184.
- Gburek, W. (1977). "Predictability of effects of a severe local storm in Pennsylvania". In: *Journal of Applied Meteorology* 16.2, pp. 136–144.
- Genuchten, M. T. van (1980). "A Closed-form Equation for Predicting the Hydraulic Conductivity of Unsaturated Soils". In: *Soil Science Society of America Journal* 44.5, p. 892.
- Georgek, J. L. et al. (2017). "Using tracer-derived groundwater transit times to assess storage within a high-elevation watershed of the upper Colorado River Basin, USA". In: *Hydrogeology Journal* 26, pp. 467–480.
- Gilbert, J. M. et al. (2016). "Global spatial sensitivity of runoff to subsurface permeability using the active subspace method". In: *Advances in Water Resources* 92, pp. 30–42.
- Gruber, N. and J. N. Galloway (2008). "An Earth-system perspective of the global nitrogen cycle." In: *Nature* 451.7176, pp. 293–6.
- Gupta, H. V. et al. (2009). "Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling". In: *Journal of Hydrology* 377.1-2, pp. 80–91.
- Haahti, K. et al. (2015). "Distributed hydrological modeling with channel network flow of a forestry drained peatland site". In: *Water Resources Research* 51, pp. 5974–5997.

- Haitjema, H. and S Mitchell-Bruker (2005). “Are water tables a subdued replica of the topography?” In: *Groundwater* 43.6.
- Harman, C. J., A. S. Ward, and A. Ball (2016). “How does reach-scale stream-hyporheic transport vary with discharge? Insights from rSAS analysis of sequential tracer injections in a headwater mountain stream”. In: *Water Resources Research* 52.9, pp. 1–20.
- Harman, C. J. (2015). “Time Variable Transit Time Distributions and Transport: Theory and Application to Storage-dependent Transport of Chloride in a Watershed”. In: *Water Resources Research* 50.1, pp. 1–30.
- Harman, C. and M Kim (2014). “An efficient tracer test for time-variable transit time distributions in periodic hydrodynamic systems”. In: *Geophysical Research Letters* 41.5, pp. 1567–1575.
- Hay, L., S. Markstrom, and C Ward-Garrison (2011). “Watershed-scale response to climate change through the twenty-first century for selected basins across the united states”. In: *Earth Interactions* 15.17, pp. 1–37.
- Heidbüchel, I., P. a. Troch, and S. W. Lyon (2013). “Separating physical and meteorological controls of variable transit times in zero-order catchments”. In: *Water Resources Research* 49.11, pp. 7644–7657.
- Heidbüchel, I. et al. (2012). “The master transit time distribution of variable flow systems”. In: *Water Resources Research* 48.6.
- Hepner, C. S., K. Loague, and J. E. VanderKwaak (2007). “Long-term InHM simulations of hydrologic response and sediment transport for the R-5 catchment”. In: *Earth Surface Processes and Landforms* 34, pp. 155–161.
- Homer, C. G. et al. (2015). “Completion of the 2011 National Land Cover Database for the conterminous United States-Representing a decade of land cover change information”. In: *Photogrammetric Engineering and Remote Sensing* 81.5, pp. 345–354.
- Hooper, R. P., N. Christophersen, and N. E. Peters (1990). “Modelling Streamwater Chemistry as a Mixture of Soilwater End-members - An Application to the Panola Mountain Catchment, Georgia, U.S.A.” In: *Journal of Hydrology* 116, pp. 321–343.
- Horton, R. (1933). “The role of infiltration in the hydrologic cycle”. In: *Eos, Transactions American Geophysical Union* 14.1, pp. 446–460.

- Hrachowitz, M and H. Savenije (2013). “A decade of Predictions in Ungauged Basins (PUB)—a review”. In: *Hydrological Sciences Journal* 58.6, pp. 1198–1255.
- Hrachowitz, M et al. (2009). “Using long-term data sets to understand transit times in contrasting headwater catchments”. In: *Journal of Hydrology* 367.3-4, pp. 237–248.
