

Using a logistic regression model to delineate channel network in southeast Australia

X.Y. Sun^{a,b}, C.J. Thompson^a and B.F.W. Croke^{a,b}

^a *Integrated Catchment Assessment and Management Centre (iCAM), Fenner School of Environment and Society (FSES), The Australian National University, Canberra
ACT 0200, Australia*

^b *Mathematical Sciences Institute (MSI), The Australian National University, Canberra
ACT 0200, Australia*

Email: xiaoying.sun@anu.edu.au

Abstract: The potential to delineate the location along a slope at which channels initiate is important for understanding hydrologic and geomorphic processes governing headwater streams. Most work assumes a uniform input of precipitation across the catchment, and every cell would receive the same volume of water. In reality, sites at higher elevations receive more rainfall, and tend to have smaller contributing area and stream length. In this paper, a channel initiation point (CIP) model is developed. The CIP model estimates channel initialisation based on a logistic regression (LR) technique. An LR relationship is applied because of its flexibility in assumptions where a discrete variable can be considered. By incorporating the accumulated rainfall surface into the LR, resulting drainage areas reflect hydrologic and geomorphic influences on channel initiation. The study area is part of the Lower Cotter experimental catchment, a headwater alpine catchment located in the Brindabella region in south-eastern Australia. The aim is to test the capability of the CIP model in estimating the channel network, capturing channel heads and disconnected channels. The estimated channel network is compared to that obtained using a classical method on the basis of a constant area threshold. The CIP model performs well in identifying channel and non-channel cells while improving channel head localisation and extraction of channel continuity. Overall, the CIP model can be considered as a valid alternative to commonly-applied traditional methods for channel network extraction from Digital Elevation Models (DEMs), in addition to considering hydrologic impacts on channel initiation.

Keywords: *Disconnected channel, channel initiation, accumulated rainfall, logistic regression, digital elevation model*

1. INTRODUCTION

Accurate identification and mapping of headwater streams is essential for physically-based characterisation of hydrologic processes in small catchments (Melville and Martz, 2004). Channel network morphology consists of two principal components: a) the channel head or channel initiation point (CIP) and b) the continuous incised channel (Hooke, 2003). Identification of the degree of channel connectivity is important for understanding the linkages between river reaches, the influence of sediment sources on channel morphology and hydrologic mechanisms (Hooke, 2003). Small errors in interpreting channel continuity would have a significant impact on the channel length, stream order, and drainage density (Garbrecht and Martz, 2000). The drainage density is indicative of the texture of the area, providing a useful numerical measure of subcatchment dissection and runoff potential (Horton, 1945).

The modeled channel extents vary according to the accumulation threshold values defined for channel initiation. The traditional approach for channel delineation is to apply a topographic threshold from Digital Elevation Models (DEMs) to define the initiation point beyond which the channel network is chosen (Dietrich *et al.*, 1993; Tarboton, 1997). These approaches assume that the flow direction is only dependent on topographic features. In most situations, the physical location of channel heads is related to other factors such as soil properties, climatic environment and land use (Montgomery and Dietrich, 1989; Prosser and Abernethy, 1996). Therefore, a unique value of threshold cannot characterise all types of channel and does not necessarily predict the actual channel heads (Tarolli and Dalla Fontana, 2009; Passalacqua *et al.*, 2010). Recent studies consider specific geomorphic characteristics of topographic surface directly derived from DEMs to choose the most suitable threshold to map channels (Gallant and Wilson, 2000; Tarboton and Ames, 2001; Tarolli and Tarboton, 2006). However, these constant threshold methods do not consider the hydrology (Furbish *et al.*, 2009; Thommeret *et al.*, 2010).

Montgomery and Dietrich (1989) and Dietrich and Dunne (1993) found that the location of channel heads changes rapidly in response to increased runoff per unit area. However, these approaches assume a uniform input of precipitation across the catchment, and every cell would receive the same volume of water. If the rainfall input is not uniform, the sourcing of water units will vary across the landscape in distribution and quantity. Consequently, the input is independent from the general flow direction map and constitutes a separate layer which acts as an independent parameter for precipitation. Recent studies suggest the use of logistic regression (LR) can produce the most accurate channel networks (Heine *et al.*, 2004; Jaeger *et al.*, 2007). The logistic regression model (LRM) is applied in this paper because of its flexibility in assumptions where discrete variables can be incorporated (Tabachnick and Fidell, 1996).

