*iEMSs 2008:International Congress on Environmental Modelling and Software*
*Integrating Sciences and Information Technology for Environmental Assessment and Decision Making*
*4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings*
*M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)*
*International Environmental Modelling and Software Society (iEMSs), 2008*

# Spatially Distributed Identification of Debris Flow Source Areas by Credal Networks

**A. Salvetti** [a], **A. Antonucci**[b], **and M. Zaffalon**[b]

[a]*Politecnico di Milano, Dipartimento di Elettronica e Informazione,*
*Via Ponzio 34/5, 20123 Milano, Italy.*
*Dipartimento del Territorio, Cantone Ticino,*
*Viale Franscini 17, 6500 Bellinzona, Switzerland.*
salvetti@elet.polimit.it, andrea.salvetti@ti.ch

[b]*Istituto Dalle Molle di Studi sull'Intelligenza Artificiale,*
*Galleria 2 CH-6928, Manno-Lugano, Switzerland.*
{alessandro,zaffalon}@idsia.ch

**Abstract:** Debris flows represent a very destructive natural hazard, affecting buildings, transport infrastructures, and, very often, causing human losses in mountain regions. That makes the identification of potential source areas of debris flows inside a watershed particularly important. In this paper we present a general identification procedure based on the *credal network* (that is an imprecise probabilistic graphical model generalizing Bayesian networks) originally introduced by Antonucci et al. [2004]. That model is significantly improved by a more refined description of the meteorological and hydrological processes contributing to the debris flow initiation. As a counterpart of such improvement, the model pays a slight increase in terms of computational time for identifications. That does not prevent its extensive, spatially distributed, application to whole basins, thanks to a preliminary deterministic analysis that rejects local areas where the triggering of a debris flow cannot take place. The overall procedure is tested for a debris flow prone watershed in Southern Switzerland. The model detects the areas in the basin more prone to debris flow initiation and also shows that different rainfall return periods produce different patterns of hazard in the basin. That makes it possible with this procedure to determine the return period of the critical rainfall that triggers debris flow as a result of channel-bed failure in a specific point along the drainage network.

***Keywords:*** Debris Flow; Geomorphologic Theory; Geographic Information System; Imprecise Probabilities; Credal Networks.

## 1 INTRODUCTION

*Debris flows* (Section 2) represent a very destructive natural hazard, affecting buildings, transport infrastructures, and, very often, causing human losses in mountain regions. As recently pointed out by Berti and Simoni [2005], the triggering mechanisms and the causal relationships for the whole process are still partially unknown. Thus, human expertise together with an analysis of historical data are still necessary to support any deterministic model for the identification of potential source areas of debris flow.

A *credal network* (which is an imprecise probabilistic graphical model as described in Section 3) has been introduced by Antonucci et al. [2004] in order to fuse into a single coherent framework the model of Takahashi [1991] with expert qualitative judgments and historical data. In this paper we first improve this probabilistic model with a more refined description of the *drainage network* delineation and the *maximum peak runoff* formation (see Section 4). We also allow for more freedom in the description of the observation of the triggering factors by showing how sets of probability mass functions can model the so-called *soft evidence* provided by vague observations.

Overall, such improved flexibility in the modeling phase has a counterpart when we compute the level of risk corresponding to the available evidence by appropriate updating algorithm. The required computational time slightly increases. Nevertheless, we show that our approach can be properly employed for spatially distributed identifications on extensive areas. This is made faster by the rejection of the points of the basin where the geomorphological conditions prevent any debris flow initiation. Finally, with the support of a detailed GIS analysis, we test this procedure for a debris flow prone watershed in Southern Switzerland (Section 5). The results indicate that the model detects the areas of the basin more prone to debris flow initiation and produces different hazard patters according to different rainfall events.

## 2 DEBRIS FLOWS

Debris flows are gravity-induced mass movement intermediate between landslides and water floods. They are composed of a mixture of water and sediment with a characteristic mechanical behavior varying with water and soil content. According to Costa [1984], prerequisite conditions for most debris flows include an abundant source of unconsolidated fine-grained rock and soil debris, steep slopes, a large but intermittent source of moisture (rainfall or snowmelt), and sparse vegetation. As mentioned in Griffiths et al. [2004], several investigation have focused on debris flows initiation and frequency. Benda and Dunne [1997] approached the modeling of spatial and temporal variability of sediment yields, Glade [2005] focused on existing links between debris-flow hazard and geomorphology. Several hypotheses have been formulated to explain mobilization of debris flows and this aspect still represents a research field. The triggering mechanism of the identification procedure presented in this paper is based on the theoretical model proposed by Takahashi [1991], although a more sophisticated explanation of the triggering of debris flow by channel-bed failure has been recently proposed by Armanini and Gregoretti [2005], which consider the exposure of a single particle to the stream flow and, explicitly, the flow velocity profile. For the purposes of this study the Takahashi's theory is regarded as appropriate and this triggering theory is further coupled with geological, hydro-meteorological and topographic factors, which all contribute to the definition of channel-bed failure. Unfortunately, not all the triggering factors considered by this model can be directly observed in the field. Takahashi's theory will therefore offer the deterministic skeleton for our model that will be integrated with probabilistic knowledge according to the methods described in the following section.