- Hrachowitz, M et al. (2010). “Gamma distribution models for transit time estimation in catchments: Physical interpretation of parameters and implications for time-variant transit time assessment”. In: *Water Resources Research* 46.10.
- Hrachowitz, M et al. (2013). “What can flux tracking teach us about water age distribution patterns and their temporal dynamics?” In: *Hydrology and Earth System Sciences* 17, pp. 533–564.
- Hrachowitz, M. et al. (2015). “Transit time distributions , legacy contamination and variability in biogeochemical 1/f scaling : How are hydrological response dynamics linked to water quality at the catchment scale?” In: *Hydrological Processes* 29.25, pp. 5241–5256.
- Hrachowitz, M. and M. P. Clark (2017). “HESS Opinions: The complementary merits of competing modelling philosophies in hydrology”. In: *Hydrology and Earth System Sciences* 21.8, pp. 3953–3973.
- Hrachowitz, M. et al. (2016). “Transit times - the link between hydrology and Transit times - the link between hydrology and water quality at the catchment scale”. In: *WIREs Water* 3.5, pp. 629–657.
- Hyer, K. E. et al. (2016). *Spatial and Temporal Variation of Stream Chemistry Associated With Contrasting Geology and Land-Use Patterns in the Chesapeake Bay Watershed: Summary of Results from Smith Creek, Virginia; Upper Chester River, Maryland; Conewago Creek, Pennsylvania; and D.* USGS, p. 211.
- Jasechko, S. et al. (2016). “Substantial proportion of global streamflow less than three months old”. In: *Nature Geoscience* 9.2, pp. 126–129.
- Jones, E, T Oliphant, and P Peterson (2014). *SciPy: open source scientific tools for Python*.
- Jones, J. and C. Woodward (2001). “Newton–Krylov-multigrid solvers for large-scale, highly heterogeneous, variably saturated flow problems”. In: *Advances in Water Resources* 24.7, pp. 763–774.

- Jothityangkoon, C, M Sivapalan, and D. L. Farmer (2001). “Process controls of water balance variability in a large semi-arid catchment : downward approach to hydrological model development”. In: *Journal of Hydrology* 254.1-4, pp. 174–198.
- Kauffman, G. et al. (2011). *Socioeconomic Value of the Chesapeake Bay Socioeconomic Value of the Chesapeake Bay Watershed in Delaware*. Tech. rep. March. Delaware Department of Natural Resources and Environmental Control’s Division of Watershed Stewardship.
- Kemp, W. et al. (2005). “Eutrophication of Chesapeake Bay: historical trends and ecological interactions”. In: *Marine Ecology Progress Series* 303, pp. 1–29.
- Kim, M. et al. (2016). “Transit time distributions and StorAge Selection functions in a sloping soil lysimeter with time-varying flow paths: Direct observation of internal and external transport variability”. In: *Water Resource Research* 52.9, pp. 7105–7129.
- Kirchner, J. W. (2015a). “Aggregation in environmental systems: catchment mean transit times and young water fractions under hydrologic nonstationarity”. In: *Hydrology and Earth System Sciences Discussions* 12.3, pp. 3105–3167.
- Kirchner, J. W. (2009). “Catchments as simple dynamical systems: Catchment characterization, rainfall-runoff modeling, and doing hydrology backward”. In: *Water Resources Research* 45.2, pp. 1–34.
- Kirchner, J. W., X. Feng, and C. Neal (2000). “Fractal stream chemistry and its implications for contaminant transport in catchments”. In: *Nature* 403.6769, pp. 524–527.
- (2001). “Catchment-scale advection and dispersion as a mechanism for fractal scaling in stream tracer concentrations”. In: *Journal of Hydrology* 254.1-4, pp. 82–101.
- Kirchner, J. (2006). “Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology”. In: *Water Resources Research* 42.3.
