EFFECTS OF TOPOGRAPHIC DEPRESSIONS ON OVERLAND FLOW: SPATIAL PATTERNS AND CONNECTIVITY

by

Feng Yu

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy

Department of Earth, Atmospheric, & Planetary Sciences

West Lafayette, Indiana

December 2018

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Jon Harbor, Co-Chair

Department of Earth, Atmospheric, & Planetary Sciences

Dr. Marty Frisbee, Co-Chair

Department of Earth, Atmospheric, & Planetary Sciences

Dr. Jane Frankenberger

Department of Agricultural & Biological Engineering

Dr. Bernard Engel

Department of Agricultural & Biological Engineering

Approved by:

Dr. Linda Lee

Head of the Graduate Program

To my mother,

who is always standing on my side in the journey

ACKNOWLEDGMENTS

I am profoundly grateful to have the opportunity to work with my advisor Dr. Jon Harbor in my Ph.D. research and other exciting projects. His inspiring guidance and advice made this research possible and he always reminded me to look at broader impacts. I also highly appreciate his support and encouragement on my graduate school journey.

I would like to express my deep gratitude to the members of my advisory committee, Dr. Marty Frisbee, Dr. Jane Frankenberger, and Dr. Bernard Engel, for their constructive comments and suggestions based on different perspectives.

My sincere thanks go to Dr. Claire Buffaut, the research hydrologist at USDA-ARS, for her generous assistance in a field study of the watershed related to this dissertation. I also would like to thank Larry Theller, who taught me the best way of navigating the powerful world of GIS.

I acknowledge all of my friends and colleagues who helped to make this dissertation possible, and also appreciate the assistance from the technicians and secretaries at ESE and EAPS.

This research was financially supported by a Purdue ESE Lynn Fellowship and a College of Science Cagiantas Fellowship.

TABLE OF CONTENTS

LIST OF TABLES 8
LIST OF FIGURES 9
ABSTRACT 12
CHAPTER 1. INTRODUCTION 14
1.1 Background 14
1.2 Objectives 18
1.3 References 20
CHAPTER 2. A GPU-ACCELERATED SPATIAL PATTERN ANALYSIS ALGORITHM FOR LARGE-SCALE, HIGH-RESOLUTION 2D/3D HYDROLOGIC CONNECTIVITY METRICS USING ARRAY VECTORIZATION AND NEURAL NETWORK CONVOLUTIONS IN A MACHINE LEARNING FRAMEWORK 24
2.1 Abstract 24
2.2 Introduction 25
2.3 The Algorithm 29
2.3.1 Detailed review of current ICSLs concept and the need for a new algorithm 29
2.4 OMNI 32
2.4.1 Step 1: Initializing variables and settings 32
2.4.2 Step 2: Reading the gridded patterns into CPU memory as NumPy arrays and data preprocessing 34
2.4.3 Step 3: Extracting eligible location indices from the NumPy array and converting to MXNet GPU array to compute the pair-based lag distances 35
2.4.4 Computing the connectivity function and OMNI 38
2.5 CARD 41
2.6 TOPO 41
2.6.1 Step 2.5: Validating the topographic determined flow pathways 42
2.6.2 Dealing with surface depressions and flat areas in the DEM 47
2.7 Test results and discussions 49
2.8 Conclusions 58
2.9 Code availability and Test Data 59

2.10 References 60
CHAPTER 3. THE EFFECTS OF TOPOGRAPHIC DEPRESSIONS ON MULTI-SCALE OVERLAND FLOW CONNECTIVITY: A HIGH-RESOLUTION SPATIAL AND TEMPORAL PATTERN ANALYSIS APPROACH BASED ON CONNECTIVITY STATISTICS 63
3.1 Abstract 63
3.2 Introduction 64
3.3 Materials and Methods 67
3.3.1 Overview 67
3.3.2 Watershed description 68
3.3.3 Hydrological and meteorological data 69
3.3.4 Spatial data 71
3.3.5 DRM implementation 72
3.3.6 Connectivity metrics 73
3.3.7 Thresholding overland flow patterns 77
3.4 Results and discussion 78
3.4.1 Role of surface depressions on overland flow connectivity across spatial scales 78
3.4.2 Macro connectivity stages represented spatially based on $\Delta\tau(h)$ 81
3.4.3 Macro connectivity stages represented temporally based on OMNI(t), the spatial extent and their potential correlation with rainfall and streamflow discharge 83
3.4.4 $\Delta\tau(h)$ plots shapes and the interaction between overland flow and depressions at multiple spatial scales 84
3.4.5 Spatial distributions for the depression related changes of connectivity magnitude 85
3.5 Conclusions 86
3.6 References 89
CHAPTER 4. IMPACTS OF TOPOGRAPHIC DEPRESSIONS ON WATERSHED-SCALE HYDROLOGIC RESPONSE: CONNECTIVITY AND GRID RESOLUTION 95
4.1 Abstract 95
4.2 Introduction 95
4.3 Materials and Methods 98
4.3.1 Data resolution and model selection 98
4.3.2 Experimental design 99

4.3.3 Watershed description and hydrologic datasets 101
4.3.4 Spatial Data 102
4.3.5 DEM and grid resolution 102
4.3.6 Connectivity metrics 103
4.4 Results and discussion 106
4.5 Conclusions 115
4.6 References 117
CHAPTER 5. CONCLUSIONS, IMPLICATIONS, AND FUTURE WORK 124
APPENDIX A. DIRECTION RAINFALL METHOD, 2D HYDRAULIC MODEL SELECTION AND THE GPU-ACCELERATED CONNECTIVITY FUNCTION ALGORITHM 128
APPENDIX B. ILLUSTRATIONS OF FILL DOMINATED (FD) AND SPILL DOMINATED (SD) OVERLAND FLOW PROCESSES 138

LIST OF TABLES

Table 2.1 Variables used in the computation of connectivity function 33

Table 2.2 Recommended variable “broadcdp” values for tested GPU cards. Note that these are based on single precision (32-bit float point) computation 34

Table 2.3 Weighting kernels used in MXNet GPU array convolution operator. Note that the brackets represent each row of the 9x9 kernel window. Each column of the kernel window is separated by comma (,). 44

Table 2.4 Lengths and the bearings for the flowlines in pattern 3..... 54

Table 2.5 Connectivity function curve tabular results for test pattern 1..... 55

Table 2.6 Connectivity function curve tabular results for test pattern 2..... 55

Table 2.7 Connectivity function curve tabular results for test pattern 3..... 56

Table 2.8 OMNI and TOPO connectivity metrics computation results for test patterns 1-3. 56

Table 2.9 Topographic determined connectivity function curve tabular results or test pattern 2 (5m resolution). Columns in bold represent the lag-distance range bins which include the lengths for the flow lines. Rows in bold represent the bearings for the flow lines categorized into 16 cardinal/ordinal/secondary intercardinal direction 57

Table 3.1 Hydrological and spatial datasets and corresponding sources used In DRM implementation 72

Table 3.2 Representation of different slope trend on $\Delta\tau(h)$ plots for each macro connectivity stages (median values). Note that Group 4 (Stage δ) is omitted because of insufficient data points..... 82

Table 4.1 Grid resolutions for selected hydrologic/ hydraulic modeling publication 104

Table 4.2 Maximum depression storage volumes (D_{imax}) for DEM_{Ori} and DEM_{Fill} at different grid resolutions and cumulative rainfall amounts for each storm event..... 111

Table 4.3 Distributions of downslope directions at GCEW for DEM_{ori} and DEM_{fill} based on aspect analysis in ArcGIS. The dominant direction is East which is perpendicular to the direction of major stream channels: North. Removal (Filling) of depressions “dilutes” the difference between downslope directions because there is a significant increase in flat area extents. 114

LIST OF FIGURES

Figure 2.1. Example of flow pathway pattern and computation of connectivity function on a 15x15 gridded domain. Positive values indicate flow depth (unit in cm) and -1 indicates cells of "NoData". (a) The original spatial pattern; (b) A threshold of k=0.1 is applied to the pattern so each cell of the pattern is categorized into "high" – above or equal to the threshold (in light blue) and "low" below the threshold (in yellow). The pair of cells A and B (in red) are connected in the domain by a contiguous pathway of "high" cells based on eight neighboring cells. Domain outlet is marked as dark blue; (c) Each connected region is given a unique index number, and any pair of cells of the same index number are connected; the cells of "low" (in yellow) are converted to 0s; (d) Connectivity function graph and the omnidirectional connectivity (OMNI) for the flow pattern. OMNI is the area under curve and computed by adding the total area of a series of trapezoids (yellow, green, orange and cyan). 30

Figure 2.2 Two scenarios of determining the connectivity between point A and B based on a side view of a topography section. A and B are connected in OMNI because contiguous cells of "high" exist between them. However, their connectivity in TOPO is different. (a) A and B are not connected in TOPO because B is not at the one-way downslope flow direction of A (blocked), even if the depression is filled; (b) A and B are connected in TOPO because the large-scale gradient between A and B permit the flow from A to B when the depression is filled. 32

Figure 2.3 (a) The 16 Cardinal, Original and Secondary intercardinal directions; (b) The axis orientation of the NumPy (and MXNet) Array. The Cartesian coordinates is labeled in each cell and separated by the comma as x,y (in form of tuples). 35

Figure 2.4 Programming flow chart illustrates the computational steps for array manipulation (step 2 and 3 in OMNI), based on the flow pattern in Figure 1c as an example. Array 0 (Figure 1c): 2D array of shape 15x15. Array 1 (Region Z): Cells of Array 0 where > 0. Array 2 (Region G-Z): Cells of Array 0 where =0. t is the maximum index number of the clustered regions. In this example, t=4. Stars (*) indicate the paired tuples for computational operator (for Euclidean distance). 39

Figure 2.5 Directions and orientation of CARD in the array coordinate. The connected direction between paired two points: A and B is calculated by $\alpha = arctan(x1 - x2)2 + (y1 - y2)2$. β is the output direction rotated counterclockwise for 90 degree. 41

Figure 2.6 a) The definition of connected cells included from the eight neighboring cells of M. In this paper, we define the cells can be connected from all eight neighboring cells, i.e., shaded cells P_0, P_1 ... P_7, in a counterclockwise sequence starting from E. b) Definition of eight triangular facets in a counterclockwise sequence starting from E, represented by the yellow triangles. (The facet index is indicated in yellow text with black background). 43

Figure 2.7 Example of convolution operator in MXNet for neighbor 0 (East) at location M (1,1), in green. The computation procedure is 0×0+0×-1+0×0+1×0+1×1+1×0+2×0+2×0+2×0=0. This generates the result at the centered cell M and recorded in a separated array as dxi..... 44

Figure 2.8 Example of the "Breadth first search" method for identifying flow downslope traversed cells. The blue arrows represent the computed flow directions in dj for each cell. The flow directions for blank cells are undetermined. We start from location "0" and trace flow downslope. The numbers are the sequence of identified downslope cells. Each color represents one step: yellow

– step 1, grass green – step 2, orange – step 3, red – step 4, purple –step 5, gray – step 6, blue – step 7 and dark gold –step 8. 46

Figure 2.9 Test Patterns and DEM for connectivity metric OMNI, CARD and TOPO in a square shaped domain of 5000m x 5000m at 5m and 2m resolution. The blue circle (in a and b) and the yellow lines (in b and c) represent "high" regions (Z) of isolated water patches and the connectivity pathways in between. The diameter of each circle is 200m and they are uniformly distributed in the domain. The origin of these flow lines is at the location x=100 and y=100. The length of each line and bearing (from N-S direction) is listed in Table 2. DEM used for testing TOPO is presented in d. Note that in a-c, "low" regions are the areas in blank (white). 50

Figure 3.1 (a) Location of GCEW and instrumentation (Raingauges, weather stations and streamflow weir locations), and (b) location of Salt River Basin in Missouri. 69

Figure 3.2 Three major types of depressions in GCEW. (a) Pond and wastewater lagoon on each side of a road and DEM at the same location; (b) Ditch by a road embankment and DEM at the same location; (c) Small depression (puddle) in low land areas at the riparian zone covered by vegetation and DEM at the same location. 70

Figure 3.3 3D overlay of original DEM (top) and depression filled DEM (bottom). Surface depressions presented here include (a) wastewater treatment lagoon and (b) pond. Note that the internal areas of the depressions are filled with flat surfaces and the edges are smoothed (marked by red circles). 73

Figure 3.4 Connectivity function $\tau(h)$ and the integrated connectivity scale lengths (ICSLs). 74

Figure 3.5 OMNI(t) (solid line) and connectivity extent range counts $ext_{\tau}(t)$ (dash line) plots for the selected 5 storm events at GCEW: (a) Event 1: Jun 20, 1981 (AEP 5-20%); (b) Event 2: Jul 04, 1998 (AEP 20-50%); (c) Event 3: Jun 03, 2001 (AEP 20-50%); (d) Event 4: Jun 10, 2006 (AEP 1-5%); and (e) Event 5: Sep 03, 2008 (AEP 20-50%). Rainfall hyetographs and stream discharge hydrographs are stacked on time axis as comparison. 75

Figure 3.6 Difference of the connectivity function $\tau(h)$ plots between the filled DEM and the original DEM $\Delta\tau(h)=\tau_{fill}(h)-\tau_{ori}(h)$ for individual temporal snapshots at hourly intervals for each storm event (a-e: event 1-5). Lightness gradient of the blue represents the relative time points during the event from dark (start) to light (end). 79

Figure 3.7 Hierarchical structure of depressions with nested topology on a cross section of a topography. The process of fill, merge, and spill are also included. Blue arrows represent the dominate filling process and orange arrows represent the dominate spilling process. (a) Base topography; (b) Depressions in the 4^{th} hierarchy (Blue); (c-d) Depressions in the 3^{rd} hierarchy (Yellow); (e) Depressions in the 2^{nd} hierarchy (Green); and (f) Depressions in the 1^{st} hierarchy (Blue). Split of the water levels at connected flow extent indicate the changes of hierarchy. Depressions at higher hierarchy usually have larger surface areas. 80

Figure 4.1 Box plots of the storage volume (a) and drainage area (b) for depressions in DEMs directly interpolated (labeled "Direct", in read) from LiDAR data and TUFLOW resampled (labeled "Resample", in blue) from 1 m resolution grid. The difference between these two interpolation methods are minor. Note that the direct interpolation method is Natural Neighbor, and TUFLOW uses a TIN based interpolation method to downgrade the resolution. 100

Figure 4.2 RSCfi function plots for the three storm events at 2 m, 5 m and 10 m grid resolutions. (a) Event 1 in 1981, (b) Event 2 in 2006 and (c) Event 3 in 2008 108

Figure 4.3 Correlations between depression storage volume and drainage area based on DEM_{Ori} (Orange square) and DEM_{Fill} (Green triangle). (a-b) 2 m resolution; (c-d) 5 m resolution; (e-f) 10 m resolution. Distortion of drainage areas are disproportionally small for depressions of large storage volume, and the degree of this distortion is increased after depression removal (filling). 110

Figure 4.4 Distributions of downslope directions at GCEW for DEM_{ori} and DEM_{fill} based on aspect analysis in ArcGIS. The dominant direction is East which is perpendicular to the direction of major stream channels: North. Removal (Filling) of depressions "dilutes" the difference between downslope directions because there is a significant increase in flat area extents. 112

Figure 4.5 NRCfi function plots for the three storm events at 2 m, 5 m and 10 m grid resolutions. (a) Event 1 in 1981, (b) Event 2 in 2006 and (c) Event 3 in 2008. 114

ABSTRACT

Author: Yu, Feng. PhD
Institution: Purdue University
Degree Received: December 2018
Title: Effects of Topographic Depressions on Overland Flow: Spatial Patterns and Connectivity
Committee Chair: Jon Harbor, Marty Frisbee

Topographic depressions are naturally occurring low land areas surrounded by areas of high elevations, also known as “pits” or “sinks”, on terrain surfaces. Traditional watershed modeling often neglects the potential effects of depressions by implementing removal (mostly filling) procedures on the digital elevation model (DEM) prior to the simulation of physical processes. The assumption is that all the depressions are either spurious in the DEM or of negligible importance for modeling results. However, studies suggested that naturally occurring depressions can change runoff response and connectivity in a watershed based on storage conditions and their spatial arrangement, e.g., shift active contributing areas and soil moisture distributions, and timing and magnitude of flow discharge at the watershed outlet. In addition, recent advances in remote sensing techniques, such as LiDAR, allow us to examine this modeling assumption because naturally occurring depressions can be represented using high-resolution DEM. This dissertation provides insights on the effects of depressions on overland flow processes at multiple spatial scales, from internal depression areas to the watershed scale, based on hydrologic connectivity metrics. Connectivity describes flow pathway connectedness and is assessed using geostatistical measures of heterogeneity in overland flow patterns, i.e., connectivity function and integral connectivity scale lengths. A new algorithm is introduced here to upscale connectivity metrics to large gridded patterns (i.e., with > 1,000,000 cells) using GPU-accelerated computing. This new algorithm is sensitive to changes of connectivity directions and magnitudes in spatial patterns and is robust for large DEM grids with depressions. Implementation of the connectivity metrics to overland flow patterns generated from original and depression filled DEMs for a study watershed indicates that depressions typically decrease overland flow connectivity. A series of macro connectivity stages based on spatial distances are identified, which represent changes in the interaction mechanisms between overland flow and depressions, i.e., the relative dominance of fill and spill, and the relative speed of fill and formation of connected pathways. In addition, to study the role of spatial

resolutions on such interaction mechanisms at watershed scale, two revised functional connectivity metrics are also introduced, based on depressions that are hydraulically connected to the watershed outlet and runoff response to rainfall. These two functional connectivity metrics are sensitive to connectivity changes in overland flow patterns because of depression removal (filling) for DEMs at different grid resolutions. Results show that these two metrics indicate the spatial and statistical characteristics of depressions and their implications on overland flow connectivity, and may also relate to storage and infiltration conditions. In addition, grid resolutions have a more significant impact on overland flow connectivity than depression removal (filling).

CHAPTER 1. INTRODUCTION

This introduction provides the background and research need for this study which focuses on the effects of topographic depressions on overland flow connectivity and streamflow discharge response at multiple spatial scales. The objectives of this Ph.D. dissertation are also included, and each subsequent chapter addresses one key objective, presented in the form of a manuscript submitted (or to be submitted) to a scientific journal.

1.1 Background

In digital elevation models (DEMs), topographic depressions are a single cell or group of cells that are surrounded by neighboring cells of higher elevations. In distributed or semi-distributed watershed modeling, overland flow routing across a terrain, typically represented by a DEM, cannot be interrupted by depressions. Otherwise, flow routing would converge towards depressions and cannot be routed out of these depressions until they are filled, and only loops internally (Martz and Garbrecht, 1998; Rieger, 1998; Burrough *et al.*, 2015). Thus, depressions may lead to incorrect derivation of hydrologic parameters, such as flow accumulation, upslope contributing areas and flow directions, and may prevent or delay flow from reaching the watershed outlet. However, almost all DEMs include depressions. Typically, these depressions are removed before any model computation, based on the implicit assumption that depressions are artifacts rooted in the uncertainty of elevation data collection and DEM interpolation methods, rather than natural occurrences on the terrain surface (Marks *et al.*, 1984; O'Callaghan and Mark, 1984; Jenson and Domingue, 1988, Hutchinson, 1989).

The assumption that depressions are artifacts was a reasonable and pragmatic approach when DEMs and the original data used to derive DEMs, e.g., contour maps, were at relatively coarse spatial resolutions, e.g., > 30 m horizontal accuracy (Moore *et al.*, 1988; Jenson, 1991). Previous works have argued that few naturally-occurring depressions have a spatial scale of > 100 m^2, except in glaciated or karst terrain (Tarboton *et al.*, 1991; Ford, 1997), and this approach avoided significant problems in model development and applications. However, recent advances in remote sensing techniques have significantly reduced the gap between DEM resolution and the scale of

naturally-occurring depressions. For example, Light Detection and Ranging (LiDAR) on an airborne platform uses highly accurate laser altimetry to collect elevations for terrain surfaces with horizontal resolutions of < 1 m at watershed scales, up to thousands of square kilometers (Antolín and Brovelli, 2007; Shan and Toth, 2008). LiDAR elevation data provide information on naturally-occurring depressions at scales of several square meters and above. With such data, the assumption that depressions are all artifacts is no longer justified, and it is possible to include depression storage in watershed models. Depression storage plays a crucial role in overland flow processes and in the distributions of water and water-mediated pollutants and sediments (e.g., Ahmed et al., 2011; Darboux et al., 2002; Einsiedl, 2005). Removal of depressions may significantly alter distributions of terrain attributes, such as elevation, slope and profile curvature, which relate to topographic derivatives used in model computations (Lindsay and Creed, 2005).

In addition, depression removal methods have impacts on the accuracy of hydrologic derivatives from DEMs (Woodrow *et al.*, 2016). Depression filling based on the D8 flow algorithm proposed by Jenson and Domingue (1988) has been widely used by modelers because of its efficiency and availability in GIS and terrain processing software packages, such as ArcGIS, QGIS and SAGAGIS (Conrad et al., 2015; QGIS Development Team, 2015; ESRI Environmental Systems Research Institute, 2016). However, filling often generates large flat areas and unrealistic straight flow lines, and may produce significant changes of terrain morphology (Martz and Garbrecht, 1999; Planchon and Darboux, 2001; Nardi *et al.*, 2008). Breaching and other depression removal methods have been shown to be a better alternative because they produce more minimal changes to the original DEMs and a higher likelihood of maintaining naturally occurring depressions (Tianqi *et al.*, 2003; Lindsay, 2016). Land evolution process models based depression removal have been compared with traditional morphologic only depression removal methods and may yield a better physical representation of slopes and flow directions (Grimaldi *et al.*, 2007).

Despite growing awareness of the impact of depression removal on characteristics of topography and key hydrologic derivatives, little work to date has focused on the direct effects of depression removal on surface water processes and watershed responses. Incorporating depressions, e.g., depression storage capacity, in a watershed model, even if only empirically, may reduce modeling uncertainty and simplify calibration parameters (Amoah *et al.*, 2013). Depressions may cause

significant changes in surface water budgets and sedimentation at the watershed scale by trapping water and sediments, which may explain gaps between field or plot and watershed scale measurements (Almendinger *et al.*, 2014). These prior works have exclusively focused on modeling outputs, e.g., discharge at the watershed or field outlets, and have not investigated in detail the spatial and temporal distributions of the hydrologic processes that drive these outputs. Indeed, weather and surface-water flow data are not traditionally measured simultaneously and continuously over an entire watershed domain (Beven and Westerberg, 2011). Rather most data come from measurements conducted on a "point" basis, which are sparsely located within or at the field, plot, subwatershed and watershed boundaries or outlets. Therefore, much modeling efforts has focused on interpreting the dynamics and processes that convert and fit the input parameters (weather data, topography, and other landscape characteristics) to the observations obtained at the outlet, typically flow data. Depending on the intended uses of the modeling, this approach is suitable for projects in which the goal is estimation, conceptualization and calibration of runoff and streamflow discharge under storms of various magnitudes (K. Ajami *et al.*, 2004; Barco *et al.*, 2008; Easton *et al.*, 2008) or predicting the long-term effects of changes in a watershed environment, such as land use changes (Bhaduri *et al.*, 2001; Tang *et al.*, 2005; Gassman *et al.*, 2009). However, depressions are complex topographic features which are spatially distributed in a non-uniform manner and that can change over time as a result of erosion and sedimentation (Weng, 2002; Trevisani and Rocca, 2015). To study the impact of depressions on watershed hydrological processes, a "white box" approach is needed that explicitly captures the spatial and temporal effects of depressions on hydrologic processes across multiple scales and the connections between these scales.

Some field and plot scale studies have approached the impact of complex topographic features in surface water models using connectivity characteristics for spatially explicit flow patterns. For example, microtopographic characteristics such as slope and roughness because of depressions have been investigated in terms of depression storage filling states and the ratio of connected areas in flow patterns (the relative surface connection function, RSCf) (Antoine *et al.*, 2009; Peñuela *et al.*, 2015). The role of depression storage on overland flow transfer and triggering behavior has been also examined in terms of water percolation on a rough surface using a simplified empirical modeling approach to explicitly simulate overland flow patterns and connectivity variability at a

plot scale (Darboux *et al.*, 2002a). The concept of connectivity has been central to the study of hydrologic responses pattern heterogeneity, e.g., flow transfer in surface or subsurface connected structures (Western *et al.*, 2001; Knudby and Carrera, 2005; Ali and Roy, 2010; Renard and Allard, 2013). Definitions of connectivity vary in the current literature, but typically include three key aspects: landscape features (e.g., topography), hydrologic processes (e.g., surface or subsurface flow) and linkages (Ali and Roy, 2009). Landscape features and linkages determine the structure of connectivity while the interaction between hydrologic processes and the structure determine the function of the connectivity (Turnbull *et al.*, 2008). Overland flow processes on topography can be reflected on changes in spatial pattern, which is interpreted through the variability of connectivity. Surface and subsurface flow pattern connectivity is highly non-linear with frequent state changes at the grid cell scale, because of random distributions of soil wetness conditions or rapid formation of hydraulic connectedness (Lehmann *et al.*, 2007; Peñuela *et al.*, 2015). Runoff generations where fill and spill occurs because of depressions is also highly non-linear, which often shows a threshold behavior (e.g., Darboux *et al.*, 2002). To quantify the connectivity state changes in hydrologic patterns, traditional geostatistics have be previously tested, e.g., these based on variogram analysis, but often failed to represent connectivity instead of only spatial correlations. A two-point connectivity function originally proposed by (Allard, 1993) has been successfully used to study connectivity patterns for soil moisture, overland flow and subsurface flow (Western *et al.*, 2001; Meerkerk *et al.*, 2009; Renard and Allard, 2013). The advantage of this type of connectivity statistic is that it explicitly accounts for linked pathways and directions in hydrologic patterns. However, the current implementation of this type of connectivity statistic has been limited to plot or field scales that have a small number of cells, in part because of computational limits.

This PhD study is a first attempt to investigate the effects of depressions on connectivity and runoff response for single storm events using spatially and temporally explicit overland flow patterns from an internal depression to a watershed scale. The statistical concept of connectivity metrics used here is similar to Western *et al.*, 2001, but a new algorithm is created to make the implementation to large grids and spatial scales feasible, including a 2D omnidirectional and a 3D topographically-determined directional connectivity function. In addition, connectivity metrics which are sensitive to changes in grid resolutions and runoff response at watershed outlet are also introduced. This study is designed to contribute to more complete understanding of feedback

mechanisms between small-scale topographic features, i.e., depressions, and overland flow processes in a watershed.

1.2 Objectives

1. Create a new and accelerated algorithm for computing two-point connectivity statistics for overland flow patterns. This algorithm should be capable of processing patterns on large grids, i.e., with > 1,000,000 cells, within a timeframe of minutes to days (depending on the grid size), and improvement of current conceptual basis for quantifying connectedness in a hydrologic pattern. Apply a directional filter to the 2D omnidirectional connectivity function to separate the flow pathway directions based on cardinal and intercardinal directions. Develop a more robust version of the 3D topographically-determined (3D) directional connectivity for complex morphologic features, i.e., topographically with depressions. The issue of over dispersion or convergence should be eliminated by introducing a more constrained routing method. Then test the robustness and the sensitivity of the new algorithm using overland flow patterns on a hypothetical landscape. (Chapter 2)

2. Investigate the role of topographic depressions on connectivity based directly on spatially and temporally explicit overland flow pathway patterns across multiple scales in a watershed for single storm events. Generate the overland flow patterns from high-performance and fully distributed 2D hydraulic models using observed hydroclimatic and topographic datasets for the study watershed. Specifically, identify possible threshold behaviors that define watershed connectivity states changes and the role of depressions on overland flow processes in such changes. Investigate whether there are changes in spatial distributions of connectivity related to that of depressions in a watershed. (Chapter 3)

3. Investigate whether grid resolution significantly changes overland flow and runoff response (at watershed outlet) to depressions and possible threshold behavior, based on connectivity, and the spatial and statistical distributions of depression characteristics, e.g., depression storage volumes and drainage areas. (Chapter 4)

4. Summarize the major findings of this Ph.D. work on spatial and temporal explicit overland flow response to depressions based on changes in connectivity states (pattern heterogeneity) and dependency in scales. Discuss limitations of this work and implications for future research. (Chapter 5)

1.3 References

Ahmed Z, Rao DRM, Reddy KRM. 2011. Sustainable storm water management - An evaluation of depression storage effect on peak flow. In *Proceedings of the International Conference on Green Technology and Environmental Conservation, GTEC-2011*336–340. DOI: 10.1109/GTEC.2011.6167690

Ali GA, Roy AG. 2009. Revisiting hydrologic sampling strategies for an accurate assessment of hydrologic connectivity in humid temperate systems. *Geography Compass* DOI: 10.1111/j.1749-8198.2008.00180.x

Ali GA, Roy AG. 2010. Shopping for hydrologically representative connectivity metrics in a humid temperate forested catchment. *Water Resources Research* **46** (12) DOI: 10.1029/2010WR009442

Allard D. 1993. On the connectivity of two random set models: the truncated Gaussian and the Boolean. In *Geostatistics Tróia'92*Springer; 467–478.

Almendinger JE, Murphy MS, Ulrich JS. 2014. Use of the Soil and Water Assessment Tool to Scale Sediment Delivery from Field to Watershed in an Agricultural Landscape with Topographic Depressions. *Journal of Environment Quality* **43** (1): 9 DOI: 10.2134/jeq2011.0340

Amoah JKO, Amatya DM, Nnaji S. 2013. Quantifying watershed surface depression storage: determination and application in a hydrologic model. *Hydrological Processes* **27** (17): 2401–2413

Antoine M, Javaux M, Bielders C. 2009. What indicators can capture runoff-relevant connectivity properties of the micro-topography at the plot scale? *Advances in Water Resources* **32** (8): 1297–1310 DOI: 10.1016/j.advwatres.2009.05.006

Antolín R, Brovelli MA. 2007. LiDAR data filtering with GRASS GIS for the determination of digital terrain models. *I Jornadas de SIG Libre* (1): 13 Available at: http://www.sigte.udg.edu/jornadassiglibre2007/comun/2pdf/4.pdf

Barco J, Wong KM, Stenstrom MK. 2008. Automatic Calibration of the U.S. EPA SWMM Model for a Large Urban Catchment. *Journal of Hydraulic Engineering* **134** (4): 466–474 DOI: 10.1061/(ASCE)0733-9429(2008)134:4(466)

Beven K, Westerberg I. 2011. On red herrings and real herrings: Disinformation and information in hydrological inference. *Hydrological Processes* **25** (10): 1676–1680 DOI: 10.1002/hyp.7963

Bhaduri B, Minner M, Tatalovich S, Harbor J. 2001. Long-Term Hydrologic Impact of Urbanization: A Tale of Two Models. *Journal of Water Resources Planning and Management* **127** (1): 13–19 DOI: 10.1061/(ASCE)0733-9496(2001)127:1(13)

Burrough PA, McDonnell R, McDonnell RA, Lloyd CD. 2015. *Principles of geographical information systems*. Oxford University Press.

Conrad O, Bechtel B, Bock M, Dietrich H, Fischer E, Gerlitz L, Wehberg J, Wichmann V, Böhner J. 2015. System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. *Geoscientific Model Development* **8** (7): 1991–2007 DOI: 10.5194/gmd-8-1991-2015

Darboux F, Davy P, Gascuel-Odoux C. 2002. Effect of depression storage capacity on overland-flow generation for rough horizontal surfaces: water transfer distance and scaling. *Earth Surface Processes and Landforms* **27** (2): 177–191 DOI: 10.1002/esp.312

Darboux F, Gascuel-Odoux C, Davy P. 2002. Effects of surface water storage by soil roughness on overland-flow generation. *Earth Surface Processes and Landforms* **27** (3): 223–233 DOI: 10.1002/esp.313

Easton ZM, Fuka DR, Walter MT, Cowan DM, Schneiderman EM, Steenhuis TS. 2008. Re-conceptualizing the soil and water assessment tool (SWAT) model to predict runoff from variable source areas. *Journal of Hydrology* **348** (3–4): 279–291 DOI: 10.1016/j.jhydrol.2007.10.008

Einsiedl F. 2005. Flow system dynamics and water storage of a fissured-porous karst aquifer characterized by artificial and environmental tracers. *Journal of Hydrology* **312** (1–4): 312–321 DOI: 10.1016/j.jhydrol.2005.03.031

ESRI Environmental Systems Research Institute. 2016. ArcGIS Desktop: Release 10.4. *Redlands CA*

Ford DC. 1997. Principal features of evaportte karst in Canada. *Carbonates and Evaporites* DOI: 10.1007/BF03175798

Gassman PW, Williams JR, Wang X, Saleh A, Osei E, Hauck L, Izaurralde C, Flowers J. 2009. The Agricultural Policy Environmental Extender (APEX) model: An emerging tool for landscape and watershed environmental analyses. *Center for Agricultural and Rural Development (CARD) Publications, Ames*

Grimaldi S, Nardi F, Benedetto F Di, Istanbulluoglu E, Bras RL. 2007. A physically-based method for removing pits in digital elevation models. *Advances in Water Resources* **30** (10): 2151–2158 DOI: 10.1016/j.advwatres.2006.11.016

Hutchinson, M. F. (1989). A new procedure for gridding elevation and stream line data with automatic removal of spurious pits. *Journal of Hydrology*. https://doi.org/10.1016/0022-1694(89)90073-5

Jenson SK. 1991. Applications of hydrologic information automatically extracted from digital elevation models. *Hydrological Processes* DOI: 10.1002/hyp.3360050104

Jenson SK, Domingue JO. 1988. Extracting topographic structure from digital elevation data for geographic information system analysis. *Photogrammetric Engineering and Remote Sensing* **54** (11): 1593–1600 DOI: 0099-1112/88/5411-1593$02.25/0

K. Ajami N, Gupta H, Wagener T, Sorooshian S. 2004. Calibration of a semi-distributed hydrologic model for streamflow estimation along a river system. In *Journal of Hydrology*112–135. DOI: 10.1016/j.jhydrol.2004.03.033

Knudby C, Carrera J. 2005. On the relationship between indicators of geostatistical, flow and transport connectivity. *Advances in Water Resources* **28** (4): 405–421 DOI: 10.1016/j.advwatres.2004.09.001

Lehmann P, Hinz C, McGrath G, Tromp-van Meerveld HJ, McDonnell JJ. 2007. Rainfall threshold for hillslope outflow: An emergent property of flow pathway connectivity. *Hydrology and Earth System Sciences* DOI: 10.5194/hess-11-1047-2007

Lindsay JB. 2016. Efficient hybrid breaching-filling sink removal methods for flow path enforcement in digital elevation models. *Hydrological Processes* **30** (6): 846–857 DOI: 10.1002/hyp.10648

Lindsay JB, Creed IF. 2005. Removal of artifact depressions from digital elevation models: Towards a minimum impact approach. *Hydrological Processes* **19** (16): 3113–3126 DOI: 10.1002/hyp.5835

Marks D, Dozier J, Frew J. 1984. Automated Basin Delineation from Digital Elevation Data. *Geo-Processing*

Martz LW, Garbrecht J. 1998. The treatment of flat areas and depressions in automated drainage analysis of raster digital elevation models. *Hydrological Processes* **12** (6): 843–855 DOI: 10.1002/(SICI)1099-1085(199805)12:6<843::AID-HYP658>3.0.CO;2-R

Martz LW, Garbrecht J. 1999. An outlet breaching algorithm for the treatment of closed depressions in a raster DEM. *Computers and Geosciences* **25** (7): 835–844 DOI: 10.1016/S0098-3004(99)00018-7

Meerkerk AL, van Wesemael B, Bellin N. 2009. Application of connectivity theory to model the impact of terrace failure on runoff in semi-arid catchments. *Hydrological Processes* **23** (19): 2792–2803 DOI: 10.1002/hyp.7376

Moore ID, O'Loughlin EM, Burch GJ. 1988. A contour-based topographic model for hydrological and ecological applications. *Earth Surface Processes and Landforms* DOI: 10.1002/esp.3290130404

Nardi F, Grimaldi S, Santini M, Petroselli A, Ubertini L. 2008. Hydrogeomorphic properties of simulated drainage patterns using digital elevation models: The flat area issue. *Hydrological Sciences Journal* DOI: 10.1623/hysj.53.6.1176

O'Callaghan JF, Mark DM. 1984. The extraction of drainage networks from digital elevation data. *Computer Vision, Graphics, and Image Processing* **27** (2): 247 DOI: 10.1016/S0734-189X(84)80047-X

Peñuela A, Javaux M, Bielders CL. 2015. How do slope and surface roughness affect plot-scale overland flow connectivity? *Journal of Hydrology* **528**: 192–205 DOI: 10.1016/j.jhydrol.2015.06.031

Planchon O, Darboux F. 2001. A fast, simple and versatile algorithm to fill the depressions of digital elevation models. In *Catena*159–176. DOI: 10.1016/S0341-8162(01)00164-3

QGIS Development Team. 2015. QGIS Geographic Information System. *Open Source Geospatial Foundation Project* DOI: http://www.qgis.org/

Renard P, Allard D. 2013. Connectivity metrics for subsurface flow and transport. *Advances in Water Resources* **51**: 168–196 DOI: 10.1016/j.advwatres.2011.12.001

Rieger W. 1998. A phenomenon-based approach to upslope contributing area and depressions in DEMs. *Hydrological Processes* **12** (6): 857–872 DOI: 10.1002/(SICI)1099-1085(199805)12:6<857::AID-HYP659>3.0.CO;2-B

Shan J, Toth CK. 2008. *Topographic Laser Ranging and Scanning: Principles and Processing*. DOI: 10.1201/9781420051438

Tang Z, Engel BA, Pijanowski BC, Lim KJ. 2005. Forecasting land use change and its environmental impact at a watershed scale. *Journal of Environmental Management* **76** (1): 35–45 DOI: 10.1016/j.jenvman.2005.01.006

Tarboton DG, Bras RL, Rodriguez-Iturbe I. 1991. On the extraction of channel networks from digital elevation data. *Hydrological Processes* **5** (1): 81–100 DOI: 10.1002/hyp.3360050107

TIANQI AO, TAKEUCHI K, ISHIDAIRA H, YOSHITANI J, FUKAMI K. 2003. Development and application of a new algorithm for automated pit removal for grid DEMs. *Hydrological Sciences Journal* **48** (6): 985–997 DOI: 10.1623/hysj.48.6.985.51423

Trevisani S, Rocca M. 2015. MAD: robust image texture analysis for applications in high resolution geomorphometry. *Computers & Geosciences* **81**: 78–92 DOI: 10.1016/j.cageo.2015.04.003

Turnbull L, Wainwright J, Brazier RE. 2008. A conceptual framework for understanding semi-arid land degradation: ecohydrological interactions across multiple-space and time scales. *Ecohydrology* DOI: 10.1002/eco.4

Weng Q. 2002. Quantifying Uncertainty of Digital Elevation Models Derived from Topographic Maps. *Advances in spatial data handling*: 403–418 DOI: 10.1007/978-3-642-56094-1_30

Western AW, Blöschl G, Grayson RB. 2001. Toward capturing hydrologically significant connectivity in spatial patterns. *Water Resources Research* **37** (1): 83–97 DOI: 10.1029/2000WR900241

Woodrow K, Lindsay JB, Berg AA. 2016. Evaluating DEM conditioning techniques, elevation source data, and grid resolution for field-scale hydrological parameter extraction. *Journal of Hydrology* DOI: 10.1016/j.jhydrol.2016.07.018

CHAPTER 2. A GPU-ACCELERATED SPATIAL PATTERN ANALYSIS ALGORITHM FOR LARGE-SCALE, HIGH-RESOLUTION 2D/3D HYDROLOGIC CONNECTIVITY METRICS USING ARRAY VECTORIZATION AND NEURAL NETWORK CONVOLUTIONS IN A MACHINE LEARNING FRAMEWORK

Manuscript submitted to Water Resources Research

2.1 Abstract

The connectivity function measures the probability of the "connectedness" between any two points based on their lag-distances. "Connectedness" is defined as the existence of linked flow pathways between the two points. Watershed-scale connectivity function improve our understanding of hydrologic behaviors and the rainfall-runoff models. However, previous research has primarily focused on plot or field scale connectivity function because of limited data availability or computational constraints. Advances in remote sensing and high-performance computing techniques provide the potential to analyze high-resolution hydrologic patterns across broad spatial scales. Based on the concept of connectivity function, we propose a new algorithm for large-scale, high-resolution spatial grids. The key innovations include: 1) Parallelization of the computation by implementing array vectorization instead of single-cell based recursions; 2) Incorporation of a cardinal/diagonal/secondary intercardinal directions indicator for inspecting the general trend of connectivity patterns; 3) Improvement of the topographic connectivity by using efficient neural network convolutions operator to determine the flow directions and examine the impact of surface morphologic features, such as flat areas and surface depressions, on the linked pathways. We implemented and tested the algorithm in the MXNet environment, a widely-used machine learning framework with Python binding, to leverage its GPU accelerated high-performance computing capacity. Results suggested that the algorithm can be used on large grids (> 1,000,000 cells) and is sensitive to the connectivity states changes of hydrologic patterns and also robust to complex surface topography. We hope that our algorithm paves the way for investigating large-scale and system-wise connectivity behaviors in watershed hydrology.

2.2 Introduction

In watershed hydrology, connectivity is generally viewed in terms of pathways of water and water-mediated substances, such as pollutants and sediments, between spatial locations in a watershed during a finite temporal interval (Bracken and Croke, 2007). Research to date has focused on three connectivity types: terrain connectivity, soil moisture connectivity, and flow process connectivity (Ali and Roy, 2010), and connectivity analysis has been recognized as a promising approach to improving our understanding of watershed-scale hydrologic behavior. For example, runoff response and transfer of water-mediated substances have a positive correlation with the magnitude of connectivity (Tockner *et al.*, 1999; Heathwaite *et al.*, 2005; Knösche, 2006). Recent works on connectivity are based on landscape features and topographic gradients, and are interpreted using pattern analysis techniques, geostatistics, percolation theory or graph theory (Knudby and Carrera, 2005; Aurousseau *et al.*, 2009; Gascuel-Odoux *et al.*, 2009; Janzen and McDonnell, 2015). Connectivity has been considered as a potential alternative framework for interpreting watershed runoff generation mechanisms. Phillips et al. (2011) suggest that watershed models would better describe the physical process controls and have a higher model accuracy if model conceptualization is based on the connectedness of linear or non-linear temporal and spatial network for a large number of small-scale reservoirs, and constrained by landscape and environmental factors such as topography, soil moisture content, and vegetation.

Previous works on algorithms of hydrologic connectivity have primarily used the connectivity function, also known as "the Integral Connectivity Scale Lengths (ICSLs)", to quantify and characterize connectivity direction and magnitude for spatial patterns in soil moisture, overland flow, and subsurface flow (Trinchero *et al.*, 2008; Meerkerk *et al.*, 2009; Ali and Roy, 2010; Renard and Allard, 2013). ICSLs measure connectivity between two locations, either omni-directionally or in directionally specific ways, that incorporate geometrically determined trajectories such as pathways on surface topography or in subsurface material (Western *et al.*, 2001; Renard and Allard, 2013). ICSLs have also been applied to watersheds in various hydroclimatic regimes (e.g., Ali and Roy, 2010). The use of ICSLs in spatial pattern analysis has a close counterpart in traditional geostatistical techniques that are based on spatial continuity, i.e., the experimental variogram (Cressie, 1993). Spatial continuity examines the correlation strength of relevant spatial data at pairs of locations separated by lag-distances based on a search radius,

regardless of their "connectedness". ICSLs are also lag-distance based metrics, but with a specific emphasis on the connectedness in a pattern. ICSLs explicitly describe the spatial variability of connectedness based on the lag-distances (within a tolerance range) in the spatial pattern domain and measures the probability for multiple paired locations (cells) connected by a continuous pathway of locations instead of only between one particular pair of locations as in variogram based continuity (Western *et al.*, 2001). One of the most attractive aspects of the ICSL approach is simplicity: it collapses the spatial connectivity states of the target spatial domain across different lag-distances into a simple indicator value: A fully connected pattern has an ICSL value equivalent to the maximum lag distance measured the pattern domain, while a fully disconnected pattern has an ICSL value of zero. This is not without its own bias but it provides modelers with a direct and intuitive way of measuring and comparing the connectivity within or between hydrologic systems.

Theoretically, ICSLs can be "scaled up" from a plot or field scale to a watershed scale simply by extending the spatial coverage of the patterns being computed. For spatial continuity, however, a complex trial and error process is often needed that consists of shifting the search radius to determine the optimum threshold of the kernel window that is the most representative of the pattern (Cressie, 1993; Trevisani *et al.*, 2009). In addition, directional indictors can be added more efficiently and flexibly to ICSLs, e.g., based on surface or subsurface topography, to demonstrate the impedance effects (physical constraints) on the connected pathways, instead of cardinal directions only in spatial continuity. Thus, ICSLs can potentially be computed in one-, two- or three spatial dimensions, instead of one or two dimensions only for continuity matrices. ICSLs can also capture temporal variability of connectivity states during hydrologic events, i.e., storms and flooding, by converting temporal "snapshots" of the relevant patterns (such as overland flow, surface runoff pathways or inundation patterns) into a series of integrated indicator values within the event duration.

Research to date in quantifying hydrologic connectivity has almost exclusively focused on plot or field scale connectivity patterns derived from gridded lattice structures without georeferencing and with restricted spatial resolution (e.g., 500m) or sampling points (e.g., 15m/point) (Western *et al.*, 2001; Ali and Roy, 2010). Few attempts have been made to scale up ICSLs computations to watershed-scales and high-resolution gridded patterns

(Knudby and Carrera, 2005; Meerkerk *et al.*, 2009; Ali and Roy, 2010). High resolution connectivity studies at a watershed scale can provide the critical information on hydrologic processes and controls for rainfall-runoff modeling and water resources management practice that may have been missed in work using much smaller and coarser spatial and temporal resolutions. The only published work to date which applies ICSLs on a watershed-scale that we are aware of is Ali et al. 2010 and their study area was a small-scale watershed (0.051 km2). The two most likely reasons for the lack of work to date on connectivity of high-resolution hydrologic gridded patterns and large spatial scales include: 1) Sample collection for hydrologic patterns encompassing large spatial and temporal scales requires considerable time and resources; 2) Analyses of large amounts of spatial data require considerable computational capacity that may be beyond the reach of projects with modest resources.

