

**Who's afraid of the Saint Venant Equations? (Everyone, so I replaced it with
Machine Learning)**

by

Octavia Crompton

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Earth and Planetary Science

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Inez Fung, Chair
Professor Sally Thompson
Professor Bruce Buffet

Fall 2018

The dissertation of Octavia Crompton, titled Who's afraid of the Saint Venant Equations? (Everyone, so I replaced it with Machine Learning), is approved:

Chair _____

Date _____
Date _____
Date _____

University of California, Berkeley

**Who's afraid of the Saint Venant Equations? (Everyone, so I replaced it with
Machine Learning)**

Copyright 2018
by
Octavia Crompton

Abstract

Who's afraid of the Saint Venant Equations? (Everyone, so I replaced it with Machine Learning)

by

Octavia Crompton

Doctor of Philosophy in Earth and Planetary Science

University of California, Berkeley

Professor Inez Fung, Chair

This thesis addresses the development of a coupled 2D Saint Venant Equation - 1D Richards equation (SVE-R) model, emulation of its output using a machine learning algorithm, and the application of the resulting emulations to predict storm-driven overland flow and infiltration in patchily vegetated dryland systems. The motivation for this work lies in the critical importance of runoff-runon processes in the function of dryland ecosystems: evaluating how these processes work in real or designed settings is crucial for assessing the vulnerability to desertification or the probability of success of management or restoration plans. The SVE-R equations describe the physics of runoff-runon processes, and are solved here by coupling a finite element SVE solver developed by Bradford and Katopodes [1999] to the Richards equation solver developed by Celia et al. [1990]. The coupled SVE-R model is validated against analytical solutions to the kinematic wave equations [Giraldez and Woolhiser, 1996]. I firstly show that the hydrological outcomes of interest in dryland systems (i.e. the distribution of infiltration, the runoff ratio and peak flow velocities) were insensitive to the choice of a roughness closure scheme. I then develop a machine learning (ML) approach to emulate its predictions. This approach involves training random forest (RF) regressors on collections of pre-computed SVE-R model runs to predict the spatial patterns of infiltration and maximum flow velocity. The RF emulator is very accurate and is several orders of magnitude faster than the SVE-R model. To demonstrate the utility of the ML emulation approach for simulating and managing dryland ecosystems, I present two case studies. Firstly, I apply RF regression to represent within-storm processes in an ecohydrological model of spatial patterning of vegetation on the landscape. For the first time, I show that the modeled spatial patterns of vegetation are sensitive to storm characteristics such as duration and intensity. Secondly, I develop a web-based tool to assess plant water availability and erosion risk at hillslope scales for user-supplied vegetation patterns, which embeds random forest regressors trained on a representative library of storm and hillslope conditions. This case study demonstrates how the combination of machine learning and a web tool can make otherwise challenging predictive insights available to a broad set of potential users.

Contents

Contents	i
List of Figures	iii
List of Tables	x
1 Introduction	1
2 A coupled overland flow- infiltration model	11
2.1 Introduction	11
2.2 Richards equation	12
2.3 The Saint Venant Equations	13
2.4 Coupling the SVE and Richards Equation	14
2.5 Model domain and boundary conditions	15
2.6 Implementation	17
2.7 Model Validation with the Kinematic Wave Approximation	19
3 Hydrological sensitivity to the representation of surface roughness in shallow overland flows	21
3.1 Introduction	21
3.2 Methods	25
3.3 Results	29
3.4 Discussion	32
4 Emulating the SVE-R model with machine learning	36
4.1 Introduction	36
4.2 Methods	39
4.3 Results and Discussion	49
4.4 Conclusions	51
5 Storm-scale hydrology in pattern forming models	53
5.1 Introduction	53
5.2 Methods	56

5.3 Results	66
5.4 Conclusions	75
6 A Web Tool for Vulnerability Assessment and Restoration Planning.	77
6.1 Introduction	77
6.2 Methods and Results	81
6.3 Discussion	86
7 Conclusion	89
A Roughness schemes	91
A.1 Roughness scheme details	91
A.2 Resistance sensitivity to scheme	92
B Predicting Velocity with Random Forest Regression	94
C Random Forest Methods	99
C.1 Feature Parameter Selection	100
C.2 Random Forest Parameters	102
C.3 Discussion and Conclusion	105
Bibliography	106

List of Figures

1.1	Distribution of drylands in the world [Bastin et al., 2017]	1
1.2	Examples of conserved (top) and degraded landscapes in the Negev Desert, Israel [The Sustainability Laboratory].	3
1.3	Overland flow in Mesquite desert shrubland after high-intensity rainstorm event [The ECHO Project].	4
1.4	Definition sketch for 1D free surface as represented by the SVE equations. FS is the free surface of the flowing water [Brutsaert, 2005].	5
1.5	schematic of the coupled SVE-Richards equation model.	5
1.6	Aerial view of a striped vegetation in Niger, where vegetation appears in dark and lighter pixels represent bare soil. The distance between successive vegetated bands varies between 60 and 120 meters [Wikimedia Commons, 2015]. In many dryland regions, vegetation patches form regular spatial patterns such as spots, labyrinths or stripes, and runoff-runon mechanisms are considered integral to the formation of this coherent patterning [Rietkerk et al., 2002].	8
1.7	Screen shot of the web app homepage, which prompts the user to upload a vegetation map and select hillslope and storm characteristics. The images show a bird-eye view of an example vegetation map (left) and a binarized version, rotated so that the top of the image matches the top of the hillslope (right).	9
1.8	Screen shot of the web app, showing the random forest-predictions of cumulative infiltration (left) and maximum velocity (right). The images show a birds-eye view, where the top of the hillslope coincides with the top of the image.	10
2.1	(A) Simulated wetting front for a 20 minute storm with $p = 10\text{cm/hr}$ and (B) drainage following the end of the storm. The soil is uniform, with $K_s = 4 \text{ cm/hr}$ and van Genuchten parameters as given in Table 2.1 (non-seal layer). The soil is initially at field capacity ($H = -342 \text{ cm}$). The lower boundary condition is free drainage.	13
2.2	Schematic to illustrate the boundary condition cases for Richards equation. PI is the potential infiltration rate, which is computed to determine whether the rain intensity exceeds the infiltration capacity of the soil (in which case ponding occurs).	15

2.3	Schematic of the model domain. The darker shading under the non-vegetated regions represents the surface crust. $K_{s,V}$ and $K_{s,B}$ represent the saturated hydraulic conductivities of undisturbed soil and surface crust, respectively.	16
2.4	Hydrographs from multiple simulations run in parallel, for factorial combination of hillslope parameters. This figure is for illustrative purposes only.	18
2.5	Sample model output from the coupled SVE-R model showing (A) the input vegetation field, (B) the resultant infiltration map and (C) the maximum flow velocities at the end of the rain. The hillslope parameters are: $K_s=2.0$ cm/s, $p=2.5$ cm/s, $S_o=2\%$. In (B), the darker regions correspond to vegetation patches with higher infiltration rates than the surrounding bare soil areas. Greater infiltration depths are observed down-slope of the bare soil areas, where the run-on source area is larger. In (C), the faster flow velocities are found over the bare soil areas, and where there is less upslope vegetation.	18
3.1	Illustration of the independent roles of the landscape roughness parameterization (α) and the $U-h$ scaling (m) on predicted kinematic relationships between velocity and depth. (A) shows a landscape where α is ‘matched’ to represent a common roughness surface between a turbulent ($m = 0.5$) and mixed ($m = 1$) flow regime. Conversely, (B) shows turbulent flow on two landscapes with different roughness values. The α -matching approach is landscape and discharge dependent, and details are outlined in the methods section.	24
3.2	Example model set-up and infiltration output. (A) map of topography (grey-scale) with spatially random vegetation field (green circles indicate vegetated 0.5×0.5 m cells). (B) Infiltration map resulting from a 20 min storm with $p = 2.7$ cm/hour.	26
3.3	$U-h$ scaling in non-kinematic flows. In (A), ΔS is the difference in between topographic and simulated friction slope (%), and Δm is the difference between the prescribed and simulated $U-h$ scaling. (B) and (C) show example $U-h$ plots where the flow is more and less kinematic, respectively. The simulation parameters are: cylinder array scheme, 5cm/hr rain, 50% vegetation cover and (B) 10% slope (C) 1% slope.	29
3.4	Darcy-Weisbach f evaluated for varying (A) slope and (B) Reynolds number at the outlet, illustrating the implications of various roughness schemes for the interpretation of flow resistance. In (B), Reynolds number was varied by changing the rainfall intensity. The $U-h$ and $U-S_o$ scaling associated with each roughness scheme imply a specific relationship between flow resistance, slope (A) and Reynolds number (B), independent of the surface roughness (α).	30
3.5	Example simulation hydrographs for sparse vegetation (A) and dense vegetation (B), showing good agreement. Hillslope parameters are $p=5$ cm/hr, $S_o=0.01$, $K_{S,bare}=0.34$ cm/hr and (A) $\phi_V=0.1$, (B) $\phi_V=0.5$. Other parameters are listed in Table 3.2.	31

3.6 Box plots show the difference of each simulation from the paired Manning simulation with the same hillslope parameters: (A) the percent difference in infiltration fractions, (B) the hydrograph rising and recession times, (C) the mean and maximum velocities at the end of the rain. (D) shows the mean values for each scheme of all the parameter cases. Abbreviations indicate roughness schemes: Cylinder array (CA), Darcy-Weisbach (DW), Manning's equation (M), Transitional (T) and Laminar (L). In column L, values in brackets include the omitted unstable simulations.	32
4.1 The hillslope on the left represents the SVE-R model domain, including vegetation patterns, storm characteristics, soil parameters and hillslope gradient. The hillslope on the right shows an SVE-R model prediction of infiltration depth, representing the target variable. Each point in the SVE-R model domain is transformed into a 1D vector of features that represent the surrounding neighborhood. The random forest regressor independently operates on each point to predict the outcome.	38
4.2 Patterns were randomly generated from arrays of random numbers between 0 and 1 (A), which were binarized to obtain the non-smoothed ($\sigma = 0$) vegetation patterns (B). The target densities, ϕ_V , were achieved by adjusting the threshold value used to binarize the arrays. The patch length scales were adjusted by applying a Gaussian filter to the binary patterns and re-binarizing to preserve the desired fractional vegetation cover (C). The standard deviation of the Gaussian kernel, σ , determines the characteristic patch length scale.	41
4.3 Illustration of six randomly generated vegetation maps in the random forest training space. Parameter are: $\phi_V=0.2$ (top row), $\phi_V=0.6$ (bottom row), and $\sigma = 0$ (left column), $\sigma = 2$ (center column), and $\sigma = 4$ (right column).	42
4.4 Mean along-slope patch length (A) and bare soil length (B) for a range of smoothing length scales σ , specifying the standard deviation for the Gaussian kernel.	42
4.5 Infiltration maps for a 20 minute storm with $p = 4.8 \text{ cm/hr}$, $S_o = 2\%$, $K_{S,V} = 1.5 \text{ cm/hr}$ and $K_{S,B} = 0.15 \text{ cm/hr}$, using the vegetation maps from Figure 4.3.	43
4.6 Illustrating how the 2D SVE-R domain is deconstructed into grid-cell representations of each point, for several select features. Descriptions of these features are provided in Table 4.2.	44
4.7 Select examples of random forest features: (A) distance to the nearest upslope bare cell, (B) across-slope distance to the nearest bare soil cell, (C) length of the adjacent upslope bare soil patch. Panels D-F show the upslope vegetation fraction computed for several window lengths (12, 20 and 28, respectively). For clarity, values in bare cells are set to 0 in panels D-F.	46
4.8 This figure illustrates the use of the Gaussian filter to smooth the feature maps: (A) the distance to the nearest upslope bare cell, UB , smoothed with a Gaussian filter with (B) $\zeta = 1$ and (C) $\zeta = 2$	47

4.9 Cumulative infiltration maps for a 30 minute storm with (A) $p = 2.5 \text{ cm/hr}$ (B) $p = 7.5 \text{ cm/hr}$. Other parameters are the same between panels: $S_o = 2\%$, $K_{S,V} = 4 \text{ cm/hr}$ and $K_{S,B} = 0.4 \text{ cm/hr}$. Darker areas correspond to vegetated patches.	47
4.10 Schematic illustrating how RF regressors are trained and tested. The data are divided into train and test sets at the hillslope level, as opposed to randomly selecting cells from all the patterns. Spatial maps of vegetation and I are for illustrative purposes only, and all \mathbf{X} and Y (including with subscripts) represent matrices and column vectors, respectively.	48
4.11 Example random forest prediction of a simulation from the test set ($\phi_V = 0.4$, $\sigma = 2$, $t_r = 10$ minutes and $p = 2.4 \text{ cm/hr}$). Panel A shows the SVE model output, and panel B shows the corresponding RF prediction.	50
4.12 A box plot summarizing the NRMSE, grouped by patch length scale (σ). Hatches indicate the test-set simulations ($\sigma = 2$). As expected, the error is generally higher for the test set.	50
4.13 The NRMSE decreases with increasing rain depth (A), and increases slightly with increasing rain duration (B).	51
4.14 (A) A high intensity storm ($t = 10 \text{ min}$, $d = 1.6 \text{ cm}$) where the RF performance is very good, with NRMSE(veg)= 0.03 and NRMSE(bare)= 0.05. (B) For a low intensity storm ($t = 10 \text{ min}$, $d = 0.4 \text{ cm}$) with the same vegetation pattern, the RF performance is comparatively poor, with NRMSE(veg)= 0.21 and NRMSE(bare)= 0.13. This comparison illustrates how the RF performance is best in high intensity storms where the distance to the divide is the dominant predictor, and worse in low intensity cases where runoff-runon patterns dominate.	52
4.15 (A) SVE-Richards versus RF predictions of the hillslope-mean infiltration fractions. The test set is plotted in orange. (B) SVE-R model simulation times versus trained RF prediction times. The RF is several orders of magnitude faster.	52
5.1 Models predict how ecosystems undergo a predictable sequence of patterns as resource availability decreases, after Rietkerk et al. [2004]. This solid lines represent equilibrium densities, and arrows represent state transitions between self-organized patchy and homogeneous states.	55
5.2 Vegetation patterns produced by the R/G model for a range of annual rainfall inputs. The domain size is $800 \times 400 \text{ m}$, with 2 m grid resolution, and the final time is 10,000 days. The initial conditions are the same for all cases, with biomass initialized as a randomly-generated binary array in which $P = 20 \text{ g m}^{-2}$ in 50% of the grid cells and 0 elsewhere. Parameters are $\alpha = 1 \text{ day}^{-1}$, $k_2 = 25 \text{ g m}^{-2}$, $\nu = 25 \text{ m day}^{-1}$, and $W_0 = 0.1$, and otherwise listed in Table 5.1.	57
5.3 Vegetation patterns produced by the R/G-SVE-R model for a range of storm durations, with $d = 1.2 \text{ cm}$, and $R = 440 \text{ mm}$. The final time is 1200 days, and the domain size is $200 \times 100 \text{ m}$ with 2 m grid resolution.	60

5.4 Storm climatology here is considered at ‘within year’ and ‘within-storm’ timescales. Within each year, storms are assumed to occur with a specified occurrence interval (τ) and depth (d). A storm that takes place is then specified by its duration (t_{storm}), its intensity (p) and the total depth. For simplicity and illustrative purposes only, storms were assumed to be identically sized and evenly distributed through time; real locations would be characterized additionally by seasonality and a distribution of waiting times and storm depths.	61
5.5 A schematic illustrating how the R/G-emulator model parameters could be mapped to the Rietkerk-Guttal model, by passing though a multidimensional space of predicted features (here, biomass and wavelength).	65
5.6 Vegetation patterns simulated by the SVE-R (A) and the RF emulator (B) versions of the pattern forming model after 1200 days (108 storms). Storm parameters: $t_{storm} = 30$ min, $d = 1.6$ cm, $R = 440$ mm.	66
5.7 Comparing the predictions of the R/G-SVE-R and the R/G-emulator models for pattern wavelength (A) biomass density (B), and peak biomass (C).	67
5.8 Box plots compare the sensitivity of pattern wavelength (panel A), biomass density (panel B), and peak biomass (panel C) associated with the SVE-R and emulator versions of the R/G model. The boxes with hatches show R/G-emulator model predictions, and boxes without show the R/G-SVE-R model predictions.	68
5.9 Patterns obtained for all storms (as listed in Table 5.2A) and $R = 444$ mm. Wavelength increases with decreasing storm duration (left to right panels) and with increasing storm depth (upper to lower panels), both corresponding to increasing storm intensity. For the highest intensity storms (e.g. $t_{storm} = 10$ min, $d = 1.6$ cm), the patterns do not reach the divide.	70
5.10 Patterns obtained for the same rainfall intensity ($p = 2.4$ cm/hr) and annual rainfall ($R = 440$ mm), for different combinations of d and t_{storm} , to illustrate that variations in pattern morphology are not solely attributable to variations in storm intensity. With storm intensity held constant, pattern wavelength and peak biomass increase with increasing storm depth and duration.	71
5.11 Illustrating the effects of varying R , for fixed $d = 1.2$ cm and $t_{storm} = 20$ min. Increasing R corresponds to decreasing wavelength and increasing biomass density.	71
5.12 Illustrating the effects of changing the random seed, with all other parameters fixed ($R = 360$ mm, $d = 1.2$ cm, and $t_{storm} = 30$ min). The final patterns differ in the details, but are qualitatively very similar.	72
5.13 Sensitivity of the pattern morphology to R , d and t_{storm} , including wavelength (A), biomass density (B), peak biomass (C) and infiltration fraction (D). Bars indicate the range of values spanned with different initial conditions (random seeds), and lines drawn between points are as visual aides only. In most cases, biomass density and peak biomass increase with R and d , and decrease with t_{storm} , with exceptions where the pattern morphology transitions from stripes to spots as storm depth increases.	73

5.14 For lower annual rainfall inputs, pattern morphology changes from stripes to spots with increasing rain depth ($R = 290$ mm, $t_{storm} = 30$ min)	74
5.15 Scatter plots of biomass density and wavelength for (A) the R/G-emulator and (B) the Rietkerk-Guttal models, where marker colors show the peak biomass. Peak biomass is better correlated with biomass density in the R/G-emulator model than the original R/G model.	75
6.1 Landscape structure and ecological functions are tightly coupled in drylands, where patch-mosaic structure facilitates runoff-runon resource redistribution from bare to vegetated areas. Restoring patch-mosaic structure supports key ecosystem services, including controlling erosion, reducing flood risk, and sustaining vegetation. Because patch-mosaic functionality is derived from within-storm hydrological processes, these ecosystem services can be related to quantitative predictions about single storm events: the maximum overland flow velocity (U_{max}), the hillslope runoff volume (Q_{total}), and spatial distribution of infiltration (I) are measures of erosive power, flood risk and water supply to the vegetation, respectively.	79
6.2 Schematic to illustrate hillslope contour planting.	81
6.3 Screenshot of the web app homepage, showing a user uploaded vegetation map (left) and the binarized version. Three additional steps are required before the image is input to the random forest model: the user is asked to (1) rotate the image that the top of the frame coincides with the top of the hillslope, (2) adjust the grid size, so that one pixel represents 1 m, and (3) adjust the threshold value distinguishing vegetated from bare soil areas.	84
6.4 Screenshot of the web app results, showing the random forest predicted infiltration depth (left) and maximum overland flow velocity (right). Summary statistics are displayed below.	85
6.5 (A) Annual infiltration depth in vegetated areas as a function of contour spacing. The dashed horizontal line shows the estimated total plant water requirements per unit area. (B) The estimated maximum flow velocity as a function of the spacing between contours.	86
B.1 Maximum velocity maps for a 20 minute storm with $p = 9.6$ cm/hr, $S_o = 2\%$, $K_{S,V} = 1.5$ cm/hr and $K_{S,B} = 0.15$ cm/hr, using the vegetation maps from Figure 4.3.	95
B.2 Examples of RF features used to make predictions for bare soil areas: (A) distance to the nearest upslope vegetated cell, (B) across-slope distance to the nearest vegetated cell, (C) the length of the upslope vegetated patch. Panels D-F show the upslope vegetation fraction computed with window lengths 12, 20 and 28, respectively (values in vegetated areas are set to 0 in panels D-F for clarity). Scattered green points indicated vegetated cells.	96

B.3 Example comparison between the SVE-R and RF predictions of U_{max} for a hill-slope from withheld set (vegetation density $\phi_V = 0.4$, $\sigma = 2$, $t_r = 10$ min and $d = 1.2$ cm). The SVE-R simulated U_{max} (panel A) and the corresponding RF prediction (panel B) are visually very similar.	97
B.4 A box plot summarizing the NRMSE, grouped by patch length-scale (σ), with hatches to indicate the test-set ($\sigma = 2$). The NRMSE is generally higher for the test set, and decreases with increasing σ	97
B.5 NRMSE grouped by rain depth (panel A) and duration (panel B). The NRMSE is insensitive to rain depth, and increases slightly with increasing rain duration.	98
C.1 NRMSE versus training time for the pattern (panel A) and web-app (panel B) simulations. SVE-R simulations, with the upslope windowing lengths case indicated by marker color. Marker sizes are proportional to the memory use, suggesting low sensitivity to the number of window lengths. The RF regressors were trained with $\zeta = 2$, $\text{max_depth} = 15$ and $\text{min_sample_split} = 20$	102
C.2 The sensitivity of the NRMSE and training time to ζ for the pattern (panel A) and web-app (panel B) simulations. In panel A, smoothing with $\zeta = 2$ or $\zeta = 3$ results in the best performance, with respect to NRMSE and training time. Smoothing does not significantly improve the RF predictions in the web-app case (panel B), most likely because the steeper slope results in flow more closely aligned along the slope gradient. In both case, smoothing at multiple length scales does not improve performance. Marker size shows file size, indicating low sensitivity to the smoothing length scale.	103
C.3 Points represent each combination of max_depth and min_sample_split , indicating very little sensitivity in terms of performance to the RF specification. The $\text{max_depth} = 10$ results, circled in green, are the same for all values of min_sample_split . Increasing tree complexity beyond $\text{max_depth} = 15$, $\text{min_sample_split} = 20$ (red circles), does not lead to improvements in accuracy.	104

List of Tables

2.1	Van Genuchten parameters of seal and underlying soil layers. θ_S and θ_R are the saturated and residual soil water content, respectively.	17
2.2	Summary of the model validation simulations. Differences between model simulations and analytical predictions are given for the runoff hydrographs (Q) and infiltration fractions (f_I). Note that the differences are given as percentages.	20
3.1	Selected roughness schemes to describe the friction slope from the literature. Note the distinct surface roughness parameterizations (α) and scaling between velocity (U) and depth (h) in each scheme. The Poiseuille, Manning and Darcy-Weisbach schemes are well-known, and parameters in the expressions include the kinematic viscosity of water ν and acceleration due to gravity (g), Manning's roughness coefficient n , and the Darcy-Weisbach resistance factor f . Parameters in the remaining schemes are equation-specific constants described in Appendix A.	22
3.2	Parameters for the model simulations. Where multiple parameters are listed, the cases were run factorially to explore all parameter combinations.	27
4.1	Parameters in the RF training space. Where multiple parameters are listed, the cases were run factorially to explore all parameter combinations.	40
4.2	Features used to fit the RF regressors. Gaussian smoothing was applied to all of the features, with the exception of the hillslope-mean fractional vegetation cover, and both smoothed and original versions were included as features.	45
5.1	Rietkerk/Guttal model parameters, descriptions and values. Values are listed for parameters that appear in the modified version of the R/G model presented here.	58
5.2	Rainfall intensity values (in cm/hr) corresponding to the parameter space of rainfall depths and durations ($p = d/t_{storm}$).	61

5.3	Parameters for three sets of model simulations: (B) the SVE-R simulations to train the random forest regressors, (C) the pattern simulations to validate the RF emulation approach, which use both R/G-SVE-R and R/G-emulator models, and (D) the pattern simulations to explore the sensitivity to storm climatology. Within each section, where multiple parameters are listed, the cases were run factorially to explore all parameter combinations. Note that these factorial combinations included the storm parameters listed in section (A). Duplicate parameter listings are included where necessary for clarity. Landscape parameters (hillslope gradient and soil parameters) are explicitly prescribed in the SVE-R simulations, and implicitly prescribed in the pattern simulations through the RF regressors.	62
5.4	Parameters for the Rietkerk-Guttal model simulations. The parameter cases were run factorially to include all parameter combinations.	66
6.1	Parameter space of the SVE-R simulations used to train the RF regressors.	82
6.2	Storm duration - intensity combinations. Intensities are drawn from the Australian Bureau of Meteorology IFD curves, and range from a 1 year Perth storm to a 5 year Darwin storm.	83
C.1	Parameters for the two subsets of SVE-R simulation used to select the feature parameters (ζ and λ) and RF parameters (max_depth and min_sample_split). Where Case A and Case B simulations differ, duplicate parameters are included for clarity.	100
C.2	Window lengths used to compute average upslope vegetation fraction. Window lengths are specified as the number of grid cells.	101
C.3	Select NRMSE scores for the ‘best’ (max_depth = 15, min_sample_split = 20) and ‘best if space limited’ (max_depth = 10, min_sample_split = 20) parameter cases.	105

Chapter 1

Introduction

Drylands, including arid, semi-arid, and dry subhumid areas, cover approximately 40% of Earth's land surface (see Figure 1) and support more than 38% of the total global population [Reynolds et al., 2007, Sarukhan et al., 2005]. They are characterized by limited and highly variable rainfall, making vegetation cover vulnerable to disturbances such as drought, overgrazing, and wildfire. Such disturbance may result in abrupt and irreversible land degradation [Guttal and Jayaprakash, 2007, van de Koppel et al., 2002, Cluff et al., 1983, James and Carrick, 2016], commonly known as desertification [Stavi and Lal, 2015, Bestelmeyer et al., 2013, Sarukhan et al., 2005]. Vegetation recovery in disturbed drylands ecosystems is often slow and uncertain due to unpredictable rainfall during seed germination and seedling establishment [Bashan et al., 2012, Fehmi et al., 2014].

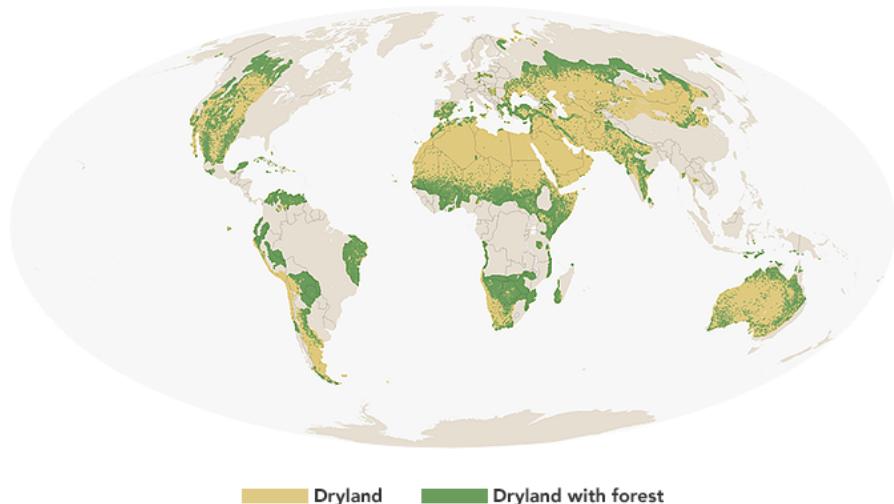


Figure 1.1: Distribution of drylands in the world [Bastin et al., 2017]

Consequently, land degradation has affected 10 to 20% of the global drylands, impacting

approximately 1.5 billion people [Stavi and Lal, 2015, UN, 1995, D'Odorico et al., 2013, Ecosystems and Well-Being, 2005]. The degradation of drylands has profound social, environmental and economic implications [Reynolds et al., 2007], including the generation of as much as 25% of global greenhouse gas emissions, increased soil erosion and flooding, forced migration of rural populations, health risks from dust pollution [Bashan et al., 2012], and loss of agricultural productivity, with implications for global food security [Lal, 2004]. Worryingly, drylands are projected to experience disproportionate climate change impacts, including rising temperatures, longer dry spells, and greater climate variability [Huang et al., 2016], which would be expected to accelerate degradation. Figure 1.2 shows examples of restored and degraded landscapes from the Negev Desert in Israel, where overgrazing and unsuccessful forestry approaches have resulted in extensive land degradation [Mussery et al., 2013, Portnov and Safriel, 2004].

The environmental, societal and economic costs of land degradation have prompted global interest in developing tools and policies to restore degraded landscapes and prevent further degradation [Bestelmeyer et al., 2013]. In 2012, the Rio+20 conference agreed on the goal of global land degradation neutrality (LDN) [Lal et al., 2012]. LDN asserts that there should be no net increase in the area of degraded lands worldwide, and that this can be achieved by reducing desertification in un-degraded regions while increasing restoration efforts to recover already degraded land [Easdale, 2016, Safriel, 2017]. In 2015, the Conference of the Parties to the United Nations Convention on Combating Desertification (UNCCD) established an intergovernmental working group to examine all the available options to achieve LDN in drylands [Grainger, 2015].

A major challenge to LDN remains the lack of a strong scientific footing to support both protection and restoration efforts [Tsunekawa et al., 2014]. Determining the risks, indicators and reversibility of landscape degradation poses practical difficulties [Bestelmeyer et al., 2013, 2015, Chen et al., 2015], and identifying and developing practical solutions to reverse large scale dryland degradation remains a challenge [James and Carrick, 2016]. Many initiatives to restore degraded drylands fail [Josa et al., 2012], due to factors such as poor vegetation selection or planting densities [Bashan et al., 2012], or the inherent variability in rainfall and water availability [Fehmi et al., 2014]. In some cases, re-vegetation has resulted in unforeseen effects on soils, water availability or ecosystem function. For example, afforestation has been widely used on the Loess Plateau of China with the goal of arresting erosion, yet the long-term results of afforestation in some cases have promoted ecosystem deterioration and increased erosion [Jiao et al., 2012, Cao, 2008]. Similarly, afforestation efforts in Niger have failed to reproduce structure and function of healthy ecosystems, and have even caused further degradation [Galle et al., 2001, Menaut and Walker, 2001]. There is thus a pragmatic need for both the protection of at-risk ecosystems and the design of restoration strategies to be informed by an understanding of dryland function.

Water is the main limiting resource in dryland ecosystems, meaning that ecosystem function is often strongly tied to the way the landscape mediates the transport and storage of water to be utilized by plants [Maestre et al., 2016]. Many dryland ecosystems are strongly dependent on the lateral redistribution of rainfall before infiltration, so that vegetation ac-



Figure 1.2: Examples of conserved (top) and degraded landscapes in the Negev Desert, Israel [The Sustainability Laboratory].

cesses more water resources than those which fall immediately above the canopy or root zone. This redistribution can occur because vegetation is typically sparse, growing in a two-phase mosaic of vegetated patches and bare ground interspaces. The bare soils often have low infiltration capacity due to the formation of physical or biological surface crusts [D'Odorico et al., 2009, van de Koppel et al., 2002, Assouline, 2004]. Conversely, vegetation canopies protect the soil surface from rain splash-driven surface sealing and limit growth of biological soil crusts by shading the soil [Belnap, 1990], while root activity creates macropore structures, so that infiltration capacities remain high in vegetated areas.

The differences in infiltration capacity between vegetated and bare sites cause patchily-vegetated landscapes to form a mosaic of sources and sinks of water and other resources [Cantón et al., 2011]. Overland flow is generated on bare ground where soil infiltration capacities are low, then flows downslope to vegetated sites where it infiltrates (see Figure 1.3). This is known as a runoff-runon mechanism, providing additional water, nutrients and soil that sustain vegetation growth [Schlesinger and Pilmanis, 1998, Thompson et al., 2011]. Plant access to water is thus dependent on the spatial distribution of vegetation on a hillslope in addition to climate, soil and topographic properties, making the assessment of vegetation vulnerability to drought or the viability of a proposed restoration strategy a fundamentally hydrological problem [Cantón et al., 2011]. Runoff volume and timing, and consequently sediment production and soil erosion, are similarly sensitive to the spatial configuration of



Figure 1.3: Overland flow in Mesquite desert shrubland after high-intensity rainstorm event [The ECHO Project].

vegetated and bare areas on a hillslope [Bromley et al., 1997, Ludwig et al., 2005, Saco et al., 2007]. Given an LDN framework that emphasizes the restoration and protection of ecosystem services in drylands, vegetation configuration and its impact of runoff generation should be considered a central issue: it impacts ecosystem biodiversity and productivity (supporting services), downstream water availability (provisioning services), and prevention of erosion and flash flooding (regulating services), while likely having indirect effects on a wide range of other provisioning, regulating and cultural ecosystem services [Lal et al., 2012, Grainger, 2015, White et al., 2002, Maestre et al., 2016]. Hydrologic models that can represent runoff production on these landscapes would thus be a useful contribution to restoration and conservation planning under the LDN umbrella.

The very shallow flows (<5 cm) that arise on flat or gently sloping dryland terrain are well represented by the Saint Venant equations (SVE) [Wang et al., 2015]. The SVE, also known as the shallow water equations, derive from depth-averaging the Navier-Stokes equations [Brutsaert, 2005], and represent the mass and momentum conservation laws for the depth-averaged flow. The SVE are shown here in their one-dimensional form:

$$\frac{\partial h}{\partial t} + \frac{\partial(Uh)}{\partial x} = p - i \quad (1.1)$$

$$\frac{\partial U}{\partial t} + U \frac{\partial U}{\partial x} + g \frac{\partial h}{\partial x} + g(S_f - S_o) + \frac{U(p - i)}{h} = 0 \quad (1.2)$$

where h is the depth of flow, U is the depth-averaged flow velocity, S_o is the bed-slope, x is the spatial coordinate and t is the time, and g is the acceleration due to gravity. S_f is the friction slope (or energy gradient), which represents the effect of bed and other shear stresses on retarding the flow. The lateral boundary conditions are rainfall inputs (p) and infiltration losses (i). Figure 1.4 shows a definition sketch for the 1D SVE. The local acceleration ($\frac{\partial U}{\partial t}$) describes the change in velocity over time, the convective acceleration ($U \frac{\partial U}{\partial x}$) describes an

acceleration caused by some change in velocity through space, and the pressure gradient term ($g \frac{\partial h}{\partial x}$) describes how pressure changes with position. The local and convective acceleration terms make up the inertia terms of the SVE.

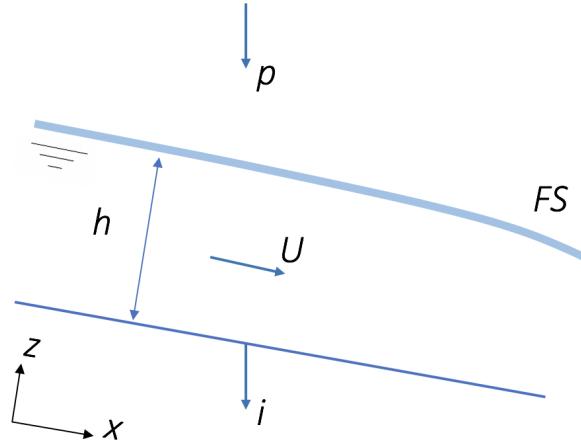


Figure 1.4: Definition sketch for 1D free surface as represented by the SVE equations. FS is the free surface of the flowing water [Brutsaert, 2005].

Implementing the SVE requires a closure model relating the friction slope S_f to h , U and the roughness characteristics of the local land surface. The boundary conditions associated with rainfall on the free boundary, and infiltration on the soil boundary, must also be specified. In this thesis, rainfall is assumed to occur in finite-duration events of known intensity. Infiltration is represented with Richards equation [Richards, 1931] and dynamically coupled to the SVE, as described in detail in Chapter 2 and illustrated schematically in Figure 1.5.

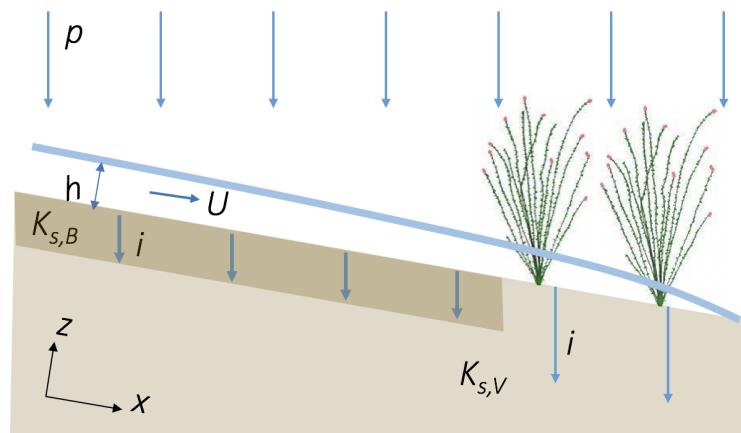


Figure 1.5: schematic of the coupled SVE-Richards equation model.

Coupled SVE-Richards equation models have not been widely implemented to simulate dryland hillslopes, because they are computationally intensive and difficult to parameterize. Other approaches include simplifying the representation of overland flow with the kinematic or diffusive wave approximations [Mügler et al., 2011, Thompson et al., 2011]. Unfortunately, these approximations make assumptions that may break down in low slope and heterogeneous landscapes. For example, the kinematic wave approximation approximates the friction slope with the topographic slope (i.e. $S_f = S_o$), making it unsuitable for modeling flat terrain or highly unsteady flows. The diffusive wave approximation neglects the inertia terms in the SVE and cannot represent supercritical flows, which may arise when variable infiltration rates or microtopography cause local acceleration. Other simplified models include terrain following models that account for flow routing but not fluid dynamics. The most commonly used is the Connectivity of Runoff Model (CRUM), in which water is routed across the surface using the multiple flow path algorithm FD8 [Quinn et al., 1991, Reaney et al., 2007, 2014]. Given the importance of the persistence and timing of flows for redistributing rainfall through space and controlling infiltration depth, these models are likely better suited for bulk runoff predictions than for assessing vegetation water availability. Thus, in spite of its complexity, the SVE-Richards coupling is attractive for retaining high physical fidelity to the dynamic flow processes.

Its use, however, raises a number of challenges, including (1) representing landscape complexity, (2) simulating spatially and temporally varying flow fields, and (3) parameterizing the surface roughness. I largely address the first challenge by adopting a simplified parameterization of the landscape based on a two-phase vegetated and bare mosaic, assuming that parameters are constant for each phase. The second issue is circumvented by using a previously validated finite volume numerical solver for the SVE. The third issue, relating to the parameterization of surface roughness, is explicitly addressed in Chapter 3.

Roughness elements on a landscape can be explicitly represented with high resolution, computationally intensive models (see e.g. Kim et al. [2012]), but this requires detailed information about the land surface that is not generally available. Thus, a parameterization of roughness elements is required [Smith et al., 2007, Cea et al., 2014]. There are, however, a broad diversity of roughness schemes that purport to represent the interaction of overland flow with its environment. These schemes generally derive from different conceptual models of flow, and are associated with very different parameterization requirements. For example, the widely-used Darcy-Weisbach friction factor has its origins in pipe flow, Manning's equation is based on empirical observations of channel flow, and the Poisseuille equation is obtained by solving the Navier Stokes equation for plane parallel flow [Brutsaert, 2005, Smith et al., 2007]. Other schemes have been developed to explicitly represent flow through vegetation [Katul et al., 2011] or through cylinder arrays that approximate vertically oriented vegetation stems [Tanino and Nepf, 2008, James et al., 2004]. How best to parameterize the effects of sub-grid scale roughness elements on the flow is a challenging, unresolved question [Smith et al., 2007], and the uncertainty introduced into results by the selection of a specific roughness scheme has not generally been quantified. I address this issue explicitly in Chapter 3. I investigate the sensitivity of relevant hydrological outcomes

to the selection and parameterization of the roughness scheme, and show that in practice, there is minimal sensitivity to scheme selection. Hydrological predictions are sensitive to the surface roughness parameterization (α), but not to the functional form of the S_f equation.

Another challenge associated with using the coupled SVE-Richards equation model to investigate water availability on dryland slopes lies in knowledge exchange with practitioners. Applying this type of model requires training in computer coding and hydraulics, and such requirements are not reasonable for many practitioners. I address this challenge by developing a method to emulate the model outcomes with machine learning. Chapter 4 introduces a machine learning approach to predict the spatial distributions of cumulative infiltration and maximum flow velocity on hillslopes as a function of the spatial patterns of the vegetation.

The use of machine learning to emulate complex model results is employed to lower the computational demands of otherwise intensive models used (for example) in weather forecasting and climate change assessment [Castruccio et al., 2014, Moonen and Allegrini, 2015]. Emulation involves replacing the physical model with a statistical or other simplified relationship between model inputs and outputs [Reichert et al., 2011]. Machine learning techniques are increasingly popular for these purposes, owing to rapid improvements in the quality and usability of machine learning algorithms [Goldstein and Coco, 2015]. Machine learning emulators have been used in hybrid environmental models, which integrate physics-based modeling (e.g. solving the geophysical fluid dynamics equations in global climate models) with machine learning components (e.g. neural network emulation of physical processes like radiation, convection, clouds and turbulence) [Krasnopsky and Fox-Rabinovitz, 2006]. As an advance over traditional curve-fitting approaches, machine learning has been used to emulate empirical results and gain insight into the physical processes, for example, in the mean velocity of open-channel flow through submerged arrays of rigid cylinders [Tinoco et al., 2015].

In this application, I use machine learning to emulate how the spatial arrangement of vegetation on the hillslope influences the resultant spatial distribution of infiltration and peak overland flow velocities. Chapter 4 discusses how a random forest approach was selected, implemented and validated to emulate the SVE-Richards equation model simulations. I show that the trained random forest regressors are nearly as accurate as the physics-based solutions, while reducing the computational requirements by several orders of magnitude. Chapter 4 also discusses the limitations of the approach, specifically (1) extrapolation beyond the training space of vegetation patterns (2) interpolation between hillslope features and storm characteristics and (3) challenges associated with the nearly unbounded diversity of vegetation patterns in the training space.

Having developed the emulation method, two applications are explored.

First, Chapter 5 shows how emulation can overcome the timescale separation that makes long-term modeling of dryland vegetation challenging. Perennial dryland vegetation is often long-lived and slow growing, so that its dynamics evolve over decadal to century-long timescales. Given the computational demands of modeling storm-driven overland flow, resolving multiple storm events with timescales of minutes to hours in multi-decadal simulations is computationally expensive. Model emulation with machine learning can overcome

this challenge. Chapter 5 presents an example of modeling the spatial self-organization of vegetation into large-scale patterns (see Figure 1.6). Applied to pattern-forming models, the emulation approach allows vegetation morphology to be related to storm-scale characteristics; something that has been ignored in most modeling of such systems to date.

