

**Distributed GIS-based Hydrologic Modelling
of Santa Catalina Island (CA), USA**

January 1999

B.SC.(HONS) GEOGRAPHY

Marco Ruocco

Contents

1	Abstract	5
2	Introduction	6
3	The hillslope system	8
4	Model structure	13
4.1	Overview	13
4.2	RUSLE	15
4.3	Channel Network Extraction	18
4.4	SCS Runoff	21
4.5	Conversion tables	23
5	Assembling the model	26
5.1	Topography	26
5.2	Soil	27
5.3	Vegetation	29
5.4	Precipitation	32
5.5	Fieldwork component	38
6	Results	40
7	Critique of the model	41
8	Conclusions	48
9	Acknowledgements	49

List of Figures

1	Location map of Santa Catalina Island (CA). Sources: (1) SmallWorld DEM Set, Questar Productions, (2) LargeWorld DEM Set, Questar Productions, (3) IFF 8-bit Z-Buffer derived from Catalina DEM, Catalina Conservancy. Scale bars indicative and not in perspective. Rendered with World Construction Set v2, Questar Productions.	7
2	The model of the geomorphological system. Diagram structure concept from [34].	9
3	Roadside and hillslope gullies, Santa Catalina Island.	12
4	Model construction diagram.	14
5	The definition of gradient and implications on RUSLE S factor.	17
6	The 8-directions flow model, adapted from [7]. The flow direction for the central cell is set according to the maximum drop, which in this case is straight downwards. The 8 numbers (from 1 to 128) indicate the code used internally to assign the flow direction to each cell. Rendered with VRLI VistaPro.	20
7	The base graph of the SCS Runoff method, from [26]	22
8	Runoff CN for hydrologic soil and cover complexes, from Soil Conservation Service (1972) and [6]	24
9	Cover type hydrologic classification, from Soil Conservation Service (1972) and [6]	24
10	Soil texture to K factor conversion nomograph, from [6].	26
11	The cluster vegetation map. Sources: ESRI Arc/INFO and ArcView3 on SPOT multiband images.	29
12	SPOT 3+1 bands stack	31
13	Procedure diagram for the generation of the precipitation surfaces.	33
14	Time series graph for rainfall daily totals, Avalon	34
15	Log Pearson distribution RI graph for Avalon	35

16	The aspect component of vegetation as taken from the field: the more vegetated slopes are facing north.	39
17	Extracted channel networks at 300 and 150 threshold accumulation values, and extracted watersheds of order 3, 4 and 5. Source: ESRI ArcView3, ESRI Arc/INFO and original Catalina DEM from Catalina Conservancy.	42
18	Morphometric statistics of Catalina extracted drainage network.	43
19	Slope map of Santa Catalina. The slopes shaded in blue represent the areas within the range of RUSLE base plots. Source: ESRI ArcView3 and Catalina DEM from Catalina Conservancy.	45

List of Tables

1	Order of variability of RUSLE parameters	46
---	--	----

1 Abstract

A hydrologic modelling system based on GIS (Geographic Information System) technology has been applied to the island of Santa Catalina (CA) to assess the extent of runoff processes and to estimate soil erosion. Digital and analogic data for topography, vegetation cover, precipitation patterns and soil characteristics have been assembled, adapted and organised in a distributed database in ESRI Arc/INFO format. The processing modules based on AML (Arc Macro Language) scripts were used for the semi-automatic extraction of stream networks from the Digital Elevation Model (DEM) representing topography, to calculate the rainfall/runoff ratios according to the SCS (Soil Conservation Service) method, and to estimate soil erosion potential using an adapted version of RUSLE (Revised Universal Soil Loss Equation). Remote Sensing techniques of cluster analysis have been used to produce a vegetation map, while the extracted channel networks and related morphometric analysis addressed analytically the hydrologic characteristics of the island. Attention has been devoted to the conceptual formulation of the model as a process of simplification from the broader geomorphological system. An assessment of model validity is conducted by considering the several processing steps in some detail from both conceptual and informational perspectives, and RUSLE in particular is considered in terms of the risks associated with the generalisation of the model from the designed range of applicability.

2 Introduction

Santa Catalina is the third largest island of the Channel Islands of California, located 40 miles from Los Angeles (see Figure 1). The extraordinary landscapes of the island provide the contextual framework for the presence of several unique expressions of Mediterranean type ecosystems [3]. Santa Catalina forms a geographical continuity with the other two major Channel Islands Santa Cruz and Santa Rosa, comparable in terms of climate, vegetation type and morphology. This study is centered on the phenomena of runoff and erosion which characterise the slopes of Santa Catalina, exposed to the impact of oceanic storms and, in historical times, to grazing. The need for precise assessment and prediction, in view of the deployment of conservation strategies, required the implementation of a hydrological model based on GIS (Geographic Information System) technology. The characteristics of the island, such as topography, vegetative cover, soil and precipitation patterns have been captured in a digital format and organised in a GIS database, from which processed information could be extracted. The modelling system employed here is based upon the methods taught in [22]. The first task of the document is to report the process of data preparation, data processing and result analysis in a structured manner. The second task is to expand the information base from the narrow view of the model, by including systematic interpretations of the underlying geomorphologic processes. This step, in turn, will allow to consider the model itself as the result of a continuous process of simplification and implementation which can be investigated and criticized.

At the broadest scale, the procedure of model construction is approached in subsequent stages of knowledge selection, formal definition, conceptual implementation and technical implementation. The document attempts to consider the topics in the same progression. The preliminary formulation of a hillslope-based geomorphological system provides a general framework for the following discussions. The model is then articulated and formalised as a propagation of this base system. A dual view of the model, conceptual and informational, is used, is used implicitly in the attempt of monitoring instances of data degradation, visualising the effort of model construction, and inter-

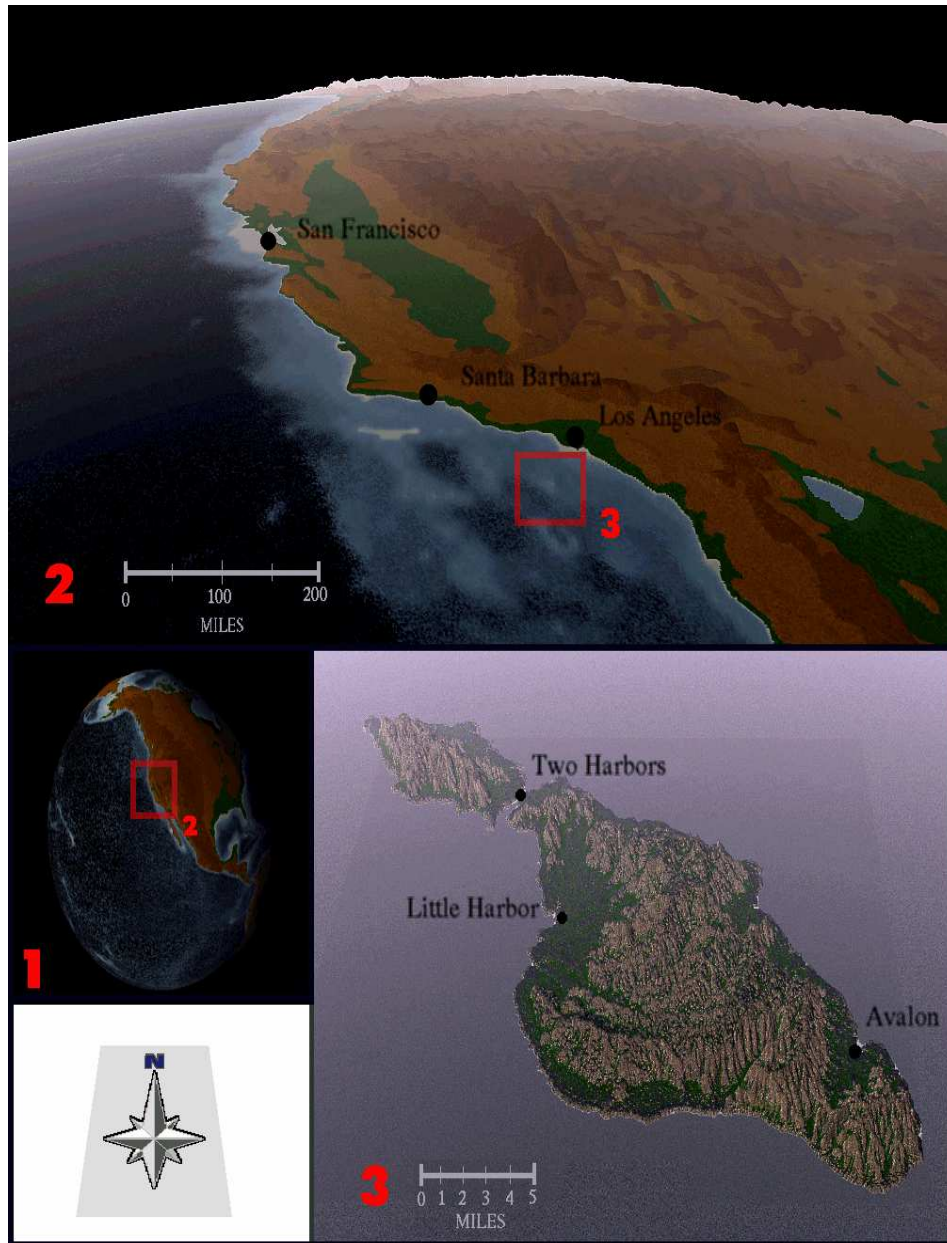


Figure 1: Location map of Santa Catalina Island (CA). Sources: (1) SmallWorld DEM Set, Questar Productions, (2) LargeWorld DEM Set, Questar Productions, (3) IFF 8-bit Z-Buffer derived from Catalina DEM, Catalina Conservancy. Scale bars indicative and not in perspective. Rendered with World Construction Set v2, Questar Productions.

preting the assumptions and implications made at each step. Processing units and data units are conceptually and visually separated, while the nature of their interaction is defined explicitly. The perspective of the analysis descends into single processing units to investigate internal structures and intermediate sources of mathematical relations, in the attempt of gaining a minimal depth for linking conceptual assumptions with technical implementation. Another objective then is to provide a general validity assessment for the results of the model.

3 The hillslope system

The analysis of Santa Catalina Island requires a prior understanding of the general geomorphological context in which the processes under study take place. The problem can be approached from a variety of scales, but the most suitable choice appears to focus the attention at the hillslope level. Such a framework allows to delineate the details of the subprocesses occurring at the small scale, integrating at the same time several hillslope systems in a single landscape unit.

At this stage, a first simplified model results useful in isolating, defining and relating the various key attributes of landscape elements and hydrologic processes within a common hillslope framework. The base model has the form of a flow model (see diagram 2), because it actually describes the flow of a water particle across a series of discrete processes (interception, infiltration, routing) taking place in discrete locations (atmosphere, vegetation, soil). Qualitative statements will delineate the internal specifications for each location, while the hydrologically relevant parameters are isolated and output to other parts of the model to assess their influence on geomorphic processes. Attempts of quantification offer specific insights into the details of system behavior, and indicate the available instruments of prediction which can be incorporated in other, more sophisticated models.

At the hillslope scale of analysis, precipitation can be reduced to the status of an external input variable. Quantity and intensity of rain are the factors which summarize its geomorphic influence [6]. Water is the primary element in the hydrologic cycle, and

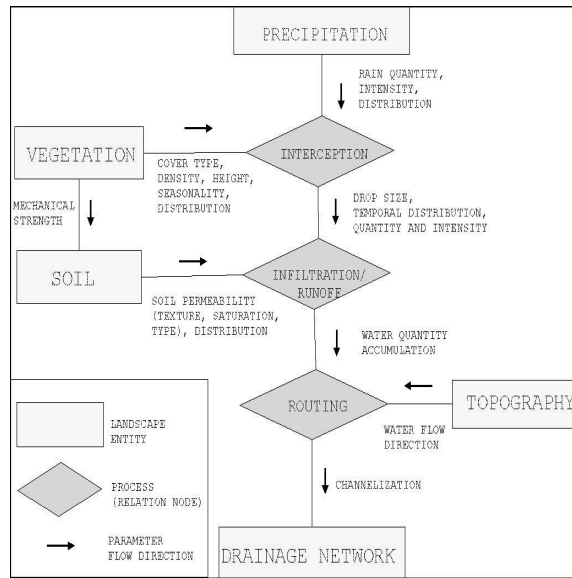


Figure 2: The model of the geomorphological system. Diagram structure concept from [34].

its fluctuation propagates to the rest of the system modifying its internal functioning (storage, buffering, transfer) as well as the final outputs. A pulsed versus continuous precipitation input, for example, might control the activation and relative importance of internal system processes. The preferential pulsed input over Catalina, driven by the approach of discrete oceanic storms, might decrease the relative importance of storage processes, with implications on the required model complexity [22].