Contemporary advances in GIS (Geographic Information System), remote sensing, and high-performance computing techniques provide ways to reduce or eliminate several limitations with implementing ICSLs and providing high resolution watershed-scale data (DeVantier and Feldman, 1993; Correia and Rego, 1998; Siart *et al.*, 2009; Vieux, 2016). Satellite imagery from the Worldview sensor or airborne Light Detection and Ranging (LiDAR) point clouds collected by Leica-Geosystems LiDAR sensors, can provide elevation measurements for digital elevation model (DEM) generation at spatial resolutions of < 2m and areal scales of > 1000km2 (Gehrke *et al.*, 2008; Saldaña *et al.*, 2012; Hu *et al.*, 2016). Overland flow patterns and inundation extents have also been collected through high-resolution aerial imagery in real time (Schumann *et al.*, 2007; Patro *et al.*, 2009). In addition, in-situ hydrologic data acquisition at watershed scales has become increasingly automated, allowing denser sampling in space and time, and telemetry allows rapid and frequent transmission of data (Shaw *et al.*, 2010). These diverse data sets can be integrated within a GIS to extend our current understanding of connectivity to a watershed or even at larger spatial or temporal scales, because the locations of nodes (elements) on the squared lattice-based patterns can be transformed and georeferenced into a 2D/3D projected coordinate system, which is managed and utilized by a GIS. This allows for high-resolution and spatially and temporally distributed overland flow patterns covering an entire watershed at timescales down to hours or minutes and differentiating areas such as hillslopes, floodplains, and stream channels, which is critical for connectivity studies.

Here we present the results of work to overcome practical and computational limitations of ICSLs algorithms in the current literature. Our algorithm is based on the mathematical concept of ICSLs proposed by Western et al. 2001, Meekerk et al. 2009 and Ali et al. 2010. We created a new programming workflow and boosted the computational efficiency by introducing array vectorization and neutral network convolution operators. Array vectorization eliminate any single-cell based loops and greatly boosted the computational and memory use efficiency. Neural network convolutions can efficiently generate flow directional grid in short amount of time for large scale and high-resolution DEM. We also made major improvements to incorporate additional directional indicators for the watershed hydrologic patterns. These includes: 1) incorporate a cardinal/diagonal/secondary intercardinal direction indicator that allows the model to represent the evolving trends of connectivity directions in hydrologic patterns spatially and temporally at a large spatial scale; 2) revise the topographically-determined connectivity by allowing a multiple flow direction routing method and computation of connectedness for topography with "difficult" surface morphologic features, including flat areas and surface depressions. Therefore, the effects of impedance (constraint) for connected water flow pathways because of these flat areas and surface depressions can be explicitly quantified. We believe that this new ICSLs algorithm will be more realistic and robust in reflecting actual flow pathways for various surface morphologic scenarios than the original implementation.

We implemented our enhanced algorithm in a machine learning framework, MXNet (Chen *et al.*, 2015), which is widely used in machine learning and deep learning applications, such as ecological pattern analysis, geological features classification and spatial object recognition (Chen *et al.*, 2017; Díez Hermano, 2017; Li *et al.*, 2018). Note that the primary reason for using MXNet is not to directly adopt a machine learning method, but rather leverages heavily parallel GPU computational capacity for time efficiency, i.e., to allow for applications to large-scale spatial patterns with cell numbers > 1,000,000. Our algorithm is generic (imperative) in style so that it can be adapted to different programming languages. Currently, the algorithm is designed to run on a single GPU device. To test our algorithm, we implemented the Python-MXNet code on a hypothetical landscape of 25 km2 with gridded raster datasets for high-resolution (2m and 5m) surface overland flow patterns and DEMs.

2.3 The Algorithm

2.3.1 Detailed review of current ICSLs concept and the need for a new algorithm

The omnidirectional ICSLs (OMNI) or omnidirectional connectivity function τ(h±r) (Allard, 1994; Western *et al.*, 2001) represents the probability that any pair of cells (elements or locations) with lag-distances hs within a tolerance range ±r are connected by contiguous neighboring cells. This is based on spatial patterns of hydrologic characteristics or states of interest, such as soil moisture conditions or surface flow pathways (e.g., Fig. 2.1a). τ(h±r) is computed by labeling each individual cell on a 2D gridded pattern as "high" or "low" based on a predefined threshold k (Fig. 2.1b). The threshold can either be time-variant based on specific percentiles of multivariate distributions of the variables of interest, or time-invariant based on a fixed physical indicator, such as a percentage of volumetric soil moisture (Anderson *et al.*, 2006a; Ali and Roy, 2010). Here we use a time-invariant threshold as a demonstration. Our example is an overland flow depth pattern with a 0.1m depth threshold to differentiate between very shallow sheet flow and concentrated/channel flow (USDA-NRCS, 1986). Cells labeled as "high" are connected if any of their eight neighboring cells in horizontal, vertical and diagonal directions are also labeled "high", which indicates a physically active status of flow such as concentrated/channel flow sufficient to affect watershed-scale hydrologic responses. The algorithm then randomly "scans" and pairs any two cells labeled "high" in the spatial pattern to test for connectivity until all possible pairs are found. A pair of cells is connected if there is a linked pathway of contiguous "high" cells between them (e.g., cell A and B in Fig 1b). τ(h) calculates the probability that any randomly paired cells x and x+h separated by a lag distance h (with tolerance range ±r) are connected. It is omnidirectional because the pathways connecting between x to x+h are always valid regardless of their trajectories and impedance factors, e.g., terrain morphological barriers. Let G be all the cells in the spatial pattern and Z be the regions where all cells are "high", the connectivity function is:

$$\tau(h) = P(x \leftrightarrow x + h \mid x \in Z, x + h \in G) \quad (1)$$

OMNI is computed as the integral area under the curve of the connectivity function across different lag-distances (Fig. 2.1d). At a watershed scale, OMNI represents the connectivity states for the measured hydrologically relevant patterns, such as overland flow pathways, at a particular time.

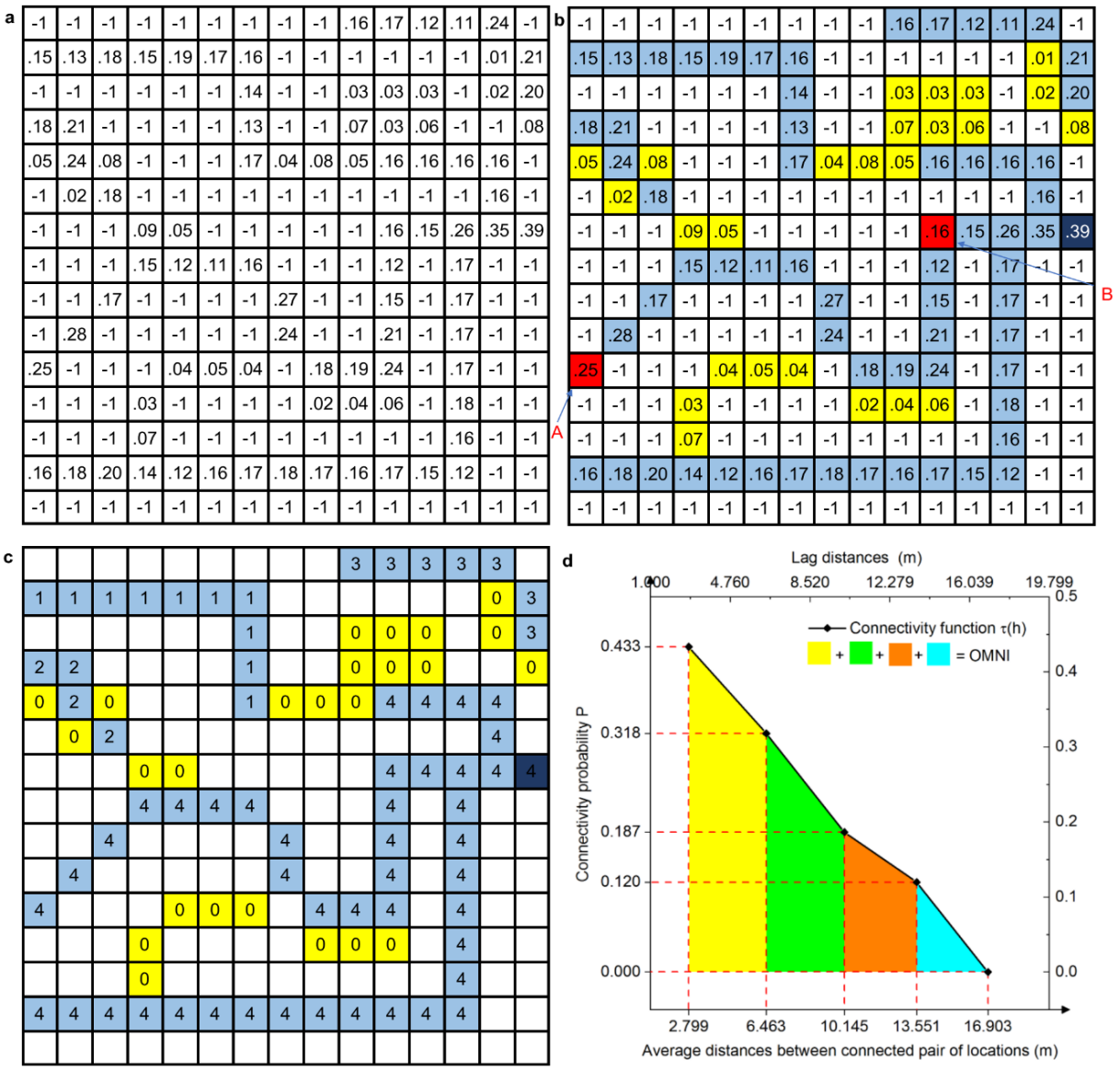


Figure 2.1. Example of flow pathway pattern and computation of connectivity function on a 15x15 gridded domain. Positive values indicate flow depth (unit in cm) and -1 indicates cells of "NoData". (a) The original spatial pattern; (b) A threshold of k=0.1 is applied to the pattern so each cell of the pattern is categorized into "high" – above or equal to the threshold (in light blue) and "low" below the threshold (in yellow). The pair of cells A and B (in red) are connected in the domain by a contiguous pathway of "high" cells based on eight neighboring cells. Domain outlet is marked as dark blue; (c) Each connected region is given a unique index number, and any pair of cells of the same index number are connected; the cells of "low" (in yellow) are converted to 0s; (d) Connectivity function graph and the omnidirectional connectivity (OMNI) for the flow pattern. OMNI is the area under curve and computed by adding the total area of a series of trapezoids (yellow, green, orange and cyan).

In this paper, we believe that a spatial direction indicator is also needed to describe the anisotropy, similar to a directional experimental variogram, i.e., directionally-biased spatial trends of flow resistance/assistance that characterize topographic patterns in high-resolution DEMs. Directions of connectivity patterns can be incorporated into OMNI based on the Cardinal/ordinal/secondary intercardinal directions (CARD) for the connected trajectories, similar to the "directions" defined in the directional variogram. CARD measures the general trend of cardinal directions of the connected pairs in addition to the probability of connectivity measured by OMNI, which also captures and quantifies the spatial anisotropy of the entire pattern. TOPO is computed in the same manner as OMNI but adds a physical constraint to the connected trajectories between the paired locations based on topography, i.e., trajectories of linked flow pathways connecting two paired locations x and x+h must follow either a one-way continuously downslope or upslope routing direction on the DEM, otherwise they are not valid (Fig. 2.2a). In work to date, two alternatives for computing the downslope or upslope flow pathways (a simplified flow routing method) are available: 1) Assign the two steepest gradients from the eight neighboring cells as the downslope or upslope directions (Western *et al.*, 2001); 2) Treat all lower/higher elevation cells among the eight neighboring cells as down/upslope (Meerkerk *et al.*, 2009; Ali and Roy, 2010). However, these flow direction assignments are likely to either underestimate or overestimate flow directions by creating unrealistically condensed flow, or overly dispersed flow paths. In addition, neither of these methods can adequately detect and process down/upslope directions on complex surface morphologic features, including flat areas and surface depressions. These features may still connect flow, depending on the large-scale topographic gradient, but are excluded in the current TOPO algorithm (Fig. 2.2b).

Therefore, a more robust flow routing algorithm is needed for a more accurate estimation of TOPO. In our work we use the Triangular Multiple Flow Direction (MD∞) (Seibert and McGlynn, 2007). MD∞ allows the assignment of flexible and multiple flow directions with a dispersion control parameter, i.e., an exponential weighting function, instead of restricted one flow direction (Quinn *et al.*, 1991; Seibert and McGlynn, 2007). Flow directions of MD∞ are also not restricted to flow directions at two contiguous neighboring cells. Flat areas and surface depressions that are identified can be processed further based on the large-scale topographic trend of the DEM, which determines whether flow is possible between the paired cells.

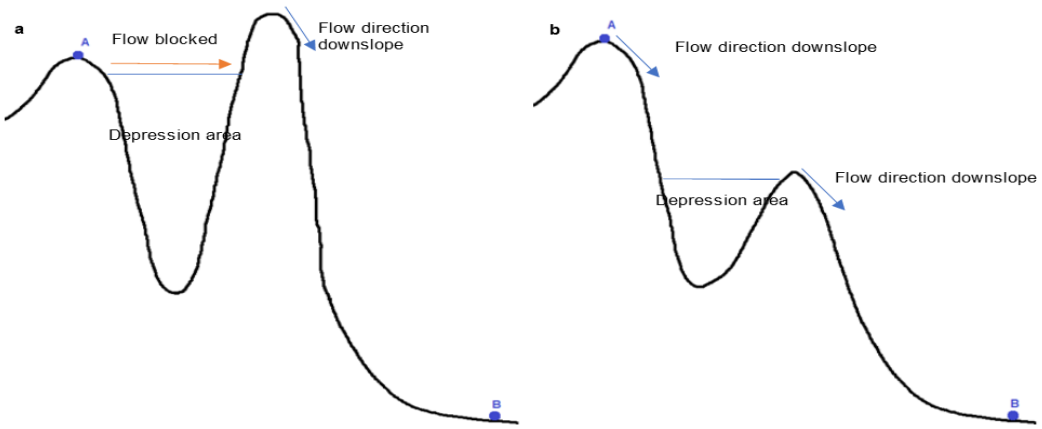


Figure 2.2 Two scenarios of determining the connectivity between point A and B based on a side view of a topography section. A and B are connected in OMNI because contiguous cells of "high" exist between them. However, their connectivity in TOPO is different. (a) A and B are not connected in TOPO because B is not at the one-way downslope flow direction of A (blocked), even if the depression is filled; (b) A and B are connected in TOPO because the large-scale gradient between A and B permit the flow from A to B when the depression is filled.

If so, the flow will be routed (or switch "on") outside these problematic features and connected to flow lines from nearby regions and eventually connected to the other cell in a pair. Otherwise, flow directions remain "unresolved" (or switch "off") and the paired cells are not connected. This functionality is useful for connectivity studies associated with the impedance effects of "difficult" topographic features, i.e., surface depressions and flat areas, at larger spatial gradient. Note that we do not ensure the accuracy of routing pathways within surface depressions and flat areas in this paper since our goal is to validate the linkage between pairs of cells and compute connectivity function rather than model the physical flow mass and momentum transverse along the pathways.

2.4 OMNI

2.4.1 Step 1: Initializing variables and settings

The algorithm requires user input for several variables as predefined parameters (Table 2.1).

Variable *broadcdp* is the computation depth which determines the number of cells extracted from the gridded patterns that are included in one iteration of computation for lag-distances between paired cells. The reason for limiting the total number of cells in one iteration is to avoid GPU memory overload. The recommended value of *broadcdp* is between 5,000 and 7,000 based on the memory capacity of the GPU card (Table 2.2). Variable *resolution* is the spatial resolution, i.e., cell size, of the gridded patterns. Variable *threshold* is used for categorizing the value of each cell

in the patterns in to “high” and “low”. Variable h_{min} and h_t are the minimum lag-distance and the tolerance range. The exact values of h_{min}, h_t, and *threshold* depend on the objective of the connectivity computation and the hydrologically relevant patterns that are being evaluated. In this paper, we use surface flow pathways as an example. The recommended values are h_{min}=68m, h_t=±22m (225 ft ± 75 ft) and *threshold=0.1m* for overland flow patterns, as a way of separating very shallow sheet flow from concentrated or channel flow based on horizontal flow travel distances (NRCS, 1986). Variable I_{max} is the maximum possible lag distance and is estimated automatically based on the spatial dimension of the gridded pattern as the hypotenuse length of the domain’s enclosing shape (rectangle). The information for the domain’s shape can be found in the metadata of the raster data, which shows the spatial coordinates of the four corners of the domain. Therefore, the length at NE-SW or NW-SE directions, i.e., the maximum possible distance l_{max} of the domain can also be computed. Then, l_{max} is divided by a series of range bins of equal interval. The total number of range bins n_{bin} are defined by the lag distance h and the tolerance h_t, such that:

$$n_{bin} = \frac{l_{max}}{h_{min} + 2|h_t|} (Rounded\ to\ the\ closest\ integer) \quad (2)$$

Note that n_{bin} must be a constant for all the patterns computed because the values of OMNI computed for the same spatial pattern based on different n_{bin} are not directly comparable.

Table 2.1 Variables used in the computation of connectivity function

Variable name	Explanation	Recommended Value
broadcdp	Computational depth (number of elements included in one iteration)	See Table 2
threshold	The threshold to separate the input pattern into “high” and “low”, i.e., eligible and not eligible for connection	Based on the physical character for the input hydrologic relevant patterns
h_{min}	The minimal lag-distance measured (sensitivity) for the pattern connectivity	Based on the physical character for the input hydrologic relevant patterns
h_t	Lag-distance tolerance range, usually the accuracy or uncertainty of the lag-distance h allowed in the computation	Based on the physical character for the input hydrologic relevant patterns
resolution	Spatial resolution of the input pattern grid (in raster format)	Refer to the GIS raster file’s metadata
NoData	Missing data value of the input pattern grid (in raster format)	Refer to the GIS raster file’s metadata
seriesname	The pattern grid file names	Name of watershed or location + time of the input pattern

Table 2.2 Recommended variable "broadcdp" values for tested GPU cards. Note that these are based on single precision (32-bit float point) computation

GPU card	Recommended "broadcdp" value
K80/Titan Z	5000
GTX 1080 Ti	6200
Tesla P100 (12G)	6500
Tesla V100	7500

2.4.2 Step 2: Reading the gridded patterns into CPU memory as NumPy arrays and data preprocessing

The gridded patterns of surface flow pathways in GIS raster data format are read as 2D NumPy arrays into CPU memory and their original spatial arrangement of cells is preserved as the array index locations (Cartesian coordinates). Note that the positive of the x and y axis is in a NumPy array index is toward S and E respectively (Fig. 2.3a). The variable *threshold* is applied to the NumPy arrays so cells above the threshold are assigned the value of 1s ("high" for concentrated and channel flow) and cells below the threshold are assigned the value of 0s ("low" for very shallow sheetflow). Cells of missing data value or "Nodata" are assigned the value of 1s. In the original algorithm proposed by Western et al. (2001), searching for the paired cells and computing their pair-wise lag-distances requires loops for each cell. This may be intuitive but significantly reduces the efficiency of the algorithm and is also not memory optimized for a large number of cells. In this paper, we assume that any connected pairs can only be found within a spatial clustered region, i.e., each cell is connected to another cell at least with one of the eight neighboring cells. Therefore, we reduce the computational intensity of the pairing process by limiting the searching boundary to each spatial clustered region assigned a unique region index. The clustered regions are generated using the SciPy module function "label" in Python. We firstly define the kernel shape of the eight neighboring cells in terms of which one(s) are defined as "connected" in a grid. Then, SciPy module function "label" scans the cells of 1s to group connected cells (with a value of "high" or 1) using sequential integers as the region index starting from *1* to *t*. Then, cells of 1s are replaced by the regional index number (e.g., Fig. 2.1c). The preprocessing of the NumPy array is completed and to simplify the later description, we name the processed array *Array 0*.

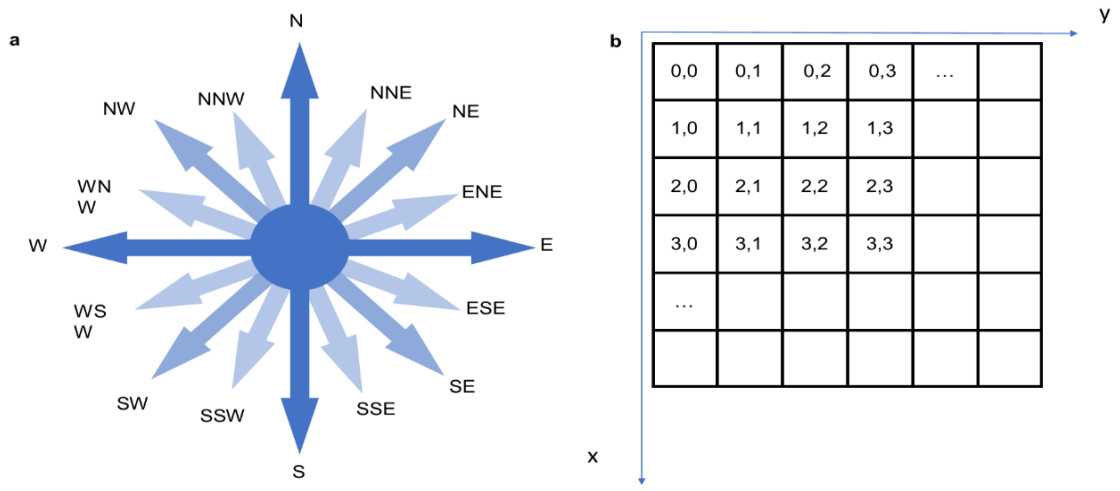


Figure 2.3 (a) The 16 Cardinal, Original and Secondary intercardinal directions; (b) The axis orientation of the NumPy (and MXNet) Array. The Cartesian coordinates is labeled in each cell and separated by the comma as x,y (in form of tuples).

2.4.3 Step 3: Extracting eligible location indices from the NumPy array and converting to MXNet GPU array to compute the pair-based lag distances

The lag-distances between pairs of cells are computed as Euclidean distances. This requires the Cartesian coordinates for both cells in a pair in Array 0. We extract the coordinates of Array 0 using NumPy function "where" for cells of ts (in Z) and cells of 0s (in G-Z). To efficiently compute the lag-distances between any random pair of cells (in G, regardless of the connectedness) and between connected pair of cells (in Z) in the pattern domain, we use the array vectorization instead of loops. The most efficient way of vectorizing NumPy arrays is "broadcasting", i.e., scalar-based matrix operations between arrays of different dimensions (shapes). Array broadcasting between two arrays is equivalent to generate the full permutations between the elements of each array, i.e., each element in one array is paired with each element in another array, regardless of the redundancy in order. Array broadcasting can also be achieved in MXNet and it further leverages the highly paralleled computation capacity on a GPU card to significantly accelerate the computation. GPU memory is usually more limited compared with CPU (system) memory. Therefore, we extract the index coordinate array of Array 0 and slice it into smaller "chunks" before converted to MXNet GPU array.

In the algorithm implementation, the Cartesian coordinates of Array 0 are extracted into two NumPy arrays, based on their values: one array represents cells with value > 0 (in Z and named as

Array 1) and the other represents cells with value = 0 (in G-Z and named as Array 2). Note that we made some minor modifications to accommodate for the difference in the axis orientations of the MXNet GPU array compared with the commonly used 2D Cartesian coordinate system (e.g., Fig 2.3b). For any 2D NumPy or MXNet array, the coordinates in the first dimension are always the row index number and the coordinates in the second dimension are always the column index number in a tuple, i.e., (row#, column#). Also, the origin of the coordinates is set at the top left corner. Array 1 and Array 2 are divided into small segments and assigned a unique segment index before being converted GPU arrays to prevent the GPU memory from overflow and to facilitate further manipulations. The length of these segments, i.e., the number of the array elements "sliced" in each iteration, is equal to *broadcdp*. We name the n_{th} segment of Array 1 as Sn_{A1}, and the m_{th} segment of Array 2 be Sm_{A2}. Sn_{A1} and Sm_{A2} are then transferred into MXNet GPU arrays and being vectorized (broadcast), which generates two-elements permutations. Because each element of Sn_{A1} or Sm_{A2} represents the locations of cells within the clustered regions in Array 0 (> 0) or the "low" regions in Array 0 (=0), this pairing procedure is similarly as recursively searching the paired locations between Array 1 (cells in Z) and Array 2 (cells in G excluding Z, or G-Z) without loops. Similarly, cells of Array 1 are separated into groups of different spatial clustered regions as Array 1_t, where t is the clustered region index number. Each Array 1_t is segmented and searched with itself again to compute the total pairs within Z, where the cell values are > 0 (cells in the clustered regions). We add additional logic to determine if two segments from Array 1, namely Sn_{A1} and Sn_{A1}', are identical to exclude the redundant element-wise pairing between the same pair of switched order, e.g., (x,y) and (y,x). Sn_{A1} and Sn_{A1}' are identical only if their segment index numbers are the same. If this condition is met, i.e., n=n', the searching process only include the two-element combinations instead of permutations between Sn_{A1} and Sn_{A1}'. The procedure above is based on Western et al. 2001 to generate the pairs between Z and G: one location is the cells with value "high" in Array 1 (1s, in Z) and the other location is the cells with value "high" or "low" in Array 1 or Array 2 (1s and 0s, in G). In this paper, we separate G into Z and G-Z because the redundancy in permutation can be eliminated between two segments Sn_{A1} and Sn_{A1}' in Z. The goal of this step is to collect all the possible random and connected pairs and their corresponding lag-distances. The lag-distances are computed as Euclidean distances using MXNet hypotenuse function "broadcast_hypot" with:

$$H = hypot\big((x_{a1} - x_{a2}), (y_{a1} - y_{a2})\big) \tag{3}$$

which is equivalent to $H = \sqrt{(x_{a1} - x_{a2})^2 + (y_{a1} - y_{a2})^2}$.

where x_{a1} and y_{a1} are row# and column# of Sn_{A1}; and x_{a2}, y_{a2} are row# and column# of Sm_{A2} (or Sn_{A1}'). Note that we only use the relative scale and distance in the pairing and computation process based on the 2D array coordinates. In both NumPy and MXNet, the arrays being vectorized need to have the same shape (lengths in each dimension) on at least one axis (along a particular dimension). Therefore, to ensure that the number of elements for at least one axis is equal, we expanded the shape of Sn_{A1} and Sm_{A2} by adding another axis of size 1. Assume that the shape of SnA1 and SmA2 are (n,1) and (m,1), the broadcasted array H includes all the possible distances computed based on the paired locations, in terms of all the possible permutations (combinations) for the elements between Sn_{A1} and Sm_{A2}. Then we collapse (flatten) H to the first axis, i.e., convert H to a one dimensional array of shape (n*m,), so that all the possible distance h_1, h_2,..h_{n*m} ($h_{n*n'}$) between the paired locations can be assigned into different range bins. We repeat the vectorization process above between two segments from Array 1 and Array 2 (or Array 1) respectively until all the segments of Array 1 and Array 2 have been included. This is achieved by generating the permutations between the indexed segments of Array 1 and Array 2, i.e., we match two segments from Array 1 and Array 2 as 1&1, 1&2, 2&1, 1&3, 3&1, …, n&m, where n and m are the total number of segments in Array 1 and Array 2. The total counts of collected lag-distances tc_G and the mean of lag-distances $ld_{G,mean}$ in each range bin is recorded. The above procedure extracts all the possible pairs of cells in the patterns in region G for each lag-distance range bin, regardless of the connectedness. In each iteration between Sn_{A1} and Sm_{A2}, the accumulated total lag-distance $ld_G = tc_G * ld_{G,mean}$ is also computed for each range bin.

Similarly, we slice Array 1_t based on *broadcdp* within each spatial clustered region where the elements equal to 1, 2, … and t (t is the max index number assigned for the clustered regions) and generate n segments for each region. These segments are indexed and the permutations between any two segments are generated, i.e., 1&2, 1&3, 1&4, 2&3, …, a&a, where n is the total number of segments in Array 1_i. Then, we determine if the index number of the two segments are equal, e.g., 1&1, 2&2, …, a&a. If this condition is met, permutations are reduced to combinations. This is achieved by sequentially exclude the previously paired elements from further pairing between the two segments. Otherwise, the element-based broadcasting is still permutations, the same as the segment pairing process between Array 1 and Array 2 above. We iterate this starting from spatial

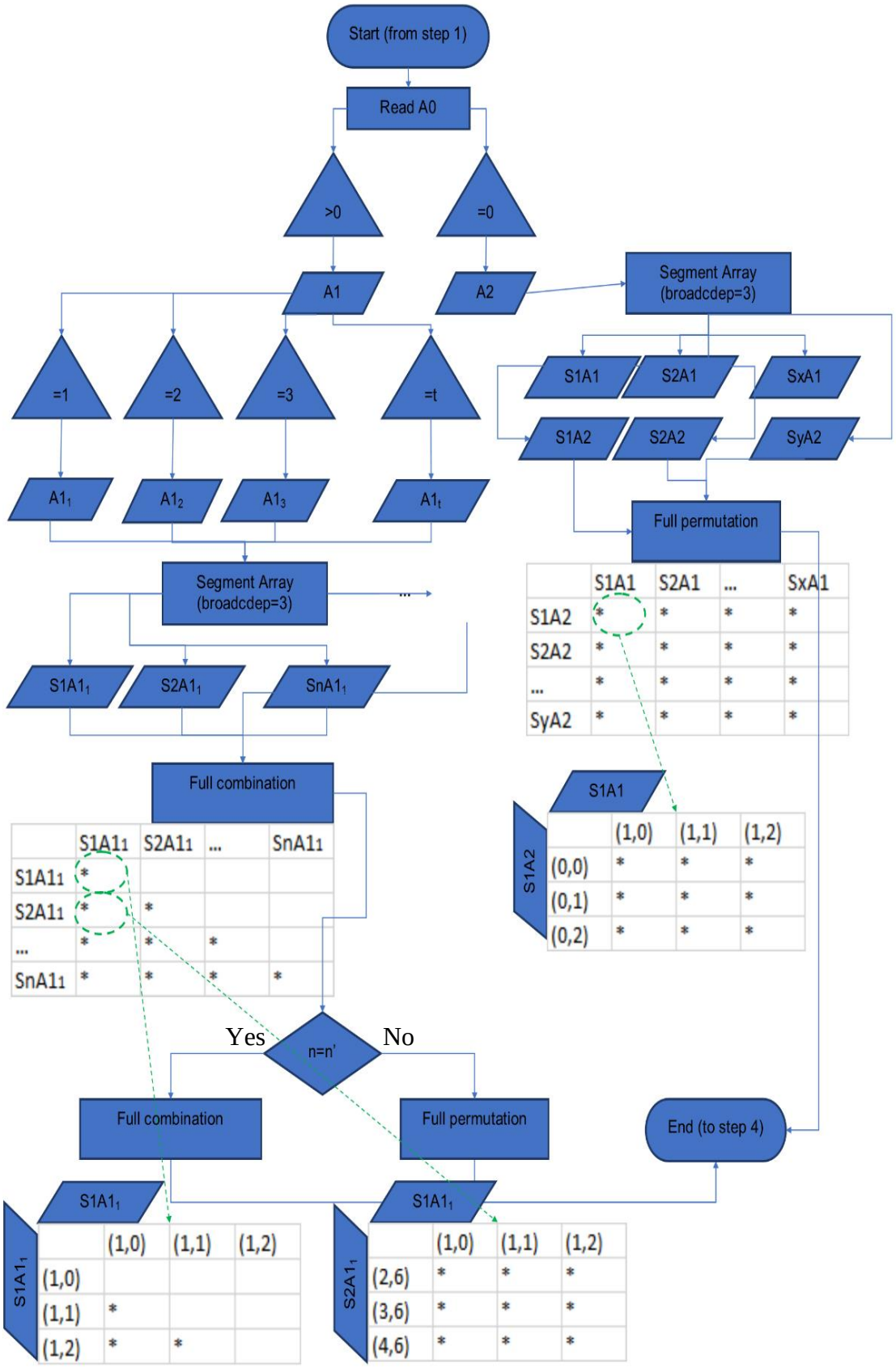
clustered region 1 until t has been reached and compute the combinations between two segments from the same clustered region t. The total counts of lag-distances tc_Z and the mean lag-distances between connected pairs $ld_{z,mean}$ in each range bin is collected and returned. The above procedure extracts the *connected* pairs labeled "high" in each range bin, i.e., the pairs are in the same spatial clustered region (with the same region index number).

Now, this step completes the collection of all the possible lag-distances and the corresponding counts in each range bin. The detailed illustration for the entire array manipulation process, including the pairing between two segments and between elements of paired segments in this step, can be found in Fig. 2.4.

2.4.4 Computing the connectivity function and OMNI

In each range bin, we assign the aggregated values for the total number of pairs (regardless of their connectedness) and the accumulated total lag distances in each range bin as $TC_G = \sum_1^{n\times m} tc_G$ and $LD_G = \sum_1^{n\times m} ld_{G,mean}$, where n and m are the total number of segments in Array 1 and Array 2. Similarly, for each range bin, we assign the total number of connected pairs as $TC_Z = \sum_1^{t\times a} tc_z$, where a is the total number of segments in Array 1_t, and t is the total number of spatial clustered regions. The probability of the connectedness in each range bin is calculated as $P = TC_Z/TC_G$ and the average lag-distance for each range bin is calculated as $l_{mean}=LD_G/TC_G$. Then, P is plotted against the average lag distance for each range bin and connected as a line graph (Fig. 2.1d). The area under the curve is the integral of the connectivity function: OMNI. We treat the shape under the connectivity function as a series of trapezoids which have their parallel sides touching each other and the total area of these trapezoids is OMNI (Fig. 2.1d).

Figure 2.4 Programming flow chart illustrates the computational steps for array manipulation (step 2 and 3 in OMNI), based on the flow pattern in Figure 1c as an example. Array 0 (Figure 1c): 2D array of shape 15x15. Array 1 (Region Z): Cells of Array 0 where > 0. Array 2 (Region G-Z): Cells of Array 0 where =0. t is the maximum index number of the clustered regions. In this example, t=4. Stars (*) indicate the paired tuples for computational operator (for Euclidean distance).


Start (from step 1)
Read A0
>0
=0
A1
A2
Segment Array (broadcdep=3)
=1
=2
=3
=t
S1A1
S2A1
SxA1
S1A2
S2A2
SyA2
A1₁
A1₂
A1₃
A1t
Full permutation
Segment Array (broadcdep=3)
S1A1₁
S2A1₁
SnA1₁
Full combination
Yes
n=n'
No
Full combination
Full permutation
End (to step 4)
S1A1
S1A1₁

2.5 CARD

CARD is computed by appending a cardinal direction indicator for the connected paired cells in the step of calculating mean lag-distances of connected pairs. The direction is computed by the slope of a line connecting the paired locations (Fig. 2.5) and then converted to degrees [-90,90] based on the inverse tangent function. To remove negative values, we added 90 degrees to all the results so the range becomes [0,180]. Note that the degree values should be interpreted as the counterclockwise deviation from the N-S direction and 0 degrees is equivalent to 180 degrees (Fig. 2.5). We computed the average cardinal directions in degrees for all the *connected* paired locations and then categorized the mean value of these directions for each corresponding lag-distance range bin as the general direction trend, into the 16 cardinal/ordinal/secondary intercardinal directions. Therefore, the directional trend for the connected pairs in each lag distances range bin can be plotted in addition to P in connectivity function *τ(h)*. All the tabulated connectivity function values τ(h), i.e., average lag-distances, probability of connectivity, the cardinal direction for each range bin, OMNI and computational time are written to .csv files.

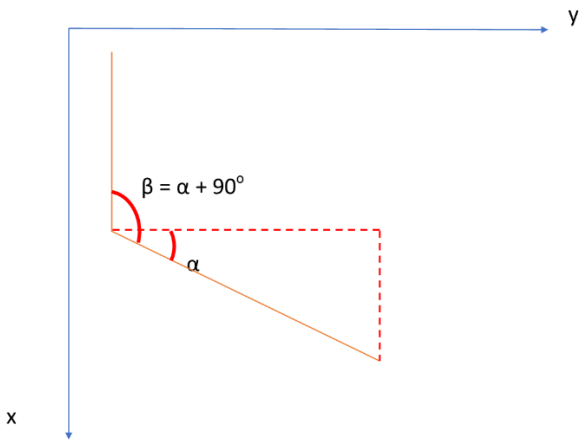


Figure 2.5 Directions and orientation of CARD in the array coordinate. The connected direction between paired two points: A and B is calculated by $\alpha = arctan\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$. β is the output direction rotated counterclockwise for 90 degree.

2.6 TOPO

TOPO follows the same procedures of OMNI except that before it proceeds to step 3, a flow pathway verification step is added to ensure that the connected pathways are topographically valid, i.e., the connected pathways between any pair of locations are only counted as connected if they

follow the downslope flow routing directions on the topography. Thus, TOPO requires a DEM representing the terrain topography of the same spatial domain and at the same projected coordinate system. If the trajectories of the linked pathways between the connected pairs on the flow pattern (generated from Step 3) completely follow (overlap) any of the topographically determined flow pathways without interruption, then we define these pairs as being topographically connected. Since this step is between step 2 and step 3 in computing OMNI, it is named step 2.5.

2.6.1 Step 2.5: Validating the topographic determined flow pathways

The DEM in raster format is converted to a MXNet GPU array which is named Array DEM. The flow direction grid including the downslope directions for each cell in Array DEM are generated based on the method of Seibert and McGlynn, (2007) and stored in an indicator grid Array d. We use MD∞ instead of D∞ because its workflow is more suitable for adopting the convolution operator. To increase the computational efficiency, we limit the tracing of downslope flow accumulation areas within each spatial clustered region in Array 1_t only. The same locations of Array 1_t and Array DEM are identified based on their spatial overlay in the same projected coordinate system. The lag distances are computed similar as OMNI except that one cell of the pairs is always fixed and is the cell where the downslope tracing starts. Flats and the lowest point(s) in surface depressions are identified simultaneously with computation of flow direction grid based on MD∞. Then, the complete spatial coverage for each surface depressions is delineated using an upslope contribution area searching routine. Note that large GPU memory may be required to generate the flow direction grid and the flats/surface depressions grid depending on the spatial scale and resolution of the DEM. We precompute these grids based on the DEM data using a GPU card with sufficient amount of memory, e.g., from a server platform or cloud environment, and save them on disk before precedes to further steps. The detailed programming procedure for TOPO, including the modified CARD, is described below.

First, the flow direction grid and computing downslope accumulation areas in each spatially clustered region are generated. The DEM raster file of the same spatial domain as the flow pattern is read into memory as a NumPy array using the Gdal geospatial library and converted to a MXNet GPU array: Array DEM. Two additional GPU arrays are also generated to store the value of

Cartesian coordinates of the Array DEM as Array X (along rows) and Array Y (along columns). Then, for each cell M, eight triangular facets based on the centers of eight neighboring cells are defined, starting from E (0 degree) and rotating counterclockwise back to E (360 degree, see Fig. 2.6a), the same as in D∞ and MD∞ (Tarboton, 1997; Seibert and McGlynn, 2007). The eight facets and neighboring cells are indexed by eight integer numbers in a counterclockwise sequence as facet 0 to facet 7 and neighbor 0 to neighbor 7 (Fig. 2.6b). Three empty Python lists of length 8 are initiated to store the Arrays of difference between the values of centered cells (M) and two contiguous neighboring cells for each of the eight triangular facets, e.g., P_1 and P_2 in facet 2, based on the Array X, Array Y and Array DEM. We name the arrays d_{xi}, d_{yi} and d_{zi}, where i is the neighboring cell index number from 0 to 7. For example, d_{x2} indicates the difference of the x axis for the neighboring cell to the east: $d_{x2}=x_M-x_{P1}$ (Fig. 2.6b).

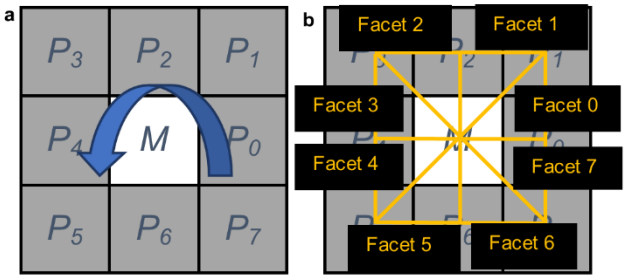


Figure 2.6 a) The definition of connected cells included from the eight neighboring cells of M. In this paper, we define the cells can be connected from all eight neighboring cells, i.e., shaded cells P_0, P_1 ... P_7, in a counterclockwise sequence starting from E. b) Definition of eight triangular facets in a counterclockwise sequence starting from E, represented by the yellow triangles. (The facet index is indicated in yellow text with black background).

Instead of looping through each cell of M in Array X, Array Y and Array Z to compute the d_{xi}, d_{yi} and d_{zi}, we use the neutral network convolutions based on kernel weighted operators in MXNet. Convolutions are often used as an efficient and optimized approach in machine learning algorithms for image processing of large amount of data. Convolution operator is based on a 3x3 kernel window, i.e., a square of total 9 cells of specific configuration, as a "weight" and implements the computation procedure for each cell of Array X, Array Y and Array DEM from the top left to the lower right (except the cells at the edges). We designed a series of 8 kernel structures and each one corresponds to only one neighboring cell (Table 2.3). For example, for weighted kernel 0, it computes the d_{xi} for the cell on the east (Fig. 2.7). The values of the d_{xi} is computed as 0x0+0x-1+0x0+1x0+1x1+1x0+2x0+2x0+2x0=1.

Table 2.3 Weighting kernels used in MXNet GPU array convolution operator. Note that the brackets represent each row of the 9x9 kernel window. Each column of the kernel window is separated by comma (,).

Weighting kernel series	Corresponding Neighboring Cell	Kernel window structure
Weight[0]	E	[0, 0, 0] [0, 1, -1] [0, 0, 0]
Weight[1]	N-E	[0, 0, -1] [0, 1, 0] [0, 0, 0]
Weight[2]	N	[0, -1, 0] [0, 1, 0] [0, 0, 0]
Weight[3]	N-W	[-1, 0, 0] [0, 1, 0] [0, 0, 0]
Weight[4]	W	[0, 0, 0] [-1, 1, 0] [0, 0, 0]
Weight[5]	S-W	[0, 0, 0] [0, 1, 0] [-1, 0, 0]
Weight[6]	S	[0, 0, 0] [0, 1, 0] [0, -1, 0]
Weight[7]	S-E	[0, 0, 0] [0, 1, 0] [0, 0, -1]

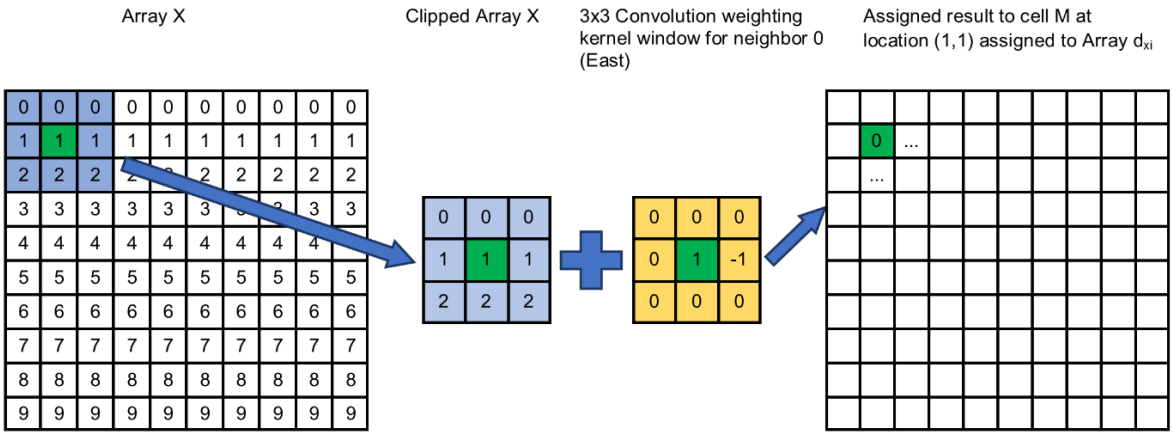


Figure 2.7 Example of convolution operator in MXNet for neighbor 0 (East) at location M (1,1), in green. The computation procedure is 0×0+0×-1+0×0+1×0+1×1+1×0+2×0+2×0+2×0=0. This generates the result at the centered cell M and recorded in a separated array as dxi

A similar convolution procedure is implemented on Array Y and Array DEM and the results are stored at each cell locations in Array d_{xi}, Array d_{yi} and Array d_{zi}. Then, for each triangular facet, we compute the normal vector between the centered cell M and the two contiguous neighboring cells as defined in Seibert and McGlynn, 2007:

$$n_j = \begin{bmatrix} n_{xj} \\ n_{yj} \\ n_{zj} \end{bmatrix} = \begin{bmatrix} dz_i dy_{i+1} - dz_{i+1} dy_i \\ dz_i dx_{i+1} - dz_{i+1} dx_i \\ dy_i dx_{i+1} - dy_{i+1} dz_i \end{bmatrix} \quad (4)$$

where j represents the facet index number from 0 to 7. n_{xj}, n_{yj} and n_{zj} are stored separately in three GPU arrays. Next, the flow directions and the slopes for each triangular fact are computed based on n_{xj}, n_{yj} and n_{zj} and stored in GPU array d_j and s_j for each centered cell M, as in Seibert and McGlynn, 2007:

$$d_j = \begin{cases} 0 & n_{xj} = 0 \; and \; n_{yj} \geq 0 \\ \pi & n_{xj} = 0 \; and \; n_{yj} \leq 0 \\ \frac{\pi}{2} - arctan\left(\frac{n_{yj}}{n_{xj}}\right) & n_{xj} > 0 \\ \frac{3\pi}{2} - arctan\left(\frac{n_{yj}}{n_{xj}}\right) & n_{xj} < 0 \end{cases} \quad (5)$$

$$s_j = -tan\left(arccos\left(\frac{n_{zj}}{\sqrt{n_{xj}^2 + n_{yj}^2 + n_{zj}^2}}\right)\right) \quad (6)$$

Note that the convolutions cannot apply to the edge of Array X, Y and DEM, i.e., the arrays d_j and s_j have the dimension of only that of Array X, Y and DEM minus 2. We extend the shape of Array X, Y and DEM by appending zeros to the edges so the dimensions of d_j and s_j are consistent with Array X, Y and DEM. This implicitly assumes that the flow directions on the edge of the DEM are always out of the domain. The edge appending can be achieved by using a "pad" parameter in the convolutions operator. Similar to Seibert and McGlynn 2007, we made further modifications for these cases when the flow direction of the facet computed, d_j, is not within 45 degrees of the corresponding facet or both the elevation of neighboring cells on the facet (e.g., p1 and p2) are higher than the centered cell M. For the former case, the flow direction d_j is assigned to the largest gradient of the two from M toward p_1 or p_2, i.e., the maximum of dz_i and dz_{i+1}, and the slope s_j is assigned as the corresponding slope between M and p_1 or p_2. For the later case, both of the directions are eliminated and d_j and s_j at the corresponding locations of centered cells M are converted to -999, which represents "NoData". The last step of updating d_j is to ensure that for the downslope directions directly pointing towards the centers neighboring cells, i.e., $\pi/4$ or $(j+1)\pi/4$, only the adjacent facets with the same directions, i.e., $d_j=d_{j+1}$ or $d_j=d_{j-1}$ are maintained. Otherwise these values in d_j are eliminated and converted to -999 as "NoData".