The aim of this paper is to develop a CIP model based on LR, and to test model performance in extracting channels heads, channels, and detecting disconnected channels. The study area is part of the Lower Cotter experimental catchment which has rainfall, discharge and LiDAR data available. These existing data sets provided the opportunity to explore and test new analysis procedures for mapping channel networks

2. METHOD

2.1. Study Site

The study sites are located in an alpine setting on the Brindabella range (184 km²), south-eastern Australia (Fig. 1).

These watersheds are selected for three reasons:

- a) The availability of LiDAR DEM and detailed independent field based network location datasets, in order to show the accuracy of the proposed method for forested environment

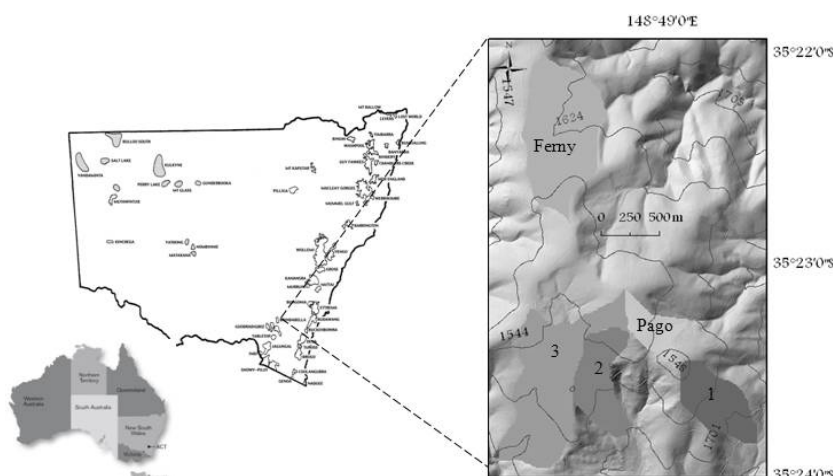


Figure 1. Location of the study area on the Lower Cotter experimental catchment

with a complex morphology.

- b) The geology is simple and homogenous, thereby minimising the influence of geologic variability on surface hydrology, while providing topographic variation in slope, curvature and aspect.
- c) Variation in rainfall due to orographic effect.

Pago catchment is the calibration site and it drains 0.15 km² of moderately sloped (15°) area, with elevation range from 1209 to 1368 m. Ferny and catchments 1, 2 and 3 are validation sites. They drain a combined area of 2.3 km², have an elevation range from 1024 to 1368 m and an average slope of 14.3°. Average annual precipitation is 980.4 mm for the study site. The LiDAR-derived DEM (1 m resolution) was resampled using the bilinear method to 10 m resolution for the CIP model.

2.2. Physical Settings

Terrain variables used to fit the CIP model were derived from the DEM for the study site. Rainfall data at 1-minute interval are available from 1971 to 1985, and were aggregated into average annual rainfall. The D∞ flow algorithm (Tarboton, 1997) was applied to the DEM, constituting an aggregated rainfall layer which acts as a parameter for precipitation. The rainfall surface assumes a rainfall accumulation direction based on the DEM. The accumulated rainfall surface was generated using:

$$AP_i = \frac{n \cdot AP_{i-1} + P_i}{n+1} \tag{1}$$

where AP_i = accumulated rainfall at receiving cell i , P = precipitation, and n is the number of cells draining to the upstream cell ($i-1$).

Where rainfall estimates are required for a single grid, values of the accumulated rainfall surface are first derived, and then used to generate rainfall sequences. The output table contains a row for each raster cell in the study site and a column for each of the variables of interest for use in statistical packages.

2.3. Model Description

The CIP model was based on the techniques described in Conrad *et al.* (2003) except they use the LR to identify where ephemeral flow begins and where it carries water in direct response to precipitation. The calibration process was carried out as shown in Fig. 2. To build the calibration data table for LR, physical parameters (Table 1) were sampled at each cell in the calibration site.

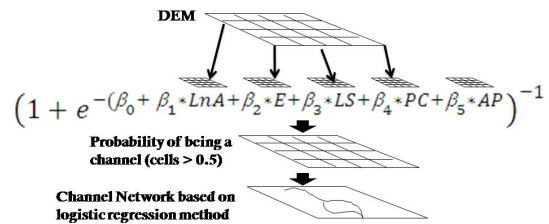


Figure 2. Diagram of a statistical spatial analysis procedure used to conduct the logistic regression

Table 1 Dependent and independent variables used in regression model predicting the occurrence of channels

Variable	Description
CI (1/0)	Presence/absence of a channel
LNA	Natural log of flow accumulation area (m ²)
E	Elevation (m)
LS	Local Slope (degree)
PC	Plan Curvature (m ⁻¹)
AP	Accumulated Rainfall (mm)

Note. Profile curvature was not included due to its high correlation with plan curvature.