- (2015b). “Aggregation in environmental systems: seasonal tracer cycles quantify young water fractions, but not mean transit times, in spatially heterogeneous catchments”. In: *Hydrology and Earth System Sciences Discussions* 12.3, pp. 3059–3103.
- Klaus, J. and J. J. McDonnell (2013). “Hydrograph separation using stable isotopes: Review and evaluation”. In: *Journal of Hydrology* 505, pp. 47–64.

- Kling, H., M. Fuchs, and M. Paulin (2012). “Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios”. In: *Journal of Hydrology* 424-425, pp. 264–277.
- Kollet, S. and R. Maxwell (2006). “Integrated surface–groundwater flow modeling: A free-surface overland flow boundary condition in a parallel groundwater flow model”. In: *Advances in Water Resources* 29.7, pp. 945–958.
- Kollet, S. J. et al. (2017). “The integrated hydrologic model intercomparison project, IH-MIP2: A second set of benchmark results to diagnose integrated hydrology and feedbacks”. In: *Water Resources Research* 53.1, pp. 867–890.
- Krysanova, V. and J. G. Arnold (2008). “Advances in ecohydrological modelling with SWAT—a review”. In: *Hydrological Sciences Journal* 53.5, pp. 939–947.
- Lau, W. K. M., H. T. Wu, and K. M. Kim (2013). “A canonical response of precipitation characteristics to global warming from CMIP5 models”. In: *Geophysical Research Letters* 40.12, pp. 3163–3169.
- Li, Q. et al. (2008). “Simulating the multi-seasonal response of a large-scale watershed with a 3D physically-based hydrologic model”. In: *Journal of Hydrology* 357.3-4, pp. 317–336.
- Lin, H. (2006). “Temporal Stability of Soil Moisture Spatial Pattern and Subsurface Preferential Flow Pathways in the Shale Hills Catchment”. In: *Vadose Zone Journal* 5.1, p. 317.
- Lindeburg, K. (2011). “Silty Mantles and Fragipans in Pennsylvania Soils”. PhD thesis. Pennsylvania State University.
- Lindsey, B. D. et al. (2003). *Residence Times and Nitrate Transport in Ground Water Discharging to Streams in the Chesapeake Bay Watershed*. U.S Geological Survey.
- Linker, L. C. et al. (2013a). “Computing Atmospheric Nutrient Loads to the Chesapeake Bay Watershed and Tidal Waters”. In: *Journal of the American Water Resources Association* 49.5, pp. 1025–1041.
- Linker, L. C. et al. (2013b). “Development of the Chesapeake Bay Watershed Total Maximum Daily Load Allocation”. In: *Journal of the American Water Resources Association* 49.5, pp. 986–1006.
- Loague, K. et al. (2005). “Further testing of the integrated hydrology model (InHM): Event-based simulations for a small rangeland catchment located near Chickasha, Oklahoma”. In: *Hydrological Processes* 19.7, pp. 1373–1398.



- Maier, N., L. Breuer, and P. Kraft (2017). “Prediction and uncertainty analysis of a parsimonious floodplain surface water-groundwater interaction model”. In: *Water Resources Research* 53.9, pp. 7678–7695.
- Małozzewski, P. and a. Zuber (1982). “Determining the turnover time of groundwater systems with the aid of environmental tracers”. In: *Journal of Hydrology* 57.3-4, pp. 207–231.
- Maxwell, R. M. and S. J. Kollet (2008). “Quantifying the effects of three-dimensional subsurface heterogeneity on Hortonian runoff processes using a coupled numerical, stochastic approach”. In: *Advances in Water Resources* 31.5, pp. 807–817.
- Maxwell, R. M. and N. Miller (2005). “Development of a Coupled Land Surface and Groundwater Model”. In: pp. 233–247.
- Maxwell, R. M. et al. (2015). “The imprint of climate and geology on the residence times of groundwater”. In: *Geophysical Research Letters* 43.2, pp. 701–708.
- Maxwell, R. M. et al. (2016). *ParFlow User’s Manual*. Tech. rep.