After the flow direction grids are computed, downslope flow accumulation areas are delineated within each of the spatial clustered region of Array 1_t. One cell within the spatially-clustered region is assigned as the starting cell k_n and the downslope accumulation areas C_n are traced based on the flow direction indicated by d_j. The cells in C_n are marked as 1 in a MXNet GPU Array Flowdown (anywhere else are 0s). Array Flowdown is sent back to CPU as NumPy array and the coordinates of the cells where the value are 1s are extracted to compute the connectivity function in a similar

manner as described in Step 4. Note that for TOPO, one cell in the connectivity function computation is *always* the starting location. Thus, no combinations between two segments of Array 1_t are searched but we simply match the starting cells and all the cells in one segment of Array 1_t. This process repeats for all the cells in the same clustered region x. The rational is that because any pair of cells within the same spatial clustered region are connected, any other locations within C_n paired with the starting cell k_n and also within the same spatial clustered region are also topographically connected. Therefore, the algorithm we designed can efficiently count the pairs and compute the lag distances between k_n and any other locations within C_n by limiting the pairing process within the spatial clustered regions. Note that we do not apply any weight on the flow directions because MD∞ and D∞ are identical in most cases.

To delineate C_n using a memory and stack efficient way in TOPO, we assume that the traverse of flow cascading downslope in C_n on the topography is equivalent to the "tree branches" rooted at the starting (upslope) location k_n. This allows us to use a parallel, non-recursive strategy, i.e., the "breadth first search" (BFS) method. BFS searches all the neighboring cells at the current depth (the closest neighbors at one-cell distance), and then the neighbors at a two-cell distance, until all the tree branches (downslope cells) are visited (Fig. 2.8). This is implemented by creating a stack

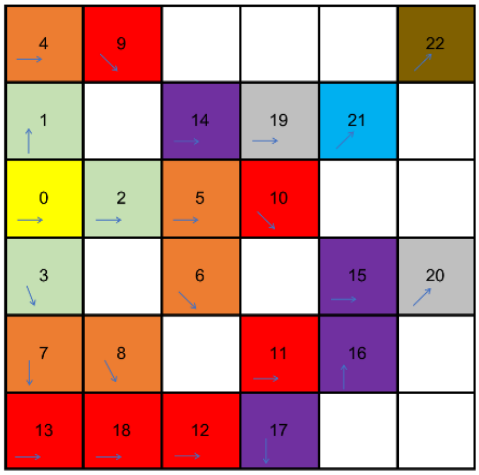


Figure 2.8 Example of the "Breadth first search" method for identifying flow downslope traversed cells. The blue arrows represent the computed flow directions in dj for each cell. The flow directions for blank cells are undetermined. We start from location "0" and trace flow downslope. The numbers are the sequence of identified downslope cells. Each color represents one step: yellow – step 1, grass green – step 2, orange – step 3, red – step 4, purple –step 5, gray – step 6, blue – step 7 and dark gold –step 8.

control list and a marker list. The stack control list is used to append new locations that need to be processed to search for downslope cells, i.e., in the current eight neighbors, and to remove the locations that have already been processed. It is also used to indicate the termination of the iteration when no cell (location) remains in the stack list. The marker list is used to exclude the locations already visited in the previous iterations and avoid repeated searching.

Because TOPO includes a full plane direction from the starting cell K_n towards the other cell in a pair instead of a half plane and omnidirectional direction connecting a pair, we modified the flow direction indicator of CARD to better differentiate the directions on the connected pathways and reflect the impedance of topography. Therefore, the range of CARD in TOPO is [0,360) degrees and we assume that the vector pointing toward East is zero degree. In the algorithm implementation, the flow direction indicator CARD in TOPO is computed as:

$$d_{dir} = \begin{cases} 0 & d_{xi} = 0 \; and \; d_{yi} \geq 0 \\ \pi & d_{xi} = 0 \; and \; d_{yi} \leq 0 \\ \dfrac{3\pi}{2} - arctan\left(\dfrac{d_{yi}}{d_{xi}}\right) & n_{xi} > 0 \\ \dfrac{\pi}{2} - arctan\left(\dfrac{d_{yi}}{d_{xi}}\right) & n_{xi} < 0 \end{cases} \quad (7)$$

Then, the histogram of CARD values is computed for each lag-distance range bin based on the 16 cardinal/ordinal/secondary intercardinal directions.

2.6.2 Dealing with surface depressions and flat areas in the DEM

MD∞ cannot derive correct flow directions in surface depressions (pits) or flat areas (areas of equal elevation), but it can help identify and label cells in these areas for further processing. The TOPO algorithm is designed compute the connectivity states based on existing flow patterns, either from observations or modeling, instead of physically simulating water mass, momentum and energy transfer. Therefore, to simplify the computation of TOPO, we made two assumptions: 1) the water flow has sufficient mass, momentum and energy to transfer through these surface depressions and flat areas if their pathways are already on the flow pattern; 2) the starting point K_n in TOPO can only connect cells within C_n of same or lower elevations. We created a two-step process to delineate these surface depression catchments based on the parameters obtained in MD∞. First, the lowest point(s) of surface depressions are identified in these cells based on the total counts of

negative and zero values in z component of the normal vector n_{zj}, which represents the elevation difference between the eight neighboring cells P_0, P_1 and P_i (i=7) and the centered cell M. Positive means P_i is higher than M, negative means P_i is lower than M and zero means P_i is the same elevation as M. Therefore, if the sum of negative counts of n_{zj} is 8, i.e., no downslope directions, then we determine that the cell(s) are the lowest point(s) in the surface depressions. If the sum of zero counts of n_{zj} is ≥1, then we determine that the cell(s) are the flat areas (including these in depression catchments). To differentiate the flat areas which may be inside a depression catchment, these identified flat areas are grouped into separated spatial regions based on their elevations and their neighboring connectedness, i.e., a flat area "patch" includes cells of the same elevations and also does not overlap with any other patches. If all the cells in a particular flat area patch do not intercept any a positive value in $n_{zj,}$ then this patch is in a depression catchment. Next, the boundary of the surface depressions is delineated by using an upslope direction searching method, i.e., reversely finding the upslope cells of these lowest point(s) or flats based on n_{zj}. We use the BFS, similar as that for downslope direction computation, but is a complete upslope "climbing" process without excluding any "ineligible" upslope cells. This step stops when there is no more upslope direction exist. The edges of the aggregated upslope contributing area for all the lowest point(s) is the preliminary boundary of the surface depression catchments. We name these boundary cells as b_i. Then if cell(s) of lower elevations (p_i) exist outside of the boundary, the pour points are determined as the cell(s) of lowest elevations of b_i that are adjacent to p_i. If cell(s) of lower elevations (p_i) only exist inside of the boundary, the pour points are determined as the cell(s) of lowest elevations of b_i. Each delineated surface depression catchment is indexed and accompanied by its spatial coverage and pour point locations in two Python lists.

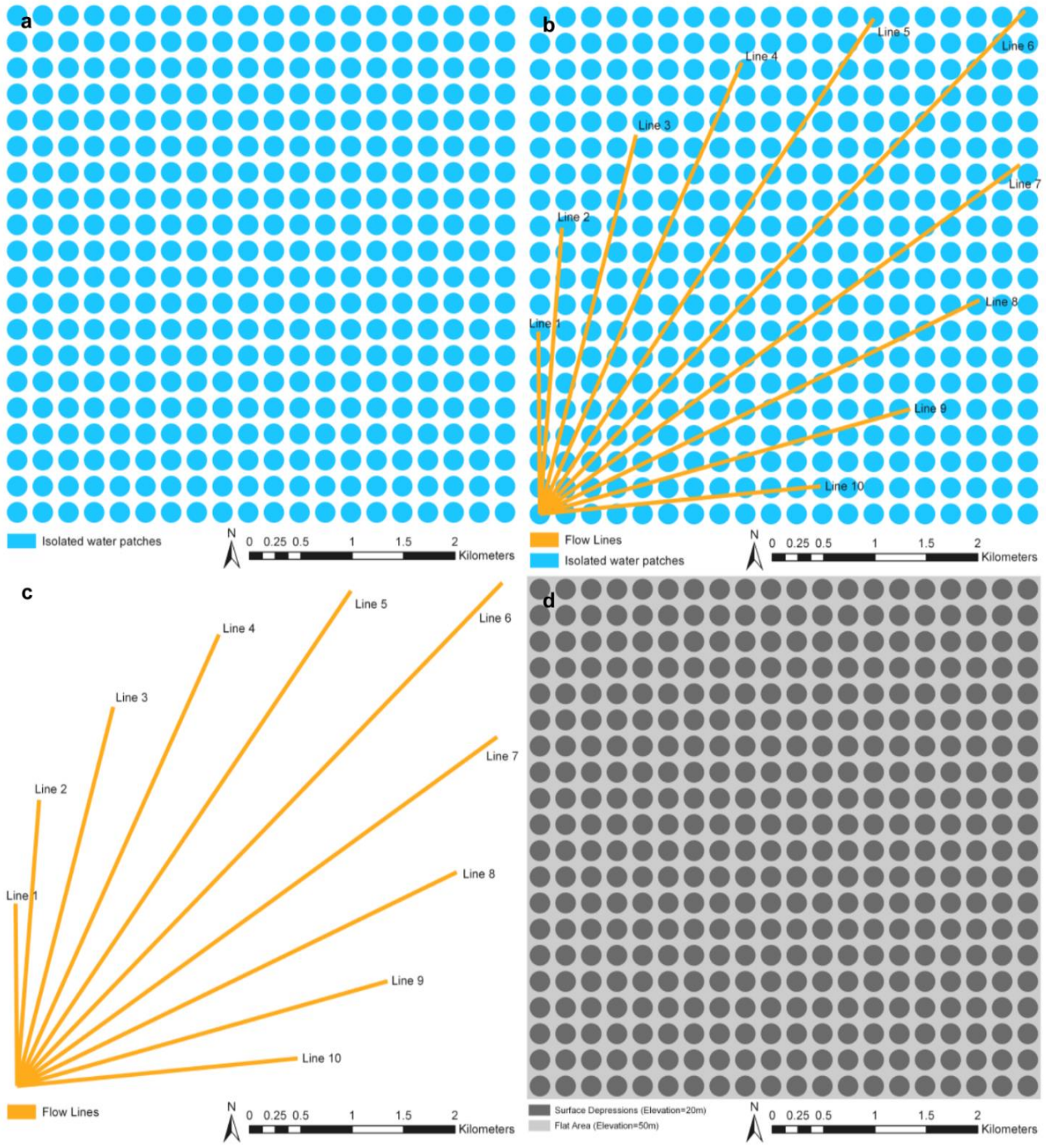
Next, in each spatial clustered region of Array 1_t, we search the downslope accumulation area C_n (store in array Flowdown and labeled with 1) from the starting point k_n. If any surface depression(s) catchment intercepts with Flowdown where the cell values are 1s, a further step is needed to determine if the surface depression(s) can or cannot block the further downslope flow cascading for K_n. This is determined by comparing the elevation of pour point(s) hb_i and the start point hk_n. If $hk_n \geq hb_{i,}$ then the downslope flow cascading continues until intercept with other depression catchment(s) or the edge of the spatial clustered region. Flow connects from K_n to the entire depression area and then continues to cascade downslope from the pour point. Otherwise ($hk_n <$

hb_i), the downslope accumulation for the starting point k_n ceases in the depression catchment area. Finally, the downslope accumulation areas of the pour point(s) are also included in C_n and the cells of array Flowdown in these areas are updated to 1s. Once this downslope searching procedure is completed, the starting point k_n is paired with every cell in array Flowdown where the value is > 0 for the lag distance computation. Therefore, the topographic determined pathways can represent the "on" or "off" states for the connectivity between any pair of cells based on the large-scale gradient (see illustrations in Fig. 2.2). This approach allows us to flexibly determine the aggregated effects, i.e., the magnitude of resistance, of surface depressions on connected pathway(s).

2.7 Test results and discussions

The new OMNI, CARD and TOPO connectivity algorithm has been tested on a hypothetical landscape with spatial resolutions of 2m and 5m. The "hydrologically relevant" data used in the test are the pattern of flow depths in a landscape domain of size 5,000 m x 5,000 m. The total number of cells for the 2m and 5m resolution domain are 6,125,625 and 980,100 respectively. Three hypothetical flow patterns at 2m and 5m resolutions have been tested to demonstrate the robustness and sensitivity of the OMNI, CARD and TOPO algorithm (Fig. 2.9a-c). Each pattern represents a time snapshot of overland flow for this domain. The total counts of "high" cells for pattern 1, pattern 2 and pattern 3 at 5 m resolution are 505,600, 537,606 and 66,079 respectively, and at 2 m resolution are 3,142,400, 3,343,734 and 412,942 respectively. The rests are the "low" cells for each pattern. In pattern 1, the "high" cells are located in circular patches (diameter of 200m) and are uniformly distributed in the domain but isolated from each other (Fig. 2.9a), representing a landscape with internal drainage catchments. In pattern 3, we use a series of 10 yellow lines representing a landscape with connected overland flow pathways. Each line is at different directions, i.e., of bearing angels from 0 to 90 degrees from N-S direction, and lag-distances, i. e., from 1745 m (Line 1 and Line 10 at 0 or 85 degrees) to 6745 m (Line 6 at 45 degrees) (Fig. 2.9c and Table 2.4). All of these flow lines connected at the cell with coordinate of (4900,100) at the SW corner of the domain. These flow lines are used to demonstrate the scenario of fully connected pattern as the comparison to pattern 1. In pattern 2, these flow lines connect the isolated patches at different directions, representing a landscape with connected flow pathways between the isolated "high" regions (patches). We use two colors (blue and yellow) in pattern 1 and pattern 2 to label the regions of isolated drainage sinks and the connected flow pathways.

Figure 2.9 Test Patterns and DEM for connectivity metric OMNI, CARD and TOPO in a square shaped domain of 5000m x 5000m at 5m and 2m resolution. The blue circle (in a and b) and the yellow lines (in b and c) represent “high” regions (Z) of isolated water patches and the connectivity pathways in between. The diameter of each circle is 200m and they are uniformly distributed in the domain. The origin of these flow lines is at the location x=100 and y=100. The length of each line and bearing (from N-S direction) is listed in Table 2. DEM used for testing TOPO is presented in d. Note that in a-c, “low” regions are the areas in blank (white).


a
Isolated water patches
N
0 0.25 0.5 1 1.5 2
Kilometers
b
Line 1
Line 2
Line 3
Line 4
Line 5
Line 6
Line 7
Line 8
Line 9
Line 10
Flow Lines
Isolated water patches
N
0 0.25 0.5 1 1.5 2
Kilometers
c
Line 1
Line 2
Line 3
Line 4
Line 5
Line 6
Line 7
Line 8
Line 9
Line 10
Flow Lines
N
0 0.25 0.5 1 1.5 2
Kilometers
d
Surface Depressions (Elevation=20m)
Flat Area (Elevation=50m)
N
0 0.25 0.5 1 1.5 2
Kilometers

In Figures 2.9a to c, blue and yellow regions are the “high” areas which indicate that flow depth is above a predefined threshold, e.g., 0.1m, and the white regions indicate areas of “low” areas. The elevations in the DEM for the landscape are presented in Figure 2.9d. Note that to demonstrate the capacity of TOPO in dealing with surface depressions and flat areas, the areas of blue circular patches in pattern 1 or 2 are located in the depression catchments of cylinder shapes with bottom elevation of 20m. Areas of white in pattern 1 or 2 (“low” cells) are contiguous flat regions in the DEM with elevation of 50m. Note that no pour point exists outside these depression catchments.

The 5m test results of OMNI for pattern 1 (isolated circular patches) show that the lag-distances for connected pairs are limited to 350 m, which only includes the pairs of cells within the patches (diameter 200m, and blank space or “low” region between patches of 50m) (Table 2.5). In contrast, the 5m test results for pattern 2 (patches connected with flow lines) show the lag-distances for connected pairs from 350m to 7,000m. In each lag-distance range, the connectivity probability P is between 3% to 25% (Table 2.6). P reaches the maximum of 24.7% at 6,650-7,000m and the minimum of 3.3% at 3,850-4,550m. This is because the length of Line 10 is at 6,750m which connects the most of the cells in “high”, and no flow line is between 3,850 and 4,550m. Also, P slightly decrease and then increases with the lag-distance. This is because more isolated patches are being connected at larger lag-distances than the smaller lag-distances by these flow lines, e.g., line 6 connects 20 isolated patches while line 1 only connects 6. Meanwhile, for the lag-distance in between the two extremes, i.e., lag-distance range bin 0 and 20, the proportion of white or “low” areas included in the computation also increases and reaches the maximum at around 3,850-4,550m (bin 12 to 13). This is also the lag-distance range with the lowest probability of connectedness. In addition, P for pattern 2 is always higher than that for pattern 3 because the yellow flow lines connect the isolated water patches approximately in a uniform manner at different lag distances. Similarly, the results of CARD for pattern 1 is also limited to the 350 m and include only one radius of the isolated patches, as expected (Table 2.5). An exact 90-degree means that the connected direction is the average between uniform distribution of [0-180], which indicates no particular direction can be identified. In other lag-distance ranges, the directions remain exactly zeros (the initial values). The average flow directions for pattern 2 starts from 86 at a lag distance of 350 and decreases to 45 at a lag distance of 7000m (Table 2.6). This is because as the lag-distances increase, directions of the flow lines connecting the patches shift from N-S or

E-W (86 or 0) directions to N-E direction (45 degrees) (line 6). A similar trend can be found in pattern 3, where the directions shift from 77 to 45 degrees (Table 2.7). The 2 m resolution patterns provide similar results to the 5 m resolution patterns (Table 2.5-2.7). OMNI captures the increasing connectivity trend of the patterns: 2, 47 and 84 for pattern 1, 3 and 2 respectively (Table 2.8).

TOPO and CARD algorithm for pattern 2 at 5m resolution based on the hypothetical DEM has been tested for the landscape domain (Fig. 2.9d) as the demonstration for its capacity of dealing with flat areas and surface depressions, and its sensitivity for topographic determined flow directions. TOPO for pattern 2 is 47 which is slightly larger than half of OMNI for pattern 2 (84) (Table 2.8). This indicates that the surface depressions significantly reduce the connectivity for pattern 2. The connectivity function $\tau(h)$ of TOPO is always lower than OMNI (Table 2.6 and 2.9), which indicates the impedance effects of the surface depressions on connected flow pathways. The magnitude of the difference between $\tau(h)$ of TOPO and OMNI decreases as the lag-distance increases. This is because surface depressions can only impede the flow pathways within these isolated circular areas. The largest lag-distance range of 6650-7000m identifies the existence and the correct spatial directions of the longest flow line: Line 6. This is confirmed by the non-zero $\tau(h)$ values in the range of 6650-7000m and the NE and SW directions (Table 2.9). Note that $\tau(h)$ for NE is always lower than SW for the lag-distances above 4,900m. This may relate to the non-uniform weight of flow lines pattern, because the origin of all these lines are located at SW corner of the domain. Relatively fewer number of pairs can have lag-distance above 4,900m excepts those on flow line 5, 6 and 7 with one cell of the pairs at the NE part of the domain. For these pairs, the major trend of the connectivity directions is towards SW instead of NE based on how they connect on line 5, 6 or 7 and other shorter flow lines. $\tau(h)$ for NE becomes approximately equal as SW in the range bin 4,550 to 4,900m, and then start to lower than SW below lag-distance 4,550m and continue to decrease. This is because pairs with lag-distances lower than 4,550m have at least one cell located at SW part of the domain, i.e., the probability of one cell located at SW increase with the decrease of lag-distance. Also, the range bin 5600-5,950m and 4,550-4,900m detect the existence of flow line 5 & 7 and 4 & 8, for both their lengths and connectivity directions. For example, P shifts from 0 to 0.003 at range bin 5600-5,950m at ENE/WSW and NNE/SSW direction respectively. Other shorter flow lines, i.e., line 1, 2, 3, 9 and 10, are not so obviously identified because their lengths are not exclusively represented.

Table 2.4 Lengths and the bearings for the flowlines in pattern 3

Line index	Distance (m)	Bearing from N-S direction (degree)	Full Plane direction from E (degree) counterclockwise
Line 1	1745	0	90 & 270
Line 2	2745	5	85 & 275
Line 3	3745	15	75 & 285
Line 4	4745	25	65 & 295
Line 5	5745	35	55 & 305
Line 6	6745	45	45 & 315
Line 7	5745	55	35 & 325
Line 8	4745	65	25 & 335
Line 9	3745	75	15 & 345
Line 10	2745	85	5 & 355

Table 2.5 Connectivity function curve tabular results for test pattern 1.

Range bin index	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Lag-distance ranges (m)	0-350	350-700	700-1050	1050-1400	1400-1750	1750-2100	2100-2450	2450-2800	2800-3150	3150-3500	3500-3850	3850-4200	4200-4550	4550-4900	4900-5250	5250-5600	5600-5950	5950-6300	6300-6650	6650-7000
Average lag distances (m)	90.874	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Connected Probability 5m Resolution	0.0583	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Connected Probability 2m Resolution	0.0587	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Average connected directions 5m resolution	90	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Average connected directions 2m Resolution	90	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
CARD	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Table 2.6 Connectivity function curve tabular results for test pattern 2.

Range bin index	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Lag-distance ranges (m)	0-350	350-700	700-1050	1050-1400	1400-1750	1750-2100	2100-2450	2450-2800	2800-3150	3150-3500	3500-3850	3850-4200	4200-4550	4550-4900	4900-5250	5250-5600	5600-5950	5950-6300	6300-6650	6650-7000
Average lag distances (m)	188.579	545.115	884.913	1230.955	1577.848	1925.195	2273.381	2620.527	2968.881	3316.578	3664.552	4017.020	4364.724	4712.092	5060.116	5406.587	5751.620	6097.211	6428.302	6717.431
Connected Probability 5mResolution	0.145	0.091	0.090	0.082	0.074	0.069	0.062	0.056	0.049	0.041	0.036	0.033	0.033	0.036	0.053	0.069	0.087	0.118	0.182	0.247
Connected Probability 2m Resolution	0.145	0.091	0.090	0.082	0.074	0.069	0.062	0.056	0.049	0.041	0.036	0.033	0.033	0.036	0.053	0.069	0.087	0.118	0.183	0.246
Average connected directions 5m resolution	86.465	87.878	88.863	85.949	83.998	82.480	79.669	74.937	69.488	61.539	52.988	46.847	44.494	44.942	45.083	45.148	45.378	45.672	45.509	45.112
Average connected directions 2m Resolution	86.468	87.878	88.863	85.950	84.001	82.482	79.672	74.939	69.489	61.542	52.991	46.849	44.501	44.945	45.088	45.148	45.378	45.672	45.509	45.112
CARD	N	N	N	N	N	N	N	NNE	NNE	NNE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE

Table 2.7 Connectivity function curve tabular results for test pattern 3.

Range bin index	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Lag-distance ranges (m)	0-350	350-700	700-1050	1050-1400	1400-1750	1750-2100	2100-2450	2450-2800	2800-3150	3150-3500	3500-3850	3850-4200	4200-4550	4550-4900	4900-5250	5250-5600	5600-5950	5950-6300	6300-6650	6650-7000
Average lag distances (m)	206.909	526.453	858.396	1195.322	1533.114	1870.861	2210.461	2547.272	2886.451	3224.491	3561.706	3903.739	4240.598	4578.255	4919.158	5253.684	5588.488	5932.014	6258.622	6535.795
Connected Probability 5m Resolution	0.089	0.068	0.060	0.052	0.046	0.041	0.038	0.033	0.030	0.026	0.023	0.021	0.021	0.021	0.029	0.035	0.036	0.048	0.084	0.192
Connected Probability 2m Resolution	0.089	0.068	0.060	0.052	0.046	0.041	0.038	0.033	0.030	0.026	0.023	0.021	0.021	0.021	0.029	0.035	0.036	0.048	0.085	0.189
Average connected directions 5m resolution	77.894	87.699	88.125	84.268	81.537	79.623	75.514	70.615	66.052	59.067	53.201	46.529	44.589	44.653	44.589	44.746	44.737	44.930	45.212	45.425
Average connected directions 2m Resolution	77.891	87.679	88.138	84.293	81.555	79.646	75.533	70.634	66.059	59.071	53.198	46.533	44.600	44.671	44.605	44.757	44.739	44.935	45.217	45.425
CARD	N	N	N	N	N	N	NNE	NNE	NNE	NNE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE

Table 2.8 OMNI and TOPO connectivity metrics computation results for test patterns 1-3. Note that for TOP, only pattern 2 at 5 m resolution has been tested.

Pattern and resolution		Computing time (sec)	OMNI	Computing time (sec)	TOPO
Pattern 1	5m resolution	8,555	1.818	--	--
	2m resolution	123,035	1.956	--	--
Pattern 2	5m resolution	7,597	84.577	123,889	48.895
	2m resolution	109,255	84.681	--	--
Pattern 3	5m resolution	6,891	47.400	--	--
	2m resolution	99,171	47.568	--	--

Table 2.9 Topographic determined connectivity function curve tabular results or test pattern 2 (5m resolution). Columns in bold represent the lag-distance range bins which include the lengths for the flow lines. Rows in bold represent the bearings for the flow lines categorized into 16 cardinal/ordinal/secondary intercardinal direction

Range bin index	1	2	3	4	**5**	6	7	**8**	9	10	**11**	12	13	**14**	15	16	**17**	18	19	**20**
Flow Lines in range bin					**1**			**2, 10**			**3, 9**			**4, 8**			**5, 7**			**6**
Lag-distance ranges (m)	0-350	350-700	700-1050	1050-1400	**1400-1750**	1750-2100	2100-2450	**2450-2800**	2800-3150	3150-3500	**3500-3850**	3850-4200	4200-4550	**4550-4900**	4900-5250	5250-5600	**5600-5950**	5950-6300	6300-6650	**6650-7000**
Average lag distances (m)	188.579	545.115	884.913	1230.955	**1577.848**	1925.195	2273.381	**2620.527**	2968.881	3316.578	**3664.552**	4017.020	4364.724	**4712.092**	5060.116	5406.587	**5751.620**	6097.211	6428.302	**6717.431**
E Line 10	0.071	0.075	0.083	0.091	**0.088**	0.091	0.100	**0.083**	0.075	0.070	**0.049**	0.027	0.013	**0.000**	0.000	0.000	**0.000**	0.000	0.000	**0.000**
ENE Line 8,9	**0.034**	**0.080**	**0.087**	**0.094**	**0.098**	**0.096**	**0.108**	**0.117**	**0.119**	**0.148**	**0.156**	**0.156**	**0.165**	**0.126**	**0.071**	**0.033**	**0.003**	**0.000**	**0.000**	**0.000**
NE Line 5,6,7	**0.037**	**0.085**	**0.087**	**0.088**	**0.092**	**0.104**	**0.113**	**0.127**	**0.135**	**0.152**	**0.177**	**0.198**	**0.223**	**0.266**	**0.286**	**0.355**	**0.397**	**0.344**	**0.166**	**0.118**
NNE Line 3,4	**0.031**	**0.078**	**0.086**	**0.087**	**0.094**	**0.096**	**0.100**	**0.113**	**0.131**	**0.129**	**0.132**	**0.127**	**0.115**	**0.086**	**0.060**	**0.027**	**0.003**	**0.000**	**0.000**	**0.000**
N Line 1,2	**0.020**	**0.055**	**0.063**	**0.071**	**0.071**	**0.079**	**0.074**	**0.071**	**0.061**	**0.051**	**0.032**	**0.017**	**0.005**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
NNW	0.019	0.048	0.048	0.047	**0.043**	0.042	0.037	**0.031**	0.020	0.008	**0.003**	0.000	0.000	**0.000**	0.000	0.000	**0.000**	0.000	0.000	**0.000**
NW	0.017	0.044	0.040	0.036	**0.030**	0.024	0.022	**0.015**	0.012	0.006	**0.001**	0.000	0.000	**0.000**	0.000	0.000	**0.000**	0.000	0.000	**0.000**
WNW	0.016	0.040	0.038	0.031	**0.027**	0.026	0.023	**0.019**	0.018	0.010	**0.003**	0.000	0.000	**0.000**	0.000	0.000	**0.000**	0.000	0.000	**0.000**
W Line 10	**0.047**	**0.043**	**0.044**	**0.045**	**0.040**	**0.039**	**0.041**	**0.037**	**0.038**	**0.036**	**0.026**	**0.025**	**0.007**	**0.006**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
WSW Line 8,9	**0.108**	**0.055**	**0.048**	**0.048**	**0.055**	**0.059**	**0.062**	**0.072**	**0.082**	**0.090**	**0.110**	**0.111**	**0.121**	**0.125**	**0.127**	**0.084**	**0.029**	**0.000**	**0.000**	**0.000**
SW Line 5,6,7	**0.104**	**0.060**	**0.054**	**0.051**	**0.056**	**0.066**	**0.072**	**0.081**	**0.097**	**0.124**	**0.162**	**0.196**	**0.212**	**0.250**	**0.320**	**0.415**	**0.550**	**0.656**	**0.834**	**0.882**
SSW Line 3,4	**0.112**	**0.056**	**0.045**	**0.048**	**0.056**	**0.064**	**0.065**	**0.076**	**0.086**	**0.094**	**0.103**	**0.119**	**0.132**	**0.135**	**0.135**	**0.086**	**0.018**	**0.000**	**0.000**	**0.000**
S Line 1,2	**0.100**	**0.066**	**0.065**	**0.066**	**0.064**	**0.049**	**0.051**	**0.049**	**0.045**	**0.041**	**0.031**	**0.023**	**0.008**	**0.006**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
SSE	0.103	0.069	0.067	0.061	**0.057**	0.048	0.032	**0.028**	0.025	0.013	**0.006**	0.001	0.000	**0.000**	0.000	0.000	**0.000**	0.000	0.000	**0.000**
SE	0.094	0.072	0.068	0.061	**0.061**	0.053	0.038	**0.024**	0.017	0.008	**0.002**	0.000	0.000	**0.000**	0.000	0.000	**0.000**	0.000	0.000	**0.000**
ESE	0.105	0.077	0.079	0.073	**0.067**	0.063	0.062	**0.055**	0.040	0.019	**0.007**	0.000	0.000	**0.000**	0.000	0.000	**0.000**	0.000	0.000	**0.000**
All Directions	0.097	0.037	0.035	0.033	**0.029**	0.026	0.024	**0.021**	0.020	0.018	**0.016**	0.016	0.018	**0.022**	0.036	0.052	**0.072**	0.108	0.178	**0.245**
Diff with pattern 2	-0.048	-0.054	-0.055	-0.049	**-0.045**	-0.043	-0.038	**-0.035**	-0.029	-0.023	**-0.020**	-0.017	-0.015	**-0.014**	-0.017	-0.017	**-0.015**	-0.010	-0.004	**-0.002**

The computing time of OMNI for test patterns 1, 2, and 3 were 8,555, 7,597, and 6,891 seconds at 5 m resolution and 123,035, 109,255, and 99,171 seconds at 2m resolution respectively. The computing time of TOPO for test pattern 2 is 123,889 seconds at 5m resolution. TOPO provides a more realistic and robust way of measuring hydrologic pattern connectivity determined by topography, including the complex features such as surface depressions and flat areas. The trade-off is that TOPO takes significant more time to than OMNI. Note that the computational time is not linearly related to the total grid size because it is also significantly affected by the spatial configuration of the pattern in "high" (Z) and "low" (G-Z). Therefore, all three patterns at the same spatial resolution with the same total number of cells, but the computational time for each pattern is different.

2.8 Conclusions

This paper demonstrates an efficient algorithm of computing hydrologic connectivity for large-scale and high-resolution patterns in a reasonable timeframe. The workflow of this algorithm is fully automated and requires minimum user interference. Also, it is easy to implement the code on various platforms and hardware configurations. The probability of connectedness P between any pair of cells can be plotted against the lag-distances as the connectivity function graph as $\tau(h)$ (see example in Fig. 2.1d), which provides a spatially explicit representation of hydrosystem level connectivity status for each time snapshot. We hope this algorithm paves the way of investigating large-scale, e.g., watershed scale, connectivity using high resolution spatial data.

The emergent characters of the watershed hydrologic processes can only be revealed at large scale and depends on the variability of the complex "connectedness" between these small-scale factors spatially and temporally. The landscape factors, e.g., surface depressions, may also be contributing to the shifts in connectivity at watershed scale. These small-scale factors are often considered the source of high uncertainly in the hydrologic models (e.g., Antoine et al., 2009; Bracken et al., 2013). Connectivity at hydrosystem level reflects the complex aggregation of feedback/interaction dynamics between these small-scale factors, but it has been difficult to find a unified indicator to inspect the relationships they have with the physical processes. The new OMNI, TOPO, and CARD algorithm provide a connectivity metric to intuitively and directly compare the aggregated and large-scale effects of small-scale factors contributing to the watershed hydrologic processes

and their interactions with the landscape in a short timeframe. We will implement OMNI, TOPO and CARD connectivity metrics in a real-world watershed and demonstrate the effect of small-scale morphological features, such as surface depressions, on the overland flow and runoff response by the changes of hydrosystem level or large-scale connectivity status, e.g., watershed scale, as a case study.

Although our algorithm makes the computation of large grid connectivity feasible within a reasonable timeframe, a major limitation of the current algorithm is that the computational time is not a simple linear function of the number of cells in the grid. This can reduce computational speed when a complex and high-resolution pattern is being analyzed. Also, this algorithm does not account for the actual water mass transfer between the paired locations. Future work to improve the algorithm will focus on 1) Incorporate water and water mediated substances transfer between any pair of cells as a physically restrictive condition to determine the material connectedness; 2) Parallelization of the algorithm across multiple GPUs to obtain the higher computational speed increases; 3) Incorporate a network distance based lag-distance computation so that the subsurface flow connectivity can also be estimated, where the trajectories of the pathways cannot be measured in a 2D only Euclidean space.

2.9 Code availability and Test Data

OMNI-CARD, and TOPO-CARD algorithm is developed and implemented in MXNet – a machine learning framework with Python binding (Version > 1.2.0). Computer code is under Apache 2.0 license and separated into two scripts: one for OMNI-CARD and one for TOPO-CARD, and can be accessed from the author's GitHub repository at the link below:
https://github.com/codefortheplanet/Connectivity_Algorithm
Instructions regarding the use of the computer code and test datasets can also be found in the repository.

2.10 References

Ali, G. A., & Roy, A. G. (2010). Shopping for hydrologically representative connectivity metrics in a humid temperate forested catchment. *Water Resources Research*, *46*(12). https://doi.org/10.1029/2010WR009442

Allard, D. (1994). Simulating a geological lithofacies with respect to connectivity information using the truncated Gaussian model. In *Geostatistical simulations* (pp. 197–211). Springer.

Anderson, B. G., Rutherfurd, I. D., & Western, A. W. (2006). An analysis of the influence of riparian vegetation on the propagation of flood waves. *Environmental Modelling and Software*, *21*, 897–902. https://doi.org/10.1016/j.envsoft.2005.04.027

Antoine, M., Javaux, M., & Bielders, C. (2009). What indicators can capture runoff-relevant connectivity properties of the micro-topography at the plot scale? *Advances in Water Resources*, *32*(8), 1297–1310. https://doi.org/10.1016/j.advwatres.2009.05.006

Aurousseau, P., Gascuel-Odoux, C., Squividant, H., Trepos, R., Tortrat, F., & Cordier, M. O. (2009). A plot drainage network as a conceptual tool for the spatial representation of surface flow pathways in agricultural catchments. *Computers & Geosciences*, *35*(2), 276–288.

Bracken, L. J., & Croke, J. (2007). The concept of hydrological connectivity and its contribution to understanding runoff-dominated geomorphic systems. *Hydrological Processes*, *21*(13), 1749–1763. https://doi.org/10.1002/hyp.6313

Bracken, L. J., Wainwright, J., Ali, G. A., Tetzlaff, D., Smith, M. W., Reaney, S. M., & Roy, A. G. (2013). Concepts of hydrological connectivity: Research approaches, Pathways and future agendas. *Earth-Science Reviews*. https://doi.org/10.1016/j.earscirev.2013.02.001

Chen, K., Fu, K., Gao, X., Yan, M., Sun, X., & Zhang, H. (2017). Building extraction from remote sensing images with deep learning in a supervised manner. In *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)* (pp. 1672–1675). https://doi.org/10.1109/IGARSS.2017.8127295

Chen, T., Li, M., Li, Y., Lin, M., Wang, N., Wang, M., et al. (2015). MXNet: A Flexible and Efficient Machine Learning Library for Heterogeneous Distributed Systems. *Neural Information Processing Systems, Workshop on Machine Learning Systems*.

Correia, F., & Rego, F. (1998). Coupling GIS with hydrologic and hydraulic flood modelling. *Water Resources …*, 229–249. https://doi.org/10.1023/a:1008068426567

Cressie, N. A. C. (1993). *Statistics for Spatial Data (revised edition)*. *New York*. https://doi.org/10.2307/2533238

DeVantier, B. a., & Feldman, A. D. (1993). Review of GIS Applications in Hydrologic Modeling. *Journal of Water Resources Planning and Management*, *119*(2), 246–261. https://doi.org/10.1061/(ASCE)0733-9496(1993)119:2(246)

Díez Hermano, S. (2017). Machine learning regularity representation from biological patterns: a case study in a Drosophila neurodegenerative model.

Gascuel-Odoux, C., Aurousseau, P., Cordier, M.-O., Durand, P., Garcia, F., Masson, V., et al. (2009). A decision-oriented model to evaluate the effect of land use and agricultural management on herbicide contamination in stream water. *Environmental Modelling & Software*, *24*(12), 1433–1446.

Gehrke, S., Morin, K., Downey, M., Boehrer, N., & Fuchs, T. (2008). Semi-global matching: an alternative to lidar for dsm generation? *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XXXVIII-B1*, 1–6.

Heathwaite, A. L., Quinn, P. F., & Hewett, C. J. M. (2005). Modelling and managing critical source areas of diffuse pollution from agricultural land using flow connectivity simulation. *Journal of Hydrology*, *304*(1–4), 446–461. https://doi.org/10.1016/j.jhydrol.2004.07.043

Hu, F., Gao, X. M., Li, G. Y., & Li, M. (2016). DEM extraction from worldview-3 stereo-images and accuracy evaluation. In *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*. https://doi.org/10.5194/isprsarchives-XLI-B1-327-2016

Janzen, D., & McDonnell, J. J. (2015). A stochastic approach to modelling and understanding hillslope runoff connectivity dynamics. *Ecological Modelling*, *298*, 64–74. https://doi.org/10.1016/j.ecolmodel.2014.06.024

Knösche, R. (2006). Organic sediment nutrient concentrations and their relationship with the hydrological connectivity of floodplain waters (River Havel, NE Germany). *Hydrobiologia*, *560*(1), 63–76. https://doi.org/10.1007/s10750-005-0983-x

Knudby, C., & Carrera, J. (2005). On the relationship between indicators of geostatistical, flow and transport connectivity. *Advances in Water Resources*, *28*(4), 405–421. https://doi.org/10.1016/j.advwatres.2004.09.001

Li, R., Liu, W., Yang, L., Sun, S., Hu, W., Zhang, F., & Li, W. (2018). DeepUNet: A Deep Fully Convolutional Network for Pixel-Level Sea-Land Segmentation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 1–9. https://doi.org/10.1109/JSTARS.2018.2833382

Meerkerk, A. L., van Wesemael, B., & Bellin, N. (2009). Application of connectivity theory to model the impact of terrace failure on runoff in semi-arid catchments. *Hydrological Processes*, *23*(19), 2792–2803. https://doi.org/10.1002/hyp.7376

Patro, S., Chatterjee, C., Mohanty, S., Singh, R., & Raghuwanshi, N. S. (2009). Flood inundation modeling using MIKE FLOOD and remote sensing data. *Journal of the Indian Society of Remote Sensing*, *37*(1), 107–118. https://doi.org/10.1007/s12524-009-0002-1

Quinn, P., Beven, K., Chevallier, P., & Planchon, O. (1991). The prediction of hillslope flow paths for distributed hydrological modelling using digital terrain models. *Hydrological Processes*, *5*(1), 59–79. https://doi.org/10.1002/hyp.3360050106

Renard, P., & Allard, D. (2013). Connectivity metrics for subsurface flow and transport. *Advances in Water Resources*, *51*, 168–196. https://doi.org/10.1016/j.advwatres.2011.12.001

Saldaña, M. M., Aguilar, M. A., Aguilar, F. J., & Fernández, I. (2012). DSM extraction and evaluation from Geoeye-1 stereo imagery. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*. https://doi.org/10.5194/isprsannals-I-4-113-2012

Schumann, G., Hostache, R., Puech, C., Hoffmann, L., Matgen, P., Pappenberger, F., & Pfister, L. (2007). High-resolution 3-D flood information from radar imagery for flood hazard management. *IEEE Transactions on Geoscience and Remote Sensing*. https://doi.org/10.1109/TGRS.2006.888103

Seibert, J., & McGlynn, B. L. (2007). A new triangular multiple flow direction algorithm for computing upslope areas from gridded digital elevation models. *Water Resources Research*. https://doi.org/10.1029/2006WR005128

Shaw, E. M. E., Beven, K. J. K., Chappell, N. A. N., & Lamb, R. (2010). *Hydrology in Practice, Fourth Edition*. *Book*. https://doi.org/10.1016/0022-1694(94)90040-X

Siart, C., Bubenzer, O., & Eitel, B. (2009). Combining digital elevation data (SRTM/ASTER), high resolution satellite imagery (Quickbird) and GIS for geomorphological mapping: A multi-component case study on Mediterranean karst in Central Crete. *Geomorphology*, *112*(1–2), 106–121. https://doi.org/10.1016/j.geomorph.2009.05.010

Tarboton, D. G. (1997). A new method for the determination of flow directions and upslope areas in grid digital elevation models. *Water Resources Research*, *33*(2), 309–319. https://doi.org/10.1029/96WR03137

Tockner, K., Pennetzdorfer, D., Reiner, N., Schiemer, F., & Ward, J. V. (1999). Hydrological connectivity, and the exchange of organic matter and nutrients in a dynamic river-floodplain system (Danube, Austria). *Freshwater Biology*, *41*(3), 521–535. https://doi.org/10.1046/j.1365-2427.1999.00399.x

Trevisani, S., Cavalli, M., & Marchi, L. (2009). Variogram maps from LiDAR data as fingerprints of surface morphology on scree slopes. *Natural Hazards and Earth System Science*, *9*(1), 129–133. https://doi.org/10.5194/nhess-9-129-2009

Trinchero, P., Sánchez-Vila, X., & Fernàndez-Garcia, D. (2008). Point-to-point connectivity, an abstract concept or a key issue for risk assessment studies? *Advances in Water Resources*, *31*(12), 1742–1753. https://doi.org/10.1016/j.advwatres.2008.09.001

USDA-NRCS. (1986). Urban Hydrology for Small Watersheds TR-55. *U.S. Department of Agriculture Natural Resource Conservation Service Conservation Engineering Division Technical Release 55*, 164. https://doi.org/Technical Release 55

Vieux, B. (2016). *Distributed Hydrologic Modeling Using GIS*. *Distributed Hydrologic Modeling Using GIS SE - 1* (Vol. 38). https://doi.org/10.1007/978-94-015-9710-4_1

Western, A. W., Blöschl, G., & Grayson, R. B. (2001). Toward capturing hydrologically significant connectivity in spatial patterns. *Water Resources Research*, *37*(1), 83–97. https://doi.org/10.1029/2000WR900241

CHAPTER 3. THE EFFECTS OF TOPOGRAPHIC DEPRESSIONS ON MULTI-SCALE OVERLAND FLOW CONNECTIVITY: A HIGH-RESOLUTION SPATIAL AND TEMPORAL PATTERN ANALYSIS APPROACH BASED ON CONNECTIVITY STATISTICS

Manuscript submitted to Hydrological Processes

3.1 Abstract

In watershed modeling, the traditional practice of arbitrarily filling topographic depressions in digital elevation models (DEMs) has raised concerns because of the advancement of high-resolution remote sensing techniques, including aerial laser altimetry (LiDAR), which can identify real depressions that impact overland flow. We examined the connectivity statistics approach, i.e., connectivity function and Integral connectivity scale lengths, for quantifying the effects of depressions on overland flow processes across spatial scales. Connectivity statistics are implemented using GPU-accelerated computing techniques. We compared connectivity of overland flow patterns for five storm events between LiDAR-derived DEMs with (original) and without (filled) depressions. At least four macro connectivity stages were identified which correspond to connectivity states at different spatial scales: internal, local, partial watershed, and full watershed. These connectivity stages represent different overland flow response mechanisms, based on the states of depression storage capacity and connectedness between depressions. Results indicate that depressions can change the probability that any two points (grid locations) in a watershed are connected from -18% to +5%. Watershed-scale connectivity is less likely to occur when depressions are present, and vice versa for local-scale connectivity. The shapes of connectivity function plots are mostly convex, but concave sections also appear during the rising limb of most events, linked to shifts in the relative speed between filling and formation of connected pathways and the locations of depressions. Spatial distributions of connectivity changes indicate that increased connectivity is typically located in floodplain areas adjacent to the stream channel, i.e., the “riparian zone”, while decreased connectivity are typically located in the hillslope areas. Therefore, this study suggests that depressions are “nontrivial” in watershed modeling and their impacts on overland flow should not be neglected. Connectivity statistics at large spatial

scales provide new insights for characterizing the accumulative effects of interactions mechanisms between overland flow and small-scale topographic features.