Interception constitutes the first linking process between atmosphere and landscape. It involves the subtraction of rainfall water to runoff processes by effective storage on barks and canopy, and direct evaporation, by which moisture is returned to the atmosphere before reaching the ground [6]. It has therefore the effect of reducing the input totals to the hydrologic system, and of storing the water inputs in local buffers with the final result of retarding and redistributing in time the arrival of water to the ground. The governing variables of interception are foliage density of topmost canopy and lower undergrowth layers, while secondary parameters are leaves form and seasonal variations in case of deciduous versus conifer forest cover [6]. Apart from these water balance

aspects, interception has an influence over erosive processes in the form of changes of raindrop concentration and raindrop terminal velocity. A series of linear decay relationships between canopy cover, height and erosion have been proposed. More specific low canopy and high canopy erosion experiments showed exponential decay of erosion in the former and linear decay in the latter [24].

Infiltration is the process by which precipitation enters the soil. The soil acts as a filter that determines the path by which rainwater is discharged from a catchment, since water that does not infiltrate initiates overland runoff, while water that enters the soil moves much more slowly underground [24]. The infiltrated water is subdivided in Subsurface Storm Flow (SSF) reaching the water table, and Saturation Excess Overland Flow (SOF), which, after sub-horizontal movement across the hillslope, emerges again at the surface by spring sapping. Infiltration capacity is defined as the maximum flux of water across the given soil surface (i.e. the maximum throughput of the interface). The infiltration rate quantifying this moisture transfer process is expressed in units of depth per unit time, similarly to rainfall intensities. They refer to the depth of a sheet of water that would soak into the soil in the chosen time frame [6]. Typical infiltration curves show a rapid initial infiltration rate which drops quickly to some constant value. The factors governing this behavior are the increase of water content in the soil, and the alteration of the soil structure during a rainfall event. The increase in moisture level causes saturation of the soil which reduces the hydraulic gradient near the surface, a process enhanced by the low permeability of lower soil layers and the accumulation of throughflow from upslope. Also, changes in the soil surface, such as the reduction of pore sizes generated by washed-in clays, inhibit further infiltration [8].

The shape of the infiltration capacity curve is influenced by rainfall intensity and duration, and most importantly by soil characteristics as texture, structure, depth, proportion of clay minerals, vegetation and land use [8] [6]. Intense precipitation might pack the soil surface reducing the pore spaces between particles, and prolonged rainstorms may saturate the soil completely. Coarse textured soils such as sands, or soils with high proportions of organic matter, have large pores and a loose structure for fast

water drainage (high potential throughput), while clays tend to retard drainage (low potential throughput). The depth of the soil profile determines the actual moisture storage capacity to which surface throughputs are linked, so that shallow soil profiles developed in clays may present limited storage as well as low throughput potential [6]. The initial soil moisture conditions are determinant in the behavior of soils [8], in that they offset the hydraulic conditions to a position more or less close to saturation. Vegetation protects soil from packing by raindrops and provides organic matter for the formation of soil aggregates which facilitate infiltration. Changes in soil cover consequent to land use policies, such as grazing or forest removal, may alter dramatically the infiltration capacity and cause soil erosion [6]. When the throughflow capacity of the soil surface is surpassed by rain intensity, or in case of soil storages saturation, overland runoff will be produced on the hillslope. The flow is formed as a sheet or film of water, but at a critical distance downslope, when runoff has reached the threshold depth, erosion will be initiated in the form of rills. The actual routing of water over the landscape is determined by topography. Slope angle regulates the extent of the downslope component of water gravity, and consequently water velocity, which in turn influences erosional processes. High slopes imply high flow velocities and deep rill incisions. Slope aspect controls the direction of flow, since water particles move along the shortest paths defined by maximum drop. Curvature, the first derivative of slope, affects the shape and spatial variability of the shortest paths of water motion, and according to which axis is considered (longitudinal or transversal) it is termed plan curvature or profile curvature. The former indicates the convergence of flow lines along highest slope directions, or the divergence of flow away from ridges or topographic highs, and it is therefore a control upon channel formation and water routing. The latter considers the variability of slope angles, and the consequent acceleration and deceleration of flow along a single hillslope profile. Considering the implication of velocity changes upon energy transfers, transversal curvature can be used as a surrogate for the determination of areas of erosion (acceleration) and deposition (deceleration) [7].

The process of erosion is strongly linked with the hydrologic cycle, but it is worth-



Figure 3: Roadside and hillslope gullies, Santa Catalina Island.

while to consider it in some more detail. Erosion is a three-stages process: detachment, transport and deposition [24]. Detachment is caused by rainsplash (the impact of rain-drops) which induces consolidation and subsequently disruption and dispersion of soil particles. Experiments have shown that rainsplash erosion varies with the square of the instantaneous rainfall intensity. Other detachment factors include weathering (mechanical, wetting-drying, frost induced, or biochemical), tillage practice and trampling of people and livestock, wind and running water. In the latter case, detachment occurs when the resistance of the particle to motion equals the force of the flow. Resistance has been modelled by the Shields number using parameters as water and particle densities, particle size and shear velocity of the flow. Its validity is limited by the exclusion from the formula of additional factors such as rainsplash effects and cohesion. A more general relationship relates the rate of detachment to grain sizes as a parabolic curve, where the opposite extremes indicate the higher resistance of respectively clays, characterised by internal cohesion, and coarse particles, for which the gravitational force becomes dominant. The transport of soil can be distinguished in areal action processes (such as rainsplash and surface runoff of "infinite" width (overland runoff), and concentrated flow processes in channels (from rills to a wide gullies), which also implicitly delineate the interrill erosion zone (overland water film runoff and rainsplash) [24].

The erosional power is a function of the two main variables of potential energy and velocity, and along the profile of a hillslope a sequence of areas of acceleration and erosion can be expected according to their interplay [24]. The several factors and models have been grouped and represented in synthetic indexes, constituting a sort of intermediate layer between theory and empirical observations. For examples, the index of agent erosivity is based on rainfall intensity, which is empirically linked to kinetic energy, in turn used as a surrogate of splash, overland and rill erosion processes. Soil erodibility is similarly expressed in the K index (considered later), which expresses soil loss per unit of rainfall erosivity, which is then again linked to the same base variable (i.e. rainfall intensity as a surrogate of kinetic energy) [24]. The use of indexes is advantageous but causes a conceptual separation from the underlying processes and a certain degree of error in the determination of actual index values when no direct field observation is available.

4 Model structure

4.1 Overview

The conceptual model is a structured description that provides a common understanding of system organisation, behavior and nomenclature [21]. In this section the Catalina model is presented in its various data and processing components. The set of entities and relationships considered at this stage are a subset of the specifications of the more general hillslope model previously illustrated. The major simplification is the limitation of water routing to direct runoff, by which infiltrated water simply exits the system and does not contribute to runoff in the form of Saturation-Excess Overland Flow. The model does not include the process of storage by the soil, while the dynamicity of the processes of infiltration (such as time dependent variations) is substituted by a static conceptualisation. The process of model building can be traced in terms of information flows, data corruption and format conversion. The conceptual characterisation of the model is therefore accompanied by the technical specifications of the data structures,

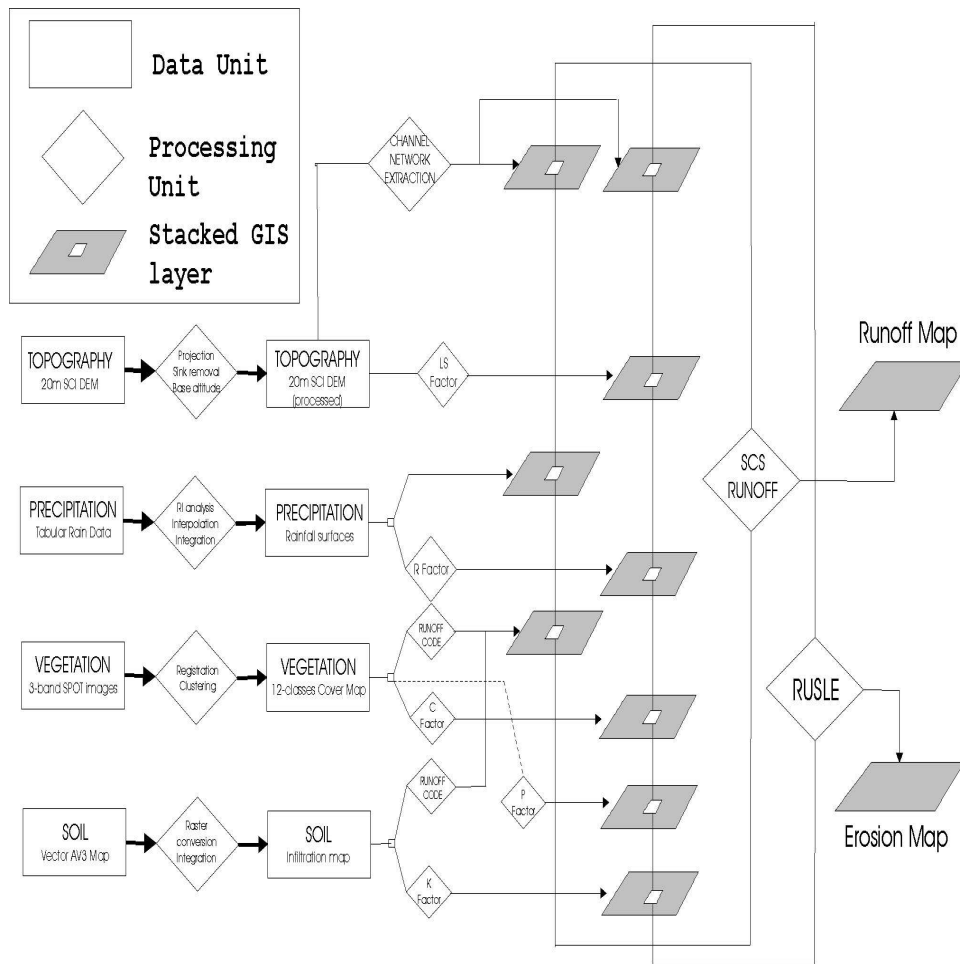


Figure 4: Model construction diagram.

and by the analysis of the conversion and processing procedures carried out on the data. The problem of providing a complete distributed parameters database for the entire island is solved by differentiating the units of hydrologic significance into independent data layers. The data structure employed in the model is raster, which therefore defines the landscape units as square grid cells of uniform size. This is maintained throughout the model, from the topography, represented by a DEM, to the vegetation, soil and precipitation layers, converted in preprocessing stage to raster structure.

The diagram in Figure 4 offers a general schematisation of the process of model construction. The data components are imported in a variety of formats, from surfaces

to vector coverages to tabular data. The objective of most of the preprocessing is to transform the sources into the internal "language" of the processing modules. The algorithmic component is based upon the main RUSLE, SCS Runoff and Channel Extraction modules, plus the conversion tables. They will be illustrated here in sequence.

4.2 RUSLE

RUSLE (Revised Universal Soil Loss Equation) is an empirically derived equation predicting average annual erosion. Its governing parameters have been obtained from the regression analysis of the erosion values registered at more than 10000 test sites. It is therefore an empirical black-box method, composed by subparameters or indexes which are not purely process oriented. RUSLE is also the most recent development of the older USLE (Universal Soil Loss Equation), from which differs in several aspects [29].