- Maxwell, R. (2013). “A terrain-following grid transform and preconditioner for parallel, large-scale, integrated hydrologic modeling”. In: *Advances in Water Resources* 53, pp. 109–117.
- Maxwell, R. and A. Tompson (2006). “SLIM-FAST: a user’s manual, Lawrence Livermore National Laboratory, Livermore, California”. In:
- Maxwell, R. et al. (2014). “Surface-subsurface model intercomparison: A first set of benchmark results to diagnose integrated hydrology and feedbacks”. In: *Water resources research* 50, pp. 1531–1549.
- McDonnell, J. J. and K Beven (2014). “Debates - The future of hydrological sciences: A (common) path forward? A call to action aimed at understanding velocities, celerities and residence time distributions of the headwater hydrograph”. In: *Water Resource Research* 50.6, pp. 5342–5350.
- McDonnell, J. J. et al. (2010). “How old is streamwater? Open questions in catchment transit time conceptualization, modelling and analysis”. In: *Hydrological Processes* 24.12, pp. 1745–1754.
- McDonnell, J. J. (2003). “Where does water go when it rains? Moving beyond the variable source area concept of rainfall-runoff response”. In: *Hydrological Processes* 17.April, pp. 1869–1875.
- McDonnell, J. (2014). “The two water worlds hypothesis: ecohydrological separation of water between streams and trees?” In: *WIREs Water*.

- McGlynn, B., J. McDonnell, and D. Brammer (2002). "A review of the evolving perceptual model of hillslope flowpaths at the Maimai catchments, New Zealand". In: *Journal of Hydrology* 257.1, pp. 1–26.
- McGuire, K. J., D. R. DeWalle, and W. J. Gburek (2002). "Evaluation of mean residence time in subsurface waters using oxygen-18 fluctuations during drought conditions in the mid-Appalachians". In: *Journal of Hydrology* 261.1-4, pp. 132–149.
- McGuire, K. and J. J. McDonnell (2006). "A review and evaluation of catchment transit time modeling". In: *Journal of Hydrology* 330.3-4, pp. 543–563.
- McMillan, H. et al. (2012). "Do time-variable tracers aid the evaluation of hydrological model structure? A multimodel approach". In: *Water Resources Research* 48, W05501.
- Meals, D. W., S. a. Dressing, and T. E. Davenport (2010). "Lag time in water quality response to best management practices: a review." In: *Journal of environmental quality* 39.1, pp. 85–96.
- Meter, K. V. et al. (2016). "The nitrogen legacy: emerging evidence of nitrogen accumulation in anthropogenic landscapes". In: *Environmental Research Letters* 11.035014.
- Meyerhoff, S. B. and R. M. Maxwell (2011). "Quantifying the effects of subsurface heterogeneity on hillslope runoff using a stochastic approach". In: *Hydrogeology Journal* 19.8, pp. 1515–1530.
- Mirus, B. et al. (2009). "A hypothetical reality of Tarrawarra-like hydrologic response". In: *Hydrologic Processes* 23.November 2008, pp. 2267–2274.
- Monteith, J. L. (1965). "Evaporation and environment". In: *Symposia of the Society for Experimental Biology* 19, pp. 205–234.
- Morgenstern, U., M. K. Stewart, and R. Stenger (2010). "Dating of streamwater using tritium in a post nuclear bomb pulse world: continuous variation of mean transit time with streamflow". In: *Hydrology and Earth System Sciences* 14.11, pp. 2289–2301.
- Morgenstern, U et al. (2015). "Using groundwater age and hydrochemistry to understand sources and dynamics of nutrient contamination through the catchment into Lake Rotorua, New Zealand". In: *Hydrology and Earth System Sciences* 19.2, pp. 803–822.
- Moriiasi, D., J. Arnold, and M. V. Liew (2007). "Model evaluation guidelines for systematic quantification of accuracy in watershed simulations". In: *Transactions of the ASABE* 50.3, pp. 885–900.