3.2 Introduction

In hydrologic modeling, topography is a key input and has a significant impact on hydrologic/hydraulic features in both surface and subsurface rainfall-runoff models. Digital elevation models (DEMs) are the numerical representation of topography in a structural grid or triangular irregular network (TIN) format and have been extensively used as a fundamental input in watershed modeling (see examples in Cormen et al., 2001; Correia and Rego, 1998; DeVantier and Feldman, 1993; Miller and Semmens, 2002; Xu et al., 2014; Zhang and Montgomery, 1994). Initially, topography was estimated from paper contour maps and manually imported into computers because of the lack of highly accurate terrain survey techniques (Moore *et al.*, 1988; Jenson, 1991). DEMs derived from these spatially coarse datasets typically include topographic depressions, termed "pits" or "sinks", that appeared as a single or contiguous set of cells of low elevations surrounded by neighboring cells of higher elevations (Jenson and Domingue, 1988; Martz and Garbrecht, 1998; Rieger, 1998). Depressions are problematic in distributed models because flow is "trapped" and no outward flow direction is available to route and accumulate flows downslope, until the depressions are completely filled with additional rainfall and overflow (Arnold, 2010). To simplify analysis for practical purposes, early modeling work treated all depressions in DEMs as artifacts. The assumption was that all of these depressions resulted from inaccurate elevation measurements in the original contour maps or bias of interpolation methods (Band, 1986; Lindsay and Creed, 2005; Vaze *et al.*, 2010). In addition, it was noted that the scale of naturally occurring depressions is typically much less than that of the DEM spatial resolution (O'Callaghan and Mark, 1984; Jenson and Domingue, 1988; Hutchinson, 1989). Thus, depressions are completely removed in DEM data preprocessing prior to modeling, which has been an "acceptable necessity" (Hutchinson, 1989; Jenson and Domingue, 1988; O'Callaghan and Mark, 1984; Wechsler, 2007).

In recent decades, the spatial resolution of elevation data has greatly improved along with lower costs and easier accessibility, e.g., online data repositories including aerial laser altimetry (LiDAR) and satellite radar altimetry (Toutin, 1995; Hirano *et al.*, 2003; Berry *et al.*, 2007; Lemmens, 2008;

Shan and Toth, 2008). It is now possible to capture high-resolution (horizontal < 1 m) surface topography based on spatially dense point clouds (> 10 points/1 m^2) at a watershed scale (10-1000 km^2) or structural grid data (with a horizontal resolution of 30 m) at intercontinental or global scales (Farr and Kobrick, 2000; Shan and Toth, 2008; Siart *et al.*, 2009). Although high-resolution elevation data are also not free from errors (Van Genderen, 2011), with data scales consistent with the scales of real depressions, the assumption that all surface depressions are spurious and mere artifacts because of measurement error or interpolation bias is clearly no longer justifiable (Lindsay and Creed, 2006). Removal (filling) of depression comes with the risk of eliminating naturally occurring depressions and "overly" smoothing the topography. Indeed, depressions can play a crucial role in surface or subsurface flow processes (Tockner *et al.*, 1999; Darboux *et al.*, 2002a; Einsiedl, 2005; Ahmed *et al.*, 2011). Elimination or an overly simplified representation of depressions for high-resolution DEMs is problematic in modeling, because of a major impact on connectivity and runoff response, even at plot and field scales, i.e., through spatially and temporally changes of flow depth, velocity and direction, and distribution distance/timing of runoff responses at the field or plot outlet (Zhang and Cundy, 1989; Dunkerley, 2003; Antoine *et al.*, 2009). The widely used "quick fix", using a constant water storage parameter for treating depressions, is not sufficient: runoff may occur, and the active contributing area may increase even *before* the surface storage capacity is fully filled (I. D. Moore and C. L. Larson, 1979; Antoine *et al.*, 2009). Depressions can potentially increase storage capacity and roughness of topography and delay the overland flow response (Martin and Valeo, 2008). Incorporating rill/depression storage in plot scale topography has been shown to significantly increase the accuracy of runoff initiation timing and residence time distributions for surface and subsurface flow (Frei and Fleckenstein, 2010). Soil moisture is also likely to change because depressions change the water potential at a local scale (Lane *et al.*, 2004).

However, depressions are complex morphological features that have been difficult to dynamically and explicitly incorporate into watershed modeling conceptualization, instead of statistical and implicit estimations (Mekonnen *et al.*, 2016; Nasab *et al.*, 2017). The concept of "connectivity" which has been introduced to hydrology from other disciplines, e.g., ecology, and provides an alternative way to examine the emergent effects of plot or field scale depressions, including their spatial arrangements, on hydrologic responses at larger spatial scales, e.g., hillslope or watershed.

Although no unified definition of connectivity currently exists, the most consistent description is that "Connectivity is the linkages (or connectedness) between structural patterns of landscape units and their interaction with functional patterns of watershed processes such as runoff response, and fluxes of water or water mediated substances and energy transfer within the spatial and temporal extent of a hydrologic system, e.g., a watershed" (Bracken *et al.*, 2013). Connectivity has been recognized as a promising way to improve our current understanding of hydrologic processes, which can only be seen at the watershed scale as the dynamic and cumulative effects of small-scale process interactions and feedback (Antoine *et al.*, 2009; Bracken *et al.*, 2013). Depressions have been shown to have a strong correlation with hydrologic connectivity, depending on their unique spatial configurations (Antoine *et al.*, 2009, 2011a; Peñuela *et al.*, 2016). Overland flow transfer distance or "connectivity length" has been computed using a water "walker" conditional function on each cell of a series of field plots of random crater-like and low-relief DEMs, from the most upslope location to the field outlet (Darboux *et al.*, 2002a). Connectivity has also been measured based on the spatial configurations of microtopographic depression structures and their storage changes during fill and spill (Antoine *et al.*, 2009; Tromp-Van Meerveld and McDonnell, 2006). The characteristics of stream network connectivity has been used to represent topographic driven watershed runoff responses (Jencso and McGlynn, 2011).

To date, however, few attempts have been made to directly incorporate connectivity into interpreting hydrologic processes based on observations or modeling outputs. Rather, much of the work to date has been focused on the development of the statistical model of connectivity, e.g., patterns of soil moisture or properties, without explicitly identifying the interaction mechanism between structural factors and flow processes, e.g., impacts of changes of topographic features on surface and subsurface flow connection , and almost exclusively focused on the plot or field scale (Western *et al.*, 2001; Zehe *et al.*, 2007; Meerkerk *et al.*, 2009; Renard and Allard, 2013). In addition, few published works have provided a clear quantification of hydrologic connectivity at a watershed-scale using high resolution spatial data (e.g., Ali and Roy, 2010). The most possible explanation is that for flow connectivity at larger spatial scales and higher resolutions, many of the approaches used at field or plot scales would be extremely computationally intensive for modeling and/or impractical to collect observed data in the field.

This study is a first attempt to investigate the role of depressions on connectivity at multiple spatial scales, from an internal depression scale (~5 m^2) to a watershed scale (~75 km^2) using spatially and temporarily explicit overland flow patterns for single storm events. The patterns at high-resolution (5 m spatial scale and 60 min temporal snapshots) are generated using 2D hydraulic models with a direct rainfall method (DRM) and high-performance computing techniques. DRM has been widely used in flood modeling for planning and emergency response, but has been used far less in hydrologic modeling. DRM has also been used as an efficient alternative to provide a comparison of key hydrologic process parameters, e.g., overland flow networks, time of concentration, which may highlight the need for further calibration in traditional watershed modeling (Taaffe *et al.*, 2011; Hall, 2015). A detailed description of DRM and model settings are included in the Appendix A. In this study, we analyze the effects of depressions on overland flow pattern connectivity at for several storm events of different annual exceedance probability (AEP) by running simulations with and without depressions using DRM. Connectivity concept is based on the two-point statistics proposed by Western *et al.*, 2001, which explicitly account for the connectedness between paired locations of heterogeneous hydrologic patterns. We create a new implementation to upscale the connectivity measures to large grids. The high-resolution view of connectivity at larger spatial scales provides us with an unprecedented opportunity to answer the following questions: 1) Do depressions play a major role in overland flow connectivity at the multiple spatial scales? 2) What are the implications of connectivity on the feedback dynamic from small-scale processes, e.g., depression filling (as sink) and spilling (as sources)? 3) What is the temporal variability of connectivity on the cumulative spatial scales for a single storm event? 4) What are the other implications that connectivity can be represented in terms of the overland flow processes? 5) What are the spatial distributions of connectivity changes caused by depressions, and what are the potential practical implications for watershed modeling?

3.3 Materials and Methods

3.3.1 Overview

We first provide a detailed description of a study watershed, including its hydrologic characteristics, instrumentation, and data availability. Then, we describe the DRM implementation, including the rationale for modeling approach selection, parameterization and calibration. Finally,

we give a brief description of the theoretical background of connectivity statistics. The workflow of this paper includes five major steps: 1) Spatial data preprocessing for LiDAR point clouds, land use, and soil map. 2) DRM parameterization, modeling, and calibration: The DRM is implemented in a 2D hydraulic modeling software package, TUFLOW. 3) Flow pattern filtering and thresholding: Map outputs from TUFLOW are used as flow patterns and labeled into binary "high" or "low" cells. 4) Connectivity statistics computation: we created a new algorithm for the $\tau(h)$ and ICSLs to quantify connectivity states. 5) Results analysis: we compared the computed $\tau(h)$ and ICSLs between flow patterns generated based on a) an original DEM and, b) the original DEM after modification by a depression-filling algorithm.

3.3.2 Watershed description

The Goodwater Creek Experimental Watershed (GCEW), located in northeastern Missouri, has a total area of approximately 75 km2 and is a subwatershed of the Salt River Basin, a 6,400 km2 headwater source tributary to Mark Twain Lake (E J Sadler et al., 2015, Fig. 3.1a-b). Mark Twain Lake is the major reservoir for the local regions public water supply (Lerch et al., 2008). GCEW has flat and gentle rolling topography with slopes of mostly 0 – 3% (Long-Branch and South Fork Salt River Project data). The major soil types are Wisconsin and Illinoian loess overlaying pre-Illinoian glacial till. Argillic horizons of 40% - 60% smectitic clays are formed by illuviation of high clay content loess. The claypan layer is approximately 50-60 cm below soil surface (Soil Survey Staff, 2016). Soil texture is dominated by clay loam and silty clay loam, with hydrologic soil groups of C and D (Steiner et al., 2009). Land use is mostly agriculture, and the major crops include soybean, corn, and sorghum (Baffaut et al., 2015). The city of Centralia has approximately 2/3 of its urban area located at the south end of GCEW (~ 4 km2) (Long-Branch and South Fork Salt River Project data). Annual precipitation is approximately 1,000 mm and 30% becomes streamflow. Runoff accounts for 85% of the total streamflow. Subsurface drainage is not used in this area because of the practical issues of installing drainage in or below a claypan layer (Baffaut et al., 2015). In a field study of GCEW, we found three major types of surface depressions, including both artificial and natural occurrence (Fig. 3.2a-c): 1) Ponds and wastewater treatment lagoons; 2) Ditches by railroad and road embankments; 3) Small depressions (puddles) in low land areas, mostly located at the riparian zone, which is immediately adjacent to the stream. Formation of naturally occurring depressions (puddles) in these areas are likely linked to bank erosion and

sediment deposition processes. Also, frequent bank failure causes detachment of root balls of woody plants from the soil surface, which can also generate depressions. Accessible locations of artificially and naturally occurring depressions at GCEW that are identified in the LiDAR-derived DEM are verified in this field study.

3.3.3 Hydrological and meteorological data

The USDA-ARS established hydroclimatic instrumentation at GCEW in 1969 and has been recording long-term flow, weather, and water quality data since then (Sadler *et al.*, 2015b, 2015a; Fig. 3.1a). Continuous 5-minute stream flow data for the GCEW watershed are recorded at three weirs on the main stream channel of Goodwater Creek. Plot and field scale runoff measurements are also available (Baffaut *et al.*, 2015). Currently, 9 rain gauges are in use at GCEW. Weather station records include solar radiation, air and soil temperature, saturated and actual vapor pressure, wind speed and direction, and precipitation. Both the breakpoint and daily data are available for all the hydroclimatic records and can be accessed from the USDA-ARS STEWARDS (Sustaining the Earth's Watersheds – Agricultural Research Database System) 3.0 data portal (Steiner et al., 2009, also see weblink in Table 3.1).

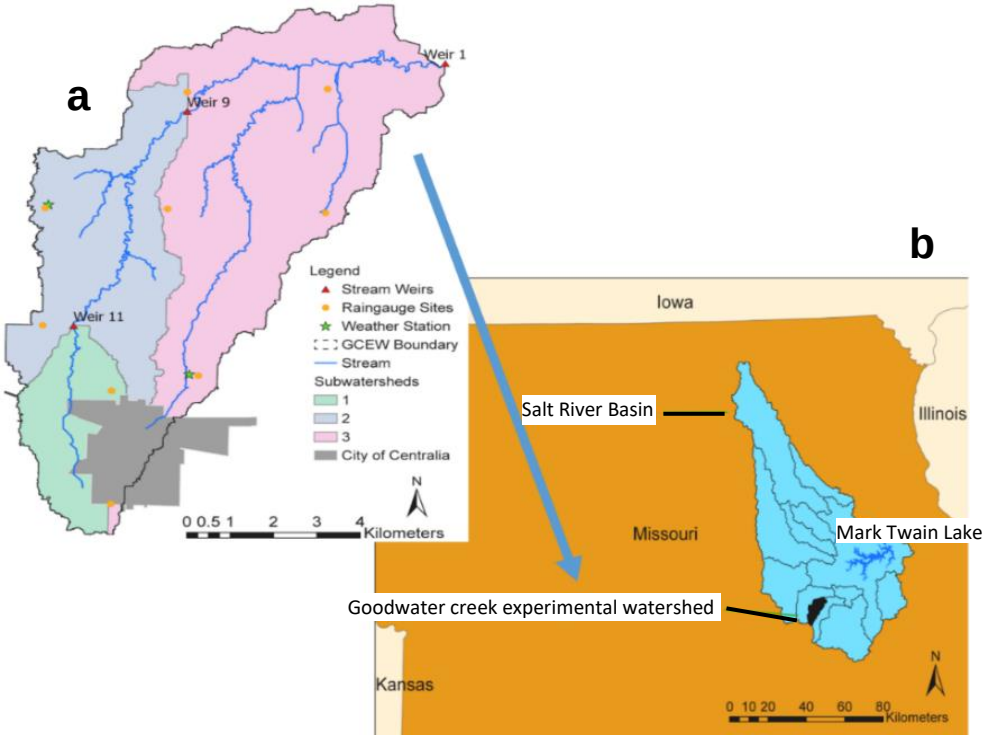


Figure 3.1 (a) Location of GCEW and instrumentation (Raingauges, weather stations and streamflow weir locations), and (b) location of Salt River Basin in Missouri.

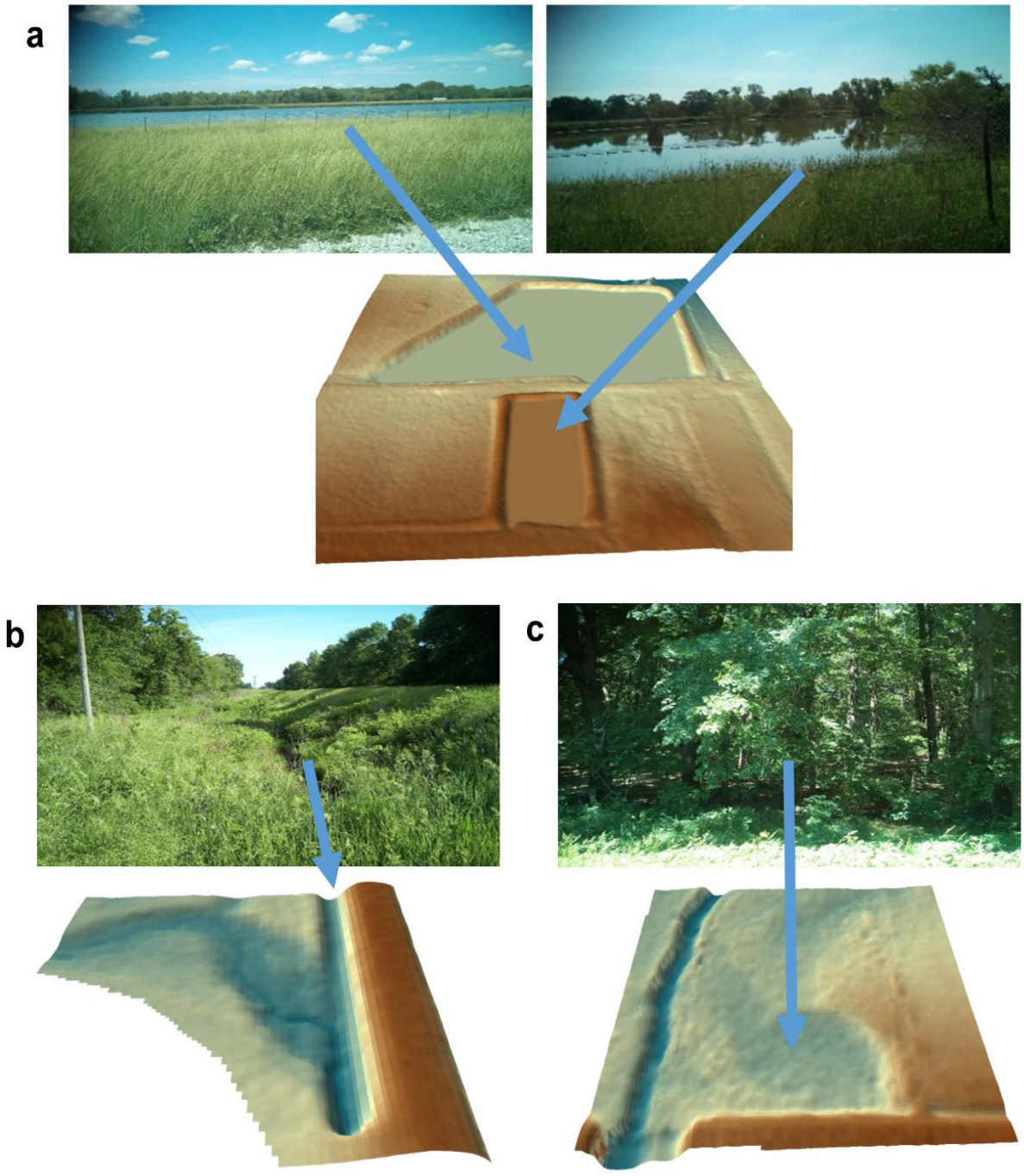


Figure 3.2 Three major types of depressions in GCEW. (a) Pond and wastewater lagoon on each side of a road and DEM at the same location; (b) Ditch by a road embankment and DEM at the same location; (c) Small depression (puddle) in low land areas at the riparian zone covered by vegetation and DEM at the same location.

In this study, the breakpoint data were aggregated into 15 min intervals as model inputs for single event simulation. Breakpoint flow weir data were aggregated into hourly intervals for model

calibration. We selected five events for this study that included three groups of annual exceedance probability (AEP): high (1% to 5%), medium (5% to 20%) and low (20% to 50%), based on a flood frequency analysis for each hydrologic year (from the beginning of April to the end of March).

3.3.4 Spatial data

We derived a DEM for GCEW using a high-resolution LiDAR point cloud (nominal post-spacing = 0.5 m) from the Long-Branch and South Fork Salt Creek Watershed Project data repository (USDA-NRCS, 2012). We did not use a DEM grid from the USGS National Elevation Dataset (NED) because NED currently does not have complete 1 m or 1/9 arcsecond (~3 m) DEM spatial coverage for the GCEW area (5 m resolution grid was used here and more details are included in section 3.3.5). In a preliminary study, we also found that the large number of small depressions in the riparian zone were mostly being filled in the NED. Therefore, in this study, we directly used LiDAR data to generate a DEM with no artificial modification of the measured elevations (except filtering out the nonground returns) to maximize the preservation of any real depression. In this study, the DEM was generated from LiDAR point cloud data using Natural Neighbor interpolation (implemented in ArcGIS) to preserve local-scale depressions while not creating spurious topographic features that are not already present in the data (ESRI, 2016; Sibson, 1981). We use this LiDAR-derived DEM to test the effects of depression removal on modeled overland flow patterns. A robust filtering algorithm is required to preprocessing LiDAR data that can separate ground from nonground signal returns and noise with minimal error (Liu, 2008; Shan and Toth, 2008). Robust Hierarchical Filtering (RHF) is a well-established approach for extracting ground return signals from complex landscapes, especially below vegetation cover (Pfeifer and Mandlburger, 2008). RHF was implemented using the SCOP++ software package (Pfeifer, 2001; Trimble Navigation Limited, 2016). The baseline values of Manning's roughness coefficient n for each land use type was assigned based on NLCD land use classification, and soil water physical characteristics , including USDA soil types, saturated hydraulic conductivity, suction head, and porosity, were estimated based on USDA soil texture classification retrieved from gSSURGO database (Homer *et al.*, 2015b; Soil Survey Staff, 2016). In situ and spatial data sources and their corresponding modeling parameters for DRM implemented in TUFLOW are included in Table 3.1.

Table 3.1 Hydrological and spatial datasets and corresponding sources used In DRM implementation

Dataset	Parameter in DRM modeling	Source	Link
Precipitation	Precipitation	USDA-ARS STEWARD 3.0 Portal	https://www.nrrig.mwa.ars.usda.gov/stewards/stewards.html
Stream stage and discharge	Hydrograph calibration		
LiDAR	DEM grid	Long-Branch and South Fork Salt River Project	http://www.msdis.missouri.edu/data/lidar/
Land use classification	Manning's n coefficient	National Land Cover Dataset (NLCD)	https://www.mrlc.gov/
Soil water physical parameters	Green-Ampt infiltration	gSSURGO database	https://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/survey/geo/?cid=nrcs142p2_053628

3.3.5 DRM implementation

After initial testing, we selected flow patterns at 5 m resolution to balance a predictable and practical connectivity statistics computational time with the maximum possible spatial resolution and model accuracy (best match to observed flow records). The original 1 m LiDAR-derived DEM was interpolated to 5 m using a triangular irregular network (TIN) transformation algorithm in TUFLOW. Rainfall is distributed across the watershed domain using nine Thiessen polygons based on the locations of the rain gauges. We used the Green-Ampt model as the infiltration loss method (Green and Ampt, 1911). Manning's roughness coefficients are used as bed resistance in TUFLOW. The computational time step was set to adaptive (for automatic adjustment), and the initial time step was 0.5 s. The results were calibrated based on measured streamflow discharge hydrographs at the watershed outlet. In addition, 5 locations were found where deep concentrated flow occurs adjacent to road embankments in the modeling results. We verified the presence of culverts at 3 corresponding locations depending on land accessibility in a field study (coordinates in Lat, Lon: 39°12'50.99"N, 92° 8'43.43"W; 39°17'36.37"N, 92° 7'17.34"W; and 39°12'44.33"N, 92° 9'10.11"W), and then manually added these culverts to the model. For the connectivity statistics computation, we used hourly time snapshots of flow depth patterns grids for the 5 storm events. Note that we used both original (DEM_{Ori}) and depression-filled (DEM_{Fill}) DEMs in the TUFLOW modeling to generate overland flow patterns. The DEM filling method used is based on Planchon and Darboux, 2002, which applies a uniformly distributed water level to the entire DEM,

which is then gradually drained out to and left in the areas of depressions (Fig. 3.3a-b). Depression filling was implemented in ArcGIS 10.5 (ESRI Environmental Systems Research Institute, 2016).

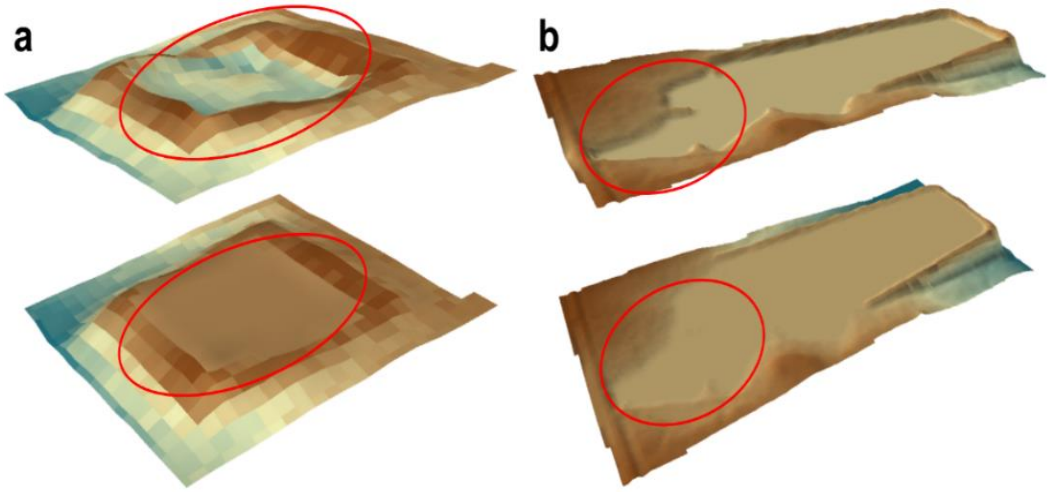


Figure 3.3 3D overlay of original DEM (top) and depression filled DEM (bottom). Surface depressions presented here include (a) wastewater treatment lagoon and (b) pond. Note that the internal areas of the depressions are filled with flat surfaces and the edges are smoothed (marked by red circles).

3.3.6 Connectivity metrics

The connectivity function τ(h) and ICSLs are connectivity statistics for multiple two-point (locations) connectedness based on percolation theory and are used to characterize the connectivity of hydrologic patterns, e.g., overland and subsurface flow pathways, antecedent soil moisture patterns, and saturated thalwegs (Meerkerk et al., 2009; Western et al., 2001 Grayson et al., 1997; James and Roulet, 2007). *τ(h)* is a lag distance dependent geostatistic that describes the probability of any pair of locations being connected, on the condition that a continuous pathway exists in the pattern between these pairs (Ali and Roy, 2010). *τ(h)* requires that a physical threshold is applied to the pattern and that cell values are labeled into two states: "high" (valid) and "low" (low). Connected pairs of locations are identified as those that have a continuous pathway with all "high" cells between them. Note that areas without valid data (missing or no data) are omitted from the calculation. The connected probability for each pair of at lag distance h within a tolerance $\pm h_t$ is calculated as the ratio between the number of connected pairs and the total number of pairs within $h \pm h_t$ regardless of the connectedness. One of ICSLs, an omnidirectional connectivity (OMNI), assume that the pathways connecting the paired locations are *always* valid regardless of their trajectories. If G is all the cells in the spatial pattern (excluding the no data cells), Z are the regions

where all cells are labeled "high" and h represents the lag distances between any pair of locations, then Western et al., (2001) proposed that the connectivity function at lag distance h within a tolerance $\pm h_t$ is:

$$\tau(\mathrm{h}) = Probability\ (x \leftrightarrow x + (h \pm h_t) |\ x\ \in Z, x + (h \pm h_t) \in G) \quad (1)$$

OMNI is calculated by integrating all the lag distances in the connectivity function $\tau(h)$ as:

$$\mathrm{OMNI} = \int_0^{\infty} \tau(\mathrm{h}) dh \quad (2)$$

Note that the lag distances are measured as Euclidean distance. The maximum possible lag distances h_{max} are determined as the maximum distance within the watershed boundary, i.e., the largest possible distance for any paired location within the flow pattern domain. Then, h_{max} is binned into a series of distance range bins, and the number of range bins n_{bin} is defined by the maximum lag distance $\mathrm{h_{max}}$ and the tolerance h_t: $n_{bin} = h_{max}/(h+2h_t)$ rounded to the closet integer. In each range bin, a is the total number of pairs that have a lag distance of h within that range, and b is the total number of pairs from a that are connected. The connected probability is calculated as P = b/a. P is plotted on the vertical axis against lag distance h as a line graph as shown in Fig. 3.4. The area under the curve is the integral of the connectivity function: OMNI, which describes the connectivity states across different lag distances of flow patterns at a particular time point, i.e., a time "snapshot". Alternatively, to describe the temporal variations of connectivity states during an entire storm event across all lag distances, OMNI is plotted on the vertical axis against time (at hourly interval, see Fig. 3.5a-e).

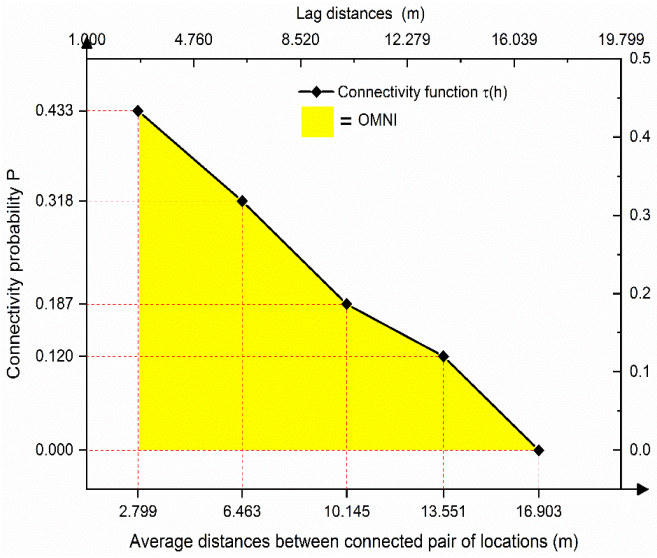


Figure 3.4 Connectivity function τ(h) and the integrated connectivity scale lengths (ICSLs).

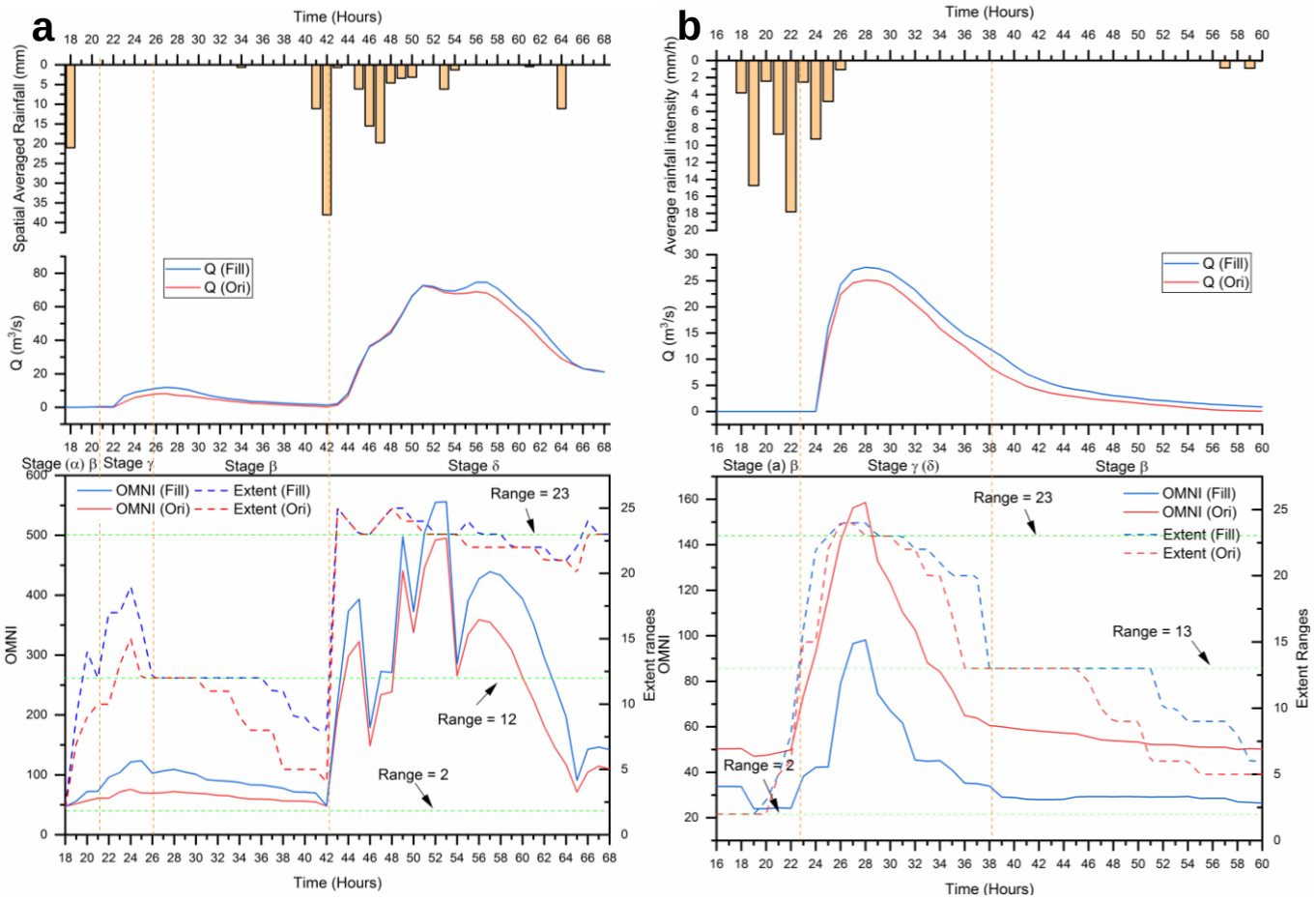
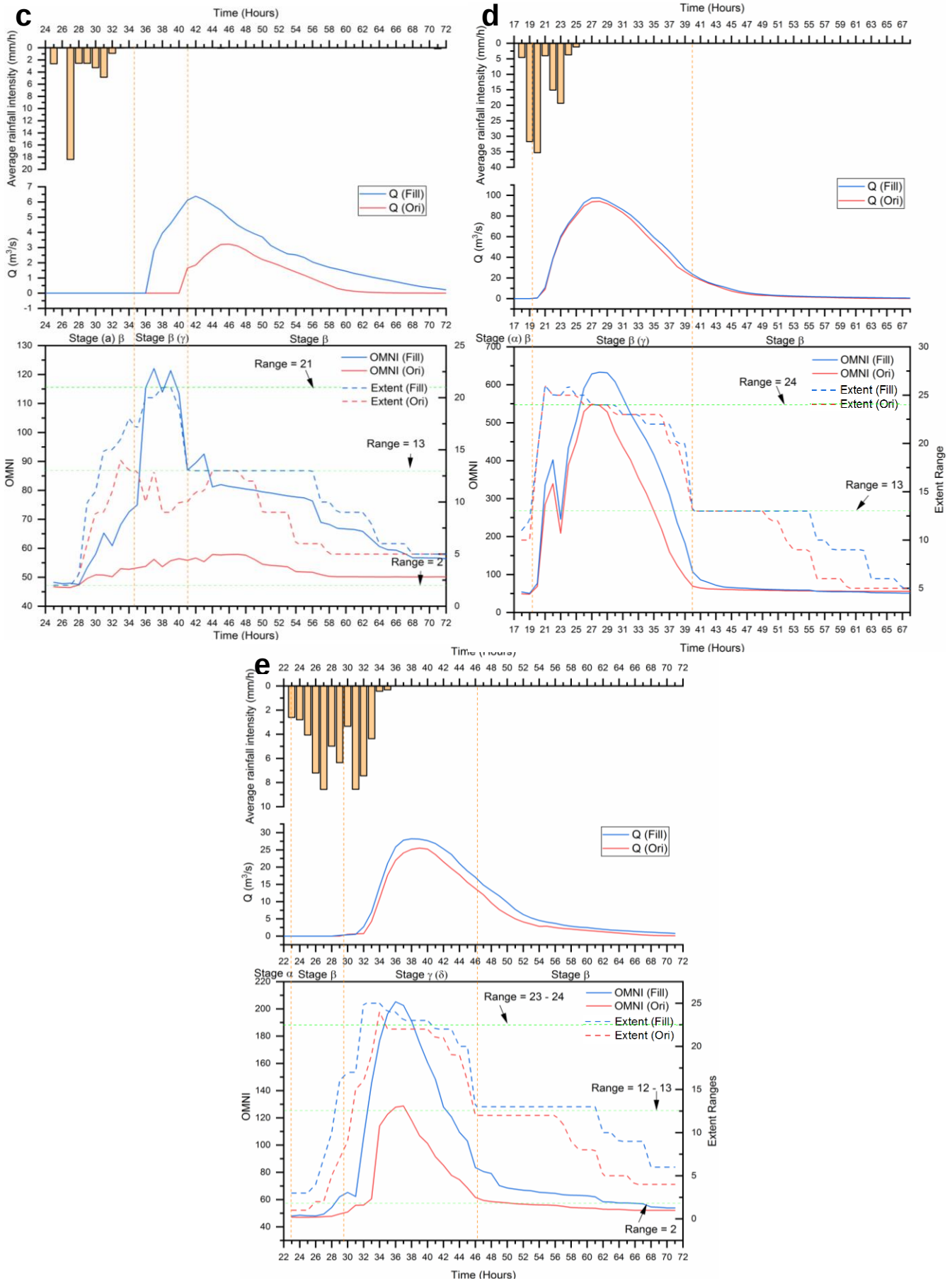


Figure 3.5 OMNI(t) (solid line) and connectivity extent range counts $ext_{\tau}(t)$ (dash line) plots for the selected 5 storm events at GCEW: (a) Event 1: Jun 20, 1981 (AEP 5-20%); (b) Event 2: Jul 04, 1998 (AEP 20-50%); (c) Event 3: Jun 03, 2001 (AEP 20-50%); (d) Event 4: Jun 10, 2006 (AEP 1-5%); and (e) Event 5: Sep 03, 2008 (AEP 20-50%). Rainfall hyetographs and stream discharge hydrographs are stacked on time axis as comparison.

Figure 3.5 continued



The effects of depressions can be reflected in the difference in $\tau(h)$ graph between the original and depression filled DEM and also in temporal variations in *OMNI(t)*. To simplify the discussion in this paper, we name the connectivity range bins using their sequential positions on the $\tau(h)$ graph, i.e., the 1st range bin (#1) represents the lag-distance range from 0 m to 701 m, and the 2nd range bin (#2) represents the lag-distance range from 701 m to 1402 m, and so on. We created a new algorithm to make the computation for large numbers of grid cells feasible (8,913,300 cells at 5 m resolution). Note that in this study, to explicitly account for the effects of depressions (and depression removal) on multi-scale connectivity, we used differences in the connectivity function plot $\tau(h)$ between a filled DEM (DEM_{Fill}) that has minimum amount of depressions, and the original DEM (DEM_{Ori}) that preserves all the depressions captured in the LiDAR data: $\Delta\tau(h)=\tau_{fill}(h)-\tau_{ori}(h)$ for each hourly snapshot of flow depth patterns. Detailed rationale and description of this new algorithm and a link to the source code repository are included in Appendix A.

3.3.7 Thresholding overland flow patterns

The threshold of "high" and "low" states for overland flow depth patterns is dependent on the DRM implementation in 2D hydraulic modeling. The Manning's roughness coefficient n is assumed constant and is used as an estimation of bed friction term in the momentum equation of SWE (see details in Appendix A). However, for very shallow flow depth *D*, n cannot be treated as a constant because the hydraulic resistance significantly increases with decreasing flow depth, especially when $D/k < 4$, where *k* is roughness height (e.g., Anderson *et al.*, 2006; Charbeneau *et al.*, 2009; Allan, 2014). The average roughness height of the original LiDAR-derived DEM is 0.0263 m, so assigning a constant Manning's n value as the unifying bed resistance parameter may not be reasonable when $D < 0.105$ m. In addition, only concentrated flow and channel flow are likely to impact connectivity in overland flow patterns at a watershed scale, so only these are set as "high" in the connectivity statistics computation (e.g., Meerkerk *et al.*, 2009). Commonly accepted shallow concentrated flow depths are between 3 cm and 15 cm without a well-defined channel, which is also the transition depth to open channel flow (USDA-NRCS, 2004). We conducted a one-way ANOVA test for the sensitivity of OMNI to the threshold of flow depths, including 2, 4, 6, 8, 10, 12, 14, 16, and 18 cm. No statistically significant difference was found between these flow depths and variability in OMNI for the test patterns at a 95% confidence

interval level. Therefore, we selected 10 cm as the threshold separating “high” and “low” states to eliminate cells with high uncertainty because of the DEM accuracy and limitations in DRM model parameterization.

3.4 Results and discussion

3.4.1 Role of surface depressions on overland flow connectivity across spatial scales

The sign and magnitude of $\Delta\tau(h)$ indicates shifts of interaction mechanisms between depressions and overland flow, and shifts of magnitude in the impacts of depressions on connectivity probability between any two points (cells) ($x \leftrightarrow x+h$), respectively, and at different spatial scales, denoted by lag-distances. Statistics shows that 98.3% of depressions at GCEW have a surface area between 25 m^2 to and 350 m^2 (1 to 14 grid cells), and 95.7% have a storage volume below 11.27 m^3. Thus, the distribution of the depression characteristics is highly skewed towards the low surface areas and volumes. In this study, the concept of depression hierarchy is also discussed and refers to the nested topological relationship between depressions of different scales (Fig. 3.7a-e). Depression hierarchy often relates to the spatial and temporal sequence of depression filling, merging, spilling and splitting process and thus directly changes the overland flow pathway connectivity (Chu *et al.*, 2013; Wu *et al.*, 2018). Surface areas are the most likely indicators of the nested hierarchy of depressions.

The magnitude of $\Delta\tau(h)$ ranges from -0.05 to 0.18, i.e. that depressions can decrease connectivity probability P up to 18% or increase it up to 5%. $\Delta\tau(h)$ plots for the 5 storm events are mostly positive at different spatial scales and points in time, which demonstrates that the connectivity probability between any two locations is mostly higher for flow patterns on DEM_{Fill}. This is to be expected because overland flow must fill the depressions before the connected flow pathways and extent becomes prevalent, and so overland flow mass and momentum are reduced, i.e., from the losses to depression storage, extended ponding time and surface roughness, which lead to decreasing the spatial extent of connected flow pathways. Interestingly, negative values of $\Delta\tau(h)$ occur between lag distances of 9,816 m and 15,425 m for event 1 and 4, and also between 0 m and 1,402 m for event 4. In these cases, the presence of depressions is increasing the connectivity probability. Most of these negative $\Delta\tau(h)$ occur during the recession phase of events.

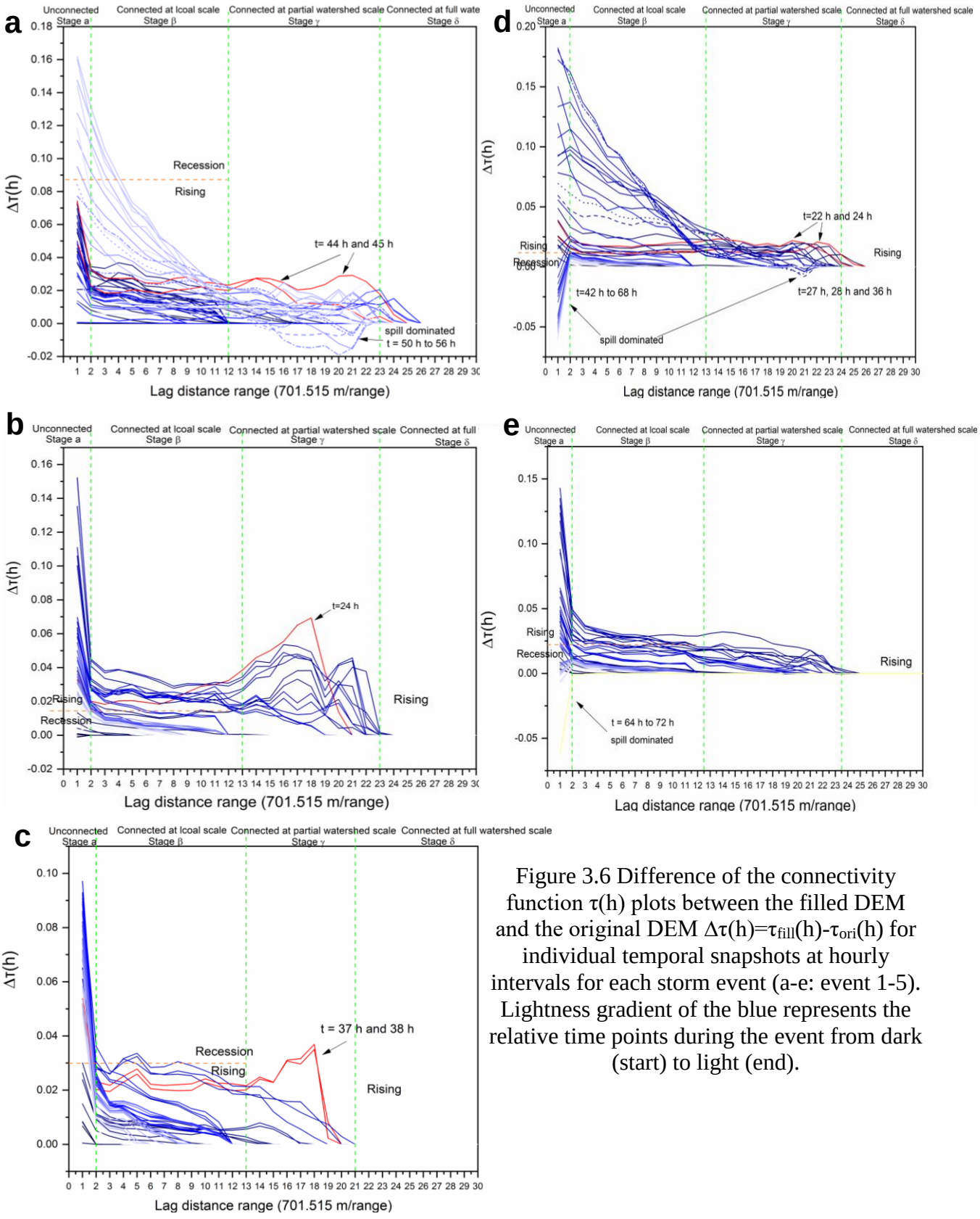


Figure 3.6 Difference of the connectivity function τ(h) plots between the filled DEM and the original DEM $\Delta\tau(h)=\tau_{fill}(h)-\tau_{ori}(h)$ for individual temporal snapshots at hourly intervals for each storm event (a-e: event 1-5). Lightness gradient of the blue represents the relative time points during the event from dark (start) to light (end).