## 3 CREDAL SETS AND CREDAL NETWORKS FOR UNCERTAIN REASONING

If a complete deterministic model of the relations between some variables cannot be provided, probabilistic approaches should be considered instead. Probabilistic graphical models like Bayesian nets seem to be particularly suited for situations where some conditional independence relations hold between the different variables. Here, we consider *credal nets* (Cozman [2000]), which are a generalization of Bayesian nets based on the fundamental notion of *credal set*.

Given a categorical random variable $X$,[1] we denote by $\mathcal{X}$ the set of the possible values of $X$, while $x$ denotes a generic element of $\mathcal{X}$. The notation $(X = x)$ denotes an event that is true if and only we know that $X$ is in the state $x$. This is clearly the most informative scenario we can consider for $X$. Otherwise, our knowledge about the actual state of $X$ can be modeled by a probability mass function $P(X)$. There are also situations where a single *precise* numerical value for the probability $P(X = x)$ cannot be easily assessed. In these cases, a more realistic model of our knowledge about $X$ could be a *credal set* $K(X)$, i.e., a closed convex set of probability mass functions over $X$. As noted by Reichert [1997], such *imprecise probability* (Walley [1991]) approach offers important advantages for environmental studies, where a vague prior knowledge

[1] All the quantities considered in this paper are regarded as random variables which assume only finitely many values. Therefore, in the case of continuous variables, a discretization should be preliminary done.

of processes is very usual, due to incomplete observation of the process or to the impossibility of gathering enough data. Furthermore, expert knowledge and qualitative beliefs can be more easily considered in the general framework of sets of probability distributions.

A credal set $K(X)$ can be specified by a set of *probability intervals* $I_X := \{I_x : I_x := [l_x, u_x], 0 \leq l_x \leq u_x \leq 1, x \in \mathcal{X}\}$, from which $K(X)$ is clearly obtained from $\{P(X) : P(x) \in I_x, x \in \mathcal{X}, \sum_{x\in\mathcal{X}} P(x) = 1\}$. Accordingly, we can ask experts to assess sets of probability intervals (and more generally linear constraints on probabilities). Probability intervals can be also inferred from data by the *imprecise Dirichlet model* according to Walley [1996], a generalization of Bayesian learning from multinomial data based on an imprecise probability model of prior ignorance.

Inference over a credal set is intended as the computation of lower and upper expectations over all the mass functions of a credal set. For example the lower probability is defined as $\underline{P}(x) := \min_{P(X)\in K(X)} P(x)$, and similarly for the upper probability $\overline{P}(x)$. Remarkably, a set of mass functions, its convex hull and its *extreme mass functions* produce the same lower and upper expectations and probabilities. Conditioning with credal sets is done by element-wise application of Bayes' rule. The posterior credal set is the union of all posterior mass functions. Denote by $K(X|y)$ the set of mass functions $P(X|Y = y)$, for generic variables $X$ and $Y$.

A *credal network* is a graphical model where each node (or variable) of a directed acyclic graph is associated with a credal set for any configuration of the variable's parents (see Cozman [2000]); informally, credal nets are equivalent to sets of Bayesian networks with the same graph. The graph codes strong dependencies by the so-called *strong Markov condition*: every variable is strongly independent of its non-descendant non-parents given its parents. A generic variable, or node of the graph, $X_i$ holds the collection of conditional credal sets $K(X_i|\pi_{X_i})$, one for each possible joint state $\pi_{X_i}$ of its parents $\Pi_{X_i}$. We assume that the credal sets of the net are *separately specified* Walley [1991]: this implies that selecting a mass function from a credal set does not influence the possible choices in others. A credal network defines a joint credal set $K(\mathbf{X})$, which is called the *strong extension* of a credal network. This is the convex hull of the set of joint mass functions $P(\boldsymbol{X}) = P(X_1, \ldots, X_t)$, over the $t$ variables of the net, that factorize according to:

$$P(x_1, \ldots, x_t) = \prod_{i=1}^{t} P(x_i\,|pa\,(X_i)\,) \quad \forall (x_1, \ldots, x_t) \in \times_{i=1}^{t} \mathcal{X}_i \tag{1}$$

Here $pa\,(X_i)$ is the assignment to the parents of $X_i$ consistent with $(x_1, \ldots, x_t)$; and the conditional mass functions $P(X_i\,|pa\,(X_i)\,)$ are chosen in all the possible ways from the respective credal sets.