- Morita, M. and B. C. Yen (2002). “Modeling of Conjunctive Two-Dimensional Surface-Three-Dimensional Subsurface Flows”. In: *Journal of Hydraulic Engineering* 128.2, pp. 184–200.
- Murphy, J. M. et al. (2007). “A methodology for probabilistic predictions of regional climate change from perturbed physics ensembles”. In: *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences* 365.1857, pp. 1993–2028.
- Najjar, R. G. et al. (2010). “Potential climate-change impacts on the Chesapeake Bay”. In: *Estuarine, Coastal and Shelf Science* 86.1, pp. 1–20.
- Nakicenovic, N. et al. (2000). *Special report on emissions scenarios: a special report of Working Group III of the Intergovernmental Panel on Climate Change*. Tech. rep. Cambridge, UK: Intergovernmental Panel on Climate Change.
- Nash, J. and I Sutcliffe (1970). “River Flow Forecasting Through Conceptual Models Part 1 - A Discussion of Principles”. In: *Journal of Hydrology* 10.
- Neal, C. (1997). “A view of water quality from the Plynlimon watershed”. In: *Hydrology and Earth System Sciences* 1.3, pp. 743–753.
- Neal, C, J Kirchner, and B Reynolds (2013). “Plynlimon research catchment high-frequency hydro-chemistry data”. In: *NERC-Environmental Information Data Centre*.
- Neal, C and J. Kirchner (2000). “Sodium and chloride levels in rainfall, mist, streamwater and groundwater at the Plynlimon catchments, mid-Wales: inferences on hydrological and chemical controls”. In: *Hydrology and Earth System Sciences Discussions* 4.2, pp. 295–310.
- Neal, C. and P. T. W. Rosier (1990). “Chemical studies of chloride and stable oxygen isotopes in two conifer afforested and moorland sites in the British uplands”. In: *Journal of Hydrology* 115.1-4, pp. 269–283.
- Needelman, B. a. et al. (2004). “Surface Runoff along Two Agricultural Hillslopes with Contrasting Soils”. In: *Soil Science Society of America Journal* 68.3, p. 914.
- Neitsch, S. et al. (2011). *Soil & Water Assessment Tool Theoretical Documentation Version 2009*. Tech. rep. Blackland Research Institute, pp. 1–647.
- Niemi, A. (1977). “Residence time distributions of variable flow processes”. In: *The International Journal of Applied Radiation and Isotopes* 28, pp. 855–860.

- Nohara, D. et al. (2006). “Impact of Climate Change on River Discharge Projected by Multimodel Ensemble”. In: *Journal of Hydrometeorology* 7.5, pp. 1076–1089.
- Page, T et al. (2007). “Modelling the chloride signal at Plynlimon , Wales , using a modified dynamic TOPMODEL incorporating conservative chemical mixing ( with uncertainty )”. In: *Hydrol. Process.* 307. August 2006, pp. 292–307.
- Palanaippan, M et al. (2010). *Clearing the waters: a focus on water quality solutions*. Tech. rep. UN Environment Programme.
- Pangle, L. A. et al. (2017). “The mechanistic basis for storage-dependent age distributions of water discharged from an experimental hillslope”. In: *Water Resources Research* 53.4, pp. 2733–2754.
- Paniconi, C. and M. Putti (2015). “Physically based modeling in catchment hydrology at 50: Survey and outlook”. In: *Water Resources Research* 51.9, pp. 7090–7129.
- Pedregosa, F. et al. (2011). “Scikit-learn: Machine Learning in Python”. In: *Journal of Machine Learning Research* 12, pp. 2825–2830.
- Peters, N., D. Burns, and B. Aulenbach (2014). “Evaluation of High-Frequency Mean Streamwater Transit-Time Estimates Using Groundwater Age and Dissolved Silica Concentrations in a Small Forested Watershed”. In: *Aquatic Geochemistry* 20.2-3, pp. 182–202.