The main RUSLE equation is so structured:

$$A = 0.224 \times R \times K \times L \times S \times C \times P \quad (1)$$

where A = the soil loss in $\frac{kg}{m^2}$, R = the rainfall erosivity factor, K = the soil erodibility factor, L = the slope length factor, S = the slope gradient factor, C = the cropping management factor and P = the erosion control practice factor. The R factor is based on the values for the 30 minutes intensity of a storm with high recurrence interval (RI). RUSLE is based on the extrapolation of this base quantity that becomes representative of the overall rainfall erosivity effect over the entire yearly time frame. A correlation has been traced between the characteristics of an instant event and the long term erosion, even if they are consistently different concepts. K determines the inherent tendency of the soil to erosion, and it constitutes the rate of erosion per unit of erosion index from a standard plot. When there is not a K number for a given soil type, the value is obtained by using special nomographs relating texture to particle size. L and S are the topographic factors of the model, directly related to water hydrology. Slope length L indicates the distance between the top of the slope and the area of deposition or entrance into a stream channel. It is expressed as a length ratio with the base experimental plot

with the equation

$$L = \left(\frac{x}{22.13}\right)^m$$

where x is the slope length and m is an exponent that varies from 0.2 to 0.5, according to the range of slope variation from less than 1% to more than 5%. The sampling was based mostly on slopes of no more than 9% steepness, or approximately 6 degrees [24]. According to [6], the base data extends to 20%, or 11 degrees, beyond which the values are based on extrapolation. An additional subfactor introduced in RUSLE is the susceptibility of the soil to rill erosion as a function of L . This is not included in the actual algorithmics of the model, which instead computes the power function with exponent values similar, but not identical, to the ones proposed in [15]. Slope gradient S is the second topographic variable used in RUSLE, and its relationship with erosion is approximately linear. The weight of S in the overall estimated results is much higher than the L factor, and a 10% percent variation in slope gives an increase in soil loss of 20%. The base data in the case of L ranges from 10 to 400 feet (3 to 120 meters), beyond which the values are extrapolated. It is calculated with the formula illustrated in [15]

$$S = \frac{0.43 + 0.30s + 0.043s^2}{6.613} \quad (2)$$

where S is the slope gradient factor and s is the gradient per cent, which expanded becomes

$$S = 0.065 + 0.045s + 0.0065s^2 \quad (3)$$

When expressing the slope in degrees, s can be considered as the percentage version of $\tan \frac{AB}{BC}$, where AB and BC are respectively the horizontal and vertical components of slope (see Figure 5). After simple calculations the final equation becomes

$$S = 0.65 + 4.5 \tan s + 65(\tan s)^2 \quad (4)$$

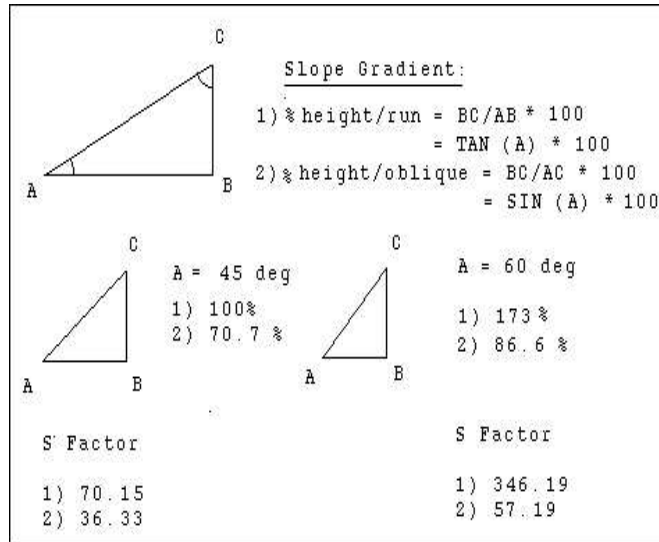


Figure 5: The definition of gradient and implications on RUSLE S factor.

In this case, s is the slope gradient in decimal degrees. The actual implementation of the S -factor algorithm actually uses $\sin(s)$ instead of $\tan(s)$ [22]. This seems due to the interpretation of gradient as the difference in height divided by the oblique distance (Figure 5), used by USLE developers. Other instances of this incorrect interpretation can be found for example in [24]. According to the definition, the gradient is a vector expressing the derivative of a function in that particular direction [1]. In the context of topography, the vector direction is vertical, and therefore the gradient represents the variation in height in relation to the horizontal distance. This is correctly expressed in the $\tan(s)$ interpretation of gradient. Interestingly, the general theory linking slope with erosion states that erosion is proportional to the tangent of slope [24]. It might be that USLE developers used a quadratic form of the sin of the slope angle as a surrogate for this general power relation based on tan, and the definition of gradient has been "adapted" to this empirical necessity. In practice the curve behavior at higher slopes is more realistically defined with sin, especially considering the asymptotic increase of tan. RUSLE is based on a more linear $S \rightarrow$ slope relation, but it has not been possible to know which one in particular [29]. In any case, the model here implemented seems

to use the old USLE S -factor equation.

The cropping management factor C represents the ratio of soil loss from a specific cover condition to the soil loss from a tilled condition, maintaining constant soil type, slope and rainfall. In its original agricultural implementation it includes the interrelated effects of cover, crop sequence, productivity level, growing season length and rainfall distribution [24]. In the most recent RUSLE implementation, C is calculated as a result of subfactors PLU (prior land use), CC (canopy), SC (surface cover) and SR (surface roughness). This subdivision allows a more analytical approach to the designation of the overall C value [29]. Unfortunately the model here implemented uses the oldest classification method offered by USLE, as implied by the use of pre-RUSLE conversion tables by US Soil conservation Services, presented in [6] and considered later.

The erosion control practise factor P is the ratio of soil loss using the specific agricultural practice compared with the soil loss using up-and-down hill culture. It includes practices as contouring, contour strip-cropping, and terracing. It is probably the most agricultural specific factor, and also one of the least reliable [29]. In RUSLE new data has extended the sample base to reevaluate the effects of contouring, and P factors have been developed to reflect conservation practices on rangeland. As for the C factor, the use of conversion tables in [6] does not allow the access to these new RUSLE features, and P does not seem to be defined outside agricultural contexts.

In conclusion, the model used here practically implements USLE, the predecessor of RUSLE, similar in structure but devoid of the new extended features. This is mainly due to the use of pre-RUSLE classificatory literature. The USLE model used here is actually a modified version adapted for functioning with distributed databases: in other words, the calculation is repeated for each 20×20 meters cell of the landscape.

4.3 Channel Network Extraction

Topography is the main variable governing the flow of water over a landscape. Potential energy and water movement depend in fact upon altitude and its variation in space, according to the laws of kinematics. A topographic model should be chosen so that the

most important altimetric information governing the flow of water is captured and maintained valid even if in simplified form [2]. From a Digital Elevation Model (DEM) it is possible to extract channel networks by numerical simulation of the forces operating on reality.

At the microscale, a single drop of water would flow downslope along the direction of greatest drop. This process of downward movement is repeated until the water droplet reaches the bottom of a valley or the stream channel. Translated into a digital raster representation, the movement of a water droplet located in a central cell can be modelled as a transition to one of the 8 neighbouring destination cells [7], as shown in Figure 6. Slope can be derived from the DEM using local interpolation methods [18], and according to the direction of greatest vertical difference, the local direction of flow is determined. Internally, a number is assigned to the central cell defining its extracted hydrologic behavior. In order to compensate for the greater distance along diagonal directions, a weighting factor of $\sqrt{2}$ is added so that the correct gradient (this time based on tan) can be computed. Alternative solutions to the movement along a single direction include the distribution of water to all the neighbouring cells in quantities proportional to slope, as seen in some models of subglacial hydrology [30]. If the cell is lower than all the other eight neighbouring cells, it causes internal drainage and it is then termed *sink* or *pit*. Most of the times, sinks are artifacts of the DEM, and they can be removed by automatic functions of detection and hydrologic correction [7].

A general map of flow directions can be generated by repeating the flow direction process iteratively across the entire landscape (by means of a 3×3 "moving window", so that every cell is considered in all the eight possible positions). Assuming a uniform input of precipitation across the entire landscape, every cell would be the starting position for the same amounts of water units. If instead the rainfall input is not uniform, the sourcing of water units will vary across the landscape in distribution and quantity. In either case, the input is independent from the general flow direction map and constitutes a separate layer which acts as a complete parameter for precipitation. When the input water is routed on the DEM it generates a map of flow accumulation, where some cells

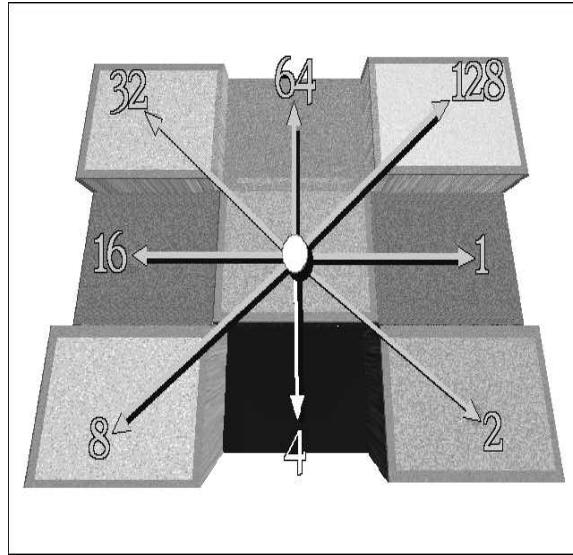


Figure 6: The 8-directions flow model, adapted from [7]. The flow direction for the central cell is set according to the maximum drop, which in this case is straight downwards. The 8 numbers (from 1 to 128) indicate the code used internally to assign the flow direction to each cell. Rendered with VRLI VistaPro.

will contain more water units than others because they constitute the relative outlet of multiple upstream cells. When the rainfall input layer is uniform and equal to 1, the value at each cell reflects the number of upstream cells flowing into it, and it is an index of the size of the flow of water in that cell. The necessary operation before delineating the channel networks is to decide the threshold value of accumulation beyond which a flow path is considered a proper channel to be mapped as such. Small values (around 50 cells) capture fine details and single headwaters, while large values (around 300 cells) produce a more coarser network of only prominent features. The transition from colluvial hollows to actual channels on the real landscape indicates the suitable threshold value. The final extracted networks are then dependent on the grid size, the routing algorithm and the choice of accumulation values: the use of different values in the three components might produce substantial differences [27] [32].

The extraction of watershed boundaries works in a similar way. A watershed is de-

fined as an area which drains water and other substances to a common outlet or pour point of concentrated drainage [7]. The algorithm for watershed definition consists in a recursive climbing from the pour point (and connected channel network cells) to the upslope contributing cells. If the cells considered during an iteration have no contributing cells (no inflow of water), the drainage divide is reached and the external boundary of the catchment can be delineated [18]. The pour points can be manually set on the map, or they can be defined automatically on the boundary between the landscape and the cells with *NODATA* values on the raster. The internal boundaries of the watersheds vary according to the accumulation threshold values defined for the channels. As an additional classification measure, the channel networks and relative watersheds can be defined with network ordering methods, such as the one of Strahler. Since different orders are stored in different datasets, they can be considered separately to derive statistical data about basin morphometry.

4.4 SCS Runoff

A quantitative, empirical method for estimating runoff given incoming rainfall and soil characteristics was developed by the Soil Conservation Service (SCS) under the name of TR-55 (Technical Release 55) [26]. The formula is derived from a decomposition of the hypothetical rainfall and runoff curves into descriptive parameters, as shown in Figure 7.