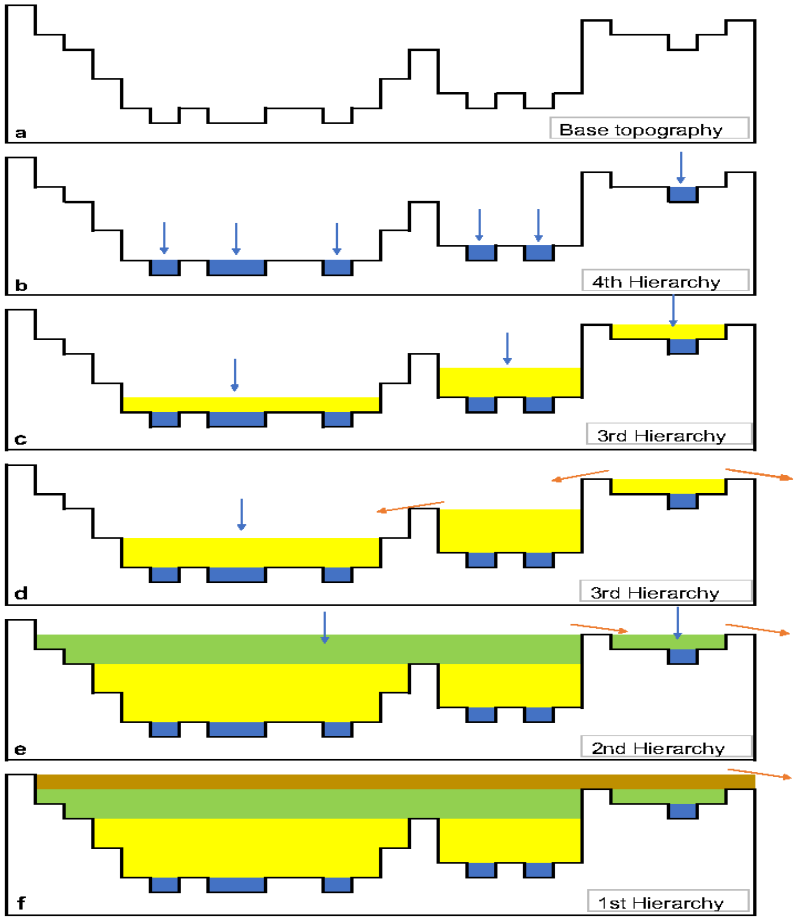


Figure 3.7 Hierarchical structure of depressions with nested topology on a cross section of a topography. The process of fill, merge, and spill are also included. Blue arrows represent the dominate filling process and orange arrows represent the dominate spilling process. (a) Base topography; (b) Depressions in the 4th hierarchy (Blue); (c-d) Depressions in the 3rd hierarchy (Yellow); (e) Depressions in the 2nd hierarchy (Green); and (f) Depressions in the 1st hierarchy (Blue). Split of the water levels at connected flow extent indicate the changes of hierarchy. Depressions at higher hierarchy usually have larger surface areas.

We inspected the overland flow patterns of DRM for these cases and found that two explanations could contribute depends on the lag-distance: 1) For large lag distance, e.g., 9,816 m to 15,425 m, when spill of overland flow from depressions (spill dominated, SD) is a dominant source of water, a larger extent of connected flow pathways is maintained for a longer period than the case when there are minimum amount depressions (DEM_{Fill}); 2) For small lag distance, e.g., 0 to 1402 m, connected overland flow pathways are also maintained at larger extent for a longer period on DEM_{Ori} than these on DEM_{Fill}, which relate to both the SD process and the extended ponding time

of depressions. We included the overland flow patterns of fill-dominate (FD) and SD processes for event 1 and event 4 at several locations in the Appendix B.

3.4.2 Macro connectivity stages represented spatially based on $\Delta\tau(h)$

Spatially, the magnitude of $\Delta\tau(h)$ is negatively correlated with lag-distances: $\Delta\tau(h)$ is the highest at the lowest possible lag distance range (0 m to 701 m), and decrease to zero at the largest possible lag distances within the watershed (14,721 m to 18,230 m). The speed of decrease (slope of $\Delta\tau(h)$) differs and four different situations are identified and labeled as α, β, γ, δ (Fig 3.6a-e and Table 3.2; but group δ rarely occurs and so is not included in the trend summary data). These situations relate to macro connectivity stages at the watershed scale, and thresholds between these stages can also be identified from the temporal variability of the connected pattern extent ($ext_{\tau(t)}$, see Fig. 3.5a-e bottom panels). Note that the connected pattern extent is defined as the maximum number of range bins that the connected overland flow pattern "snapshot" can reach spatially, given that $\tau(h) > 0$. The stage changes are signified by a stepwise shape in $ext_{\tau(t)}$ with each stage accompanied by a rapid increase/decrease and then a period of stability: Stage α: No significant overland flow pathways have been developed, and only water storage and internal flows within depressions are connected. Stage β: overland flow pathways only connect depressions at the local scale and are mostly within the floodplain. Stage γ: overland flow pathways connect many depressions at the watershed scale, including both across the floodplain and along part of the channels, and the drainage from these connected depressions reaches the watershed outlet. Stage δ: overland flow pathways connect almost all depressions at the watershed scale, including both the floodplain and channels, and drainage from all surface depressions reaches the watershed outlet. The lag-distances of macro connectivity increases from α, β, γ, to δ. The transition from stage α to β occurs between 701 m and 1,402 m which, for GCEW, is the average distance between the hillslope area and the floodplain adjacent to the stream channel. The transition from stage β to γ occurs between 7,712 m and 8,640 m which, for the study watershed is the average distance from the hillslope area plus half the length of the main stream channel. The transition from stage γ to δ occurs between 14,724 m and 17,280 m which, for the study watershed, is the average distance from the hillslope area plus the total length of the main stream channel. Note that all of the distances are estimated in 2D Euclidean space.

Table 3.2 Representation of different slope trend on Δτ(h) plots for each macro connectivity stages (median values). Note that Group 4 (Stage δ) is omitted because of insufficient data points

	Group 1 (Stage α)	Group 2 (Stage β)	Group 3 (Stage γ)
Event 1	-0.0214	-0.00182	-0.00105
Event 2	-0.0179	-0.000894	-0.00319
Event 3	-0.0486	-0.00136	-0.00268
Event 4	0.0267	-0.000601	-0.00135
Event 5	-0.0150	-0.000704	-0.00112

Typically, macro connectivity stages during a storm event would occur sequentially α→β→γ→δ→γ→β→α, but a storm event does not necessarily represent all stages, depending on the amount of rainfall amount, and the presence/absence of depressions. The stage transition sequence for event 1 is α→β→γ→β→γ→δ. For events 2 to 5 the sequence is α→β→γ→δ→γ→β. These sequences apply to the overland flow patterns on both DEM_{ori} and DEM_{fill}, with the exception of event 3, which the overland flow patterns on DEM_{Ori} is α→β→γ→β with only a short appearance of stage γ and no stage δ. Event 3 has the smallest accumulated rainfall of 35.58 mm among the 5 storm events. This suggests that storm of lower rainfall amount is more sensitive to the presence of depressions and can have an impact on the sequence of macro connectivity stages. For all 5 storm events, stages α and δ are very short duration because these stages occur when there is relatively low (< 2 mm/h) or high (> 18 mm/h) rainfall intensity, which relate to the start or the peak of a storm event. Note that small amount of water still retains in the stream channels and isolated depressions at the end of each storm event after the spill and split process, so the connectivity stage does not return to α.

In each macro connectivity stage, the impacts that depressions have can also be differentiated and reflected on the $\Delta\tau(h)$ plots: the magnitude of $\Delta\tau(h)$ differs significantly at the transition from stage α to β, especially in the rising limbs. In stage α, depressions typically decrease the connectivity probability ($|\Delta\tau(h)|$) by a magnitude of 0.1, and in stage β by a magnitude of 0.01. Because the scales of most depressions are below 720 m, at the lag distance range bin #1 (< 720 m and in stage α), depressions dominantly play the role of sinks. The magnitude of $\Delta\tau(h)$ rapidly reduces between range bins #2 and #3, between 720 m – 2160 m (stage β), especially in the rising limbs, because the depressions are not being connected by extensive overland flow

pathways across larger spatial scales yet, with or without depressions. Conversely, in a storm's late phase (recession limb) the difference in $\Delta\tau(h)$ between stage α and β is much smaller, because during the late phase, 1) most of the local depression storage areas are "filled" and connected at a larger spatial extent; 2) The role of depressions has been reversed, i.e., from fill-dominated (sinks) (FD) to spill-dominated (sources) (SD) and hence connectivity is less sensitive to the decline in rainfall and upslope overflow. For example, the negative sign of $\Delta\tau(h)$ from t=42 h to t=68 h and t=64 h to t=72 h in event 4 and 5 means that the overland flow patterns generated from the DEM_{Fill} have less connectivity probability than those generated from the DEM_{Ori} for lag distances smaller than 1402 m. Between these durations, most of the surface depressions are filled and connected, and the depressions transitioned from FD→SD. Therefore, depressions *increase* the overland flow connectivity in this situation, and the magnitude of such increase is higher than that for the removal of depressions, which is the most obvious within local scale. This finding is consistent with other studies that have described how, in recession limbs, flow is slowly draining (spill) out of low laying areas (mostly depressions) (e.g., Trigg *et al.*, 2013).

3.4.3 Macro connectivity stages represented temporally based on OMNI(t), the spatial extent and their potential correlation with rainfall and streamflow discharge

To investigate temporal changes in connectivity behavior, we examined *OMNI(t)*, connectivity extent $ext_{\tau(t)}$ (the number of lag distance range bins in connectivity function $\tau(h)$ that are not zeros), stream discharge hydrographs at the watershed outlet, and rainfall hyetographs (Fig. 3.5a-e, 1st and 2nd panel). The connectivity extent for overland flow patterns on DEM_{Fill} is primarily greater than or equal to that for DEM_{Ori}, i.e., connected overland flow pathways typically reach a larger spatial extent when depressions are removed (filled) (Fig. 3.5a-e, bottom panel). Connectivity extent is an indicator of macro connectivity threshold behavior as discussed above. The transition between connectivity stages can be identified by stepwise changes in $ext_{\tau(t)}$: a rapid increase or decrease in stage changes from low to high (α→β→γ(→δ)) or high to low ((δ→)γ→β→α). A relatively stable phase follows stage changes during high to low, but not during low to high. Early in a storm event, depressions mainly act as "sinks" and the overland flow pathways between these depressions have not yet been formed and there is minimal impact of rainfall intensity on OMNI. The transition α→β→γ(→δ) occurs in early phase of a storm event (rising limb) and corresponds to a switch from "isolated filling" (IF) to "connected filling" (CF) based on the hierarchical structure of

depressions, i.e., with the increase of connectivity spatial scales, flow pathways begin to connect between depressions in a lower hierarchy (already filled) while depressions at a higher hierarchy are still being filled. The transition of (δ→)γ→β→α occurs in late phase of a storm event (falling limb) and corresponds to a mixed phase of both 1) FD→SD and 2) CF→IF. Because statistics indicate that the majority of depressions are in lower hierarchy, the spatial extent of connected pathways expands rapidly during the transitions of α→β→γ(→δ). However, such expansion is stabilized for a certain duration once the macro connectivity stages enter the transitions of (δ→)γ→β→α when depressions are connected across all levels of the hierarchy. Depressions then becomes "sources" instead of "sinks", i.e., FD→SD, and the spill of water generate overflow and maintains the connected overland flow pathways for a longer period even if rainfall is declining and ceases. Any additional rainfall or increase in intensity is highly effective in generating excess runoff and expanding the spatial extent of connected overland flow pathways. Only after the "sources" are depleted, the effects of elimination of rainfall begin to become identifiable and represented by the sharp decreases in connectivity extent, and eventually lead to the transition from CF→IF.

Note that the transition of FD→SD can occur in either stage α or γ, but the magnitude of impact on OMNI(t)ori or OMNI(t)fill are different in these two cases. For example, in event 4, the negative values of Δτ(h) occur at t=42 to 68 h and t=27 to 28 h. However, the magnitude of Δτ(h) is lower in stage γ (-0.01 during t = 27 to 28 h) compared with stage β (-0.05 during t = 42 to 68 h) (Fig. 3.6d). This is because the effect of SD is more marked when connected pathways are restricted within high proximity to the depressions and so creating new pathways because of spill process can significantly increase connectivity. This effect is less obvious at the partial watershed scale because 1) a larger lag distances and thereafter a lager connected extent is expected from the spill; 2) lower floodplain areas and stream flow are also been included in the connectivity computation, which "dilutes" the effect of SD.

3.4.4 Δ*τ(h)* plots shapes and the interaction between overland flow and depressions at multiple spatial scales

Δ*τ(h)* plots predominately have a convex shape, which means that the magnitude of connectivity probability increases because of removal (filling) of depressions is initially high, when depressions

of low hierarchy are being filled during the internal and local scale (stage α and β) early in a storm, and then the magnitude of connectivity probability increases is reduced during partial and full watershed scale (stage γ and δ). Formation of connected overland flow pathways is much easier in internal/local scale than the partial/full watershed scale for overland flow on both DEM_{Ori} and DEM_{Fill} because of the highly skewed statistics of the scales of depressions and the predominate presence at lower hierarchy discussed above. However, some $\Delta\tau(h)$ plots have concave sections (marked in red, Fig. 3.6a-c, e), which means that magnitude of connectivity probability increases from removal (filling) of depressions are larger at the partial watershed scale than at the local scale. Interestingly, all of these cases correspond to the start of stream flow discharge and are at a point near the maximum of $ext_{\tau(t)}$. Thus, the concave shape may relate to a shift in the relative speed of depression filling and extending connected flow pathways because of the hierarchical structure of naturally occurring depressions and the connectedness between depressions. At the start of stream flow discharge, overland flow on DEM_{Ori} are at the transition of IF→CF, which means that depressions at a lower hierarchy have mostly been filled and the extension of the connected flow pathways is occurring faster than the filling within the lag-distances range of stage β. Therefore, mass and momentum losses to depressions is being compensated and $\Delta\tau(h)$ between DEM_{Ori} and DEM_{Fill} is lower at stage α and β. At partial or complete watershed scale (stage γ or δ), mass and momentum losses to depressions for overland flow on DEM_{Ori} are being accumulated because of the larger travel distances for overland flow. Thus, the effects of depressions become prominent and $\Delta\tau(h)$ increase. The thresholds for the start of concave sections are between stage β and stage γ (8412 m to 9113 m) for events 1 to 5, corresponding to the average distance between the watershed boundary and the floodplain areas immediately adjacent to the stream channels, i.e., the riparian zone. For example, in event 1, the rankings of $\Delta\tau(h)$ at hours t=44 and 45 have changed from 9 and 11 in stage β to 1 and 2 in stage γ, respectively. We found that depressions are spatially concentrated in these areas. Because the connected overland flow pathways between any two points (x↔x+h) must extend cross this area to have a lag-distance in stage δ and γ, a significant increase of mass and momentum losses occurs and causes the significant increase of $\Delta\tau(h)$.

3.4.5 Spatial distributions for the depression related changes of connectivity magnitude

Both *OMNI(t)* and $\Delta\tau(h)$ are predominately higher for overland flow patterns generated from DEM_{Fill} than that of DEM_{Ori}. We inspected the corresponding spatial locations of the temporally

aggregated magnitude of connectivity changes based on the snapshots within the watershed boundary and found that the increases of connectivity are mostly located in floodplain areas with a close proximity to the stream channel, e.g., the riparian zone. This is expected because 1) terrain LiDAR signal reflection in the riparian zone is susceptible to the high noise of the canopy tops and the accuracy of the filtering algorithm is lower and 2) the filling of depressions overly smoothed or "flattened" these areas so that the flow pathways are easier to connect. In contrast, on hillslope parts of the watershed, we found less overall connectivity for overland flow patterns on DEM_{Fill} than that on DEM_{Ori}. The most likely explanation is that the SD phase dominates in these areas for the majority of the event snapshots.

3.5 Conclusions

Connectivity at large scale is vital for understanding the aggregated effects of small-scale factors and their feedbacks on physical processes. In this study, we demonstrate that the connectivity statistics, i.e., connectivity function $\tau(h)$ and the omnidirectional integrated connectivity scale lengths (OMNI-ICSLs), are important indicators of overland flow response mechanisms to depressions across multiple spatial scales in a spatially and temporally explicit manner, and their potential correlations with rainfall and streamflow discharge. To achieve this, we implemented a new algorithm for $\tau(h)$ and OMNI for overland flow patterns generated using 2D hydraulic modeling on a low-relief agricultural watershed. We found that:

(1) Depressions mostly decrease, but in some cases can increase overland flow connectivity, based on the probability that any two points (locations), i.e., $x \leftrightarrow x+h$, are connected within the watershed boundary. The magnitude of such increase/decrease is between -18% and +5%, depending on the spatial scale between these two points and stage during a storm event. Depressions tend to decrease local connectivity (within depression areas) at maximum magnitude and then gradually reduce at larger spatial scales until reaching watershed scale. For the most duration of storm events, depressions decrease the connectivity because depressions in a DEM are acting as sinks of water and increasing infiltration losses by extended ponding time, and thus results in decreased the connected overland flow pathway. Depressions increase connectivity occasionally during the late phase (recession limb) of relatively high magnitude storm events (AEP < 20%), when

depressions are acting as sources of water. This also explains why the magnitude of $\Delta\tau(h)$ is lower in the recession limb than that in rising limb, because depression sources initially delay the impact of a cease in rainfall and overflow from upslope areas.

(2) A single event storm has four macro connectivity stages based on the spatial scales which also determine different magnitude of impacts on overland flow connectivity: An internal stage (connections within depression areas, stage α, < 701 m), localized stage (stage β, < 7,712 m or 8,640 m), partial watershed connected stage (stage γ, < 14,724 m or 17,280 m) and full watershed connected stage (stage δ, < 21,045 m). However, the thresholds separating these stages are determined by rainfall amount/intensity and other topographic features in addition to depressions. The significance of depressions is the most prominent for low magnitude rainfall events at GCEW because of the presence of numerous small-scale depressions. Stage α and stage δ only occur in a short duration for any storm events because of they require very low and high storm intensity. The transitions between stages, i.e., α→β→γ(→δ) and (δ→)γ→β→α, mostly occur at early phase (rising limb) and late phase (recession limb) of the storm events respectively, and progress at different speed because of the highly skewed distributions towards depressions of lower hierarchy and the shits from fill dominate (FD) to spill dominate (SD) for the interaction between overland flow and depressions.

(3) At the same point in time, *OMNIs(t)* are always higher for the overland flow patterns generated from DEM_{Fill} than that of DEM_{Ori}, i.e., depressions reduce the probability of connectivity at integrated spatial scales.

(4) In most cases, Δτ(h) plots show a convex shape with the maximum at connectivity stage α and decreasing to zero at stage δ. This is expected because the sensitivity of overland flow to depressions are most prominent at stage α, i.e., more prone to the shifts between FD and SD. However, for events 1 to 4, the Δτ(h) plots include concave sections at stage γ and δ. This increase in the expansion of the overland flow extent may link to the condensed distribution of small-scale depressions at riparian zone and occurs at the transition of IF→CF, corresponding to the start of increase in streamflow discharge. At GCEW, connectivity has been overestimated by depressions removal in these areas.

(5) The spatial distributions of connectivity increases/decrease because of removal (filling) of depressions are not uniform within a watershed. Increases in connectivity are mostly located in the floodplain immediately adjacent to stream, i.e., the "riparian zone", as this is an area where depressions mostly occur at GCEW. The decrease in connectivity are mostly located in hillslope areas because of the higher likelihood of SD condition.

This work shows that the common practice of depression removal (filling) in watershed modeling can have significant implications for simulated overland flow processes at multiple spatial scales based on high-resolution and LiDAR-derived DEMs. Depressions may change spatial heterogeneity of overland flow patterns through different interaction mechanisms, i.e., fill and spill. The impacts of depressions depend on macro connectivity stages which are based on connectivity extents (lag distance ranges). In this study, we also demonstrated that other functional connectivity factors, such as rainfall intensity, may have more control on the macro connectivity stages than depressions alone. Other landscape features, such as vegetations and land use patterns, may also have impacts on the connectivity of overland flow. Future work needs to include more structural and functional connectivity factors in the analysis and their corresponding implications for overland flow processes. In addition, this work used a separate 2D hydraulic modeling approach to generate the overland flow pattern. A more efficient approach would incorporate a simple estimation of excess and losses relationship between the paired locations directly (e.g., Ali *et al.*, 2018) in the connectivity metrics, without the need for a fully distributed modeling for overland flow.

3.6 References

Ahmed Z, Rao DRM, Reddy KRM. 2011. Sustainable storm water management - An evaluation of depression storage effect on peak flow. In *Proceedings of the International Conference on Green Technology and Environmental Conservation, GTEC-2011*336–340. DOI: 10.1109/GTEC.2011.6167690

Ali G, Oswald C, Spence C, Wellen C. 2018. The T-TEL Method for Assessing Water, Sediment, and Chemical Connectivity. *Water Resources Research* **54** (2): 634–662 DOI: 10.1002/2017WR020707

Ali GA, Roy AG. 2010. Shopping for hydrologically representative connectivity metrics in a humid temperate forested catchment. *Water Resources Research* **46** (12) DOI: 10.1029/2010WR009442

Allan T. 2014. A Study of the Varibility Versus the Assumed Constancy of Manning's n.Utah State University.

Anderson BG, Rutherfurd ID, Western AW. 2006. An analysis of the influence of riparian vegetation on the propagation of flood waves.pdf. *Environmental Modelling and Software* **21**: 897–902 DOI: 10.1016/j.envsoft.2005.04.027

Antoine M, Javaux M, Bielders C. 2009. What indicators can capture runoff-relevant connectivity properties of the micro-topography at the plot scale? *Advances in Water Resources* **32** (8): 1297–1310 DOI: 10.1016/j.advwatres.2009.05.006

Antoine M, Javaux M, Bielders CL. 2011. Integrating subgrid connectivity properties of the micro-topography in distributed runoff models, at the interrill scale. *Journal of Hydrology* **403** (3): 213–223 DOI: 10.1016/j.jhydrol.2011.03.027

Arnold N. 2010. A new approach for dealing with depressions in digital elevation models when calculating flow accumulation values. *Progress in Physical Geography* **34** (6): 781–809 DOI: 10.1177/0309133310384542

Baffaut C, Sadler EJ, Ghidey F. 2015. Long-Term Agroecosystem Research in the Central Mississippi River Basin: Goodwater Creek Experimental Watershed Flow Data. *Journal of Environment Quality* **44** (1): 18 DOI: 10.2134/jeq2014.01.0008

Band LE. 1986. Topographic Partition of Watersheds with Digital Elevation Models. *Water Resources Research* **22** (1): 15–24 DOI: 10.1029/WR022i001p00015

Berry PAM, Garlick JD, Smith RG. 2007. Near-global validation of the SRTM DEM using satellite radar altimetry. *Remote Sensing of Environment* **106** (1): 17–27

Bracken LJ, Wainwright J, Ali GA, Tetzlaff D, Smith MW, Reaney SM, Roy AG. 2013. Concepts of hydrological connectivity: Research approaches, Pathways and future agendas. *Earth-Science Reviews* **119**: 17–34 DOI: 10.1016/j.earscirev.2013.02.001

Charbeneau RJ, Jeong J, Barrett ME. 2009. Physical modeling of sheet flow on rough impervious surfaces. *Journal of Hydraulic Engineering* **135** (6): 487–494

Chu X, Yang J, Chi Y, Zhang J. 2013. Dynamic puddle delineation and modeling of puddle-to-puddle filling-spilling-merging-splitting overland flow processes. *Water Resources Research* **49** (6): 3825–3829 DOI: 10.1002/wrcr.20286

Cormen TH, Leiserson CE, Rivest RL. 2001. *Introduction to Algorithms , Second Edition*. DOI: 10.2307/2583667

Correia F, Rego F. 1998. Coupling GIS with hydrologic and hydraulic flood modelling. *Water Resources*: 229–249 DOI: 10.1023/a:1008068426567

Darboux F, Davy P, Gascuel-Odoux C. 2002. Effect of depression storage capacity on overland-flow generation for rough horizontal surfaces: water transfer distance and scaling. *Earth Surface Processes and Landforms* **27** (2): 177–191 DOI: 10.1002/esp.312

DeVantier B a., Feldman AD. 1993. Review of GIS Applications in Hydrologic Modeling. *Journal of Water Resources Planning and Management* **119** (2): 246–261 DOI: 10.1061/(ASCE)0733-9496(1993)119:2(246)

Dunkerley DL. 2003. Determining friction coefficients for interrill flows: The significance of flow filaments and backwater effects. *Earth Surface Processes and Landforms* **28** (5): 475–491 DOI: 10.1002/esp.453

Einsiedl F. 2005. Flow system dynamics and water storage of a fissured-porous karst aquifer characterized by artificial and environmental tracers. *Journal of Hydrology* **312** (1–4): 312–321 DOI: 10.1016/j.jhydrol.2005.03.031

ESRI Environmental Systems Research Institute. 2016. ArcGIS Desktop: Release 10.4. *Redlands CA*

Farr TG, Kobrick M. 2000. Shuttle radar topography mission produces a wealth of data. *Eos* **81** (48): 583–585 DOI: 10.1029/EO081i048p00583

Frei S, Fleckenstein JH. 2014. Representing effects of micro-topography on runoff generation and sub-surface flow patterns by using superficial rill/depression storage height variations. *Environmental Modelling and Software* **52**: 5–18 DOI: 10.1016/j.envsoft.2013.10.007

Van Genderen JL. 2011. *Airborne and terrestrial laser scanning*. Taylor & Francis Group. DOI: 10.1080/17538947.2011.553487

Grayson RB, Western AW, Chiew FHS, Blöschl G. 1997. Preferred states in spatial soil moisture patterns: Local and nonlocal controls. In *Water Resources Research*2897–2908. DOI: 10.1029/97WR02174

Green WH, Ampt GA. 1911. Studies on Soil Phyics. *The Journal of Agricultural Science* **4** (1): 1–24

Hall J. 2015. Direct rainfall flood modelling: The good, the bad and the ugly. *Australian Journal of Water Resources* **19** (1): 74–85

Hirano A, Welch R, Lang H. 2003. Mapping from ASTER stereo image data: DEM validation and accuracy assessment. In *ISPRS Journal of Photogrammetry and Remote Sensing*356–370. DOI: 10.1016/S0924-2716(02)00164-8

Homer CG, Dewitz JA, Yang L, Jin S, Danielson P, Xian G, Goulston J, Herold ND, Wickham JD, Megown K. 2015. Completion of the 2011 National Land Cover Database for the Conterminous United States – Representing a Decade of Land Cover Change Information. *Photogrammetric Engineering and Remote Sensing* **81** (5): 345–354 Available at: https://cfpub.epa.gov/si/si_public_record_report.cfm?dirEntryId=309950

Hutchinson MF. 1989. A new procedure for gridding elevation and stream line data with automatic removal of spurious pits. *Journal of Hydrology* **106** (3–4): 211–232 DOI: 10.1016/0022-1694(89)90073-5

I. D. Moore ID, C. L. Larson CL. 1979. Estimating Micro-Relief Surface Storage from Point Data. *Transactions of the ASAE* **22** (5): 1073–1077 DOI: 10.13031/2013.35158

James AL, Roulet NT. 2007. Investigating hydrologic connectivity and its association with threshold change in runoff response in a temperate forested watershed. *Hydrological Processes* **21** (25): 3391–3408 DOI: 10.1002/hyp.6554

Jencso KG, McGlynn BL. 2011. Hierarchical controls on runoff generation: Topographically driven hydrologic connectivity, geology, and vegetation. *Water Resources Research* DOI: 10.1029/2011WR010666

Jenson SK. 1991. Applications of hydrologic information automatically extracted from digital elevation models. *Hydrological Processes* DOI: 10.1002/hyp.3360050104

Jenson SK, Domingue JO. 1988. Extracting topographic structure from digital elevation data for geographic information system analysis. *Photogrammetric Engineering and Remote Sensing* **54** (11): 1593–1600 DOI: 0099-1112/88/5411-1593$02.25/0

Lane SN, Brookes CJ, Kirkby MJ, Holden J. 2004. A network-index-based version of TOPMODEL for use with high-resolution digital topographic data. *Hydrological Processes* **18** (1): 191–201 DOI: 10.1002/hyp.5208

Lemmens M. 2008. Leica Geosystems, HDS. *GIM International*

Lerch RN, Sadler EJ, Kitchen NR, Sudduth KA, Kremer RJ, Myers DB, Baffaut C, Anderson SH, Lin C-H. 2008. Overview of the Mark Twain Lake/Salt River basin conservation effects assessment project. *journal of soil and water conservation* **63** (6): 345–359

Lindsay JB, Creed IF. 2005. Removal of artifact depressions from digital elevation models: Towards a minimum impact approach. *Hydrological Processes* **19** (16): 3113–3126 DOI: 10.1002/hyp.5835

Lindsay JB, Creed IF. 2006. Distinguishing actual and artefact depressions in digital elevation data. *Computers & Geosciences* **32** (8): 1192–1204 DOI: 10.1016/j.cageo.2005.11.002

Liu X. 2008. Airborne LiDAR for DEM generation: some critical issues. *Progress in Physical Geography* **32** (1): 31–49 DOI: 10.1177/0309133308089496

Martin Y, Valeo C, Tait M. 2008. Centimetre-scale digital representations of terrain and impacts on depression storage and runoff. *CATENA* **75** (2): 223–233 DOI: 10.1016/j.catena.2008.07.005

Martz LW, Garbrecht J. 1998. The treatment of flat areas and depressions in automated drainage analysis of raster digital elevation models. *Hydrological Processes* **12** (6): 843–855 DOI: 10.1002/(SICI)1099-1085(199805)12:6<843::AID-HYP658>3.0.CO;2-R

Meerkerk AL, van Wesemael B, Bellin N. 2009. Application of connectivity theory to model the impact of terrace failure on runoff in semi-arid catchments. *Hydrological Processes* **23** (19): 2792–2803 DOI: 10.1002/hyp.7376

Mekonnen BA, Mazurek KA, Putz G. 2016. Incorporating landscape depression heterogeneity into the Soil and Water Assessment Tool (SWAT) using a probability distribution. *Hydrological Processes* DOI: 10.1002/hyp.10800

Miller S, Semmens D. 2002. GIS-based hydrologic modeling: the automated geospatial watershed assessment tool. *Proceeding of the Second Federal Interagency Hydrologic Modeling Conference*: 1–12 Available at: http://www.epa.gov/nerlesd1/land-sci/agwa/pdf/pubs/agwa-conference.pdf

Moore ID, O'Loughlin EM, Burch GJ. 1988. A contour-based topographic model for hydrological and ecological applications. *Earth Surface Processes and Landforms* DOI: 10.1002/esp.3290130404

Nasab MT, Singh V, Chu X. 2017. SWAT modeling for depression-dominated areas: How do depressions manipulate hydrologic modeling? *Water (Switzerland)* DOI: 10.3390/w9010058

O'Callaghan JF, Mark DM. 1984. The extraction of drainage networks from digital elevation data. *Computer Vision, Graphics, and Image Processing* **27** (2): 247 DOI: 10.1016/S0734-189X(84)80047-X

Peñuela A, Darboux F, Javaux M, Bielders CL. 2016. Evolution of overland flow connectivity in bare agricultural plots. *Earth Surface Processes and Landforms*: n/a-n/a DOI: 10.1002/esp.3938

Pfeifer N. 2001. Derivation Of Digital Terrain Models In The Scop++ Environment. *OEEPE Workshop on Airborne Laserscanning and Interferometric SAR for Digital Elevation Models, ,*: 13 Available at: http://www.lidar.com.br/arquivos/dem-scop-laser.pdf

Pfeifer N, Mandlburger G. 2008. LiDAR Data Filtering and DTM Generation. In *Topographic Laser Ranging and Scanning*CRC Press; 307–334. DOI: doi:10.1201/9781420051438.ch11

Planchon O, Darboux F. 2002. A fast, simple and versatile algorithm to fill the depressions of digital elevation models. In *Catena*159–176. DOI: 10.1016/S0341-8162(01)00164-3

Renard P, Allard D. 2013. Connectivity metrics for subsurface flow and transport. *Advances in Water Resources* **51**: 168–196 DOI: 10.1016/j.advwatres.2011.12.001

Rieger W. 1998. A phenomenon-based approach to upslope contributing area and depressions in DEMs. *Hydrological Processes* **12** (6): 857–872 DOI: 10.1002/(SICI)1099-1085(199805)12:6<857::AID-HYP659>3.0.CO;2-B

Sadler EJ, Lerch RN, Kitchen NR, Anderson SH, Baffaut C, Sudduth KA, Prato AA, Kremer RJ, Vories ED, Myers DB, et al. 2015a. Long-Term Agroecosystem Research in the Central Mississippi River Basin: Introduction, Establishment, and Overview. *Journal of Environment Quality* **44** (1): 3 DOI: 10.2134/jeq2014.11.0481

Sadler EJ, Sudduth KA, Drummond ST, Vories ED, Guinan PE. 2015b. Long-Term Agroecosystem Research in the Central Mississippi River Basin: goodwater creek experimental watershed weather data. (Special Section: LTAR in the Central Mississippi river basin.). *Journal of Environmental Quality* **44**: 13–17 Available at: https://www.crops.org/publications/jeq/articles/44/1/13

Shan J, Toth CK. 2008. *Topographic Laser Ranging and Scanning: Principles and Processing*. DOI: 10.1201/9781420051438

Siart C, Bubenzer O, Eitel B. 2009. Combining digital elevation data (SRTM/ASTER), high resolution satellite imagery (Quickbird) and GIS for geomorphological mapping: A multi-component case study on Mediterranean karst in Central Crete. *Geomorphology* **112** (1–2): 106–121 DOI: 10.1016/j.geomorph.2009.05.010

Sibson, R. 1981. A Brief Description of Natural Neighbor Interpolation. In: Barnett, V., Ed., *Interpreting Multivariate Data*, John Wiley & Sons, New York, 21-36.

Soil Survey Staff. 2016. Soil Survey Geographic (SSURGO) Database for [U.S.]. *United States Department of Agriculture accessed [01/09/20016]* Available at: http://soildatamart.nrcs.usda.gov

Steiner JL, Sadler EJ, Wilson G, Hatfield JL, James D, Vandenberg B, Chen J-S, Oster T, Ross JD, Cole K. 2009. STEWARDS watershed data system: System design and implementation. *Transactions of the ASABE* **52** (5): 1523–1533

Taaffe F, Gray S, Sharma A, Babister MK. 2011. The ineptitude of traditional loss paradigms in a 2D direct rainfall model. *33rd Hydrology and Water Resources Symposium* (July): 434–441

Tockner K, Pennetzdorfer D, Reiner N, Schiemer F, Ward J V. 1999. Hydrological connectivity, and the exchange of organic matter and nutrients in a dynamic river-floodplain system (Danube, Austria). *Freshwater Biology* **41** (3): 521–535 DOI: 10.1046/j.1365-2427.1999.00399.x

Toutin T. 1995. Generating DEM from stereo images with a photogrammetric approach: Examples with VIR and SAR data. *EARSeL Advances in Remote Sensing* **4** (2): 110–117

Trigg MA, Michaelides K, Neal JC, Bates PD. 2013. Surface water connectivity dynamics of a large scale extreme flood. *Journal of Hydrology* DOI: 10.1016/j.jhydrol.2013.09.035

Trimble Navigation Limited. 2016. Trimble SCOP++ Software Available at: http://trl.trimble.com/docushare/dsweb/Get/Document-789933/022516-022A_Inpho_SCOP++_TS_A4_0915_LR.pdf [Accessed 25 February 2016]

Tromp-Van Meerveld HJ, McDonnell JJ. 2006. Threshold relations in subsurface stormflow: 2. The fill and spill hypothesis. *Water Resources Research* **42** (2): 1–11 DOI: 10.1029/2004WR003800

USDA-NRCS. 1986. Urban Hydrology for Small Watersheds TR-55. *U.S. Department of Agriculture Natural Resource Conservation Service, Conservation Engineering Division, Technical Release 55*, 164. https://doi.org/Technical Release 55

USDA-NRCS. 2012. Long Branch and South Fork Salt Creek Watersheds Project Database. *U.S. Department of Agriculture Natural Resource Conservation Service*. Available at http://www.msdis.missouri.edu/data/lidar/ [Accessed 30 January 2017]

Vaze J, Teng J, Spencer G. 2010. Impact of DEM accuracy and resolution on topographic indices. *Environmental Modelling & Software* **25** (10): 1086–1098

Wechsler SP. 2007. Uncertainties associated with digital elevation models for hydrologic applications: a review. *Hydrol. Earth Syst. Sci* **11**: 1481–1500 Available at: www.hydrol-earth-syst-sci.net/11/1481/2007/ [Accessed 8 June 2016]

Western AW, Blöschl G, Grayson RB. 2001. Toward capturing hydrologically significant connectivity in spatial patterns. *Water Resources Research* **37** (1): 83–97 DOI: 10.1029/2000WR900241

Wu Q, Lane CR, Wang L, Vanderhoof MK, Christensen JR, Liu H. 2018. Efficient Delineation of Nested Depression Hierarchy in Digital Elevation Models for Hydrological Analysis Using Level-Set Method. *Journal of the American Water Resources Association* DOI: 10.1111/1752-1688.12689

Xu X, Li J, Tolson BA. 2014. Progress in integrating remote sensing data and hydrologic modeling. *Progress in Physical Geography* **38** (4): 464–498 DOI: 10.1177/0309133314536583

Zehe E, Elsenbeer H, Lindenmaier F, Schulz K, Blöschl G. 2007. Patterns of predictability in hydrological threshold systems. *Water Resources Research* DOI: 10.1029/2006WR005589

Zhang W, Cundy TW. 1989. Modeling of Two-Dimensional Overland Flow. *Water Resources Research* **25** (9): 2019–2035 DOI: 10.1029/WR025i009p02019

Zhang W, Montgomery DR. 1994. Digital elevation model grid size, landscape representation, and hydrologic simulations. *Water Resources Research* **30** (4): 1019–1028 DOI: 10.1029/93WR03553

CHAPTER 4. IMPACTS OF TOPOGRAPHIC DEPRESSIONS ON WATERSHED-SCALE HYDROLOGIC RESPONSE: CONNECTIVITY AND GRID RESOLUTION

4.1 Abstract

To investigate how topographic depression removal (filling) and grid resolution affects hydrologic response at the watershed scale, we used revised versions of two functional connectivity indicators typically used at field or plot scales: the Relative Surface Connectivity Function (RSCf) and Normalized Runoff Connectivity Function (NRCf). The revised indicators provide good estimates of geostatistical characteristics of depressions and their correlations with connectivity states for single storm events of different magnitudes, based on maximum depression storage, similarity between the function plots, e.g., shapes and slopes, and discontinuity in the trends. Depression removal (filling) changes these characteristics and shifts the interaction between depressions and overland flow process. In addition, we observed that the effects of grid resolution are even more significant than depression removal during three rainfall storm events of different magnitude for the study area, Goodwater Creek Experimental Watershed in northeastern Missouri.

4.2 Introduction

Topography is a key parameter in any hydrologic/hydraulic model application because of its impact on key physical processes (DeVantier and Feldman, 1993; Zhang and Montgomery, 1994; Correia and Rego, 1998; Miller and Semmens, 2002; Aryal and Bates, 2008; Beven, 2012). Topography physically determines flow routing and other hydraulic structures, such as channel morphology, and these are significant controls on mass, momentum and energy transfer, wave form movement and resistance (Tarboton *et al.*, 1991; Horritt and Bates, 2002a; Wallin and Johansson, 2002; Metz *et al.*, 2011). The numerical representation of topography is a DEM (digital elevation model) which provides a mathematical estimation of the terrain morphology. DEM quality has a direct impact on model performance.

In the early days of hydrologic/hydraulic modeling, the impact of DEM quality was rarely investigated because modelers typically only had access DEMs estimated from the elevations of

spatially coarse contour maps with high uncertainty. Also, modeling at that time was for coarse estimation of hydrologic response at watershed outlet aggregated from large spatial scales. It was reasonable to assume that "imperfections" in the DEMs were either error in field surveys or from biased interpolation process (O'Callaghan and Mark, 1984; Band, 1986; Lindsay and Creed, 2005). The most prevalent form of DEM "imperfection" are topographic depressions: single cell or contiguous regions of low elevations, also called "pits" or "sinks". Flow cannot be routed out of these regions until they are filled, and interruptions in flow routing become problematic in modeling because of overconcentrated flow near depressions, changes of soil moisture content distribution and significantly reduced flow discharge at watershed outlets (Lane *et al.*, 2004). Therefore, for simplification and practical expediency, early modeling treated all surface depressions as artifacts. Depressions are removed in data preprocessing (hydrologic conditioning) (O'Callaghan and Mark, 1984; Jenson and Domingue, 1988; Hutchinson, 1989), and this was viewed as an "acceptable necessity" in the hydrologic modeling community (Wechsler, 2007).

Recent advances in remote sensing techniques have significantly improved the spatial resolution of DEMs from sensors on multiple platforms, such as spaceborn stereo images, airborn laser altimetry (LiDAR), and spaceborn radar altimetry (Toutin, 1995; Hirano *et al.*, 2003; Berry *et al.*, 2007; Shan and Toth, 2008). These techniques produce high-resolution data for surface topography using spatially dense elevation measurements (Siart *et al.*, 2009). Because DEMs have resolutions comparable to the scales of real topographic depressions, it is no longer appropriate to assume that all the depressions are artifacts, and this brings the tradition of depression removal into question (Lindsay and Creed, 2006). Real depressions can play a critical role in watershed-scale hydrologic process and response. For example, depression storage filling and spilling dynamics not only determine the distribution, direction and magnitude of overland flow, but also water mediated substances and sediments (e.g., Ahmed et al., 2011; Darboux et al., 2002a; Einsiedl, 2005). The current "fix" to account for depressions by applying a constant storage value and reduction in the water budget before runoff occurs does not explicitly include the impacts of depression characteristics and the spatial arrangements that are now widely available and thus less realistic. In fact, runoff can occur before depression storage is completely filled, and the active contributing area may increase simultaneously (I. D. Moore and C. L. Larson, 1979; Antoine *et al.*, 2009) on surfaces of varying roughness (Darboux *et al.*, 2002b). The spatial configuration of depressions

also has strong correlation with hydrologic connectivity, and can cause a non-negligible impact on the distribution and timing of catchment runoff responses (Zhang and Cundy, 1989; Dunkerley, 2003; Antoine *et al.*, 2009). The variability of connectivity states (by functional indicators) has been related to the delay effects of depressions on overland flow because of the transitions from graduate filling of depressions to the "breakthrough" value when most of depressions are connected and significantly increase the hydrograph at outlet (Antoine *et al.*, 2009; Peñuela *et al.*, 2016).

Hydrologic/hydraulic model performance can also be impacted by the spatial resolution of DEMs. Models using DEMs with higher spatial resolutions often represent hydrologic features more accurately, e.g., stream morphology or topographic index, but do not necessarily produce a closer match with observations than models using lower resolution DEMs (Wolock and Price, 1994; Ghaffari, 2011; Yang *et al.*, 2014). In addition, higher DEM resolutions also increase the model computational intensity, which may make a model simulation too expensive or unachievable. A balanced approach would be to find the maximum resolution that can be modeled, with the minimum loss of performance. Prior work has suggested that 10 m might be the optimal resolution for deriving hydrologic parameters from DEM in watershed modeling (see examples in Zhang and Montgomery, 1994; Hancock, 2005; Yang *et al.*, 2014). The spatial resolution of DEMs is a confounding factor that interacts with depressions in terms of impacts on hydrologic connectivity and watershed flow process. The storage capacity and spatial configurations (topologic relationships) of depressions may be significantly different for DEMs of different resolutions. For example, Alvarez-Mozos et al., (2011) found that the maximum depression storage (DS_{max}) had a nonmonotonic response to shifts of DEM resolution, slopes, and tillage density in an agricultural watershed. Changes in depression characteristics may lead to changes in spatial distributions of drainage and flow pathways, and thus shift hydrologic connectivity (Yang and Chu, 2013). Changes in DEM resolution and connectivity eventually lead to changes in runoff response, which have been found at inter-rill, hillslope, and field scales (Darboux *et al.*, 2002b; Antoine *et al.*, 2011b; Yang and Chu, 2015).

However, prior work has not explicitly quantified the impacts of depressions and their interactions with DEM resolution on hydrologic response at a watershed scale, using high-resolution DEM

data. Work to date has focused on theoretically generated or field measured DEMs at plot or field scales. One reason that evaluating the impact of depressions and DEM resolution on watershed hydrology has been challenging is because it requires a fully distributed and robust modeling approach which is sensitive to changes in detailed morphology in a high-resolution DEM without issues such as instability or mass balance inaccuracy. In addition, the DEM generation process should be aimed at preserving the original terrain surface morphology, including depressions, to the maximum possible degree. With such a DEM and modeling, it is possible to examine the impacts of depressions, and the interactions with DEM resolution on spatial and temporal patterns of overland flow and connectivity at a watershed scale.

Therefore, the goal of this study is to investigate the impact of surface depressions on hydrologic responses at the watershed scale and the role of DEM spatial resolution in changing these responses. Specific objectives of this study include investigating:

1) The correlation between DEM resolution and the characteristics of depressions, i.e., spatial distributions and geomorphologic characters (surface drainage areas, storage volumes);
2) The effects of depression storage on watershed scale hydrologic responses using DEMs of different grid resolution, i.e., surface drainage areas and streamflow discharge, based on connectivity, and the potential threshold behavior.

4.3 Materials and Methods

4.3.1 Data resolution and model selection

High-resolution surface elevation data at GCEW were accessed from LiDAR data repository of Long-Branch and South Fork Salt River Project and used to generate the DEM for hydrologic modeling. However, the large number of DEM grid cells (> 1,000,000) generated is impractical to use in traditional fully distributed hydrologic models because of extremely high computing costs and long run times. An alternative approach to address this issue involves recent advances in the field of computational hydrology, particularly 1D/2D hydraulic modeling based on Shallow Water Equations (SWEs) combined with high-performance computing. Software packages for hydrologic modeling (e.g., MIKE FLOOD and TUFLOW), have recently introduced “heavily

paralleled" computation (Syme, 2001; Patro *et al.*, 2009) that leverages the highly parallel computational capacity of GPU devices to significantly reduce modeling time without making any simplifications regarding the numerical solution schemes. SWEs can still be fully resolved for each cell but at < 1/100th of the time as traditional CPU computation (Huxley and Syme, 2016). Another advantage of these modeling packages is that they incorporate a "Direct Rainfall Method" (DRM) to simplify the conceptualization of watershed modeling and implementation of SWEs (Caddis *et al.*, 2008; Clark *et al.*, 2008; Hall, 2015). In the DRM, rainfall is directly "mapped" onto the topography and converted to overland flow. Then, flow is completely routed on the topography in the entire watershed domain to the outlet without any lumping assumptions. In addition, the computational cost is relatively low because high capacity GPU devices are widely available. Details of DRM are provided in Appendix A.

4.3.2 Experimental design

The first step is DEM and high-resolution flow pattern generation. In this study, to efficiently upscale the overland flow simulation to a watershed scale within a reasonable time frame, we implemented DRM in the TUFLOW high performance computing (HPC) version to generate spatially explicit overland flow patterns at GCEW. The LiDAR point cloud elevation data (point density 3.89/m^2) was filtered and then used to generate a 1m DEM based on the Natural Neighbor method to maximize the preservation of local-scale depressions and implemented in ArcGIS (Sibson, 1981; ESRI; 2016). The 1 m DEM was resampled using a triangular irregular network (TIN) interpolation method in TUFLOW to produce lower resolution (2m, 5m, 10m) DEMs. Other spatial and hydrologic data, i.e., NLCD land use classifications, SSURGO soil survey mapping and observed rainfall records at GCEW are used as input parameters for DRM to generate overland flow patterns (Homer *et al.*, 2015a; Sadler *et al.*, 2015a; Soil Survey Staff, 2015). Depression-filled DEMs were also used as a comparison with the original DEMs for each spatial resolution, and depression filling was implemented in ArcGIS 10.5 (ESRI Environmental Systems Research Institute, 2016) DRM was implemented in TUFLOW.