- Pianosi, F. et al. (2016). “Sensitivity analysis of environmental models: A systematic review with practical workflow”. In: *Environmental Modelling and Software* 79, pp. 214–232.
- Pinder, G. (1973). *An iterative digital model for aquifer evaluation*. Tech. rep. U.S. Geological Survey.
- Pollock, D. (1988). “Semianalytical Computation of Pathlines for Finite Difference Models”. In: *Ground Water* 26.6, pp. 743–750.
- Pyke, C. R. et al. (2008). *Climate Change and the Chesapeake Bay: State-of-the-Science Review and Recommendations. A Report from the Chesapeake Bay Program Science and Technical Advisory Committee (STAC)*. Tech. rep. STAC.
- Qu, Y. and C. J. Duffy (2007). “A semidiscrete finite volume formulation for multiprocess watershed simulation”. In: *Water Resources Research* 43.8, pp. 1–18.
- Remondi, F. et al. (2018). “Water Flux Tracking with A Distributed Hydrological Model to Quantify Controls on the Spatio-Temporal Variability of Transit Time Distributions”. In: *Water Resources Research*, pp. 1–19.

- Richards, L. A. (1931). "Capillary Conduction of Liquids Through Porous Mediums". In: *physics* 1.5, pp. 318–333.
- Rinaldo, A et al. (2011). "Catchment travel time distributions and water flow in soils". In: *Water Resources Research* 47.7, W07537.
- Rinaldo, A. et al. (2015). "Storage selection functions: A coherent framework for quantifying how catchments store and release water and solutes". In: *Water Resources Research* 6, p. 446.
- Robinson, J. S. and M. Sivapalan (1997). "Temporal scales and hydrological regimes : Implications for flood frequency scaling". In: *Water Resource Research* 33.12, pp. 2981–2999.
- Rodhe, A, L Nyberg, and K Bishop (1996). "Transit times for water in a small till catchment from a step shift in the oxygen 18 content of the water input". In: *Water Resources Research* 32.12, pp. 3497–3511.
- Rodriguez, N., K. McGuire, and J. Klaus (2018). "Time-Varying Storage – Water Age Relationships in a Catchment with a Mediterranean Climate". In: *Water Resources Research*.
- Rooij, R. de, W. Graham, and R. M. Maxwell (2013). "A particle-tracking scheme for simulating pathlines in coupled surface-subsurface flows". In: *Advances in Water Resources* 52, pp. 7–18.
- Roth, K. et al. (1991). "Transport of chloride through an unsaturated field soil". In: *Water Resources Research* 27.10, pp. 2533–2541.
- Sanford, W. E. and J. P. Pope (2013). "Quantifying groundwater's role in delaying improvements to Chesapeake Bay water quality". In: *Environmental science & technology* 47.23, pp. 13330–8.
- Sanford, W. E. et al. (2012). *Simulation of Groundwater Flow in the Shallow Aquifer System of the Delmarva Peninsula, Maryland and Delaware*. Tech. rep. USGS, p. 68.
- Schall, J., E. Richardson, and J. Morris (2008). *Introduction to Highway Hydraulics*. Tech. rep.
- Schoups, G. and J. A. Vrugt (2010). "A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian errors". In: *Water Resources Research* 46.10, pp. 1–17.
- Schubel, J. and D. Hirschberg (1978). "Estuarine graveyards, climatic change, and the importance of the estuarine environment". In: *Estuarine Interactions* 285303.
- Sebben, M. L. et al. (2013). "On the testing of fully integrated surface-subsurface hydrological models". In: *Hydrological Processes* 27.8, pp. 1276–1285.

- Seck, A., C. Welty, and R. Maxwell (2015). "Spin-up behavior and effects of initial conditions for an integrated hydrologic model Alimatou". In: *Water Resource Research* 51, pp. 2188–2210.