The main variables are so defined: P is rainfall (inches); P_e is effective storm runoff after initial abstraction; Q is runoff (inches); I_a is initial abstraction (inches), i.e. rainfall intercepted, infiltrated or stored in surface depressions before runoff begins; F is the actual retention after runoff begins (inches), variable with time and equal to $P_e - Q$; S is the potential maximum retention (inches), a constant defining the maximum retention possibly occurring for a given storm. The equation linking rainfall and runoff is obtained in the following equation:

$$Q = P_e \times \frac{F}{S} \tag{5}$$

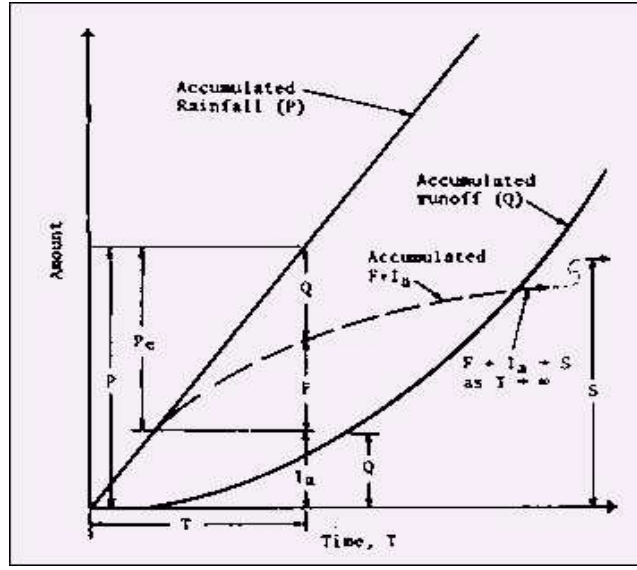


Figure 7: The base graph of the SCS Runoff method, from [26]

The ratio $\frac{F}{S}$ between retention and maximum retention here determines the separation between runoff and actual runoff. Substituting $F = P_e - Q$

$$Q = P_e \times \frac{P_e - Q}{S} = \frac{P_e^2}{P_e + S} \quad (6)$$

The actual runoff is the amount of rainfall minus initial abstraction. Substituting $P_e = P - I_a$

$$Q = \frac{(P - I_a)^2}{(P - I_a) + S} \quad (7)$$

By analysing rainfall and runoff data from experimental small watersheds, the relationship between I_a and S has been found as a constant with value $I_a = 0.2 \times S$. Then the equation becomes

$$Q = \frac{(P - 0.2 \times S)^2}{P + 0.8 \times S} \quad (8)$$

The factor of maximum retention S is determined by

$$S = \left(\frac{1000}{CN}\right) - 10 \quad (9)$$

where CN is the Curve Number, a constant value defined for every different combination of soil infiltration capacity and cover type, and grouped in tables such as in [6]. The final equation becomes

$$Q = \frac{P - 0.2 \times \left(\frac{1000}{CN} - 10\right)^2}{P + 0.8 \times \left(\frac{1000}{CN} - 10\right)^2} \quad (10)$$

which is equivalent to the equation used by the AML script to compute runoff on a cell to cell basis. The pivot of the entire equation is CN , determined by converting the available data according to the classification tables. The method of determination of CN is probably based on regression on the entire range of field data. A component missing in the present application of the model is the use of the Antecedent Soil Moisture condition factor, an additional factor that shifts curve values to accommodate for the initial level of saturation of the soil. The intermediate value II is used instead [26].

There are three important passages determining the accuracy of the SCS Runoff model: the first is the fit of the analytic representation to physical reality; the second is the accuracy of SCS CN regressions; the third is the process of conversion of data into CN numbers.

4.5 Conversion tables

The conversion tables are an element of major importance in the information flow of the model. Their role is to convert data into representations meaningful to the algorithmic modules previously described. At this point it is interesting to investigate what kind of data conversion is carried out, and what are the consequent implications for data quality. Several conversion tables, developed by the US Soil Conservation Service and presented in [6], are used for the determination of RUSLE C factor for vegetation, K factor for the soil and CN (Curve Numbers) for soil and vegetation combined.

For the soil component of the CN factor, the available information was ordinal (4-classes) infiltration data and the table requested ordinal (4 classes) infiltration data,

LAND USE OR COVER	TREATMENT OR PRACTICE	HYDROLOGIC CONDITION	HYDROLOGIC SOIL GROUP			
			A	B	C	D
Fallow	Straight row	Poor	77	86	91	94
Row crops	Straight row	Poor	72	81	88	91
	Straight row	Good	67	78	85	89
	Contoured	Poor	70	79	84	88
	Contoured	Good	65	75	82	86
	Contoured and terraced	Poor	66	74	80	82
Small grain	Contoured and terraced	Good	62	71	78	81
	Straight row	Poor	65	76	84	88
	Straight row	Good	63	75	83	87
	Contoured	Poor	63	74	82	85
	Contoured	Good	61	73	81	84
Close-seeded legumes or rotation meadow	Contoured and terraced	Poor	61	72	79	82
	Contoured and terraced	Good	59	70	78	81
	Straight row	Poor	66	77	85	89
	Straight row	Good	58	72	81	85
	Contoured	Poor	64	75	83	85
Pasture or range	Contoured	Good	55	69	78	83
	Contoured and terraced	Poor	63	73	80	83
	Contoured and terraced	Good	51	67	76	80
	Poor	68	79	86	89	
	Fair	49	69	79	84	
Meadow (permanent)	Good	39	61	74	80	
	Contoured	Poor	47	67	81	88
	Contoured	Fair	25	59	75	83
	Contoured	Good	6	35	70	79
Woodlands (farm woodlots)	Good	30	58	71	78	
	Poor	45	66	77	83	
	Fair	36	60	73	79	
Farmsteads	Good	25	55	70	77	
	Poor	59	74	82	86	
Roads, dirt		72	82	87	89	
Roads, hard-surface		74	84	90	92	

Figure 8: Runoff *CN* for hydrologic soil and cover complexes, from Soil Conservation Service (1972) and [6]

VEGETATIVE COVER	HYDROLOGIC CONDITION
Crop rotation	Poor: Contain a high proportion of row crops, small grains, and fallow.
	Good: Contain a high proportion of alfalfa and grasses.
Native pasture or range	Poor: Heavily grazed or having plant cover on less than 50% of the area.
	Fair: Moderately grazed; 50-75% plant cover.
	Good: Lightly grazed; more than 75% plant cover.
	Permanent Meadow: 100% grass cover.
Woodlands	Poor: Heavily grazed or regularly burned so that litter, small trees, and brush are destroyed.
	Fair: Grazed but not burned; there may be some litter.
	Good: Protected from grazing so that litter and shrubs cover the soil.

Figure 9: Cover type hydrologic classification, from Soil Conservation Service (1972) and [6]

with the only difference that the ranges were defined in a non-linear numerical fashion in the input case, and according to a qualitative subdivision in the case of the table. This slight mismatch might introduce a degree of error difficult to quantify. In the case of the CN factor for vegetation, the great variability of cover type characteristics had to fit into a 2 class subdivision (*Rangeland* and *Woodland*) with 3 subclasses each (*poor*, *medium* and *good*) - see Figure 8. The tables were originally compiled with a particular attention to land use (grazing) and undergrowth aspects of vegetation, which could not be directly assessed on a cover type basis. The mismatch at this point is due to different information sources, and the reclassified data content may be substantially different from the original. A similar situation is found in the case of the C factor, which depends on percentage cover, cover height and undergrowth type, while the available information was less parametric and also had a distributed component which affected the estimation of cover percentages. Finally, the table for the K factor conversion (see Figure 10) was actually a nomograph relating grain size to a range of values for K , empirically obtained from a base of approximately 20 measurements. The visual estimation of K on the graph implied a range of error of approximately 50% throughout the textural spectrum, which also was not defined with the same textural succession as the one in the available dataset.

The conversion tables are the stage in the model where the intervention of qualitative estimation is most influential, and coincidentally where the data undergoes a process of redefinition only partially supported by available data or field observation. The problem is principally due to a difficulty in adapting to the format of classification set by the Soil Conservation Tables, which are not specifically designed for the type of application presented in this context.

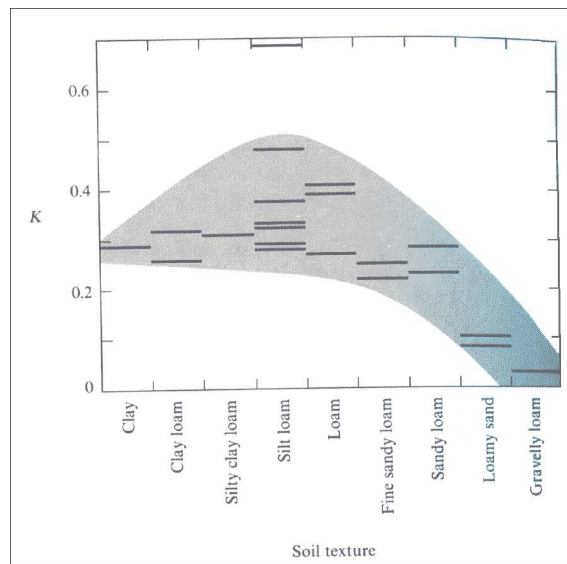


Figure 10: Soil texture to K factor conversion nomograph, from [6].

5 Assembling the model

5.1 Topography

The Catalina DEM consists of an array of 1398 x 1016 binary values referring to the altitude in feet from sea level. The spatial resolution (i.e., the spacing of each elevation value) is 20 meters. The original DEM was registered in latitude/longitude according to the UTM coordinate system using the Clarke 1927 datum. The source maps were the standard USGS quadrangle sheets that were scanned and digitized along the contours by the Catalina Conservancy. An Arc/INFO proprietary Countour-to-DEM utility carried out the interpolation at the required resolution. The exact method employed by Arc/INFO is difficult to deduce, but it was probably a specific algorithm possibly to be used in combination with other data sources, such as stream network coverages, to avoid the occurrence of artifacts [7]. A standard quantification of error introduced in the process, such as RMSE (Round Mean Square Error) was not available, but by means of manual testing the DEM resulted within the published standards of accuracy [Bushing, personal communication].

The preprocessing procedure started with the modification of the DEM attribute values to metric units, by re-projecting the coverage and multiplying the values by means of map algebra. The rectangular DEM had a floor altitude of 0 meters without bathymetry. By setting the DEM floor at 0.1 meters, the flat sea surface at 0 meters could be removed and the landscape could be clipped precisely at the coastline boundary. This permitted a reduction of the database size (subsequently propagated to the other layers, similarly clipped with a polygonal outline of the DEM) and a more precise definition of the study area. The numerical distinction of the sea-land boundary did not affect the accuracy of the digital coastline, as could be seen by comparing the opposite shores at Two Harbors with the other reference maps. The hydrologic correction of the DEM consisted in the detection and removal of sinks. The sink maps were indicative of the kind of production artifacts present in the DEM. In some cases, sinks appeared to follow closely the contour lines on the DEM, especially along steep slopes. This is probably due to the interpolation method which might have used higher order polynomial surfaces, resulting in very curved interpolated profiles, generating intermediate cells at lower altitude than the original, closely spaced contours.

5.2 Soil

The soil database had been assembled by the Center for Natural Areas in 1976, based upon the US Soil Conservation Service mapping in 1955. The resulting product was an ArcView3 format polygon shapefile with multiple attributes. These included *Permeability* (4 classes defined by numerical ranges of variable size), *Texture* (8 nominal classes based upon *Medium-Fine-Coarse* subdivision, plus relative quantity of gravel), *Erosion* (7 nominal classes such as *Gullying* and a range of intensity values from *Slight* to *Very Severe* erosion) and *Land Use* (8 nominal classes such as Cultivated, Urban and Scrub). The objective was to extract information on soil permeability, to be used in the runoff estimation, and on soil texture, for the RUSLE equation. Direct information on permeability was available in the soil coverage for approximately 85% of the island. The polygon coverage had to be first imported into Arc/INFO, and then converted to the

raster structure on which the rest of the model was based. This required the specification of the attribute of interest in the Polygon Attribute Table (PAT) in the INFO database, and then the generation of a raster layer using the geometry of the polygon coverage and the actual cell value for the specific attribute considered (in this case, permeability).