2.4. Model Calibration

The LRM was applied to the DEM to predict the channels as absent (coded 0) or present (coded 1). For channel network delineation, only independent variables correlated with the response and found to be statistically significant ($P < 0.05$) were kept in the model. Parameters LNA and E could not be used to create the equation for determining the probability of channel initiation based on a 0.05 significance level. Therefore, the LRM was simplified where the parameter LS , PC , and AP are selected. Table 2 provides the calibration results of the CIP. PC was highly significant in the LRM. The negative coefficient on PC indicates that the concavity characteristic of valley cross sections is associated with channel incision. Overall, the LRM predicting the probability that a cell contains a channel performed very well (0.731) with the high -2 Log likelihood and chi-square statistic (Table 3). Pseudo R-Square U (uncertainty coefficient) indicates the CIP model predicted 58.6% of the variability in the variable.

Table 2 Results of logistic regression using presence of a channel as the dependent variable

Variable	Coefficient	Standard Error	Significance	exp(β)	95% Confidence Interval for exp(β)	
					Lower	Upper
Constant	72.726	22.596	<0.001			
LS	-0.380	0.049	<0.001	0.684	0.622	0.752
PC	-2.052	0.159	<0.001	0.128	0.094	0.176
AP	-0.065	0.020	<0.001	0.937	0.902	0.974

Table 3 Statistical summary of the CIP model

Significance	-2 Log Likelihood (Hosmer and Lemeshow, 1989)	Chi Square	Pseudo R-Square U
0.731	328.771	447.722	0.586

Accordingly, the CIP model applied for this paper is:

$$Z = 72.726 - 0.38 \times LS - 2.052 \times PC - 0.065 \times AP \quad (2)$$

The response function where is the estimated probability of channel presence:

$$\Pi = \frac{1}{1+e^{-Z}} \quad (3)$$

2.5. Field Surveys

Field surveyed data on channel head location and continuous channels for use in the CIP model were collected in summer of 2010 and winter of 2011. GPS points were collected in both drainage systems along selected channels. The consequent channel network depends on the decision rules and criteria for channel head identification (Drummond, 1974). This paper was based on a field survey which followed specific criteria to define the channel head and hence its location. In the published literature and government grey literature on streams, there is considerable variation in criteria used to describe the elements of the fluvial system. This paper distinguishes between headwater stream (stream head) and channel head which are often referred to interchangeably. The stream head (starting point of surface water expression) is difficult to predict, because it may extend up an unchanneled swale above a channel head by overland flow, or shift down as the water table shrinks when weather is dry. The channel head or CIP is the starting point of channel network (Montgomery and Dietrich, 1989).

This study applies the following criteria to determine the starting point of the channel network. These criteria also enable the channel network to be delineated by hydro-geomorphic modelling.

- a) The *Protection of the Environment Operations Act 1997* (Forestry Commission of NSW, 2011) defined the channel head as the upstream limit of concentrated flow that carries only water in direct response to precipitation and one or a combination of the following:
 - i. Bank height/channel depth > 0.03m depth
 - ii. Channel width at bankfull from 1 to 5 m
 - iii. Evidence of active erosion or deposition
- b) Continuous channel extending for at least 5 m downstream from the channel head (Jaeger *et al.*, 2007).

2.6. Model Performance Measures

For comparison, channel networks were also derived for the study sites by using a constant area threshold (CAT). The application of CAT was carried out using the TauDEM extension within ArcGIS (Tarboton, 1997).

Modelled channels are assessed for channel network accuracy against field-mapped channels and include seven accuracy measures that describe the channel network.

- a) The channel length was obtained by measuring all the drainage in a subcatchment.
- b) A top-down stream order system (Strahler, 1952) is used to classify stream segments based on the number of upstream tributaries.
- c) The drainage density was computed from total identified channels taken from GIS attribute tables and dividing them by the drainage area which was held constant for all networks of the catchment.
- d) The channel head is represented by the end at the streamlines extracted. In the case in which the field survey did not identify the channel head, that particular location was excluded.
- e) The distance from the field-surveyed channel head is defined as minimal distance that between the modelled channel head and the field-surveyed channel head.
- f) The number of cells that are accurately classified as channels or non-channels.