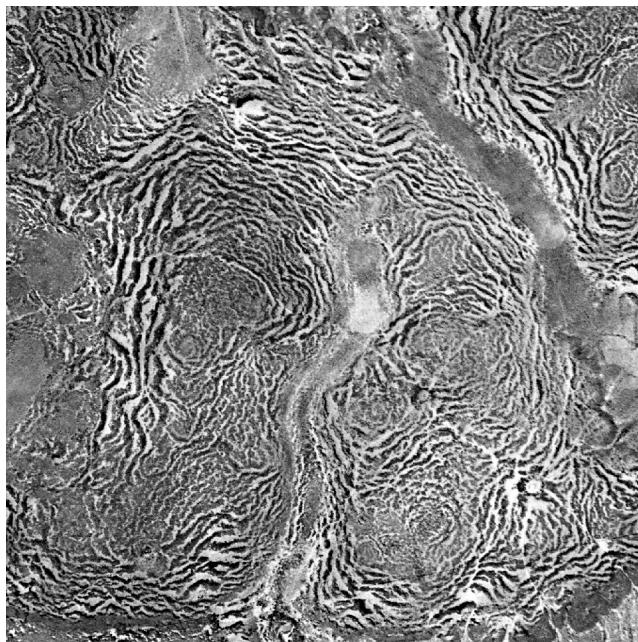


Figure 1.6: Aerial view of a striped vegetation in Niger, where vegetation appears in dark and lighter pixels represent bare soil. The distance between successive vegetated bands varies between 60 and 120 meters [Wikimedia Commons, 2015]. In many dryland regions, vegetation patches form regular spatial patterns such as spots, labyrinths or stripes, and runoff-runon mechanisms are considered integral to the formation of this coherent patterning [Rietkerk et al., 2002].

Chapter 6 introduces a second machine learning application, aimed at addressing the knowledge gap between practitioners (e.g. land managers, farmers, landscape architects, restoration ecologists, government agencies, NGOs) and hydrologists/engineers with the tools to study water movement in drylands. Despite the known importance of overland flow in arid systems, the tools to simulate storm-driven overland flow are not readily available to practitioners seeking to assess vulnerability or plan restoration projects. Consequently, practitioners are not using quantitative predictions for overland flow and infiltration from hydrological models [Hewett et al., 2018].

Chapter 6 outlines the development of a web app to assess infiltration capacity and erosion risk at the hillslope scale, for user-specified hillslope parameters, storm characteristics and vegetation distribution. The web app uses a library of trained random forest regressors covering a large parameter space of hillslope gradients, storm durations and intensities, soil

parameters (via the saturated hydraulic conductivity), and vegetation patterns. It prompts the user to upload a vegetation map and select hillslope parameters from dropdown menus (Figure 1.7), runs the relevant random forest regressor, and returns the random forest predictions (Figure 1.8). Predictions include (1) spatial maps of infiltration and maximum flow velocity and, (2) estimates of whether the vegetation will have enough water, provided the user also specifies the annual mean rainfall and estimated plant water requirements. Chapter 6 describes the design of this tool, specifically the determination of a representative parameter space on which to train the random forests.

Welcome to the SHADE App!

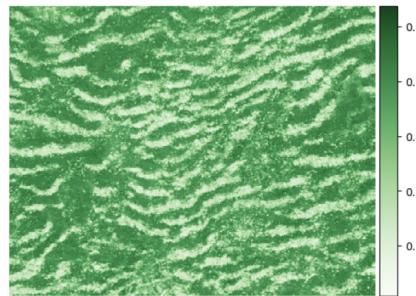
The Storm and Hillslope Assessment for Dryland Ecosystems (SHADE) app is a tool to assess plant water availability and erosion risk using storm-scale hydrological predictions.

For more information: [Home](#) [Instructions](#)

Landscape features added

1. Upload vegetation map

veg file selected : *WAsmall.png* [Change file](#)



2. Binary image

Threshold: 0.40 Pixels per meter: 3.00 [Update](#)



Hillslope dimensions = 166m x 150m
Vegetation density = 0.62

3. Input storm and landscape features

Select a typical storm duration and update to view available storm intensities:

Typical storm duration : hr [Update duration](#)

Typical storm intensity : cm/hr

Ksat : cm/hr

slope : %

[Update features](#)

4. Feature summary

Storm duration = 0.5 hr

Intensity = 2.5 cm/hr

Rain depth = 1.25 cm

Ksat = 2.5 cm/hr

Slope = 0.5%

[Run random forest model](#) or [Reset form](#)

Figure 1.7: Screen shot of the web app homepage, which prompts the user to upload a vegetation map and select hillslope and storm characteristics. The images show a bird-eye view of an example vegetation map (left) and a binarized version, rotated so that the top of the image matches the top of the hillslope (right).

Limitations to this approach are also discussed in Chapter 6. First, it requires the user to specify a typical storm for their region, which is difficult given the high variability of

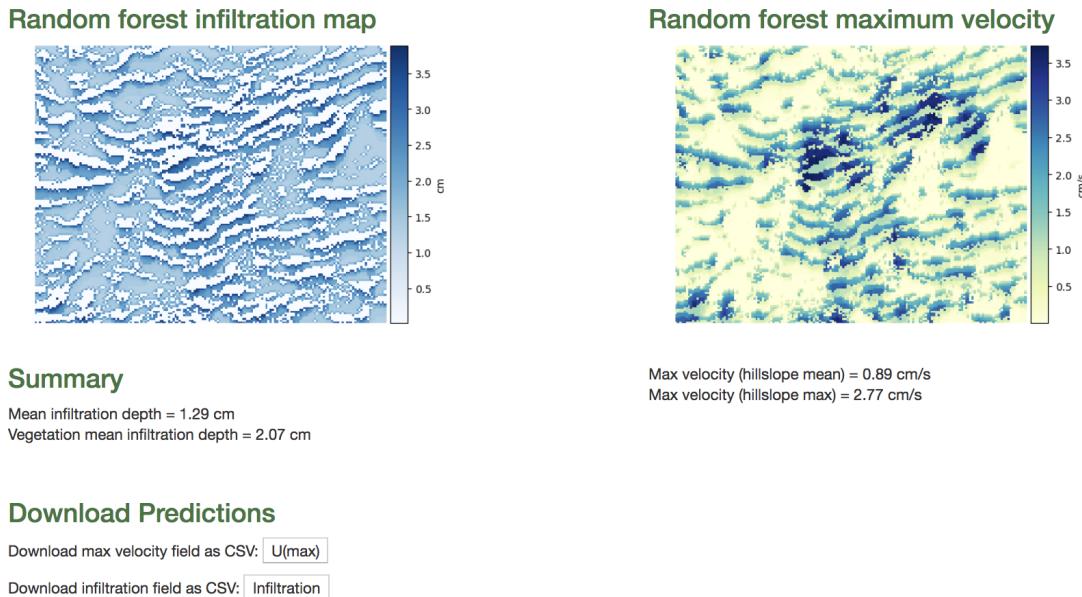


Figure 1.8: Screen shot of the web app, showing the random forest-predictions of cumulative infiltration (left) and maximum velocity (right). The images show a birds-eye view, where the top of the hillslope coincides with the top of the image.

storms in most dryland regions. Second, it requires the user to choose among a relatively small number of slope, soil and storm parameters (a total of 256 parameter combinations). Nevertheless, the library of random forests represents a total of 5120 simulations, each taking 2-48 hours to run. Despite its limitations, the web app is unique in providing information about the spatial distribution of infiltration and overland flow. With further development, the representative parameter space could be broadened, particularly with the addition of non-planar topographies (e.g. convergent topographies).

In summary, this dissertation describes the development and validation of a coupled SVE-Richards equation model, designed to represent storm-driven overland flow on dryland hillslopes with patchy vegetation (Chapter 2). The remaining chapters discuss the model sensitivity to the roughness parameterization (Chapter 3), emulation of the model outcomes with machine learning (Chapter 4), and two applications of the machine learning approach: pattern forming models (Chapter 5) and a web app (Chapter 6). The conclusion will summarize the findings and discuss the potential for future research.

Chapter 2

A coupled overland flow- infiltration model

2.1 Introduction

Overland flow is common in locations with low soil permeability, commonly found in mountainous, arid, urban or agricultural landscapes [Descroix et al., 2007, Dunne, 1983, Li et al., 2014]. Overland flow occurs at the expense of infiltration, and thus water supply to vegetation, and is responsible for soil erosion and flash flooding [Abrahams et al., 1994, Bracken et al., 2008]. Predicting overland flow is therefore broadly relevant to flood prediction, erosion assessment, and the prediction of water balance partitioning (runoff v.s. infiltration) [Hallema et al., 2016, Cantón et al., 2011]. Some of these outcomes can be assessed with simplified or steady-state models, such as the kinematic or diffusive wave approximations. For example, erosion assessment requires accurate prediction of the maximum flow velocities, which could be approximated by steady state conditions in many cases. Determining the spatial patterns of infiltration to assess plant water availability, however, requires representing the transient flow conditions, as infiltration patterns arise largely from the runoff or infiltration of ponded water after the end of a storm.

The basic equations describing overland flow are the Saint-Venant Equations (SVE), also known as the shallow water equations. These equations are derived from the depth-averaged Navier-Stokes equations, under the assumption of shallow flow (i.e. that the horizontal length scales are much larger than the vertical ones). The standard approach for modeling infiltration in soils is Richards equation, a nonlinear partial differential equation representing the movement of water through unsaturated soils under the action of capillarity and gravity [Richards, 1931]. The SVE and Richards equation are coupled through the boundary conditions at the soil surface.

The infiltration and overland flow components of the model are described in Sections 2.2 and 2.3, and Section 2.4 summarizes how these components are dynamically coupled. Section 2.5 describes the model domain of patchily vegetated hillslopes, where vegetated patches are

distinguished from bare ground by surface roughness and soil infiltration capacities. Sections 2.6 and 2.7 present implementation details, sample results and model validation with an analytical solution to the kinematic wave approximation [Giraldez and Woolhiser, 1996].

2.2 Richards equation

For the short timescales of rain showers, infiltration can be approximated as a 1D process in the vertical direction. With the assumption of 1D flow, a fully coupled, 3D approach can be replaced with many, independent 1D calculations. In its mixed water content-capillary head form, the 1D (vertical) Richards equation is:

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

where K is the unsaturated hydraulic conductivity, H is the matric head, z is the elevation above a vertical datum, and θ is the volumetric water content. The first term on the right represents the effects of capillarity, and second term represents the gravity-driven flux [Farthing and Ogden, 2017]. θ is related to H via a water retention curve such as the Van Genuchten [1980] equation:

$$\theta(H) = \theta_R + \frac{\theta_S - \theta_R}{[1 + (\alpha|\phi|)^n]}^{1-1/n} \quad (2.2)$$

where θ_S and θ_R are the saturated and residual soil water content, respectively, n is a measure of the pore size distribution, and α is related to the inverse of the air entry suction.

The surface boundary condition depends on whether water is accumulated above the surface: a Dirichlet boundary condition (fixed H) is applied if water is ponded, and a Neumann boundary condition (fixed flux) is applied otherwise. The lower boundary condition can be applied as no flux or free drainage (zero gradient). A no flux boundary condition represents an impermeable layer such as hardpan, which causes the soil to eventually saturate for all rainfall cases. A free drainage boundary condition represents a freely draining soil profile, which is often appropriate for the situation where the water table lies far below the simulation domain [Simunek et al., 2005].

There are several challenges to implementing Richards equation: there is no closed-form analytical solution, numerically solving Richards equation is computationally expensive and not guaranteed to converge [Short et al., 1995], and many numerical methods do not ensure mass conservation [Farthing and Ogden, 2017]. The 1D algorithm developed by Celia et al. [1990] is implemented here because it offers near perfect mass conservation. The algorithm involves a backward Euler implementation of the mixed $H - \theta$ formulation of Richards equation, and uses modified Picard iterations to improve mass conservation. It has become a standard numerical approach, used in many production codes such as the USDA Hydrus-1D Richards equation solver [Simunek et al., 2005]. Figure 2.1 shows an example simulation of a wetting front during a 20 minute rainstorm (panel A) and drainage after the end of storm (panel B).

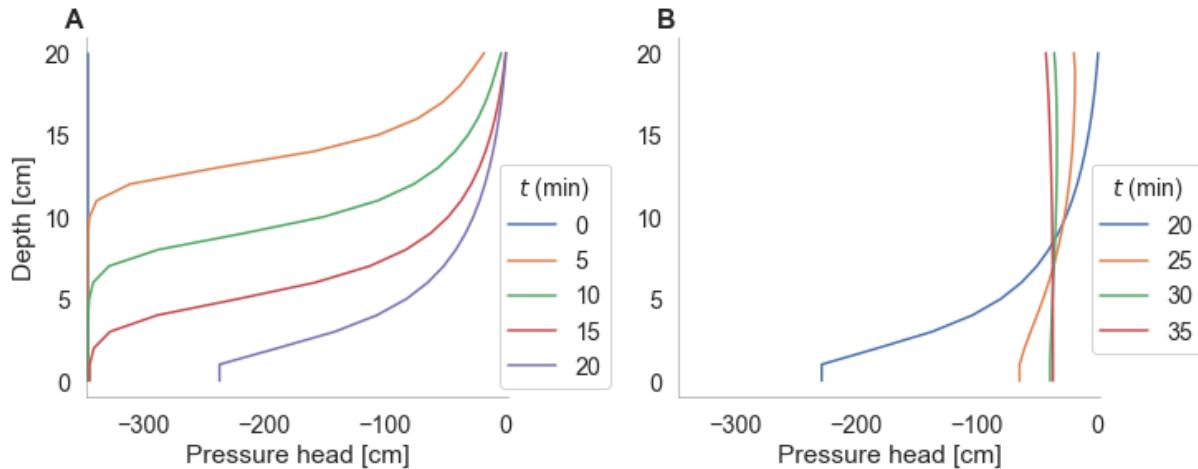


Figure 2.1: (A) Simulated wetting front for a 20 minute storm with $p = 10\text{cm/hr}$ and (B) drainage following the end of the storm. The soil is uniform, with $K_s = 4 \text{ cm/hr}$ and van Genuchten parameters as given in Table 2.1 (non-seal layer). The soil is initially at field capacity ($H = -342 \text{ cm}$). The lower boundary condition is free drainage.

2.3 The Saint Venant Equations

The integral form of the 2D Saint Venant Equations can be written:

$$\frac{\partial}{\partial t} \int_{\Omega} \mathbf{U} d\Omega + \oint_{\partial\Omega} (\mathbf{F} dy - \mathbf{G} dx) = \int_{\Omega} \mathbf{S} d\Omega \quad (2.3)$$

where $\mathbf{U} = (h, hU, hV)^T$, Ω is the volume element over which the equations are integrated, and

$$\mathbf{F} = \begin{pmatrix} hU \\ hU^2 + \frac{1}{2}gh^2 \\ hUV \end{pmatrix}; \quad \mathbf{G} = \begin{pmatrix} hV \\ hUV \\ hV^2 + \frac{1}{2}gh^2 \end{pmatrix}$$

where h is the flow depth, and U and V are the vertically-averaged velocities in the x - and y - directions, respectively. The first equation represents mass conservation, and the second and third represent momentum conservation. S is the source term, given as:

$$\mathbf{S} = \begin{pmatrix} p - i \\ -gh\frac{\partial z}{\partial x} - ghS_{f,x} + \frac{(p-i)U}{2} \\ -gh\frac{\partial z}{\partial y} - ghS_{f,y} + \frac{(p-i)V}{2} \end{pmatrix} \quad (2.4)$$

where i is the infiltration rate, and $S_{f,x}$ and $S_{f,y}$ are the x and y components of the friction slope, S_f , which represents the effect of bed and other shear stresses on the flow. Specification of the S_f is required to close the SVE, and S_f is parameterized as a function of the flow conditions (local depth and velocity) and roughness properties of the local land surface.

To solve the 2D SVE, the finite volume model of Bradford and Katopodes [2001] was selected because it conserves volume and can handle wetting and drying in the domain. Finite volume methods use the integral form of the SVE, which are solved in computational cells, and evaluate the fluxes between cell faces by solving a Riemann problem [Bradford and Katopodes, 2001]. The model uses the monotone upstream scheme for conservation laws to achieve second-order spatial accuracy, in conjunction with predictor-corrector time-stepping to provide a second-order accurate solution.

Boundary conditions are applied through the use of ghost cells that are adjacent to the boundary, but outside of the domain. The boundary conditions are specified by placing the appropriate values in the ghost cell, so the boundary fluxes can be solved for in the same manner as the fluxes between non-boundary cells. Typical boundary conditions are closed (solid wall), open (outflow) and specified inflow. At an outflow boundary, h , U and V are linearly extrapolated from the domain. At a wall boundary, velocities are specified in the ghost cell such that the perpendicular velocity at the wall is zero and the parallel velocity is unchanged. For a supercritical inflow boundary, h , U and V are specified in the ghost cells, while for a subcritical inflow boundary, only two independent quantities may be specified and the remaining quantity is extrapolated from the flow domain [Bradford and Katopodes, 1999].

While the Bradford and Katopodes [2001] scheme offers many advantages, stability constraints limit the range of conditions that can be simulated. Flow instabilities tend to arise for large gradients in the water surface, due either to steep topographic slopes or sharp contrasts in infiltration capacity. Such instabilities can be avoided with sufficiently small timesteps (i.e. 0.5 ms for a 30% slope); however, this can lead to very long simulation times. For example, simulating a 60 minute storm with 0.5 ms timesteps on a 100×50 domain requires approximately 35 hours.

2.4 Coupling the SVE and Richards Equation

Overland flow and infiltration are dynamically coupled. At the soil surface, the infiltration rate is determined by the presence or absence of water, in addition to a number of factors such as the soil type and moisture profile. Likewise, the flow depth at a given location is determined by the infiltration rate, lateral advection and rainfall inputs. To represent this connection in the model, the SVE and Richards equation are coupled through their respective boundary conditions.

The model components are coupled at each grid cell in two steps: the depth from the SVE solver provides the surface boundary condition to the Richards equation solver, and

the infiltration rate from the Richards equation solver is used by the SVE source term. This requires that several cases be accounted for: (1) no rain and no ponding, (2) rain but no ponding, (3) ponding (with or without rain). In case (1), a no flux boundary condition is applied at the surface. In case (2), the Richards equation solver computes a potential infiltration rate (PI), defined as the infiltration rate that would occur with $H=0$ at the surface, and compares this value to the rainfall intensity, p . If p exceeds the potential infiltration rate, ponding begins and the boundary condition switches to case (3). Otherwise, the potential infiltration rate is greater than p , and the infiltration rate is equal to p . Finally, in case (3), the upper boundary condition H is equal to the ponding depth h . These cases are schematically illustrated in Figure 2.2.

Because of the difference in the coordinate systems of the SVE and Richards equation, there can be a difference in the surface area available for infiltration in the z-direction versus the vertical. For steep slopes, this could result in an overestimation of gravity driven infiltration, which only occurs in the vertical. To correct for this, the infiltration flux could be multiplied by a correction term of $[1 - \cos(\beta)]K(\theta)dx^2$, where β is the slope angle.

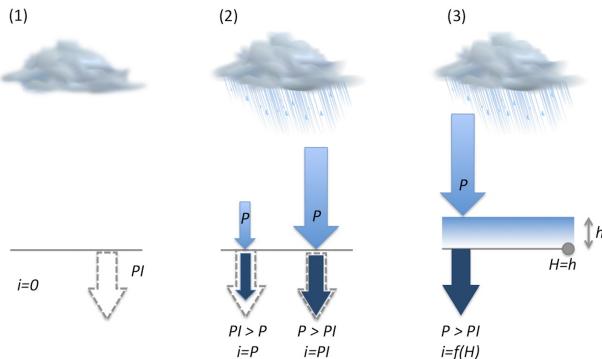


Figure 2.2: Schematic to illustrate the boundary condition cases for Richards equation. PI is the potential infiltration rate, which is computed to determine whether the rain intensity exceeds the infiltration capacity of the soil (in which case ponding occurs).

2.5 Model domain and boundary conditions

The explored modeling domain is a patchily vegetated, arid hillslope, which is loosely based on the Lehavim Long-Term Ecological Research site [Chen et al., 2013], located in the Negev desert, in the lower border of the semiarid zone of Israel. The site is characterized by hilly terrain, emergent rock formations, and sparse vegetation (approximately 25% perennial vegetation cover) with scattered dwarf shrubs. The soils are brown lithosols and arid brown loess, and prone to physical crust formation [Chen et al., 2013]. Infiltration rates are low in bare-ground areas, due to the formation of physical surface crusts [Assouline, 2004], but are higher under vegetation cover due to root activity and protection of the soil surface against

rainsplash by the canopy [Thompson et al., 2010]. The landscape is spatially organized as a two-phase mosaic of shrub and bare open soil patches [Svoray et al., 2008].

In the model simulations, the landscape is similarly represented as a two-phase mosaic of bare and vegetated cells, where the presence or absence of vegetation determines the surface roughness and soil infiltration capacity (see schematic in Figure 2.3). The soil domain is described first, followed by the overland flow domain.

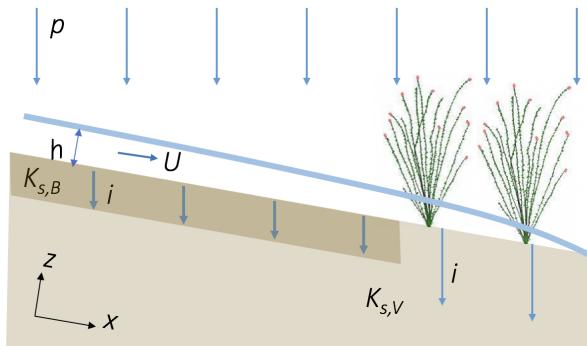


Figure 2.3: Schematic of the model domain. The darker shading under the non-vegetated regions represents the surface crust. $K_{s,V}$ and $K_{s,B}$ represent the saturated hydraulic conductivities of undisturbed soil and surface crust, respectively.

Soil domain

Bare soil is represented by a two-layer model, with a thin, low-permeability layer representing the surface crust. The vegetated patches and lower layer of the bare soil (i.e. the undisturbed soil) have the same infiltration parameters. The saturated hydraulic conductivities of surface crust and undisturbed soil are independently specified, and differ by one to two orders of magnitude. The van Genuchten parameters are from the Lehavim field site, listed in Table 2.1. The initial soil moisture content is set to field capacity ($H = -342$ cm).

The lower soil boundary is free draining; however, the plants are assumed to uptake all water that infiltrates in vegetated patches. Consequently, the soil depth is interpreted as the depth over which the soil is resolved by the Richards equation solver, rather than as the physical soil depth (from the perspective of plant water availability). Sensitivity analyses indicate that resolving the upper 20 cm versus 40 cm of the soil results in minimal differences, so for computational purposes, the soil depth is set to 20 cm in most simulations.

Overland flow domain

The simulation domain is restricted to rectangular planar hillslopes, with slope angles ranging from 0.5 to 30%. The upper boundary is closed (no flux) to simulate a ridge, and the lower

Table 2.1: Van Genuchten parameters of seal and underlying soil layers. θ_S and θ_R are the saturated and residual soil water content, respectively.

	seal layer	undisturbed soil
n	1.43	1.47
α_ν	0.0078	0.0096
θ_S	0.450	0.472
θ_R	0.0394	0.0378

boundary is open to represent a channel (or more hillslope). The lateral boundaries are also closed. The hillslope length-to-width ratio is 2:1 in most simulations, where the longer hillslope length is motivated by the more interesting (non-symmetric) along-slope boundary conditions.

The surface roughness is smaller for bare soil areas than for vegetated patches, reflecting the greater resistance to flow exerted by the vegetation. Manning's equation is used to specify the surface resistance in Chapters 5 and 6, following a sensitivity analysis to the choice of a roughness scheme in Chapter 3.

2.6 Implementation

The Richards equation solver is the more time-intensive component of the model, because it involves a matrix inversion for each time step and grid cell. For computational efficiency, the overland flow and infiltration components are gridded to different time steps. The Richards equation solver updates the infiltration rates every 2 seconds, and the infiltration rate is held constant between updates. The SVE time steps are restricted by the CFL condition to ensure numerical stability, and are consequently much smaller (0.5-10 ms depending on the slope angle). Sensitivity analyses indicate that the simulation outcomes are insensitive to the SVE time step, provided the model is stable.

The SVE solver uses a minimum depth threshold, ϵ , below which the momentum equations are not solved. Following Bradford and Katopodes [2001], $\epsilon = 0.2\text{mm}$. Sensitivity tests (comparing $\epsilon = 0.2, 0.5$ and 0.1 mm) indicate that model performance is best for this $\epsilon = 0.2\text{mm}$, in terms of model stability and hydrograph quality.

The model simulations are run on the Berkeley supercomputer Savio, enabling the simulation of up to 20 cases in parallel. The core model is written in Fortran, with wrapper functions in Python to interface with Savio and deploy simulations in parallel (using the multiprocessing library). Most simulations are run for factorial combinations of hillslope and storm parameters. For example, Figure 2.4 shows the runoff hydrographs for a set of 80 simulations, with factorial combinations of 4 rainfall rates, 5 vegetation densities and 4 patch length scales. This figure is included to illustrate how the model is constructed to run numerous simulations in batches, and the parameter details are not included.

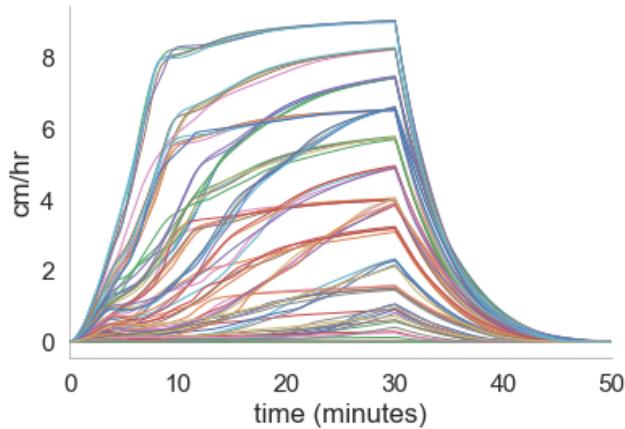


Figure 2.4: Hydrographs from multiple simulations run in parallel, for factorial combination of hillslope parameters. This figure is for illustrative purposes only.

As an example, Figure 2.5 displays a simulation of a 30 minute, 2.5 cm/hr storm on a 2% incline with randomly-generated vegetation patterns. Panel A shows a bird's eye view of the vegetation field and topography, with the top of the hillslope at the top of the page. Panels (B) and (C) show the cumulative infiltration depth and maximum flow velocity velocities, respectively.

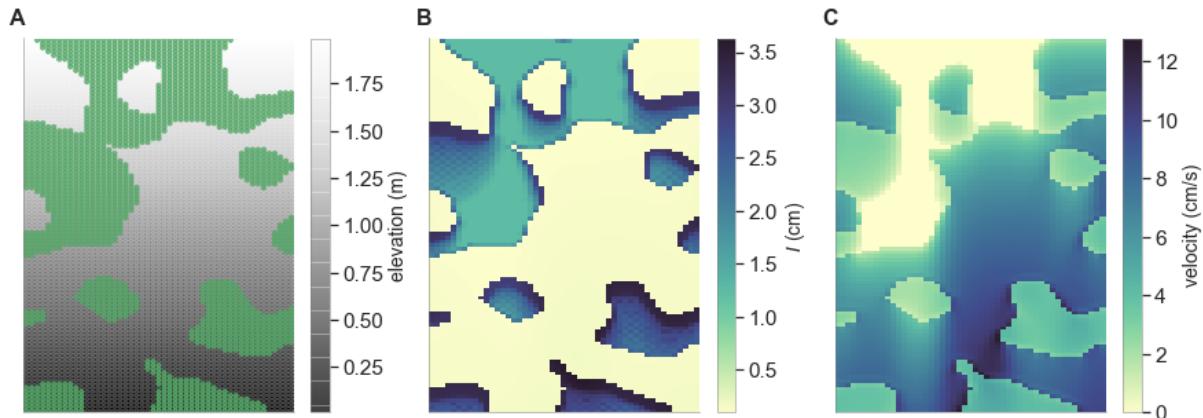


Figure 2.5: Sample model output from the coupled SVE-R model showing (A) the input vegetation field, (B) the resultant infiltration map and (C) the maximum flow velocities at the end of the rain. The hillslope parameters are: $K_s=2.0$ cm/s, $p=2.5$ cm/s, $S_o=2\%$. In (B), the darker regions correspond to vegetation patches with higher infiltration rates than the surrounding bare soil areas. Greater infiltration depths are observed down-slope of the bare soil areas, where the run-on source area is larger. In (C), the faster flow velocities are found over the bare soil areas, and where there is less upslope vegetation.

2.7 Model Validation with the Kinematic Wave Approximation

Model validation is concerned with the coupling of the SVE and Richards equation components, as these components have both been previously validated. For this, an analytical solution for a coupled infiltration - runoff problem was employed. This validation method only explores the kinematic regime, which is close to most of the real world cases of interest (see Chapter 3). While there is no reason to expect the validity of the model to change for non-kinematic cases, the coupled model has not been tested for strongly non-stationary flows.

The model was validated against the analytical solution of Giraldez and Woolhiser [1996], using the method of characteristics to solve for runoff on a plane under constant rainfall and infiltration specified by Philips equation [Philips, 1957]:

$$i = \frac{1}{2}St^{-1/2} + K_s \quad (2.5)$$

where S is the soil sorptivity and K_s is the saturated hydraulic conductivity.

The validation simulations consist of 20 minute rainstorms on a homogeneous planar hill-slope with roughness specified by Manning's equation and infiltration solved with Richards equation. The sorptivity in Philips equation is determined with a 1D Richards equation solver, by matching the cumulative infiltration at 20 minutes. Table 2.2 summarizes the model parameters and simulation results, with the discrepancies between model simulations and analytical predictions summarized by the percent differences between runoff hydrographs (Q) and infiltration fractions (f_I). Q ranges from 0.4% to 1.4% and f_I ranges from -3.9% to 4.6%, showing good agreement. Perfect agreement is not expected, as the method of characteristics uses the kinematic wave approximation with Philips equation, as opposed to the SVE with Richards equation.

Validation of the model is light, involving only a kinematic test case, because the model consists of two pre-existing, extensively tested components [Bradford and Katopodes, 2001, Celia et al., 1990]. Validation of the Bradford and Katopodes [2001] overland flow model included a dam break problem on an irrigation basin with complex topography, demonstrating the model's ability to handle a wetting front, complex topography and non-kinematic flow conditions. The only change from the pre-existing models to the current one is that the SVE and Richards equation are dynamically coupled. Further testing of the model is ongoing with experimental runoff data.

Table 2.2: Summary of the model validation simulations. Differences between model simulations and analytical predictions are given for the runoff hydrographs (Q) and infiltration fractions (f_I). Note that the differences are given as percentages.

p	S_o	$K_{s,bare}$	Q	f_I
2.7	1.0	0.034	0.7	1.7
2.7	1.0	0.340	0.1	0.1
2.7	5.0	0.034	1.0	2.6
2.7	5.0	0.340	0.9	-0.8
2.7	10.0	0.034	1.1	3.5
2.7	10.0	0.340	1.1	-0.9
2.7	20.0	0.034	1.1	4.6
2.7	20.0	0.340	1.2	-1.1
5.0	1.0	0.034	0.8	-3.9
5.0	1.0	0.340	0.4	-0.0
5.0	5.0	0.034	0.9	-3.2
5.0	5.0	0.340	0.9	-1.4
5.0	10.0	0.034	1.0	-2.4
5.0	10.0	0.340	1.1	-1.9
5.0	20.0	0.034	1.1	-1.3
5.0	20.0	0.340	1.4	-2.1

Chapter 3

Hydrological sensitivity to the representation of surface roughness in shallow overland flows

3.1 Introduction

The Saint Venant Equations do not form a closed system of equations, and require the specification of the friction slope, S_f , which represents the effect of bed and other shear stresses on the flow. To close the equations, S_f is parameterized as a function of the flow conditions (local depth h and velocity u) and environmental factors contributing to frictional dissipation.

The specification of S_f for real surfaces has produced a bewildering array of roughness schemes [Brutsaert, 2005, Kirstetter et al., 2016, Mügler et al., 2011, Katul et al., 2011, Cheng and Nguyen, 2010], relating energy dissipation (the loss of kinetic energy to the production of turbulent eddies) to surface roughness (a physical measure or parameterization of the elevation variability of the local land surface obstructing flow), and flow resistance (a measure of how the local land surface resists flow, given by e.g. a friction factor). Derived from empirical and theoretical studies, these roughness schemes cover a variety of conceptual representations of the surface roughness and its interaction with the flow (see Table 3.1). Navigating the choice of such schemes is difficult: the sensitivity of predictions to the choice of scheme is hard to assess, because the different schemes are associated with very different parameterization requirements for the same land surface, which impedes inter-comparison. The evaluation of individual schemes against experimental data is also challenging, with many researchers finding that flow resistance parameters, which ideally would represent fixed properties of a land surface, vary with the depth and velocity of individual flow fields [Abrahams et al., 1994, Lawrence, 1997, Smith et al., 2007, Zheng et al., 2012, Pan et al., 2016]. Consequently, selecting and parameterizing the ‘correct’ roughness scheme for a given application remains something of an art form.

Table 3.1: Selected roughness schemes to describe the friction slope from the literature. Note the distinct surface roughness parameterizations (α) and scaling between velocity (U) and depth (h) in each scheme. The Poiseuille, Manning and Darcy-Weisbach schemes are well-known, and parameters in the expressions include the kinematic viscosity of water ν and acceleration due to gravity (g), Manning's roughness coefficient n , and the Darcy-Weisbach resistance factor f . Parameters in the remaining schemes are equation-specific constants described in Appendix A.

Name	Resistance equation	Roughness (α)	References
Poiseuille	$U = \frac{1}{\alpha} h^2 S_f$	$\frac{3\nu}{g}$	Brutsaert [2005]
			Kirstetter et al. [2016]
Manning	$U = \frac{1}{\alpha} h^{2/3} S_f^{1/2}$	n	Brutsaert [2005]
			Smith et al. [2007]
Darcy-Weisbach	$U = \frac{1}{\alpha} h^{1/2} S_f^{1/2}$	$\sqrt{\frac{f}{8g}}$	Brutsaert [2005] Cea et al. [2014]
Depth-dependent	$U = \frac{1}{\alpha} h S_f^{1/2}$	$n_o h_o^{1/3}$	Mügler et al. [2011]
Manning			Jain et al. [2004]
Transitional/ Mixed	$U = \frac{1}{\alpha} h S_f^{1/2}$	α	Brutsaert [2005] Maheshwari [1992]
Poggi	$U = \frac{1}{\alpha} h^{1/2} S_f^{1/2}$	$\frac{\beta}{\sqrt{g}} \exp\left[\frac{H_c}{2\beta^2 L_c}\right]$	Katul et al. [2011]
Cylinder array	$U = \frac{1}{\alpha} S_f^{1/2}$	$\sqrt{\frac{C_d m D}{(1-\rho) 2g}}$	Cheng and Nguyen [2010] Wang et al. [2015]

In this chapter, I aim to demonstrate that recourse to a simplified situation: kinematic flow, facilitates inter-comparison of roughness schemes and allows an assessment of sensitivity to the selection of a scheme, at least for situations where the exposure of the flow to roughness elements is similar for all flow depth conditions. Under kinematic conditions, the momentum equation (Equation 1.2) reduces to a force balance between acceleration due to gravity and flow resistance by frictional dissipation:

$$S_f - S_o = 0 \quad (3.1)$$

By inspection of the roughness schemes presented in Table 3.1, the local flow velocity, depth, friction slope and roughness coefficient are related via the kinematic resistance equation [Brutsaert, 2005]:

$$U = \frac{1}{\alpha} h^m S_f^\eta \quad (3.2)$$

where α describes the roughness properties of the surface over which flow occurs. η describes how the flow resistance scales with hillslope gradient, and is typically 1/2 for turbulent flow and 1 for laminar flow.

The exponent m is related to the flow regime: for turbulent flow over homogeneous, planar surfaces, it derives from the vertical profile of the horizontal velocity $u(z)$:

$$\frac{u(z)}{u^*} = A \left(\frac{z}{z_o} \right)^M \quad (3.3)$$

where u^* is the friction velocity, z_o is a characteristic length scale of the surface, and A and M are constants (see, for example, Chen [1991] or Brutsaert [2005]). M is related to m as: $M = m + 1/2$. For laminar conditions, plane-parallel flow can be solved exactly from the Navier Stokes equations, yielding a parabolic velocity profile and $m=2$. For heterogeneous surfaces with emergent roughness elements or flow obstructions the origin of m is more complicated, and it may not be directly relatable to the vertical velocity profile, but rather represents an emergent property of the flow. One example is the situation where exposure to roughness elements (and thus α) decreases with flow depth, such that the value of m depends on both flow regime and the vertical profile of roughness [Jain et al., 2004, Mügler et al., 2011, Lawrence, 1997].

Combining Equations 3.1 and 3.2:

$$U = \frac{1}{\alpha} h^m S_o^\eta \quad (3.4)$$

Equation 3.4 is derived from the approximation that the frictional retardation of the flow is equal and opposite to the downslope component of water weight. For kinematic flow on a uniform slope, the prescription of a roughness scheme enforces a fixed relationship between flow depth and velocity, with the effects of the slope gradient and surface roughness independently specified. Figure 3.1 illustrates how surface roughness (α) and flow regime (m) change the nature of this relationship for a fixed slope. The units and interpretation of α are flow-regime specific (indicated by the value of m), and cannot be directly related to the physical characteristics of a surface. For a given flow regime, α should ideally be a function of surface characteristics alone, with increasing surface roughness corresponding to increasing α values.

There are many examples of more complex (non-power law) roughness schemes, for example, the Colebrook-White or Keulegan friction factor parameterizations [Cea et al., 2014]. In most cases, such schemes could be approximated by a power law form for a sufficiently restricted range of flow depths and velocities [Chen, 1991]. Furthermore, the analysis presented here could be generalized to non-polynomial forms of the resistance equation, and the results would likely be similar to those presented here.

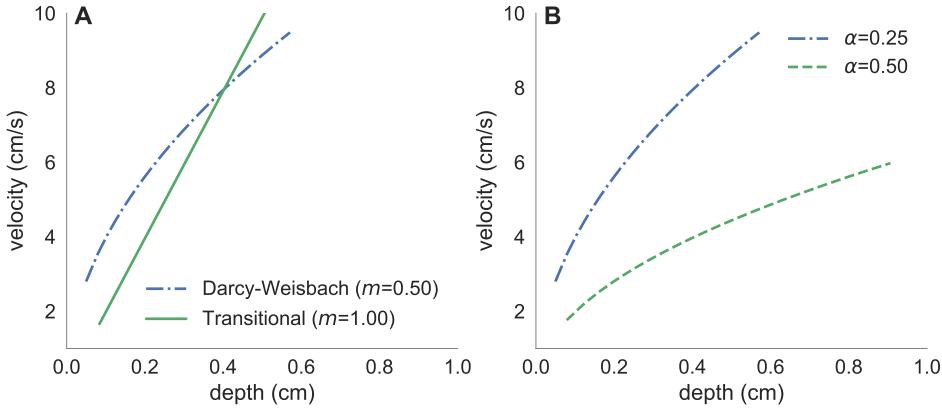


Figure 3.1: Illustration of the independent roles of the landscape roughness parameterization (α) and the $U-h$ scaling (m) on predicted kinematic relationships between velocity and depth. (A) shows a landscape where α is ‘matched’ to represent a common roughness surface between a turbulent ($m = 0.5$) and mixed ($m = 1$) flow regime. Conversely, (B) shows turbulent flow on two landscapes with different roughness values. The α -matching approach is landscape and discharge dependent, and details are outlined in the methods section.

In this study, I make three, interrelated arguments that derive from the well-known kinematic resistance equation:

The first is that Equation 3.4 imposes a scaling between velocity and depth ($U-h$ scaling) that provides a reasonable approximation to flow in non-kinematic situations, specifically unsteady, non-homogeneous flows on hillslopes with patchy vegetation. For uniform flow over smooth terrain, the value of m is determined by the shape of the vertical velocity profile in the flow, while for complex surfaces, m may be an effective parameter.

Secondly, if such $U-h$ scaling persists for the range of flow conditions encountered in shallow overland flow, then controlling for the form of the $U-h$ scaling relation is a necessary precondition when assessing the roughness characteristics of any real slope. A corollary of this argument is that, if the $U-h$ scaling is not controlled for, flow-dependent behavior in the inferred resistance may be mistakenly attributed to varying surface roughness. For example, in studies of overland flow resistance, a common area of inquiry is the relationship between resistance to overland flow and the Reynolds number, where the resistance is most commonly expressed as the Darcy-Weisbach friction factor (f), but other forms such as Manning’s n or the Chézy coefficient C are also common. f relates the head loss due to friction to the average kinetic energy of the fluid flow, and has several advantages as a measure of resistance relative to n and C : it is dimensionless, applicable to both laminar and turbulent flow regimes, and interpretable as a drag coefficient in some cases [Ferguson, 2007]. In combination with the flow continuity equation, Equation 3.4 can be expressed as a power law relationship between f and Re with the exponent value determined by m (see Appendix A). Controlling for this implied relationship can help to distinguish complex $f\text{-Re}$

relationships that arise as a consequence of the flow regime from f -**Re** relationships arising from more complex processes, such as changing inundation of surface roughness elements [Lawrence, 1997], deformation of the vegetation with [James et al., 2004, Järvelä, 2002], or flow heterogeneities such as turbulent eddies or filaments [Dunkerley, 2003].

Finally, if U - h scaling is controlled for, different roughness schemes can be parameterized to represent equivalent surfaces, allowing for rigorous inter-comparison. With a common representation of the surface roughness (via parameterization of α), I show that important hydrological outcomes, including water balance partitioning, flow velocity and the runoff hydrograph characteristics, display little sensitivity to the selection of a roughness scheme.

These findings offer some guidance for practitioners aiming to make predictions about the consequences of shallow overland flow on hillslopes. Principally (i) accounting for flow regime when assessing the behavior of surface resistance can avoid mathematical artifacts in the resulting f -**Re** relations, and (ii) that modelers should focus effort on the parameterization, rather than selection, of surface roughness schemes, as uncertainty is more likely to arise from such parameterization than from the details and conceptual formulation of these schemes. While complex, depth- and flow-dependent roughness behaviors will continue to be identified in overland flow settings [Smith et al., 2007, Smith, 2014, Lawrence, 1997, Dunkerley, 2003], these findings may assist in distinguishing these real sources of complexity from artifacts of common analytical approaches.

3.2 Methods

I modeled a series of rainstorms on patchily-vegetated hillslope surfaces using the shallow water equations coupled to a one-dimensional Richards equation solver to represent infiltration. The shallow water equation solver is identical to that used by Bradford and Katopodes [2001], and the Richards equation solver is that outlined by Celia et al. [1990] (see Chapter 2). The explored modeling domain is a patchily vegetated, planar, arid hillslope, which is loosely modeled after the Lehavim Long-Term Ecological Research site [Chen et al., 2013]. Vegetation is included as a binary field, and used to specify variable infiltration rates and roughness parameters. Infiltration rates are low in bare-ground areas, due to the formation of physical surface crusts [Assouline, 2004, Assouline et al., 2015], but are higher under vegetation cover due to root activity and protection of the soil surface against rainsplash by the canopy [Thompson et al., 2010]. For a storm simulation, the model output includes the overland flow fields, runoff hydrograph and infiltration map. Figure 3.2 presents an example model domain with vegetation (A) and cumulative infiltration depth (B) to illustrate. The model was used in conjunction with solutions to the kinematic wave approximation to address the three arguments described in Section 3.1.