As the permeability attribute coverage was not complete, for the remaining 25% of the island the *Texture* attribute was used as a surrogate. After the same procedure of polygon coverage import and rasterization, a similar classification task had to be carried out. In this case, the 8 input classes were subdivided to fit the same 4-way classification of infiltration. While a spectrum of *Fine-Medium-Coarse* could have been seen as an analogy of low to very high permeability, the additional element of *Gravel* content complicated the problem. One of the risk was to establish an arbitrary hierarchy, by which, for example, *Medium Cobbly* was considered to be experiencing lower infiltration than plain *Coarse* soil. Another problem was to estimate the extent of polygons with explicit *Gravel* "tags", because the placing of one very popular "tagged" class into a level 2 infiltration instead of level 3 could have unnaturally unbalanced the dataset. In conclusion the *Medium* textures (4 classes) were split in two representing slow and moderate infiltration, while *Fine* and *Coarse* were placed at the extremes. The resulting distribution might be inconsistent with the previous classification using direct infiltration values. Permeability and texture polygons covered 95% of the island. For the remaining 5% a standard average infiltration value was used. A possibility could have been to use geology as a data source to infer the permeability of the overlying soil horizon, but this resulted objectively not feasible for the long succession of several assumptions required to link bedrock with specific surface characteristics. Possible gaps in the geology layer for specific classes could also have required additional coverages, so that the averaged background resulted to be a reasonable choice. The three grids, permeability, texture and background, were integrated into one layer with discrete values (in thousands) ranging from 1000 to 4000, as requested by the SCS-Runoff tables. The raster map of textural characteristics could be directly converted into the RUSLE *K* factor.

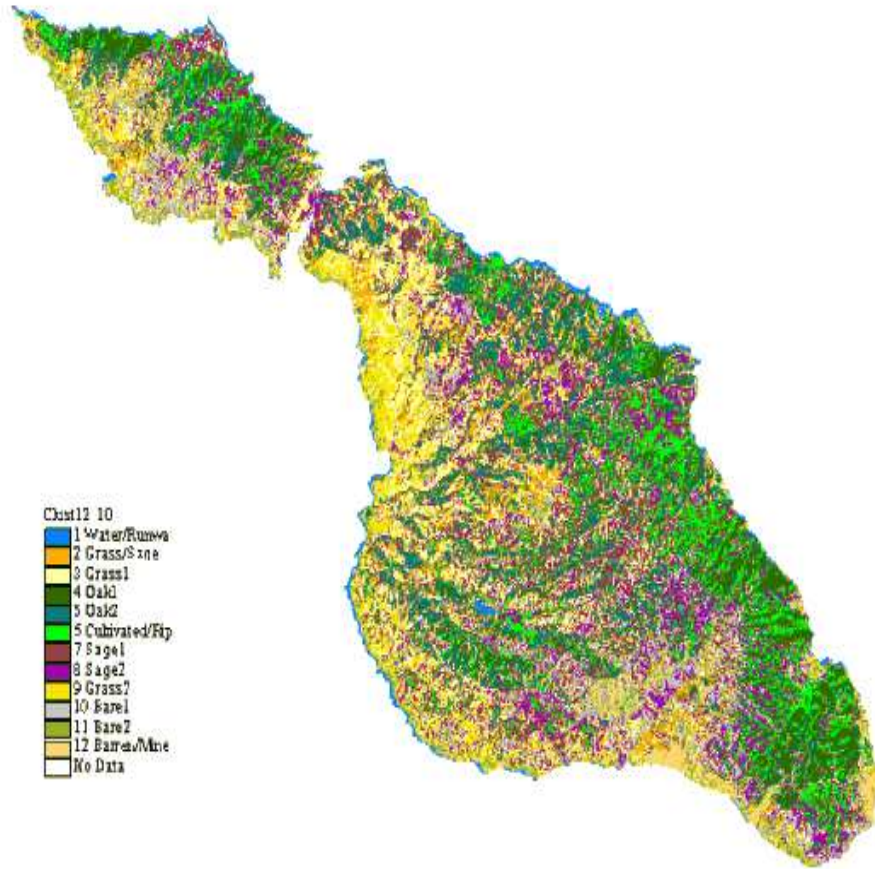


Figure 11: The cluster vegetation map. Sources: ESRI Arc/INFO and ArcView3 on SPOT multiband images.

5.3 Vegetation

The source of information for vegetation was a multiband (Infrared, Red and Green) satellite SPOT image of October 1993. Also, a vegetation coverage was provided by Catalina Conservancy in Arc/INFO readable format: it has been used first on the field as reference, and then as an alternative source in the classification of cover type.

The preprocessing of the source SPOT bitmaps consisted in removing the atmospheric filtering effect, importing the images into Arc/INFO environment and clipping the landscape with the vectorized outline of the DEM, so that the ocean could be

removed from further analyses. The image registration was carried out using the built-in Arc/INFO *CONTROLPOINTS* tool which permitted an interactive overlay of the source image with the reference coverages. The Road Network coverage (produced by the Catalina Conservancy using differential GPS) and the extracted stream network, were used as references. The operation consisted in linking through on-screen digitising the visible and distinguished features appearing in both source and reference coverages. The road network appeared to be the most useful reference, because the road cuts were sharper and more precise (only a few pixels wide) than the incisions of stream valleys. The control points were chosen mostly along the coastline, at road interruptions and sharp bends. The registration information was stored in 33 individual links, an adequate quantity considering that an increase in the number of links did not seem to affect the match between the coverages. The three SPOT images were rewarped to the new geometry by a Linear, Nearest Neighbour algorithm. Linearity proved to be more reliable than Quadratic or Cubic fits after testing the three options at the edges of the landscape along the coastline boundary north of Avalon. Nearest Neighbour resampling has been preferred to other schemes because it only shifts pixels without modifications in brightness, minimizing the differences with the original image [20].

An extra NDVI (Normalized Difference Vegetation Index) image was generated from the first three, using map algebra. The NDVI image is an arithmetic enhancement of the response of vegetation in Remote Sensing imagery, and it is based on the shape of the spectral response curve of vegetation across several emission bands [12].

The images (including NDVI) were grouped in a stack (see Figure 12), a logical construct of Arc/INFO that relates vertically different coverages on a cell to cell basis [7]. Each corresponding group of cells constitutes the so-called measurement vector, the unit of analysis for the multispectral classification schemes. A statistical cross comparison of grid vectors groups the pixels according to the chosen number of categories (assumed to contain normally distributed populations). The classifier algorithm must be trained to determine the actual statistical characteristics of the populations (mean and covariance). This may be carried out by supervised training, presenting to the computer

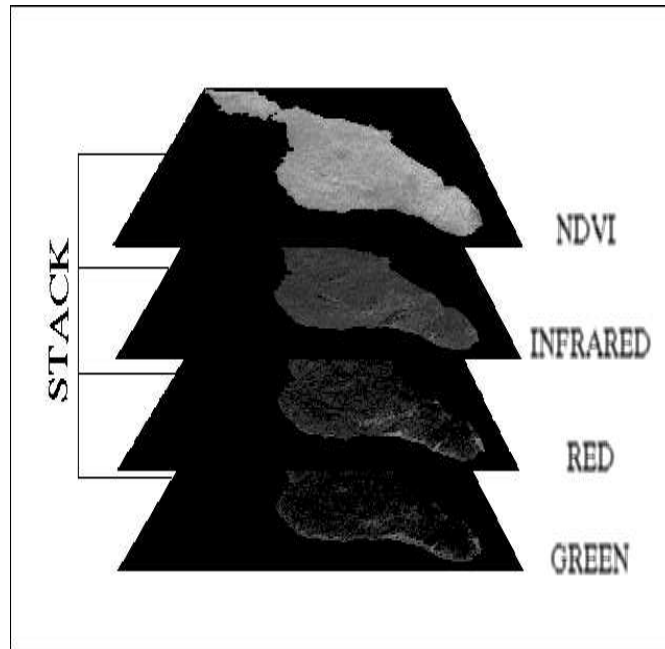


Figure 12: SPOT 3+1 bands stack

spectral signatures from a relatively small set of pixels of well known origin, or by unsupervised training, where the computer is allowed to select the population characteristics by performing a process called *Clustering*. The unsupervised method requires the user to specify the expected number of categories, while the computer selects the method and the selection parameters to partition the spectral signature space accordingly. After the partitioning, the individual clusters can be displayed, and it is possible to link each cover type to the corresponding cluster [10].

Several options for the number of clusters were attempted (4, 7, 12 and 15), but 12 was considered as the most suitable for the reconstruction of the vegetation cover over the island, balancing resolution with accuracy and readability of cluster maps. Each cluster was then assigned to a specific cover type, using all the ancillary information available (paper maps, digital vegetation map, field pictures). The assigned classes were 12 in total and in particular *Water*, *Grass* (3 subtypes), *Oak* (2 subtypes), *Cultivated*, *Sage* (2 subtypes), *Bare* (2 subtypes) and *Barren/Mine*. The very specific cover types

like *Water*, *Cultivated* and *Barren/Mine* were recorded with high spatial precision, identifying small reservoirs, narrow fields and delimited land-use types signalled on paper maps and written notes from fieldwork. The other classes were somewhat less defined, and the great internal variability, already experienced on the field, resulted in the proliferation of cover subclasses with mixed or intermediate attributes. The reclassification of the vegetation map obtained from clustering was carried out using the standard tables for vegetative covers by the Soil Conservation Service, published in [6].

5.4 Precipitation

The process for the generation of rainfall surfaces is illustrated in Figure 13. The sources of precipitation data consisted in tabular values of the annual precipitation values for four stations on Santa Catalina (Avalon, Two Harbours, Middle Ranch, Airport) and daily totals limited to the city of Avalon, available in [3].

Both data sources have been imported into a spreadsheet environment. A discrete time-series precipitation chart for the city of Avalon could be plotted out, so that major storms events were highlighted as distinguished pulses against the rainfall background. Four major storms could be isolated, and a preliminary study of the hydrograph shapes hinted to the character of the oceanic storm fronts passing over the island: sudden bursts of rainfall, without preliminary precipitation build-ups and quick fading after the peaks of the perturbation. Two statistical analyses were carried out on this data. The first was a Weibull Recurrence Interval (RI), for which the major storm events were ordered in decreasing order of magnitude, and the respective recurrence interval was derived from the formula

$$\text{RI (years)} = \frac{n + 1}{m} \quad (11)$$

where n is the number of years on record and m is the magnitude rank. The second method was based on the Log Pearson distribution. The daily precipitation data was aggregated to form yearly totals, by subdividing the year according to the wet season from September to August. The values were then processed by extracting logarithms

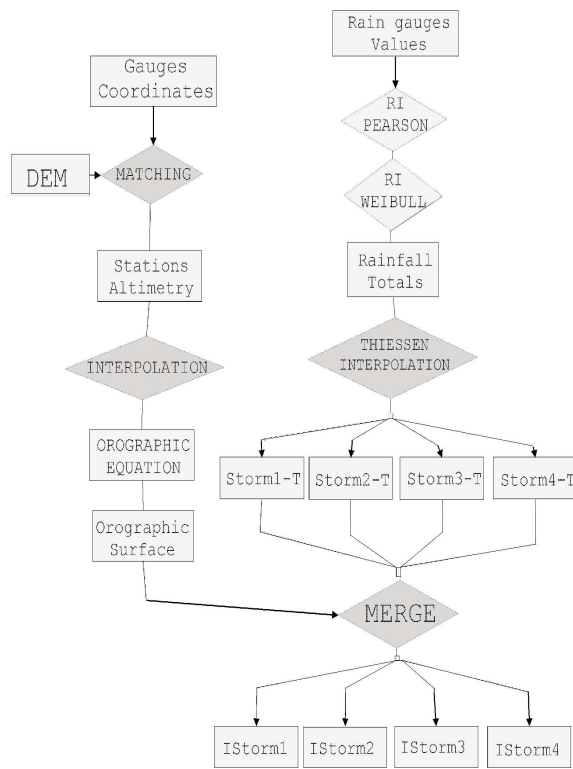


Figure 13: Procedure diagram for the generation of the precipitation surfaces.