- g) The percentage of correctly identified channel and non-channel cells (errors of omission and commission).

3. RESULT AND DISCUSSION

3.1. Channel Network Extraction

The CAT used to define the drainage area at which channels were initiated was set at 6670 grid cells (0.07 km²) for the study site. Selection of this threshold value was assisted by comparisons of the resulting flow accumulation surface with the field-surveyed channel head in the calibration site.

The CIP model was used to predict the probability of channel presence for each cell of the DEM in the study area. The resultant data layer can be used to generate channels by designating a threshold probability as indicating the presence of channels.

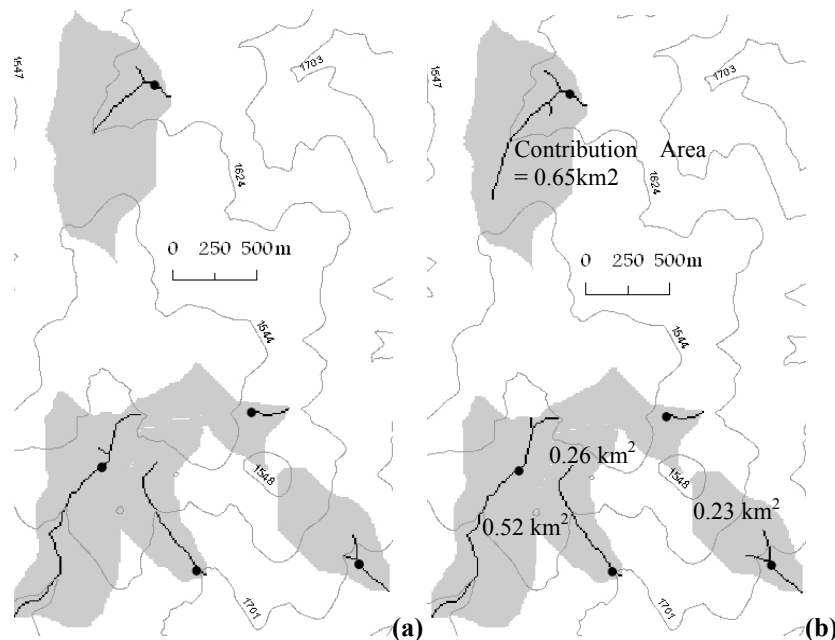


Figure 3. Channel network with channel heads observed in field (dot mark): (a) CIP model (b) TauDEM

A probability threshold of 0.5 (i.e. cells with values > 0.5 are considered part of the channel network) was selected to represent the modelled channel network. Converting the modelled channel raster cells into a topologically correct vector drainage network was completed by tracing the downstream cost path of each channel-designated cell to the outlet point. Fig. 3 illustrates channels identified using the CIP model compared with TauDEM derived channels.

3.2. Performance Measures

Table 2 shows five accuracy measures that describe the modelled channel network in the Ferny catchment.

Table 2 Predictors derived from the channel network extraction methods in Ferny catchment

Hydrologic Predictor	CAT	CIP
Total channel length (km)	1.71	1.13
Stream order	2	2
Drainage density (km/km ²)	1.68	1.11
(% error)	(+165)	(+75)
Number of channel head identified	3	2
Distance from the field-surveyed channel head (km)	0.78	0.50

Using a CAT overestimates total channel length and drainage density 165 %. The over-prediction in the validation site is due to the relative smaller contribution area (0.07 km²) calibrated in Pago catchment. In contrast, the CIP modelled channel show slightly lower overall length (Fig. 3). Therefore, the CIP model results in a more accurate identification of channel extent.

The CAT method is valid for mapping locations of higher-order (4th order) streams in large basins if sufficient resolution DEM is available (James and Hunt, 2010). However, the stream order was insensitive to the method in the study site.

The drainage density derived from the CAT method is greater than that from the CIP model. The CIP modelled channel network suggests that the delivery of water and sediment is less efficient than would be expected based on channel networks derived from the TauDEM. The CIP model appears to have the potential to improve on the CAT method, suggesting the channel initiation is likely to be influenced by hydrology and geomorphology in the catchment.

In TauDEM, the channel was traced continuously downstream from the initiation point with no possibility of detecting the presence of discontinuity. The CIP model decides which channels to trace based on the presence of a channel in each cell (Eq. 3). This suggests that the chosen thresholding criteria serve the purpose of identifying disconnected channels.