- Sheshukov, A. Y. et al. (2011). "High Spatial Resolution Soil Data for Watershed Modeling: 1. Development of a SSURGO- ArcSWAT Utility". In: *Journal of Natural & Environmental Sciences Sciences Sciences* 2.2, pp. 15–24.
- Sklash, M. G. and R. N. Farvolden (1979). "The Role Of Groundwater In Storm Runoff". In: *Journal of Hydrology* 43, pp. 45–65.
- Soulsby, C et al. (2015). "Stream water age distributions controlled by storage dynamics and nonlinear hydrologic connectivity: Modeling with high-resolution isotope data". In: *Water Resource Research*, pp. 7759–7776.
- Spalding, R. F. and M. E. Exner (1993). "Occurrence of Nitrate in Groundwater—A Review". In: *Journal of Environment Quality* 22.3, p. 392.
- Spear, R. C. and G. M. Hornberger (1980). "Eutrophication in peel inlet-II. Identification of critical uncertainties via generalized sensitivity analysis". In: *Water Research* 14.1, pp. 43–49.
- Srivastava, V. et al. (2014). "Insights on geologic and vegetative controls over hydrologic behavior of a large complex basin - Global Sensitivity Analysis of an integrated parallel hydrologic model". In: *Journal of Hydrology* 519.PB, pp. 2238–2257.
- STAC (2013). *Incorporating Lag-Times Into the Chesapeake Bay Program*. STAC Publication 13-004.
- Sulis, M., C. Paniconi, and M. Camporese (2011). "Impact of grid resolution on the integrated and distributed response of a coupled surface-subsurface hydrological model for the des Anglais catchment, Quebec". In: *Hydrological Processes* 25.12, pp. 1853–1865.
- Svensson, T., G. M. Lovett, and G. E. Likens (2012). "Is chloride a conservative ion in forest ecosystems?" In: *Biogeochemistry* 107.1-3, pp. 125–134.
- Torres, R. et al. (1998). "Unsaturated zone processes and the hydrologic response of a steep, unchanneled catchment". In: *Water Resources Research* 34.8, p. 1865.
- Troch, P. et al. (1993). "Evaluation of a Distributed Catchment Scale Water Balance Model". In: *Water Resource Research* 29.6.
- Tromp-van Meerveld, H. J and J. J. McDonnell (2006). "Threshold relations in subsurface stormflow : 2 . The fill and spill hypothesis". In: *Water Resources Research* 42.WO2411, pp. 1–11.

- Uchida, T., I. Tromp-Van Meerveld, and J. J. McDonnell (2005). “The role of lateral pipe flow in hillslope runoff response: An intercomparison of non-linear hillslope response”. In: *Journal of Hydrology* 311.1-4, pp. 117–133.
- Urban, J. (1977). “The Mahantango Creek watershed: evaluating the shallow ground-water regime.” In: *Watershed research in eastern North America*. Ed. by D. L. Correll. Washington, D.C.: Smithsonian Institution, pp. 251–275.
- van der Velde, Y et al. (2012). “Quantifying catchment-scale mixing and its effect on time-varying travel time distributions”. In: *Water Resources Research* 48.6, W06536.
- van der Velde, Y. et al. (2014). “Consequences of mixing assumptions for time-variable travel time distributions”. In: *Hydrological Processes* 29.16, pp. 3460–3474.
- Van Meter, K. J., P. Van Cappellen, and N. B. Basu (2018). “Legacy nitrogen may prevent achievement of water quality goals in the Gulf of Mexico”. In: *Science* 430.3, pp. 1–7.
- VanderKwaak, J. E. and K. Loague (2001). “Hydrologic-response simulations for the R-5 catchment with a comprehensive physics-based model”. In: *Water Resources Research* 37.4, pp. 999–1013.
- Vogel, R. M. and N. Fennessey (1994). “Flow-duration curves. I: New Interpretation and Confidence Intervals”. In: *J. Water Resour. Plann. Manage.* 120.4, pp. 485–504.