### 3.1 Updating Credal Networks

Credal networks are often used as *expert systems*. The available evidential information $x_E$ about the variables $X_E$ that have been observed is first gathered. Then, we compute the posterior lower (and upper) probabilities for a queried variable $X_q$, say $\underline{P}(X_q|x_E)$ (and similarly for the upper) with respect to the network strong extension. The evidence $X_E = x_E$ is said to be *strong* as it provides completely informative information about the state of their variables. Updating of credal networks with strong evidence is a hard task, but a number of approximate (and even exact in some special cases) algorithms can be employed.

Moreover, it is possible to consider situations where the result of the observation of a variable is vague, and we cannot obtain a strong evidence about the state of the observed variable. Nevertheless, such *soft evidence* about the state of the variable could be expressed by a credal set. This is a generalization of Jeffrey's updating to imprecise probability. This soft evidence can be easily embedded in the structure of a credal network if the observed variable is a root node: it suffices to replace the unconditional credal set for the variable with that corresponding to the observation. This approach will be applied to the credal network for debris flow evaluation presented in this paper in order to model the vague observation of *Granulometry*.

## 4 Spatially Distributed Identification of Source Areas

Antonucci et al. [2004] proposed a credal network for a single-point analysis of debris flow initiation. We present a significantly improved version of that model obtained with a refined description of the meteorological and hydrological processes contributing to the debris flow initiation. The new credal network is based on the directed acyclic graph in Fig. 1 and expresses the causal relationships between the topographic and geological characteristics, and hydrological preconditions, which are recognized as triggering factors. The key nodes (denoted as shadowed nodes in Fig. 1) of the network are the *Effective Soil Water Capacity*, which reflects the influences of the soil and the geological characteristics of the area, the *Basin Response Function*, related to the topographic properties of the watershed and whose footprint can be detected in the hydrologic response of the basin, the *Peak Flow*, which summarizes the interactions among the hydro-meteorological factors and topographic and geologic preconditions, the *Theoretical Debris Thickness*, obtained by the Takahashi's theory, and the *Available Debris Thickness*, which considers the influence of the topography on the debris availability. The leaf node *Movable Debris Thickness*, which is defined as the depth of debris likely to be transported downstream during a flood event, is our proxy for the risk level in the specific point along the drainage network where we have collected evidence about the triggering factors. The ranges $0$–$10\mathrm{cm}$, $10$–$30\mathrm{cm}$, $>30\mathrm{cm}$ for that thickness are assumed to indicate respectively a low, medium and high level of risk. In this section only topics related to the novel improvements of the credal network regarding the *Stream Power Index* and the *Maximum Peak Runoff* are described in detail, while for a more comprehensive description of each node the reader is referred to Antonucci et al. [2004].

**Stability of a Debris Cluster** The effect of a water depth on the movable debris quantity is based on the equilibrium of forces acting on a debris cluster under different conditions. According to Takahashi [1991], the local slope $\theta$ for which debris-flow formation can take place obeys the following constraint:

$$\frac{c^*\Delta}{\frac{4}{3}+c^*\Delta}\tan\phi \leq \tan\theta \leq \frac{c^*\Delta}{1+c^*\Delta}\tan\phi, \tag{2}$$

where $\phi$ is the *Friction Angle* corresponding to the actual level of *Granulometry*, $c^*$ is the volume concentration of the particles, and $\Delta$ is the ratio between the mixture and the water density. For the points of the basin whose values of $\phi$ and $\theta$ do not satisfy the constraint in Eq. (2), either the cluster is not completely saturated and, if unstable at high slope angles, produces a landslide or the process that takes place is the ordinary solid transport (Dietrich et al. [2006]), and therefore we drop the relative point from the potential source areas of this hazard without any further analysis.