Next, we analyzed structural characteristics of surface depressions, including drainage areas, storage volume and total numbers of depressions, in terms of their statistical variability at different spatial resolutions. These indicators are computed based on DEMs exported from TUFLOW

resampled DEMs at 3 different resolutions: 2 m, 5 m and 10 m. We used the ArcHydro add-on in ArcGIS to obtain these structural characteristics for each depression and then analyzed their distributions at the watershed scale using SPSS (ESRI Environmental Systems Research Institute, 2016; IBM SPSS Inc., 2017). Because these lower resolution DEM grids (2 m, 5 m and 10 m) resampled by TUFLOW were based on the 1 m grid instead of the original LiDAR data, we created a test set of lower resolution grids which were directly interpolated from the LiDAR data for a sub-region of the study area, and compared the statistical distribution of the drainage areas and storage volumes of the depressions between these two set of data. These two datasets are very similar (Figure 4.1a-b), and so we concluded that DEM grids generated by TUFLOW at 2 m, 5 m and 10 m resolution, which were resampled from the 1 m resolution grid, have depressions of similar drainage areas and storage volumes to grids interpolated directly from the LiDAR data.

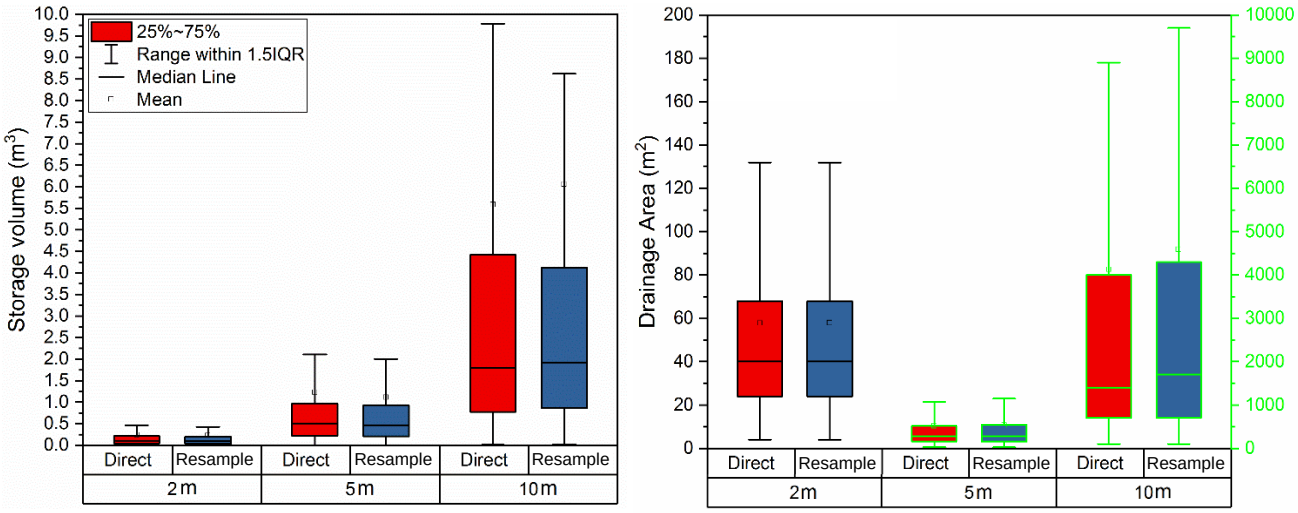


Figure 4.1 Box plots of the storage volume (a) and drainage area (b) for depressions in DEMs directly interpolated (labeled "Direct", in read) from LiDAR data and TUFLOW resampled (labeled "Resample", in blue) from 1 m resolution grid. The difference between these two interpolation methods are minor. Note that the direct interpolation method is Natural Neighbor, and TUFLOW uses a TIN based interpolation method to downgrade the resolution.

Last, TUFLOW modeling results in the form of spatially explicit overland flow extent for each hourly time "snapshot", were used to examine potential threshold behavior represented by connectivity changes, and any variability in the hydrograph at the watershed outlet, corresponding to the changes of the interaction mechanism between depressions and overland flow. We used a revised version of two functional connectivity indicators: Relative Surface Connection function (RSCf) and Normalized Runoff Connection Function (NRCf) to capture the filled depression storage, discharge at outlet and the connected flow extents in a spatially and temporally explicit

manner (Antoine *et al.*, 2009; Grimm and Chu, 2018). Changes in RSCf and NRCf may also correspond to changes triggered by thresholds in the timing and magnitude of runoff response. All of these analyses include a comparison of the TUFLOW modeling results using both the original DEMs and DEMs with depressions removed (filled).

Note that in this study we implemented DRM in TUFLOW as a practical and feasible alternative to lumped and semi-distributed models for overland flow pattern generation. DRM was originally designed to simulate flooding and inundation extent, and so focuses on water mass and momentum conservations for overland flow based on topography. Subsurface flow processes and the interface with surface flow are not represented in DRM, and so this approach is most appropriate for watersheds where surface water is the main contributor to overland and streamflow discharge. In addition, slope gradients at large scale should be low to minimize the impact of distortions to overland flow patterns except depressions. In this study, GCEW meets these restrictions: minimal groundwater interaction because of a claypan layer (very low hydraulic conductivity and as an impervious layer) under shallow soil depths (50cm to 60cm) in low-relief topography (0% - 3%).

4.3.3 Watershed description and hydrologic datasets

The 75 km^2 Goodwater Creek Experimental Watershed (GCEW) is a subwatershed of the 6,400 km^2 Salt River Basin (E J Sadler et al., 2015). Topography in GCEW has slopes of mostly 0 – 3% (Long-Branch and South Fork Salt River Project data). Soil in GCEW typically has high clay content and the major soil types are clay loam and silty clay loam, with hydrologic soil groups of C and D (Steiner *et al.*, 2009). GCEW located in continental climate zone with a strong seasonality. Artificial subsurface drainage is installed because of the claypan layer (Baffaut *et al.*, 2015). In a field study of GCEW, we found three major types of depressions: 1) Ponds and wastewater treatment lagoons; 2) Ditches by railroad and road embankments; 3) Small depressions (puddles) in low areas, mostly in the riparian zone. Bank erosion and bank failure often lead to the detachment of root balls of woody plants, leading to small depressions in the riparian zone. Other mechanisms to produce small depressions in the riparian zone include scour and deposition by overbank flow. For those small depressions included in the riparian zone in the DEM that were not visible in the field, they are most likely errors related to high frequency noise in LiDAR data caused by canopy tops. We obtained stream flow and rain gauge data (Baffaut *et al.*, 2015) from

the USDA-ARS STEWARDS (Sustaining the Earth's Watersheds – Agricultural Research Database System) 3.0 data portal (Steiner et al., 2009). We selected three events of annual exceedance probability (AEP) at high (1% to 5%), medium (5% to 20%) and low (20% to 50%), respectively. Further details regarding the watershed environment and hydrologic datasets are provided in Chapter 3.

4.3.4 Spatial Data

A DEM for GCEW was derived from the high-resolution LiDAR point cloud (nominal post-spacing = 0.5 m) downloaded the Long-Branch and South Fork Salt Creek Watershed Project data repository (USDA-NRCS, 2012). The DEMs from USGS National Elevation Dataset (NED) were not used because (1) No complete coverage of 1 m or 1/9 arcsecond (~3 m) DEMs for GCEW was available (5 m NED only available in Alaska. Lower resolutions, e.g., > 10 m are not used here. More details regarding grid resolution can be found in section 4.3.5); (2) Small depressions in the riparian zone were found mostly being smoothed in NED. In this study, we aimed to preserve the real depressions detected by the original LiDAR data without any artificial modifications. A robust LiDAR filtering algorithm was required which can separate ground returns from nonground returns and noise with minimal error (Liu, 2008; Shan and Toth, 2008). We used the Robust Hierarchical Filtering (RHF) algorithm which is especially suitable for terrain surface with vegetation cover (Pfeifer and Mandlburger, 2008). We used SCOP++ software package to implement RHF (Pfeifer, 2001; Trimble Navigation Limited, 2016). NLCD land use types were used to assign Manning's roughness coefficient n , and USDA soil texture types in gSSURGO were used to estimate soil water physical characteristics (Homer *et al.*, 2015b; Soil Survey Staff, 2017).

4.3.5 DEM and grid resolution

In this study, a sufficient DEM resolution is needed to capture rapid surface morphological changes, especially the edge of high slope change areas such as depressions and channel edges (Hall, 2015; Huxley and Syme, 2016). In GCEW, the channel width of the main stream is between 8 and 20 m. Therefore, any resolution above 10 m can't be used for the entire watershed (less than 2 grids cover the channel width) and so we used a 1D/2D link approach instead: 1D model for stream channel

area using a 1 m grid, linked to 2D model using a coarser resolution grid for floodplain and hillslope areas.

Widely used hydrologic/hydraulic models were run with a variety of grid resolutions (Table 4.1). For 1D/2D hydraulic models, DEM resolutions are mostly between 1m and 60m. For lumped and semi-distributed hydrologic models, DEM resolutions are mostly between 30m and 50m and are spatially lumped to some degree, e.g., aggregated homogenous hydrologic response units (HRU), upslope drainage area and/or landscape discretization based on hydrologic similarity. Lumped or semi-distributed models are often used for larger scale watersheds between 100 km^2 and 1000km^2 (Campling *et al.*, 2002; K. Ajami *et al.*, 2004; Barco *et al.*, 2008; Goderniaux *et al.*, 2009; Kiesel *et al.*, 2010) and 1D or 2D hydraulic models are often used for smaller scale watersheds between 1km^2 and 100km^2 (Horritt and Bates, 2002b; Phillips *et al.*, 2005; Carrivick, 2006; Pappenberger *et al.*, 2007; Hunter *et al.*, 2008; Erpicum *et al.*, 2010; Fabio *et al.*, 2010; Kalyanapu *et al.*, 2011; Smith *et al.*, 2012; Yang *et al.*, 2014; Jarihani *et al.*, 2015). The exceptional cases where 1D/2D hydraulic models are used for large scale watersheds use lower grid resolution, e.g., 90m (Wilson *et al.*, 2007; Patro *et al.*, 2009). Overall, most works use DEMs at resolutions between 10m and 50m. Studies have shown that peak discharge, peak height, travel time, terminal water storage and flood extent, are not sensitive to grid resolution greater than 30m (Jarihani *et al.*, 2015) or between 30m and 50m (Horritt *et al.*, 2006; Wu *et al.*, 2008). However, model sensitivity to resolutions finer than 10m has rarely been discussed in current literature and the spatial scales for 66.7% depressions are between 1m to 10m at GCEW. Therefore, we examined resolution effects on overland flow using grid sizes of 2m, 5m, and 10m here. Note that the 1 m resolution grid is not used here because of the practical issues of long simulation time and large memory requirement (> 100 GB).

4.3.6 Connectivity metrics

To quantify the effects of depressions on overland flow connectivity and watershed scale runoff response, as well as their interactions with DEM resolutions, connectivity metrics are needed to interpret the spatial and temporal variability of overland flow patterns and streamflow discharge at the watershed outlet. The metrics should be sensitive to grid resolution changes.

Table 4.1 Grid resolutions for selected hydrologic/ hydraulic modeling publication

Model conceptualization	Grid Size m	Total simulation area km^2 or length km	No. of cells	Model	Article
Fully Distributed	30m	480km^2	533333	HydroGeoSphere	Goderniaux et al. 2009
Semi distributed	30m	1645km^2	18277777	SAC-SMA (Sacramento soil moisture accounting model)	Ajami et al. 2004
Semi distributed	25m	50 km^2	80000	SWAT (Soil and Water Assessment Tool)	Kiesel et al. 2010
Semi distributed	50m	379km^2	151600	TOPOMODEL	Campling et al. 2002
Lumped	30m	217 km^2	241111	SWMM (Storm Water Management model)	Barco et al., 2008
2D Hydraulic modeling	4m	1.65km^2	100000	XP-SWMM and TUFLOW Classic	Philips et al. 2005
1D Hydraulic modeling	30 - 50m	60km	1200 to 12107	HEC-RAS, TELEMAC-2D and LISFLOOD-FP	Horritt and Bates 2002
2D	10m	500km^2	5000000	SOBEK	Carrivick 2006
2D	30m	16000km^2	8710000	TUFLOW GPU	Jarihani et al. 2015
2D	9.36m	62km^2	708864	Flood2D GPU	Kalyanapu et al. 2010
1D	50m	20km	73472	LISFLOOD-FP	Smith et al. 2012
2D	1- 60m	320 km^2	32000000 to 88889	HSPF (Hydrological Simulation Program – Fortran)	Yang et al. 2014
2D	1 - 4m	1.256km^2	450000	WOLF2D	Erpicum et al. 2010
2D	50m	~5km^2	2000	LISFLOOD-FP 2D	Pappenberger et al. 2007
2D	25m	~29km^2	87945	Using Aronica et al. (1998)	Fabio et al. 2010
1D/2D	90m	6800km^2	839506	MIKE Flood	Patro et al. 2009
1D	90m	260km	414000	LISFLOOD-FP	Wilson et al. 2007
2D	2m	0.4km^2	100000	TUFLOW Classic, DIVAST-TVD, TRENT, JFLOW, and LISFLOOD-FP	Hunter et al. 2008

In Chapter 3, we used the two-point connectivity statistics, the connectivity function τ(h) and integral connectivity scale lengths (ICSLs), as the connectivity metrics. However, in a preliminary test we found that these metrics are not sensitive to the connectivity changes at different grid resolutions between overland flow patterns generated from original and depression removed (filled) DEMs. Therefore, connectivity is represented based on a revised version of the relative surface connection function (RSCf) which has been used to describe the effects of the depression storage on overland flow connectivity at plot and field scales (Antoine *et al.*, 2009; Peñuela *et al.*, 2013, 2016). RSCf measures the surface areas which are hydraulically connected to the outlet (SA in m^2) as a function of depression storage capacity being filled (DS in m^3/m^3) at a certain time point and is sensitive to grid resolution changes because both SA and DS change with resolution. The effects of depressions on watershed runoff response at the outlet is represented based on a revised version of the Normalized Runoff Connectivity Function (NRCf), which measures the accumulated overland runoff (CR, in mm) at an outlet as a function of depression storage capacity being filled (Grimm and Chu, 2018). Note that to eliminate dependency on the spatial scale of the modeling domain, SA, DS and CR are normalized by the area of the domain (AD in m^2), the maximum storage capacity (DS_{max} in m^3) and the accumulated rainfall (CP in mm) respectively, as C=SA/AD, $D=DS/DS_{max}$, and Q=CR/CP.

However, the original formulation of RSCf and NRCf for the plot or field scale is not realistic or reasonable at the watershed scale, including the assumption of 1) no infiltration and 2) instantaneous transfer of overland flow to watershed outlet. Therefore, we revised the RSCf and NRCf by 1) replacing $D=DS/DS_{max}$ with $D_i=D_{Vconnected}/DS_{max}$, where $D_{Vconnected}$ is the total storage capacity (volumes in m^3) being hydraulically connected to the watershed outlet by overland flow pathways, regardless of their deficit status; 2) replacing C=SA/AD by $C_i=C_{DA}/AD$, where C_{DA} is the depression drainage area connected to the outlet. Thus, the uncertainty of overland flow extent in DRM can be greatly reduced because we only consider the *drainage area* of the depressions instead of the exact area of overland flow extent, which may subject to variability of rainfall applied on the topography in DRM implementation. The metrics are named RSCfi and NRCfi. RSCfi represents the depression drainage areas that are being hydraulically connected to the watershed outlet (as the proportion of total watershed area) as a function of the total depression storage (as the proportion of total depression storages) being included in these connected areas,

based on the spatial coverage of overland flow extent (connected to the watershed outlet) at any given time. NRCfi describes the accumulated streamflow discharge (normalized by the accumulated rainfall amount) at the watershed outlet as a function of the total depression storage (as the proportion of total depression storages) being included in the hydraulically connected areas, based on the spatial coverage of overland flow extent (connected to the watershed outlet) at any given time. This is implemented by overlaying the TUFLOW generated overland flow extents of each hourly time snapshot on the grid of depressions and their drainage areas and finding the accumulated total storage and drainage areas at the spatial intersection.

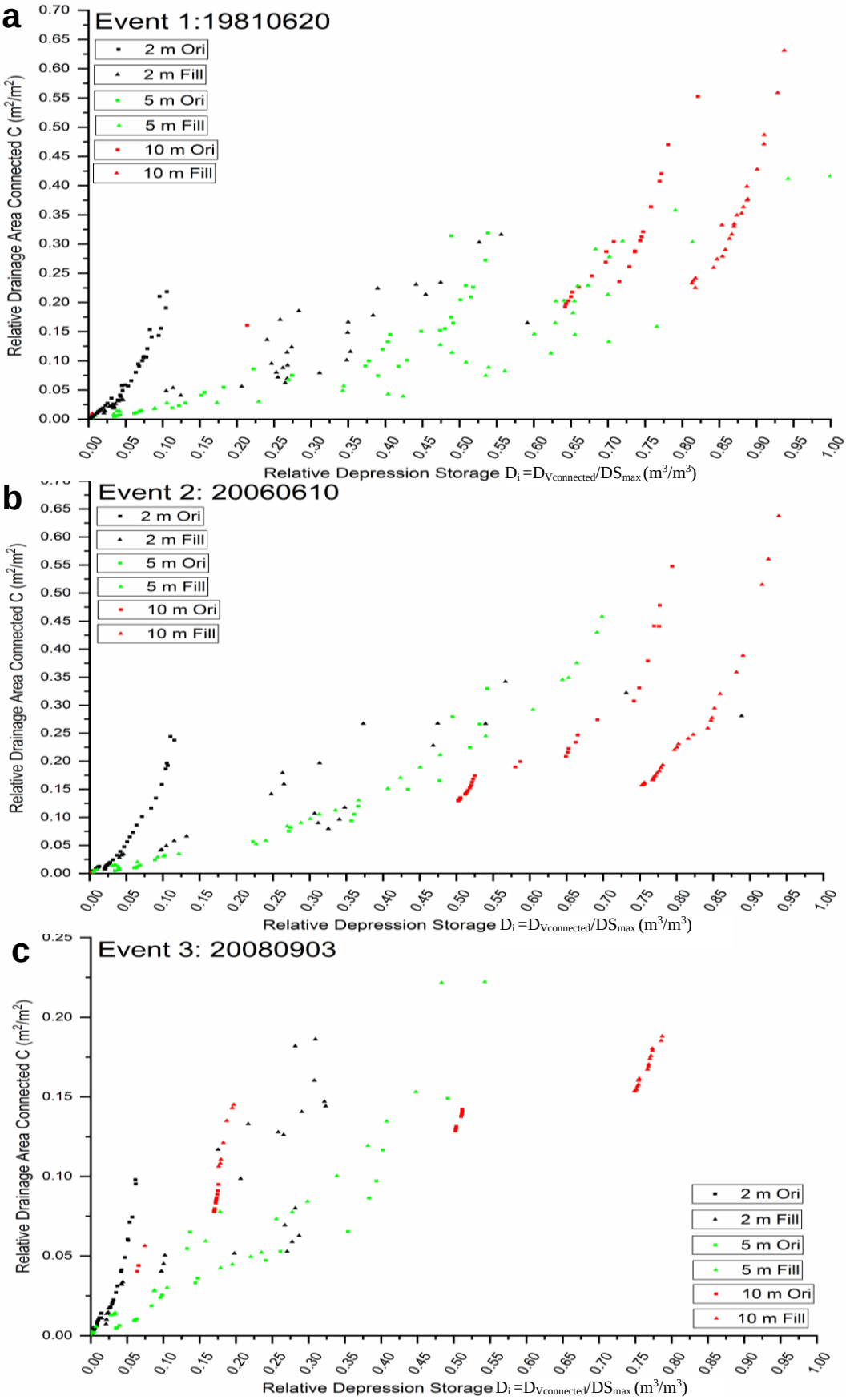
4.4 Results and discussion

Three thresholds have been discussed in RSCf: the initial ratio of surface connected (C_0), the connectivity threshold (CT) and the maximum ratio of depression storage (D_{max}). C_0 occurs where the overland flow area already connected to the outlet prior to the storm event, CT occurs where a sharp increase in C in response to a small increase in depressions $D=DS/DS_{max}$ (slope = 1), D_{max} occurs where the maximum depression storage that are filled during a storm event (Peñuela *et al.*, 2015, 2016). For RSCfi, C_0 at the start of each storm event is always zero because the DRM we used in this study assumes no upstream inflow boundary for 1D stream channel area in TUFLOW. CTs are not observed in our $RSCf_i$ function plots (Fig. 4.1a-c). The lack of obvious CT means that the threshold of "volume to breakthrough" is not applicable for $RSCf_i$ at a watershed scale because the overland flow transition time is not negligible, so overland flow waves propagation are discretized in more detail this study and cannot be assumed to reach the watershed outlet in one time step as previous works at field or plot scales (e.g., Antoine *et al.*, 2011). $RSCf_i$ function plots for all storm events at 2 m resolution show trends of smooth convex shapes based on DEM_{Ori} and more "flattened" shapes with lower slopes based on $DEM_{Fill.}$ The convex shapes of RSCf function plots for overland flow patterns on DEM_{Ori} indicate a relatively well-connected topography with large numbers of small-scale depressions (Antoine *et al.*, 2009; Appels *et al.*, 2011, 2016; Peñuela *et al.*, 2015), although with a slightly slower initial increase (before $D_i = 0.025$) in connected overland flow pathways than that of DEM_{Fill} (Peñuela *et al.*, 2016). At GCEW, the 2 m resolution DEM shows that 66.7% of depressions have a surface area between 1 m^2 to and 12 m^2 (1 to 3 grid cells), and 94.8% have a storage volume below $1m^3$. These depressions are predominantly low in the hierarchy, i.e., they are filled first, then the overflow is merged into or connected with

depressions higher in the hierarchy with significantly higher volumes. Once these small-scale depressions have been filled, the extent of connected overland flow pathways rapidly expands and thus there is a rapid increase in connectivity (after $D_i = 0.025$) which is higher than that of DEM_{Fill}. Note that a rapid increase in connectivity would not be immediately apparent on the streamflow hydrograph because of the delayed flow transfer time at the watershed scale.

We found that removal (filling) of depressions generates large numbers of flat areas which reduce the storage volumes of depressions high in the hierarchy and increase drainage areas to these depressions because of merging of depressions low in the hierarchy (Fig. 4.2a-b and Table 4.4). A slightly faster initial increase of C_i (before $D_i = 0.025$) is expected than that for DEM_{Ori} because of reduced storage volumes, but this trend reverses after D_i=0.025, where the speed of Ci increase is lower than that for DEM_{Ori}. Overland flow on DEM_{Fill} have less mass and momentum losses, and can hydraulically connect more depressions to the watershed outlet. However, such overland flow extent increases on DEM_{Fill} are slower than DEM_{Ori} (when a threshold amount depression storage D_i=0.025 are connected) because of the higher magnitude of infiltration on the larger and merged drainage areas (for depressions mostly located in the low land areas, e.g., the riparian zone), which delays the filling of depression storage and thus changes the convex shapes of $RSCf_i$ function plots towards lower gradient and "flatter" shapes. This trend is more significant at 2 m resolution than 5 m and 10 m resolution. In addition, significant residuals occur in the $RSCf_i$ function plots based on DEM_{Fill} (not on DEM_{Ori}) at both 2 m and 5 m resolution, and the magnitude of residuals increases is higher at 2 m than 5 m resolution. For DEM_{Fill}, depressions higher in the hierarchy often have larger surface areas but not necessarily larger storage volumes because some of these depressions are filled and merged, i.e., depression storage volumes have been disproportionally filled with increase of drainage areas. Therefore, there is an increase in the variability of drainage areas for the depression storage. At a 10m resolution, the difference between the shapes of the $RSCf_i$ trends is not significant (while it was for 2 m and 5 m resolution), and the trends are parallel with consistent larger gaps (not presented at 2 m or 5 m resolutions). This may be related to the artificially increased storage volume and overly "smoothed" surface topography at 10 m resolution, so that the small-scale depressions (these low in hierarchy) are eliminated. Thus, depressions with large drainage areas and lower proximity (these high in hierarchy) are retained and DEM grids are not sufficiently sensitive to the difference in the speed of initialization of connected overland flow

Figure 4.2 RSCfi function plots for the three storm events at 2 m, 5 m and 10 m grid resolutions. (a) Event 1 in 1981, (b) Event 2 in 2006 and (c) Event 3 in 2008


a
Event 1:19810620
2 m Ori
2 m Fill
5 m Ori
5 m Fill
10 m Ori
10 m Fill
Relative Drainage Area Connected C (m²/m²)
Relative Depression Storage Di =DVconnected/DSmax (m³/m³)
b
Event 2: 20060610
2 m Ori
2 m Fill
5 m Ori
5 m Fill
10 m Ori
10 m Fill
Relative Drainage Area Connected C (m²/m²)
Relative Depression Storage Di =DVconnected/DSmax (m³/m³)
c
Event 3: 20080903
2 m Ori
2 m Fill
5 m Ori
5 m Fill
10 m Ori
10 m Fill
Relative Drainage Area Connected C (m²/m²)
Relative Depression Storage Di =DVconnected/DSmax (m³/m³)

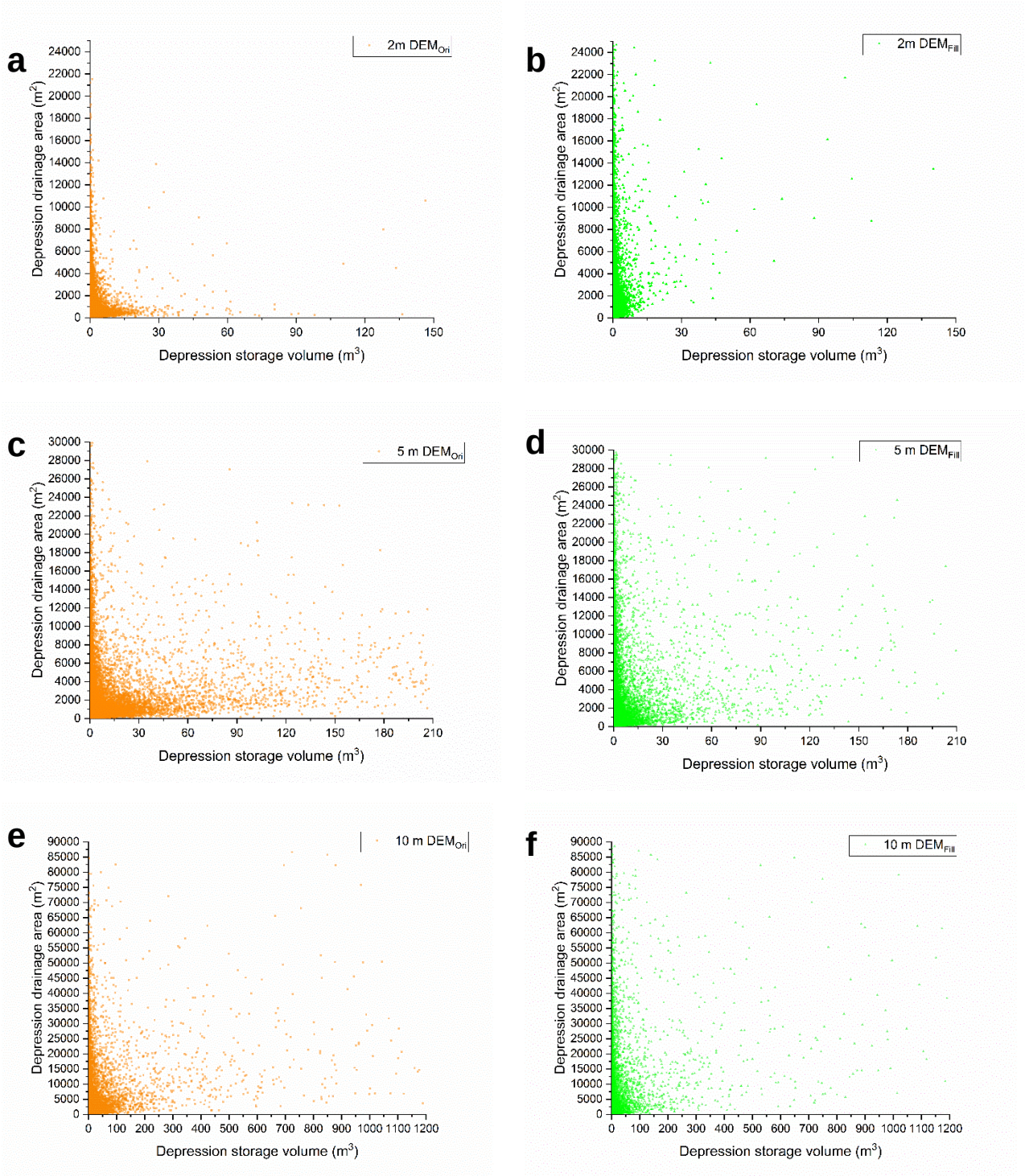


Figure 4.3 Correlations between depression storage volume and drainage area based on DEM_{Ori} (Orange square) and DEM_{Fill} (Green triangle). (a-b) 2 m resolution; (c-d) 5 m resolution; (e-f) 10 m resolution. Distortion of drainage areas are disproportionally small for depressions of large storage volume, and the degree of this distortion is increased after depression removal (filling).

paths and the structural connectivity characteristics (e.g., well-connected or not) between DEM_{Ori} and DEM_{Fill}.

D_{imax} increases with higher accumulated rainfall amounts for both DEM_{Ori} and DEM_{Fill} (Table 4.2). This means more depressions are being hydraulically connected to the watershed outlet for larger storm events. In this study, we assume that there is no erosion and thus the surface topography is static, which means that with more rain then more water is filling any depression storage and generating runoff. In this case, overflow from depressions becomes more frequent which expands the connected overland flow extent. However, the increase in C_i is much less than that of D_{imax}, regardless of the presence/absence of depressions (Fig. 4.1a-c). This indicates that, in general, depressions with larger storage volumes at GCEW have a much lower surface drainage areas, i.e., drainage areas of large volume depressions are proportionately smaller (Fig. 4.2a-f). The relationship between DA and DS matches previous work that established a the power-law relationship between surface drainage area and volume for depressions in naturally occurring topography (Abedini *et al.*, 2006; Le and Kumar, 2014). For all the three storm events and spatial resolutions, D_{imax} is higher for DEM_{Fill} than DEM_{Ori} because the removal (filling) of depressions reduces storage volumes mostly for depressions of large storage volume and thus the total storage capacity. In addition, D_{imax} increases with a decrease in grid resolution for both DEM_{Ori} and DEM_{Fill}. This is because depression storage volumes are artificially boosted in areas that are hydraulically connected to the watershed outlet at larger grid resolutions. Note that the magnitude of D_{imax} increases $RSCf_i$ do not reflect the storage deficit because the exact percentage of depressional storage being filled is unknown.

Table 4.2 Maximum depression storage volumes (D_{imax}) for DEM_{Ori} and DEM_{Fill} at different grid resolutions and cumulative rainfall amounts for each storm event.

DS_{max}	DEM_{Ori}			DEM_{Fill}		
Event year / Resolution	1981	2006	2008	1981	2006	2008
2 m	0.10	0.11	0.06	0.56	0.56	0.32
5 m	0.54	0.54	0.49	1.00	0.69	0.54
10 m	0.82	0.79	0.51	0.93	0.94	0.78
Cumulative Rainfall (mm)	143	115	61	143	115	61

The D_{imax} increase is larger for DEM_{Ori} than that for DEM_{Fill} with the increase of grid resolutions. Peñuela et al. (2013) examined border effects of changing width or length for a rectangular plot

scale on overland flow connectivity for different types of microtopography (crater, river, and random), and found that the sensitivity of the response is significantly different between width and length changes. In this study, because we only consider the watershed outlet as a border, changes of the grid resolution with a uniform increase/decrease of the width/length of cells can be considered as either width or length changes depending on the connectivity types of topography. GCEW has an eastern preferential flow direction at the watershed scale based on the spatial aspect of DEM (Fig. 4.3), while the dominant direction of the overland flow pathways, including stream flow, is North. In DEM_{Fill}, preferential flow directions at the watershed scale are less obvious because of large flat areas (Fig. 4.3). Therefore, the DEMs used in this study are a typical "river" type topography where the directions of preferential flow and the overland flow pathways are not compatible. Note that we consider the *increase* of grid resolution as similar to the *decrease* of both widths and lengths because of the similar effects of decreasing surface continuity in Peñuela et al., 2013. Increasing the grid resolution leads to a change in the weight of overland flow process between connectivity-driven and overflow-driven, i.e., more overflow is needed to connect the overland flow pathways to the watershed outlet to compensate for the loss of connected structural flow paths, thus a lower portion of total depression storage is being filled in the connected flow areas. Therefore, higher grid resolution is associated with lower $D_{imax.}$ Removal (filling) of depressions create large flat areas which dilute border effects (reduce the incompatibility between the directions of preferential and structural flow paths) so D_{imax} decreases more slowly.

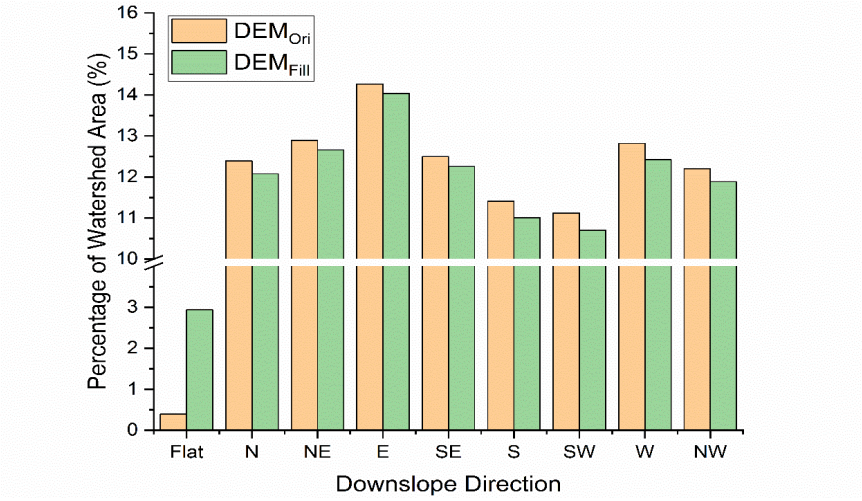


Figure 4.4 Distributions of downslope directions at GCEW for DEM_{ori} and DEM_{fill} based on aspect analysis in ArcGIS. The dominant direction is East which is perpendicular to the direction of major stream channels: North. Removal (Filling) of depressions "dilutes" the difference between downslope directions because there is a significant increase in flat area extents.

$NRCf_i$ function plots show two separated decreasing trends with obvious gaps, which represents a sudden "jump" to a higher $NRCf_i$ for all events, regardless of DEM_{Ori} or DEM_{Fill} (Fig. 4.4a-c). If infiltration excess is the major runoff generation process, $NRCf_i$ trends should monotonically decrease with D_i because the larger the depression storage volume that is included in the connected overland flow extent, the more rainfall will be contributing to filling depressions and the higher is the loss to infiltration. One explanation for the sudden "jump" is that this when the soil is fully wetted and depressions are filled within the connected area, so that any additional rainfall on the domain will rapidly increase the $NRCf_i$, i.e., the overland drainage system changes from fill dominated and high infiltration capacity (no ponding) to spill dominated and low infiltration capacity (with significant ponding). The decrease (slope) of $NRCf_i$ plots for DEM_{Ori} and DEM_{Fill} are very similar regardless of the spatial resolution, but to reach a similar magnitude of $NRCf_i$ (normalized discharge at the watershed outlet), a lower value of D_i (less depressions are included in the connected areas) is needed for DEM_{Ori} than DEM_{Fill}. This is expected because the total (and average) volumes of depression storage have been significantly decreased for DEM_{Fill} (Table 4.3). In addition, the gaps between these two trends show large discontinuity for D_i at 10 m resolution, which is not realistic because of the continuous extension of the overland flow extent in the storm events. Similar to RSCifs, NRCfis for DEM_{Fill} are more linear than DEM_{Ori}, but this is less pronounced.

In the past, watershed modeling often treated depression storage as a single and constant loss term, and as a function of total retention or excess rainfall instead of explicitly accounting for their spatial complexity, e.g., distributions and topological relationships. Thus, these approaches implicitly assume a spatial uniform distribution of depressions within a watershed boundary. The results of our study indicate that this assumption is not reasonable for either DEM_{Fill} or DEM_{Ori} given the statistical characteristics of depressions and the spatial non-linear $RSCf_i$ and $NRCf_i$ functional plots for single storm events. Moreover, grid resolution changes the spatial and statistical characteristics and functional connectivity for depressions at watershed scale. Therefore, for event-based simulations, a cautious selection of spatial resolution is vital in the parameterization, even if the only modeling goal is to simulate the streamflow discharge at the watershed outlet. In this study, a grid resolution smaller than 10 m is recommended.

Table 4.3 Statistical characteristics for depressions in DEM_{ori} and DEM_{fill}.

	2m	5m	10m
Total number of depressions in DEM_{ori}	371,984	82,885	14,444
Total volume of depressions in DEM_{ori} (m^3)	424,071	1,267,141	1,403,643
Average of DA in DEM_{ori}	193	865	4952
Average of DS in DEM_{ori}	1	15	97
Total number of depressions in DEM_{Fill}	154,091	45,709	8,729
Total volume of depressions in DEM_{Fill} (m^3)	34,077	139,032	658,863
Average of DA in DEM_{Fill}	464	1564	8174
Average of DS in DEM_{Fill}	.22	3	75

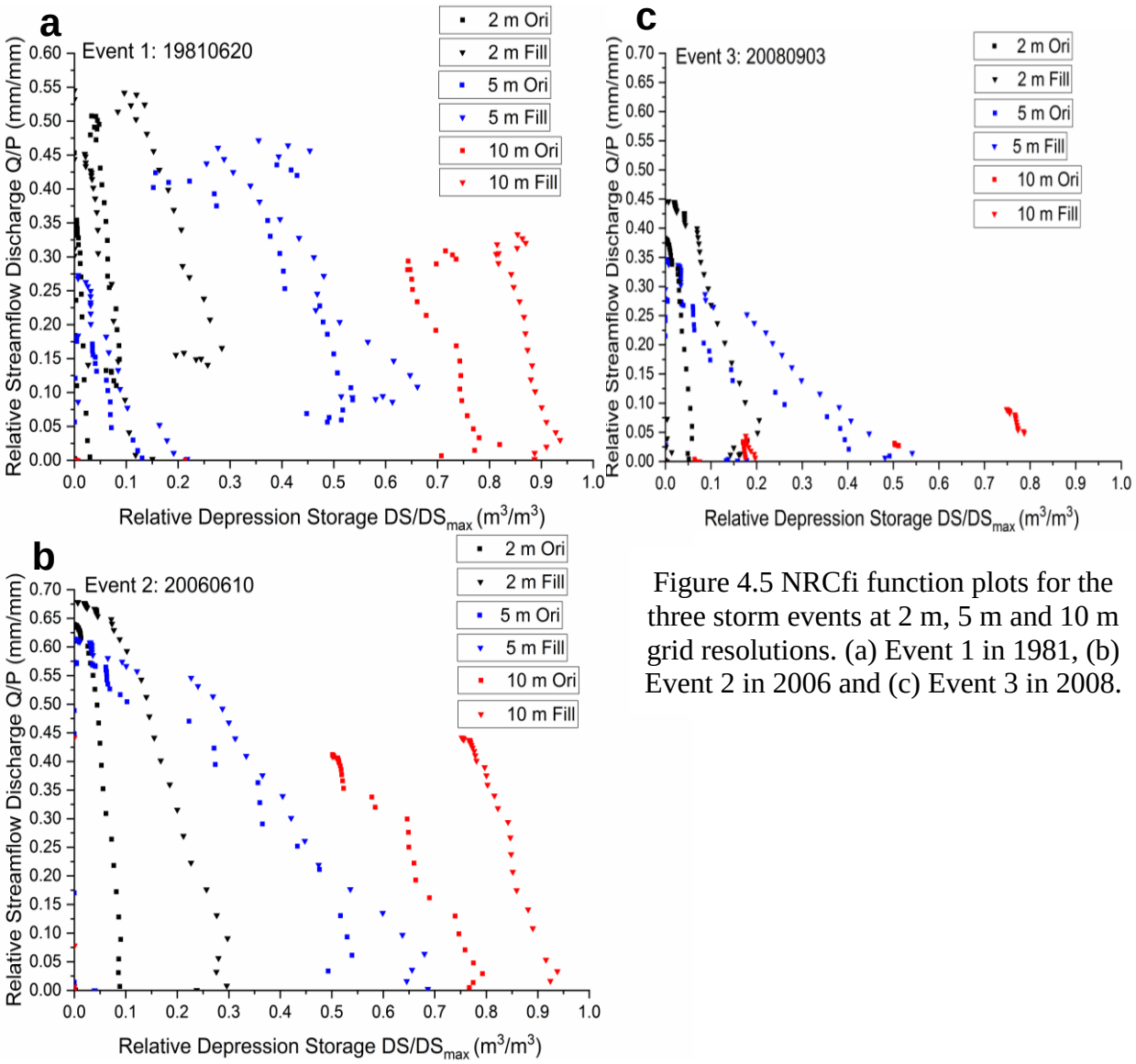


Figure 4.5 NRCfi function plots for the three storm events at 2 m, 5 m and 10 m grid resolutions. (a) Event 1 in 1981, (b) Event 2 in 2006 and (c) Event 3 in 2008.

4.5 Conclusions

Depressions change watershed scale connectivity and overland flow responses, and the effects of such changes in a model depends on the gird resolution of the DEM. This study demonstrated that revised functional connectivity indicators, $RSCf_i$ and $NRCf_i$ function plots, are important indicators for those changes and their interaction with grid resolution, without the need of an explicit computation of storage deficit at every modeling time step, which is not practical at watershed scale models based on high-resolution spatial discretization. Specifically, we found that:

The revised relative surface connection function ($RSCf_i$) for hydraulically connected drainage areas and depression storages is a functional connectivity indicator which represents the characteristics of spatial and statistical distributions for depressions, i.e., relative scales and relative topologic hierarchy, thus can be used to determine the change of topographic connectedness for overland flow based on connected depression storage at watershed scale. The convex shapes of $RSCf_i$ function plots (especially at 2 m resolution) for DEM_{Ori} indicate a well-connected topography with large amount of small depressions, which is confirmed with the statistical characteristics of depressions. Removal (filling) of depressions flattens these depressions and results in a modified topography that has smaller total storage volumes. Thus, RSCfis for DEM_{Fill} are more flattened and with lower slopes, reflecting a higher total depression storage volume in the overland flow extent that are connected to the watershed outlet but with a slower extension of the connected overland flow pathways. $RSCf_i$ for DEM_{Fill} also demonstrates the removal (filling) of depressions disproportionally reduce the depressions (high in the hierarchy) storage volumes to a higher degree because of the large “flattened” areas, while retains similar drainage areas as those in the original DEM.

In addition, larger storm events tend to increase the ratio of total depression storages (indicated by D_{imax}) in the overland flow extents that are hydraulically connected to the watershed outlet. However, RSCfi shows that the degree of increase for storage volumes is much higher than the increase for the drainage areas, regardless of DEM_{Ori} or DEM_{Fill}. This indicates that depressions with larger storage volumes at GCEW have a much lower total connected drainage area. Alternatively, changes of D_{imax} may also be linked to the connectivity types of the topography (river type for the study watershed), based on relative degree of overland flows mechanism on

topography, i.e., between connectivity-driven and overflow-driven. Under larger storm events and higher grid resolutions, the degree of increase of D_{imax} is larger than C_i. Note that a lower grid resolution means higher degree of continuity for the DEM ("smooth"). Therefore, the increase of connectivity-driven mechanism leads to a higher ratio of total depression storage D_{imax} that are being connected in the topographic structure but with less magnitude of increase of the overland flow extent connected to the outlet.

The revised normalized runoff connectivity function ($NRCf_i$) for stream discharge and depression storages is a functional connectivity indicator which represents the remaining rainfall (stream flow discharge out of topography) after the losses to the depression storage and infiltrations in the hydraulically connected drainage areas. Slopes of $NRCf_i$ determine the speed of rainfall that are being infiltrated and filled to depression storage based on a spatial explicit coverage of the overland flow extent. $NRCf_i$ function plots show a more linear trend for DEM_{Fill} than for DEM_{Ori}, indicating that the uncertainty of streamflow discharge because of depression storage is reduced after depressions have been filled. In addition, the "jumps" in the relative stream flow discharge that breaks the continuous trend of $NRCf_i$ function plots in to two parallel trends may indicate the thresholds of the fully filled depressions and the high ponding condition with in the overland flow extent connected to the watershed outlet. $NRCf_i$ function plots for both DEM_{Ori} or DEM_{Fill} also indicate that the discharge at 10 m resolution is unrealistic because of the sudden depression storage increase represented by the large gaps in D_i appeared between the two parallel trends. Therefore, to better represent the role of depressions on overland flow in watershed modeling, a grid resolution of less than 10 m is recommended.

Future work is needed to confirm the impact of grid resolution on the interaction between depressions and overland flow at watershed scales in different watershed environments, and to investigate the role of depressions on functional connectivity for subsurface flow processes. A high-performance, fully distributed and coupled modeling approach may be required that encompasses a complete surface-subsurface processes, where the contribution of subsurface flow to the watershed runoff response is not negligible.