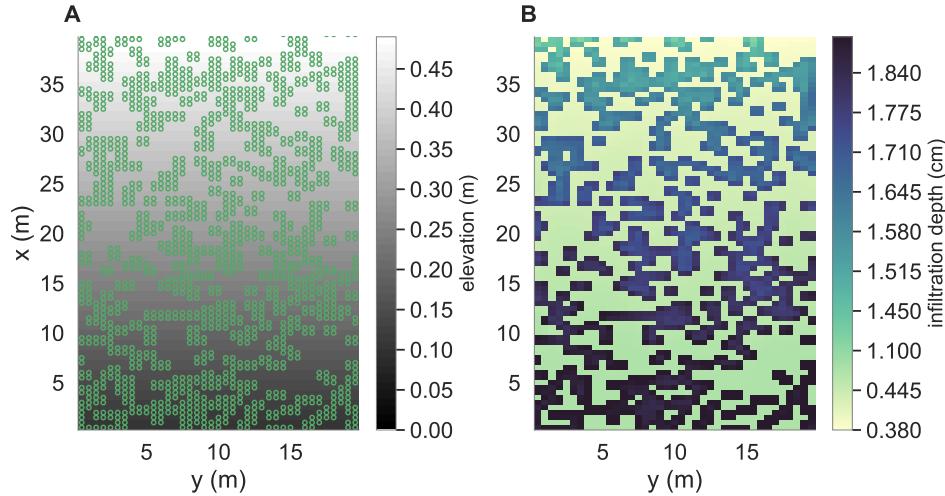


Figure 3.2: Example model set-up and infiltration output. (A) map of topography (grey-scale) with spatially random vegetation field (green circles indicate vegetated 0.5×0.5 m cells). (B) Infiltration map resulting from a 20 min storm with $p = 2.7$ cm/hour.

Does U - h scaling persist in non-kinematic flows?

The kinematic wave approximation assumes steady, uniform flow, which is not generally true for flows on complex, heterogeneous hillslopes. For realistic storm conditions and planar slopes, however, the range of flow depths and velocities are such that these departures from kinematic conditions might not cause large deviations from the U - h scaling associated with kinematic conditions. To test this hypothesis, shallow water equation simulations of a rainstorm were run for factorial combinations of the hillslope parameters shown in Table 3.2. The mean departures from kinematic conditions (ΔS) and kinematic U - h scaling (Δm) were computed. Multiple roughness schemes were included (see Table 3.1 for details), in order to explore a range of flow regimes (m values).

For each model simulation, ΔS was quantified as the mean difference between the simulated friction slope S_f^{sim} and the topographic slope for all cells at all times:

$$\Delta S = S_f^{sim} - S_o \quad (3.5)$$

Δm was computed from the slope of the $\log(U)$ - $\log(h)$ plot for all locations at all times to obtain a modeled m^{sim} , which was compared to the m specified in the model's roughness formulation:

$$\Delta m = m^{sim} - m \quad (3.6)$$

Because roughness and infiltration characteristics differ between bare and vegetated grid cells, ΔS and Δm were estimated separately for bare and vegetated sites.

Table 3.2: Parameters for the model simulations. Where multiple parameters are listed, the cases were run factorially to explore all parameter combinations.

Variable	Symbol	Values
Hydraulic conductivity (bare soil)	$K_{S,B}$	0.34, 0.034 cm/hr
Hydraulic conductivity (vegetated sites)	$K_{S,V}$	3.4 cm/hr
Slope gradient (%)	S_o	1%, 10%
Rainfall intensity	p	2.7, 5.0 cm/hr
Vegetation fraction	ϕ_V	0.1, 0.3, 0.5
Manning's n (bare soil)	n_{bare}	0.03
Manning's n (vegetated sites)	n_{veg}	0.1
Domain size	L_x, L_y	20m \times 40m
Grid resolution	dx	0.5m
Storm duration	t_{rain}	20 minutes
Roughness scheme	m	Laminar, Transitional, Manning, Darcy-Weisbach, Cylinder Array

What are the implications of $U-h$ scaling when evaluating landscape resistance behaviors?

If kinematic scaling persists, the functional form of the roughness scheme implicitly imposes a $U-h$ scaling relationship on the resulting flow. The form of this relationship has the potential to confound interpretation of landscape roughness and its variation with environmental variables.

To illustrate the potential for such confounding behavior with a simple kinematic case, synthetic, kinematic runoff data was generated using various roughness schemes with fixed roughness (constant α). The kinematic wave approximation was imposed to ensure that $U \sim h^m S_o^\eta$. The schemes include a Darcy-Weisbach scheme with constant f ($m=1/2$), a cylinder array scheme ($m=0$), a transitional scheme ($m=1$), laminar flow ($m=2$), and Manning's equation ($m=2/3$). For each of the roughness schemes, variations of the resistance with changing hillslope gradient and Reynolds number were naively examined, in order to characterize the implications of $U-h$ scaling on interpretations of flow resistance.

Are hydrological predictions sensitive to the choice of a roughness scheme (or only to its parameterization)?

To determine the sensitivity of the hydrological predictions to the choice of a roughness scheme, predictions made with different roughness schemes must be compared. To do this, the schemes need to be parameterized to represent equivalent land surface conditions. This process is referred to as parameter matching. The parameter matching approach assumes steady-state, kinematic flow on a uniformly vegetated hillslope under constant rainfall. In

this case, the steady-state flow rate $q(x)$ is determined by the rate of water input to the hillslope. For a constant rainfall rate p , and with a no-flux boundary condition at the divide, the discharge at any location x from the divide is given by:

$$q = Uh = px \quad (3.7)$$

This can be combined with Equation 3.4 to predict the flow velocity at all locations on the hillslope:

$$U^{m+1} = \frac{1}{\alpha}(px)^m S_f^\eta \quad (3.8)$$

I consider two different roughness schemes (denoted 1 and 2) and aim to find roughness values (α_1 and α_2) that represent the same hillslope conditions. This is achieved by fixing one of the roughness values, and then minimizing the RMS difference between U_1 and U_2 in Equation 3.8. A similar approach can be applied to minimize the depth differences between the schemes by combining Equations 3.4 and 3.7 to solve for h rather than U .

There are some immediate caveats to this approach for parameter matching. First, it assumes a homogeneous hillslope and steady state conditions, so parameters may not be matched in a way that minimizes differences for unsteady flow or a non-homogeneous hillslope. Second, different matched parameter values are obtained if depth differences rather than velocity differences are minimized. Finally, for several of the roughness schemes in Table 3.1, α is a function of multiple parameters, and this matching approach does not constrain specific parameter values without making further assumptions. However, the approach is broadly comparable to fitting a roughness scheme to field data, and ensures that, for a wide range of experienced flow conditions, schemes are optimized to an equivalent hillslope.

A range of planar, patchily vegetated landscapes with different slope angles and varying fractional vegetation cover were simulated, factorially varying hillslope characteristics so that 24 landscape and rainfall combinations were simulated. The vegetated patch roughness coefficients (α_{veg}) were matched to a ‘reference’ roughness scheme, taken as Manning’s equation with $n = 0.1$. For each roughness scheme and hillslope parameter case (slope, vegetation density, rainfall intensity, etc.), α_{veg} was obtained with the parameter matching approach described above. The rainfall intensity was included in Equation 3.8 as an effective rate, adjusted by the infiltration capacity of the soil ($p - \phi_V K_{S,V} - \phi_B K_{S,B}$). For bare ground, Manning’s equation with $n= 0.03$ was applied for all roughness schemes. The parameter matching approach was applied to the Poiseuille equation, despite its lack of a free parameter for flow over a smooth surface. The laminar roughness coefficient was matched because complex terrain may alter the hydraulic radius and thus the effective roughness parameter for laminar flow. Flume studies of laminar flow over natural surfaces have observed α values ranging from around 100 to 1800 [Dunkerley, 2001, Pan et al., 2016], providing justification for this approach. The model parameters are shown in Table 3.2, including those that varied and those that were held constant between simulations.

Hydrological outcomes, including the hydrograph characteristics, infiltration partitioning, and flow velocity, were computed for each simulation and compared between roughness

schemes. The hydrograph characteristics are represented by the time for the hydrograph to reach 90% of the peak flow (rising time, $t_{90\%}$), and the time after the rain ends for the hydrograph to recede to 10% of the peak discharge (recession time, $t_{10\%}$). The infiltration fraction (I) is defined as the ratio of mean infiltration depth to rainfall depth, and the vegetation infiltration fraction (I_{veg}) as the ratio of the mean infiltration depth to rainfall depth. Values of I_{veg} greater than 1.0 indicate that the mean vegetated patch infiltration depth exceeded the rainfall depth, so run-on from bare soil sites provided additional water as run-on. Flow velocities are compared using the mean and maximum velocities at the end of the rain, when the discharge is greatest and closest to ‘steady state’.

3.3 Results

Does U - h scaling persist in non-kinematic flows?

For all of the simulations, the flow was well described by the kinematic roughness approximation (small ΔS) and the associated U - h scaling (small Δm). As shown by the plot of ΔS versus Δm (Figure 3.3A), the average difference between friction and topographic slope is less than 0.1%, small by comparison to the topographic slope gradients of 1% and 10%. The differences between the simulated and prescribed m values are less than 0.08, which is significantly less than the m separation between any of the roughness schemes. Figure 3.3A shows the maximum Δm values increasing as ΔS increases, suggests that the kinematically-interpreted U - h scaling may not persist in highly non-kinematic flows; however, for a realistic range of typical overland flow conditions, kinematic U - h scaling persists despite unsteady and non-uniform flow behavior.

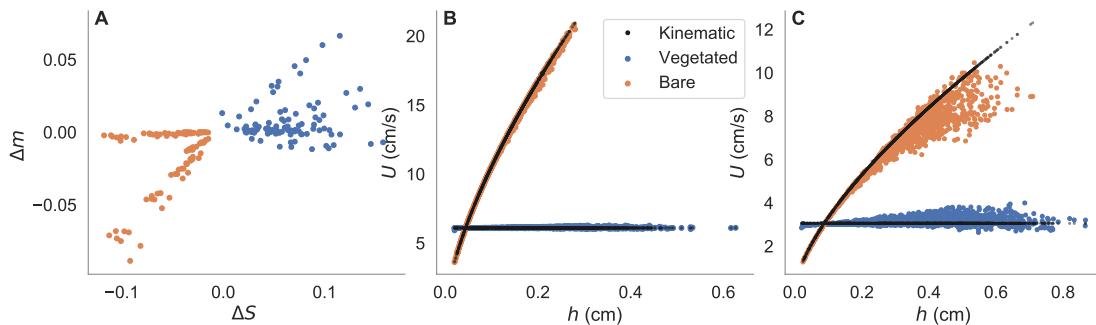


Figure 3.3: U - h scaling in non-kinematic flows. In (A), ΔS is the difference in between topographic and simulated friction slope (%), and Δm is the difference between the prescribed and simulated U - h scaling. (B) and (C) show example U - h plots where the flow is more and less kinematic, respectively. The simulation parameters are: cylinder array scheme, 5cm/hr rain, 50% vegetation cover and (B) 10% slope (C) 1% slope.

What are the implications of U - h scaling when evaluating landscape resistance behaviors?

Given that kinematic U - h scaling is likely applicable in many overland flow conditions, these scalings can be propagated into representations of landscape resistance such as the f - Re plot. Figure 3.4 shows how the flow resistance f varies with slope (A) and Reynolds number (B), depending on the choice of roughness scheme. The exception is the Darcy-Weisbach scheme with constant f (not shown), for which f is constant by construction. f is also slope-independent for laminar flow. Because α is held constant for each roughness scheme, the relationships in Figure 3.4 arise from the kinematic resistance equations associated with each scheme. Flow resistance decreases with increasing Re for schemes with $m > 1/2$, and increases with Re for the $m=0$ cylinder array scheme. In Figure 3.4A, the relationship between resistance and slope is more complicated, because it depends on both m and η .

Details about the mathematical relationship between the exponents in Equation 3.4 and the log f -log Re and log f -log S_o slopes are included in Appendix A for the overland flow scenario introduced in Section 3.2. The slopes are uniquely determined by the exponents in the kinematic resistance equation.

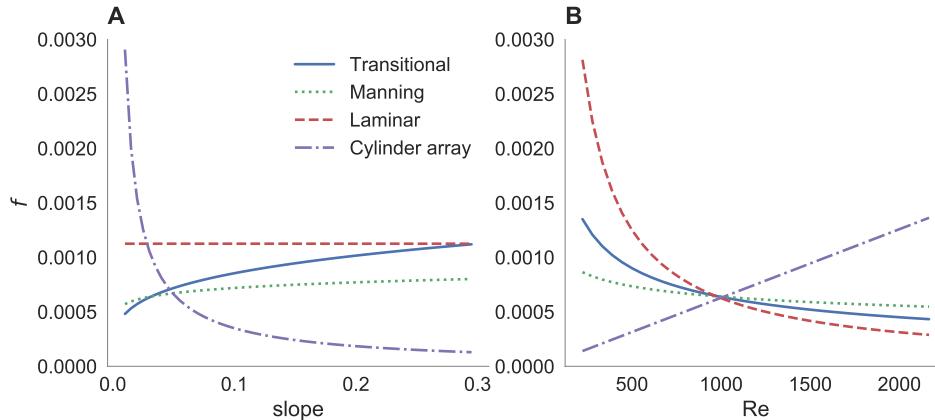


Figure 3.4: Darcy-Weisbach f evaluated for varying (A) slope and (B) Reynolds number at the outlet, illustrating the implications of various roughness schemes for the interpretation of flow resistance. In (B), Reynolds number was varied by changing the rainfall intensity. The U - h and U - S_o scaling associated with each roughness scheme imply a specific relationship between flow resistance, slope (A) and Reynolds number (B), independent of the surface roughness (α).

Are hydrological predictions sensitive to the choice of a roughness scheme?

Figure 3.5 shows example hydrographs for two modeled hillslopes, one with sparse vegetation cover (A) and one with dense vegetation cover (B), with the hydrographs produced by five separate roughness schemes following parameter matching. These results are similar to the other modeled hillslopes, which show near exact agreement in the hydrographs for hillslopes with sparse vegetation and close agreement for hillslopes with denser vegetation.

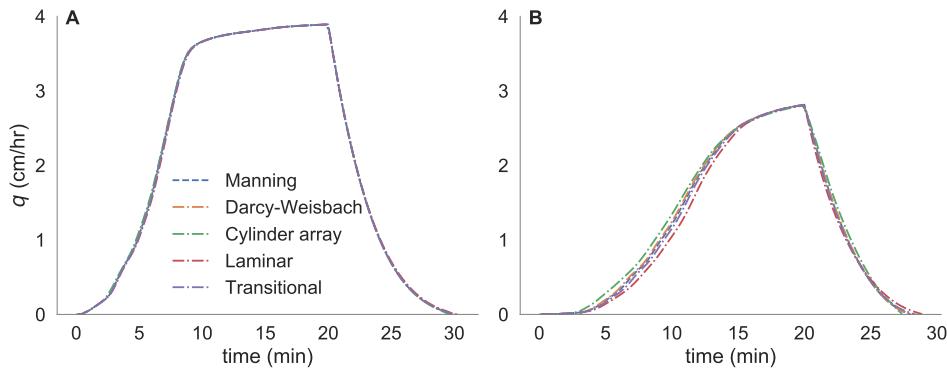


Figure 3.5: Example simulation hydrographs for sparse vegetation (A) and dense vegetation (B), showing good agreement. Hillslope parameters are $p=5$ cm/hr, $S_o=0.01$, $K_{S,bare}=0.34$ cm/hr and (A) $\phi_V=0.1$, (B) $\phi_V=0.5$. Other parameters are listed in Table 3.2.

Figure 3.6D displays the simulation hydrologic outcomes ($t_{90\%}$, $t_{10\%}$, I , I_{veg} , hillslope max and mean U) averaged across all parameter cases and grouped by roughness scheme. Four of the laminar simulations were discarded due to numerical instabilities in the velocity fields (simulations with $S_o = 0.1$ and $p = 5.0$ cm/hr), however, visual inspection of the hydrographs and infiltration maps did not show any problems. Accordingly, column L shows the laminar results with the unstable cases included in brackets.

Comparing the mean hydrograph rising and recession times by scheme, the results differ by less than 0.1 min (for the bracketed values). The infiltration fractions (I and I_{veg}) also show very close agreement, differing by less than 0.02 between schemes. For the mean and maximum velocities at the end of the rain, the mean outcomes differ by less than 0.06 cm/s for the non-laminar schemes. For the laminar velocities, the larger bracketed velocities reflect the instabilities in the flow fields, whereas the smaller, non-bracketed velocities reflect the fact that the discarded cases had a steeper slope and higher intensity rain, resulting in higher flow velocities.

In Figure 3.6A-C, the differences between paired Manning and non-Manning simulations are displayed as box-plots, grouped by roughness scheme and outcome variable. The differences range from -2% to 2% for I and I_{veg} , from -0.1 to 0.3 cm/s for the velocity fields, and from -0.4 to 0.4 min for the hydrograph rising and recession times. These differences sug-

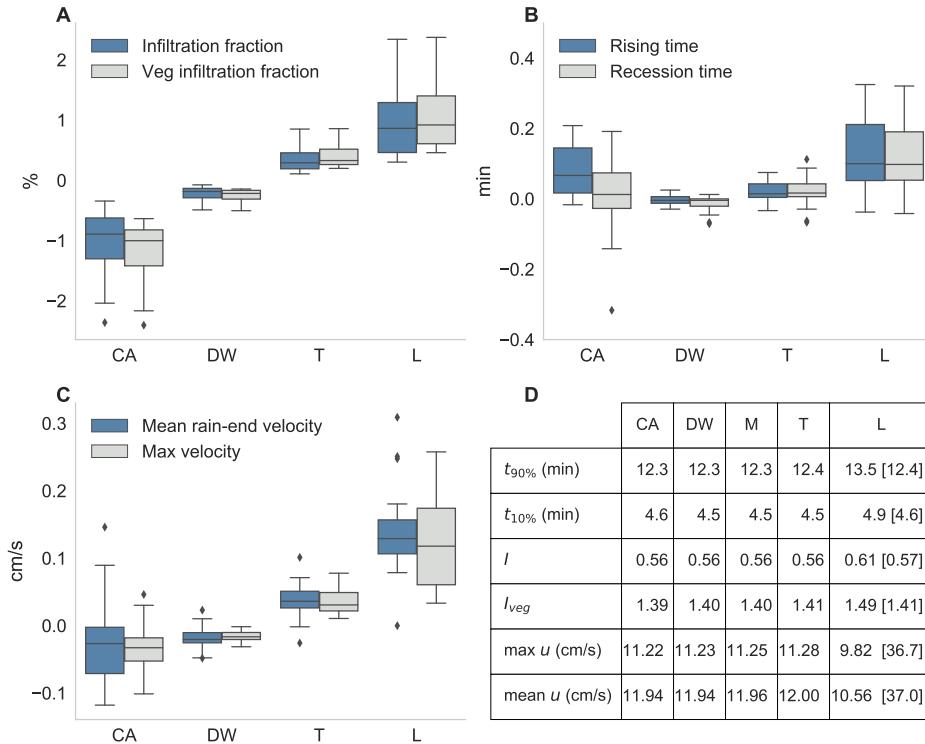


Figure 3.6: Box plots show the difference of each simulation from the paired Manning simulation with the same hillslope parameters: (A) the percent difference in infiltration fractions, (B) the hydrograph rising and recession times, (C) the mean and maximum velocities at the end of the rain. (D) shows the mean values for each scheme of all the parameter cases. Abbreviations indicate roughness schemes: Cylinder array (CA), Darcy-Weisbach (DW), Manning’s equation (M), Transitional (T) and Laminar (L). In column L, values in brackets include the omitted unstable simulations.

gest that the sensitivity to the choice of roughness scheme is small, provided the roughness parameters are appropriately matched.

3.4 Discussion

The kinematic scaling relationships are echoed in the flows simulated by the shallow water equation for a range of realistic hillslope scenarios. Flow is well described by the kinematic wave approximation and obeys the $U-h$ scaling implied by the prescribed roughness schemes, despite the conditions for applying the kinematic wave approximation not being met. The small ΔS values in Figure 3.3 show that the simulated flow fields are close to kinematic, with the difference between topographic and friction slope less than 0.1% for all simulations. The small Δm values correspondingly show that the $U-h$ scaling is close to that implied

by the kinematic resistance equation. This finding justifies the use of the kinematic wave approximation to (a) investigate how roughness scheme selection influences f -**Re** and f -slope relations, and (b) match roughness coefficients between different roughness schemes.

An important caveat to the kinematic scaling results derives from the underlying assumption that the model adequately represents the overland flow dynamics. While it introduces several non-kinematic aspects, the shallow water model does not explicitly simulate macroscopic roughness elements, microtopography, turbulent eddies or preferential flow paths. Instead, the simulated planar hillslopes are abstracted as square cells with differing roughness and infiltration parameterizations. Consequently, the kinematically-derived U - h scaling may not be applicable to other non-kinematic flows, specifically to cases with macroscopic roughness elements or non-planar topography.

Kinematic scaling persistence in overland flows could be further investigated with models that explicitly represent microtopography, vegetation or the 3D flow field. Modeling studies that explicitly represent flow obstacles such as vegetation stems [Kim et al., 2012] or microtopography [Razafison et al., 2012, Özgen et al., 2015] could examine whether these elements change the effective U - h scaling. For example, Kim et al. [2012] simulate overland flow through emergent stems represented as hexagonal obstacle cells and determine an effective Manning's n as a function of vegetation density and flow characteristics. Manning's equation is used to prescribe the base roughness, but how closely the model-simulated U - h scaling matches the Manning's equation $U \sim h^{2/3}$ scaling is not explicitly addressed.

The persistence of U - h scaling for non-kinematic conditions implies that the interpretation of roughness parameters for such overland flows should be cognizant of U - h scaling. Figure 3.4 illustrates how, for fixed roughness coefficients, different roughness schemes imply different relationships between flow resistance, hillslope gradient and **Re**. Flow resistance increases with Reynolds number for the cylinder array scheme, and the opposite relationship exists for the laminar, transitional and Manning's equations.

Of course, in real-world flows, a number of additional factors are likely to influence the observed f -**Re**. First, surface roughness may scale with flow depth: in cases of very shallow overland flows, progressive inundation of the roughness elements with increasing flow depths may alter the effective surface roughness [Lawrence, 1997, Abrahams et al., 1994]. Similarly, for higher discharge flows through vegetation, the surface roughness (more precisely, the vegetative drag coefficient) may vary with **Re** due to bending or deformation of the vegetation [Tanino and Nepf, 2008, James et al., 2004, Järvelä, 2002]. Second, f -**Re** relationships may reflect transitions in flow regime from laminar to mixed or turbulent. On heterogeneous surfaces, the development of flow filaments may contribute to increased spatial complexity and viscous dissipation in mixing regions [Dunkerley, 2003]. Additionally, turbulent eddies may develop with increasing Reynolds numbers and contribute to additional energy dissipation, altering the effective flow regime. Third, the principal sources of resistance may vary with flow depth. For example, James et al. [2004] observe a shift in the principle source of resistance from surface to vegetative drag with increasing flow depths, and an associated change in U - h scaling.

Interpretation of f -**Re** relationships has the potential to represent complex superposition

of these factors. It is therefore unsurprising that kinematic scaling arguments cannot explain the diversity of f - Re relationships observed in overland flow experiments. For flows with Reynolds numbers in the laminar flow range (<1000), studies have observed $f \sim \text{Re}^{-1}$ for a variety of surfaces (grass, bare soil, sandpaper), slopes and input methods [Dunkerley, 2001, Pan et al., 2016, Zhang et al., 2014]. This is consistent with laminar flow, and suggests the persistence of laminar $U-h$ scaling over more complex surfaces. However, other studies observe increasing f with Re [Abrahams et al., 1994, Yang et al., 2017], or non-monotonic f - Re relationships [Dunkerley, 2003]. Among these studies, there is no obvious correlation with surface type (grass, sandpaper, bare soil), vegetation type (synthetic or real), water input mechanism (rainfall vs. upslope) or experiment type (flume vs. field experiment).

Nonetheless, interpretation of experimental results can be usefully informed by understanding the f - Re relationships implied by $U - h$ scaling. For example, interpreted as a cylinder array scheme, a log f -log Re plot with slope between 0 and 1 could suggest a roughness coefficient that decreases with discharge due to vegetation bending or deformation (the kinematic wave approximation yields a log f -log Re slope of 1 for a cylinder array scheme). Interpreted as turbulent flow with constant f , the same plot would suggest a roughness coefficient that increases with discharge. The cylinder array and Darcy-Weisbach schemes lead to opposite conclusions about the roughness sensitivity to Reynolds number.

Kinematic scaling arguments offer a framework for distinguishing between the various factors controlling flow resistance, to the extent that this is possible. Measuring U and h simultaneously is an experimental challenge [Legout et al., 2012, Mügler et al., 2011, Tatard et al., 2008]; however, such measurements in addition to discharge can determine $U-h$ scaling. With $U-h$ scaling controlled for, variation in the roughness characteristics with discharge, due to submergence of roughness elements or bending of the stems, can be independently assessed. The roughness characteristics of real landscapes may truly vary with depth due to different submergence of roughness elements, but unless the $U-h$ scaling is known, such variation cannot be independently assessed.

The persistence of $U-h$ scaling implies that kinematic approaches can be used to match the parameterization of roughness schemes for a given land surface. With a simple kinematic approach, there are minimal differences in predicted hydrologic outcomes associated with the selection of a roughness scheme. That is, the results are more sensitive to roughness coefficient than to differences in the functional form of the resistance equation.

This conclusion is similar to the results of Cea et al. [2014] comparing high resolution simulations with experimental flow data on $1 \times 1\text{m}$ plaster moulds. Cea et al. [2014] explore the possibility of improving their model predictions by comparing multiple friction parameterizations, and match roughness parameters by calibration to experimental measurements. They find that all roughness parameterization give a similar agreement with their experimental data, after calibration of their respective friction parameters.

In practical applications, Figure 3.6 suggests that the selection of the correct roughness scheme is less important than appropriate parameterization of any selected scheme. In selecting a roughness scheme, other considerations can be taken into account. Some schemes will be more numerically stable than others. For example, when matched to represent equivalent

surfaces, a laminar scheme will tend to produce faster, shallower flows than a Darcy-Weisbach scheme with constant f , which may result in more numerical instabilities. For some schemes, the resistance coefficient can be related to measurable surface characteristics, while for others, calibration with runoff experiments are required. In Table 3.1, for example, the cylinder array and Poggi roughness coefficients depend on canopy features, whereas Manning's n is an empirical parameter requiring calibration [Cea et al., 2014]. Of course, efforts to relate a roughness parameter to surface characteristics (e.g. grain size) will encounter flow and slope dependencies if the U - h scaling is not fully known.

In conclusion, this study has shown that the selection of a roughness scheme, combined with the kinematic wave approximation, implies a unique U - h scaling that persists into non-kinematic flows. The extent to which such kinematic scaling persists for non-planar topography and macroscopic roughness elements is not known. This kinematic scaling persistence has implications regarding the interpretation of f - Re and f -slope relations; however, further research is needed to better disentangle the various factors known to influence these relationships. Finally, the kinematic wave approximation provides a suitable framework for matching roughness parameters, as evidenced by the close agreement between the various schemes in the simulation results. Consequently, choosing the correct roughness scheme is less important than correctly parameterizing any selected scheme. Despite this practical implication, these results do not assist in determining the correct roughness value for a given scheme. Thus, roughness parameterization remains an open question, subject to ongoing research.

Chapter 4

Emulating the SVE-R model with machine learning

4.1 Introduction

Desert landscapes are characterized by a feedback between firstly the spatial distribution of vegetated and bare sites, which determines the flow and infiltration patterns of water on the land surface, and secondly the conversion of rainfall into a spatial pattern of plant-available soil water resources, which determines the growth rates and drought mortality of vegetation. Although the plant and water systems are tightly coupled, their dynamics emerge on very different timescales: the rapid (minutes - hours) timescales associated with specific storm events, and the slow (years - centuries) timescales associated with the growth and mortality of individual slow-growing plants and the migration of vegetation communities through space. Similar timescale separations are common across biological systems [e.g. in enzyme kinetics, Briggs and Haldane, 1925, Michaelis and Menten, 1913], and typically pose challenges for analysis, because of the difficulty of resolving all the possible states of the system simultaneously. In general, to simplify the analysis of such systems, timescale separation procedures are employed, in which the state variables of interest (in this case biomass, soil water and surface water) are characterized as “slow” or “fast”, and the evolution of each subsystem (of the slow or the fast variables respectively) is solved on timescales appropriate to its rate of change, while holding the other system in pseudosteady conditions. Although there are many successful examples of applying timescale separation to the analysis of biological systems, the computational intensity of solving the “fast” overland flow processes in dryland ecohydrological systems (e.g. see Chapter 2) can make the application of timescale separation approaches intractable. Thus, to date, most models that attempt to represent the coupling of water redistribution and plant spatial organization over long timescales have adopted simplified phenomenological modeling approaches, for example, approximating overland flow as a diffusive process, and approximating storm occurrence with continuous rainfall [Rietkerk et al., 2002, HilleRisLambers et al., 2001]. The downside of such simplified ap-

proaches is that the details of the fast processes - which are explicitly omitted or averaged over by these phenomenological approaches - may be critical in setting the conditions under which the slow processes operate. This is certainly the case when considering how fast hydroclimatic processes - which I take here to encompass the “storminess” (i.e. frequency of storms), storm intensity and storm duration of the climate - influence the spatial pattern of water resources at the end of a storm. For instance, a low intensity storm is likely to infiltrate uniformly over a dryland hillslope, while a high intensity storm will generate runoff on low-permeability bare sites, resulting in a spatially variable pattern of infiltration.

Thus, dryland ecohydrology faces the challenge of representing fast within-storm processes in a computationally tractable fashion while maintaining high fidelity to the physics of runoff generation and infiltration. Here, I address this challenge by emulating the results of a coupled Saint Venant - Richards Equation (SVE-R) model with machine learning (ML): a broad suite of algorithms used to develop predictive models using a set of input data [Goldstein and Coco, 2015], typically applied to large, multidimensional data sets (e.g. social media data), where more common statistical methods such as multivariate linear regression are less successful [Pennacchiotti and Popescu, 2011]. In the earth sciences, machine learning has been applied to emulate the outcomes of complex models such as global climate models [Goldstein and Coco, 2015], and to emulate empirical observations [Tinoco et al., 2015]. Here, I use random forest (RF) regression to emulate the output of the SVE-R model for a range of storm and hillslope conditions. RF regression is an ensemble method that creates a “forest” of random, uncorrelated decision trees on sub-samples of the data to which it is fit, and uses averaging to improve the predictive accuracy of the algorithm while controlling for over-fitting [Pedregosa et al., 2011]. While a number of ML algorithms would be suitable for emulating the output of the SVE-R equations, random forests are relatively easy to train, fast to execute, and tend to perform well [Shih, 2012]. In this application, preliminary tests on SVE-R simulations indicated that RF regression outperformed other techniques, including K-Nearest Neighbors, Support Vector Machines, Neural Networks and Multivariate Linear Regression.

In order to apply random forest regression, the inputs and outputs of the SVE-R model need to be transformed into a form on which the algorithm can operate (see Figure 4.1). The inputs to the SVE-R model are two-dimensional elevation and vegetation maps, soil and roughness fields that are one-to-one correlated with vegetation cover, and lateral boundary conditions, including the storm features and domain boundaries. The SVE-R model outputs are two-dimensional overland flow and infiltration fields, which vary in space and time and which embed considerable spatio-temporal correlation based on the continuity of the flow field. Random forest regression, by contrast, is not a spatial operator, so the 2D spatial structures of the SVE-R model domain must be adapted into a set of points on which the regressor can be applied. Each point can be characterized by a vector of features (analogous to independent variables), on which the random forest operates to make predictions of the target variable for that point.

Applying random forest regression to emulating the SVE-R model thus requires transforming the SVE-R simulation domain into feature and target variables. The key challenge in

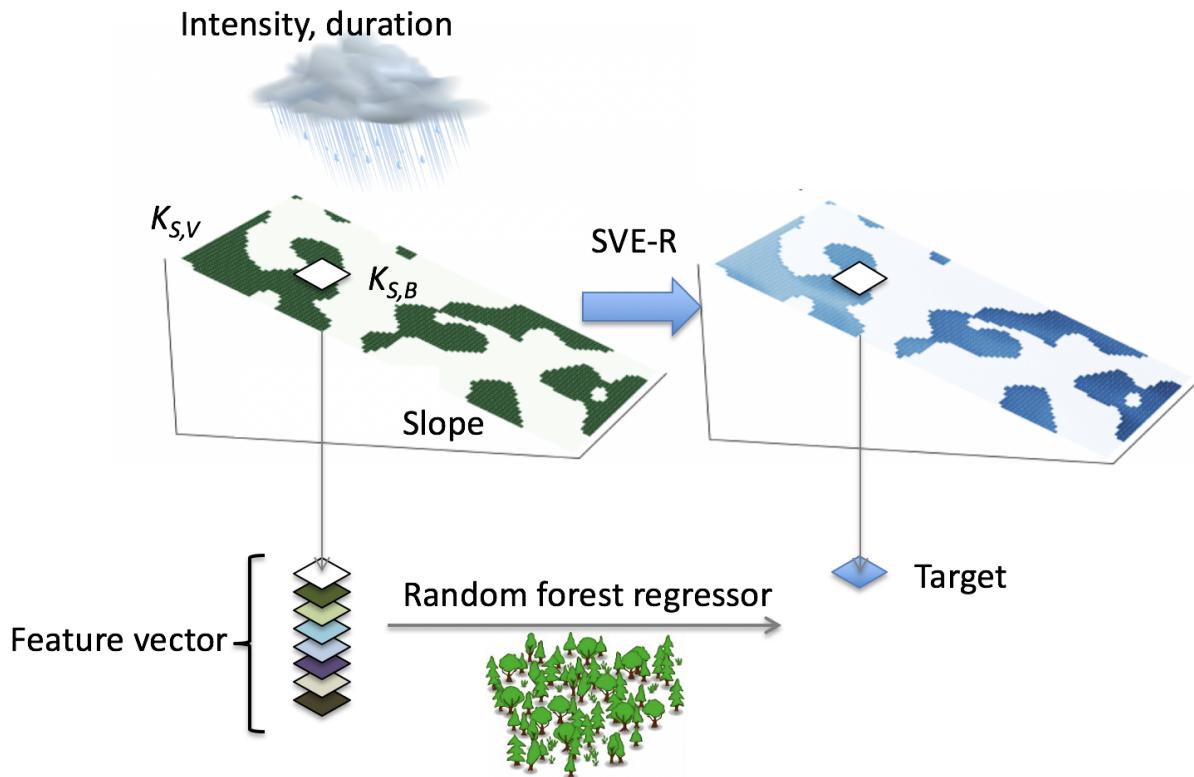


Figure 4.1: The hillslope on the left represents the SVE-R model domain, including vegetation patterns, storm characteristics, soil parameters and hillslope gradient. The hillslope on the right shows an SVE-R model prediction of infiltration depth, representing the target variable. Each point in the SVE-R model domain is transformed into a 1D vector of features that represent the surrounding neighborhood. The random forest regressor independently operates on each point to predict the outcome.

specifying the RF features is to provide enough information about the drivers of the spatial correlation of the SVE-R output through time and space that the re-assembled point-by-point predictions have fidelity to the spatial correlation in the SVE-R output. To do this, the 2D SVE-R domain is transformed into a set of 1D neighborhood descriptors for each grid cell, which enables the 2D structure of the results to be reproduced. This transformation is an example of “feature engineering”, which is broadly defined as the process of using domain knowledge to transform raw data into the best representation for the ML algorithm to learn a solution [Brownlee, 2014]. Feature engineering is a crucial step in many ML applications, often involving the transformation of complex, multifaceted data into features that reflect the underlying processes [Kubat et al., 1998]. Rather than emulate the complete time-varying SVE-R model output, the SVE-R predictions are aggregated, or transformed into peak values. Specifically, the cumulative infiltration patterns (I) and peak overland flow velocities

(U_{max}) are selected as target variables, because of their relevance to plant water availability and erosion risk, respectively.

This chapter describes the feature engineering process, the training of the random forest regressors and the performance of the regressors in emulating the SVE-R results. In particular, I ask: “how does RF regression perform in predicting cumulative infiltration fields and peak flow velocities as storm characteristics are varied for a given landscape?” Throughout the chapter, the examples that are given relate to using RF regression to emulate the cumulative infiltration fields *I*. The same approach is also used to predict the maximum overland flow velocity, as detailed in Appendix B.

4.2 Methods

A random forest regressor is trained on a collection of SVE-R model runs. This collection of model runs spans a multidimensional parameter space, including landscape parameters (hill-slope gradient, soil infiltration capacity, and surface roughness), storm parameters (duration and intensity) and vegetation characteristics. The vegetation characteristics include both hillslope-scale properties of the vegetation distribution, such as the mean fractional cover and characteristic length scales of the vegetation patches, as well as the smaller scale details of how this vegetation is specifically arranged in space. The RF regressor enables interpolation between a subset of this parameter space, while other parameters are held constant. I elected to hold storm and landscape parameters constant while varying vegetation patterns. Given the nonlinear dynamics of the SVE-R model, it is unclear what (if any) interpolation would be appropriate between landscape or storm features. Thus, RF regressors are trained to interpolate between different realistic patterns of vegetation, for fixed storm and landscape properties. As detailed in Chapters 5 and 6, the selection of fixed storms and landscapes is application-specific. The remaining steps, namely selecting different spatial vegetation patterns on which to run the SVE-R model, running the SVE-R model, engineering features to represent the vegetation patterns, and training and testing the RF regressors, are detailed in the following sections.

Constructing the domain of SVE-R simulations

Given the decision to hold storm and landscape parameters constant, an assessment of the quality of the RF regression as a function of storm characteristics requires the generation of multiple random forests, each of which predicts a hydrological outcome (infiltration or peak velocity) for an input vegetation pattern. To illustrate, this chapter presents the RF emulation approach using a collection of SVE-R simulations spanning a range of 16 unique storms for a single landscape (a 200×100 m planar hillslope with 2% slope, with saturated hydraulic conductivity under vegetation set to $K_{S,V} = 1.5$ cm/hr and for bare sites set to $K_{S,B} = 0.15$ cm/hr). All parameter values are summarized in Table 4.1, and are discussed in more detail in Chapter 5.

Table 4.1: Parameters in the RF training space. Where multiple parameters are listed, the cases were run factorially to explore all parameter combinations.

Parameter	Description	Units	Values
d	Storm depth	cm	0.4, 0.8, 1.2, 1.6
t_{storm}	Storm duration	min	10, 20, 30, 40
ϕ_V	Fractional vegetation cover	—	0.2, 0.4, 0.6, 0.8
σ	Patch length scale	—	0, 1, 2, 3, 4
$K_{S,V}$	Hydraulic conductivity (vegetated sites)	cm/hr	1.5
$K_{S,B}$	Hydraulic conductivity (bare soil)	cm/hr	0.15
S_o	Hillslope gradient	%	2.0
n_B	Manning's n (bare soil)	s m ^{-1/3}	0.03
n_V	Manning's n (vegetated sites)	s m ^{-1/3}	0.1
L_x, L_y	Domain size	m	200 × 100
dx	Grid resolution	m	2.0

With 5,000 cells in the simulation domain, there is a countable, but very large possibility space within which vegetation patterns could be constructed. I therefore specified two macroscopic pattern variables: fractional vegetation cover ϕ_V , and the characteristic patch length scale σ to represent typologies of patterns.

Vegetation patterns were specified as binary arrays of 0's and 1's to represent bare and vegetated sites, respectively, and were constructed in several steps (see Figure 4.2). First, random numbers, uniformly distributed between 0 and 1, were generated in each grid cell. To obtain a binary array with the desired fractional vegetation cover, values greater than a threshold value (ϕ_V) were set to 0 and all other values were set to 1. To adjust the patch length scale, a Gaussian filter was applied to the binary pattern using the Python library SciPy, where the standard deviation of the Gaussian kernel, σ , was used to adjust the length scale of smoothing [Jones et al., 2001]. The Gaussian-filtered patterns were then re-binarized with the same threshold approach, where the cutoff was adjusted to maintain the target fractional vegetation cover (for example, to achieve $\phi_V = 0.3$, the cutoff would be the 30th percentile of the smoothed array). The pattern training space included 4 fractional vegetation covers and 5 patch length scales, for a total of 20 unique vegetation patterns. In order to maximize the diversity of spatial patterns and to ensure reproducibility, a unique random seed (used to initialize the Python pseudo-random number generator) was set for each ϕ_V - σ combination. A subset of the training space of vegetation patterns is displayed in Figure 4.3, for a range of densities and patch length scales. Figure 4.4 shows how σ and ϕ_V together influence the mean along-slope length scales of the vegetated patches (panel A) and bare soil areas (panel B).

For every combination of storm depth and duration listed in Table 4.1, each of the 20 vegetation patterns was input to the coupled SVE-R model described in Chapter 2. The SVE-R model predictions were summarized into spatial fields of cumulative infiltration depth and maximum overland flow velocity. Figure 4.5 illustrates this step with maps of cumulative

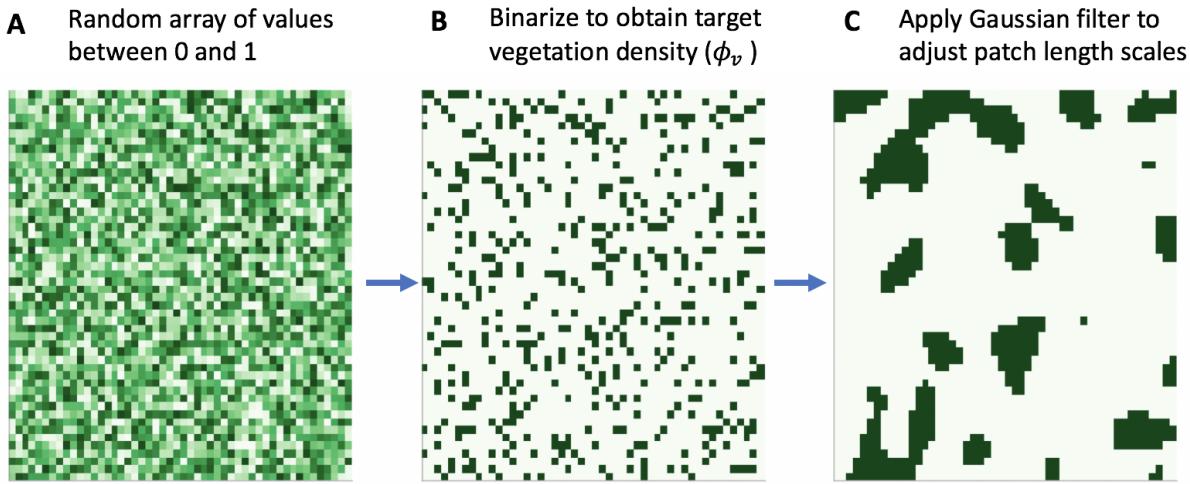


Figure 4.2: Patterns were randomly generated from arrays of random numbers between 0 and 1 (A), which were binarized to obtain the non-smoothed ($\sigma = 0$) vegetation patterns (B). The target densities, ϕ_V , were achieved by adjusting the threshold value used to binarize the arrays. The patch length scales were adjusted by applying a Gaussian filter to the binary patterns and re-binarizing to preserve the desired fractional vegetation cover (C). The standard deviation of the Gaussian kernel, σ , determines the characteristic patch length scale.

infiltration, corresponding to the vegetation patterns in Figure 4.3.

Features Engineering and Extraction

Random forest features were generated from the SVE-R model domain by decomposing each point in the 2D vegetation map into a set of 1D descriptions of its local neighborhood (see Figure 4.6). The 2D vegetation maps were thus transformed into matrices with N columns (representing the engineered features) and M rows (representing the number of grid points in the hillslope domain). These feature matrices were vertically stacked, creating a single feature matrix, \mathbf{X} , representing all of the SVE-R model simulations in a suitable form for RF regression. The infiltration (and U_{max}) fields were similarly transformed into $1 \times M$ target vectors, which were vertically stacked into a single target vector, \mathbf{Y} .