- Vrugt, J. a. et al. (2009). “Accelerating Markov Chain Monte Carlo Simulation by Differential Evolution with Self-Adaptive Randomized Subspace Sampling”. In: *International Journal of Nonlinear Sciences and Numerical Simulation* 10.3, pp. 273–290.
- Vrugt, J. A. (2016). “Markov chain Monte Carlo simulation using the DREAM software package: Theory, concepts, and MATLAB implementation”. In: *Environmental Modelling & Software* 75, pp. 273–316.
- Weiler, M and J McDonnell (2004). “Virtual experiments: a new approach for improving process conceptualization in hillslope hydrology”. In: *Journal of Hydrology* 285.1, pp. 3–18.
- Wellen, C., A.-R. Kamran-Disfani, and G. B. Arhonditsis (2015). “Evaluation of the Current State of Distributed Watershed Nutrient Water Quality Modeling”. In: *Environmental Science and Technology* 49, pp. 3278–3290.
- Whitehead, P. G. et al. (2009). “A review of the potential impacts of climate change on surface water quality”. In: *Hydrological Sciences* 54.1, pp. 101–123.

- Wilkinson, J. et al. (1997). “Major, minor and trace element composition in cloudwater and rainwater at Plynlimon”. In: *Hydrology and Earth System Sciences* 1.3, pp. 557–569.
- Wilusz, D. C., C. J. Harman, and W. P. Ball (2017). “Sensitivity of Catchment Transit Times to Rainfall Variability Under Present and Future Climates”. In: *Water Resources Research* 53.12, pp. 10231–10256.
- Wilusz, D. (2015). “Review of SWAT Modeling of Nitrate Transport in Mesoscale Catchments”. Baltimore, MD.
- Wilusz, D. and W. Ball (2014). “Simulating the Effects of intra-monthly rainfall variability on Chesapeake Bay water quality under a changing climate”. In: *Chesapeake Bay Modeling Symposium 2014*. Annapolis, MD.
- Winchell, M. et al. (2010). *ArcSWAT Interface for SWAT2009 User’s Guide*. Tech. rep. Temple: Blackland Research Center.
- Yang, J. et al. (2018). “Exploring the Dynamics of Transit Times and Subsurface Mixing in a Small Agricultural Catchment”. In: *Water Resources Research* 54.3, pp. 2317–2335.
- Yang, Q. et al. (2015). “Increased nitrogen export from eastern North America to the Atlantic Ocean due to climatic and anthropogenic changes during 1901–2008”. In: pp. 1046–1068.
- Zhang, Q. (2018). “Synthesis of nutrient and sediment export patterns in the Chesapeake Bay watershed: Complex and non-stationary concentration-discharge relationships”. In: *Science of the Total Environment* 618, pp. 1268–1283.



## Curriculum Vitae - Dano Wilusz

Daniel (Dano) was born in Hartford, CT on March 27, 1977. He pursued his PhD in the Department of Environmental Health and Engineering at the Johns Hopkins University Whiting School of Engineering in Baltimore, MD, USA. Under the guidance of co-advisors Dr. Ciaran Harman and Dr. Bill Ball, Dano studied how watershed models can be improved to help resource managers understand and simulate the terrestrial effects of climate change on water quality. Before returning to graduate school, Dano was a Foreign Affairs Officer covering water security and environmental health issues at the U.S. Department of State in Washington, DC. He held various other posts during his six years with the Department including Presidential Management Fellow; Special Advisor to the President's Malaria Initiative at the U.S. Agency for International Development; and Environment, Science, Technology, and Health Advisor to the U.S. Embassies to Bolivia and India. Before entering the civil service, Dano developed and installed semiconductor equipment for a Fortune 500 technology firm. Dano received a Bachelor of Science with Honors in Applied and Engineering Physics from Cornell University in 1999, a Master of Public Policy from the University of California at Berkeley in 2006, and a certificate from the CDC's Environmental Public Health Leadership Institute in 2012. He currently resides in Baltimore, MD, with frequent visits to his hometown in Burlington, CT.