The large distances between the field surveyed and the modelled channel initiation points mean that the area thresholding criteria are geomorphologically inappropriate for the study site. This result suggests that a unique value of threshold cannot characterise all types of channel and does not necessarily predict the actual channel heads (Tarolli and Dalla Fontana, 2009). In these test catchments, the use of LRM can produce more accurate channel networks. The CIP model indicates that local slope is statistically significant because the proportion of precipitation that occurs as surface runoff is positively correlated with slope on the landscape.

An error analysis for each of the methods is provided by the percentage of correctly identified channel and hillslope cells and errors of omission and commission (Table 3).

Table 3 Errors of omission and commission from grid cell counts for the study site

Error Predictor	CAT	CIP
Channel: Actual channel cells correctly identified (%)	82.9	100
Commission: Actual channel cells incorrectly identified (%)	17.1	0
Hillslope: non-channel cells correctly identified (%)	99.7	100
Omission: non-channel cells incorrectly identified (%)	0.3	0

The difference between channel networks is a trade-off between errors of omission and commission (James and Hunt, 2010). For example, CAT method identified a larger percentage of channel cells with higher errors of commission, but it committed lower errors of omission. It is due to the low-gradient study area which has fewer channels. The small difference in hillslope reflects that fact that total precision is driven by the number of channel cells that are correctly classified in the subcatchments.

The CIP modelled channel network provided 100 % of grid cells correctly classified, suggesting that identifying channel and non-channel cells through LRM is more accurate in the study site. The close statistics between channel mapped using CIP and the field survey provides an opportunity to consider the temporal variability (rainfall) improving channel network. Although CIP model resulted in a precise estimate of actual channel than the CAT method, this can be achieved with a near perfect compensation of false channels and missed channels.

4. CONCLUSION

Traditional approaches based on an area threshold do not recognise the importance of hydro-geomorphic processes such as rainfall required to increase the predictive quality for channel network. In this situation, the LRM can in principle demonstrate that it is possible to determine a channel head which is more accurate by considering hydrologic drivers. This paper presents a comparison of the capability of two channel extraction methods in extracting the channel network, capturing channel heads, detecting relevant channel continuity in the Lower Cotter catchment of southeast Australia. The CIP model captures channel discontinuity, while the TauDEM would have not been able to highlight the presence of these disruptions, as channels would be traced continuously throughout the basin. Therefore, the CIP model is capable to distinguish a continuous and disconnected channel that is important for forest management. The CIP model represents a new approach for advancing automatic feature extraction and thus advance the study of channelised processes in terms of better understanding of hydro-geomorphic form. There are two avenues of investigation which can be explored for further study. Firstly, due to time and resource constraints, variables used to develop the CIP model were limited to the terrain derivatives that could be measured from the DEM and rainfall records. The additional variables such as vegetation and soil properties will be considered in future work, but could not be included in this initial feasibility phase of the project. While this research only concerned the 10m DEM, it is important to study the effect of data resolution on the final outputs. It is necessary to illustrate the change in distance between modelled channel head location and field-verified channel head location for different DEM resolutions. Recently, a Light Detection and Ranging (LiDAR) derived DEM has been used for calculation of slope for headwater channel network analysis (James and Hunt, 2010). In the study site, LiDAR elevation data are available. This will be used to detect the areas of convergence related to channelised processes under forest cover, thus potentially more accurate in identifying and mapping headwater streams.

ACKNOWLEDGMENT

The authors would like to acknowledge the support of Australian Research Council (ARC). Linkage Project LP0881993