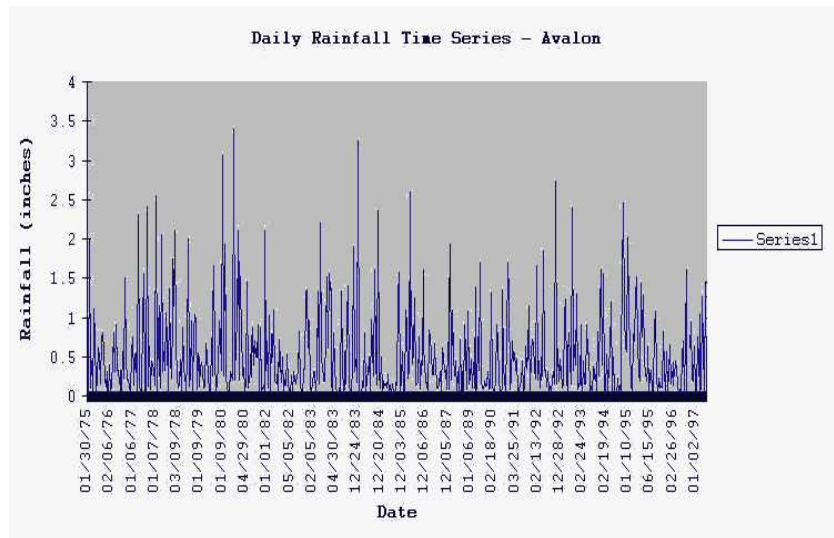


Figure 14: Time series graph for rainfall daily totals, Avalon

and evaluating the resulting distribution in terms of variance and skew factors. The final result was a magnitude ranking that permitted to identify a function relating Recurrence Interval (RI) to expected total rainfall [14]. The precipitation totals, expected from RIs of 2, 5 and 10 years, and the maximum experienced storm, were then deducted from the Log Pearson plot (Figure 15). Comparing these values to the daily time-series, the four selected storms could be identified and extracted as data to be employed in the model.

Using the co-ordinates provided by the Catalina Conservancy, the altitude of the four gauging stations could be determined from the DEM. A fifth station, El Rancho Escondido, had been discarded due to partial or inconsistent data. Each annual series was compared to the respective altitude in order to derive an equation linking altimetry with precipitation totals. The altitude versus rainfall plots for the four stations suggested consistently different equations, and a simple averaging of the linear parameters did not seem sufficient. Therefore a general equation, determined in previous studies, has been used [3]

$$y = 0.0036x + 12.405 \tag{12}$$

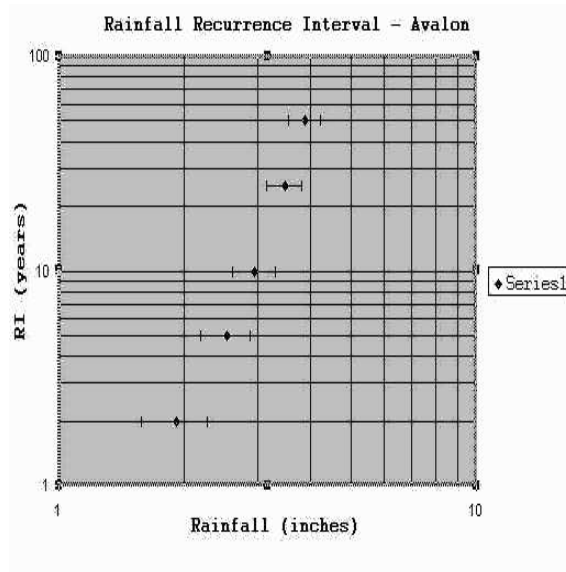


Figure 15: Log Pearson distribution RI graph for Avalon

where y is the total annual rainfall in inches, and x is the altitude in meters. Theoretical models have investigated and formalized the orographic effect upon precipitation caused by relief and topographic barriers. Using the modelling variables of p (storm inflow depth), v (velocity of storm front) and w (water content), the actual precipitation intensity appeared as a linear function of the differentials of these variables. In qualitative terms, the storm front approaching an idealised landscape profile presents in succession initial depletion of w at the first topographic barrier, then convergence and intensification progressing with air mass rise, an intensity peak close to the top of highest relief and a sudden decay characterised by rainfall spillovers [35].

This simplified model may be applied to Santa Catalina, even if the island is not symmetrical along a central ridge and presents internal topographic variability. The gauging stations result located in the initial depletion zone (Avalon, Two Harbor), in the intensification zone (Middle Ranch) and the peak intensity zone (Airport). They then provide a spatially minimal sampling coverage for capturing the orographic effect process as formulated before. The direction of approach of the storm fronts is variable, presenting a preferred WNW prevailing trend (Westerling), NW (Winter), NE (Santa

Ana) and S (Chubasco)[3]. An island-wide aspect component of the rainshadowing effect is probably not dominant in determining rainfall characteristics (a difference in interior versus exterior locations is probably more important). The numerical resolution of the given equation is probably too high compared to accuracy, while spatially it is unlikely to be more accurate than the general subdivision in three precipitation zones as suggested by the theoretical model.

The four stations provided the points for the generation of rainfall surfaces. Using the dates of the four storms previously derived from Avalon time series and RI calculations, it was possible to select the storm daily totals from the four stations for each given event (the gauging stations, beside annual values, provided also data for specific important events). The storm daily values were aggregated and averaged to obtain one representing value for each station and each storm. The interpolation scheme chosen was Thiessen interpolation, because the limited number of data points was not suitable for Kriging and Inverse Distance Weighting (IDW) methods. Moreover, Thiessen polygons were originally designed for precipitation data interpolation and so it seemed a proper case for application. The final result was a rainfall surface for each storm.

Using the orographic equation combined with the altimetric data of the DEM, another rainfall surface could be interpolated and considered in the study. By means of map algebra, a surface constituted by the expected average annual precipitation could be derived for each cell of the DEM. The resolution of the resulting surface (20 meters as for the DEM) was then much higher than the other polygonal Thiessen surfaces.

The integration of these two different datasets, the Thiessen surfaces of daily totals on one hand, and the orographic precipitation surface of yearly totals on the other hand, required the development of a special approach. The objective was to base the final rainfall surface on the Thiessen surfaces and to include some of the information stored in the orographic surface. The problem to overcome was that the former contained storm totals, while the latter yearly totals. Both the Thiessen and the orographic surfaces were then normalized to real values from 0 to 1, so that they represented only the relative indices of precipitation distribution on the landscape. Then the differences

(Δ , calculated by subtraction with map algebra) between the two normalized surfaces were computed. A normalized surface intermediate between the two source surfaces could be defined by adding a percentage of Δ (determined by a real weighting factor k , ranging from 0 to 1) to the original Thiessen surface. Finally, the output rainfall surface was obtained by de-normalising the intermediate surfaces using the original normalising parameters used for the Thiessen surface, so that the output units were still daily storm totals. For example, for Storm1T the range of precipitation values (read from the raster diagnostic data) was from 2.07 to 3.60 inches. The normalized Storm1TN was computed by

$$\text{Storm1TN} = \frac{\text{Storm1T} - 2.07}{3.60 - 2.07} \quad (13)$$

The normalized orographic surface was computed similarly, considering extremes of 12.416 and 35.448 inches

$$\text{OrographicN} = \frac{\text{Orographic} - 12.416}{35.448 - 12.416} \quad (14)$$

The Δ was calculated by difference between the two grids

$$\Delta = \text{Storm1TN} - \text{OrographicN} \quad (15)$$

The intermediate normalized surface was generated by

$$\text{IntermediateN} = k\Delta + \text{OrographicN} \quad (16)$$

The factor k can be seen as a control variable that can be regulated according to the relative confidence with the two datasets to be integrated. With $k = 1$, the IntermediateN surface is equal to the Thiessen Storm1TN surface; with $k = 0$, it is equal to OrographicN. Intermediate k values result in a linear interpolation between the source surfaces. In this case, a value of $k = 0.6$ is used to indicate a preference towards the Thiessen surfaces and a subordinate contribution of the orographic effects to the final totals. Finally, the IntermediateN surface must be de-normalized into the original units

of Storm1T, using the same extreme values as coefficients and with minimal reworking of the equation

$$\text{Storm1Final} = (3.60 - 2.07) \times \text{IntermediateN} + 2.07 \quad (17)$$

The normalisation procedure introduced two enhancements in the overall precipitation information included in the model. First, it increased the resolution to 20 meters, maintaining an alignment with the rest of the datasets, and incremented the accuracy, since the sources included in the estimates are now altimetry, yearly totals and daily totals. Second, the storm surfaces have been spatially corrected so that they comprise indirect theoretical assumptions (the orographic effect) as well as direct measurements. The rainfall surfaces have been converted to R factors by computing the 30-minutes precipitation intensity values from precipitation totals and the length of the storm.

5.5 Fieldwork component

Another source of information was the three-days fieldwork session carried out on Santa Catalina Island in collaboration with the Catalina Conservancy prior to the actual implementation of the model. It consisted in a thorough reconnaissance of the island along primary and secondary roads. The main interest was placed in the characteristics and distribution of vegetation, which would have resulted important in the successive classification of cover types from RS maps. Another interest consisted in the observation of soil types and widespread erosional features. The collected data was mainly digital photographs, supported by field sketches and additional observations, including approximate locations and camera bearings for a later reconstruction of the observations.

After extensive observation of the various vegetative features of the island, a certain lack of consistency of the source maps emerged in some areas. In particular it could not be found the correlative of *remnant* (*Oak*, *Sage*) terminology used in the map, which was responsible for distinctive subgrouping. A great variability internal to most vegetation categories was also observed, presenting for example extensive grassland patches in areas classified as woodland, or bare eroded soil instead of grasses. The rela-



Figure 16: The aspect component of vegetation as taken from the field: the more vegetated slopes are facing north.

tive inadequacy of the map justified the convenience of a Remote Sensing approach for developing the vegetation map. From an hydrologic point of view, the variability of vegetative factors controlling interception and infiltration within the same class appeared to be unsuitable for broad distinctions such as the ones imposed by the conversion tables. For example the riparian oaks presented an high undergrowth layer, which was not as much developed in the north western part of the island, even if classified in the same way in the map and having the same spectral signature. Erosion processes appeared to be dominated by rill and gullies formations, which are not considered in the current implementation of USLE (also the 20 meters cell resolution is well before the threshold of detection of such microtopographic features). With respect to the DEM, the presence of reservoirs damming stream flow might alter locally the hydrologic significance of topography. A preliminary map of extracted channel networks was used on the field, and critical locations as drainage divides or stream convergences were verified. For example, the quadruple joint convergence point near Little Harbor was compared to the extracted network and verified in accuracy. The threshold of distinction between colluvial hollows and actual channels was compared on the short streams on steep slopes at Two Harbors, and roughly confirmed the accumulation threshold of 50 represented on the map.

The digital pictures were later organised into an ArcView coverage which included the Catalina DEM, the Road network and an additional vector layer which linked the pictures to specific locations, so that they could be retrieved in a hypertextual manner. A modification in the Avenue linking script allowed for dynamic resizing of the bitmaps which could be displayed in the original dimensions.

6 Results

The Channel Network Extraction process, in conjunction with watershed delineation procedures, have produced the stream networks and watershed subdivision illustrated in Figure 17. The summarizing morphometric statistics are presented in the graphs in Figure 18. The more developed watersheds of order 4 and 5 are located in the central part of the island, and in the south-western area close to Avalon. The main stream networks converge approximately towards Little Harbor, and drain in close proximity in the western coastline. The northern peninsula presents small but well developed 4th order streams, while the eastern and southern coastline of the island, characterised by steep slopes and drainage divides in close proximity to the ocean, presents small, closely spaced 3rd (or lower) order watersheds. The statistics present regular exponential relationships between stream order and number of streams (Figure 18.1), average stream length (Figure 18.2) and sum length of streams (Figure 18.3). The average stream gradient is instead linearly dependent with stream order (Figure 18.4). In the literature, the progression of stream length and number of links according to stream order is generalized as geometrical [11]. The aggregation of data from many different networks might modify the validity of the generalization. In the case of Santa Catalina, exponential relationships fit very closely the data. The exponents of the exponential equations are indicative of the specific drainage characteristics. In the stream number relation, -1.7 denotes an extremely rapid decline of number as order increases (and therefore a dominance of lower order streams in drainage). In sum length, -0.98 is a more moderate exponent, which compared to the previous one indicates that the higher number of lower order streams does not determine the same dominance in length on the drainage system

as implied by number alone (i.e. lower order streams are many but comparatively short). 0.76 for the average length indicates a somewhat slower increase in length than expected from sum length estimates, expressing a certain degree of compression of the drainage networks which develop higher order interconnections in a limited space, also visible in Figure 17. It must be considered that the characteristics of the extracted networks vary according to the threshold accumulation values. The employed threshold of 50 produces many low order streams, while higher thresholds (for example 300) are likely to show a more compressed order spectrum (4 orders maximum) and the morphometric characteristics would be scaled accordingly.