To account for the different soil and roughness parameters of bare and vegetated cell types, separate random forests were fit for each cell type. The random forests were separately trained to make predictions for bare and vegetated cells, which were then recombined into a map of the target variable (I or U_{max}).

The feature engineering process is described here primarily for the vegetated cells, and the same general approach was used to define the bare soil features. First, the SVE-R predictions were visually inspected to identify spatial trends and patterns related to topographic position

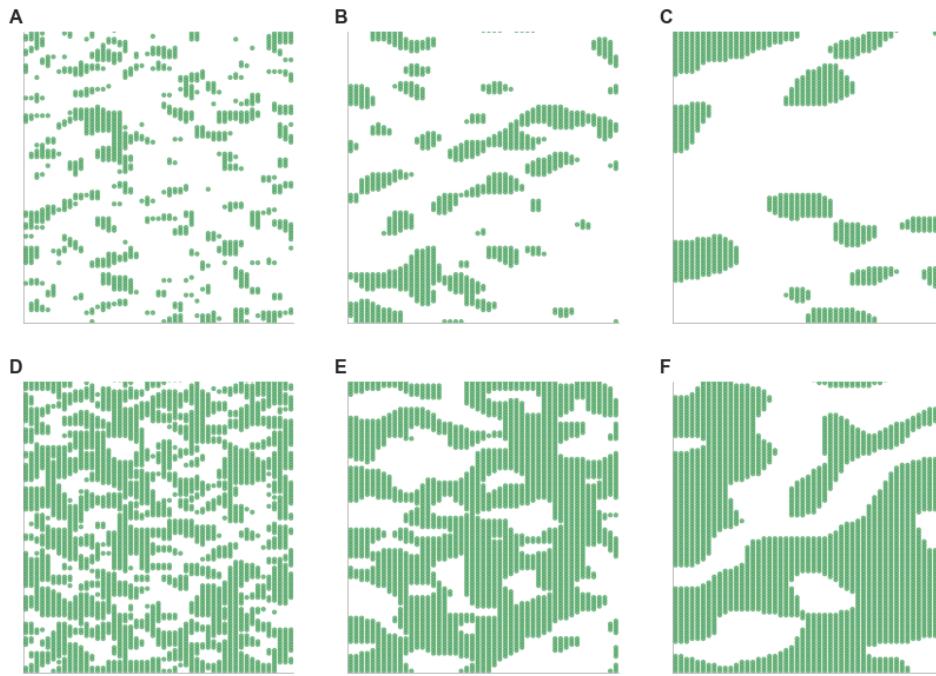


Figure 4.3: Illustration of six randomly generated vegetation maps in the random forest training space. Parameter are: $\phi_V=0.2$ (top row), $\phi_V=0.6$ (bottom row), and $\sigma = 0$ (left column), $\sigma = 2$ (center column), and $\sigma = 4$ (right column).

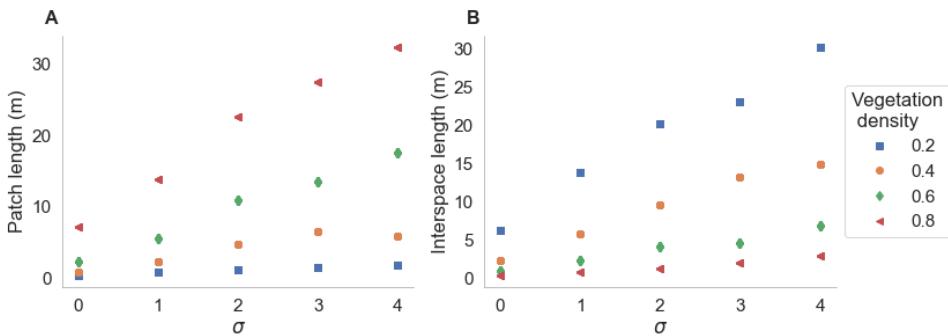


Figure 4.4: Mean along-slope patch length (A) and bare soil length (B) for a range of smoothing length scales σ , specifying the standard deviation for the Gaussian kernel.

and vegetation cover. For example, the infiltration patterns in Figure 4.5 suggest that the distance to the upslope divide, length of the nearest upslope bare soil patch, and distances to the nearest bare soil cells in the upslope, downslope and lateral directions are relevant features for predicting cumulative infiltration. Based on these observations (and similar observations for the U_{max} fields), potentially useful features were computed for each cell, as

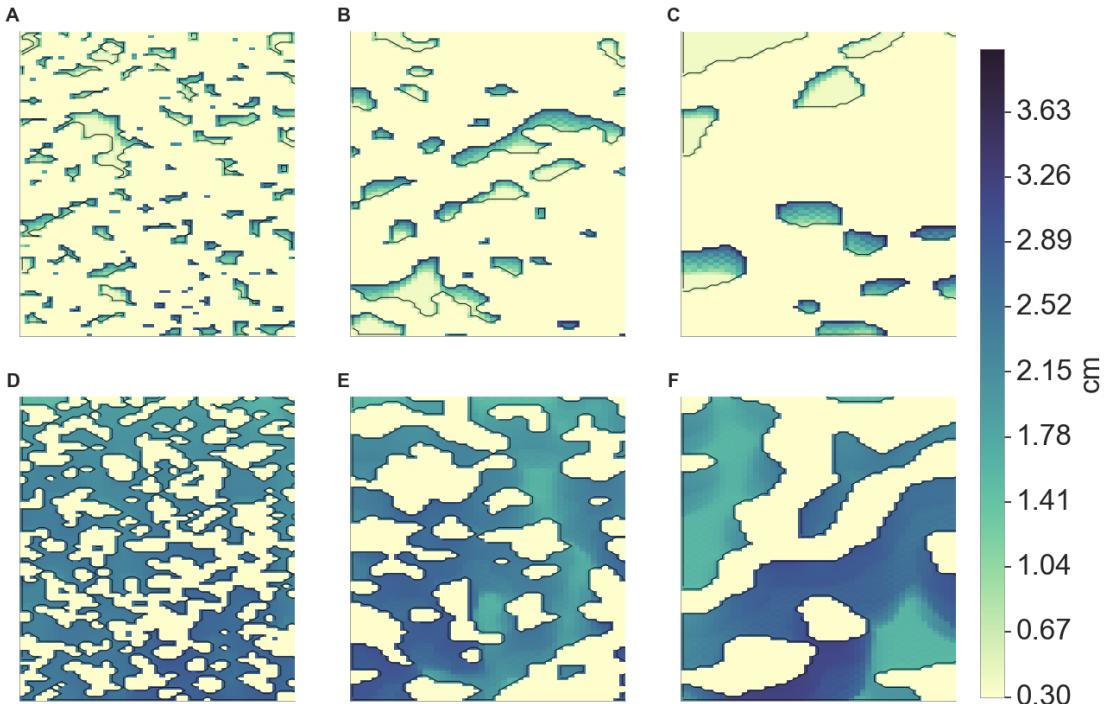


Figure 4.5: Infiltration maps for a 20 minute storm with $p = 4.8 \text{ cm/hr}$, $S_o = 2\%$, $K_{S,V} = 1.5 \text{ cm/hr}$ and $K_{S,B} = 0.15 \text{ cm/hr}$, using the vegetation maps from Figure 4.3.

illustrated in Figure 4.7.

While flow in the SVE-R simulations is primarily aligned along the hillslope gradient, there is a lateral flow as well, meaning that infiltration at a point depends on the conditions directly upslope, and also on conditions upslope to the left and right. To account for this, the engineered features were smoothed using SciPy’s Gaussian filter, while ignoring cells of the other type (i.e. smoothing of features based on the vegetation field ignored any bare soil cells, and vice versa). The smoothing length scale was adjusted by varying ζ , the standard deviation of the Gaussian kernel. Figure 4.8 shows the effect of the Gaussian filter for an example feature, distance to the nearest upslope bare soil cell, with $\zeta = 1$ (panel B) and $\zeta = 2$ (panel C). Both the original and Gaussian-filtered features were used to train the RF regressors.

Table 4.2 summarizes the engineered features used for predicting both I and U_{max} . To select the ‘feature parameters’ ζ and λ , I trained multiple versions of the random forest regressors on subsets of SVE-R simulations that spanned a wide range of landscape and storm properties. I evaluated the time required to train the RF regressors, their memory usage and predictive accuracy, in order to identify a suitable set of feature parameters. While this did not encompass a formal optimization of the feature selection, it did ensure that the selected feature parameters - namely $\zeta = 2$ and multiple λ s, ranging from 4 to 30 grid cells -

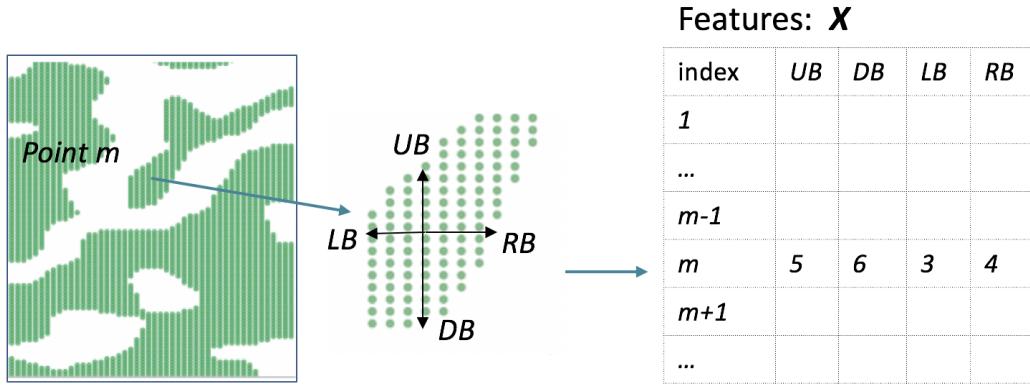


Figure 4.6: Illustrating how the 2D SVE-R domain is deconstructed into grid-cell representations of each point, for several select features. Descriptions of these features are provided in Table 4.2.

achieved both good predictive accuracy while keeping memory usage and time requirements within acceptable limits. This selection process is described in more detail in Appendix C.

The use of both original and smoothed features, as well as multiple window-lengths used to describe the upslope vegetation fraction, result in a high degree of multicollinearity within the feature space. In general, such multicollinearity does not compromise the performance of random forest models, although they may increase the risk of over-fitting the model [Shih, 2012], and challenge the interpretation of feature importance weights. Because the goal of the study is to imitate the SVE-R model, rather than to interpret the output of the RF regression to gain physical insight into the system being studied, assessing the feature importance is not a priority. Overfitting was assessed by testing the RF performance on the test set for a range of ζ s and λ s, as discussed in the Appendix B, and no signs of overfitting were observed.

The random forest regressors used a common set of features across all storm and landscape parameters, but the importance of the specific features varied with the storm and landscape simulated. For example, Figure 4.9 illustrates how the influence of the vegetation distribution on I varies with the storm intensity. Runoff-runon patterns (i.e. within patch variations in I) are more apparent for a low intensity storm (panel A), whereas upslope-downslope differences are more apparent for a higher intensity storm (panel B). In panel A, the infiltration capacity of the vegetated patches exceeds the rainfall rate ($K_{S,V} > p$), and runoff-runon mechanisms are more important than the distance to the ridge. In panel B, the rainfall intensity exceeds the soil infiltration capacity, and the entire hillslope becomes a runoff source during the storm. Consequently, the spatial pattern of infiltration is determined by what happens after the storm ends (i.e. where the ponded water infiltrates or runs off). Because the depth of ponded water and potential source area both increase with distance downslope, the water infiltrated by vegetation also increases. Similar comparisons demonstrate the influence of hillslope gradient and soil infiltration capacity on the spatial patterns of infiltration. These

Table 4.2: Features used to fit the RF regressors. Gaussian smoothing was applied to all of the features, with the exception of the hillslope-mean fractional vegetation cover, and both smoothed and original versions were included as features.

Feature	Description
<i>Vegetation features</i>	
L_V	Vegetation patch length (along-slope)
L_{UB}	Length of the nearest upslope bare soil patch
UB	Distance to the nearest upslope bare soil cell
DB	Distance to the nearest downslope bare soil cell
YB	Across-slope (lateral) distance to the nearest bare soil cell
<i>Bare soil features</i>	
L_B	Bare soil patch length (along-slope)
L_{UV}	Length of the nearest upslope vegetated patch
UV	Distance to the nearest upslope vegetated cell
DV	Distance to the nearest downslope vegetated cell
YV	Across-slope (lateral) distance to the nearest vegetated cell
<i>Common features</i>	
L_{div}	Distance to the divide
ϕ_V	Hillslope-mean fractional vegetation cover
$\phi_{V,\lambda}$	Upslope vegetation computed with window length λ
<i>Feature parameters</i>	
λ	Window lengths in the computation of the upslope vegetation fraction ($\lambda = 4, 12, 20, 28$)
ζ	Standard deviation of the Gaussian kernel used in feature smoothing ($\zeta = 2$)

observations of a strongly nonlinear relationship between vegetation, storm properties and hydrological outcome support the initial choice to train separate random forests for each combination of storm and landscape parameters.

Training and Testing the RF predictors

Figure 4.10 illustrates how the RF regressors were trained and tested. To avoid over-fitting, the simulations were divided into train and test sets, $\mathbf{X}_{train}, Y_{train}$ and $\mathbf{X}_{test}, Y_{test}$, at the hillslope level. Eighty percent of the vegetation patterns were used for training the RF regressors, and the remaining twenty percent were reserved for validation. The test set consists of the simulations with $\sigma = 2$, corresponding to an intermediate patch length scale, for the purpose of testing the ability of the RF regressor to interpolate between small and large patch sizes. The hillslope-level division was necessary in order to test the ability of the RF regressors to extrapolate to new patterns. Due to the high spatial correlation in the infiltration patterns, dividing the data into train and test sets by randomly selecting grid

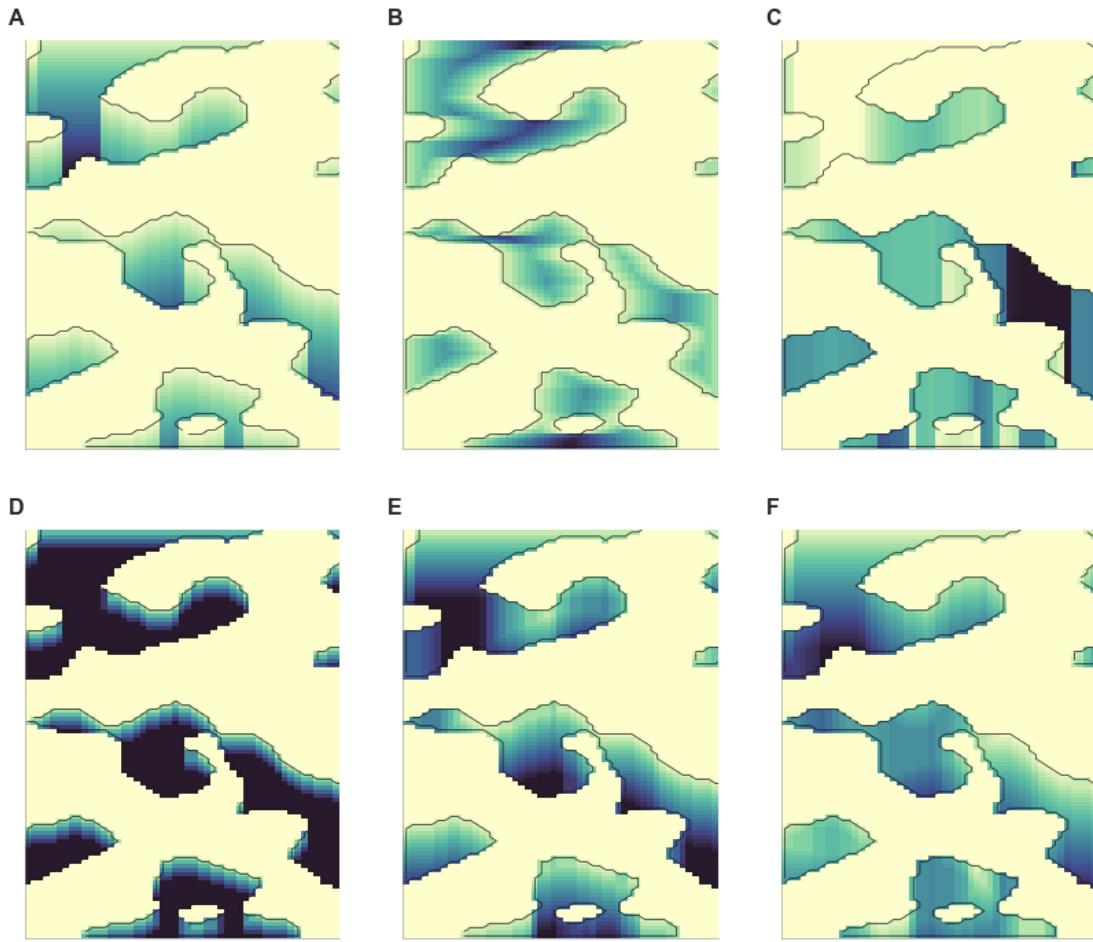


Figure 4.7: Select examples of random forest features: (A) distance to the nearest upslope bare cell, (B) across-slope distance to the nearest bare soil cell, (C) length of the adjacent upslope bare soil patch. Panels D-F show the upslope vegetation fraction computed for several window lengths (12, 20 and 28, respectively). For clarity, values in bare cells are set to 0 in panels D-F.

cells from all of the vegetation patterns resulted in misleadingly high scores on the test set.

In order to train separate RF regressors for bare and vegetated cells, the inputs were further divided into vegetated and bare soil subsets: $\mathbf{X}_{V,train}$, $Y_{V,train}$ for vegetated cells and $\mathbf{X}_{B,train}$, $Y_{B,train}$ for bare soil cells, and similarly for the test set.

The random forest regressors (Φ_V and Φ_B) were then trained using the Python machine learning library scikit-learn [Pedregosa et al., 2011]. The scikit-learn random forest regressor class takes as input the features and target (e.g. $\mathbf{X}_{V,train}$, $Y_{V,train}$), and returns an estimator object (e.g. Φ_V), which can be applied to make predictions on new data. The random forest regressors were trained by maximizing the r^2 coefficient of determination for the pre-

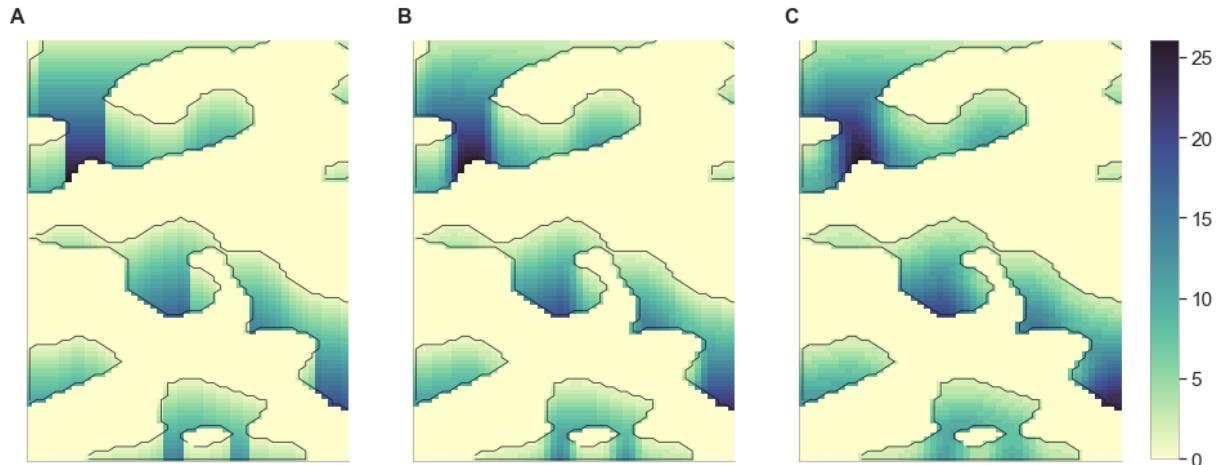


Figure 4.8: This figure illustrates the use of the Gaussian filter to smooth the feature maps: (A) the distance to the nearest upslope bare cell, UB , smoothed with a Gaussian filter with (B) $\zeta = 1$ and (C) $\zeta = 2$.

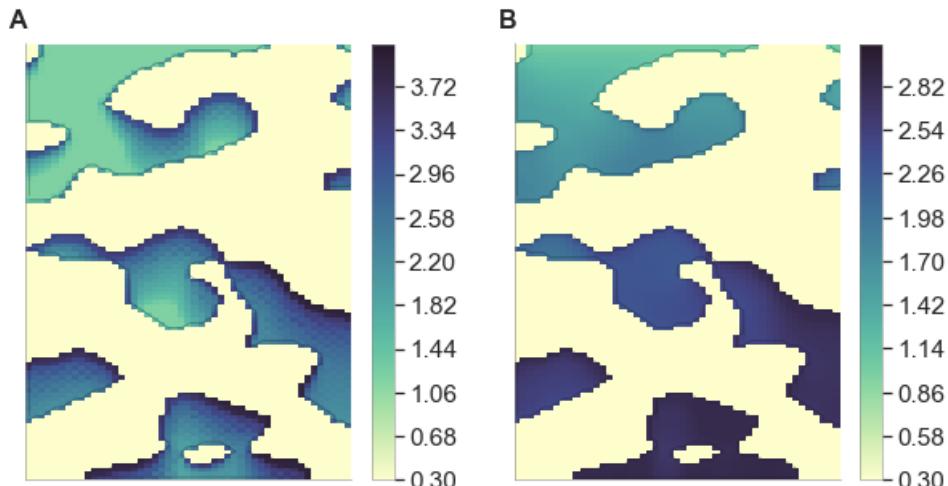


Figure 4.9: Cumulative infiltration maps for a 30 minute storm with (A) $p = 2.5 \text{ cm/hr}$ (B) $p = 7.5 \text{ cm/hr}$. Other parameters are the same between panels: $S_o=2\%$, $K_{S,V} = 4 \text{ cm/hr}$ and $K_{S,B} = 0.4 \text{ cm/hr}$. Darker areas correspond to vegetated patches.

dictions. Because all of the training patterns were combined in \mathbf{X}_{train} , the training process minimized the error of the predictions at the cell-level. Similarly, the predictive accuracy of the trained random forests was assessed at the cell level by comparing Y_{test} to the random forest predictions on \mathbf{X}_{test} .

However, to obtain a hillslope-level assessment, the RF performance was also evaluated

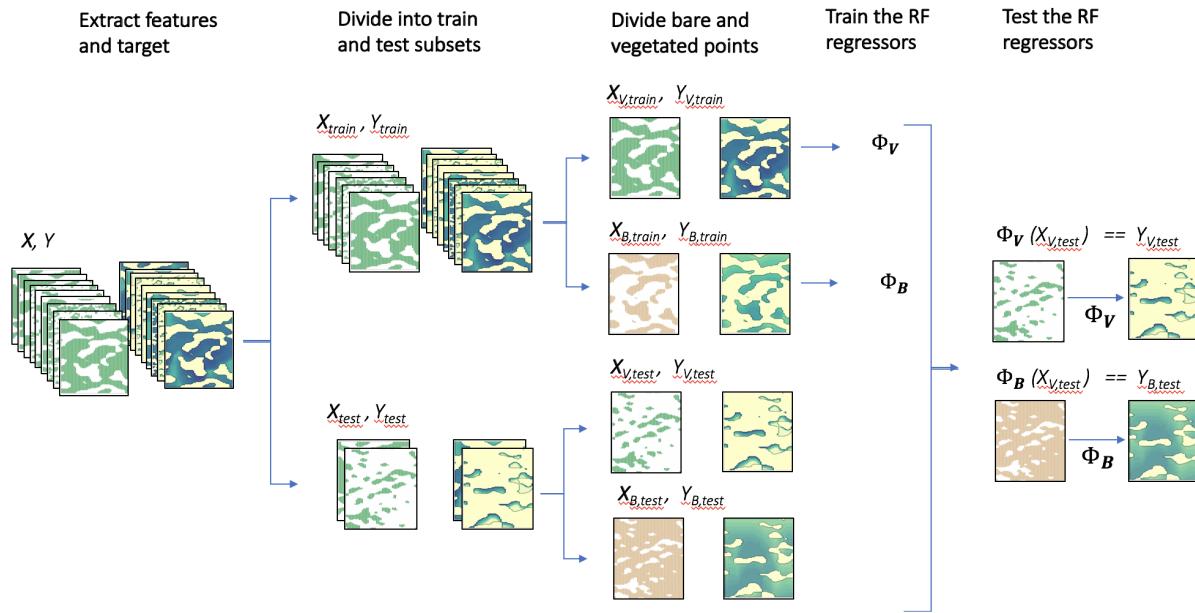


Figure 4.10: Schematic illustrating how RF regressors are trained and tested. The data are divided into train and test sets at the hillslope level, as opposed to randomly selecting cells from all the patterns. Spatial maps of vegetation and I are for illustrative purposes only, and all \mathbf{X} and \mathbf{Y} (including with subscripts) represent matrices and column vectors, respectively.

at the scale of the hillslope. Visual assessment required combining the RF predictions for vegetated and bare soil cells, and comparing to the SVE-R simulated infiltration field. The hillslope-scale evaluation metric is the normalized Root Mean Square Error (NRMSE), defined as the standard deviation of the prediction errors scaled by the mean infiltration depth. The NRMSE is reported separately for bare and vegetated cells, reflecting the separate RF regressors used to make predictions. The NRMSE is used instead of the r^2 scores because it is easier to interpret and less sensitive to outliers (particularly for the very small infiltration depths in bare soil areas). A second assessment metric is the infiltration fraction, which provides a measure of the RF success in predicting hillslope-scale outcomes. The NRMSE is reported for hillslopes in both the train and test sets, with the expectation that the NRMSE will be larger in the test set.

The random forest algorithm has several parameters (sometimes referred to as hyperparameters) that have to be set by the user before the learning process begins, such as the number of regression trees estimated to generate the random forest (`n_estimators`), the maximum depth of the trees (`max_depth`) and the minimum number of samples required to allow an internal node to be split (`min_samples_split`) [Pedregosa et al., 2011]. The sensitivity of the RF performance and characteristics to most of these parameters was low, and the default values were used. However, to reduce memory consumption, I altered the `max_depth`

and `min_sample_split` by evaluating their impact on model performance, training time and memory usage, selecting a `max_depth` of 15 and `min_sample_split` of 20, which were used for all storm and landscape types (see Appendix C for details).

Note that although the SVE-R simulations were divided into train and test sets for feature engineering, parameter selection (i.e. to fix values of ζ , λ , `max_depth` and `min-sample split`), and assessment of the overall RF performance, the final RF regressors used in Chapters 5 and 6 for specific applications were trained on all SVE-R simulations generated for those applications, using the features and parameters selected during the training phase.

4.3 Results and Discussion

Visual inspection of the results shows good agreement between the SVE-R simulations and the RF predictions. For example, Figure 4.11 presents a sample infiltration map from the test set (panel A) and its corresponding RF prediction (panel B). Figure 4.12 shows a box plot of the NRMSE for all of the simulations, grouped by σ and cell type. The mean NRMSE for the vegetated cells is approximately 5% for the train set and 10% for the test set ($\sigma = 2$), and slightly less for the bare soil cells. These performance results are comparable with other studies using ML for model emulation; for example, Verrelst et al. [2017] apply RF regression to emulate a radiative transfer model, and report NRMSE values ranging from approximately 5 to 10%. Performance is often lower when ML are used to interpret data, likely due to the greater uncertainty involved (see e.g. Cortez and Morais [2007], Blackard and Dean [1999]).

Figure 4.11 displays the NRMSE sensitivity to rain depth (panel A) and storm duration (panel B). The NRMSE decreases with increasing rain depth and increases slightly with increasing duration. These results indicate that the RF is less successful in predicting runoff-runon patterns associated with low intensity storms, and more successful for higher intensity rain where the distance to the divide is the dominant control on I . To illustrate, Figure 4.14 compares infiltration patterns for two cases, one with very low NRMSE (panel A) and the other with higher NRMSE (panel B). Lower intensity storms produce infiltration maps with greater dependence on the vegetation distribution, which makes prediction of the infiltration patterns in these cases a harder problem. While the current model performance is still good in these cases, there remains scope for further improvement by feature refinement.

The random forests are tasked with the dual challenge of predicting both within-slope water redistribution and hillslope scale outcomes. To assess the latter, Figure 4.15A compares the SVE-R model and the RF regressor predictions of infiltration fraction, showing good agreement ($r^2 > 0.99$ for train and test sets). Figure 4.15A suggests that the RF emulation approach could perform well as a component in a larger (e.g. catchment scale) model, for example, to predict the hillslope scale runoff following storms.

The main advantage of the RF emulation approach is computational speed (see Figure 4.15B). A trained RF regressor takes less than a second to make a prediction, making it several orders of magnitude faster than the SVE-R model. This has the potential to make

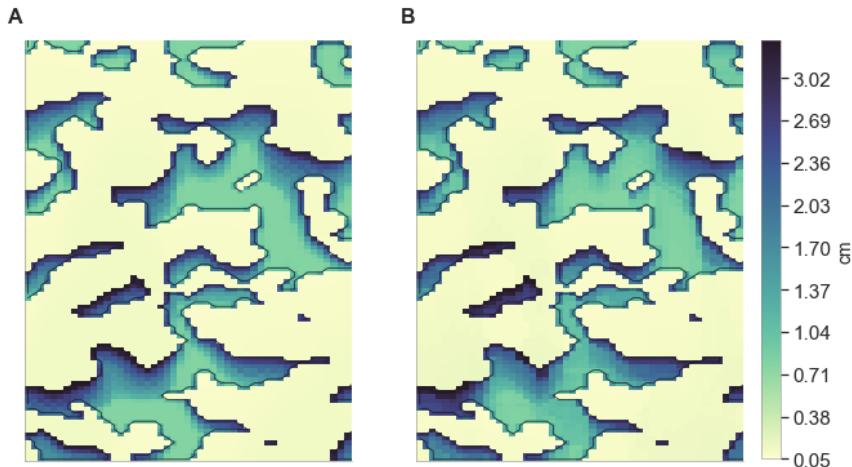


Figure 4.11: Example random forest prediction of a simulation from the test set ($\phi_V = 0.4$, $\sigma = 2$, $t_r = 10$ minutes and $p = 2.4$ cm/hr). Panel A shows the SVE model output, and panel B shows the corresponding RF prediction.

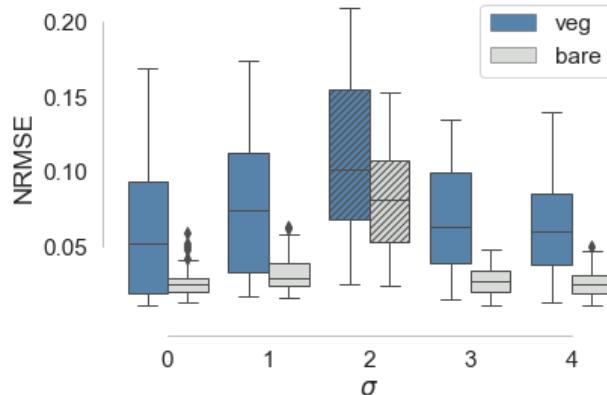


Figure 4.12: A box plot summarizing the NRMSE, grouped by patch length scale (σ). Hatches indicate the test-set simulations ($\sigma = 2$). As expected, the error is generally higher for the test set.

computationally intensive problems more manageable, for example, the investigation of larger domain sizes or the rapid assessment of many different patterns.

Of course, RF regression also has limitations, namely, the random forests are not good at interpolation between values. While it is possible to incorporate global features such as rainfall duration, slope angle and soil hydraulic conductivity directly into the random forest model, such an approach raises difficult questions about interpolation. For example, a random forest trained on 1.6 cm and 0.4 cm storms should not be used to interpolate to a

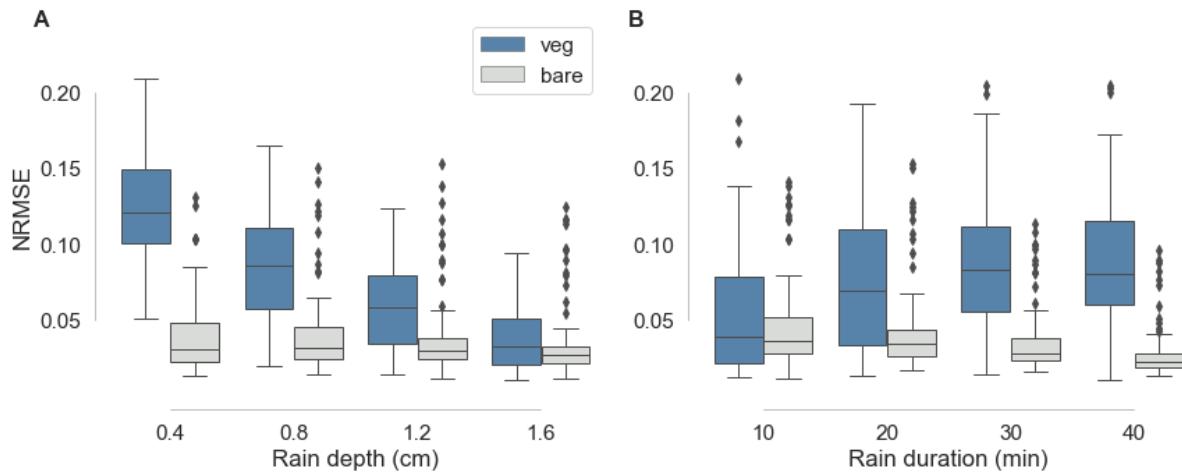


Figure 4.13: The NRMSE decreases with increasing rain depth (A), and increases slightly with increasing rain duration (B).

1.2 cm storm, due to the decision tree nature of its construction. Given the nonlinear nature of the SVE-R model, it is uncertain what, if any, type of interpolation between features is sensible. For example, linearly interpolating between the high intensity, 1.6 cm storm in Figure 4.14 panel A and the low intensity, 0.4 cm storm in panel B is unlikely to yield accurate results, due to the different controls on runoff and infiltration processes in these cases. A more general approach (i.e. one that can extrapolate between hillslope gradients or storm durations) would require dealing with the nonlinear nature of the overland flow-infiltration dynamics.

4.4 Conclusions

Random forest regression was applied to emulate the overland flow model predictions of cumulative infiltration, which is important for plant water availability. The same approach was successfully applied to predict maximum flow velocity (see Appendix B). Despite the complexity of the SVE-R simulations, emulation of the solutions with machine learning is promising. Random forest regression reduces the computational time by several orders of magnitude, from multi-hour simulations for the SVE-R model to seconds. This can enable the integration of quantitative predictions about storm-driven overland flow into ecological models or assist in landscape restoration planning. The following chapters explore such applications.

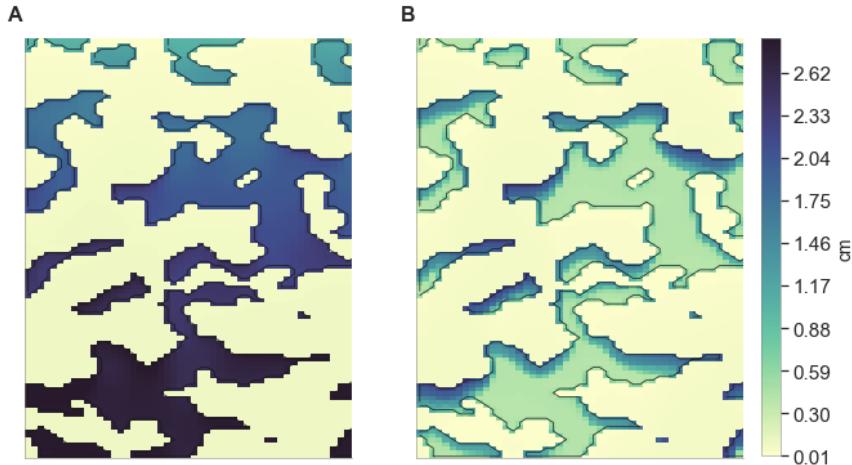


Figure 4.14: (A) A high intensity storm ($t = 10$ min, $d = 1.6$ cm) where the RF performance is very good, with NRMSE(veg)= 0.03 and NRMSE(bare)= 0.05. (B) For a low intensity storm ($t = 10$ min, $d = 0.4$ cm) with the same vegetation pattern, the RF performance is comparatively poor, with NRMSE(veg)= 0.21 and NRMSE(bare)= 0.13. This comparison illustrates how the RF performance is best in high intensity storms where the distance to the divide is the dominant predictor, and worse in low intensity cases where runoff-runon patterns dominate.

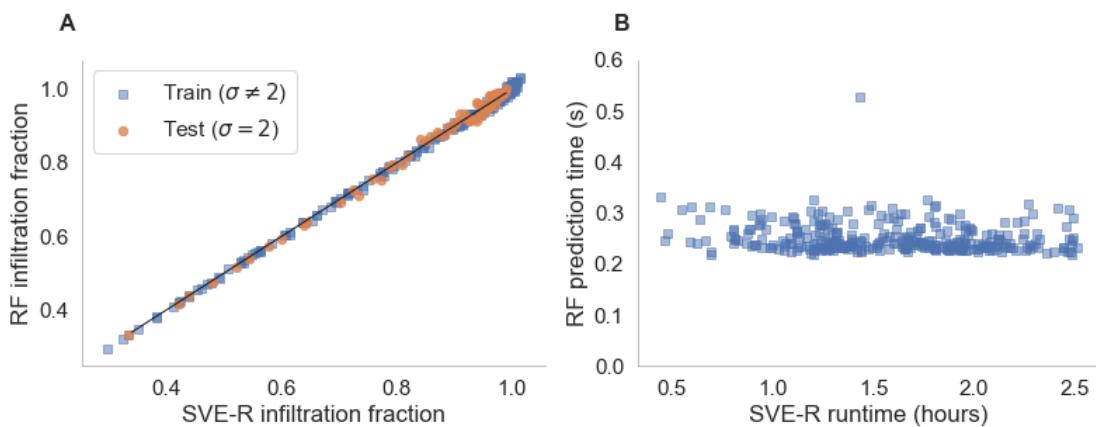


Figure 4.15: (A) SVE-Richards versus RF predictions of the hillslope-mean infiltration fractions. The test set is plotted in orange. (B) SVE-R model simulation times versus trained RF prediction times. The RF is several orders of magnitude faster.

Chapter 5

Storm-scale hydrology in pattern forming models

5.1 Introduction

In many dryland regions worldwide, vegetation patches form regular spatial patterns [Bromley et al., 1997] characterized by a more-or-less regular wavelength, and, on sloping terrain, a directionality, typically oriented along the slope contours [Penny et al., 2013]. The self-organized patterning results from interactions between plants that have opposing effects on different length scales. Plants growing close together tend to enhance each other's growth (a cooperative or facilitative interaction), while plants growing further apart tend to inhibit each other's growth (a competitive or inhibiting interaction) [Borgogno et al., 2009]. There are numerous mechanisms by which these competitive and facilitative interactions can arise [Meron, 2015], but in many cases facilitation is associated with increased infiltration capacities and thus soil water resources beneath plant canopies [Seghieri et al., 1997], while inhibition arises from below-ground root competition for water. Outside vegetation canopies, physical and biological soil crusts and seals form on many desert soils, preventing the infiltration of incoming rain and generating runoff [Belnap, 1990, Assouline, 2004]. This water then ‘runs on’ to vegetated areas where it infiltrates. This ‘patch-mosaic’ functionality of patterned landscapes - and dryland landscapes more broadly - is integral to sustaining ecosystems in areas where rainfall alone would not provide sufficient water for vegetation to grow [Tongway and Ludwig, 2001].

Patch-mosaic functionality, however, also results in a vulnerability of patterned and patchily-vegetated landscapes to degradation, since once the vegetation cover is disrupted, it may be impossible to naturally re-establish the runoff-runon mechanism that would allow new plants to persist. Mathematical models of patterned vegetation predict that gradual declines in resource availability - e.g. annual rainfall - can cause abrupt collapse of vegetation cover. This collapse, however, occurs after the landscape moves through a predictable trajectory in pattern morphologies [van de Koppel et al., 2002, Rietkerk et al., 2004] (see

Figure 5.1). This trajectory raises the appealing possibility that patterns could provide an early warning of land degradation [Kéfi et al., 2007, Rietkerk and Van de Koppel, 2008]. For this to be feasible, however, the sensitivity of pattern morphology to changes in climate and land use need to be understood, to enable clear interpretation of observations.

Unfortunately, replicating field observations of pattern morphology and its dependence on environmental conditions is difficult using most existing models of patterned systems, because the representation of biological and physical processes in these models is idealized [Rietkerk et al., 2002]. Most models of patterned systems represent biomass and hydrological dynamics using phenomenological partial differential equations [Borgogno et al., 2009]. For example, these models usually represent the discrete and stochastic process of rainfall with a continuous time-averaged value, and the runoff generated by this rainfall is described via diffusion or simple advection with a fixed wave celerity [Rietkerk et al., 2002, HilleRis-Lambers et al., 2001]. Many of these models apply periodic boundary conditions, rather than the landscape boundary conditions of no-flow at an upslope divide and water loss in a downslope channel. Such simplifications distort the water balance within phenomenological pattern forming models, complicating the relationship between the modeled and real systems. Recent high resolution modeling of within-storm processes and plant water uptake has demonstrated that storm properties and runoff processes have non-trivial impacts on modeled water availability within vegetation band locations [Paschalis et al., 2016]. Consequently, the use of phenomenological representations of storm and runoff processes in pattern forming models is likely to influence the predictions made about pattern morphology, its distribution along hillslope catenas and its relationship to storm climatology.

The simplifications do, however, also introduce a range of advantages. Notably, in many cases, they permit a range of valuable mathematical analyses of the pattern forming model [HilleRisLambers et al., 2001, Sherratt, 2005, Siteur et al., 2016]. The simplifications also improve the computational tractability of the models, principally by avoiding the computationally demanding task of simulating overland flow [Konings et al., 2011, Guttal and Jayaprakash, 2007, e.g.]. Other models simplify the system by solving directly for a steady state vegetation spatial pattern [Foti and Ramírez, 2013]. Such an approach, however, imposes a strong assumption of steady state conditions, which negates observations of migrating vegetation bands in many ecosystems [Valentin et al., 1999, Tongway and Ludwig, 1990] and may limit analysis of transient behaviors associated with desertification.