4.6 References

Abedini MJ, Dickinson WT, Rudra RP. 2006. On depressional storages: The effect of DEM spatial resolution. *Journal of Hydrology* DOI: 10.1016/j.jhydrol.2005.06.010

Ahmed Z, Rao DRM, Reddy KRM. 2011. Sustainable storm water management - An evaluation of depression storage effect on peak flow. In *Proceedings of the International Conference on Green Technology and Environmental Conservation, GTEC-2011*336–340. DOI: 10.1109/GTEC.2011.6167690

Antoine M, Javaux M, Bielders C. 2009. What indicators can capture runoff-relevant connectivity properties of the micro-topography at the plot scale? *Advances in Water Resources* **32** (8): 1297–1310 DOI: 10.1016/j.advwatres.2009.05.006

Antoine M, Javaux M, Bielders CL. 2011. Integrating subgrid connectivity properties of the micro-topography in distributed runoff models, at the interrill scale. *Journal of Hydrology* **403** (3–4): 213–223 DOI: 10.1016/j.jhydrol.2011.03.027

Appels WM, Bogaart PW, van der Zee SEATM. 2011. Influence of spatial variations of microtopography and infiltration on surface runoff and field scale hydrological connectivity. *Advances in Water Resources* **34** (2): 303–313 DOI: 10.1016/j.advwatres.2010.12.003

Appels WM, Bogaart PW, van der Zee SEATM. 2016. Surface runoff in flat terrain: How field topography and runoff generating processes control hydrological connectivity. *Journal of Hydrology* **534**: 493–504 DOI: 10.1016/j.jhydrol.2016.01.021

Aryal SK, Bates BC. 2008. Effects of catchment discretization on topographic index distributions. *Journal of Hydrology* **359** (1–2): 150–163 DOI: 10.1016/j.jhydrol.2008.06.025

Baffaut C, Sadler EJ, Ghidey F. 2015. Long-Term Agroecosystem Research in the Central Mississippi River Basin: Goodwater Creek Experimental Watershed Flow Data. *Journal of Environment Quality* **44** (1): 18 DOI: 10.2134/jeq2014.01.0008

Band LE. 1986. Topographic Partition of Watersheds with Digital Elevation Models. *Water Resources Research* **22** (1): 15–24 DOI: 10.1029/WR022i001p00015

Barco J, Wong KM, Stenstrom MK. 2008. Automatic Calibration of the U.S. EPA SWMM Model for a Large Urban Catchment. *Journal of Hydraulic Engineering* **134** (4): 466–474 DOI: 10.1061/(ASCE)0733-9429(2008)134:4(466)

Berry PAM, Garlick JD, Smith RG. 2007. Near-global validation of the SRTM DEM using satellite radar altimetry. *Remote Sensing of Environment* **106** (1): 17–27

Beven K. 2012. Hydrological Similarity, Distribution Functions and Semi-Distributed Rainfall–Runoff Models. In *Rainfall-Runoff Modelling* John Wiley & Sons, Ltd; 185–229. DOI: 10.1002/9781119951001.ch6

Caddis BM, Jempson MA, Ball JE, Syme WJ. 2008. Incorporating hydrology into 2D hydraulic models-the direct rainfall approach. *9Th Conference on Hydraulics in Water Engineering* (September)

Campling P, Gobin A, Beven K, Feyen J. 2002. Rainfall-runoff modelling of a humid tropical catchment: The TOPMODEL approach. *Hydrological Processes* **16** (2): 231–253 DOI: 10.1002/hyp.341

Carrivick JL. 2006. Application of 2D hydrodynamic modelling to high-magnitude outburst floods: An example from Kverkfjoll, Iceland. *Journal of Hydrology* **321** (1–4): 187–199 DOI: 10.1016/j.jhydrol.2005.07.042

Clark KM, Ball JE, Babister MK. 2008. Can fixed grid 2D hydraulic models be used as hydrologic models. *Proceedings of Water Down Under*: 2496–2507

Correia F, Rego F. 1998. Coupling GIS with hydrologic and hydraulic flood modelling. *Water Resources …*: 229–249 DOI: 10.1023/a:1008068426567

Darboux F, Davy P, Gascuel-Odoux C. 2002a. Effect of depression storage capacity on overland-flow generation for rough horizontal surfaces: water transfer distance and scaling. *Earth Surface Processes and Landforms* **27** (2): 177–191 DOI: 10.1002/esp.312

Darboux F, Davy P, Gascuel-Odoux C, Huang C. 2002b. Evolution of soil surface roughness and flowpath connectivity in overland flow experiments. *CATENA* **46** (2–3): 125–139 DOI: 10.1016/S0341-8162(01)00162-X

DeVantier B a., Feldman AD. 1993. Review of GIS Applications in Hydrologic Modeling. *Journal of Water Resources Planning and Management* **119** (2): 246–261 DOI: 10.1061/(ASCE)0733-9496(1993)119:2(246)

Dunkerley DL. 2003. Determining friction coefficients for interrill flows: The significance of flow filaments and backwater effects. *Earth Surface Processes and Landforms* **28** (5): 475–491 DOI: 10.1002/esp.453

Einsiedl F. 2005. Flow system dynamics and water storage of a fissured-porous karst aquifer characterized by artificial and environmental tracers. *Journal of Hydrology* **312** (1–4): 312–321 DOI: 10.1016/j.jhydrol.2005.03.031

Erpicum S, Dewals B, Archambeau P, Detrembleur S, Pirotton M. 2010. Detailed Inundation Modelling Using High Resolution DEMs. *Engineering Applications of Computational Fluid Mechanics* **4** (2): 196–208 DOI: 10.1080/19942060.2010.11015310

ESRI Environmental Systems Research Institute. 2016. ArcGIS Desktop: Release 10.4. *Redlands CA*

Fabio P, Aronica GT, Apel H. 2010. Towards automatic calibration of 2-D flood propagation models. *Hydrology and Earth System Sciences* **14** (6): 911–924 DOI: 10.5194/hess-14-911-2010

Ghaffari G. 2011. The impact of DEM resolution on runoff and sediment modelling results. *Research Journal of Environmental Sciences* **5** (8): 691–702

Goderniaux P, Brouyère S, Fowler HJ, Blenkinsop S, Therrien R, Orban P, Dassargues A. 2009. Large scale surface-subsurface hydrological model to assess climate change impacts on groundwater reserves. *Journal of Hydrology* **373** (1–2): 122–138 DOI: 10.1016/j.jhydrol.2009.04.017

Green WH, Ampt GA. 1911. Studies on Soil Phyics. *The Journal of Agricultural Science* **4** (1): 1–24

Grimm K, Chu X. 2018. Modeling of spatiotemporal variations in runoff contribution areas and analysis of hydrologic connectivity. *Land Degradation & Development* **29** (8): 2629–2643

Hall J. 2015. Direct rainfall flood modelling: The good, the bad and the ugly. *Australian Journal of Water Resources* **19** (1): 74–85

Hancock GR. 2005. The use of digital elevation models in the identification and characterization of catchments over different grid scales. *Hydrological Processes* **19** (9): 1727–1749 DOI: 10.1002/hyp.5632

Hirano A, Welch R, Lang H. 2003. Mapping from ASTER stereo image data: DEM validation and accuracy assessment. In *ISPRS Journal of Photogrammetry and Remote Sensing*356–370. DOI: 10.1016/S0924-2716(02)00164-8

Homer CG, Dewitz JA, Yang L, Jin S, Danielson P, Xian G, Coulston J, Herold ND, Wickham JD, Megown K. 2015a. Completion of the 2011 National Land Cover Database for the conterminous United States-Representing a decade of land cover change information. *Photogrammetric Engineering and Remote Sensing* **81** (5): 345–354 DOI: 10.14358/PERS.81.5.345

Homer CG, Dewitz JA, Yang L, Jin S, Danielson P, Xian G, Goulston J, Herold ND, Wickham JD, Megown K. 2015b. Completion of the 2011 National Land Cover Database for the Conterminous United States – Representing a Decade of Land Cover Change Information. *Photogrammetric Engineering and Remote Sensing* **81** (5): 345–354 Available at: https://cfpub.epa.gov/si/si_public_record_report.cfm?dirEntryId=309950

Horritt MS, Bates PD. 2002a. Evaluation of 1D and 2D numerical models for predicting river flood inundation. *Journal of Hydrology* **268** (1–4): 87–99 DOI: 10.1016/S0022-1694(02)00121-X

Horritt MS, Bates PD. 2002b. Evaluation of 1D and 2D numerical models for predicting river flood inundation. *Journal of Hydrology* **268** (1–4): 87–99 DOI: 10.1016/S0022-1694(02)00121-X

Horritt MS, Bates PD, Mattinson MJ. 2006. Effects of mesh resolution and topographic representation in 2D finite volume models of shallow water fluvial flow. *Journal of Hydrology* **329** (1–2): 306–314 DOI: 10.1016/j.jhydrol.2006.02.016

Hunter NM, Bates PD, Neelz S, Pender G, Villanueva I, Wright NG, Liang D, Falconer RA, Lin B, Waller S, et al. 2008. Benchmarking 2D hydraulic models for urban flood simulations. *Proceedings of the Institution of Civil Engineers: Water Management* **161** (1): 13–30 DOI: doi: 10.1680/wama.2008.161.1.13

Hutchinson MF. 1989. A new procedure for gridding elevation and stream line data with automatic removal of spurious pits. *Journal of Hydrology* **106** (3–4): 211–232 DOI: 10.1016/0022-1694(89)90073-5Huxley C, Syme B. 2016. TUFLOW GPU-Best practice advice for hydrologic and hydraulic model simulations. In *37th Hydrology & Water Resources Symposium 2016: Water, Infrastructure and the Environment*Engineers Australia; 195.

I. D. Moore ID, C. L. Larson CL. 1979. Estimating Micro-Relief Surface Storage from Point Data. *Transactions of the ASAE* **22** (5): 1073–1077 DOI: 10.13031/2013.35158

IBM SPSS Inc. 2017. SPSS Statistics for Windows, version 24.0. *IBM Corp. Released 2017*

Jarihani AA, Callow JN, McVicar TR, Van Niel TG, Larsen JR. 2015. Satellite-derived Digital Elevation Model (DEM) selection, preparation and correction for hydrodynamic modelling in large, low-gradient and data-sparse catchments. *Journal of Hydrology* **524**: 489–506

Jenson SK, Domingue JO. 1988. Extracting topographic structure from digital elevation data for geographic information system analysis. *Photogrammetric Engineering and Remote Sensing* **54** (11): 1593–1600 DOI: 0099-1112/88/5411-1593$02.25/0

K. Ajami N, Gupta H, Wagener T, Sorooshian S. 2004. Calibration of a semi-distributed hydrologic model for streamflow estimation along a river system. In *Journal of Hydrology*112–135. DOI: 10.1016/j.jhydrol.2004.03.033

Kalyanapu AJ, Shankar S, Pardyjak ER, Judi DR, Burian SJ. 2011. Assessment of GPU computational enhancement to a 2D flood model. *Environmental Modelling & Software* **26** (8): 1009–1016

Kiesel J, Fohrer N, Schmalz B, White MJ. 2010. Incorporating landscape depressions and tile drainages of a northern German lowland catchment into a semi-distributed model. *Hydrological Processes* **24** (11): 1472–1486 DOI: 10.1002/hyp.7607

Lane SN, Brookes CJ, Kirkby MJ, Holden J. 2004. A network-index-based version of TOPMODEL for use with high-resolution digital topographic data. *Hydrological Processes* **18** (1): 191–201 DOI: 10.1002/hyp.5208

Le PV V., Kumar P. 2014. Power law scaling of topographic depressions and their hydrologic connectivity. *Geophysical Research Letters* DOI: 10.1002/2013GL059114

Lindsay JB, Creed IF. 2005. Removal of artifact depressions from digital elevation models: Towards a minimum impact approach. *Hydrological Processes* **19** (16): 3113–3126 DOI: 10.1002/hyp.5835

Lindsay JB, Creed IF. 2006. Distinguishing actual and artefact depressions in digital elevation data. *Computers & Geosciences* **32** (8): 1192–1204 DOI: 10.1016/j.cageo.2005.11.002

Liu X. 2008. Airborne LiDAR for DEM generation: some critical issues. *Progress in Physical Geography* **32** (1): 31–49 DOI: 10.1177/0309133308089496

Metz M, Mitasova H, Harmon RS. 2011. Efficient extraction of drainage networks from massive, radar-based elevation models with least cost path search. *Hydrology and Earth System Sciences* **15** (2): 667–678 DOI: 10.5194/hess-15-667-2011

Miller S, Semmens D. 2002. GIS-based hydrologic modeling: the automated geospatial watershed assessment tool. *Proceeding of the Second Federal Interagency Hydrologic Modeling Conference*: 1–12 Available at: http://www.epa.gov/nerlesd1/land-sci/agwa/pdf/pubs/agwa-conference.pdf

O'Callaghan JF, Mark DM. 1984. The extraction of drainage networks from digital elevation data. *Computer Vision, Graphics, and Image Processing* **27** (2): 247 DOI: 10.1016/S0734-189X(84)80047-X

Pappenberger F, Frodsham K, Beven KJ, Romanovicz R, Matgen P. 2007. Fuzzy set approach to calibrating distributed flood inundation models using remote sensing observations. *Hydrology and Earth System Sciences* **11** (2): 739–752 DOI: 10.1029/2001WR001056

Patro S, Chatterjee C, Mohanty S, Singh R, Raghuwanshi NS. 2009. Flood inundation modeling using MIKE FLOOD and remote sensing data. *Journal of the Indian Society of Remote Sensing* **37** (1): 107–118 DOI: 10.1007/s12524-009-0002-1

Peñuela A, Darboux F, Javaux M, Bielders CL. 2016. Evolution of overland flow connectivity in bare agricultural plots. *Earth Surface Processes and Landforms*: n/a-n/a DOI: 10.1002/esp.3938

Peñuela A, Javaux M, Bielders CL. 2013. Scale effect on overland flow connectivity at the plot scale. *Hydrology and Earth System Sciences* **17** (1): 87–101 DOI: 10.5194/hess-17-87-2013

Peñuela A, Javaux M, Bielders CL. 2015. How do slope and surface roughness affect plot-scale overland flow connectivity? *Journal of Hydrology* **528**: 192–205 DOI: 10.1016/j.jhydrol.2015.06.031

Pfeifer N. 2001. Derivation Of Digital Terrain Models In The Scop++ Environment. *OEEPE Workshop on Airborne Laserscanning and Interferometric SAR for Digital Elevation Models, ,*: 13 Available at: http://www.lidar.com.br/arquivos/dem-scop-laser.pdf

Pfeifer N, Mandlburger G. 2008. LiDAR Data Filtering and DTM Generation. In *Topographic Laser Ranging and Scanning*CRC Press; 307–334. DOI: doi:10.1201/9781420051438.ch11

Phillips BC, Yu S, Thompson GR, Silva N De. 2005. 1D and 2D Modelling of Urban Drainage Systems using XP-SWMM and TUFLOW. *10th International Conference on Urban Drainage* (August): 21–26

Planchon O, Darboux F. 2002. A fast, simple and versatile algorithm to fill the depressions of digital elevation models. In *Catena*159–176. DOI: 10.1016/S0341-8162(01)00164-3

Rawls WJ, Brakensiek DL, Miller N. 1983. Green-ampt Infiltration Parameters from Soils Data. *Journal of Hydraulic Engineering* **109** (1): 62–70 DOI: 10.1061/(ASCE)0733-9429(1983)109:1(62)

Rawls WJ, Brakensiek DL, Saxton KE. 1982. Estimation of Soil Water Properties. *Transactions of the ASAE* **25** (5): 1316–1320 & 1328 DOI: 10.13031/2013.33720

Sadler EJ, Lerch RN, Kitchen NR, Anderson SH, Baffaut C, Sudduth KA, Prato AA, Kremer RJ, Vories ED, Myers DB, et al. 2015a. Long-Term Agroecosystem Research in the Central Mississippi River Basin: Introduction, Establishment, and Overview. *Journal of Environment Quality* **44** (1): 3 DOI: 10.2134/jeq2014.11.0481

Sadler EJ, Sudduth KA, Drummond ST, Vories ED, Guinan PE. 2015b. Long-Term Agroecosystem Research in the Central Mississippi River Basin: goodwater creek experimental watershed weather data. (Special Section: LTAR in the Central Mississippi river basin.). *Journal of Environmental Quality* **44**: 13–17 Available at: https://www.crops.org/publications/jeq/articles/44/1/13

Shan J, Toth CK. 2008. *Topographic Laser Ranging and Scanning: Principles and Processing*. DOI: 10.1201/9781420051438

Siart C, Bubenzer O, Eitel B. 2009. Combining digital elevation data (SRTM/ASTER), high resolution satellite imagery (Quickbird) and GIS for geomorphological mapping: A multi-component case study on Mediterranean karst in Central Crete. *Geomorphology* **112** (1–2): 106–121 DOI: 10.1016/j.geomorph.2009.05.010

Sibson, R. 1981. A Brief Description of Natural Neighbor Interpolation. In: Barnett, V., Ed., *Interpreting Multivariate Data*, John Wiley & Sons, New York, 21-36.

Smith RAE, Bates PD, Hayes C. 2012. Evaluation of a coastal flood inundation model using hard and soft data. *Environmental Modelling & Software* **30**: 35–46 DOI: 10.1016/j.envsoft.2011.11.008

Soil Survey Staff. 2015. Soil Survey Geographic (SSURGO) Database for [U.S.]. *United States Department of Agriculture accessed [04/16/2015]*: 1–22 Available at: http://soildatamart.nrcs.usda.gov

Soil Survey Staff. 2017. Soil Survey Staff, Natural Resources Conservation Service, United States Department of Agriculture, Gridded Soil Survey Geographic (gSSURGO) database for Missouri

Steiner JL, Sadler EJ, Wilson G, Hatfield JL, James D, Vandenberg B, Chen J-S, Oster T, Ross JD, Cole K. 2009. STEWARDS watershed data system: System design and implementation. *Transactions of the ASABE* **52** (5): 1523–1533

Syme W. 2001. TUFLOW . Two & onedimensional Unsteady FLOW Software for Rivers , Estuaries and Coastal Waters. *IEAust Water Panel Seminar and Workshop on 2d Flood Modelling*: 2–9

Tarboton DG, Bras RL, Rodriguez-Iturbe I. 1991. On the extraction of channel networks from digital elevation data. *Hydrological Processes* **5** (1): 81–100 DOI: 10.1002/hyp.3360050107

Toutin T. 1995. Generating DEM from stereo images with a photogrammetric approach: Examples with VIR and SAR data. *EARSeL Advances in Remote Sensing* **4** (2): 110–117

Trimble Navigation Limited. 2016. Trimble SCOP++ Software Available at: http://trl.trimble.com/docushare/dsweb/Get/Document-789933/022516-022A_Inpho_SCOP++_TS_A4_0915_LR.pdf [Accessed 25 February 2016]

USDA-NRCS. 2012. Long Branch and South Fork Salt Creek Watersheds Project Database. *U.S. Department of Agriculture Natural Resource Conservation Service*. Available at http://www.msdis.missouri.edu/data/lidar/ [Accessed 30 January 2017]

Wallin S, Johansson A V. 2002. Modelling streamline curvature effects in explicit algebraic Reynolds stress turbulence models. *International Journal of Heat and Fluid Flow* **23** (5): 721–730 DOI: 10.1016/S0142-727X(02)00168-6

Wechsler SP. 2007. Uncertainties associated with digital elevation models for hydrologic applications: a review. *Hydrol. Earth Syst. Sci* **11**: 1481–1500 Available at: www.hydrol-earth-syst-sci.net/11/1481/2007/ [Accessed 8 June 2016]

Wilson MD, Bates P, Alsdorf D, Forsberg B, Horritt M, Melack J, Frappart F, Famiglietti J. 2007. Modeling large-scale inundation of Amazonian seasonally flooded wetlands. *Geophysical Research Letters* **34** (15) DOI: 10.1029/2007GL030156

Wolock DM, Price C V. 1994. Effects of digital elevation model map scale and data resolution on a topography based watershed model. *Water Resources Research* **30** (11): 3041–3052 DOI: 10.1029/94WR01971

Wu S, Li J, Huang GH. 2008. A study on DEM-derived primary topographic attributes for hydrologic applications: Sensitivity to elevation data resolution. *Applied Geography* **28** (3): 210–223 DOI: 10.1016/j.apgeog.2008.02.006

Yang J, Chu X. 2013. Effects of DEM Resolution on Surface Depression Properties and Hydrologic Connectivity. *Journal of Hydrologic Engineering* **18** (9): 1157–1169 DOI: 10.1061/(ASCE)HE.1943-5584.0000731

Yang J, Chu X. 2015. A new modeling approach for simulating microtopography-dominated, discontinuous overland flow on infiltrating surfaces. *Advances in Water Resources* **78**: 80–93 DOI: 10.1016/j.advwatres.2015.02.004

Yang P, Ames DP, Fonseca A, Anderson D, Shrestha R, Glenn NF, Cao Y. 2014. What is the effect of LiDAR-derived DEM resolution on large-scale watershed model results? *Environmental Modelling and Software* **58**: 48–57 DOI: 10.1016/j.envsoft.2014.04.005

Zhang W, Cundy TW. 1989. Modeling of Two-Dimensional Overland Flow. *Water Resources Research* **25** (9): 2019–2035 DOI: 10.1029/WR025i009p02019

Zhang W, Montgomery DR. 1994. Digital elevation model grid size, landscape representation, and hydrologic simulations. *Water Resources Research* **30** (4): 1019–1028 DOI: 10.1029/93WR03553

CHAPTER 5. CONCLUSIONS, IMPLICATIONS, AND FUTURE WORK

This Ph.D. study applied a spatial pattern and connectivity approach to investigate the impact of topographic depressions and topographic depression removal (filling) on overland flow response across multiple spatial scales. A new algorithm for a widely used two-point connectivity statistic was created to enable upscaling to large grids (i.e., with > 1,000,000 cells) and generate results within minutes or days (depending on the grid size), using GPU-accelerated computing techniques. A directional filter for the cardinal and intercardinal directions were incorporated to the 2D omnidirectional connectivity function. In addition, a more sensitive and robust algorithm for 3D directional connectivity function was proposed which is able to compute connectedness of overland flow patterns under the topographic constraint represented by a DEM with depressions. Hypothetical and real landscapes were used in the connectivity analysis for a theoretical test and a case study, respectively. The case study focused on a low-relief agricultural watershed of 75 km^2, located in Northern Missouri. Surface water is the major focus of this study. DEM was computed directly from aerial LiDAR point cloud data. Spatial patterns of overland flow (depth) at hourly interval for single storm events were generated using a high-performance 2D hydraulic model with input parameters that included DEM, rainfall, and soil survey data. The connectivity algorithm was then implemented on the overland flow patterns for the study area to derive the variability of connectivity at different spatial scales and time points. Two other functional connectivity indicators which are sensitive to the effects of depressions and changes of grid resolutions were also introduced to further explore the role of depressions on overland flow connectivity as a function of resolution. Major findings of this research include:

1. The connectivity function $\tau(h)$ and integral connectivity scale lengths (ICSLs, including Omnidirectional and Directional metrics: OMNI, CARD and TOPO) provide a simple and effective way of measuring connectivity for hydrologic patterns at multiple spatial scales and time points. $\tau(h)$ and ICSLs can be used as a unified connectivity indicator to directly compare hydrologic patterns generated under a wide range of watershed environments and hydroclimatic regimes.

2. Removal (filling) of topographic depressions increases overland flow pattern connectivity in most cases, and the degree of increase depends on the spatial scales (represented by lag distances) of the locations being considered. Maximum connectivity increases are typically found at small scales (low lag distances), in particular at scales on the order of the scales of depressions. As the scale increases, the difference in connectivity caused by depressions rapidly decreases. Thus, four macro connectivity stages are identified to differentiate different levels of changes in connectivity related to depressions. Fill and spill processes associated with depressions also have an impact on the degree of the connectivity changes, with spill process maintaining or increasing connectivity, and fill process delaying or reducing connectivity. Fill often occurs in the early phase (rising limb) of a storm event and increases connectivity changes between overland flow pattern based on original and depression removed DEM. In contrast, spill often occurs in the late phase (falling limb) of a storm event and reduces connectivity changes and at some cases even reverses the role of depressions by slightly increasing connectivity. In addition, a period of rapid increase of connectivity at large distances in the late phase was found to be associated with the relative speed of fill and formation of connected pathways for depressions and their condensed locations in floodplain areas immediately adjacent to stream.

3. Topological characteristics, i.e., the hierarchical structures of depressions, are linked to the speed of transition between the macro connectivity stages mentioned above.

4. Changes of connectivity because of depressions removal (filling) are not uniformly distributed in the study watershed. Increase of connectivity are spatially concentrated in the riparian zone (areas of floodplain immediately adjacent to the stream), where a large number of small depressions are found. Depression removal created large smoothed (flattened) topography in riparian zone and it is possible that connectivity is overestimated.

5. The revised relative surface connection function (RSCfi), describes depression drainage areas as a function of storage volumes, can be used to identify detailed characteristics of spatial and statistical distributions for depressions, i.e., scales, spatial uniformity, and topologic hierarchy. RSCfi can also identify connectivity types related to the relative degree of overland flow connectedness on topography between connectivity-driven and overflow-driven at a watershed scale. The revised normalized runoff connectivity function

($NRCf_i$), describes normalized stream discharge as a function of storage volumes, may relate to the threshold of fully filled depression storage and highly ponding condition at a watershed scale. A grid resolution below 10 m is recommended to better represent the interaction mechanisms between depressions and overland flow in modeling.

Depressions change overland flow timing and magnitude in response to rainfall events, and this change may only be seen through an explicit quantification of connectivity for a hydrologic system based on spatio-temporal variability of process patterns. This study indicates that the interaction mechanism between depressions and overland flow, e.g., fill and spill, and the hierarchical structures of depressions, provides the physical foundation for connectivity changes in overland flow patterns. These changes are a direct indicator for heterogeneity of multi-scale functional responses, e.g., overland flow, which cannot be reflected in implicit analysis, such as single point stage or discharge at a watershed outlet. Note that overland flow connectivity changes because of depressions are most sensitive at small lag-distance, e.g., hillslope scale, but the impacts are spatially aggregated and extended to watershed scale. Modelers should be cautious about the "unintended consequences" of the blind removal of depressions because of changes in response heterogeneity. For example, a filled DEM can compensate the uncertainty of other modeling inputs or calibration parameters e.g., underestimated rainfall, baseflow or soil moisture condition at hillslope scale, which may provide a false impression of high model accuracy. The calibrated parameters may not be robust for storm events of significantly higher/lower magnitude or other initial watershed conditions because of the heterogeneous shifts in the connected depressions that are being included the overland flow patterns and their interaction mechanisms. Changes of connectivity internally in a hydrologic system because of depression removal also has implications on the transfer of water flux, energy and water mediated substances based on the connectedness of the overland flow pathways, which are critical for ecological and environmental functions and related decision-making in a watershed.

Future research to build on the work presented here should focus on eliminating the need for computationally intensive and fully distributed 2D hydraulic modeling as the prerequisite for generating overland flow patterns, because of the extensive simulation time required on a high-resolution DEM grid. An alternative approach could create a simplified mass and momentum

computational scheme for excesses and losses between the two connected points into the connectivity metrics, or could involve development and implementation of more efficient ways of collecting overland flow data across large spatial scales, e.g., real-time water surface detection based on remotely sensed radar data. Connectivity is a critical predictor for water transfer and water mediated substances transfer in the riparian zone, but the results of this work indicates that connectivity may be overestimated in highly vegetated riparian zones because of increased DEM errors. This needs to be examined in more detail, and coupled to explicit modeling of biogeochemical or sedimentation processes depending on connected flow pathways and depressions storage, so the effects of connectivity overestimation and possible implications for watershed management practices, i.e., the assimilation and transformation of point and non-point source pollution for downstream water quality control, can be better understood. Therefore, it would be an important first step to include the interaction mechanisms between overland flow and depressions directly into watershed modeling conceptualization based on connectivity. Moreover, filling is the only method that has been used in this study to remove depression in DEM, and other depression removal methods and their corresponding effects on overland flow could also be investigated. Finally, more detailed work is needed to confirm an optimum resolution below 10 m and to investigate the impacts of subsurface flow processes on connectivity.

APPENDIX A. DIRECTION RAINFALL METHOD, 2D HYDRAULIC MODEL SELECTION AND THE GPU-ACCELERATED CONNECTIVITY FUNCTION ALGORITHM

A.1 Brief introduction to the direct rainfall method (DRM)

A.1.1 Origin

Hydrologic and hydraulic modeling both follow the same workflow. Two key stages are typically included: 1) Hydrologic models transfer a rainfall hyetograph to a runoff hydrograph by subtracting losses and propagating the peaks and shapes (Hawkins *et al.*, 2008; Douglas-Mankin *et al.*, 2010) and then, 2) the runoff hydrograph is routed as overland and channel flows using hydraulic models, either empirically lumped, i.e., linear or nonlinear storage changes based on the continuity equation, or physically based using distributed routing methods, i.e., dynamic unsteady flow equations based on conservation laws of mass, momentum and energy, until the flow reaches the watershed outlet (Ponce and Yevjevich, 1978; Refsgaard and Storm, 1995; Singh, 1995; Rehman *et al.*, 2003). Note that all rainfall-runoff models, are essentially simplified and numerical representations of a much more complex reality. The simplification approach largely depends on modelers' own understanding of the hydrological processes watershed or studied region, often through knowledge, experience, and/or experiments, which is termed "model conceptualization" (Moradkhani and Sorooshian, 2009; Beven, 2011). In addition, specific model applications play a large part in decisions around model conceptualization (Moradkhani and Sorooshian, 2009). For hydraulic modeling, conceptualizing the hydrologic model component (in the 1st step mentioned above) is often a time consuming and labor-intensive endeavor in terms of finding a useful abstraction of reality, while projects are usually under significant budget and time constraints (Hunter *et al.*, 2007). For example, such hydrologic models may need to include different runoff generation mechanisms and be robust in different geographic and climate regimes (Boughton and Droop, 2003). In addition, these hydrologic models often require extensive model calibration to derive useful parameters (Eckhardt and Arnold, 2001). If the majority of these conceptualizations and calibrations can be eliminated, this will save a significant amount of project expense and time. For this purpose, hydraulic modelers focused on engineering applications created the so-called "Direct Rainfall Method (DRM)" to eliminate major conceptualization tasks, and have used this approach to solve engineering problems for more than a decade (Taaffe *et al.*, 2011; Hall, 2015).

A.1.2 Definition of DRM

DRM, also known as "rainfall on grid", partially or completely removes the need to use the hydrologic model component by applying rainfall directly to the computational grid. In practice, this means that the rainfall term is added directly to the shallow water equations (SWEs) (Rehman, 2011). DRM then directly routes overland flow on terrain topography based on hydrodynamics instead of linear or nonlinear storage functions, and can be directly applied without relying on the lumping schemes based on hydrologic similarity, e.g., hydrological response units (HRUs) (Singh, 1995; Rehman, 2011). DRM calculates rainfall losses using identical models as in lumped or distributed hydrologic models, e.g., Green-Ampt for infiltration (Green and Ampt, 1911) and Penman-Monteith for evapotranspiration (Priestley and Taylor, 1972), but the loss function is applied simultaneously with overland flow hydrodynamics instead of being directly removed from rainfall before it is applied to the computational grid (Taaffe *et al.*, 2011). Additionally, surface runoff in the form of dynamic unsteady flow is routed through the entire watershed based on topography, including floodplains and main stream channels. Thus, detailed information of flow rate, water depth, and velocity at each grid in the modeling domain are available, which is similar to flood inundation models but at higher spatial and temporal resolutions(Rehman *et al.*, 2003). Currently, high-resolution remote sensing data, e.g., LiDAR point clouds from the Leica scanner, can generate a DEM for use in DRM at a spatial resolution of approximately 0.5 m for large-scale watersheds with a vertical accuracy of ±0.05 m (Popescu, 2007; Collin and Hench, 2012). This allows for a sufficient representation of small morphologic features in topography, e.g., depressions, and thus can be incorporated into modeling using DRM for flow and drainage connectivity study (Jones *et al.*, 2008).

DRM requires flow routing in the entire modeling domain, e.g., a watershed, and might also apply to higher spatial and temporal resolutions than a typical flood inundation model. The computational intensity is much higher, and a longer run time is expected. This may seem to "counteract" the time saved by eliminating hydrologic components (Taaffe *et al.*, 2011). However, this issue is largely diminished by the recent emergence of high-performance computing techniques, especially Graphic Processing Unit (GPU) acceleration (Kalyanapu *et al.*, 2011; Vacondio *et al.*, 2014; Le *et al.*, 2015). GPU accelerated computation has been used in a wide range of disciplines in addition to computational hydrology, including machine learning/deep

learning, and big data analysis (Bleiweiss, 2008; Napoli *et al.*, 2014; Abadi *et al.*, 2016). GPU uses heavily paralleled computational units on a single device and can simulate large models (> 1,000,000 cells) within a reasonable timeframe, here defined as a modeling run time at the scale between 0.001 to 1 of the actual simulated event time duration, depending on the spatial resolution of computational units (usually from 100 m to 1 m) and the number of parallel GPU devices (Huxley and Syme, 2016). DRM was originally designed to speed up flood modeling practice for both fast decision-making, such as in emergency response, hydraulic design of flood prevention and dam breaking experiments reduce costs and time. In the research presented here we use DRM to aid in better understanding of watershed hydrologic processes, especially unsolved questions related to hydrologic connectivity across the watershed.

A.1.3 Comparison with lumped or semi-distributed hydrologic models

The appeal of DRM includes 1) a better representation of minor or ephemeral flowpaths than conventional lumped hydrologic models; 2) flow depth, velocity, and discharge at every computational cell for each timestep is available in the entire modeling domain; 3) flows cross watershed boundary can also be simulated (Caddis *et al.*, 2008). High-resolution, spatially and temporally explicit overland flow patterns for stream channels, riparian zones, and floodplain areas on a watershed scale are key to improving our understanding of hydrologic connectivity, similarity and the “scaling issue” for complex environmental systems on a heterogeneous landscape (Lexartza-Artza and Wainwright, 2009; Beven, 2012). This approach is particularly suitable for the purpose of this study: quantifying variability of connectivity and continuity using those patterns based on watershed topography. In addition, although many software packages (of different computational capacity) that can implement DRM exist, the results of these packages will have less variations than lumped or distributed hydrologic models because their conceptualizations are essentially the same: all based on SWEs but only differing in their numerical solution schemes (Caddis *et al.*, 2008; Rehman, 2011). The "scaling" issue of measured hydrologic data in a heterogeneous landscape can be better represented and investigated because DRM uses such a unified model conceptualization. In contrast, lumped or distributed models are based on different model conceptualization and use coarse spatial discretization, which would make it difficult to separate the uncertainty from the input data versus model conceptualization.

Most importantly, DRM enabled us to investigate the effect of depressions through the DRM dynamic loss function (Taaffe *et al.*, 2011). For lumped or semi-distributed hydrologic models, water loss are assumed to be constant or linearly correlated with rainfall or storage amount, and subtracted from rainfall before any further modeling computation is for the overland flow (e.g., Mishra, S.K. and Singh, 2003). However, for DRM, part of the water loss occurs simultaneously with overland flow modeling and directly on the surface topography, i.e., when water enters and is "retained" in these depressions. Part of the "trapped" water will not be routed to the watershed outlet. Thus, depressions act as part of the loss function, while the remaining “run-on” overflow continues to move across the modeling domain.

A.1.4 Comparison with fully distributed hydrologic models

Some fully coupled and distributed hydrologic models also use SWEs, for surface water or overland flow component, e.g., MIKE SHE, HydroGeoSphere and Vflo (Abbott *et al.*, 1986; Vieux and Ceo, 2002; Brunner and Simmons, 2012). However, these fully distributed models mostly make approximations of SWEs to reduce the computational intensity, assuming that one or more terms in the momentum equation can be neglected (Singh, 1996). The approximation was the same for the 1D/2D hydraulic models created before the mid-1990s but today’s hydraulic models rarely use such an approximation because of the need to reduce error at higher spatial resolution and the improvement of computer processing power. For a fully distributed hydrologic model, however, simplified SWEs are more practical in terms of computational feasibility because they also include other components in the hydrologic cycle, e.g., subsurface flow and groundwater flow in addition to surface water flow, and the interface between these components (Singh, 1995). The two most commonly used SWE approximations are Kinematic wave or diffusion wave. Kinematic wave or diffusion wave approximations can potentially cause limitations under a wide range of flow conditions (De Roo *et al.*, 2000), because they only consider downstream or downslope wave movement. Therefore, turbulence at a downstream boundary is not represented (Tayfur *et al.*, 1993). Although diffusive waves can be an alternative in these situations, prior research has found that convergence is often an issue even at small time step (Vieira, 1983).

In addition, fully distributed models require a large number of modeling parameters which means intensive in-situ data collection, and pre- and post-processing work. These are usually beyond a

project budget and time frame, especially for large watersheds (Refsgaard and Storm, 1995). The consequence of "overparameterization" often leads to a "blind" selection of parameters, blur their physical basis and thus weaken the importance of a few determinant physical parameters (Beven, 1996). Another issue is that different parameter sets can reach a similar goodness of fit to observations but the quality of these sets cannot be verified and compared against the characteristics of physical processes, a problem termed "equifinality" (Beven, 2006). These issues also create challenges for communicating with water resource management practitioners and policy makers outside of the hydrologic community (Tang *et al.*, 2005). In contrast, DRM uses readily available data – topography, soil, land use and meteorological data (groundwater level is optional but might be needed for watersheds with shallow or perched water tables). Meanwhile, the model conceptualization and parameterization are unified and the results are easily communicated within or outside of the hydrologic community.

A.1.5 Potential limitation and precautions of using DRM

DRM is not a "panacea" nor was DRM designed to completely replace conventional lumped or distributed hydrologic models. DRM mostly focuses only on surface water dynamics (some may include subsurface water as well) and do not explicitly simulate dynamics in other components of hydrologic cycle, e.g., groundwater, assuming the interaction with these components are static. DRM also cannot avoid some of the common pitfalls of lumped or fully distributed models. For example, Manning's roughness coefficient n is used to describe flow resistance and is difficult to measure directly in the field, so the assignment of n values is *always* subjective (Ferguson, 2010; Rehman, 2011). The uncertainty of n values can be aggravated at the watershed scale and even more so for a very shallow depth of overland flow. If the flow depth is comparable to the roughness height, n is no longer a constant (Anderson *et al.*, 2006; Charbeneau *et al.*, 2009; Allan, 2014). In this study, n values are determined from different land use types in the NLCD land use data. NLCD data are derived from Landsat imagery, and the understory vegetation at ground level is often not captured, while flow resistance for different understory vegetation is the actual indicator of Manning's n (Vieux, 2001). However, field estimation of Manning's n for each of the different land use types for a large-scale watershed is usually not feasible because of time and budget limitations. Additionally, DRM is based on hydraulic models originally intended for flood modeling use only. Therefore, watershed to be studied should meet the following assumptions: 1)

Overland flow routing in stream channels and floodplains mainly depends on surface topography; 2) overland flow is the main contributor to runoff and streamflow, and subsurface flow is negligible; and 3) groundwater table does not have significant spatial and temporal fluctuations, and there is no horizontal groundwater movement (Warren and Bach, 1992; WBM, 2017). The hydrological characteristics of the GCEW watershed meet these criteria, so DRM can be used in this study (see detailed GCEW descriptions in the main text).

A.2 DRM implementation: 2D distributed model selection

In this study, to obtain high-resolution temporally and spatially explicit overland flow patterns at the watershed scale to test for the effects of depressions, a highly efficient and robust distributed modeling method is required. Because the depth to the impervious layer (claypan) is shallow (50 cm to 60 cm) at GCEW, and to simplify model conceptualization, we did not use fully coupled and distributed models. Instead, we only focused on surface topography and infiltration-dominated overland flow. In addition, because of the watershed scale and high-resolution DEM data, the selected model should have a high-computational capacity. Thus, we adopted DRM implementation in a 2D hydraulic modeling package – TUFLOW (Syme, 2001). We used the TUFLOW HPC version in our work because of three unique features: 1) the TUFLOW HPC version dynamically and fully solves the shallow water equations (SWEs) on a uniform Cartesian gird without making any simplifying assumptions (WBM, 2017); 2) the TUFLOW HPC version is suitable for large-scale watershed simulation based on a high-resolution DEM because of its highly parallel algorithm, which is capable of leveraging the computational capacity of both CPU and GPU devices (Collecutt *et al.*, 2014); 3) the TUFLOW HPC version can reasonably simulate supercritical flows and capture hydraulic shocks due to the rapid change of small-scale topographic structures, especially at the edges of depressions where rapid gradient shifts occur. This is because the latest TUFLOW solution scheme is based on the 2nd order explicit finite volume and adopts the total variation diminishing (TVD) approach (Collecutt and Syme, 2017). The assumption that the variability of spatially and temporally explicit overland flow patterns and the discharge hydrograph at the watershed outlet are sensitive to depressions and can be simulated and tested by implementing DRM in TUFLOW.

A.3 Additional details regarding the algorithm for $\tau(h)$ and ICSLs

The algorithm of connectivity function τ(h) and integral connectivity scale lengths ICSLs proposed by Western *et al.*, 2001 used a recursive programming approach to searching for connected pairs and downslope accumulation areas, beginning with "seed" (center) cells to neighboring cells and then repeating the searching process for neighbors of the neighboring cells, and so on, i.e., the cell location(s) from the previous recursive function is used as the input for the next recursive function. This algorithm is sufficient for grids with a small number of cells. However, the high-resolution spatial patterns in this paper have a large number of cells (> 1,000,000). The original algorithm accumulates extreme complexity for the recursions and terminates the connectivity computation because it exceeds the recursion upper limit allowed for any programming language. To compute τ(h) and ICSLs more efficiently and upscale to larger grids, we created a new strategy for the algorithm implementation: each cell of input overland flow pattern grids is processed simultaneously as arrays by applying an element-wise operation instead of extracting value from and looping through individual cell. We assume that the connected pairs can only be found in the contiguous "high" regions or connected components, i.e., the spatial clusters of cells with the label "high" connected through the eight neighboring cells. Therefore, we label these clustered regions first and then make full combinations of the cells in each clustered region to make the pairing. Acceleration can then be achieved by adopting an efficient matrix multiplication and vectorization procedure. The algorithm for ICSLs is implemented in MXNet with Python binding for leveraging the parallel computational capacity on both the CPU and GPU devices. In this chapter, we calculated one of the connectivity metrics ICSLs, i.e., OMNI, for the overland flow patterns from the 5 storm events at hourly time intervals. The spatial resolution of the computational grid is 5 m, and the total grid size of the watershed domain (including the cells of "no data") is 8,913,300.

A.4 References

Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, Corrado GS, Davis A, Dean J, Devin M. 2016. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv preprint arXiv:1603.04467*

Abbott MB, Bathurst JC, Cunge JA, O'connell PE, Rasmussen J. 1986. An introduction to the European Hydrological System—Systeme Hydrologique Europeen,"SHE", 2: Structure of a physically-based, distributed modelling system. *Journal of hydrology* **87** (1–2): 61–77

Beven K. 2006. A manifesto for the equifinality thesis. In *Journal of Hydrology*18–36. DOI: 10.1016/j.jhydrol.2005.07.007

Beven K. 2012. Hydrological Similarity, Distribution Functions and Semi-Distributed Rainfall–Runoff Models. In *Rainfall-Runoff Modelling*John Wiley & Sons, Ltd; 185–229. DOI: 10.1002/9781119951001.ch6

Beven KJ. 1996. A Discussion of distributed hydrological modelling. In *Distributed Hydrological Modelling*. DOI: 10.1007/978-94-009-0257-2_13

Beven KJ. 2011. *Rainfall-runoff modelling: the primer*. John Wiley & Sons.

Bleiweiss A. 2008. GPU accelerated pathfinding. In *Proceedings of the 23rd ACM SIGGRAPH/EUROGRAPHICS Symposium on Graphics Hardware*Eurographics Association; 65–74.

Boughton W, Droop O. 2003. Continuous simulation for design flood estimation - A review. *Environmental Modelling and Software* **18** (4): 309–318 DOI: 10.1016/S1364-8152(03)00004-5

Brunner P, Simmons CT. 2012. HydroGeoSphere: A Fully Integrated, Physically Based Hydrological Model. *Ground Water* DOI: 10.1111/j.1745-6584.2011.00882.x

Caddis BM, Jempson MA, Ball JE, Syme WJ. 2008. Incorporating hydrology into 2D hydraulic models-the direct rainfall approach. *9Th Conference on Hydraulics in Water Engineering* (September)

Collecutt G, Syme B. 2017. EXPERIMENTAL BENCHMARKING OF MESH SIZE AND TIME-STEP CONVERGENCE FOR A 1ST AND 2ND ORDER SWE FINITE VOLUME SCHEME. In *Proceedings of the 37th IAHR World Congress*Kuala Lumpur, Malaysi; 5140–5150.

Collecutt G, Syme B, Nielsen C, Roberts M. 2014. Large scale distributed hydrologic modelling using a massively parallelised two dimensional finite volume hydrodynamic model. In *Hydrology and Water Resources Symposium 2014*Engineers Australia; 267.

Collin A, Hench JL. 2012. Towards deeper measurements of tropical reefscape structure using the WorldView-2 spaceborne sensor. *Remote Sensing* **4** (5): 1425–1447

Douglas-Mankin KR, Srinivasan R, Arnold JG. 2010. Soil and Water Assessment Tool (SWAT) Model : Current Developments and Applications. *American Society of Agricultural and Biological Engineers ISSN 2151-0032.* **53** (5): 1423–1431

Eckhardt K, Arnold JG. 2001. Automatic calibration of a distributed catchment model. *Journal of Hydrology* **251** (1–2): 103–109 DOI: 10.1016/S0022-1694(01)00429-2

Ferguson R. 2010. Time to abandon the Manning equation? *Earth Surface Processes and Landforms* **35** (15): 1873–1876 DOI: 10.1002/esp.2091

Green WH, Ampt GA. 1911. Studies on Soil Phyics. *The Journal of Agricultural Science* **4** (1): 1–24

Hall J. 2015. Direct rainfall flood modelling: The good, the bad and the ugly. *Australian Journal of Water Resources* **19** (1): 74–85

Hawkins R, Ward T, Woodward D, Van Mullem J. 2008. *Curve Number Hydrology*. American Society of Civil Engineers. DOI: doi:10.1061/9780784410042

Hunter NM, Bates PD, Horritt MS, Wilson MD. 2007. Simple spatially-distributed models for predicting flood inundation: A review. *Geomorphology* **90** (3–4): 208–225 DOI: 10.1016/j.geomorph.2006.10.021

Huxley C, Syme B. 2016. TUFLOW GPU-Best practice advice for hydrologic and hydraulic model simulations. In *37th Hydrology & Water Resources Symposium 2016: Water, Infrastructure and the Environment*Engineers Australia; 195.

Jones KL, Poole GC, O'Daniel SJ, Mertes LAK, Stanford JA. 2008. Surface hydrology of low-relief landscapes: Assessing surface water flow impedance using LIDAR-derived digital elevation models. *Remote Sensing of Environment* **112** (11): 4148–4158 DOI: 10.1016/j.rse.2008.01.024

Kalyanapu AJ, Shankar S, Pardyjak ER, Judi DR, Burian SJ. 2011. Assessment of GPU computational enhancement to a 2D flood model. *Environmental Modelling & Software* **26** (8): 1009–1016

Le PV V, Kumar P, Valocchi AJ, Dang H-V. 2015. GPU-based high-performance computing for integrated surface–sub-surface flow modeling. *Environmental Modelling & Software* **73**: 1–13

Lexartza-Artza I, Wainwright J. 2009. Hydrological connectivity: Linking concepts with practical implications. *Catena* **79** (2): 146–152 DOI: 10.1016/j.catena.2009.07.001

Mishra, S.K. and Singh VP. 2003. *Soil Conservation Service Number (SCS-CN) Methodology*.