The vegetation map produced from cluster analysis and classification (Figure 11) can be considered as a result on its own. Compared to the other digital vegetation map and to direct observation on the field, it appears to be much more precise in resolving the variability of vegetation and internal dishomogeneities, as the other source presented unrealistic extended zones with uniform vegetation. It is also able to capture natural uniformities such as the aspect-based difference in the northern peninsula: the south facing slopes with *Barren/Grass* cover versus wooded north facing slopes, feature also visible throughout the island (see Figure 16), the delineation of riparian corridors along streams and the predominantly forested east side of the island. The cover classification could be refined with further validation on the field, because in the current study it has been refined predominantly using the other source map.

Unfortunately, no result maps for runoff and erosion are available at this stage. Nevertheless, the critical framework proposed in this document and summarised in the next section provides general criteria for interpreting the accuracy of model results.

7 Critique of the model

A critique should be addressed mainly to the algorithmic components of the model, because they are responsible for representing and simulating the physical processes under study. The USLE equation had originally been developed for application in agricultural contexts, and the various parameters are all directly linked to agricultural entities like

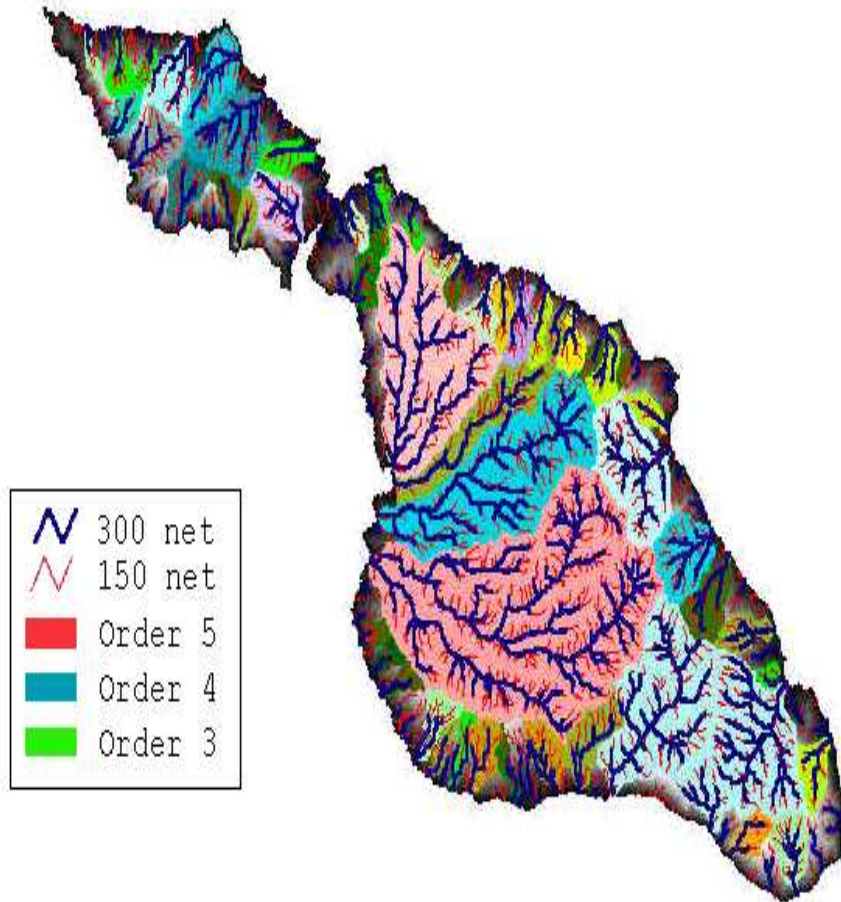


Figure 17: Extracted channel networks at 300 and 150 threshold accumulation values, and extracted watersheds of order 3, 4 and 5. Source: ESRI ArcView3, ESRI Arc/INFO and original Catalina DEM from Catalina Conservancy.

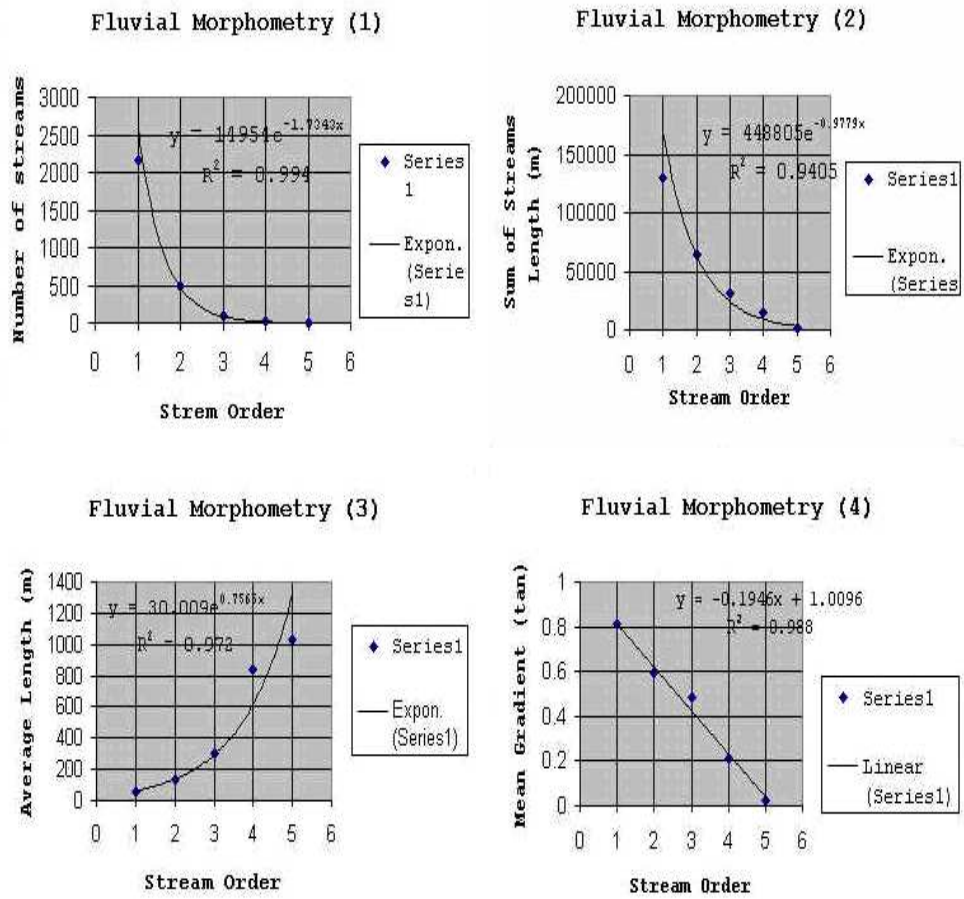


Figure 18: Morphometric statistics of Catalina extracted drainage network.

crop types and practices. As a result, the values to be inserted in the USLE equation must be referred and compared to the base set of agricultural parameters; for example the P factor is an expression of specific agricultural practices of contouring and terracing, and does not appear to have a more general meaning. When the method is applied to geomorphic studies, the process of conversion and adaptation might reduce the validity of the results. Spatially, USLE, even if defined as "Universal", is built upon a database which is restricted to the United States east of the Rocky Mountains. Moreover the base is further restricted to a range of slopes where cultivation is feasible, from 0% to 7% [8]. The diagram proposed in [6] expands the range from 3% to 20% (11 degrees). In the the slope map (see Figure 19) approximately 80% of the island results beyond the extended ranges of applicability. The implied extensive extrapolation is a factor to be considered. It is also a problem of scale, because USLE was designed for a single field and at the hillslope scale here considered, it does not take in account the sediment deposition at the base of the slope [17]. A more conceptual problem derives from the independent status of rainfall and soil factors in the USLE equation, which contradicts the relationship existing in reality between the permeability of the soil and the erosive effect of precipitation [15]. The calculation of single parameters implies a large degree of uncertainty, such as in the case of the K factor when it is not specifically defined for the given soil type.

For all of these reasons USLE should not be applied beyond the base set of conditions for which it was originally devised [17] [15][24][8]. Nonetheless, the vast database and the promised effective prediction would be advantageous gain if USLE could be extended to other environments. An assessment of the applicability of USLE to geomorphic studies is given in [33]. The tests reported in the research paper presented a difference of two orders of magnitude between measured soil loss and RUSLE predicted soil loss on reclaimed and natural hillslopes in non-extreme conditions. The original tests conducted by Wischmeier in 1976, based on 189 agricultural plots, resulted in an error level of 10%, but they were conducted in the ideal setting for USLE. The official position of the US Soil Conservation Service is extremely supportive of the validity of USLE, but does not

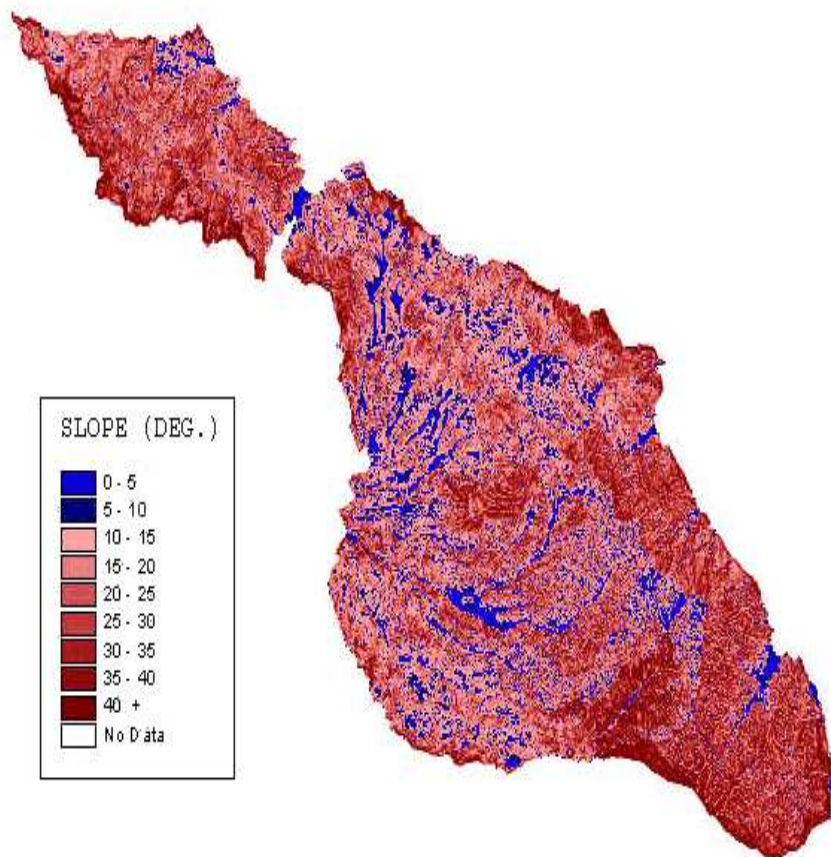


Figure 19: Slope map of Santa Catalina. The slopes shaded in blue represent the areas within the range of RUSLE base plots. Source: ESRI ArcView3 and Catalina DEM from Catalina Conservancy.