Progressing towards a fully physical description of vegetation pattern morphologies and their environmental determinants thus remains a challenging problem, with major uncertainties regarding the representation of vegetation growth [Paschalis et al., 2016], reproductive biology and dispersion [Thompson et al., 2014, Thompson and Katul, 2009, Thompson et al., 2008], drought resistance versus drought mortality pathways [Guttal and Jayaprakash, 2007, McDowell et al., 2008], and biogeochemistry and geomorphic processes in these ecosystems [McGrath et al., 2011, 2012, Saco et al., 2007]. Empirically, there is also a general lack of clarity regarding the sensitivity of pattern morphologies and their stability to changes in any of these processes. Injecting better process fidelity in any of these arenas into existing models, however, is likely to exacerbate the challenge of model tractability by invoking additional

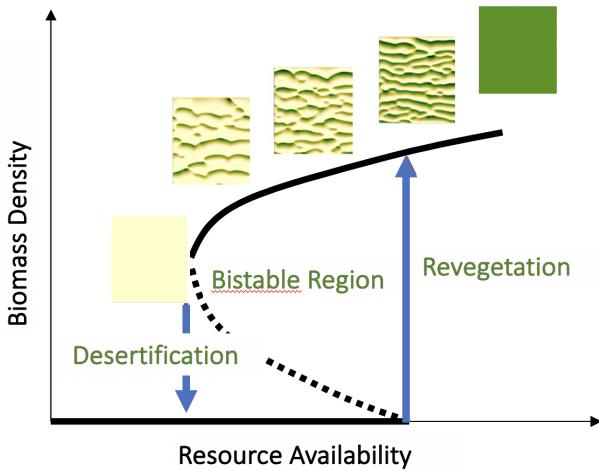


Figure 5.1: Models predict how ecosystems undergo a predictable sequence of patterns as resource availability decreases, after Rietkerk et al. [2004]. This solid lines represent equilibrium densities, and arrows represent state transitions between self-organized patchy and homogeneous states.

processes, typically with fast timescales that require high temporal resolution to be resolved by a model. Additionally, the more complex the vegetation patterning models become, the more challenging it is to relate the insights obtained from analytical treatments to more complex modeled systems. This is problematic both in terms of testing and parameterizing simple models to represent real places, and in terms of using the potentially powerful insights about pattern dynamics that can be gained from analytical treatments [Siteur et al., 2014, 2016, Foti and Ramírez, 2013, Deblauwe et al., 2012].

Thus, it would be useful to (i) reduce the computational cost of representing fast-timescale processes in vegetation patterning models, and (ii) to relate the parameterization of fast-timescale processes in such models to the mean-field or phenomenological representations adopted in analytically tractable models. While it may not be possible to achieve a one-to-one correspondence between the predictions of a phenomenological model and the predictions of a model incorporating a description of fast-timescale processes, it would still be useful to identify, for example, how a change in land surface slope or rainstorm intensity - physically measurable parameters that are used directly in the description of the fast runoff processes - could be captured by changing phenomenological parameters in a reduced complexity model.

This chapter attempts to develop a framework for the representation of fast processes associated with rainfall, specifically runoff generation and the infiltration of runoff in vegetated patches, by simulating these processes with a physics-based model, and then emulating the simulation results using the machine learning (ML) approach introduced in Chapter 4. The overland flow-infiltration model consists of the Saint Venant Equations (SVE) coupled to Richards equation (referred to here as the SVE-R model), and random forest (RF) regression is used to emulate the SVE-R model predictions. This chapter specifically addresses

the following research questions:

1. Can ML emulation of the SVE-R model predictions, specifically runoff and infiltration, replace direct predictions by the SVE-R model in a pattern forming model, and preserve the predicted vegetation patterns?
2. How does pattern morphology - specifically the mean wavelength, biomass density and peak biomass - respond to storm depth, storm duration and annual rainfall?
3. What correspondences can be drawn between the parameterization of rainfall processes in a phenomenological model and these storm features (depth, duration and annual rainfall)?

The analysis here focuses on the predictions of an existing phenomenological model, proposed by Rietkerk et al. [2002] and modified to account for intermittent rainfall conditions by Guttal and Jayaprakash [2007].

5.2 Methods

The Rietkerk/Guttal model

The Rietkerk/Guttal (R/G) model describes the dynamics of three coupled variables: plant biomass density (P ; g m⁻²), the local depth of water stored within the soil (W ; mm), and the local depth of ponded water on the land surface (O ; mm) [Rietkerk et al., 2002], as shown below:

$$\frac{\partial O}{\partial t} = r - \alpha O \frac{P + W_0 k_2}{P + k_2} + \nu \frac{\partial O}{\partial x} \quad (5.1)$$

$$\frac{\partial W}{\partial t} = \alpha O \frac{P + W_0 k_2}{P + k_2} - g_{max} \frac{W}{W + k_1} P - r_w W + D_w \nabla^2 W \quad (5.2)$$

$$\frac{\partial P}{\partial t} = \frac{1}{\tau_p(W)} \left(c g_{max} \frac{W}{W + k_1} - d \right) P + D_p \lambda^2 P \quad (5.3)$$

where

$$\frac{1}{\tau_p(W)} = \frac{1}{4} \frac{W^2 + f k_3^2}{W^2 + k_3^2} \quad (5.4)$$

The $\frac{1}{\tau_p(W)}$ term is a modification of the original Rietkerk model, introduced by Guttal and Jayaprakash [2007] to account for drought-adapted vegetation. The modification slows

the rates of biomass growth and decay during periods of low soil moisture, in order to reflect drought-adaptations in desert vegetation. Figure 5.2 shows examples of patterns produced by the R/G model for a range of annual rainfalls, illustrating how pattern wavelength increases as rainfall decreases. Descriptions of the parameters are given in Table 5.1.

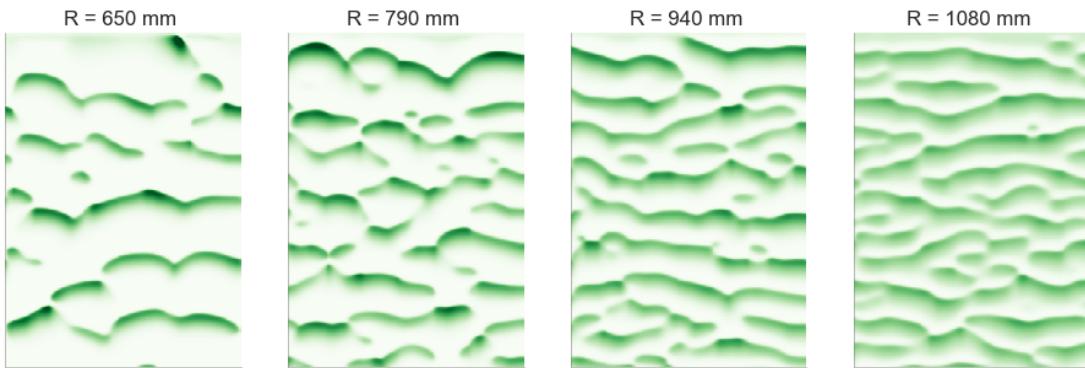


Figure 5.2: Vegetation patterns produced by the R/G model for a range of annual rainfall inputs. The domain size is 800×400 m, with 2 m grid resolution, and the final time is 10,000 days. The initial conditions are the same for all cases, with biomass initialized as a randomly-generated binary array in which $P = 20 \text{ g m}^{-2}$ in 50% of the grid cells and 0 elsewhere. Parameters are $\alpha = 1 \text{ day}^{-1}$, $k_2 = 25 \text{ g m}^{-2}$, $\nu = 25 \text{ m day}^{-1}$, and $W_0 = 0.1$, and otherwise listed in Table 5.1.

The R/G model is a phenomenological model, and many of the representations contained within are simplified [Kefi et al., 2008]. Here, the focus is on improving the representation of the surface water from that in the R/G model, which uses several simplifications: (i) a constant rainfall rate, r (mm/day), in contrast to the reality of rainfall inputs in drylands, typically as short, intense storms; (ii) a constant advection rate representing overland flow, $\nu \frac{\partial Q}{\partial x}$, where ν represents the effect of the land surface slope; and (iii) a heuristic depiction of infiltration as a function of biomass and ponded water depth. Importantly, these depictions avoid physical representations of overland flow and infiltration processes, which depend on storm-scale properties (such as rainfall intensity and storm duration) and landscape properties (such as hillslope gradient and soil infiltration capacity).

Modifications to the Rietkerk/Guttal Model

To allow a more mechanistic depiction of overland flow processes within the R/G model, I replaced the continuous rainfall in Equation 5.1 with discrete storm events. This change involves replacing Equation 5.1 with SVE-R model predictions of the runoff and infiltration processes for each storm. The infiltration term in the soil water equation (Equation 5.2), $\alpha O \frac{P+W_0 k_2}{P+k_2}$, is replaced with the SVE-R model prediction of the cumulative infiltration depth after a given storm. With this modification, the soil water equation becomes:

Table 5.1: Rietkerk/Guttal model parameters, descriptions and values. Values are listed for parameters that appear in the modified version of the R/G model presented here.

Parameter	Description	Units	Values
<i>(A) Parameters that appear in both the original R/G model and the modified version presented here.</i>			
c	Conversion factor for water uptake to plant biomass	$\text{g m}^{-2} \text{mm}^{-1}$	5
g_{max}	Maximum specific water uptake	$\text{mm g}^{-1} \text{m}^{-2} \text{day}^{-1}$	0.1
k_1	Half-saturation constant of specific plant growth and water uptake	mm	5
m	Specific rate of plant density loss due to mortality	day^{-1}	0.25
r_w	Soil water loss rate due to evaporation and drainage	day^{-1}	0.4
k_3	Soil water scale at which plant growth resumes normal metabolic activity	mm	10
f	Metabolic activity coefficient at zero water availability	—	0.04
D_p	Plant dispersal diffusion constant	$\text{m}^2 \text{ day}^{-1}$	0.01
D_w	Soil water diffusion constant	$\text{m}^2 \text{ day}^{-1}$	0.1
r	Daily rainfall rate	mm day^{-1}	—
<i>(B) R/G model parameters that are replaced by SVE-R predictions in the modified version.</i>			
α	Rate of surface water infiltration	day^{-1}	—
k_2	Plant density scale determining how surface water infiltration increases with P	g m^{-2}	—
W_0	Minimum surface water infiltration coefficient in the absence of plants	—	—
ν	Effective slope parameter	m day^{-1}	—

$$\frac{\partial W}{\partial t} = i - g_{max} \frac{W}{W + k_1} P - r_w W + D_w \nabla^2 W \quad (5.5)$$

where i is an infiltration rate, which is zero between storm events and predicted by the SVE-R model during storm events. For each storm, the SVE-R model predicts an infiltration depth (I_{storm}), which is converted to a storm infiltration rate using the R/G model time discretization: $i_{storm} = I_{storm}/dt$. The infiltration rate is non-zero for timesteps during storms, and zero for all other timesteps.

By replacing the R/G surface water equation with SVE-R simulated storms, several R/G parameters are removed from the modified R/G-SVE-R model, specifically α , W_0 , k_2 and ν . New landscape parameters are introduced, including the hillslope gradient, S_o ,

and the saturated hydraulic conductivities of bare soil and vegetated areas, $K_{S,B}$ and $K_{S,V}$, respectively. These landscape parameters are incorporated into the pattern forming model through their influence on overland flow and infiltration processes (land surface slope and soil infiltration capacity are specified in the SVE-R component of the R/G-SVE-R model, and ‘felt’ in the R/G component through the SVE-R simulated runoff and infiltration patterns). For this study, I limited the analysis to the influence of storm characteristics on pattern morphology, and assumed a single set of landscape parameters, summarized in Table 5.3B. I also imposed a fixed ratio of 0.1 between the hydraulic conductivities of bare soil and vegetated areas, which is typical of the contrasts in infiltration capacity observed in patterned landscapes [Bromley et al., 1997].

The dependencies within the R/G model are preserved by using the biomass field from the R/G model to describe the spatial pattern of vegetation in the R/G-SVE-R model, and conditioning the infiltration parameters of the SVE-R model on the vegetation distribution (see Chapter 2). However, biomass evolves as a continuous spatial field in the R/G model, and the SVE-R model treats vegetation in a binary presence/absence fashion. Therefore, to finish the coupling of the models, a threshold biomass value, P_T , was introduced to distinguish vegetation from bare soil. P_T is specified as the biomass density in the R/G model for which the infiltration rate is half of its theoretical (infinite biomass) maximum, αO , and P_T is thus related to W_0 and k_2 as $P_T = k_2/(1-2W_0)$. Interpreted as the infiltration ratio of bare soil ($P = 0$) to vegetated patches (large P), $W_0 = 0.1$ is analogous to the SVE-R prescribed infiltration capacity ratio. k_2 does not have a direct analogue in the SVE-R model, so $k_2=25 \text{ g m}^{-2}$ was selected, following Guttal and Jayaprakash [2007]. For these R/G parameters, the corresponding biomass threshold value is $P_T = 20 \text{ g m}^{-2}$.

I verified that this modified R/G model - the R/G-SVE-R model - would still produce vegetation patterns by running it for all combinations of the parameters listed in Table 5.3C (with storm properties listed in 5.3A). The vegetation distribution was initialized in each simulation as a randomly-generated binary array with density $\phi_V = 0.5$ (vegetation cover in half of the cells). In Table 5.3, the random seed s refers to the number used to initialize the Python pseudo-random number generator, which is needed to (i) ensure reproducibility and (ii) use the same initial conditions for multiple simulations. The 4 random seeds listed in Table 5.3C signify that the simulations were replicated 4 times with different initial vegetation fields. The biomass fields were initialized with $P = 0$ in bare soil cells and $P = 40 \text{ g/m}^2$ in vegetated cells (twice P_T).

Visual inspection confirmed that patterns are still produced in the R/G model modified to include discrete storms (Figure 5.3 displays a subset of the final biomass fields). Due to its high computational requirements, the R/G-SVE-R model was tested on a relatively small domain, so boundary effects are likely present. Also because of computational limitations, the simulations were run for 365 - 1600 days (108 - 192 storms). This relatively short duration was long enough for patterns to emerge, but insufficient to reach a steady state (i.e. where hillslope-mean biomass density remains constant in time).

Having coupled the SVE-R model to the R/G model and verified that the coupled system continues to produce patterns, it is possible to investigate the effects of storm characteristics

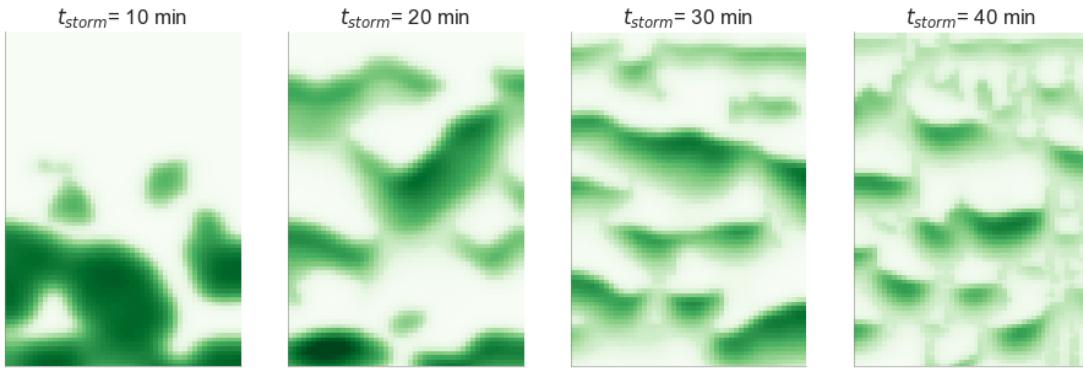


Figure 5.3: Vegetation patterns produced by the R/G-SVE-R model for a range of storm durations, with $d = 1.2 \text{ cm}$, and $R = 440 \text{ mm}$. The final time is 1200 days, and the domain size is $200 \times 100 \text{ m}$ with 2 m grid resolution.

on pattern formation dynamics. The next section describes the storm climatologies developed for this purpose.

Storm Climatology

Storm climatology is high-dimensional [Feng et al., 2017], because rainfall occurrence, intensity and duration vary as a stochastic process on timescales ranging from seconds to years. I considered storm climatology to comprise distinct properties of event occurrence on within-year timescales, and event properties on within-storm timescales, as illustrated in Figure 5.4.

For illustrative purposes, this study simplifies the variability of rainfall climatologies by simulating sequences of identical storms with a fixed recurrence interval. This avoids the additional complexities of seasonality, the distribution of recurrence intervals and depths, and their influence on pattern morphology (resolving these effects requires Monte-Carlo sampling of the recurrence interval and rainfall depth/duration distributions, which is a logical extension of the present work but beyond the current scope). Thus, I prescribe an annual precipitation total, R , which is disaggregated into different patterns of storm occurrence (storminess), as $R = Nd$, where N is the number of storms per year and d is the depth of each storm. For fixed storm depth, a given R is obtained by adjusting the storm recurrence interval, τ .

Each storm is further disaggregated into a set of possible storms, each with a fixed mean intensity and mean duration, as $d = pt_{\text{storm}}$, where p and t_{storm} represent the storm intensity (cm/hr) and duration (min), respectively. These features are expected to impact infiltration depths and spatial patterns: more runoff occurs for more intense storms, likely increasing the heterogeneity of the distribution of water resources. Different patterns of storm occurrence alter the exposure of plants to variability in soil moisture resources, and thus the simulated

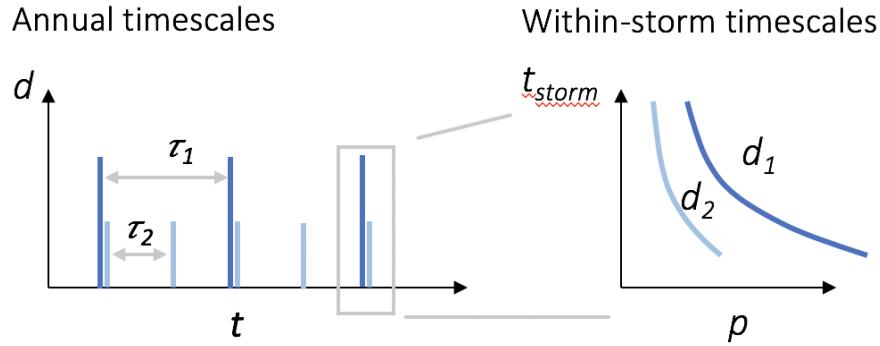


Figure 5.4: Storm climatology here is considered at ‘within year’ and ‘within-storm’ timescales. Within each year, storms are assumed to occur with a specified occurrence interval (τ) and depth (d). A storm that takes place is then specified by its duration (t_{storm}), its intensity (p) and the total depth. For simplicity and illustrative purposes only, storms were assumed to be identically sized and evenly distributed through time; real locations would be characterized additionally by seasonality and a distribution of waiting times and storm depths.

Table 5.2: Rainfall intensity values (in cm/hr) corresponding to the parameter space of rainfall depths and durations ($p = d/t_{\text{storm}}$).

Rain duration (min)	Rain depth (cm)			
	0.4	0.8	1.2	1.6
10	2.4	4.8	7.2	9.6
20	1.2	2.4	3.6	4.8
30	0.8	1.6	2.4	3.2
40	0.6	1.2	1.8	2.4

vegetation growth and spread dynamics.

To explore the interactions of total annual rainfall inputs with the timing and duration of individual storms, I formed a parameter space consisting of all possible combinations of four rainfall depths ($d = 0.4, 0.8, 1.2$ and 1.6 cm) and four storm durations ($t_{\text{storm}} = 10, 20, 30$ and 40 minutes). The parameter space thus consists of 16 unique storm events, with rain intensity $p = d/t_{\text{storm}}$, as listed in Table 5.2.

With this approach, storm climatology is defined by three variables, d , t_{storm} , and R (where $p = d/t_{\text{storm}}$ and $\tau = R/d$). Among these parameters, d and t_{storm} are specified in the SVE-R model component (within storm processes), and R is specified in the R/G model (between storm processes).

Table 5.3: Parameters for three sets of model simulations: (B) the SVE-R simulations to train the random forest regressors, (C) the pattern simulations to validate the RF emulation approach, which use both R/G-SVE-R and R/G-emulator models, and (D) the pattern simulations to explore the sensitivity to storm climatology. Within each section, where multiple parameters are listed, the cases were run factorially to explore all parameter combinations. Note that these factorial combinations included the storm parameters listed in section (A). Duplicate parameter listings are included where necessary for clarity. Landscape parameters (hillslope gradient and soil parameters) are explicitly prescribed in the SVE-R simulations, and implicitly prescribed in the pattern simulations through the RF regressors.

Parameter	Description	Units	Values
<i>(A) Storm parameters</i>			
d	Storm depth	cm	0.4, 0.8, 1.2, 1.6
t_{storm}	Storm duration	min	10, 20, 30, 40
<i>(B) Parameters for the SVE-R simulations to train the RF regressors (See Chapter 4).</i>			
Model	SVE-R		
$L_x \times L_y$	Hillslope dimensions	m	200×100
S_o	Hillslope gradient	%	2.0
$K_{S,V}$	Hydraulic conductivity (vegetated sites)	cm/hr	1.5
$K_{S,B}$	Hydraulic conductivity (bare soil)	cm/hr	0.15
<i>(C) Parameters for the pattern simulations to validate the RF emulation approach.</i>			
Models	R/G-SVE-R and R/G-emulator		
$L_x \times L_y$	Hillslope dimensions	m	200×100
R	Annual precipitation	mm	440
s	Random seed	—	0,1,2,3
$\phi_{V,0}$	Initial fractional vegetation cover	—	0.5
P_T	Threshold biomass	g m^{-2}	20.
<i>(D) Parameters for the pattern simulations to explore the climate sensitivity.</i>			
Model	R/G-emulator		
$L_x \times L_y$	Hillslope dimensions	m	800×400
R	Annual precipitation	mm	290, 365, 440, 515, 585
s	Random seed	—	0,1,2,...,9
$\phi_{V,0}$	Initial fractional vegetation cover	—	0.5
P_T	Threshold biomass	g m^{-2}	20.

Replacing the SVE-R Model with a Random Forest Emulator

Fully exploring the effects of the rainfall climatologies shown in Tables 5.2 and 5.3 on pattern morphologies requires simulating individual storms many hundreds of times. Due to the high computational demands of the SVE-R model, this is either numerically expensive or intractable. To reduce the computational demand of simulating these storms, I developed an emulation approach which replaces the use of the SVE-R model in the R/G-SVE-R model with random forest regression. This reduces the computational time by several orders of magnitude relative to the R/G-SVE-R model (see Chapter 4). A random forest regressor can be used to emulate the SVE-R model predictions in this context, because it can adapt to the variable biomass fields that will evolve as the pattern forming model is run.

The R/G-emulator model relies on trained RF regressors, which were obtained following the methods described in Chapter 4. The RF regressors can extrapolate to new vegetation patterns, but are storm and landscape-specific. Accordingly, a separate RF was trained for each of the 16 unique storms (Table 5.2) and the landscape parameters listed in Table 5.3B. To train each RF regressor, the SVE-R model was run for a set of 20 randomly generated vegetation patterns with a range of vegetation cover fractions and patch length scales (see Chapter 4). The RF regressors were trained on the SVE-R model predictions using the Python library sci-kit learn, and, once trained, predict the cumulative infiltration depth for an input binary vegetation pattern. The trained RF regressors then replaced the SVE-R model in the R/G-SVE-R model, so that the storm-scale predictions were made by an RF regressor. With this replacement, I addressed the research questions as follows:

- 1. Can ML emulation of the SVE-R model predictions, specifically runoff and infiltration, replace direct predictions by the SVE-R model in a pattern forming model, and preserve the predicted vegetation patterns?**

To validate the use of the RF regressor as a replacement for the SVE-R model, I first established the fidelity of the infiltration fields produced by RF regression to the SVE-R predictions for representative vegetation patterns (see Chapter 4). I also verified that the simulations using the RF model could extend the domain size beyond the training space to a 400×200 cell grid, also with 2 m resolution, without domain effects significantly altering the infiltration behavior. The larger domain size becomes important when exploring pattern morphology, in order to resolve patterns with long wavelengths.

Having established robust infiltration emulation, I tested the effects on pattern morphology of substituting an RF regressor for the SVE-R model, using common, random initial conditions. The SVE-R and RF emulator versions were run for every combination of the parameters listed in Table 5.3C (and the storms listed under ‘Storm parameters’). For each storm case, the simulations were replicated 4 times with different random seeds to generate the initial vegetation fields. The SVE-R and RF emulator versions of the R/G model were run for over one hundred storms (108 to 192, varying for each simulation), and the resulting pairs of patterns were visually and quantitatively compared.

Because the model undergoes strong transient conditions (particularly in evolving vegetation patterns from the initial random condition to the final patterned state), I expected to see differences in the details of the patterns in each case. Quantitative comparisons therefore focused on the mean wavelength, mean biomass density and peak biomass predicted in each case. Mean biomass density, \bar{P} , is defined as the hillslope-mean biomass, including bare soil areas, evaluated at the final time. The peak biomass, $P_{90\%}$, is defined as the 90th percentile of the within-patch biomass density. To estimate the mean wavelengths, λ , I used the Fourier windowing method developed by Penny et al. [2013], which computes the wavelengths of subsets of the pattern using a moving window. The windows are then merged to create a map of pattern wavelengths, from which the hillslope mean wavelength λ is obtained. These metrics were computed for each pattern produced by the R/G-SVE-R and R/G-emulator models.

To validate the use of the R/F-emulator model to explore the pattern sensitivity to storm climatology, I checked whether the relationships between pattern morphology and storm characteristics were preserved between the R/G SVE-R and R/G-emulator models.

2. How does pattern morphology - specifically the mean wavelength, the biomass density and peak biomass - respond to storm depth and duration?

To address this question, the R/G-emulator model was used to explore a larger parameter space, listed in 5.3D. This included a larger hillslope domain (800×400 m), and multiple values of R . For each unique storm climatology, the simulations were replicated with 10 different random seeds (initial vegetation patterns), in order to characterize the range of variability in the resultant pattern morphologies. The patterns were allowed to evolve for 10,000 days, which was sufficient for the biomass density to reach a steady state. Simulations were terminated if the vegetation field arrived at a bare or completely vegetated state.

The pattern sensitivity to storm climatology was explored by plotting the final patterns for various cross sections in the parameter space (which includes R , d , t_{storm} , and the random seed). Visual inspection assessed (i) whether the wavelengths and biomass densities varied with rainfall parameters, and (ii) the sensitivity of the pattern morphology to variation in the random seed. Visual inspection also explored whether observed trends in pattern morphology with t_{storm} and d could be explained by rain intensity alone. This step was necessary because increasing d and decreasing t_{storm} both correspond to increasing p in the storm parameter space (Table 5.2).

Summary metrics, including the pattern wavelength, mean biomass density, and peak biomass, were extracted from the biomass fields at the final timestep. For each set of rainfall parameters, final metrics were obtained by averaging over the 10 random seeds. The range of values obtained by varying the random seed was used to assess the range of pattern variability associated with each storm climatology.

3. What correspondences can be drawn between the parameterization of rainfall processes in a phenomenological model and the rainfall processes in a model that emulates storm runoff?

While the R/G-emulator model has the distinct advantage of more realistic storm-scale hydrology, it lacks the analytical tractability of the Rietkerk-Guttal model. A mapping from the R/G-emulator parameters to the R/G parameters would provide a way to recover this analytical tractability while retaining realistic storms. There are many features that one might want to replicate with such a mapping, for example, the steady state biomass density, pattern wavelength, patch size distribution and timescales of biomass growth and decay. Conceptually, an m -dimensional space of relevant pattern features exists, and parameters from both models can be mapped to this space (see Figure 5.5). By passing through this m -dimensional space, a mapping from real storm characteristics to phenomenological R/G parameters could thus be obtained. The relevant pattern features can be related to the model parameters via sensitivity analysis, but this is not known a priori for either model.

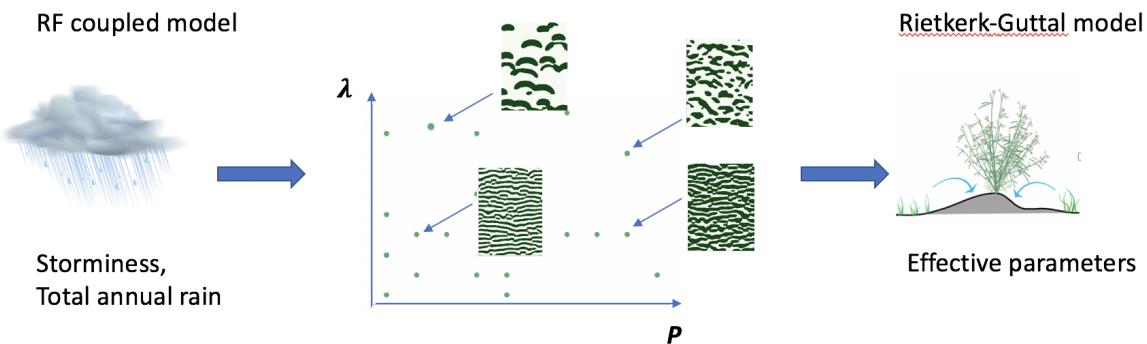


Figure 5.5: A schematic illustrating how the R/G-emulator model parameters could be mapped to the Rietkerk-Guttal model, by passing through a multidimensional space of predicted features (here, biomass and wavelength).

A general approach to obtain such a mapping is outlined here, with the caveat that the implementation specifics would depend on the goals of the end user. The potential to map the storm characteristics (d and t_{storm}) in the R/G-emulator model to the R/G parameters is examined in an example using two features: average biomass density and pattern wavelength. The R/G-emulator simulations produce a space of $\bar{P}-\lambda$ points, each corresponding to a different combination of d , t_{storm} and R . I attempted to replicate this space of $\bar{P}-\lambda$ points with the R/G model by varying R , ν and the infiltration parameters α , k_2 , W_0 (see Table 5.4 for parameter values). For parameters that the R/G and R/G-emulator models have in common, the same values were used (see Table 5.1), with the exception of R , because the models produce patterns for different ranges of annual rainfalls.

Table 5.4: Parameters for the Rietkerk-Guttal model simulations. The parameter cases were run factorially to include all parameter combinations.

Parameter	Values	Units
α	0.1, 1.0, 2.0	day ⁻¹
k_2	5, 25, 50	g m ⁻²
W_0	0.01, 0.1, 0.5	—
R	1.4, 1.8, 2.2, 2.6, 3.0, 3.4	mm day ⁻¹
ν	1, 25, 50	m day ⁻¹

5.3 Results

The research questions listed in the previous sections are addressed in the following.

1. **Can ML emulation of the SVE-R model predictions, specifically runoff and infiltration, replace direct predictions by the SVE-R model in a pattern forming model, and preserve the predicted vegetation patterns?**

An example of the visual comparison is presented in Figure 5.6 (compare panel A, the SVE-R version, and panel B, the RF emulator version). The generated patterns are visually similar, while differing in details, as expected. Scatter plots comparing the R/G-SVE-R and the R/G-emulator model predictions are shown in Figure 5.7 for wavelength (panel A), biomass density (panel B), and peak biomass (panel C). The R/G-emulator model predicts lower wavelengths than the R/G-SVE-R model for some of the longer wavelength cases. The R/G-SVE-R model also tends to predict larger peak biomass values than the R/G-emulator model.

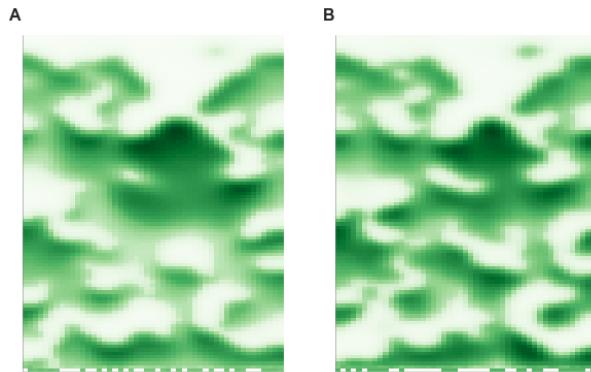


Figure 5.6: Vegetation patterns simulated by the SVE-R (A) and the RF emulator (B) versions of the pattern forming model after 1200 days (108 storms). Storm parameters: $t_{storm} = 30$ min, $d = 1.6$ cm, $R = 440$ mm.

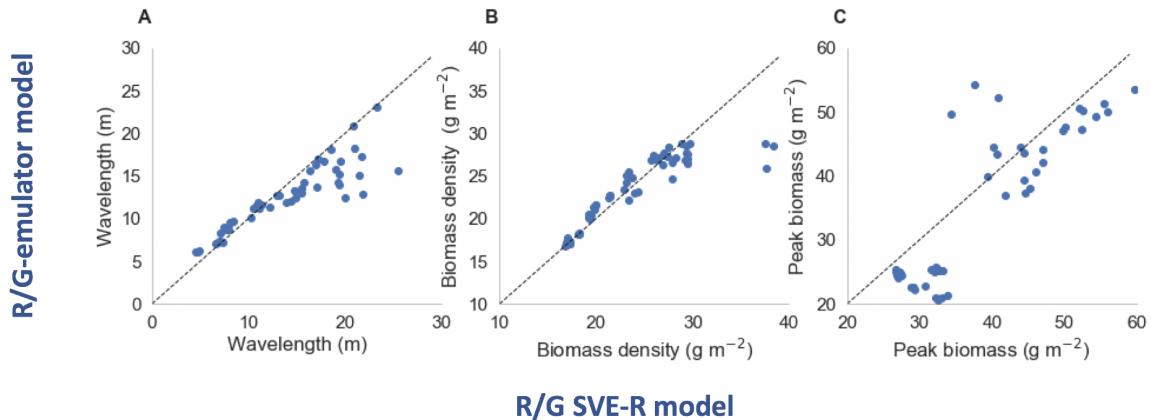


Figure 5.7: Comparing the predictions of the R/G-SVE-R and the R/G-emulator models for pattern wavelength (A) biomass density (B), and peak biomass (C).

Figure 5.8 assesses the ability of the R/G-emulator model to reproduce the R/G-SVE-R model predictions of the sensitivity of the pattern morphologies to storm depth and duration. The R/G-SVE-R model produces longer wavelengths for some short duration storms than those produced by the R/G-emulator model (panel A), specifically for the cases with $d = 0.8$ and 1.2 cm. Additionally, in the 1.6 cm, 10 min storm case, the SVE-R version did not produce patterns, while the RF emulator version did. The relationships between biomass density and storm characteristics (panel B) are very similar between the SVE-R and emulator versions, with the exception of the 1.6 cm, 10 and 20 min storms. The SVE-R model tends to predict higher peak biomass values (panel C), but the sensitivity to d and t_{storm} is preserved (also with the exception of the 1.6 cm, 10 and 20 min storms). Overall, the results are similar enough that the trends found with the R/G-emulator model are inferred to be a good representation of the trends associated with the R/G-SVE-R model.

Computational limitations prevent running the R/G-SVE-R model until steady state. Despite this limitation, the results suggest that the SVE-R and RF emulator versions produce similar pattern morphologies, as quantified by the wavelength, biomass density and peak biomass. Because simulations using the SVE-R and emulator versions started from the same initial conditions, Figures 5.7 and 5.8 quantify the divergence between these versions after $365 - 1600$ days. Thus, since the goal was to assess the suitability of RF emulation as a replacement of the SVE-R model, the computational restrictions are less problematic than if the goal were to assess the final pattern morphologies.

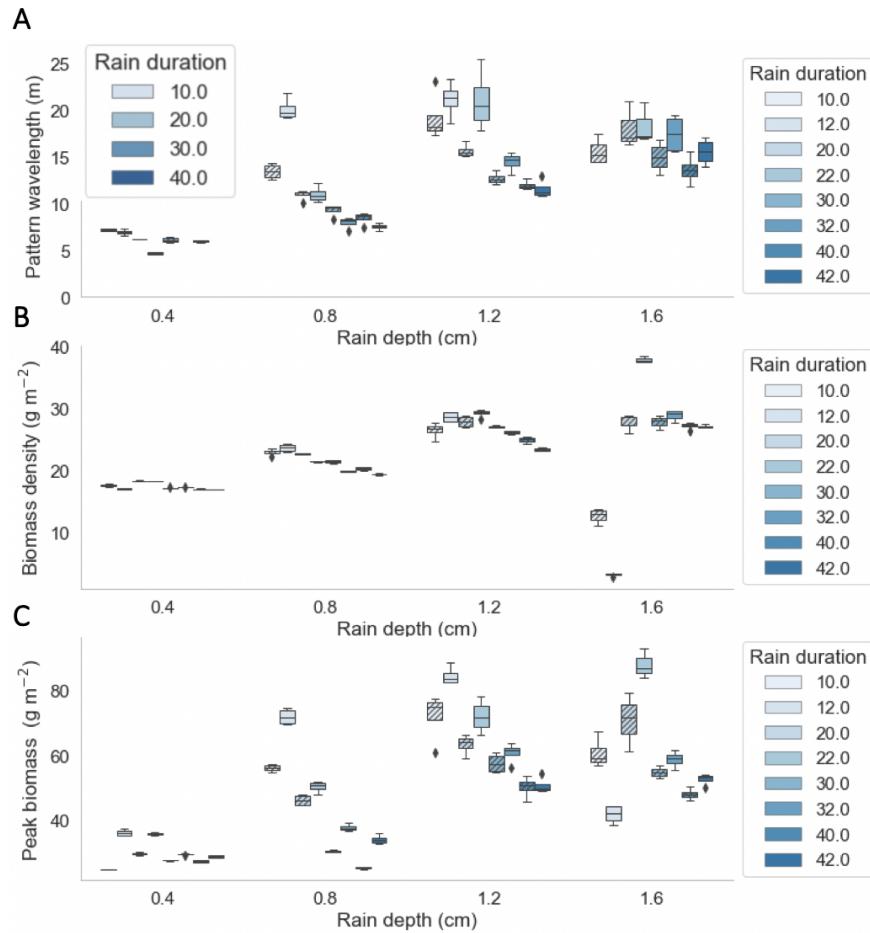


Figure 5.8: Box plots compare the sensitivity of pattern wavelength (panel A), biomass density (panel B), and peak biomass (panel C) associated with the SVE-R and emulator versions of the R/G model. The boxes with hatches show R/G-emulator model predictions, and boxes without show the R/G-SVE-R model predictions.

2. How does pattern morphology - specifically the mean wavelength, biomass density, peak biomass - respond to storm depth and duration?

The sensitivity of the pattern morphology to storm depth, duration and annual rainfall can be visualized by plotting the final patterns for different cross-sections of the storm parameter space (R , d and p). Figure 5.9 illustrates this sensitivity in a grid of all 16 storms, with annual rainfall held constant at 440 mm. For fixed storm duration, pattern wavelengths increase with increasing storm depth. Conversely, with depth held constant, pattern wavelengths increase with decreasing duration (more intense) storms. To confirm that these trends in pattern morphology cannot be explained by rainfall intensity alone, Figure 5.10 compares

4 pattern morphologies with the same storm intensity ($p = 2.4 \text{ cm/hr}$) and annual rainfall ($R = 440 \text{ mm}$). With storm intensity held constant, pattern wavelength and peak biomass increase with increasing storm depth and duration, indicating that variations in pattern morphology are not solely attributable to variations in storm intensity.

For fixed storm characteristics, increasing annual rainfall produces shorter wavelength patterns with larger peak biomass densities, as illustrated by Figure 5.11.

Figure 5.12 compares the final patterns for different random seeds, with all other parameters the same. The patterns differ in the details, but are qualitatively very similar, indicating that the sensitivity of the pattern morphology to initial conditions is small for this parameter case.

Figure 5.13A summarizes the mean wavelength sensitivity to R , d and t_{storm} , confirming the trends visible in Figures 5.9 and 5.11. Pattern wavelength increases with increasing R , in agreement with many previous studies [Borgogno et al., 2009]. Biomass density increases with d and R and decreases with t_{storm} in most cases (panel B), however, there are several cases where the biomass density sharply decreases with increasing rain depth (e.g. 30-40 minute storms with $R=290 \text{ mm}$). In these cases, there is a similar decrease in peak biomass density (panel C) and infiltration fraction (panel D). Figure 5.14 shows the corresponding vegetation maps for one such case ($R = 290 \text{ mm}$, $t_{\text{storm}} = 30 \text{ min}$), which reveals that the collapse in biomass reflects a change in pattern morphology from stripes to spots. These results may suggest an alternate path to desertification: for the same annual rainfall, a landscape may progress from striped to spotted and eventually bare if the plants are unable to access the rainfall inputs. With less frequent, more intense storms, more water is lost as runoff and plant water resources decline.

For the highest intensity storms, the patterns do not reach the divide (e.g. $t_{\text{storm}} = 10 \text{ min}$, $d = 1.6 \text{ cm}$ in Figure 5.9), presumably because the runoff source area is limited by the no-flow upper boundary and the vegetation does not receive sufficient water inputs. McGrath et al. [2012] observed a similar effect from a no-flow boundary at the divide, in a study of the effects of micro-topography on pattern formation.

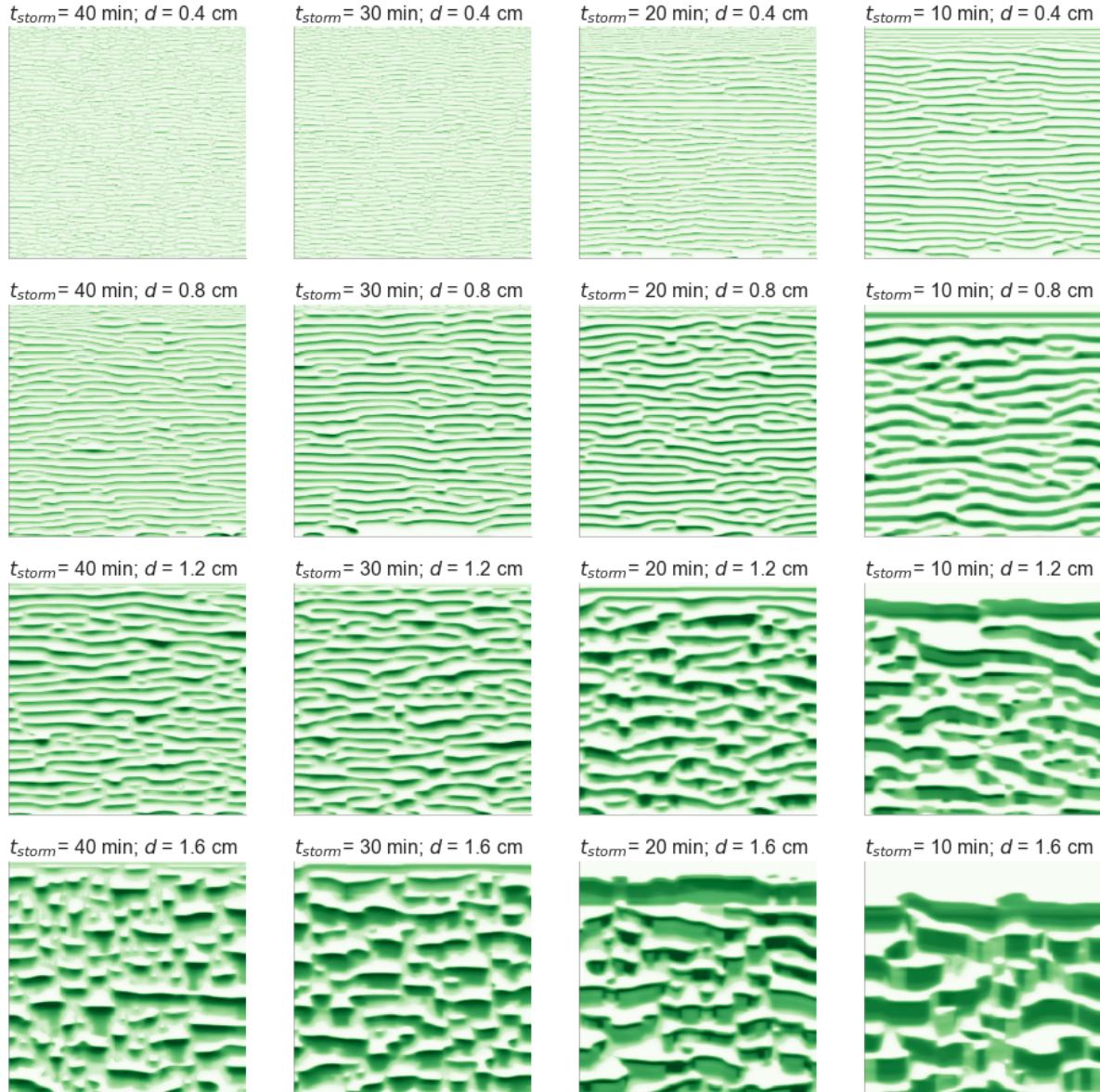


Figure 5.9: Patterns obtained for all storms (as listed in Table 5.2A) and $R = 444 \text{ mm}$. Wavelength increases with decreasing storm duration (left to right panels) and with increasing storm depth (upper to lower panels), both corresponding to increasing storm intensity. For the highest intensity storms (e.g. $t_{storm} = 10 \text{ min}, d = 1.6 \text{ cm}$), the patterns do not reach the divide.