Moradkhani H, Sorooshian S. 2009. General review of rainfall-runoff modeling: model calibration, data assimilation, and uncertainty analysis. In *Hydrological Modelling and the Water Cycle*Springer; 1–24.

Napoli C, Pappalardo G, Tramontana E, Zappalà G. 2014. A cloud-distributed GPU architecture for pattern identification in segmented detectors big-data surveys. *The Computer Journal* **59** (3): 338–352

Ponce VM, Yevjevich V. 1978. Muskingum-Cunge method with variable parameters. In *Journal of the Hydraulics Division, American Society of Civil Engineers*.

Popescu SC. 2007. Estimating biomass of individual pine trees using airborne lidar. *Biomass and Bioenergy* **31** (9): 646–655

Priestley CHB, Taylor RJ. 1972. On the Assessment of Surface Heat Flux and Evaporation Using Large-Scale Parameters. *Monthly Weather Review* **100** (February): 81–92 DOI: 10.1175/1520-0493(1972)100<0081:OTAOSH>2.3.CO;2

Refsgaard JC, Storm B. 1995. MIKE SHE. In *Computer Models of Watershed Hydrology*809–846.

Rehman H. 2011. Rainfall-on-Grid Modelling-a Decade of Practice. In *Proceedings of the 34th World Congress of the International Association for Hydro-Environment Research and Engineering: 33rd Hydrology and Water Resources Symposium and 10th Conference on Hydraulics in Water Engineering*Engineers Australia; 372.

Rehman HU, Zollinger MW, Collings GB. 2003. Hydrological vs Hydraulic Routing Possibilities with Two-dimensional Hydraulic Modelling. In *28th International Hydrology and Water Resources Symposium: About Water; Symposium Proceedings* Institution of Engineers, Australia; 3.

De Roo APJ, Wesseling CG, Van Deursen WPA. 2000. Physically based river basin modelling within a GIS: the LISFLOOD model. *Hydrological Processes* **14** (11–12): 1981–1992

Singh VP. 1995. *Computer models of watershed hydrology*. Available at: http://www.cababstractsplus.org/abstracts/Abstract.aspx?AcNo=19961904762

Singh VP. 1996. *Kinematic wave modeling in water resources, surface-water hydrology*. John Wiley & Sons.

Syme W. 2001. TUFLOW . Two & onedimensional Unsteady FLOW Software for Rivers , Estuaries and Coastal Waters. *IEAust Water Panel Seminar and Workshop on 2d Flood Modelling*: 2–9

Taaffe F, Gray S, Sharma A, Babister MK. 2011. The ineptitude of traditional loss paradigms in a 2D direct rainfall model. *33rd Hydrology and Water Resources Symposium* (July): 434–441

Tang Z, Engel BA, Pijanowski BC, Lim KJ. 2005. Forecasting land use change and its environmental impact at a watershed scale. *Journal of Environmental Management* **76** (1): 35–45 DOI: 10.1016/j.jenvman.2005.01.006

Tayfur G, Kavvas ML, Govindaraju RS, Storm DE. 1993. Applicability of St. Venant equations for two-dimensional overland flows over rough infiltrating surfaces. *Journal of Hydraulic Engineering* **119** (1): 51–63

Vacondio R, Dal Palù A, Mignosa P. 2014. GPU-enhanced finite volume shallow water solver for fast flood simulations. *Environmental modelling & software* **57**: 60–75

Vieira JHD. 1983. Conditions governing the use of approximations for the Saint-Venant equations for shallow surface water flow. *Journal of Hydrology* **60** (1–4): 43–58

Vieux BE. 2001. Hydraulic Roughness: Land Use and Cover. In *Distributed Hydrologic Modeling Using GIS*Springer; 1–17.

Vieux JE, Ceo P. 2002. Vflo™: A REAL-TIME DISTRIBUTED HYDROLOGIC MODEL. *Engineering* (2002): 1–12

Warren IR, Bach H. 1992. MIKE 21: a modelling system for estuaries, coastal waters and seas. *Environmental Software* **7** (4): 229–240

WBM BMT. 2017. TUFLOW User Manual. *Brisbane, Queensland*

Western AW, Blöschl G, Grayson RB. 2001. Toward capturing hydrologically significant connectivity in spatial patterns. *Water Resources Research* **37** (1): 83–97 DOI: 10.1029/2000WR900241

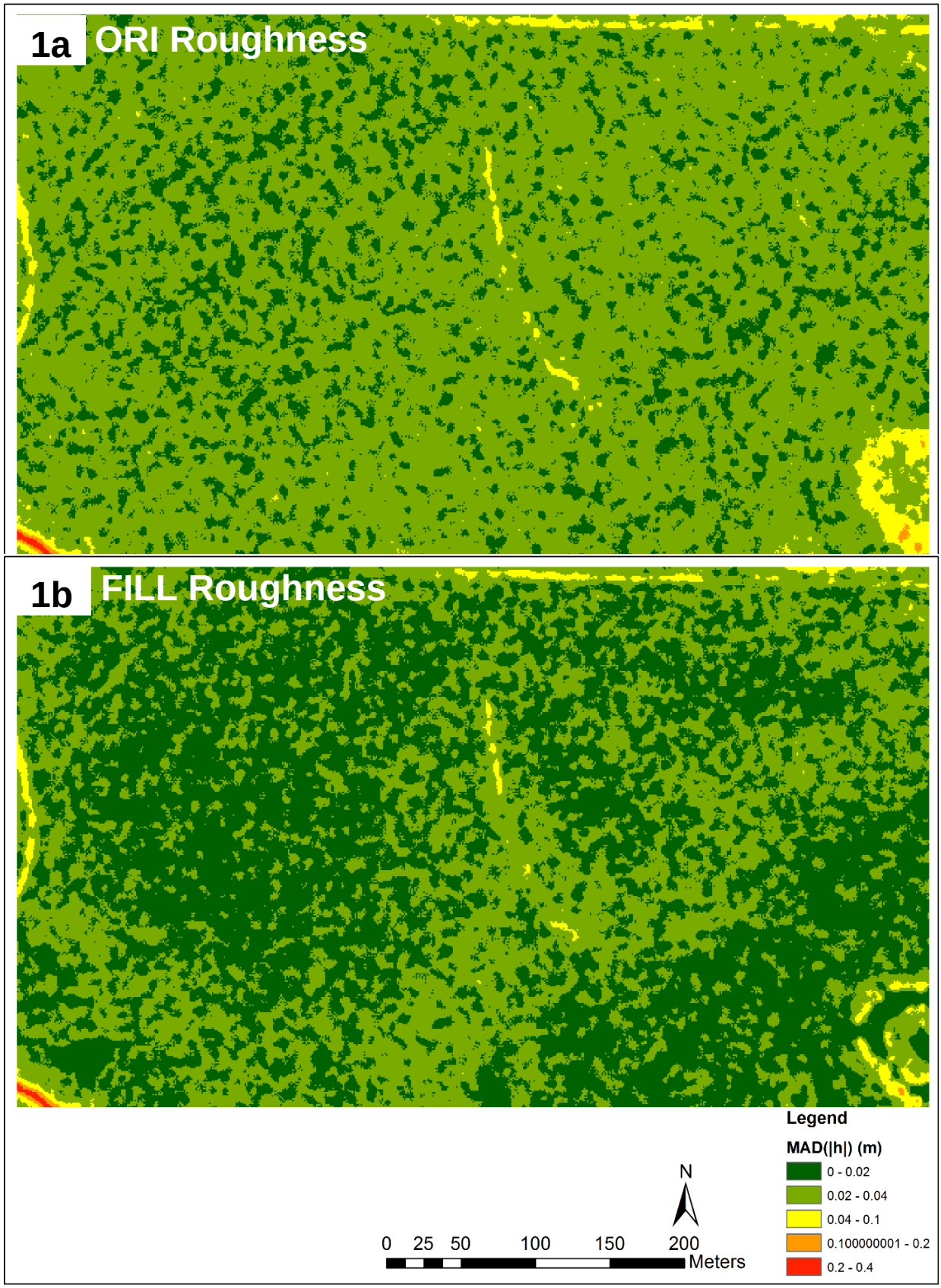
APPENDIX B. ILLUSTRATIONS OF FILL DOMINATED (FD) AND SPILL DOMINATED (SD) OVERLAND FLOW PROCESSES

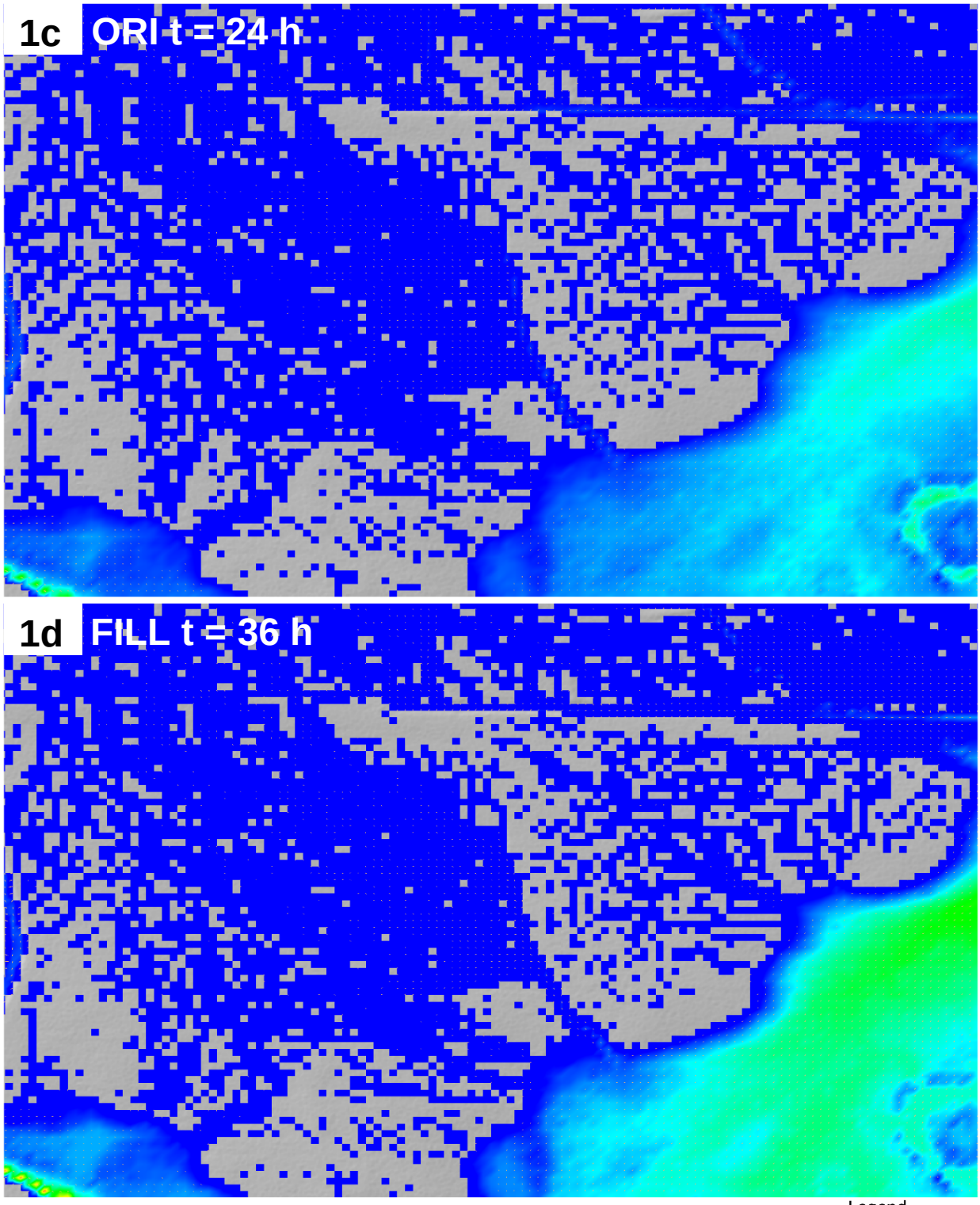
This appendix provides supporting information to illustrate overland flow patterns in fill dominated (FD) and spill dominated (SD) conditions for event 4 (Fig. B.1) and event 1 (Fig. B.2) mentioned in the main text. Surface roughness height is measured by the roughness indicator MAD (Median Absolute Differences) on a 3x3 neighborhood. MAD is a robust metric for evaluating surface roughness characteristics and especially suitable for highly non-stationary spatial data, e.g., surface morphology, at high resolution (Trevisani and Rocca, 2015). MAD is an improved metric compared to the traditional variogram analysis approach.

B.1 References

Trevisani S, Rocca M. 2015. MAD: robust image texture analysis for applications in high resolution geomorphometry. *Computers & Geosciences* **81**: 78–92 DOI: 10.1016/j.cageo.2015.04.003

Figure B.1 MAD roughness distribution map (Omnidirectional magnitude) at an example location in a riparian zone close to the stream for (a) original DEM and (b) depression removed (filled) DEM. The spatial density of low surface roughness areas (in green) is significantly increased after depression removal. Flow depth patterns (c-h) for event 4 at the same location. Patterns of FD process ($\tau(h) > 0$) generated on DEM_{Ori} (c) and DEM_{Fill} (d) at t = 24 h. Patterns of SD process ($\tau(h) < 0$) generated on DEM_{Ori} (c) and DEM_{Fill} (d) at t = 36 h. Patterns of SD process ($\tau(h) < 0$) generated on DEM_{Ori} (e) and DEM_{Fill} (f) at t = 62 h. OMNI(t) and the difference of OMNI(t) between patterns on DEM_{Ori} and DEM_{Fill}. ΔOMNI(t) has been calculated for each pattern. Overland flow connectivity (visualized as connected extent of blue and green cells) between patterns at same time point for FD and SD processes are different and can be measured by changes in OMNI(t). For FD process, ΔOMNI(t) is the highest (positive), while for SD process, ΔOMNI(t) is reversed (negative) at t = 36 h and t = 62 h. Note that the gray color represents the hillside of DEM on which the flow depth pattern is overlaid.

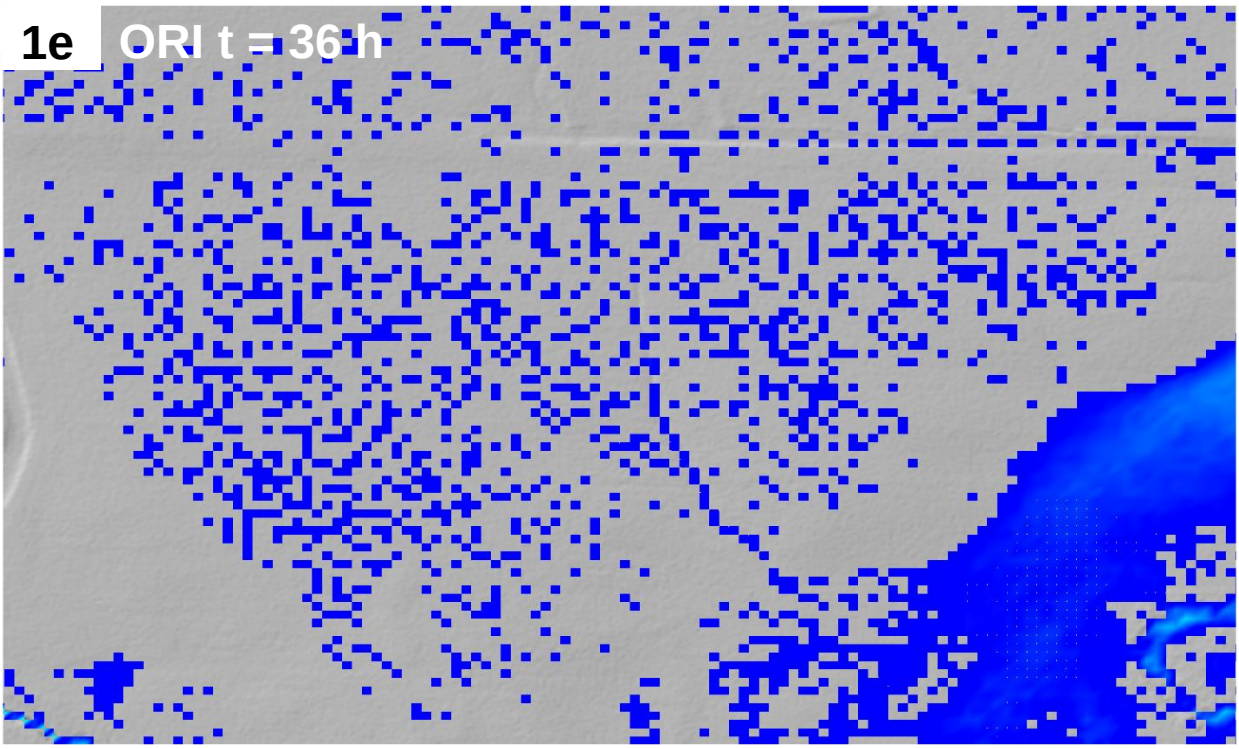

1a ORI Roughness
1b FILL Roughness
Legend
MAD(|h|) (m)
0 - 0.02
0.02 - 0.04
0.04 - 0.1
0.100000001 - 0.2
0.2 - 0.4
N
0 25 50 100 150 200
Meters

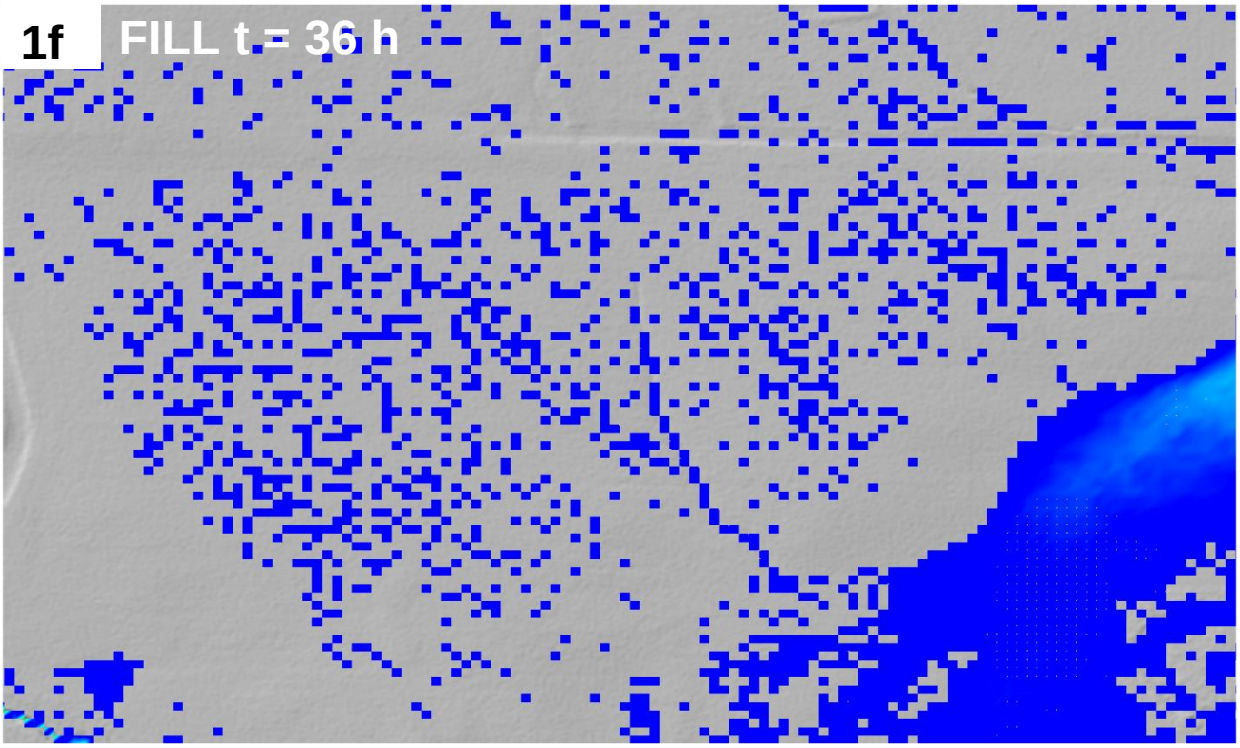


$OMNI(t)_{1c}$ = 66.943

$OMNI(t)_{1d}$ = 73.665

$\Delta OMNI(t)\%$= 10.04





OMNI(t)$_{1e}$ = 0.133

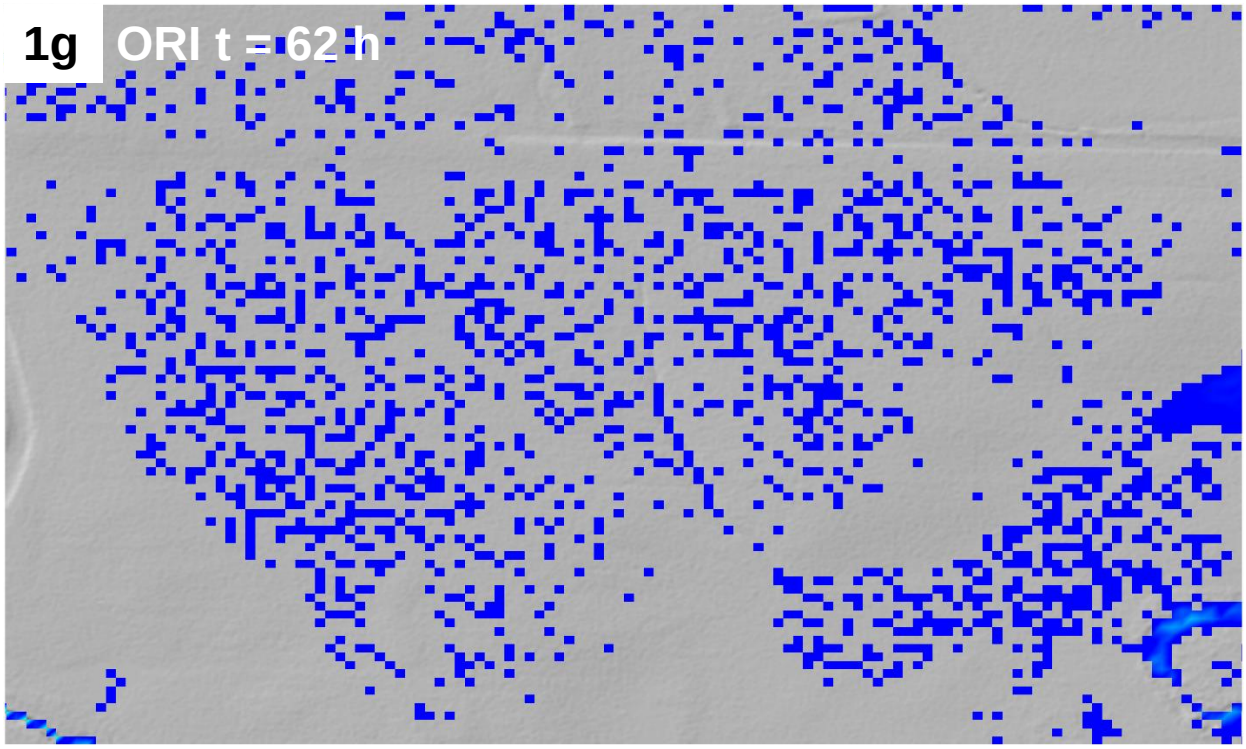
OMNI(t)$_{1f}$ = 0.0746

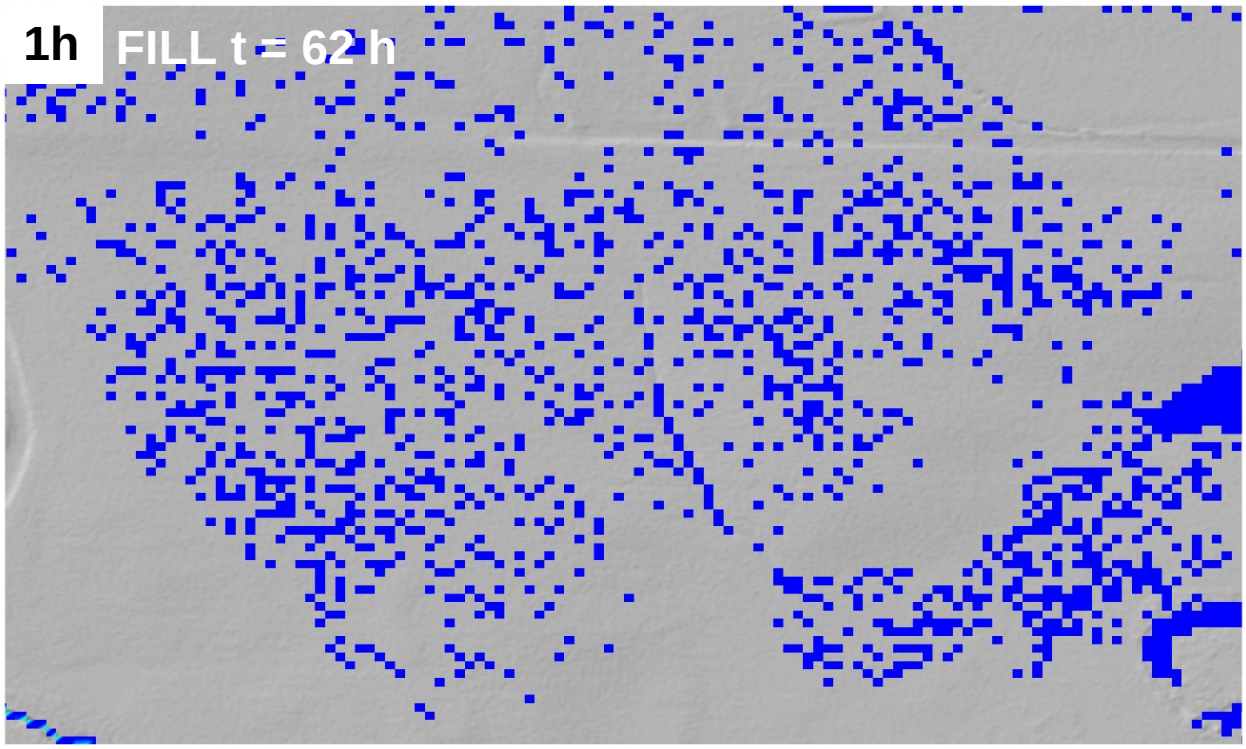
ΔOMNI(t)%= -43.91

Legend

Flow Depth (m)
0.100
0.825
1.550
2.275
3.000

50 0 50 100 150 200 m





$OMNI(t)_{1g}$ = 0.0433

$OMNI(t)_{1h}$ = 0.0368

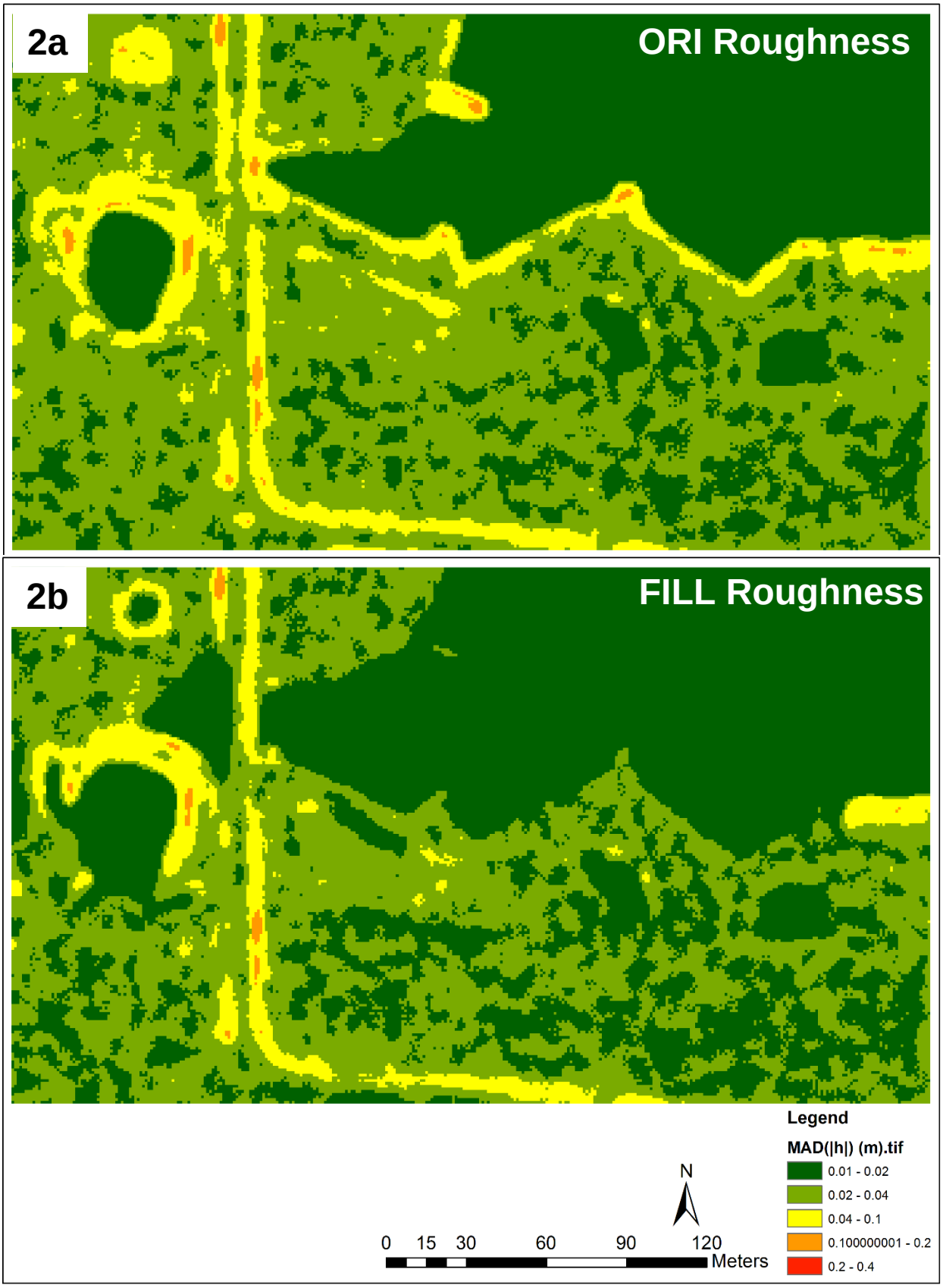
ΔOMNI(t)%= -15.01

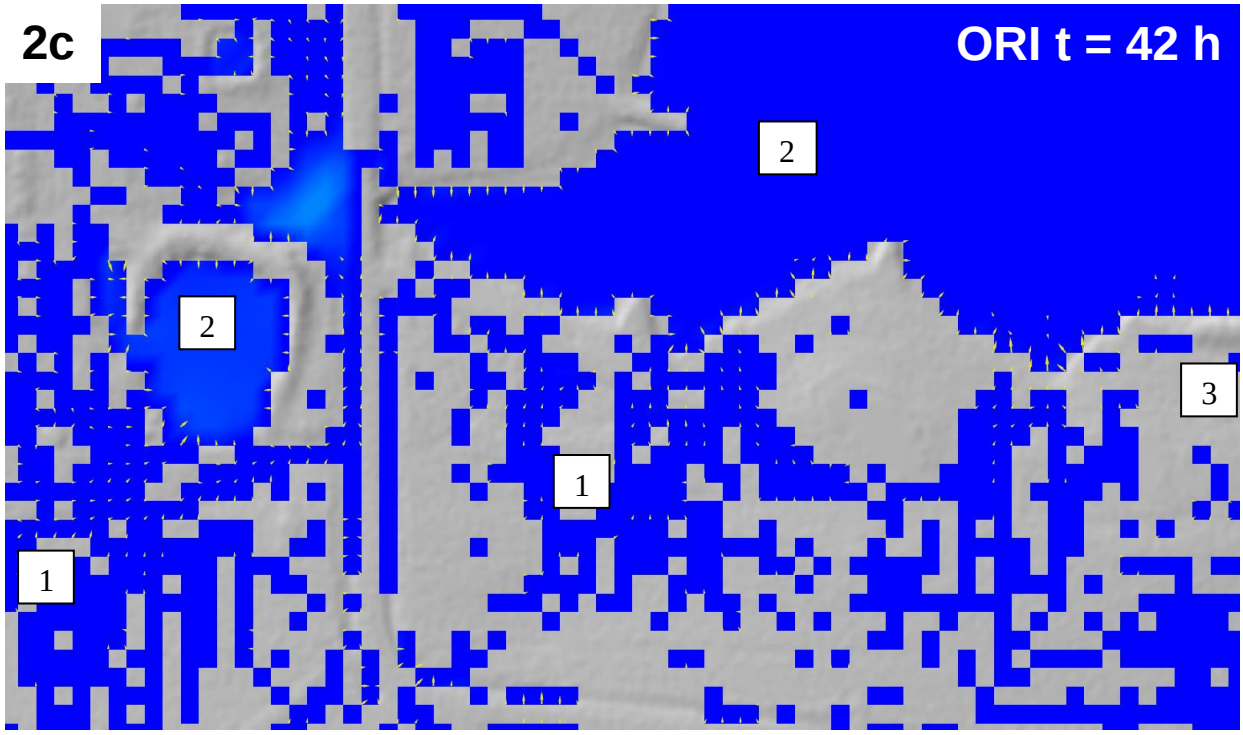
Legend

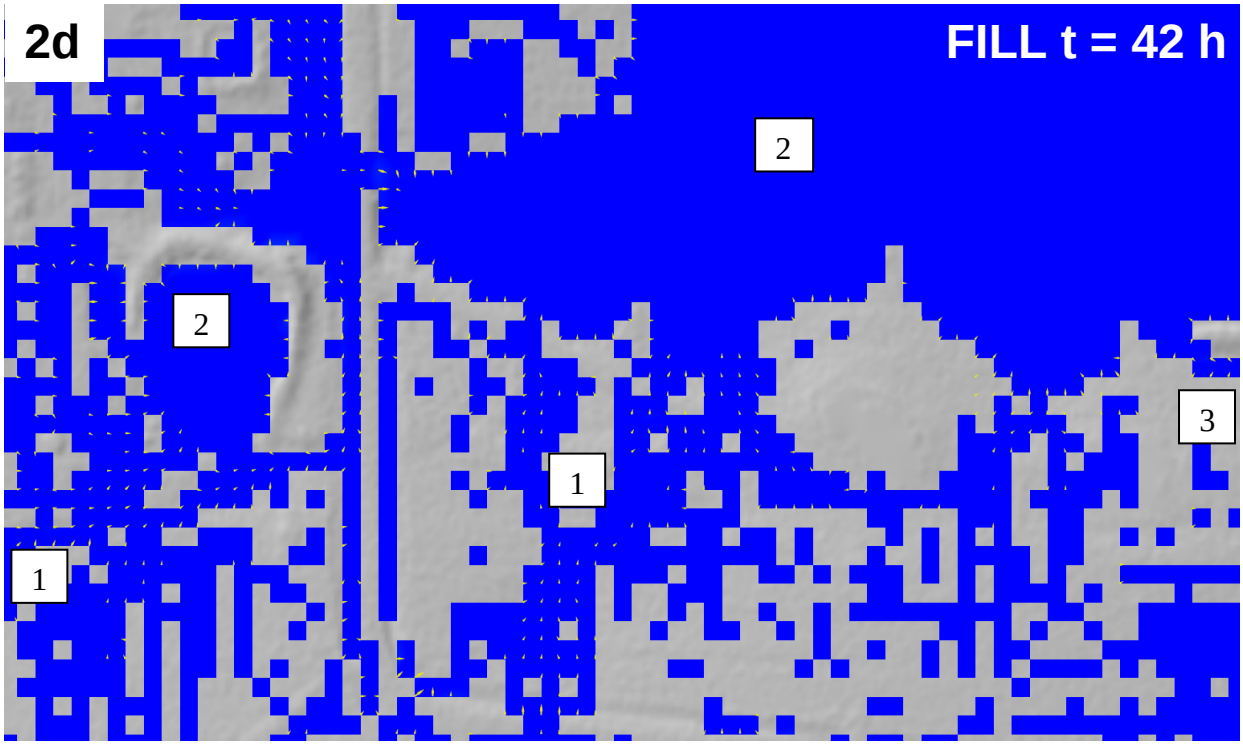
Flow Depth (m)
0.100
0.825
1.550
2.275
3.000

50 0 50 100 150 200 m

Figure B.1 MAD roughness distribution map (Omnidirectional magnitude) for (a) original DEM and (b) depression removed (filled) DEM at an example location near a wastewater treatment lagoon and a pond. Surface roughness is mainly reduced at the edges of lagoons, ponds, and roads. Flow depth patterns (c-h) for event 1 at the same location. Patterns of FD process ($\tau(h)>0$) generated on DEM_{Ori} (c) and DEM_{Fill} (d) at t = 42 h. Patterns of SD process ($\tau(h) < 0$) generated on DEM_{Ori} (c) and DEM_{Fill} (d) at t = 51 h. Patterns of SD process ($\tau(h) < 0$) generated on DEM_{Ori} (e) and DEM_{Fill} (f) at t = 54 h. OMNI(t) and the difference of OMNI(t) between patterns on DEM_{Ori} and DEM_{Fill}. ΔOMNI(t) has been calculated for each pattern. Orange arrows represent the directions and magnitude (head and shaft lengths) of flow velocity. Locations of significant flow velocity differences are circled in red, which are related to changes in fill and spill processes for depressions in different hierarchy level (marked in squared textbox: 1 - 3) between patterns on DEM_{Ori} and DEM_{Fill}. Overland flow connectivity (visualized as connected extent of blue and green cells) between patterns at same time point for FD and SD processes are different and can be measured by changes in OMNI(t). For FD process, ΔOMNI(t) is the highest, while for SD process, ΔOMNI(t) reduces at t = 51 h and t = 54 h. Note that gray color represents the hillside of DEM on which the flow depth pattern is overlaid.


2a
ORI Roughness
2b
FILL Roughness
Legend
MAD(|h|) (m).tif
0.01 - 0.02
0.02 - 0.04
0.04 - 0.1
0.100000001 - 0.2
0.2 - 0.4
N
0 15 30 60 90 120
Meters





$OMNI(t)_{2c}$ = 27.859

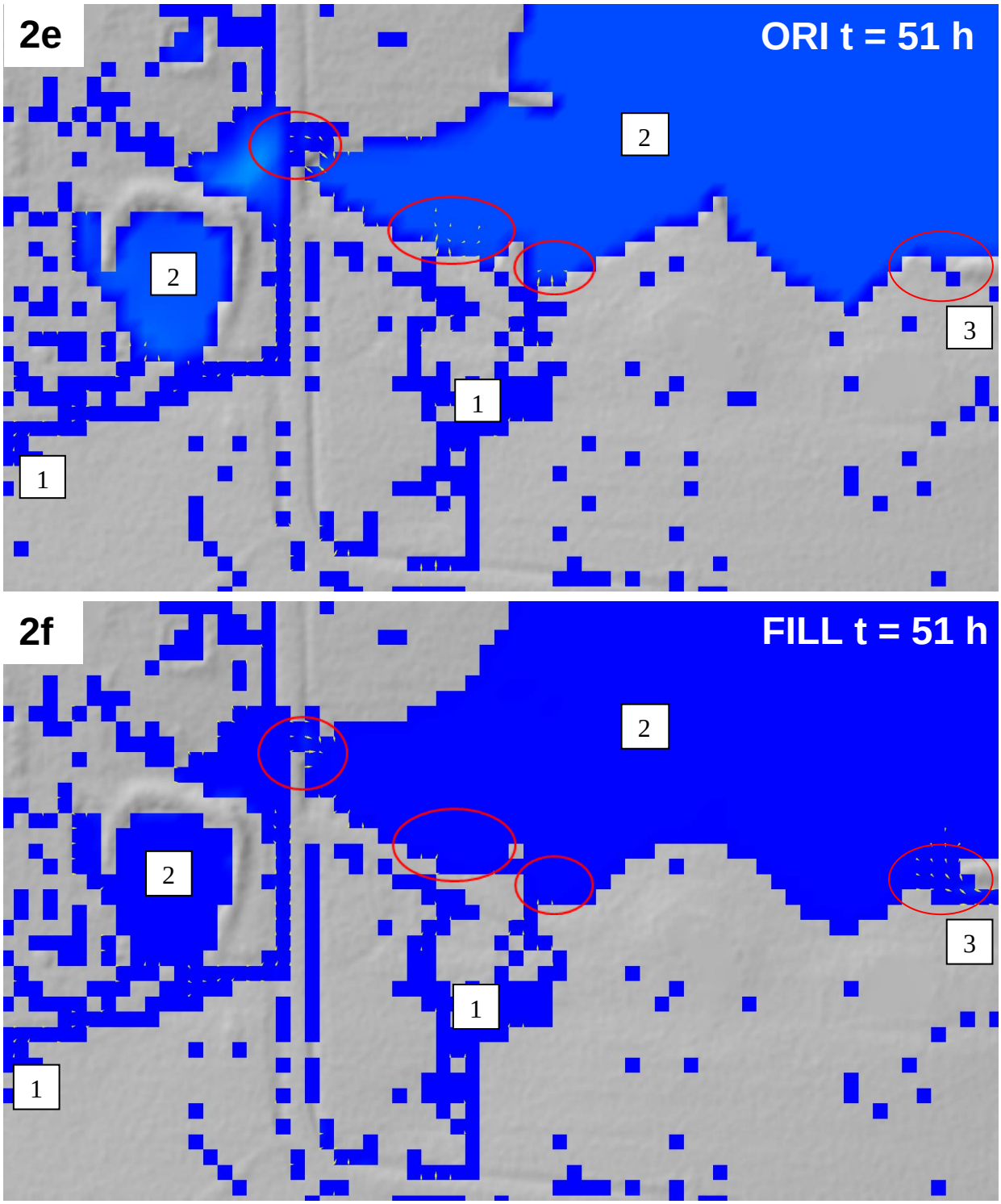
$OMNI(t)_{2d}$ = 33.298

$\Delta OMNI(t)\%$= 19.52

Legend

Flow Depth (m)
0.100
0.825
1.550
2.275
3.000

50 0 50 100 150 200 m



$OMNI(t)_{2e}$ = 38.271

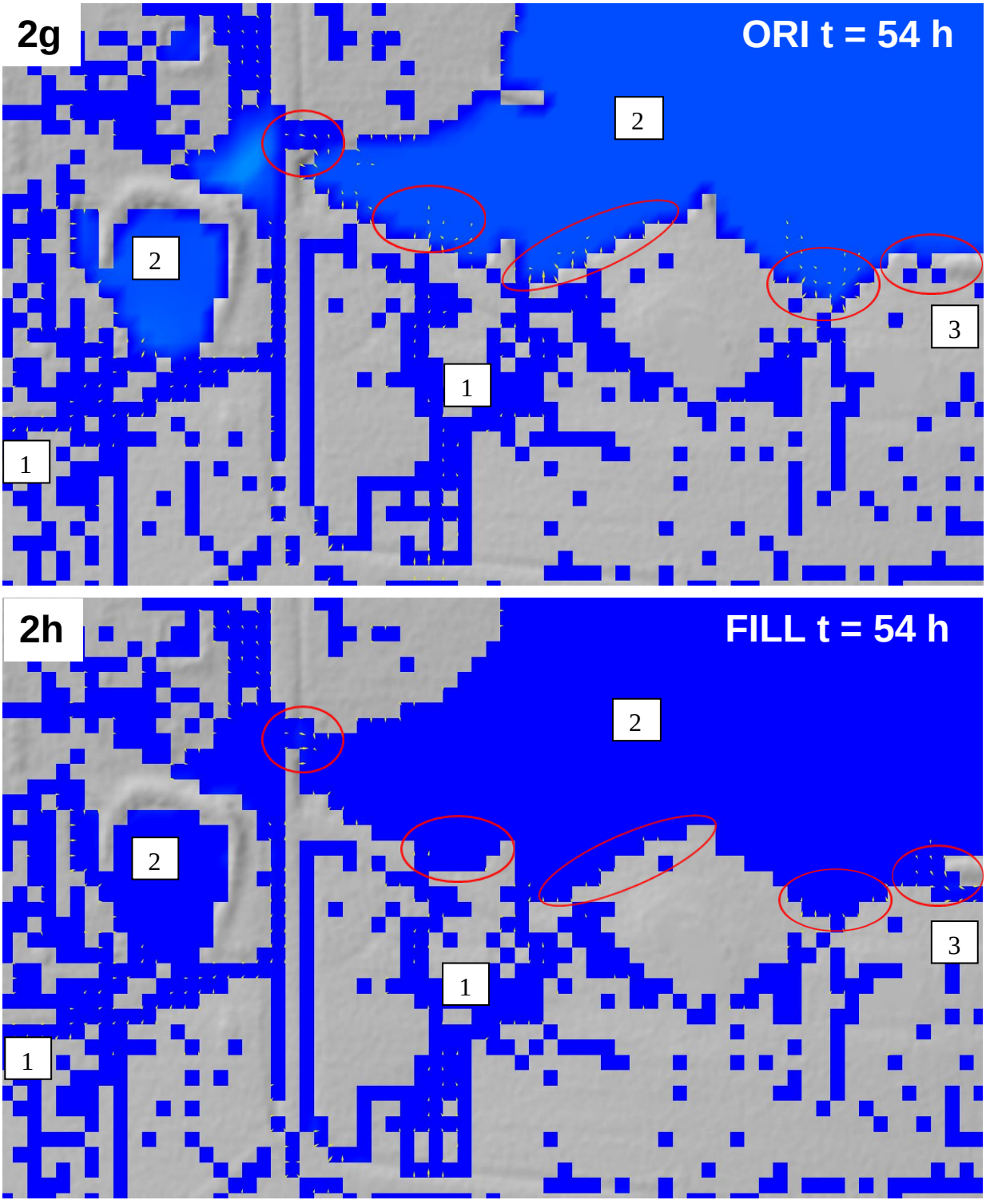
$OMNI(t)_{2f}$ = 44.003

$\Delta OMNI(t)\%$= 14.98

Legend

Flow Depth (m)
0.100
0.825
1.550
2.275
3.000

50 0 50 100 150 200 m



$OMNI(t)_{2g}$ = 30.669

$OMNI(t)_{2h}$ = 32.643

$\Delta OMNI(t)\%$= 6.44

Legend

Flow Depth (m)
0.100
0.825
1.550
2.275
3.000

50 0 50 100 150 200 m