Factor	MIN	MAX	Variability
K	0	0.5	10
L	0	5 (For 1 km)	10
S	0.065	31.35 (For 60 deg.)	10^3
C	0.003	0.45	10^2
P	0.1	0.90	10

Table 1: Order of variability of RUSLE parameters

give indications of applicability or official estimations of error. Interestingly it seems to suggest to consider predicted soil loss data only in a relative sense as a broad indication for the improvement of agricultural practice [33], and this might imply that the accuracy of the results is much lower than the resolution obtainable from the equation.

In the limited context of the Catalina model, the application of USLE might be summarised by considering that S factor results from risky extrapolation in 80% of the cases; that K factor introduces an error range of approximately 50% when using the nomograph; that the C factor is obtained from tables adaptable only with difficulty with general source data; and finally P factor has no defined sense in a non-agricultural contexts. The overall composite error might be of the order of 50%, and therefore the results have only a broad validity in terms of general spatial distribution.

Table 1 represents each RUSLE factor separately with minimum and maximum values (derived from tables and formulas) and relative oscillation bands expressed in terms of order of magnitude. The only factor omitted from the table is R , which depends on precipitation intensity of relevant storm events and also varies greatly accordingly to the method of calculation employed [24]. Approximately, a variability of 10 is considered realistic. The slope factor S is by large the one which introduces the greatest variability in the results. The others, taken individually, are somewhat subordinated to the role of modifying the slope factor values. According to variability, factors can then be grouped in a hierarchical order, determining a order 3 variation class (S), a secondary order 2 variation class (C) and four tertiary order 1 variation classes (R , K , L and P). It must

be considered that the factors are multiplied and not summed together, and therefore the order 3 dominance of S should be compared to the composite weight of the other factors, equivalent to 6. Also each factor has the potential capacity, derived from the relation based on multiplication, of imposing drastical changes to the final equation values. On the other hand, while slope is consistently mapped throughout the landscape as a continuous function of topography, the other factors are spatially discrete or less regular (as seen from the technical specifications of the coverages), and they might tend to reciprocally compensate their relative differences, and result of subordinated importance at the large scale. The ranges of values in Table 1 are also based on the extremes read from the general SCS tables, and on real landscapes they might result narrower. Their relative influence would then be reduced even more when compared to the S factor, which covers the entire order 3 variability range within moderated slope values. Also, considering the ranges in Table 1, there is not the possibility of a sudden zeroing of the erosion values, caused by the zeroing of a single factor (the only exception seems to be K , but this is only due to the particular representation of boundary values in the nomograph in Figure 10). Therefore, the erosion patterns traced by slope are never completely disrupted. The final outcome of the dominance of the S factor is a certain similarity between slope maps and correspondent erosion maps. If the final erosion maps, because of validity and accuracy constraints, are further degraded to ordinal (*High/Medium/Low*) level, the secondary and tertiary factors become even more marginal, because internal micro variability is eliminated by the coarse regrouping of values. The consistency of the slope factor would emerge visually and numerically. Therefore the slope map in Figure 19 has a comparable amount of spatial information about erosion as a degraded erosion map. For numerical estimates the full application of the model is required, even if previous considerations about numerical accuracy of the model would then apply.

The SCS Runoff method depends heavily on the validity of the CN curve numbers, and the same problem of classification of RUSLE factors applies. SCS runoff produces useful information relative to spatial distributions and generalised rainfall/runoff ratios,

but the accuracy of the results is not the one implied by the resolution of the data used in the model, or by the numerical precision of runoff estimates. The Channel Network Extraction process relies on the algorithms of Arc/INFO, but the assessment of validity can be approached by considering the general principles of cell based flow and channel definition and it has been carried out elsewhere [27] [32].

The transferability of this modelling system to other locations, such as to the other major Channel Islands (Santa Rosa and Santa Cruz) is difficult to assess. Certainly, if the model has been considered applicable to Santa Catalina, the transition to other islands, from the model perspective, would be limited to a change of sources (topography, soil, precipitation data, SPOT images), with the only limitation that the same infrastructure present on Catalina would not be available on other islands (there are less rain gauges on Santa Cruz, for example) with consequences on model reliability. The pulsed precipitation inputs characterise all the Channel Islands and justify the use of a model not based on saturated conditions and groundwater balance. The size of the island under study, and the DEM array size, should be adequate for the effectiveness of the channel extraction process. The slope ranges and the cover types would still be located outside the ranges of confidence of USLE applicability, and all the other problems would apply as well.

8 Conclusions

The modelling system employed in this study offers an example of how different digital data sources can be integrated in a single distributed database to run hydrologic simulations and obtain runoff and erosion estimates. The construction of the model has been traced from the initial abstraction from the base hillslope system, to the formulation of the model, and finally to the actual implementation. The necessary simplification included the limitation of runoff processes to Hortonian overland type only, and the use of empirical submodels (SCS Runoff and RUSLE) which employ generalising indexes of limited versatility instead of lower level physical conceptualisations of processes. Several different techniques have been used to gather and relate the necessary information,

ranging from cluster analysis on RS imagery, to the interpolation and integration of rainfall tabular data, to the conversion of the digital soil map and the validation of the data from field observations. Each independent data layer produced results such as the hydrologically corrected DEM, the RS-sourced vegetation map, and the extracted channel networks with delineated watersheds. These descriptive components contribute to trace a first analytic framework for the hydrology of Santa Catalina, from which additional modelling can be carried out. The RUSLE and SCS Runoff modules, considered in some detail, required the preprocessed data to pass through the filter of conversion tables. This step has been identified to imply the use of USLE instead of RUSLE equation in the calculations, and the introduction of a certain degree of mismatch between the original landscape representation and the internal representation. The application of RUSLE in the considered environment implies risky extrapolations, mainly due to its specialization in agricultural problems, even if no precise boundaries to its applicability could be found in the literature. The results of SCS-Runoff and RUSLE, while valuable for distributions and general spatial patterns, should be taken with the necessary attention, considering all the different procedures which this document has contributed to map.

9 Acknowledgements

The Catalina Project, constituting the core of this dissertation, has been carried out in collaboration with many individuals. Thanks to Dr. Leal A.K. Mertes for offering the opportunity of working on the project, for the concession of the SPOT images, for the digital camera used on the field, and for general guidance in the initial parts of the implementation. Thanks to Christine Martella for her collaboration on the field, in the interpretation of cluster maps and for additional remote sensing analysis which unfortunately could not be included in this document. Thanks to Dr William Bushing, Frank Starkey and Peter Schuyler of the Catalina Conservancy, for all sorts of data and information (vegetation, soil and roads map, and orographic equation for precipitation), for the kind hospitality on the beautiful island, and for the professional assistance in

the exciting exploration of the island, especially along washed out roads in the interior. Thanks to Dr A. Morrison, Dr J. Drummond and the IT Technicians of the Department of Geography and Topographic Science, University of Glasgow, for facilitating the technical transition of the project from the Department of Geography, University of California at Santa Barbara. Thanks to my family for the moral support during my extended stay in California and throughout the development of the dissertation. In the typesetting of the document, the hints about LaTeX commands received from my brother Sergio and Sebastiano Vigna have been helpful.

References

- [1] Alonso, M., Finn, E.J. (1992) *Physics*, Addison-Wesley
- [2] Beven, K., Kirkby, M.J. (1993) *Channel Network Hydrology*, John Wiley and Sons Ltd.
- [3] Bushing, W. Santa Catalina Island Web Site <http://www.catalinas.net/seer>.
- [4] Chorley, R.J., Hagget, P. (1967) *Models in Geography*, Methuen & co. LTD.
- [5] Doebelin, E.O. (1976) *System Modeling and Response*, John Wiley & Sons.
- [6] Dunne, T., Leopold, L. B. (1978) *Water in environmental planning*, W.H. Freeman and Company.
- [7] ESRI (1990) *Help for Cell-Based Modeling Arc/INFO v7* On Line manual.
- [8] Gerrard, A.J. (1981) *Soils and Landforms*, George Allen & Unwin.
- [9] Goossens, M., Mittelbach, F., Samarin, A. (1994) *The LaTeX companion*, Addison-Wesley Publishing Company.
- [10] Hord, R.M. (1982) *Digital Image Processing of Remotely Sensed Data*, Academic Press.
- [11] Ingle, S. and Stopp (1984) *The river basin*, Cambridge University Press
- [12] Jensen, J.R. (1994) *Introductory Digital Image Processing: A Remote Sensing perspective*, Prentice Hall Series in Geographic Information Science.
- [13] Kaposi, A, Myers, M. (1994) *Models and Measures*, Springer-Verlac.
- [14] Keller, E. (1998) *G144 Rivers*, Department of Geography, University of California at Santa Barbara, Santa Barbara (CA) USA.
- [15] Kirkby, M. J., Morgan, R. P. C., (1980) *Soil Erosion*, John Wiley & Sons Ltd.

- [16] Lindsay, K.A. and Nimmo, J.J.C. (1994) *A First Guide to LaTeX* issued in Mathematical Modelling 2M, Department of Mathematics, University of Glasgow.
- [17] Lorup, J.K., Styczen, M. *Soil erosion Modelling* in Abbot, M. B., Refsgaard, J.C. (1996) *Distributed Hydrological Modelling*, Kluwer Academic Publishers.
- [18] Mather, P. M. (1993) *Geographical Information Handling - Research and Applications*, John Wiley and Sons Ltd.
- [19] Matko, D., Zupancic, B., Karba, R. (1992) *Simulation and modeling of continuous systems*.
- [20] Meijerink, A.M.J., De Brouwer, H.A.M., Mannaerts, C.M., Valenzuela, C. (1994) *Introduction to the use of Geographic Information Systems for practical hydrology*, The International Institute for Aerospace Survey and Earth Sciences (ITC).
- [21] Menner, W. A. (1995) *Introduction to Modeling and Simulation*, John Hopkins APL Technical Digest, Volume 16, Number 1.
- [22] Mertes, L.A.K. *Computational methods for watershed analysis* Course G151/251, Department of Geography, University of California at Santa Barbara, Santa Barbara (CA) USA.
- [23] Minshull, R. (1975) *An introduction to Models in Geography*, Longman.
- [24] Morgan, R.P.C. (1995) *Soil Erosion & Conservation (2nd edition)*, Longman
- [25] Morgan, R.P.C. (1996) in *Advances in Hillslope Processes vol 1*, Anderson M. G., Brooks, S.M., 1996 John Wiley & Sons.
- [26] North Carolina State University, *Hydraulic Engineering Web Site* [On-Line] <http://courses.ncsu.edu/classes/nr300/www/lecture/runoff-scs.htm>
- [27] Quinn, P., Beve, K., Chevallier, P., Planchon, O. (1992) *The prediction of hillslope flow for distributed hydrological modelling using digital terrain models*, in Beven,

- K.J. and Moore I.D. *Terrain Analysis and Distributed Modelling in Hydrology*, John Wiley & Sons.
- [28] Refsgaard, J.C. *Terminology, modelling, protocol and classification of hydrological model codes* in Abbot, M. B., Refsgaard, J.C. (1996) *Distributed Hydrological Modelling*, Kluwer Academic Publishers.
- [29] Renard, K.G. *et al.* (1991) *RUSLE Revised Universal Soil Loss Equation*, Journal of Soil and Water Conservation.
- [30] Richards, K. *et al.* (1996) *An integrated approach to modelling hydrology and water quality in glacierized catchments*, Hydrological Processes vol.10.
- [31] Singh, V.P. and Fiorentino, M. (1996) *Geographical Information systems in Hydrology*, Kluwer Academic Publishers.
- [32] Tarboton, D.G., Bras, R.L., Rodriguez-Iturbe, I. (1992) *On the extraction of channel networks from digital elevation data*, in Beven, K.J. and Moore I.D. *Terrain Analysis and Distributed Modelling in Hydrology*, John Wiley & Sons.
- [33] Toy, T.J. and Osterkamp, W.R., (1995) *The applicability of RUSLE to geomorphic studies*, Journal of Soil and Water Conservation.
- [34] Watt, D.A. (1996) *Structure Diagrams in CS2A Software Engineering* Department of Computing Science, University of Glasgow.
- [35] Wisener, C.J. (1970) *Hydrometry*, Chapman and Hall Ltd.