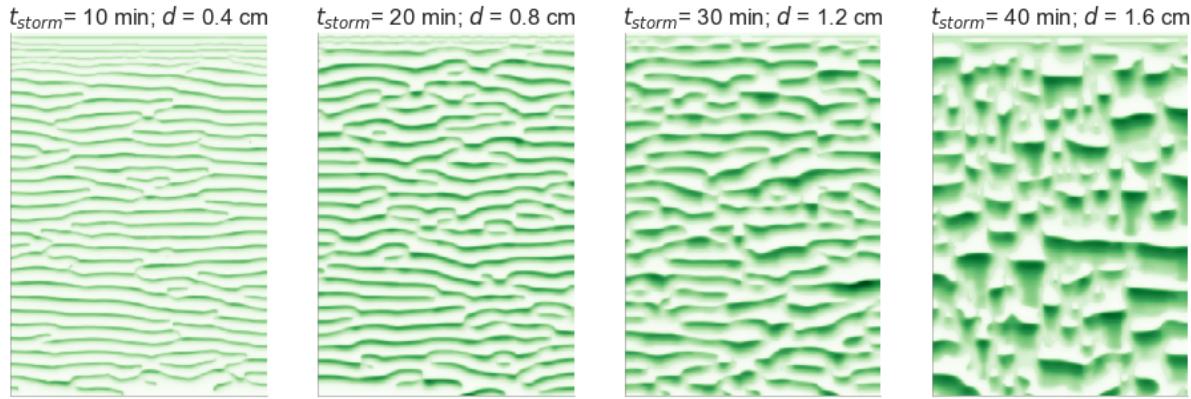


Figure 5.10: Patterns obtained for the same rainfall intensity ($p = 2.4 \text{ cm/hr}$) and annual rainfall ($R = 440 \text{ mm}$), for different combinations of d and t_{storm} , to illustrate that variations in pattern morphology are not solely attributable to variations in storm intensity. With storm intensity held constant, pattern wavelength and peak biomass increase with increasing storm depth and duration.

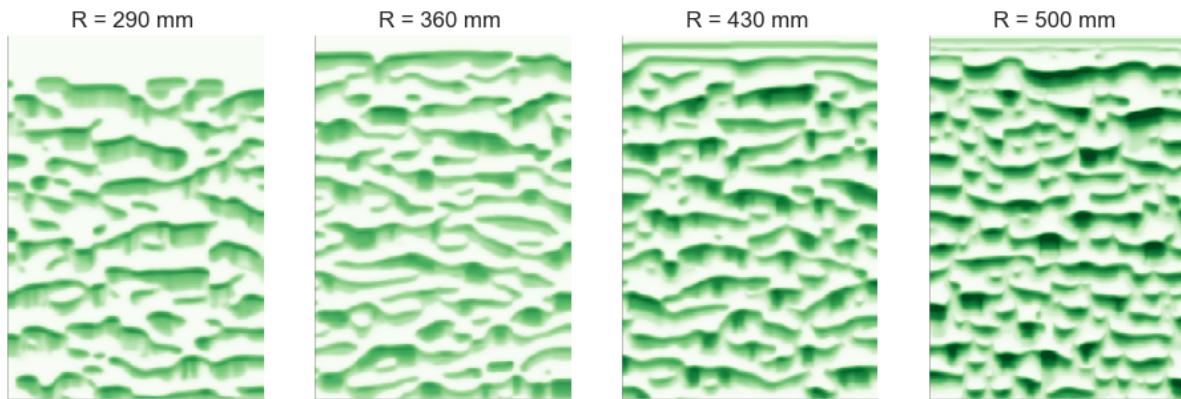


Figure 5.11: Illustrating the effects of varying R , for fixed $d = 1.2 \text{ cm}$ and $t_{storm} = 20 \text{ min}$. Increasing R corresponds to decreasing wavelength and increasing biomass density.

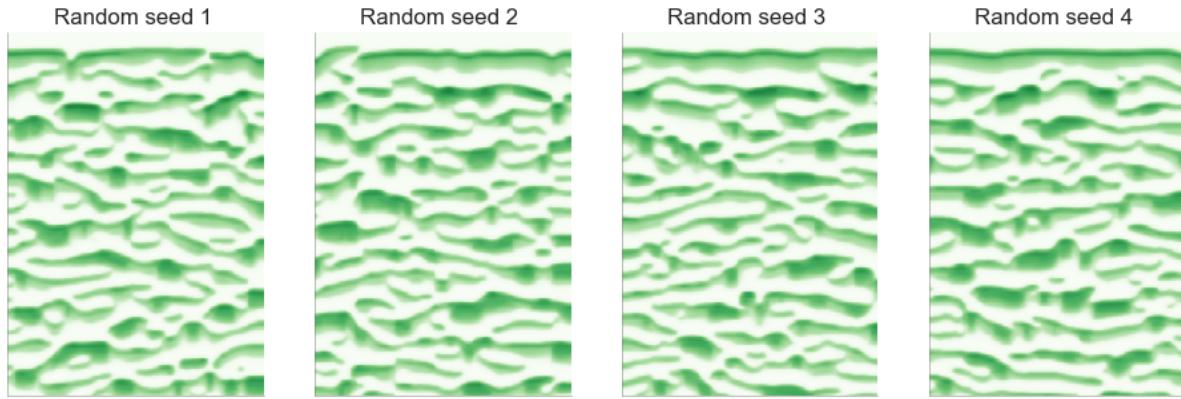


Figure 5.12: Illustrating the effects of changing the random seed, with all other parameters fixed ($R = 360$ mm, $d = 1.2$ cm, and $t_{storm} = 30$ min). The final patterns differ in the details, but are qualitatively very similar.

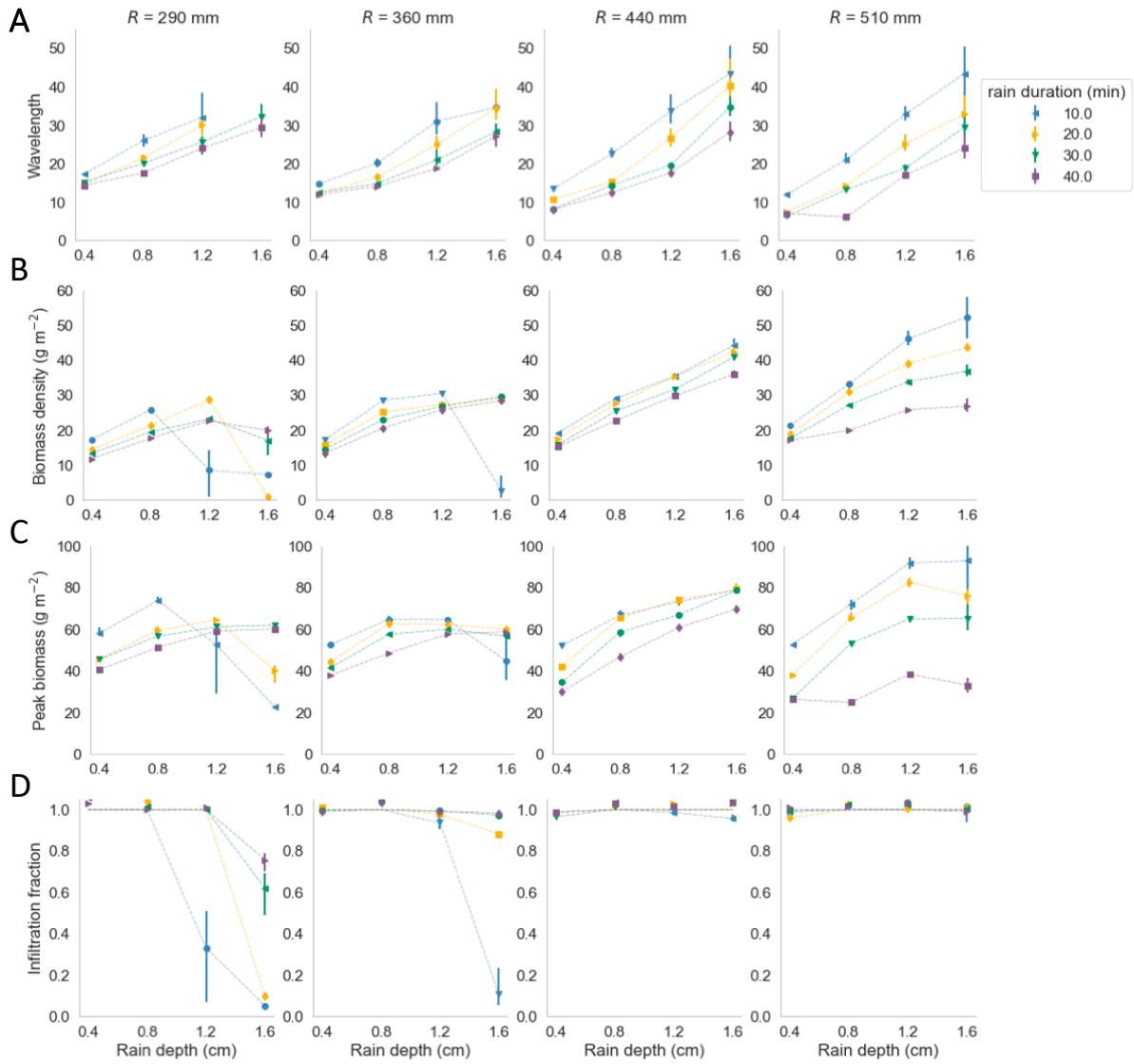


Figure 5.13: Sensitivity of the pattern morphology to R , d and t_{storm} , including wavelength (A), biomass density (B), peak biomass (C) and infiltration fraction (D). Bars indicate the range of values spanned with different initial conditions (random seeds), and lines drawn between points are as visual aides only. In most cases, biomass density and peak biomass increase with R and d , and decrease with t_{storm} , with exceptions where the pattern morphology transitions from stripes to spots as storm depth increases.

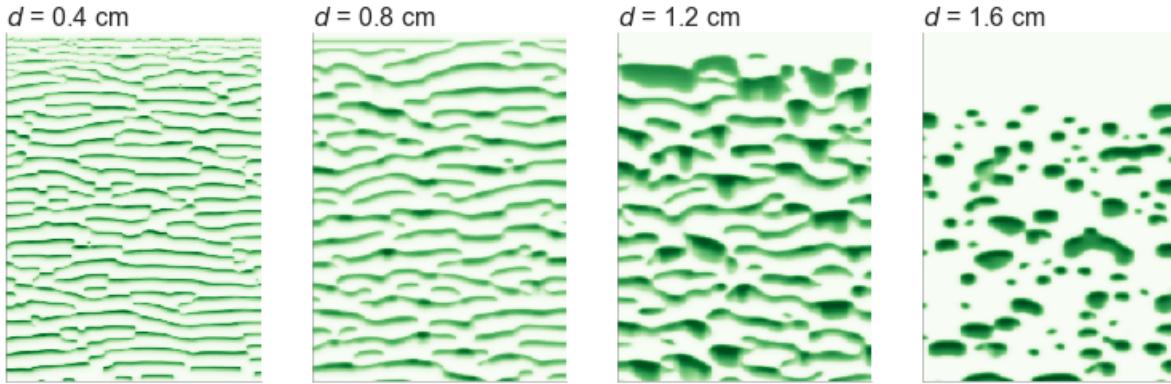


Figure 5.14: For lower annual rainfall inputs, pattern morphology changes from stripes to spots with increasing rain depth ($R = 290 \text{ mm}$, $t_{\text{storm}} = 30 \text{ min}$).

3. What correspondences can be drawn between the parameterization of rainfall processes in a phenomenological model and the rainfall processes in a model that emulates storm runoff?

To visually assess the potential of a mapping from R/G-emulator parameters to R/G parameters, scatter plots of the biomass density - wavelength ($\bar{P}-\lambda$) values generated with the R/G-emulator and R/G models are compared in Figure 5.15. There is clearly no unique mapping from R/G-emulator parameters (panel A) to R/G parameters (panel B), because the R/G-emulator patterns span a larger range of biomass densities. In panel A, multiple specifications of R , d and t_{storm} will produce comparable biomass-wavelength values, suggesting that including additional landscape features may be necessary to produce a unique mapping from R/G to R/G-emulator parameters.

The range of R/G parameters represented by Figure 5.15B is large, but I have not confirmed that it comprehensively spans the pattern forming domain of either of the R/G or the R/G-emulator models. Consequently, the existence or lack of existence of a mapping between R/G and R/G-emulator cannot be definitively determined at this stage. This lack of clarity reflects the early stage of development of both the R/G and the R/G-emulator model. While the establishment of the full domain of stable pattern forming solutions in each model as a function of their input parameters is an obvious next step, mapping this domain is highly non-trivial [Siteur et al., 2014, 2016]. As such, I have elected to treat it as lying beyond the scope of this initial study. With these caveats, the lack of overlap between the pattern properties of the R/G model and those of the R/G-emulator model is nonetheless striking, and is suggestive of non-trivial changes in modeled pattern dynamics associated with a more comprehensive representation of within-storm processes.

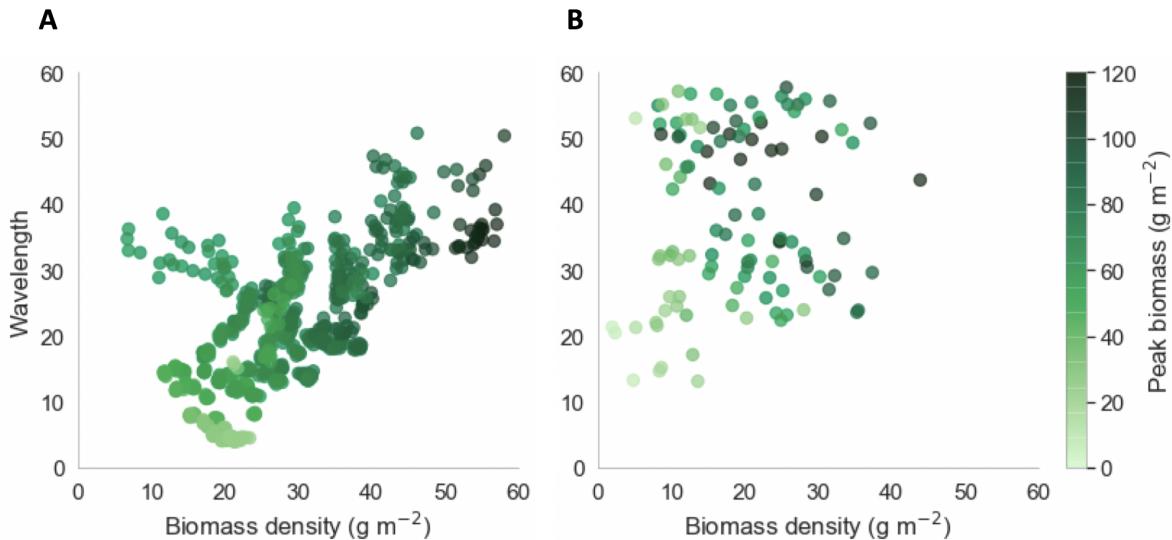


Figure 5.15: Scatter plots of biomass density and wavelength for (A) the R/G-emulator and (B) the Rietkerk-Guttal models, where marker colors show the peak biomass. Peak biomass is better correlated with biomass density in the R/G-emulator model than the original R/G model.

5.4 Conclusions

This chapter established that machine learning provides a viable pathway towards including greater process complexity within models that exhibit a strong timescales separation between interdependent processes. By replacing a phenomenological representation of overland flow with a machine learning-emulated representation of a surface hydrological model, I was able to introduce within-storm processes to vegetation pattern forming systems without compromising model runtime. The properties of the patterns generated by the full hydrological model and those generated by the ML emulator were similar in terms of wavelength and biomass distribution on the landscape. Minor differences did arise, which are attributed to the nonlinearity of the pattern forming model, and the potential for small errors in the ML emulator to be amplified by these nonlinearities. Despite these differences, the patterns produced by the ML emulator were similar enough to those produced by the physical model to enable the exploration of a large parameter space of storm climatologies. For the first time, I was able to demonstrate the sensitivity of pattern morphology to storm properties. Specifically, increased storm depths and reduced storm durations were associated with longer wavelengths, higher biomass density and higher peak biomass. In dry climates (i.e. those with lower annual rainfall totals), the pattern morphology was also sensitive to increased intensity, fragmenting from bands to spots. No similar behavior can be produced with the R/G model in the absence of a representation of storm properties. This observation raises

the possibility of an alternative path to desertification, where for fixed R , increasing storm intensity causes more water to be lost to runoff, leaving plants unable to use this water. This path to desertification could be important given that in many locations, climate change is projected to increase storminess [Knapp et al., 2008, Wang et al., 2012]. Future work will further explore this phenomenon by refining the storm parameter space, including simulations for lower R values.

The focus of this chapter was on a more explicit representation of storm properties; however, the same methods could be applied to characterize the pattern sensitivity to the hillslope gradient or soil infiltration capacity, although this would require training additional random forest regressors. Because $K_{S,V} : K_{S,B}$ and S_o have direct analogues in the R/G model (W_o and ν , respectively), exploring hillslope and infiltration controls on morphology would be an interesting avenue for future research. There remain many challenges to making quantitative predictions about real world patterns with the modified R/G model, because this model still has many phenomenological aspects, including the representation of soil water and biomass evolution. Replacing these phenomenological descriptions with more process-based equations, however, is less a challenge due to time-scale separation, and more associated with the need for a better understanding of plant physiology in dryland systems. Nevertheless, the work presented here makes an important stride towards a more realistic representation of the processes driving vegetation pattern formation.

Chapter 6

A Web Tool for Vulnerability Assessment and Restoration Planning.

6.1 Introduction

Land degradation, defined as the deterioration in the quality of land, its topsoil, vegetation, and/or water resources, is a global problem with adverse implications for biodiversity, greenhouse gas emissions, food security, poverty eradication, socio-economic stability and sustainable development [Keenelyside et al., 2012, Johnson et al., 1997, Eswaran et al., 2001]. Rather than allowing degraded lands to persist in their damaged state, recent international agreements, including the land degradation neutrality (LDN)s target adopted by the parties to the United Nations Convention to Combat Desertification [Lal, 2004, Bestelmeyer et al., 2013], advocate for the restoration of degraded lands. Current interpretations of restoration are more nuanced than a literal objective of returning landscapes and ecosystems to some pre-degraded, ‘natural state’ [Palmer et al., 2016, Suding et al., 2015, Hobbs et al., 2009]. Today, restoration ecologists recognize that such a return is likely impossible - whether due to uncertainty in natural recovery processes, the background of ongoing environmental change or irreversible physical or biotic modifications (e.g. urbanization) [Chazdon, 2014, Higgs et al., 2014, Hobbs et al., 2009, Suding et al., 2015]. Contemporary restoration efforts embrace a holistic approach which recognizes that restoration involves tradeoffs between multiple goals, including ecological integrity, historical fidelity, and societal needs [Suding, 2011, Palmer et al., 2016, Higgs et al., 2014]. To navigate these tradeoffs, Suding et al. [2015] proposed four principles for ecological restoration: (i) ecological integrity, which describes the ability of an ecosystem to support and maintain ecological processes, (ii) self-sustainability, which requires restoration efforts to establish systems that are resilient in the long term, (iii) historical perspective, which can provide a reference for determining feasible restoration goals, even where returning to past ecological conditions is impossible, and (iv) cultural values, which recognize the importance of ecosystems to people through the provision of ecosystem services, such as improved water quality, drought and flood buffering, genetic

diversity, and carbon sequestration.

Although land degradation impacts all terrestrial biomes [Blaikie and Brookfield, 2015, Keenelyside et al., 2012, Le et al., 2016], dryland degradation, often known as desertification, is particularly widespread and damaging. The degradation of drylands has profound social, environmental and economic consequences [Reynolds et al., 2007], including increased soil erosion and flooding, the generation of as much as 25% of global greenhouse gas emissions, forced migration of rural populations, health risks from dust pollution [Bashan et al., 2012], and loss of agricultural productivity, with implications for global food security [Lal, 2004]. Desertification has a long history [Barker, 2002], and consequently there have been many attempts to restore degraded drylands. These efforts have not always been successful. Challenges include the maintenance of biodiversity [Veldman et al., 2015], species selection for dryland climates [Cao, 2008], and challenges associated with tradeoffs between different restoration targets [Bremer and Farley, 2010, Ciccarese et al., 2012]. Consequently, there remains considerable scientific uncertainty about best practices for implementing dryland restoration projects [Menz et al., 2013, Suding, 2011].

The holistic approach to restoration defined by Suding et al. [2015] places an emphasis on restoring ecological functions, particularly those associated with key ecological processes (or supporting ecosystem services) such as nutrient cycling, primary production and decomposition [Palmer et al., 2016]. Typically, these ecological functions are tightly coupled to the physical structure of the landscape [Suding et al., 2015, Larkin et al., 2006]. This is particularly true in drylands where rainfall is insufficient to support continuous vegetation cover, and the landscape consists of vegetated patches surrounded by bare soil areas [Schlesinger and Pilmanis, 1998]. This ‘patch-mosaic’ structure facilitates the redistribution of water and other resources, primarily by overland flow following rainfall events [Schlesinger and Pilmanis, 1998]. Bare sites generate runoff on compacted, crusted or sealed soil surfaces [Assouline, 2004, Belnap, 2006, Rietkerk et al., 2000], while vegetated locations slow water flow and enable infiltration [Thompson et al., 2010]. This ‘runoff-runon’ mechanism provides additional water, nutrients and soil to vegetation, creating ‘islands of fertility’ in the desert landscape that enable plant growth and are critical to ecosystem function and biodiversity [Schlesinger and Pilmanis, 1998, Noy-Meir, 1979, Menaut and Walker, 2001].

Given the critical role of runoff-runon resource redistribution mechanisms for drylands, it follows that efforts to restore dryland ecosystem function should necessarily incorporate the restoration of the patch-mosaic landscape structure [Palmer et al., 2016]. Several successful restoration efforts lean on this principle: for example, Tongway and Ludwig [1996] simulated natural patchiness in mulga bushland in Australia by placing brush piles along slope contours to fragment surface runoff and generate resource sinks. After three years, soil under branches had more nutrients, more soil moisture, and less extreme soil temperatures. Similar approaches, using pits, mounds or depressions to retain water have been successfully used in restoration activities in Israel, Texas and Niger [Boeken and Shachak, 1994, Whisenant, 2002, Whisenant et al., 1995]. Ideally, restored patch-mosaic landscapes should be self sustaining [cf Suding et al., 2015], and should support key ecological functions, including sustaining vegetation, controlling erosion, and mitigating flooding [Bromley et al., 1997, Li et al., 2008,

Ludwig et al., 2005], as illustrated in Figure 6.1. Of course, these are minimum, rather than complete criteria for restoration success, focused on the physical functioning of the landscape, and the constraints these will place on other restoration choices.

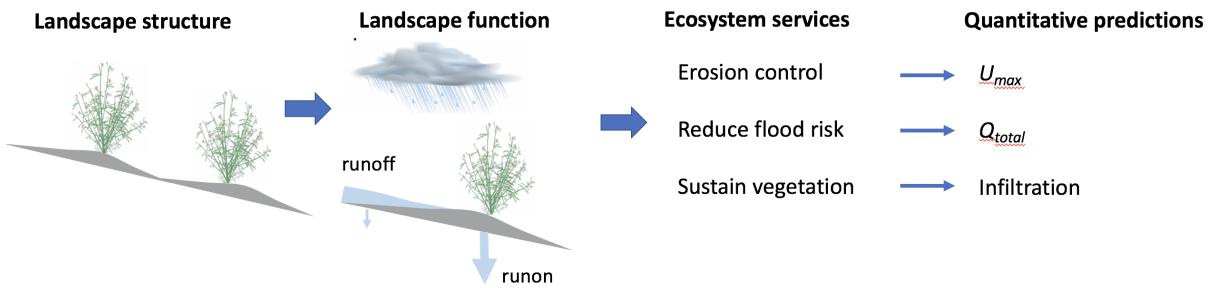


Figure 6.1: Landscape structure and ecological functions are tightly coupled in drylands, where patch-mosaic structure facilitates runoff-runon resource redistribution from bare to vegetated areas. Restoring patch-mosaic structure supports key ecosystem services, including controlling erosion, reducing flood risk, and sustaining vegetation. Because patch-mosaic functionality is derived from within-storm hydrological processes, these ecosystem services can be related to quantitative predictions about single storm events: the maximum overland flow velocity (U_{max}), the hillslope runoff volume (Q_{total}), and spatial distribution of infiltration (I) are measures of erosive power, flood risk and water supply to the vegetation, respectively.

Patch-mosaic functionality is derived from a series of fast, within-storm, hydrological processes. The design of landscape restorations that aim to restore such functions should therefore be informed by a detailed understanding of these processes. As Figure 6.1 shows, the functionality of the patch-mosaic system is determined by the outcomes of single storm events: erosion mitigation by the reduction of overland flow velocities U_{max} , reductions in flash flood risk by reductions in the hillslope hydrograph Q_{total} , and supply of sufficient water to vegetation by the spatial distribution of infiltration, I , relative to the location of plants. Designing a patch-mosaic restoration strategy thus incorporates a non-trivial hydrological problem that requires quantitative predictions about storm scale overland flow and infiltration dynamics [Cantón et al., 2011, Bracken and Croke, 2007]. While mean-field estimates of suitable planting density can be readily made based on climate and plant water requirements, these have major limitations due to the nonlinear relationships between storm, soil and slope properties, in particular, the distribution of vegetation patch sizes, and the resulting hydrological outcomes. For example, vegetation patches that are too small or sparsely distributed through space can result in a highly connected landscape which sheds significant runoff. If planting is too dense, resource redistribution may be insufficient to meet plant needs, resulting in plant mortality or growth inhibition - a mode of restoration failure that has been observed in large regions of the Loess Plateau of China [Shao et al., 2018, Feng et al., 2016, Guangquan et al., 2010]. It would be preferable, therefore, to simulate hydrolog-

ical outcomes for specific proposed planting strategies in order to evaluate these outcomes against plant water requirements and other design criteria (e.g. avoiding flow velocities high enough to promote surface erosion). Unfortunately, such simulations are complex, requiring the use of advanced, computationally intensive models to solve nonlinear coupled equations such as the Saint-Venant and Richards Equations (Chapter 2). The specialized nature of the knowledge and equipment (e.g. supercomputers) needed to use such models presents a major barrier to their adoption by restoration ecologists and landscape designers. Overcoming this barrier is a scientific challenge to improving dryland restoration practice, and forms the main goal of this chapter, which presents a web-based tool to predict hydrological outcomes at the storm scale as a function of the spatial arrangement of vegetation (equivalently, high infiltration capacity areas) within a lower infiltration capacity bare soil matrix.

I selected a web-based platform to support these predictions based on the growing interest in using such tools for the democratization of knowledge and to promote shared decision-making across diverse arenas in land stewardship, restoration and development. Such tools address, for example, rapid assessment of erosion risk [Hewett et al., 2018], optimizing the location of microhydropower infrastructure [Müller et al., 2016], and, with respect to dryland management, geographic range matching tools to support seed selection for rangeland restoration [Doherty et al., 2017]. The tool presented here is the Storm and Hillslope-scale Assessment for Dryland Ecosystems (SHADE) web app, which is intended to provide information about the spatial distribution of plant available water following storm conditions, as well as an index of the risk of surface erosion during the runoff process for user-specified landscape, vegetation and storm characteristics. Users supply a spatial distribution of vegetation (e.g. an aerial photograph or a planting design), and select soil, slope and storm parameters most pertinent to their site's landform and climate. The web app applies the machine learning emulator approach introduced in Chapter 4 to the spatial vegetation pattern and predicts the distribution of infiltration and maximum overland flow velocities for a single storm. The web app draws on a library of trained random forests (RF) spanning a broad parameter space of hillslope gradients, storm durations and intensities, soil parameters, and vegetation patterns, enabling rapid assessment of hydrological outcomes associated with specific design or restoration decisions.

This chapter firstly presents the web app, and then presents an example use-case in which restoration activities on a degraded and barren hillslope consist of as soil amendment (ripping and mulching), followed by contour planting of vegetation, using the early successional Australian species *Acacia saligna* as an example. As illustrated in Figure 6.2 key design considerations in this instance include the width and spacing of the amended contour bands. If bands are too widely spaced, overland flow velocities may increase to the point of risking surface erosion. If bands are too closely spaced, runoff from bare sites may not provide sufficient additional water (beyond rainfall) to support vegetation growth. The web app is used to estimate the appropriate contour spacing and to quantify the benefits of the soil amendment, considering plant water availability and erosion control as important outcomes.

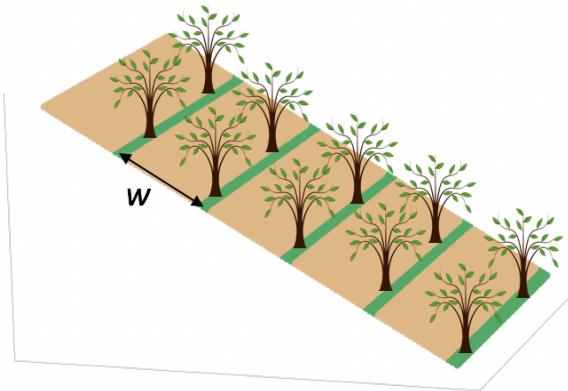


Figure 6.2: Schematic to illustrate hillslope contour planting.

6.2 Methods and Results

Implementation

The web app is written in Flask, a Python micro-framework [Ronacher, 2010], and hosted by Heroku, a cloud application platform that enables developers to deploy web applications. Data hosting for the random forest files is provided by the Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI) on their HydroShare platform. The web app can be found at: <https://hillslope.herokuapp.com/>.

The random forest regressors are trained on a collection of SVE-R model simulations, using the methods described in Chapter 4. The SVE-R model domain is restricted to planar hillslopes with dimensions 100×50 m and 1 m resolution. The landscape has a patch-mosaic structure, where each grid cell is classified as vegetated or bare. The RF regressors are designed to extrapolate to new vegetation patterns and hillslope dimensions, but not to interpolate between storm characteristics, hillslope gradients or soil infiltration capacities. Thus, the specification of the SVE-R model parameter space required to train the RF regressors is an important component of the web app design, as described in the following section.

Once parameters were specified, the SVE-R simulations were run on the Berkeley supercomputer Savio in batches of 80 simulations. The SVE-R results were used to train separate RF regressors for I and U_{max} for each combination of landscape and storm characteristics in the SVE-R simulations, as outlined in Chapter 4. The web app accesses the trained RF regressors directly, based on a user specified set of landscape and storm parameters ($K_{S,V}$, p , t_{storm} and hillslope gradient), and uses them to predict I and U_{max} .

Table 6.1: Parameter space of the SVE-R simulations used to train the RF regressors.

Parameter	Description	Units	Values
S_o	Slope gradient	%	0.5, 2.0, 10, 30
ϕ_V	Vegetation density	—	0.2, 0.4, 0.6, 0.8
$K_{S,V}$	Saturated hydraulic conductivity (vegetated sites)	cm/hr	2.5, 5.0, 7.5, 10.0
t_{storm}	Storm duration	hr	0.5, 1, 3, 6
p	Rainfall intensity	cm/hr	(see Table 6.2)
n_B	Manning's n (bare soil)	s m ^{-1/3}	0.03
n_B	Manning's n (vegetated sites)	s m ^{-1/3}	0.1
L_x	Hillslope length	m	100
L_y	Hillslope width	m	50
dx	Grid resolution	m	1.0

Parameter space design

The parameter space consists of a representative range of storms, hillslope gradients, and soil parameters as summarized in Tables 6.1 and 6.2. Storm durations include 0.5, 1, 3 and 6 hours, with the longest duration primarily restricted by the computational feasibility of running the SVE-R model. A range of rainfall intensities were obtained from the Australian Bureau of Meteorology Intensity-Frequency-Duration (IFD) curves [Ball et al., 2016], based on data from two seasonally rainy, dryland cities: Perth and Darwin. The lowest intensity storm corresponds to a 1-year storm in Perth, and the highest intensity to a 5-year storm in Darwin. I did not model low intensity storms that would not produce significant runoff on natural surfaces; equivalently, I did not model truly extreme storms, which are too infrequent to be a reasonable basis for estimating plant water availability.

Soil properties for the vegetated patches were set to span a range of typical saturated hydraulic conductivity values for clay loam to sandy loam soils [Carsel and Parrish, 1988]. I neglected sandy soils, where infiltration rates are likely to be so high that little runoff is generated, and I neglected heavy clays, which conversely have such low infiltration capacities that limited infiltration would occur. The water retention curve was described using the van Genuchten model, using representative parameters measured on a loamy desert soil by Chen et al. [2013]. Rather than specifying bare soil infiltration capacities independently of the vegetated sites, I specified a ratio of vegetated : bare soil infiltration capacities, $K_{S,V}:K_{S,B}$, which I set to 100:1. Although a wide range of values for this ratio is reported in the literature [Valentin et al., 1999, Chen et al., 2013, Thompson et al., 2010], this 100:1 ratio is consistent with reported measurements for strongly sealed or crusted soils [Assouline, 2004].

The parameter space covered by the web app thus consists of factorial combinations of 4 soils, 4 hillslope gradients, 4 rainfall durations, with 4 unique intensities per duration. This corresponds to 256 unique hillslope/storm parameter cases. As described in Chapter 4, the SVE-R model simulations include 20 vegetation patterns per case, for a total of 5120 total

Table 6.2: Storm duration - intensity combinations. Intensities are drawn from the Australian Bureau of Meteorology IDF curves, and range from a 1 year Perth storm to a 5 year Darwin storm.

Rain duration (hr)	Rain intensity (cm/hr)			
0.5	2.5	5.0	7.5	10
1.0	1.7	3.4	5.1	6.8
3.0	0.8	1.6	2.4	3.2
6.0	0.5	1.0	1.5	2.0

simulations.

User interface

Figure 6.3 shows a screenshot of the homepage, where the user is prompted to upload an image file of the vegetation, such as an aerial photograph or a design image. To convert the image to binary information about vegetation presence/absence as required by the random forest model, the web app converts the image to grayscale, and prompts the user to specify a threshold value to distinguish vegetated and bare sites. The user must also rotate the image to align the top of the frame with the top of the hillslope. Drop-down menus are then used to select storm intensity, storm duration, hillslope gradient, and saturated hydraulic conductivity. At this point, the user can ‘Run the random forest’ via a mouse click, and the web app returns the random forest predictions.

Figure 6.4 shows a screenshot of random forest predictions corresponding to the vegetation patterns in Figure 6.3, which are returned on the same web page and below the input data. The predictions include the infiltration map (left) and maximum overland flow velocities (right), and summary statistics for both outputs. The random forest predictions are the hydrological outcomes from a single storm event, and interpretation is required to provide detailed insights into plant water availability and erosion risk. Such interpretation is landscape, climate and species specific, and at this point in time is not included in the web app. I illustrate in the next section how such interpretation can be applied in combination with site-specific details to inform, in this case, the design of a re-vegetation project.

Contour planting example

To set up the use case, I considered restoration of a bare slope in an arid region that experiences a climate comparable to the Negev Desert in Israel, which receives an annual rainfall of 200 mm/year, distributed in eight 30 minute storms with $p = 5 \text{ cm/hr}$. I assumed a monocultural revegetation project with *Acacia saligna* (Labill.) H.L.Wendl (Golden Wreath Wattle, koojong, orange wattle), a widely used species for revegetation and agroforestry due to its early successional growth, deep root system and drought resistance [Nativ et al.,

Welcome to the SHADE App!

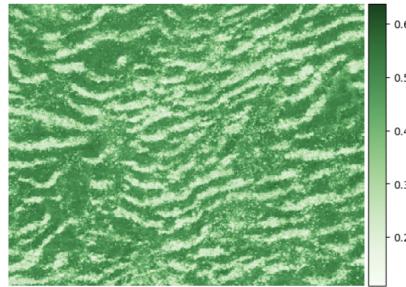
The Storm and Hillslope Assessment for Dryland Ecosystems (SHADE) app is a tool to assess plant water availability and erosion risk using storm-scale hydrological predictions.

For more information: [Home](#) [Instructions](#)

Landscape features added

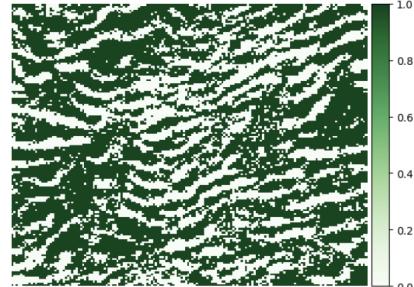
1. Upload vegetation map

veg file selected : *WAsmall.png* [Change file](#)



2. Binary image

Threshold: 0.40 Pixels per meter: 3.00 [Update](#)



Hillslope dimensions = 166m x 150m
Vegetation density = 0.62

3. Input storm and landscape features

Select a typical storm duration and update to view available storm intensities:

Typical storm duration : hr [Update duration](#)
 Typical storm intensity : cm/hr
 Ksat : cm/hr
 slope : %
[Update features](#)

4. Feature summary

Storm duration = 0.5 hr
 Intensity = 2.5 cm/hr
 Rain depth = 1.25 cm
 Ksat = 2.5 cm/hr
 Slope = 0.5%

[Run random forest model](#) or [Reset form](#)

Figure 6.3: Screenshot of the web app homepage, showing a user uploaded vegetation map (left) and the binarized version. Three additional steps are required before the image is input to the random forest model: the user is asked to (1) rotate the image that the top of the frame coincides with the top of the hillslope, (2) adjust the grid size, so that one pixel represents 1 m, and (3) adjust the threshold value distinguishing vegetated from bare soil areas.

1999]. The annual water requirements of *Acacia saligna* exceed the rainfall at the site, such that the patch-mosaic function of a desert landscape will be needed to enable a sustainable restoration design. I consider a 10% slope gradient, with loess-type soils that are prone to surface crusting. To provide additional soil moisture resources prior to planting, the soil is tilled in 3m-wide contours across the hillslope, increasing the infiltration capacity and surface roughness, slowing overland flow, reducing runoff and increasing the volume of water that infiltrates in the root zone. Tilling is assumed to increase the infiltration capacity of the soil from nearly impermeable ($K_S = 0.1$ cm/hr) to relatively high ($K_S = 10$ cm/hr), and to increase the surface roughness from $n = 0.03$ to $n = 0.1$. Seedlings are to be planted in the

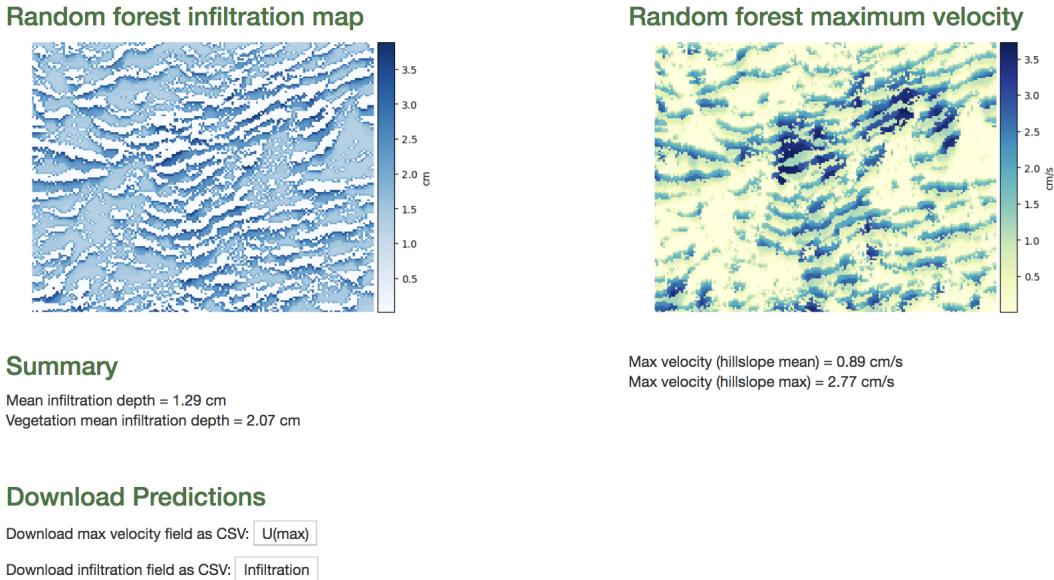


Figure 6.4: Screenshot of the web app results, showing the random forest predicted infiltration depth (left) and maximum overland flow velocity (right). Summary statistics are displayed below.

contours with 4 m along-contour spacing, with the assumption that the adult canopies will span 4 m. The question to be resolved is: what the optimal spacing of these contours is in order to meet plant water availability while preventing erosion?

Acacia saligna water requirement estimates were obtained from Nativ et al. [1999], who reported a transpiration demand of 111.2 L for 6 month old seedlings during July-October in the Negev desert, Israel. Assuming that (i) seedling leaf area was approximately 1 m², (ii) seedlings have a root zone that extends over 1 m², and (iii) July-October water demand can be scaled to the calendar year using annual potential evaporation data [Kidron, 2009], these measurements imply an annual water requirements for seedlings of 0.25 m³ per m² ground area. I assumed that seedlings and adult trees have the same water requirement per leaf area, but that adult tree canopies are closer to 25 m², with along-contour root extent limited to 2 m, due to the 4 m spacing between plants [Zegada-Lizarazu et al., 2007]. Assuming that infiltration in the bare soil areas is negligible compared to the contour area, the infiltration zone for each plant is 12 m². With these considerations, 520 mm would need to infiltrate on average within the contour to meet the water requirements of the plants.

Erosion risk for the landscape was computed by estimating the critical bed shear stress for rill erosion, τ_c from tabulated values for loess soils at 3 Pa [Yao et al., 2008, Gilley et al., 1993]. This value is at the lower end of the range reported for rill incision in bare topsoils (i.e. 1.8 - 10.6 Pa [Poesen et al., 2003]), providing a conservative estimate of a 'safe' U_{max} . The bed shear stress is estimated from the flow depth by assuming kinematic conditions

$(S_o = S_f$, see Chapter 3):

$$\tau = \rho gh S_o \quad (6.1)$$

With roughness specified by Manning's equation, flow velocity and depth are related as:

$$U = \frac{1}{n} h^{2/3} S_o^{1/2} \quad (6.2)$$

Rearranging, the critical flow velocity, U_c , can be related to the critical shear stress as:

$$U = \frac{1}{n} \left(\frac{\tau_c}{\rho g S_o} \right)^{2/3} S_o^{1/2} \quad (6.3)$$

For $\tau_c = 3$ Pa, the estimated critical flow velocity is 29 cm/s.

To evaluate the potential contour designs, I varied contour spacing from 5 to 30 m, and computed the mean annual infiltration depth in vegetated areas and the maximum flow velocities in the domain. The dependence of these outputs on contour spacing is shown in Figure 6.5. The results suggest that the water requirements of the plants could be met with a contour spacing of 20 m without the peak overland flow velocities exceeding the critical threshold. These results were applicable to the design case of a 30 minute, 5 cm/hr storm.