REFERENCES

- Conrad, R.H., Fransen, B., Duke, S., Liermann, M. and Needham, S. (2003). The development and assessment of the preliminary model for identifying fish habitat in western Washington. Draft report prepared for Washington Dept. Natural Resources Instream Scientific Advisory Group.
- Dietrich, W.E. and Dunne, T. (1993). The channel head. In: *Channel network hydrology*, (ed. by K., Beven and M.J Kirkby), 175-219, John Wiley and Sons, New York.
- Dietrich, W.E., Wilson, C.J., Montgomery, D.R. and McKean, J. (1993). Analysis of erosion thresholds, channel networks and landscape morphology using a digital terrain model. *Journal of Geology*, 101(2), 161-180.
- Furbish, D.J., Haff, P.K., Dietrich, W.E. and Heimsath, A.M. (2009). Statistical description of slope-dependent soil transport and the diffusion-like coefficient. *Journal of Geophysical Research, Earth Surface*, 114, F00A05, doi: 10.1029/2009JF001267, 19p.
- Gallant, J.C. and Wilson, J.P. (2000). Primary Topographic Attributes. In: *Terrain Analysis: Principles and Applications*, (ed. by J. P. Wilson and J. C. Gallant), 51-85, New York, John Wiley and Sons.
- Garbrecht, J. and Martz, L. (2000). Digital elevation model issues in water resources modelling. In: *Hydrologic and Hydraulic Modelling Support with Geographic Information Systems: Redlands*, (ed. by D. Maidment and D. Djokic), Environmental Systems Research Institute Press.
- Heine, R.A., Lant, C.L. and Sengupta, R.R. (2004). Development and comparison of approaches for automated mapping of stream channel networks. *Annals of the Association of American Geographers*, 94(3), 477-490.
- Hooke, J. (2003). Coarse sediment connectivity in river channel systems: a conceptual framework and methodology. *Geomorphology*, 56, 79-94.
- Horton, R.E. (1945). Erosional development of streams and their drainage basins: hydro-physical approach to quantitative morphology", *Geological Society of America Bulletin*, 56 (3), 275-370.
- Hosmer, D. and Lemeshow, S. (1989). *Applied Logistic Regression*. New York: John Wiley and Sons.
- Jaeger, K.L., Montgomery, D.R. and Bolton, S.M. (2007). Channel and perennial flow initiation in headwater streams: Management implications of variability in source-area size. *Environmental Management* 40, 775-786.
- James, L.A. and Hunt, K.J. (2010). The LiDAR-side of Headwater Streams: Mapping Channel Networks with High-Resolution Topographic data. *Southeastern Geographer*, 50(4), 523-539.
- Lashermes, B., Foufoula-Georgiou, E. and Dietrich, W.E. (2007). Channel network extraction from high resolution topography using wavelets. *Geophysical Research Letters*, 34, L23S04, doi: 10.1029/2007GL031140.
- Melville, J.K. and Martz, L.W. (2004). A comparison of data sources for manual and automated hydrographical network delineation. *Canadian Water Resources Journal*, 29(4), 267-282.
- Montgomery, D.R. and Dietrich, W.E. (1989) Channel initiation, drainage density and slope. *Water Resources Research*, 25(8), 1907-1918.
- Passalacqua, P., Do Trung, T., Foufoula-Georgiou, E., Sapiro, G. and Dietrich, W.E. (2010). A geometric framework for channel network extraction from LiDAR: Nonlinear diffusion and geodesic paths. *Journal of Geophysical Research*, 115, F01002, doi: 10.1029/2009JF001254
- Prosser, I. and Abernethy, B. (1996). Predicting the Topographic Limits to a Gully Network Using a Digital Terrain Model and Process Thresholds. *Water Resources Research*, 32(7), 2289-2298.
- Protection of the Environment Operations Act 1997*, Environment Protection Licence, Environment Protection Authority, Forestry Commission of NSW, visited July 2011, <http://www.environment.nsw.gov.au/resources/forestagreements/UNEEPLAmend23.pdf>
- Strahler, A.N. (1952). Hypsometric (area-altitude) analysis of erosional topology. *Geological Society of America Bulletin*, 63 (11), 1117-1142.
- Tabachnick, B. G. and Fidell, L. S. (1996). *Using multivariate statistics* (3rd ed.). New York: Harper Collins
- Tarboton, D.G. (1997). A New Method for the Determination of Flow Directions and Contributing Areas in Grid Digital Elevation Models. *Water Resources Research*, 33(2), 309-319.
- Tarboton, D.G. and Ames, D.P. (2001). Advances in the mapping of flow networks from digital elevation data. *World Water and Environmental Resources Congress*, May 20-24, Orlando, Florida.
- Tarolli, P. and Tarboton, D.G. (2006). A new method for determination of most likely landslide initiation points and the evaluation of digital terrain model scale in terrain stability mapping. *Hydrology and Earth System Sciences*, 10, 663-677.
- Tarolli, P. and Dalla Fontana, G. (2009). Hillslope to valley transition morphology: New opportunities from high resolution DTMs. *Geomorphology*, 113, 47-56.
- Thommeret, N., Bailly, J.S. and Puech, C. (2010). Extraction of thalweg networks from DTMs: application to badlands. *Hydrology and Earth System Sciences*, 14, 1527-1536.