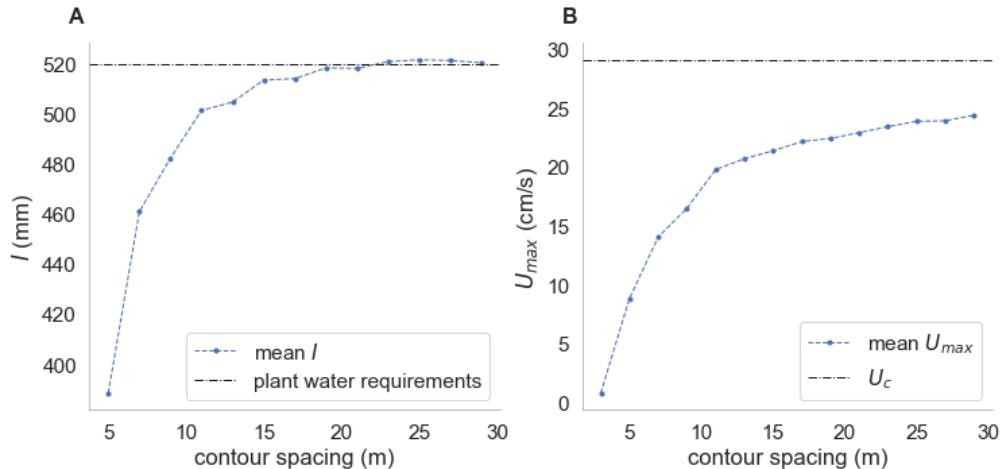


Figure 6.5: (A) Annual infiltration depth in vegetated areas as a function of contour spacing. The dashed horizontal line shows the estimated total plant water requirements per unit area. (B) The estimated maximum flow velocity as a function of the spacing between contours.

6.3 Discussion

The web app presented here offers a way to rapidly test the hydrological implications of a design of vegetation planting at hillslope scales. As illustrated above, this can be used

to inform planting densities, species selection, fractional cover, or the distribution of soil amendments, provided that site specific information about climatology (e.g. a reasonable basis for estimating a design rain storm and relating it to annual infiltration volumes), plant requirements (e.g. rooting area and plant-specific annual water requirements), soil properties and erosion risks (e.g. a reasonable basis for estimating a critical velocity) are available.

These predictions, naturally, come with some caveats. The web app is constrained by the limitations of the SVE-R model and random forest emulators that underpin it. The SVE-R model construction assumes infiltration excess overland flow, with shallow, free draining soils and thus no saturation excess, and it ignores any subsurface redistribution of water. These assumptions are not appropriate in all environments, and may be particularly inappropriate in mesic-humid regions. The web app also assumes an initial wetness in the soil that approximates field capacity. Thus runoff predicted by the web app will be over-estimated for very dry soils, and underestimated for very wet soils. This constraint is only one of several that are introduced by the need to creating a collection of pre-computed SVE-R simulations on which to train the RF regressors, which means that the parameter space is discrete and limited. The RF regressor, as approximations of the SVE-R model, are an additional source of errors. Consequently, the web app is intended to provide estimates for the user-supplied landscape, rather than precise predictions. If a given site is not sufficiently close to one of the provided parameter cases, or high accuracy predictions are required, the web app may not be an appropriate tool. For example, the web app is not recommended for landscapes with relatively permeable bare soil areas (i.e. $K_{S,B} >> 0.1 \text{ cm/hr}$). Complex topography, channelization of flow, convergence and divergence are also omitted from the planar slopes simulated by the web app.

Beyond these practical limitations, the web app is limited in its scope to a focus on hydrology. When used for restoration planning, it should be only one component of a broader framework that incorporates ecological integrity, local species diversity, resilience, sustainability, and opportunities for community engagement. It may need to be combined with other tools or assessments that account for technological constraints on implementing any design, financial viability, and community priorities [Hobbs et al., 2009, Chazdon et al., 2018], and which provide information about a broader range of ecosystem functions and services such as carbon sequestration, biodiversity etc. [Veldman et al., 2015, Shao et al., 2018]. The web app is not a replacement for stakeholder engagement, buy-in and co-production of knowledge about landscape maintenance, which are associated with improved restoration outcomes in drylands [Reynolds et al., 2007, Hewett et al., 2018]. In its current form, the web app is a prototype intended as a starting point for further communication and exchange. The web app has not yet been tested by potential users, and would likely benefit from further iterative development with feedback from practitioners. But by providing a readily accessible tool for landscape assessment and restoration planning, the web app may facilitate knowledge-sharing and collaboration between stakeholders and the scientific community [Hewett et al., 2018].

This chapter outlined the development of a web app, which provides a spatially-resolved means to assess the hydrological outcomes of storms at the hillslope scale. Integrated into

a broader framework, this application may be used to estimate plant water availability and erosion risk associated with natural, degraded or restored landscapes. While somewhat restricted in scope to a set of idealized and prototypical storm, slope and soil characteristics, I nonetheless hope that it could assist in restoration or re-vegetation efforts, as illustrated in the contour planting example. To continue improving the accessibility of predictive tools, I will work to make the SVE-R model open source, enabling practitioners to apply the full model at their site. While this option will require significantly more coding work for the user, it will remove errors associated with the RF-emulator approach and limited parameter space, and allow a broader range of landscapes, climates and soil conditions to be explored.

Chapter 7

Conclusion

This thesis presented the development of a coupled 2D Saint Venant Equation - 1D Richards equation (SVE-R) model, and its application to simulate storm-driven overland flow and infiltration in patchily vegetated dryland ecosystems. The SVE-R model coupled two components, the finite element SVE solver developed by Bradford and Katopodes [2001] and the Richards equation solver developed by Celia et al. [1990]. The coupled model was validated against analytical solutions of the kinematic wave equations [Giraldez and Woolhiser, 1996] (see Chapter 2), demonstrating that the coupling was implemented appropriately. The resulting model provided a way to represent the generation and routing of infiltration excess overland flow on patchily vegetated landscapes on within-storm timescales.

The remainder of the thesis addressed a number of practical challenges to its implementation, including the representation of energy dissipation due to frictional resistance, the computationally intensive nature of the model, and the difficulties faced by practitioners who may wish to use this model to inform design or dryland management. Firstly, I demonstrated that there is minimal sensitivity of the hydrologically relevant outputs of the SVE-R model (for sheet flow on hillslopes) to the specification of the friction slope (see Chapter 3). In addition to suggesting that modelers can select the resistance formulation of their preference without biasing results, I showed that the results are sensitive to the parameterization of such an equation. Next, I developed a method to emulate the SVE-R model predictions using machine learning, specifically random forest (RF) regression (see Chapter 4). This approach involved training random forests regressors on collections of pre-computed SVE-R model runs to predict the spatial distributions of infiltration and maximum flow velocities. Once trained, the random forest regressors were accurate, with NRMSE scores typically ranging from 5-15% on the test sets. The machine learning approach was several orders of magnitude faster than the SVE-R model, requiring less than a second to make predictions. Because this advance enabled such rapid predictions, I was able to use it to improve the representation of within-storm processes in an ecohydrological model describing the spatial patterning of vegetation on the landscape (see Chapter 5). I was able to show for the first time that modeled spatial patterns of vegetation are sensitive to within-storm characteristics, such as depth and duration, and I identified a potential new pathway to desertification

in these landscapes associated with the intensification of individual storm events. Finally, I developed a web-based tool to assess plant water availability and erosion risk at hillslope scales (see Chapter 6) that embeds random forest regressors within a library of prescribed storm and slope conditions. With these conditions fixed, users can predict infiltration, runoff and peak flow velocities for different spatial patterns of vegetation on the landscape. A prototype of this application is now available online, and I hope to iterate on its development in the future by engaging with practitioners.

Important scope for future work with these tools remains. I hope to further validate the SVE-R modeling framework by comparison to experimental results from dryland environments, and potentially, by adding Lagrangian particle tracking to tracer experiments. I also hope to generalize the use of the RF emulation approach. A major caveat to its use at present is that users can only adjust vegetation fields while holding landscape and storm characteristics constant. Assessing the conditions under which RF regressors could interpolate between landscape and storm characteristics is outstanding, and if found to be feasible, could improve the generality of the approach. This would, for instance, increase the utility of the current web app without requiring the simulation of a larger collection of SVE-R model runs and training of additional RF regressors. Of course, the most general predictions about surface hydrology on patchy hillslopes can be made via the SVE-R model itself, and to that end I hope to publish an open-source version of the code. While an open source SVE-R model would require significantly more programming experience and coding work for the user, it would also give the user more control over the model domain and remove errors associated with the RF emulation approach. To my knowledge, there are currently no open source SVE-R models implemented in Python, and such a model would be a useful tool in many applications. For example, I am currently exploring the extent to which these modeling tools can be incorporated into urban stormwater management settings, where similar runoff behaviors occur.

Restoration and conservation of dryland ecosystems is an important and often overlooked international environmental challenge. Restoration of drylands is now a centrepiece of international attempts to address this challenge, such as the Land Degradation Neutrality pledge made by the parties to the United Nations Convention on Combating Desertification (UNCCD). By reducing the barriers - whether computational or in terms of expertise required - to the adoption of hydrological predictions for dryland environments, I hope that this thesis has contributed to the ability of restoration ecologists and land managers to meet this challenge. This thesis illustrates a spectrum of approaches, from full simulation, to ML emulation and ultimately web-based platforms, which together can be leveraged to democratize information access, and hopefully improve management decisions in some of the world's most vulnerable environments.

Appendix A

Roughness schemes

Section A.1 provides additional details about select roughness schemes from Chapter 3, including the Cylinder Array, Poggi, Transitional and depth-dependent Manning schemes. Section A.2 presents the derivation of analytical expressions relating the $\log f$ - $\log \text{Re}$ and $\log f$ - $\log S_o$ slopes to the exponents in Equation 6 in Chapter 3.

A.1 Roughness scheme details

The depth-dependent Manning scheme is based on the premise that roughness decreases with increasing flow depth:

$$n = n_0 \left(\frac{h}{h_0} \right)^{-\epsilon} \quad \text{for } h < h_0 \quad \text{and} \quad n = n_0 \quad \text{for } h \geq h_0 \quad (\text{A.1})$$

where n is the depth dependent Manning coefficient. n_0 is the value of minimum Manings roughness, which corresponds to the flow depth h_0 beyond which n is assumed constant. ϵ is an exponent related to vegetation drag, yielding an effective scaling of $m = 2/3 - \epsilon$. Conceptually, n decreases with increasing flow depth to represent reduced resistance with increasing inundation of the roughness elements. This scheme is used by Jain et al. [2004] to represent flow through vegetation in a distributed rainfallrunoff model, and adapted by Mügler et al. [2011] to model flow experiments on a non-vegetated runoff plot. Jain et al. [2004] and Mügler et al. [2011] select $\epsilon = 1/3$, producing an apparent $m=1$ scaling, which is the value we use in the main text.

The transitional scheme with $m = 1$ comes from experimental studies involving high **Re** sheet flow over surfaces with short vegetation (e.g. grasses), from which m values close to unity have been observed Brutsaert [2005]. For example, in deriving the rising hydrograph, Horton [1938] suggested that $m = 1$ might represent flow that is 75% turbulent and 25% laminar. Similarly, Wooding [1965] observed that flow over irregular surfaces may vary spatially and temporally between laminar and turbulent, owing to the roughness fluctuations.

Finally, Brutsaert [2005] notes that, even if flow is turbulent near the surface, flow within the grass stems and leaves may more closely resemble laminar flow through a porous medium.

The depth-dependent Manning and transitional schemes illustrate how similar mathematical forms of the roughness scheme may arise from different conceptual models of the land surface. These schemes have the same power law scaling ($m=1$, $\eta=1/2$), despite differing physical interpretations.

The Poggi and cylinder array schemes represent flow through vegetation, and are parameterized with canopy-specific values. The Poggi scheme is derived from canopy mixing length theory [Katul et al., 2011], with f approximately constant for low-depth flows:

$$f = 8\beta^2 \exp\left(\frac{H_c}{\beta^2 L_c}\right) \quad (\text{A.2})$$

where H_c is the canopy height, β is a the momentum absorption coefficient that varies with canopy density, and L_c is an adjustment length scale parameterizing the loss of turbulent kinetic energy from advecting eddies due to their dissipation by drag element.

The cylinder array scheme is derived from flume studies representing vegetation as an array of infinitely tall cylindrical rods [Tanino and Nepf, 2008, Wang et al., 2015]. A force balance approach, wherein the weight of the fluid is balanced by vegetation resistance to flow, is used to relate the physical characteristics of the canopy to a drag coefficient [James et al., 2004, Wang et al., 2015, Tanino and Nepf, 2008]. The roughness parameter is related to the vegetative drag coefficient C_d as:

$$\alpha = \sqrt{\frac{C_d m D}{(1 - \phi) 2g}} \quad (\text{A.3})$$

where D_{veg} is the stem diameter, ϕ is the area concentration of stems within the vegetated patches, and m is the number of vegetation stems per unit ground area ($\phi_{veg} = m\pi D_{veg}^2/4$). C_d can be experimentally determined from flume experiments. For example, Cheng and Nguyen [2010] use a large library of experiments to fit C_d as a function of Reynolds number and cylinder attributes.

A.2 Resistance sensitivity to scheme

In experimental studies, the sensitivity of f to Reynolds number or slope is frequently examined with a log-log plot. Accordingly, we assume that f can be described by a power law expression of the form:

$$f = a \mathbf{Re}^b S_o^c \quad (\text{A.4})$$

where a , b and c are constant parameters. As in Section 2.3, we consider the case of steady-state, kinematic flow on a uniformly vegetated hillslope under constant rainfall.

Combining the definition of f with the general form of the resistance equation:

$$U = \left(\frac{8gh}{f} S_o \right)^{1/2} = \frac{1}{\alpha} S_o^\eta h^m \quad (\text{A.5})$$

Next, with the combined constraints of continuity and kinematic flow, h can be determined as a function of distance x from the ridge. Combining resistance and continuity equations and solving for h :

$$h^{m+1} = \alpha q S_o^{-\eta} \quad (\text{A.6})$$

Substituting Equation A.6 into Equation A.5 and simplifying:

$$f^{m+1} = (8g)^{m+1} \alpha^3 S_o^{m+1-3\eta} q^{1-2m} \quad (\text{A.7})$$

Finally, substituting $\mathbf{Re} = q/\nu$:

$$f^{m+1} = (8g)^{m+1} \alpha^3 S_o^{m+1-3\eta} (\nu \mathbf{Re})^{1-2m} \quad (\text{A.8})$$

By comparison to Equation A.4, the resistance sensitivity to Reynolds number is related to m by:

$$b = \frac{1-2m}{1+m} \quad (\text{A.9})$$

Similarly, the f sensitivity to slope is related to m and η by:

$$c = \frac{m+1-3\eta}{1+m} \quad (\text{A.10})$$

For example, for laminar flow ($m=2$, $\eta=1$), Equations A.9 and A.10 yield the familiar $f = a \mathbf{Re}^{-1}$.

Appendix B

Predicting Velocity with Random Forest Regression

The application of the RF regression approach to the maximum overland flow velocities (U_{max}) is presented here. The approach is nearly identical to that described in Chapter 4, and presented using the same collection of SVE-R simulations (see Figure B.1 for examples of the simulated U_{max} fields, for the same cases as Figure 4.5). The principal difference is in the assessment metrics: the RMSE is normalized by the maximum overland flow velocities, because very small velocities produce very large NRMSE values otherwise. The hillslope mean U_{max} is used in place of the infiltration fraction as a hillslope-scale assessment metric.

The same features were used to predict U_{max} and I , as summarized in Table 4.2. Several examples of the features engineered for bare soils are displayed in Figure B.2, analogous to Figure 4.7.

As in Chapter 4, visual inspection was used to assess the RF performance (see Figure B.3). RF performance was further assessed as a function of patch length scale (Figure B.4), and rain depth and duration (Figure B.5). The performance is slightly worse for the test set ($\sigma = 2$), as expected. The accuracy tends to improve with increasing σ (increasing patch size), possibly because the spatial correlation of the simulated flow fields also increases, making the emulation problem easier (fewer degrees of freedom). The NRMSE is relatively insensitive to rain depth, but increases slightly with increasing duration. The increase in NRMSE with increasing duration may reflect the construction of the storm parameter space such that increasing storm duration corresponds to decreasing intensity, less runoff production, and thus less spatial coherence in the overland flow field.

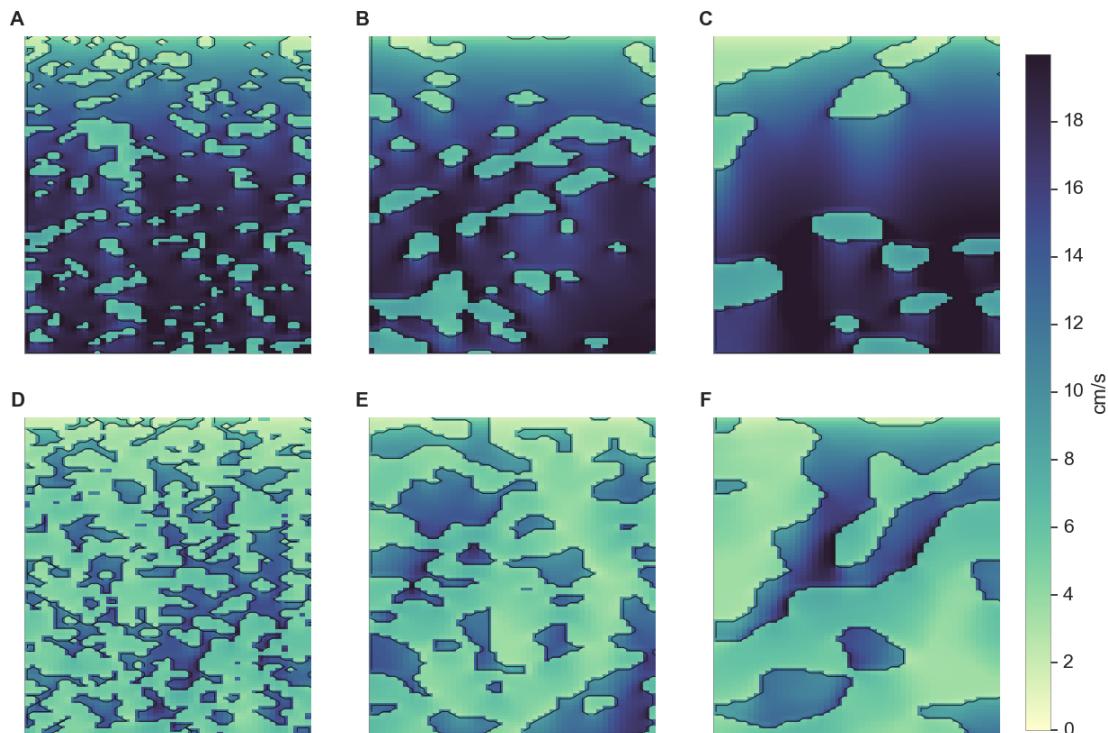


Figure B.1: Maximum velocity maps for a 20 minute storm with $p = 9.6 \text{ cm/hr}$, $S_o = 2\%$, $K_{S,V} = 1.5 \text{ cm/hr}$ and $K_{S,B} = 0.15 \text{ cm/hr}$, using the vegetation maps from Figure 4.3.

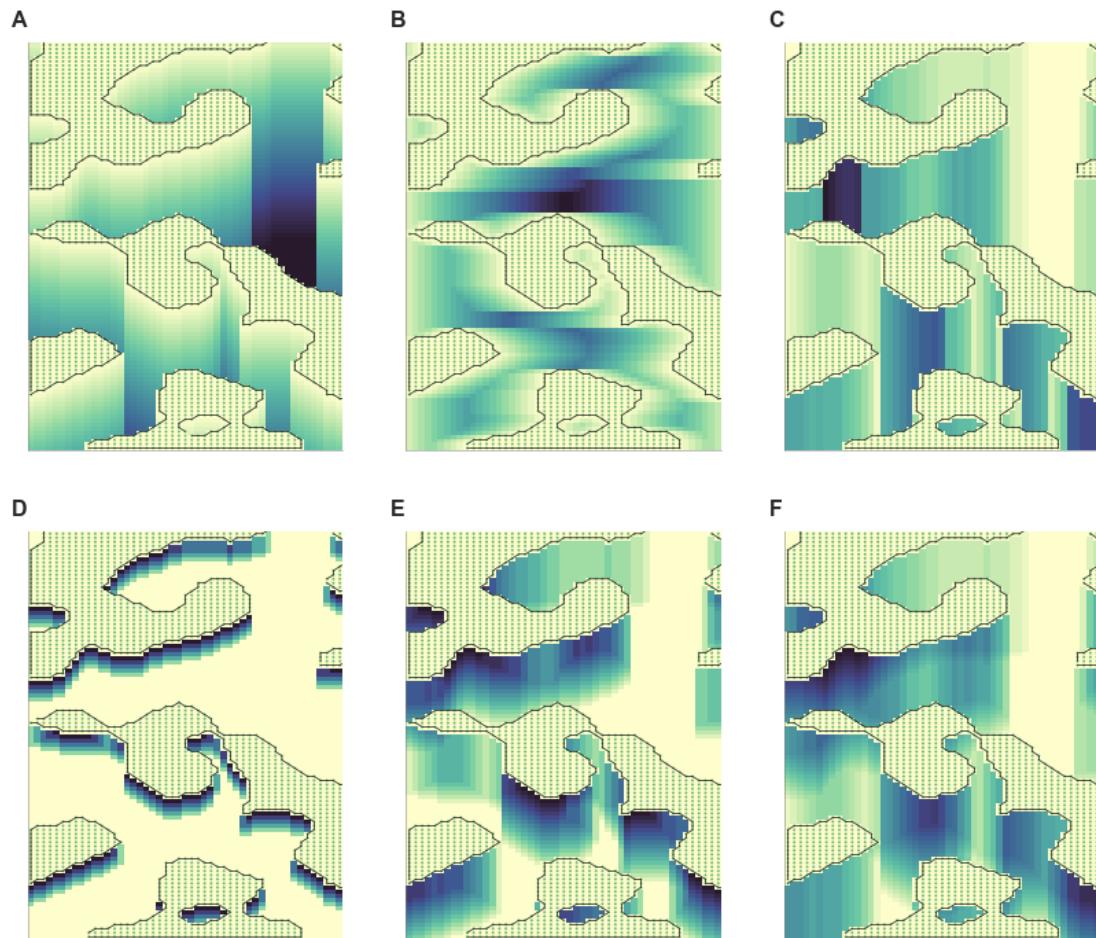


Figure B.2: Examples of RF features used to make predictions for bare soil areas: (A) distance to the nearest upslope vegetated cell, (B) across-slope distance to the nearest vegetated cell, (C) the length of the upslope vegetated patch. Panels D-F show the upslope vegetation fraction computed with window lengths 12, 20 and 28, respectively (values in vegetated areas are set to 0 in panels D-F for clarity). Scattered green points indicated vegetated cells.

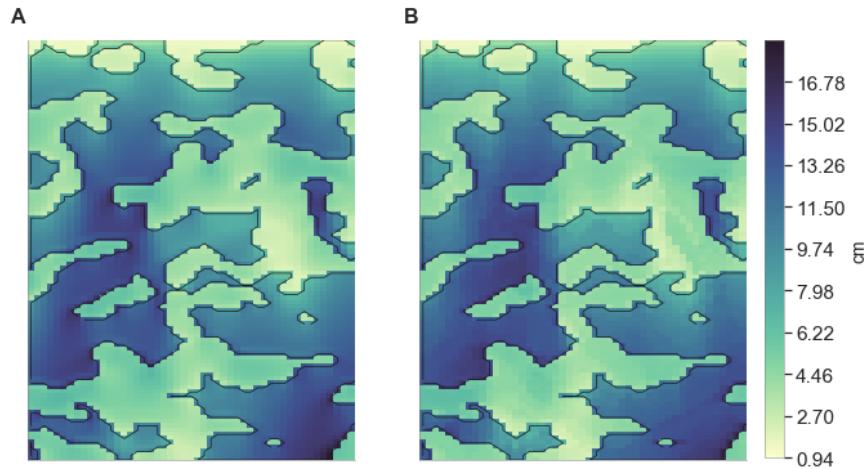


Figure B.3: Example comparison between the SVE-R and RF predictions of U_{max} for a hillslope from withheld set (vegetation density $\phi_V = 0.4$, $\sigma = 2$, $t_r = 10$ min and $d = 1.2$ cm). The SVE-R simulated U_{max} (panel A) and the corresponding RF prediction (panel B) are visually very similar.

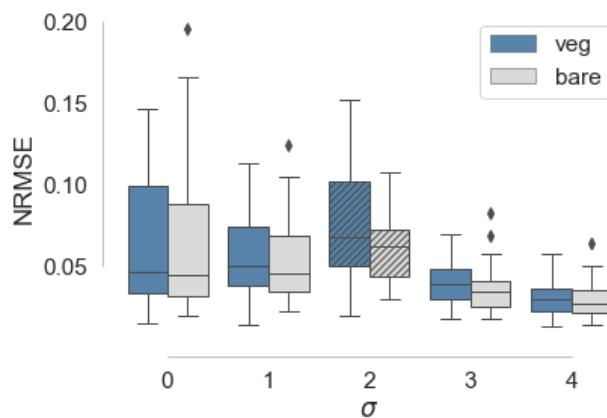


Figure B.4: A box plot summarizing the NRMSE, grouped by patch length-scale (σ), with hatches to indicate the test-set ($\sigma = 2$). The NRMSE is generally higher for the test set, and decreases with increasing σ .

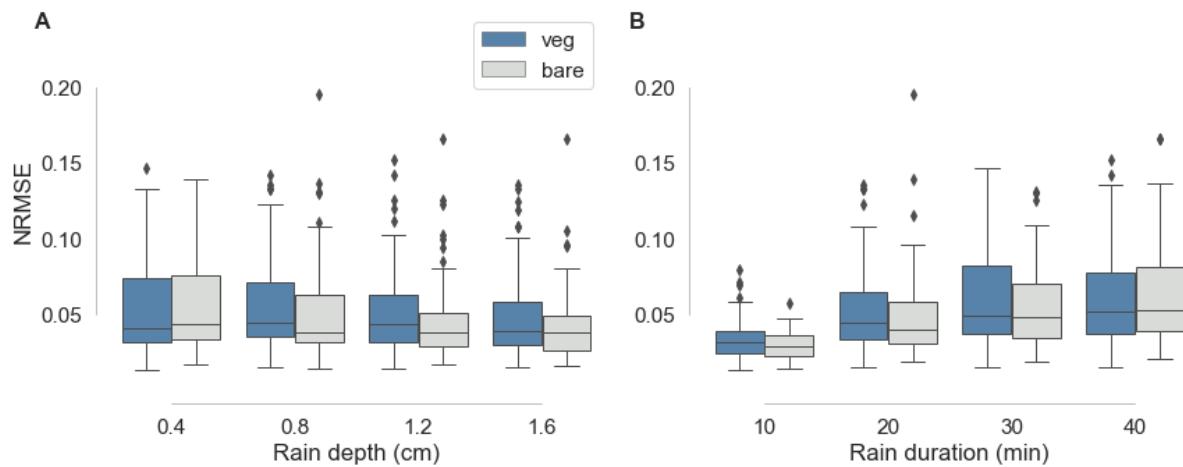


Figure B.5: NRMSE grouped by rain depth (panel A) and duration (panel B). The NRMSE is insensitive to rain depth, and increases slightly with increasing rain duration.

Appendix C

Random Forest Methods

This appendix describes the approach used to select the feature parameters (the window size λ and the smoothing kernel ζ) and the random forest parameters (max_depth and min_sample_split). Rather than optimize the RF performance, the parameter selection process qualitatively considered several factors: (i) predictive accuracy, (ii) the memory usage and (iii) the time required to fit the RF regressors to the SVE-R model output.

Two representative subsets of SVE-R simulations were selected for assessing the parameters, each consisting of 80 simulations. The first is a subset of the simulations detailed in Chapter 5, consisting of all storms with $d = 0.4$ cm, and referred to here as the ‘pattern simulations’. These simulations were also introduced in Chapter 4 to describe the RF emulator methods, and the parameter values are given in Table C.1, case A. The second subset of SVE-R simulations consists of the 30 minute storms on hillslopes with $S_o = 10\%$, $K_{S,V} = 10$ cm/hr from the Chapter 6 simulations, referred to as the ‘web-app simulations’ (see Table C.1, case B). The storm characteristics, soil parameters and slopes are sufficiently different between the pattern and web-app simulations that they are considered to span a range of landscapes and storms, and thus be appropriate for parameter testing.

Random forest performance was assessed with the NRMSE of the test set (consisting of all vegetation patterns with $\sigma=2$, as defined in Chapter 4). The NRMSE of predictions made on vegetated cells, NRMSE(veg), was used to quantify the RF performance at predicting infiltration, because vegetated areas are more relevant than bare soil areas to plant water availability and the hillslope runoff coefficient. Similarly, the RF performance for U_{max} , was assessed on the NRMSE for bare soil areas, NRMSE(bare), because overland flow velocities tend to be greater in bare soil areas and therefore more relevant to erosion risk assessment. The NRMSE(veg) is computed as the mean RMSE(veg) of the hillslopes in the test set, divided by the mean SVE-R predictions, and similarly for NRMSE(bare). In assessing the results, I examined whether adding more features or increasing tree size resulted in overfitting, which would be observed as worsening performance on the test set with increasing model complexity.

In addition to random forest predictive performance, I also evaluated the training time and the storage size of the RF regressors. This allows for assessment of the relative trade-

Table C.1: Parameters for the two subsets of SVE-R simulation used to select the feature parameters (ζ and λ) and RF parameters (max_depth and min_sample_split). Where Case A and Case B simulations differ, duplicate parameters are included for clarity.

Parameter	Description	Units	Values
<i>(Case A) Subset from the pattern (Chapter 5) simulations</i>			
S_o	Hillslope gradient	%	2.0
$K_{S,V}$	Hydraulic conductivity (vegetated sites)	cm/hr	1.5
$K_{S,B}$	Hydraulic conductivity (bare soil)	cm/hr	0.15
d	Storm depth	cm	0.4
p	Rainfall intensity	cm/hr	0.6, 0.8, 1.2, 2.4
L_x, L_y	Domain size	m	200×100
dx	Grid resolution	m	2.0
<i>(Case B) Subset from the web-app (Chapter 6) simulations</i>			
S_o	Hillslope gradient	%	10
$K_{S,V}$	Hydraulic conductivity (vegetated sites)	cm/hr	10
$K_{S,B}$	Hydraulic conductivity (bare soil)	cm/hr	0.1
t_{storm}	Storm duration	min	30
p	Rainfall intensity	cm/hr	2.5, 5, 7.5, 10
L_x, L_y	Domain size	m	100×50
dx	Grid resolution	m	1.0
<i>Common parameters</i>			
n_B	Manning's n (bare soil)	$s m^{-1/3}$	0.03
n_B	Manning's n (vegetated sites)	$s m^{-1/3}$	0.1
ϕ_V	Vegetation fraction	—	0.2, 0.4, 0.6, 0.8
σ	Patch length scale	—	0, 1, 2, 3, 4

offs between the RF performance and the computational demands associated with the RF regression. Since the use of RF regression is intended to speed up predictions and analysis, these performance metrics are pertinent to the specification of the regressors.

C.1 Feature Parameter Selection

Obtaining features from the input vegetation patterns involves a number of spatial operators, which decompose the 2D vegetation field into a cell-level, ‘neighborhood’ representation of each grid cell, as described in Chapter 4. The feature parameters, λ and ζ , are used by the spatial operators. Specifically, λ is the window length used to compute the upslope vegetation fractions, and each value of λ corresponds to a unique feature $\phi_{V,\lambda}$. ζ describes the length-scale of Gaussian smoothing, which is applied to all features (including the $\phi_{V,\lambda}$ features). Visual examples of λ can be seen in Figure 4.7 for vegetated cells and Figure B.2 for bare soil cells, and an example of ζ is given in Figure 4.8.

Table C.2: Window lengths used to compute average upslope vegetation fraction. Window lengths are specified as the number of grid cells.

Case abbreviation	Window lengths
L1	—
L2	4, 12, 20, 28
L3	4, 12, 20, 28, 36, 44, 52
L4	3, 4, 5, 6, 10, 14, 18, 22, 30, 38, 46, 54

To select the upslope window lengths, four sets of λ s were tested, as listed in Table C.2. These cases range from no values of λ (i.e. omitting upslope vegetation fraction) to 12 values of λ . In case L4, the window lengths are closely spaced for small upslope distances, with decreasing resolution for larger upslope lengths. Cases L2 and L3 have the same uniform distribution of window lengths, but case L3 includes larger λ values. There are, of course, many combinations of windowing lengths that could be tested, and the primary objective was to establish the sensitivity of the RF performance to the specification of the upslope windowing lengths.

Figure C.1 presents the RF training time versus NRMSE for cases L1 - L4, for the pattern (panel A) and web-app (panel B) simulations. Increasing the number of window lengths increases the training time, with small improvements in performance. For the pattern simulations (Panel A), NRMSE decreases from L1 to L2 for infiltration and U_{max} , but does not further decrease with more window lengths. For the web app simulations (Panel B), the infiltration predictions are not improved by including upslope vegetation density (left). There is a small decrease in U_{max} NRMSE from L2 to L3 (suggesting that, for the steeper slopes, longer length scales are relevant to predicting U_{max} in bare soil areas). The NRMSE does not increase with the addition of more window lengths in any of these cases, suggesting that the multicollinearity introduced by adding correlated features (i.e. upslope vegetation fractions computed at multiple window lengths) does not lead to overfitting. The marker sizes show the RF memory usage (Mb), and the similarity amongst the marker sizes in all cases indicates that the memory usage is insensitive to the number of window lengths.

A range of feature smoothing options was tested, including: $\zeta = 0, 1, 2, 3$, [1,2], and [1,2,3], where values in brackets indicate smoothing with multiple length scales, and 0 indicates that no feature-smoothing was applied. Figure C.2 shows the NRMSE versus training time for each option, with marker sizes scaled to show memory use. In most cases, introducing some smoothing ($\zeta = 0$ vs. $\zeta=1,2$ or 3) improves predictor accuracy, with small increases in training time (the exception is U_{max} in the web-app simulations). Introducing more than one ζ does not improve accuracy, and increases training time significantly. On the basis of these results, limiting the smoothing length scale to one value ($\zeta = 2$ or 3) was determined to be the best choice.

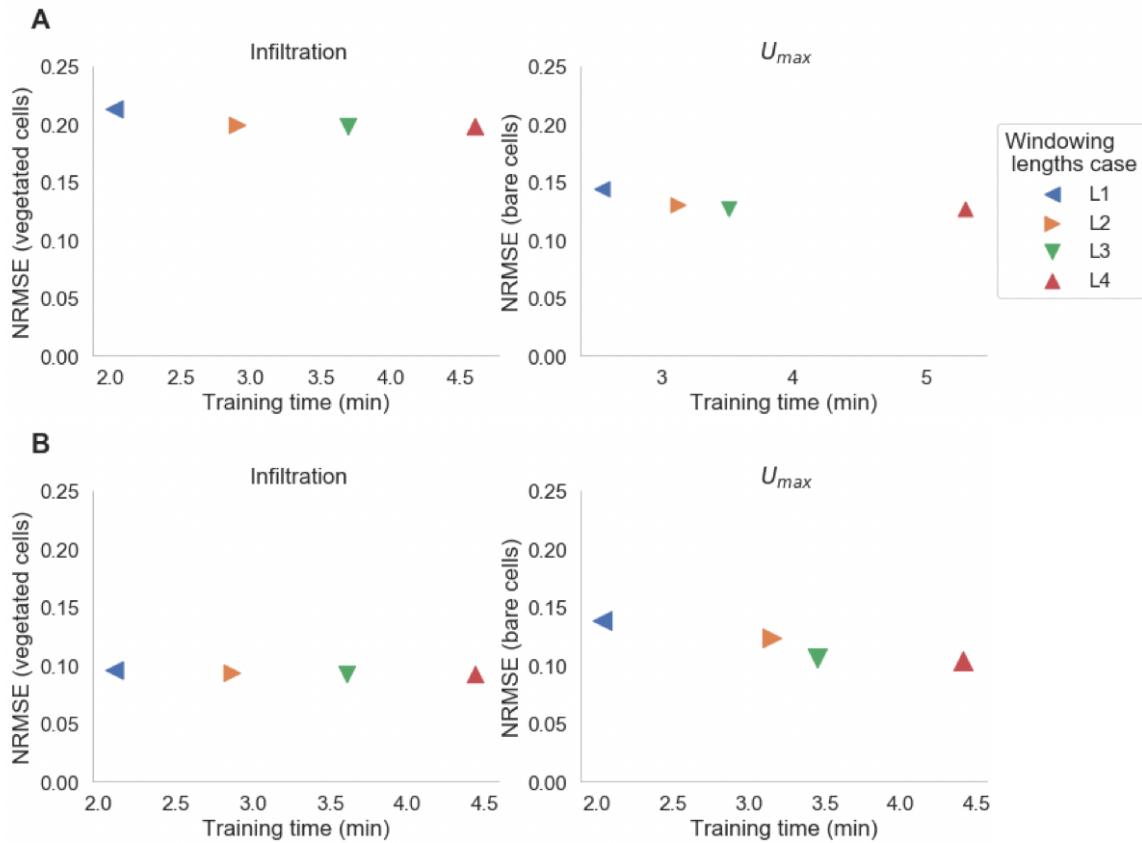


Figure C.1: NRMSE versus training time for the pattern (panel A) and web-app (panel B) simulations. SVE-R simulations, with the upslope windowing lengths case indicated by marker color. Marker sizes are proportional to the memory use, suggesting low sensitivity to the number of window lengths. The RF regressors were trained with $\zeta = 2$, `max_depth` = 15 and `min_sample_split` = 20.

C.2 Random Forest Parameters

In fitting the random forest regressors, the optimal size of the individual decision trees is an important factor. Trees that are too large risk overfitting the training data and poorly generalizing to new samples, while trees that are too small may not capture important structural information about the sample space [Hauskrecht, 2003]. Random forests are made up of a large number of decision trees, where every node in a decision tree is conditioned on a single feature, which splits the data so that similar values are grouped in the same subsets [Shih, 2012]. To grow a decision tree, each split condition is chosen to reduce the impurity measure (typically the variance for regression trees), which quantifies how well the two classes are separated. In sci-kit learn, the default values for the parameters controlling

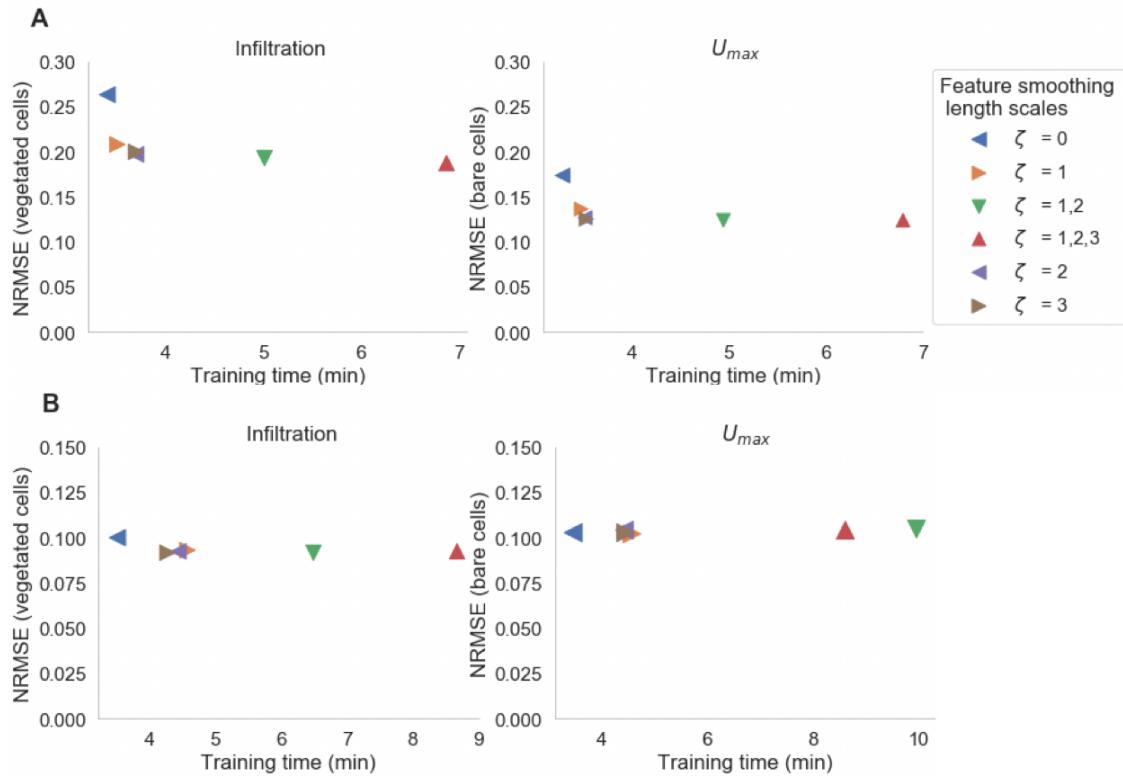


Figure C.2: The sensitivity of the NRMSE and training time to ζ for the pattern (panel A) and web-app (panel B) simulations. In panel A, smoothing with $\zeta = 2$ or $\zeta = 3$ results in the best performance, with respect to NRMSE and training time. Smoothing does not significantly improve the RF predictions in the web-app case (panel B), most likely because the steeper slope results in flow more closely aligned along the slope gradient. In both case, smoothing at multiple length scales does not improve performance. Marker size shows file size, indicating low sensitivity to the smoothing length scale.

the size of the trees (e.g. `max_depth`, `min_samples_split`) result in fully grown and unpruned trees, where each leaf node corresponds to the lowest impurity. The default parameters can result in overfitting and potentially very large trees on some data sets, so controlling the complexity and size of the trees is needed to obtain good results with respect to both predictive accuracy and memory consumption.

Preliminary tests indicated that the accuracy and memory usage were sensitive to parameters controlling the tree size, and that the best performance would be obtained by specifying these parameters to limit tree growth. While sci-kit learn provides built-in parameter tuning functionality (e.g. `GridSearchCV`), this functionality does not allow for assessment across multiple criteria (i.e. RF memory usage, RF training time). Consequently, following an initial exploratory analysis, a small parameter grid, consisting of all combinations of `max_depth`

$= 10, 15, 20$ and $\text{min_sample_split} = 10, 15, 20$, was manually tested (i.e. by fitting a RF regressor for each parameter combination, and assessing the test set NRMSE, training time and memory usage).

Figure C.3 displays the NRMSE versus memory usage for this parameter space, with marker size scaled by training time. The NRMSE and training time sensitivity to max_depth and min_sample_split are low, while memory usage ranges from approximately 8 Mb to 100 Mb with increasing RF complexity. The RF accuracy improves slightly by increasing max_depth from 10 to 15, but further increases in max_depth result in only marginal improvements in performance. The results suggest that when memory usage / storage limits are not constraining factor, $\text{max_depth} = 15$ and $\text{min_sample_split} = 20$ are good choices. Where memory or storage constraints arise, $\text{max_depth} = 10$ also performs well, for any value of min_sample_split (see Table C.3).

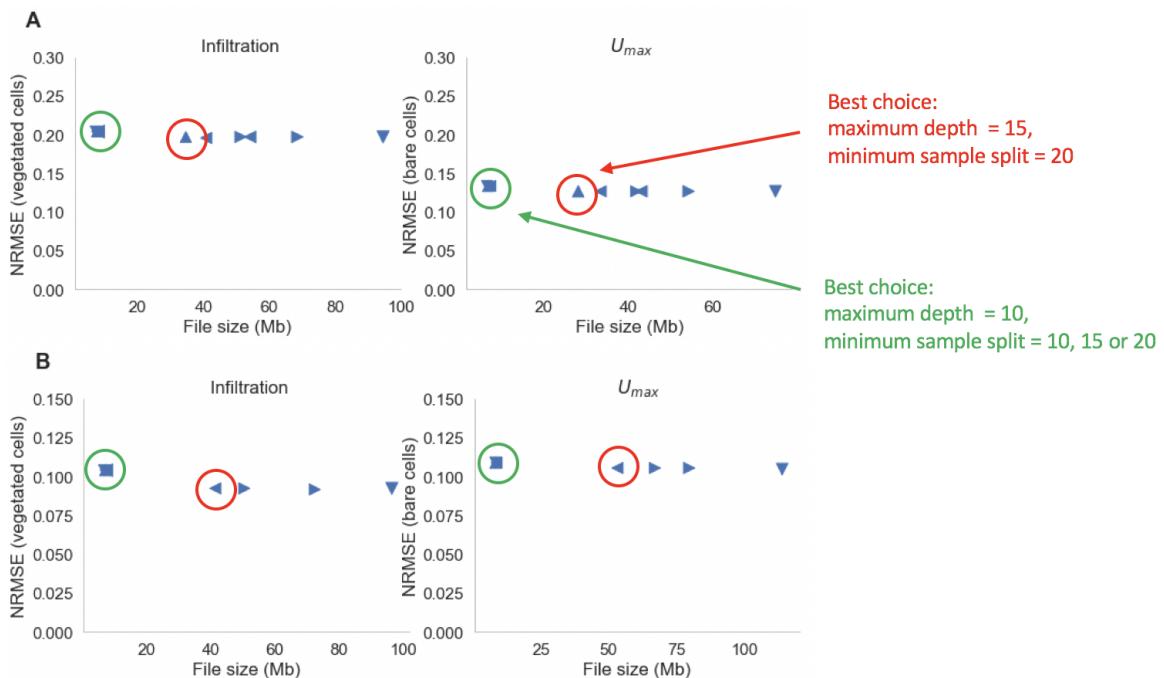


Figure C.3: Points represent each combination of max_depth and min_sample_split , indicating very little sensitivity in terms of performance to the RF specification. The $\text{max_depth} = 10$ results, circled in green, are the same for all values of min_sample_split . Increasing tree complexity beyond $\text{max_depth} = 15$, $\text{min_sample_split} = 20$ (red circles), does not lead to improvements in accuracy.

Table C.3: Select NRMSE scores for the ‘best’ (`max_depth = 15, min_sample_split = 20`) and ‘best if space limited’ (`max_depth = 10, min_sample_split = 20`) parameter cases.

Target	<code>max_depth</code>	<code>min_sample_split</code>	NRMSE(veg)	NRMSE(bare)	File size (Mb)
<i>Case A (Ch 5 subset with $d = 0.4$)</i>					
I	10	20	0.205	0.071	7.04
I	15	20	0.198	0.0678	34.4
U_{max}	10	20	0.358	0.134	6.53
U_{max}	15	20	0.348	0.128	28.1
<i>Case B (Ch 6 subset with $S_o = 10\%, K_{S,V} = 10 \text{ cm/hr}, t_{storm} = 30 \text{ min}$)</i>					
I	10	20	0.104	0.0512	6.91
I	15	20	0.0926	0.0505	41.3
U_{max}	10	20	0.172	0.109	8.14
U_{max}	15	20	0.152	0.106	53.2

C.3 Discussion and Conclusion

The sensitivity to varying the feature parameters and random forest parameters was independently assessed for each parameter, while holding all other parameters fixed. For example, $\zeta = 2$ was selected as the best smoothing length scale by assessing a range of ζ options while holding the RF parameters and λ s fixed. A more thorough approach would involve a grid search over these parameters. However, because the sensitivity of the NRMSE to the parameter specification was low, this approach was deemed sufficient.

In selecting the feature parameters and random forest parameters, I did not attempt to optimize across all potential combinations. Instead, I attempted to determine a set of parameters that would achieve good predictive accuracy, while also limiting the required storage requirements and training time. Increasing the number of features increases the required training time (see Figures C.1 and C.2), with marginal improvements in accuracy and little change in memory usage. Increasing the tree complexity (i.e. increasing `max_depth` or decreasing the `min_sample_split`) results in increased memory usage, but does not significantly influence the training time. Finally, because the RF performance is not very sensitive to the selection of these parameters, the parameters determined here to obtain the best results were used to fit the RF regressors in all Chapter 5 and Chapter 6 cases.

Bibliography

- Athol D Abrahams, Anthony J Parsons, and John Wainwright. Resistance to overland flow on semiarid grassland and shrubland hillslopes, walnut gulch, southern arizona. *Journal of Hydrology*, 156(1-4):431–446, 1994.
- S Assouline. Rainfall-induced soil surface sealing. *Vadose Zone Journal*, 3(2):570–591, 2004.
- S Assouline, SE Thompson, L Chen, T Svoray, S Sela, and GG Katul. The dual role of soil crusts in desertification. *Journal of Geophysical Research: Biogeosciences*, 120(10):2108–2119, 2015.
- JE Ball, MK Babister, R Nathan, PE Weinmann, W Weeks, M Retallick, and I Testoni. Australian rainfall and runoff-a guide to flood estimation, 2016.
- Graeme Barker. A tale of two deserts: contrasting desertification histories on rome's desert frontiers. *World Archaeology*, 33(3):488–507, 2002.
- Yoav Bashan, Bernardo G Salazar, Manuel Moreno, Blanca R Lopez, and Robert G Linderman. Restoration of eroded soil in the sonoran desert with native leguminous trees using plant growth-promoting microorganisms and limited amounts of compost and water. *Journal of environmental management*, 102:26–36, 2012.
- Jean-Francois Bastin, Nora Berrahmouni, Alan Grainger, Danae Maniatis, Danilo Mollicone, Rebecca Moore, Chiara Patriarca, Nicolas Picard, Ben Sparrow, Elena Maria Abraham, et al. The extent of forest in dryland biomes. *Science*, 356(6338):635–638, 2017.
- Jayne Belnap. Microbiotic crusts: their role in past and present ecosystems. *Park Science*, 10(3):3, 1990.
- Jayne Belnap. The potential roles of biological soil crusts in dryland hydrologic cycles. *Hydrological processes*, 20(15):3159–3178, 2006.
- Brandon T Bestelmeyer, Michael C Duniway, Darren K James, Laura M Burkett, and Kris M Havstad. A test of critical thresholds and their indicators in a desertification-prone ecosystem: more resilience than we thought. *Ecology Letters*, 16(3):339–345, 2013.

Brandon T Bestelmeyer, Gregory S Okin, Michael C Duniway, Steven R Archer, Nathan F Sayre, Jebediah C Williamson, and Jeffrey E Herrick. Desertification, land use, and the transformation of global drylands. *Frontiers in Ecology and the Environment*, 13(1):28–36, 2015.

Jock A Blackard and Denis J Dean. Comparative accuracies of artificial neural networks and discriminant analysis in predicting forest cover types from cartographic variables. *Computers and electronics in agriculture*, 24(3):131–151, 1999.

Piers Blaikie and Harold Brookfield. *Land degradation and society*. Routledge, 2015.

Bertrand Boeken and Moshe Shachak. Desert plant communities in human-made patches—implications for management. *Ecological Applications*, 4(4):702–716, 1994.

F Borgogno, P D’Odorico, F Laio, and L Ridolfi. Mathematical models of vegetation pattern formation in ecohydrology. *Reviews of Geophysics*, 47(1), 2009.

LJ Bracken, NJ Cox, and J Shannon. The relationship between rainfall inputs and flood generation in south-east spain. *Hydrological Processes: An International Journal*, 22(5):683–696, 2008.

Louise J Bracken and Jacky Croke. The concept of hydrological connectivity and its contribution to understanding runoff-dominated geomorphic systems. *Hydrological processes*, 21(13):1749–1763, 2007.

Scott F Bradford and Nikolaos D Katopodes. Hydrodynamics of turbid underflows. i: Formulation and numerical analysis. *Journal of hydraulic engineering*, 125(10):1006–1015, 1999.

Scott F Bradford and Nikolaos D Katopodes. Finite volume model for nonlevel basin irrigation. *Journal of irrigation and drainage engineering*, 127(4):216–223, 2001.

Leah L Bremer and Kathleen A Farley. Does plantation forestry restore biodiversity or create green deserts? a synthesis of the effects of land-use transitions on plant species richness. *Biodiversity and Conservation*, 19(14):3893–3915, 2010.

George Edward Briggs and John Burdon Sanderson Haldane. A note on the kinetics of enzyme action. *Biochemical journal*, 19(2):338, 1925.

J Bromley, J Brouwer, AP Barker, SR Gaze, and C Valentine. The role of surface water redistribution in an area of patterned vegetation in a semi-arid environment, south-west niger. *Journal of Hydrology*, 198(1-4):1–29, 1997.

Jason Brownlee. Discover feature engineering, how to engineer features and how to get good at it, Sep 2014. URL <https://machinelearningmastery.com/discover-feature-engineering-how-to-engineer-features-and-how-to-get-good-at-it/>.

- Wilfried Brutsaert. *Hydrology: an introduction*. Cambridge University Press, 2005.
- Yolanda Cantón, Albert Solé-Benet, J De Vente, Carolina Boix-Fayos, Adolfo Calvo-Cases, C Asensio, and Juan Puigdefábregas. A review of runoff generation and soil erosion across scales in semiarid south-eastern spain. *Journal of Arid Environments*, 75(12):1254–1261, 2011.
- Shixiong Cao. Why large-scale afforestation efforts in china have failed to solve the desertification problem, 2008.
- Robert F Carsel and Rudolph S Parrish. Developing joint probability distributions of soil water retention characteristics. *Water resources research*, 24(5):755–769, 1988.
- Stefano Castruccio, David J McInerney, Michael L Stein, Feifei Liu Crouch, Robert L Jacob, and Elisabeth J Moyer. Statistical emulation of climate model projections based on precomputed gcm runs. *Journal of Climate*, 27(5):1829–1844, 2014.
- Luis Cea, Cédric Legout, Frédéric Darboux, Michel Esteves, and Guillaume Nord. Experimental validation of a 2d overland flow model using high resolution water depth and velocity data. *Journal of hydrology*, 513:142–153, 2014.
- Michael A Celia, Efthimios T Bouloutas, and Rebecca L Zarba. A general mass-conservative numerical solution for the unsaturated flow equation. *Water resources research*, 26(7):1483–1496, 1990.
- RL Chazdon, MR Guariguata, et al. Decision support tools for forest landscape restoration: current status and future outlook. *CIFOR Occasional Paper*, (183), 2018.
- Robin L Chazdon. *Second growth: the promise of tropical forest regeneration in an age of deforestation*. University of Chicago Press, 2014.
- Cheng-lung Chen. Unified theory on power laws for flow resistance. *Journal of Hydraulic Engineering*, 117(3):371–389, 1991.
- Lin Chen, Shai Sela, Tal Svoray, and Schmuel Assouline. The role of soil-surface sealing, microtopography, and vegetation patches in rainfall-runoff processes in semiarid areas. *Water Resources Research*, 49(9):5585–5599, 2013.
- Yuxin Chen, Theodore Kolokolnikov, Justin Tzou, and Chunyi Gai. Patterned vegetation, tipping points, and the rate of climate change. *European Journal of Applied Mathematics*, 26(6):945–958, 2015.
- Nian-Sheng Cheng and Hoai Thanh Nguyen. Hydraulic radius for evaluating resistance induced by simulated emergent vegetation in open-channel flows. *Journal of Hydraulic Engineering*, 137(9):995–1004, 2010.

- Lorenzo Ciccarese, Anders Mattsson, and Davide Pettenella. Ecosystem services from forest restoration: thinking ahead. *New Forests*, 43(5-6):543–560, 2012.
- Greg J Cluff, James A Young, and Raymond A Evans. Edaphic factors influencing the control of wyoming big sagebrush and seedling establishment of crested wheatgrass. *Journal of Range Management*, pages 786–792, 1983.
- Paulo Cortez and Aníbal de Jesus Raimundo Morais. A data mining approach to predict forest fires using meteorological data. 2007.
- Vincent Deblauwe, Pierre Couteron, Jan Bogaert, and Nicolas Barbier. Determinants and dynamics of banded vegetation pattern migration in arid climates. *Ecological monographs*, 82(1):3–21, 2012.
- Luc Descroix, David Viramontes, Juan Estrada, José-Luis Gonzalez Barrios, and Jean Asse-line. Investigating the spatial and temporal boundaries of hortonian and hewlettian runoff in northern mexico. *Journal of hydrology*, 346(3-4):144–158, 2007.
- Paolo D’Odorico, Todd M Scanlon, Christiane W Runyan, Kate Abshire, Patrick Barrett, Abinash Bhattachan, James J Coloso, Ashton Erler, Janet Miller, Nathan Mitchell, et al. Dryland ecohydrology: research perspectives. *Annals of Arid Zone*, 48(3):229, 2009.
- Paolo D’Odorico, Abinash Bhattachan, Kyle F Davis, Sujith Ravi, and Christiane W Runyan. Global desertification: drivers and feedbacks. *Advances in Water Resources*, 51: 326–344, 2013.
- Kyle D Doherty, Bradley J Butterfield, and Troy E Wood. Matching seed to site by climate similarity: techniques to prioritize plant materials development and use in restoration. *Ecological Applications*, 27(3):1010–1023, 2017.
- David Dunkerley. Estimating the mean speed of laminar overland flow using dye injection-uncertainty on rough surfaces. *Earth Surface Processes and Landforms*, 26(4):363–374, 2001.
- David L Dunkerley. Determining friction coefficients for interrill flows: the significance of flow filaments and backwater effects. *Earth Surface Processes and Landforms: The Journal of the British Geomorphological Research Group*, 28(5):475–491, 2003.
- Thomas Dunne. Relation of field studies and modeling in the prediction of storm runoff. *Journal of Hydrology*, 65(1-3):25–48, 1983.
- Marcos H Easdale. Zero net livelihood degradation—the quest for a multidimensional protocol to combat desertification. *Soil*, 2(2):129–134, 2016.
- Millennium Ecosystem Assessment Ecosystems and Human Well-Being. Desertification synthesis. *Washington, DC: World Resources Institute*, 2005.

- Hari Eswaran, Rattan Lal, PF Reich, et al. Land degradation: an overview. *Responses to Land degradation*, pages 20–35, 2001.
- Matthew W Farthing and Fred L Ogden. Numerical solution of richards equation: A review of advances and challenges. *Soil Science Society of America Journal*, 2017.
- Jeffrey S Fehmi, Guo-Yue Niu, Russell L Scott, and Andrea Mathias. Evaluating the effect of rainfall variability on vegetation establishment in a semidesert grassland. *Environmental monitoring and assessment*, 186(1):395–406, 2014.
- Xiaoming Feng, Bojie Fu, Shilong Piao, Shuai Wang, Philippe Ciais, Zhenzhong Zeng, Yihe Lü, Yuan Zeng, Yue Li, Xiaohui Jiang, et al. Revegetation in chinas loess plateau is approaching sustainable water resource limits. *Nature Climate Change*, 6(11):1019, 2016.
- Xue Feng, Todd E Dawson, David D Ackerly, Louis S Santiago, and Sally E Thompson. Reconciling seasonal hydraulic risk and plant water use through probabilistic soil–plant dynamics. *Global change biology*, 23(9):3758–3769, 2017.
- Rob Ferguson. Flow resistance equations for gravel-and boulder-bed streams. *Water resources research*, 43(5), 2007.
- R Foti and JA Ramírez. A mechanistic description of the formation and evolution of vegetation patterns. *Hydrology and Earth System Sciences*, 17(1):63, 2013.
- Sylvie Galle, Joost Brouwer, and Jean-Pierre Delhoume. Soil water balance. In *Banded vegetation patterning in arid and semiarid environments*, pages 77–104. Springer, 2001.
- John E Gilley, WJ Elliot, JM Lafren, and JR Simanton. Critical shear stress and critical flow rates for initiation of rilling. 1993.
- J.V. Giraldez and D.A. Woolhiser. Analytical integration of the kinematic equation for runoff on a plane under constant rainfall rate and smith and parlange infiltration. *Water Resources Research*, (32):3385–3389, 1996.
- Evan B Goldstein and Giovanni Coco. Machine learning components in deterministic models: hybrid synergy in the age of data. *Frontiers in Environmental Science*, 3:33, 2015.
- Alan Grainger. Is land degradation neutrality feasible in dry areas? *Journal of Arid Environments*, 112:14–24, 2015.
- Liu Guangquan, Kuang Shangfu, and Tu Xiao. Water resources carrying capacity for vegetation restoration of eco-fragile region in the loess plateau [j]. *The Global Seabuckthorn Research and Development*, 8(1):13–20, 2010.
- Vishwesha Guttal and Ciriayam Jayaprakash. Self-organization and productivity in semi-arid ecosystems: implications of seasonality in rainfall. *Journal of theoretical biology*, 248(3):490–500, 2007.

- Dennis W Hallema, Roger Moussa, Ge Sun, and Steven G McNulty. Surface storm flow prediction on hillslopes based on topography and hydrologic connectivity. *Ecological Processes*, 5(1):13, 2016.
- Milos Hauskrecht. CS 2750, Lecture Notes: Machine Learning, 2003. URL: <https://people.cs.pitt.edu/~milos/courses/cs2750-Spring03/>. Last visited on 2018/10/01.
- Casper JM Hewett, Carolyn Simpson, John Wainwright, and Steve Hudson. Communicating risks to infrastructure due to soil erosion: A bottom-up approach. *Land Degradation & Development*, 29(4):1282–1294, 2018.
- Eric Higgs, Donald A Falk, Anita Guerrini, Marcus Hall, Jim Harris, Richard J Hobbs, Stephen T Jackson, Jeanine M Rhemtulla, and William Throop. The changing role of history in restoration ecology. *Frontiers in Ecology and the Environment*, 12(9):499–506, 2014.
- Reinier HilleRisLambers, Max Rietkerk, Frank van den Bosch, Herbert HT Prins, and Hans de Kroon. Vegetation pattern formation in semi-arid grazing systems. *Ecology*, 82(1): 50–61, 2001.
- Richard J Hobbs, Eric Higgs, and James A Harris. Novel ecosystems: implications for conservation and restoration. *Trends in ecology & evolution*, 24(11):599–605, 2009.
- Robert E Horton. The interpretation and application of runoff plot experiments with reference to soil erosion problems. In *Soil Science Society of America Proceedings*, volume 3, pages 340–349, 1938.
- Jianping Huang, Haipeng Yu, Xiaodan Guan, Guoyin Wang, and Ruixia Guo. Accelerated dryland expansion under climate change. *Nature Climate Change*, 6(2):166, 2016.
- Manoj K Jain, Umesh C Kothiyari, and Kittur G Ranga Raju. A gis based distributed rainfall-runoff model. *Journal of Hydrology*, 299(1-2):107–135, 2004.
- CS James, AL Birkhead, AA Jordanova, and JJ O’sullivan. Flow resistance of emergent vegetation. *Journal of Hydraulic Research*, 42(4):390–398, 2004.
- Jeremy J James and Peter J Carrick. Toward quantitative dryland restoration models. *Restoration Ecology*, 24:S85–S90, 2016.
- Juha Järvelä. Flow resistance of flexible and stiff vegetation: a flume study with natural plants. *Journal of hydrology*, 269(1-2):44–54, 2002.
- Juying Jiao, Zhenguo Zhang, Wenjuan Bai, Yanfeng Jia, and Ning Wang. Assessing the ecological success of restoration by afforestation on the chinese loess plateau. *Restoration Ecology*, 20(2):240–249, 2012.

Donald L Johnson, Stanley H Ambrose, Thomas J Bassett, Merle L Bowen, Donald E Crummey, Jeffry S Isaacson, Daniel N Johnson, Peter Lamb, Mahir Saul, and Alex E Winter-Nelson. Meanings of environmental terms. *Journal of environmental quality*, 26(3):581–589, 1997.

Eric Jones, Travis Oliphant, Pearu Peterson, et al. SciPy: Open source scientific tools for Python, 2001. URL <http://www.scipy.org/>. [Online; accessed †today‡].

Ramon Josa, Montse Jorba, and V Ramon Vallejo. Opencast mine restoration in a mediterranean semi-arid environment: failure of some common practices. *Ecological Engineering*, 42:183–191, 2012.

Gabriel G Katul, Davide Poggi, and Luca Ridolfi. A flow resistance model for assessing the impact of vegetation on flood routing mechanics. *Water Resources Research*, 47(8), 2011.

Karen Keenelyside, Nigel Dudley, Stephanie Cairns, Carol Hall, and Sue Stolton. *Ecological restoration for protected areas: principles, guidelines and best practices*, volume 18. iucn, 2012.

Sonia Kéfi, Max Rietkerk, Concepción L Alados, Yolanda Pueyo, Vasilios P Papanastasis, Ahmed ElAich, and Peter C De Ruiter. Spatial vegetation patterns and imminent desertification in mediterranean arid ecosystems. *Nature*, 449(7159):213, 2007.

Sonia Kefi, Max Rietkerk, and Gabriel G Katul. Vegetation pattern shift as a result of rising atmospheric co₂ in arid ecosystems. *Theoretical population biology*, 74(4):332–344, 2008.

Giora J Kidron. The effect of shrub canopy upon surface temperatures and evaporation in the negev desert. *Earth Surface Processes and Landforms*, 34(1):123–132, 2009.

Jongho Kim, Valeriy Y Ivanov, and Nikolaos D Katopodes. Hydraulic resistance to overland flow on surfaces with partially submerged vegetation. *Water Resources Research*, 48(10), 2012.

Geoffroy Kirstetter, Jie Hu, Olivier Delestre, Frédéric Darboux, P-Y Lagrée, Stéphane Popinet, Jose-Maria Fullana, and Christophe Josserand. Modeling rain-driven overland flow: Empirical versus analytical friction terms in the shallow water approximation. *Journal of Hydrology*, 536:1–9, 2016.

Alan K Knapp, Claus Beier, David D Briske, Aimée T Classen, Yiqi Luo, Markus Reichstein, Melinda D Smith, Stanley D Smith, Jesse E Bell, Philip A Fay, et al. Consequences of more extreme precipitation regimes for terrestrial ecosystems. *AIBS Bulletin*, 58(9):811–821, 2008.

Alexandra G Konings, Stefan C Dekker, Max Rietkerk, and Gabriel G Katul. Drought sensitivity of patterned vegetation determined by rainfall-land surface feedbacks. *Journal of Geophysical Research: Biogeosciences*, 116(G4), 2011.

- Vladimir M Krasnopolksy and Michael S Fox-Rabinovitz. A new synergetic paradigm in environmental numerical modeling: Hybrid models combining deterministic and machine learning components. *Ecological Modelling*, 191(1):5–18, 2006.
- Miroslav Kubat, Robert C Holte, and Stan Matwin. Machine learning for the detection of oil spills in satellite radar images. *Machine learning*, 30(2-3):195–215, 1998.
- R Lal, U Safriel, and B Boer. Zero net land degradation, a new sustainable development goal for rio+ 20; a report prepared for the secretariat of the unccd, 2012.
- Rattan Lal. Soil carbon sequestration to mitigate climate change. *Geoderma*, 123(1-2):1–22, 2004.
- Daniel Larkin, Gabrielle Vivian-Smith, and Joy B Zedler. Topographic heterogeneity theory and ecological restoration. *Foundations of restoration ecology*, 7:142–164, 2006.
- DSL Lawrence. Macroscale surface roughness and frictional resistance in overland flow. *Earth Surface Processes and Landforms: The Journal of the British Geomorphological Group*, 22(4):365–382, 1997.
- Quang Bao Le, Ephraim Nkonya, and Alisher Mirzabaev. Biomass productivity-based mapping of global land degradation hotspots. In *Economics of land degradation and improvement—A global assessment for sustainable development*, pages 55–84. Springer, 2016.
- Cédric Legout, Frédéric Darboux, Yves Nédélec, Alexandre Hauet, Michel Esteves, Bernard Renaux, Hervé Denis, and Stéphane Cordier. High spatial resolution mapping of surface velocities and depths for shallow overland flow. *Earth Surface Processes and Landforms*, 37(9):984–993, 2012.
- Hong-Yi Li, Murugesu Sivapalan, Fuqiang Tian, and Ciaran Harman. Functional approach to exploring climatic and landscape controls of runoff generation: 1. behavioral constraints on runoff volume. *Water Resources Research*, 50(12):9300–9322, 2014.
- XJ Li, XR Li, WM Song, YP Gao, JG Zheng, and RL Jia. Effects of crust and shrub patches on runoff, sedimentation, and related nutrient (c, n) redistribution in the desertified steppe zone of the tengger desert, northern china. *Geomorphology*, 96(1-2):221–232, 2008.
- John A Ludwig, Bradford P Wilcox, David D Breshears, David J Tongway, and Anton C Imeson. Vegetation patches and runoff–erosion as interacting ecohydrological processes in semiarid landscapes. *Ecology*, 86(2):288–297, 2005.
- Fernando T Maestre, David J Eldridge, Santiago Soliveres, Sonia Kéfi, Manuel Delgado-Baquerizo, Matthew A Bowker, Pablo García-Palacios, Juan Gaitán, Antonio Gallardo, Roberto Lázaro, et al. Structure and functioning of dryland ecosystems in a changing world. *Annual review of ecology, evolution, and systematics*, 47:215–237, 2016.

- BL Maheshwari. Suitability of different flow equations and hydraulic resistance parameters for flows in surface irrigation: A review. *Water resources research*, 28(8):2059–2066, 1992.
- Nate McDowell, William T Pockman, Craig D Allen, David D Breshears, Neil Cobb, Thomas Kolb, Jennifer Plaut, John Sperry, Adam West, David G Williams, et al. Mechanisms of plant survival and mortality during drought: why do some plants survive while others succumb to drought? *New phytologist*, 178(4):719–739, 2008.
- Gavan S McGrath, Kyungrock Paik, and Christoph Hinz. Microtopography alters self-organized vegetation patterns in water-limited ecosystems. *Journal of Geophysical Research: Biogeosciences*, 117(G3), 2012.
- GS McGrath, K Paik, and C Hinz. Complex landscapes from simple ecohydrological feedbacks. In *MODSIM2011, 19th International Congress on Modelling and Simulation*, pages 2528–2534. Model. and Simul. Soc. of Australia and New Zealand Canberra, 2011.
- J-C Menaut and B Walker. *Banded vegetation patterning in arid and semiarid environments: ecological processes and consequences for management*, volume 149. Springer Science & Business Media, 2001.
- Myles HM Menz, Kingsley W Dixon, and Richard J Hobbs. Hurdles and opportunities for landscape-scale restoration. *Science*, 339(6119):526–527, 2013.
- Ehud Meron. *Nonlinear physics of ecosystems*. CRC Press, 2015.
- Leonor Michaelis and Maude L Menten. The kinetics of the inversion effect. *Biochem. Z*, 49:333–369, 1913.
- Peter Moonen and J Allegrini. Employing statistical model emulation as a surrogate for cfd. *Environmental Modelling & Software*, 72:77–91, 2015.
- C Mügler, Olivier Planchon, J Patin, Sylvain Weill, Norbert Silvera, Patrick Richard, and E Mouche. Comparison of roughness models to simulate overland flow and tracer transport experiments under simulated rainfall at plot scale. *Journal of Hydrology*, 402(1-2):25–40, 2011.
- Marc F Müller, Sally E Thompson, and Maggi N Kelly. Bridging the information gap: A webgis tool for rural electrification in data-scarce regions. *Applied energy*, 171:277–286, 2016.
- Amir Mussery, Stefan Leu, Itamar Lensky, and Arie Budovsky. The effect of planting techniques on arid ecosystems in the northern negev. *Arid land research and management*, 27(1):90–100, 2013.
- Rotem Nativ, Jhonathan E Ephrath, Pedro R Berliner, and Yehoshua Saranga. Drought resistance and water use efficiency in acacia saligna. *Australian Journal of Botany*, 47(4):577–586, 1999.

- Immanuel Noy-Meir. Structure and function of desert ecosystems. *Israel Journal of Plant Sciences*, 28(1):1–19, 1979.
- Ilhan Özgen, Katharina Teuber, Franz Simons, Dongfang Liang, and Reinhard Hinkelmann. Upscaling shallow water models for overland flow using roughness formulations. 2015.
- Margaret A Palmer, Joy B Zedler, and Donald A Falk. *Foundations of restoration ecology*. Island Press, 2016.
- Chengzhong Pan, Lan Ma, John Wainwright, and Zhouping Shangguan. Overland flow resistances on varying slope gradients and partitioning on grassed slopes under simulated rainfall. *Water Resources Research*, 52(4):2490–2512, 2016.
- Athanasis Paschalis, Gabriel G Katul, Simone Fatichi, Gabriele Manoli, and Peter Molnar. Matching ecohydrological processes and scales of banded vegetation patterns in semiarid catchments. *Water Resources Research*, 52(3):2259–2278, 2016.
- F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.
- Marco Pennacchiotti and Ana-Maria Popescu. A machine learning approach to twitter user classification. *Icwsm*, 11(1):281–288, 2011.
- Gopal G Penny, Karen E Daniels, and Sally E Thompson. Local properties of patterned vegetation: quantifying endogenous and exogenous effects. *Phil. Trans. R. Soc. A*, 371(2004):20120359, 2013.
- JR Philips. The theory of infiltration: the infiltration equation and its solution. *Soil Sci*, 83(5):345–357, 1957.
- Jean Poesen, Jeroen Nachtergaele, Gert Verstraeten, and Christian Valentin. Gully erosion and environmental change: importance and research needs. *Catena*, 50(2-4):91–133, 2003.
- BA Portnov and UN Safriel. Prospective desertification trends in the negevimplications for urban and regional development. In *Environmental Challenges in the Mediterranean 2000–2050*, pages 123–138. Springer, 2004.
- PFBJ Quinn, K Beven, Pierre Chevallier, and Olivier Planchon. The prediction of hillslope flow paths for distributed hydrological modelling using digital terrain models. *Hydrological processes*, 5(1):59–79, 1991.
- Ulrich Razafison, Stephane Cordier, Olivier Delestre, Frédéric Darboux, Carine Lucas, and Francois James. A shallow water model for the numerical simulation of overland flow on surfaces with ridges and furrows. *European Journal of Mechanics-B/Fluids*, 31:44–52, 2012.

- SM Reaney, LJ Bracken, and MJ Kirkby. Use of the connectivity of runoff model (crum) to investigate the influence of storm characteristics on runoff generation and connectivity in semi-arid areas. *Hydrological Processes: An International Journal*, 21(7):894–906, 2007.
- SM Reaney, LJ Bracken, and MJ Kirkby. The importance of surface controls on overland flow connectivity in semi-arid environments: Results from a numerical experimental approach. *Hydrological Processes*, 28(4):2116–2128, 2014.
- P Reichert, G White, Maria J Bayarri, and E Bruce Pitman. Mechanism-based emulation of dynamic simulation models: Concept and application in hydrology. *Computational Statistics & Data Analysis*, 55(4):1638–1655, 2011.
- James F Reynolds, D Mark Stafford Smith, Eric F Lambin, BL Turner, Michael Mortimore, Simon PJ Batterbury, Thomas E Downing, Hadi Dowlatabadi, Roberto J Fernández, Jeffrey E Herrick, et al. Global desertification: building a science for dryland development. *science*, 316(5826):847–851, 2007.
- Lorenzo Adolph Richards. Capillary conduction of liquids through porous mediums. *physics*, 1(5):318–333, 1931.
- Max Rietkerk and Johan Van de Koppel. Regular pattern formation in real ecosystems. *Trends in ecology & evolution*, 23(3):169–175, 2008.
- Max Rietkerk, Pieter Ketner, Joep Burger, Bart Hoorens, and Han Olff. Multiscale soil and vegetation patchiness along a gradient of herbivore impact in a semi-arid grazing system in west africa. *Plant Ecology*, 148(2):207–224, 2000.
- Max Rietkerk, Maarten C Boerlijst, Frank van Langevelde, Reinier HilleRisLambers, Johan van de Koppel, Lalit Kumar, Herbert HT Prins, and André M de Roos. Self-organization of vegetation in arid ecosystems. *The American Naturalist*, 160(4):524–530, 2002.
- Max Rietkerk, Stefan C Dekker, Peter C De Ruiter, and Johan van de Koppel. Self-organized patchiness and catastrophic shifts in ecosystems. *Science*, 305(5692):1926–1929, 2004.
- Armin Ronacher. Flask: web development on drop at a time, 2010. URL <http://flask.pocoo.org/>. [Online; accessed 1 today].
- PM Saco, GR Willgoose, and GR Hancock. Eco-geomorphology of banded vegetation patterns in arid and semi-arid regions. *Hydrology and Earth System Sciences Discussions*, 11(6):1717–1730, 2007.
- Uriel Safriel. Land degradation neutrality (ldn) in drylands and beyond—where has it come from and where does it go. *Silva Fennica*, 51(1B):20–24, 2017.

- Jose Sarukhan, Anne Whyte, R Hassan, R Scholes, N Ash, ST Carpenter, PL Pingali, EM Bennett, MB Zurek, K Chopra, et al. Millenium ecosystem assessment: Ecosystems and human well-being. 2005.
- William H Schlesinger and Adrienne M Pilmanis. Plant-soil interactions in deserts. *Biogeochemistry*, 42(1-2):169–187, 1998.
- Josiane Seghieri, Sylvie Galle, Jean-Louis Rajot, and M Ehrmann. Relationships between soil moisture and growth of herbaceous plants in a natural vegetation mosaic in niger. *Journal of Arid Environments*, 36(1):87–102, 1997.
- Mingan Shao, Yunqiang Wang, Yongqiu Xia, and Xiaoxu Jia. Soil drought and water carrying capacity for vegetation in the critical zone of the loess plateau: A review. *Vadose Zone Journal*, 17(1), 2018.
- Jonathan A Sherratt. An analysis of vegetation stripe formation in semi-arid landscapes. *Journal of mathematical biology*, 51(2):183–197, 2005.
- Stephanie Shih. UCLA Linguistics 251: Variation in Phonology, 2012. URL: http://www-bcf.usc.edu/~shihhs/shih_randomforests.pdf. Last visited on 2018/10/01.
- David Short, Warrick R Dawes, and Ian White. The practicability of using richards' equation for general purpose soil-water dynamics models. *Environment International*, 21(5):723–730, 1995.
- Jirka Simunek, M Th Van Genuchten, and M Sejna. The hydrus-1d software package for simulating the one-dimensional movement of water, heat, and multiple solutes in variably-saturated media. *University of California-Riverside Research Reports*, 3:1–240, 2005.
- Koen Siteur, Eric Siero, Maarten B Eppinga, Jens DM Rademacher, Arjen Doelman, and Max Rietkerk. Beyond turing: The response of patterned ecosystems to environmental change. *Ecological Complexity*, 20:81–96, 2014.
- Koen Siteur, Maarten B Eppinga, Arjen Doelman, Eric Siero, and Max Rietkerk. Ecosystems off track: rate-induced critical transitions in ecological models. *Oikos*, 125(12):1689–1699, 2016.
- Mark W Smith. Roughness in the earth sciences. *Earth-Science Reviews*, 136:202–225, 2014.
- Mark W Smith, Nicholas J Cox, and Louise J Bracken. Applying flow resistance equations to overland flows. *Progress in Physical Geography*, 31(4):363–387, 2007.
- Ilan Stavi and Rattan Lal. Achieving zero net land degradation: challenges and opportunities. *Journal of Arid Environments*, 112:44–51, 2015.

Katharine Suding, Eric Higgs, Margaret Palmer, J Baird Callicott, Christopher B Anderson, Matthew Baker, John J Gutrich, Kelly L Hondula, Matthew C LaFevor, Brendon MH Larson, et al. Committing to ecological restoration. *Science*, 348(6235):638–640, 2015.

Katharine N Suding. Toward an era of restoration in ecology: successes, failures, and opportunities ahead. *Annual review of ecology, evolution, and systematics*, 42:465–487, 2011.

Tal Svoray, Rakefet Shafran-Nathan, Zalmen Henkin, and Avi Peregolotsky. Spatially and temporally explicit modeling of conditions for primary production of annuals in dry environments. *ecological modelling*, 218(3-4):339–353, 2008.

Yukie Tanino and Heidi M Nepf. Laboratory investigation of mean drag in a random array of rigid, emergent cylinders. *Journal of Hydraulic Engineering*, 134(1):34–41, 2008.

Lucile Tatard, Olivier Planchon, John Wainwright, Guillaume Nord, David Favis-Mortlock, Norbert Silvera, Olivier Ribolzi, Michel Esteves, and Chi Hua Huang. Measurement and modelling of high-resolution flow-velocity data under simulated rainfall on a low-slope sandy soil. *Journal of Hydrology*, 348(1-2):1–12, 2008.

The ECHO Project. Ecohydrological feedback dynamics at grassland-shrubland transitions in the south-western us. <https://www.uni-potsdam.de/echo/NewMexico.html>. [Online; accessed 7-September-2018].

The Sustainability Laboratory. Dryland agriculture. <http://www.sustainabilitylabs.org/ecosystem-restoration/dryland-agriculture/>. [Online; accessed 7-September-2018].

Sally Thompson and Gabriel Katul. Secondary seed dispersal and its role in landscape organization. *Geophysical Research Letters*, 36(2), 2009.

Sally Thompson, Gabriel Katul, and Sean M McMahon. Role of biomass spread in vegetation pattern formation within arid ecosystems. *Water Resources Research*, 44(10), 2008.

Sally Thompson, Gabriel Katul, Alexandra Konings, and Luca Ridolfi. Unsteady overland flow on flat surfaces induced by spatial permeability contrasts. *Advances in water resources*, 34(8):1049–1058, 2011.

Sally E Thompson, Shmuel Assouline, Li Chen, Ana Trahkenbrot, Tal Svoray, and Gabriel G Katul. Secondary dispersal driven by overland flow in drylands: Review and mechanistic model development. *Movement ecology*, 2(1):7, 2014.

SE Thompson, CJ Harman, P Heine, and GG Katul. Vegetation-infiltration relationships across climatic and soil type gradients. *Journal of Geophysical Research: Biogeosciences*, 115(G2), 2010.

- RO Tinoco, EB Goldstein, and G Coco. A data-driven approach to develop physically sound predictors: Application to depth-averaged velocities on flows through submerged arrays of rigid cylinders. *Water Resources Research*, 51(2):1247–1263, 2015.
- David J Tongway and John A Ludwig. Vegetation and soil patterning in semi-arid mulga lands of eastern australia. *Australian Journal of Ecology*, 15(1):23–34, 1990.
- David J Tongway and John A Ludwig. Rehabilitation of semiarid landscapes in australia. i. restoring productive soil patches. *Restoration Ecology*, 4(4):388–397, 1996.
- David J Tongway and John A Ludwig. Theories on the origins, maintenance, dynamics, and functioning of banded landscapes. In *Banded vegetation patterning in arid and semiarid environments*, pages 20–31. Springer, 2001.
- Atsushi Tsunekawa, Guobin Liu, Norikazu Yamanaka, and Sheng Du. *Restoration and development of the degraded Loess Plateau, China*. Springer, 2014.
- UN. United nations convention to combat desertification in those countries experiencing serious drought and/or desertification, particularly africa. 1995.
- Christian Valentin, Jean-Marc d'Herbès, and Jean Poesen. Soil and water components of banded vegetation patterns. *Catena*, 37(1-2):1–24, 1999.
- Johan van de Koppel, Max Rietkerk, Frank van Langevelde, Lalit Kumar, Christopher A Klausmeier, John M Fryxell, John W Hearne, Jelte van Andel, Nico de Ridder, Andrew Skidmore, et al. Spatial heterogeneity and irreversible vegetation change in semiarid grazing systems. *The American Naturalist*, 159(2):209–218, 2002.
- M Th Van Genuchten. A closed-form equation for predicting the hydraulic conductivity of unsaturated soils 1. *Soil science society of America journal*, 44(5):892–898, 1980.
- Joseph W Veldman, Gerhard E Overbeck, Daniel Negreiros, Gregory Mahy, Sozig Le Stradic, G Wilson Fernandes, Giselda Durigan, Elise Buisson, Francis E Putz, and William J Bond. Where tree planting and forest expansion are bad for biodiversity and ecosystem services. *BioScience*, 65(10):1011–1018, 2015.
- Jochem Verrelst, Juan Pablo Rivera, and Jose Moreno. Emulation of radiative transfer models. http://optimise.dcs.aber.ac.uk/wp-content/uploads/Session4_Jochem-Verrelst.pdf, 2017. [Online; accessed 30-September-2018].
- Lixin Wang, P d'Odorico, JP Evans, DJ Eldridge, MF McCabe, KK Caylor, and EG King. Dryland ecohydrology and climate change: critical issues and technical advances. *Hydrology and Earth System Sciences*, 16(8):2585–2603, 2012.
- Wei-Jie Wang, Wen-Xin Huai, Sally Thompson, and Gabriel G Katul. Steady nonuniform shallow flow within emergent vegetation. *Water Resources Research*, 51(12):10047–10064, 2015.

- STEVE G Whisenant. Terrestrial systems. *Handbook of Ecological Restoration Volume 1: Principles of Restoration*, pages 83–105, 2002.
- Steven G Whisenant, Thomas L Thurow, and Steven J Maranz. Initiating autogenic restoration on shallow semiarid sites. *Restoration Ecology*, 3(1):61–67, 1995.
- Robin P White, Daniel B Tunstall, and Norbert Henninger. *An ecosystem approach to drylands: building support for new development policies*. World Resources Institute Washington, DC, 2002.
- Wikimedia Commons. File:tiger bush niger corona 1965-12-31.jpg — wikipedia commons, the free media repository, 2015. URL \url{https://commons.wikimedia.org/w/index.php?title=File:Tiger_Bush_Niger_Corona_1965-12-31.jpg&oldid=156849502}. [Online; accessed 8-September-2018].
- Robin A Wooding. A hydraulic model for the catchment-stream problem: I. kinematic-wave theory. *Journal of hydrology*, 3(3-4):254–267, 1965.
- Ping-ping Yang, Hui-lan Zhang, and Chao Ma. Effects of simulated submerged and rigid vegetation and grain roughness on hydraulic resistance to simulated overland flow. *Journal of Mountain Science*, 14(10):2042–2052, 2017.
- Chunmei Yao, Tingwu Lei, William J Elliot, Donald K McCool, Jun Zhao, and Shulin Chen. Critical conditions for rill initiation. *Transactions of the ASABE*, 51(1):107–114, 2008.
- Walter Zegada-Lizarazu, Emilo Garcia-Apaza, Jhonathan Ephrath, and Pedro Berliner. Above and below ground development of acacia saligna shrubs grown under different irrigation frequencies in an arid environment. *Plant and soil*, 297(1-2):157–169, 2007.
- Guanhua Zhang, Guobin Liu, Liang Yi, and Pingcang Zhang. Effects of patterned artemisia capillaris on overland flow resistance under varied rainfall intensities in the loess plateau of china. *Journal of Hydrology and Hydromechanics*, 62(4):334–342, 2014.
- Zi-cheng Zheng, Shu-qin He, and Fa-qi Wu. Relationship between soil surface roughness and hydraulic roughness coefficient on sloping farmland. *Water Science and Engineering*, 5(2):191–201, 2012.