**Drainage Network Delineation and Debris Availability** Many authors have dealt with the capability of *Local Slope* and *Upstream Contributing Area* to account for topographic control on erosion and deposition potential in complex terrain and with the use of slope and contributing area for channel network extraction, based on critical area and slope-area threshold (e.g., Prosser and Abernethy [1996]). In this study, such a method was used to extract the channelized portion of the *Digital Elevation Model* (DEM), where debris flow initiation can appear, according to the following equation:

$$SPI = \sqrt{A}\cdot\theta, \tag{3}$$

where $SPI$ denotes the *Stream Power Index* and $A$ is the upstream area. The threshold value of $SPI$ has been identified by trials, comparing the extracted network with the drainage network on the map, where also many ephemeral channel in the upper part of the basin were included in the network. That index can be used as an indicator of the local transport capacity of a single reach along the network and, therefore, to identify channel reaches were debris material preferentially accumulates (Dalla Fontana and Marchi [2003]). Clearly, the availability of an abundant debris thickness in the drainage network is a fundamental precondition for debris initiation and we therefore developed a conceptual framework for a qualitative evaluation of the debris availability in the river bed. We assume that the debris availability is a function of the convenience capacity of the

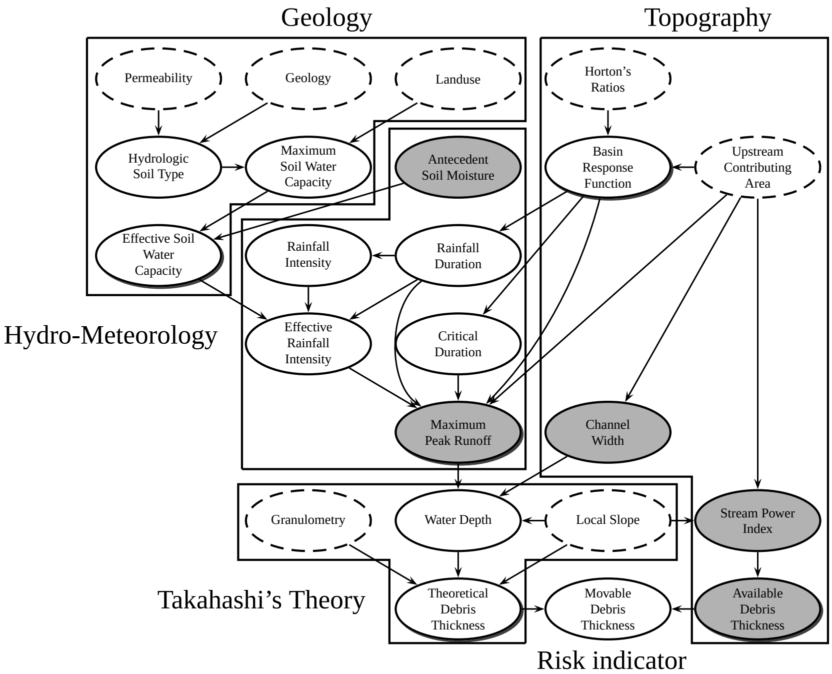


**Figure 1:** The directed acyclic graph of the credal network for hazard identification. The gray nodes denote the new factors or those for which a new quantification has been proposed with respect to the original network in Antonucci et al. [2004]. A dashed border denotes the nodes observed in the case study of Section 5.

river network associated to the $SPI$. Cells with $SPI$ value exceeding the threshold for channel initiation correspond very often to areas where bedrocks emerge and local slope is quite high, and therefore the sediment deposition is zero or very low. On the contrary, in cells where $SPI$ is much less than the selected threshold level, high deposition instead of erosion is expected and we therefore assume a high availability of debris material. These principles supported by expert knowledge have been used for an interval-valued probabilistic quantification of the node *Available Debris Thickness*.

**Maximum Peak Runoff** According to the model of the initiation mechanism considered in this study, the soil failure is induced by surface runoff and, consequently, the maximum discharge and the corresponding water depth must be estimated. Rodriguez-Iturbe and Rinaldo [1997] investigated how the variation of the characteristics of stream channel is expressed as a function of the discharge by a power law at a given cross section and also along the channel network. The parameters were estimated by using a few collected cross-section data, randomly distributed along the drainage network.

The maximum runoff along the drainage network was calculated according to the well-established theory of the *Instantaneous Unit Hydrograph* (IUH), expressed as the convolution integral of the effective rainfall input. The hydrograph shape strongly depends on the geomorphological features of the river basin, therefore, the *Geomorphologic Instantaneous Unit Hydrograph* (GIUH, Rodriguez-Iturbe and Valdes [1979]) presents obvious advantages in ungauged watersheds, since the GIUH only depends on the morphological characteristics of the watershed and the drainage network. According to this theory, the maximum peak runoff $Q$ is obtained using the following:

$$Q = \begin{cases} I' A \left[H\left(t^*\right) - H\left(t^* - t_p\right)\right] & 0 \leq t_p \leq \tau_c \\ \\ I' A & t_p > \tau_c, \end{cases} \tag{4}$$

where $I'$ is the effective rainfall intensity, $H(t)$ represents the integral of the GIUH from the beginning of the storm, $t^*$ is the critical duration at the considered point, and it is a function of the rainfall duration $t_p$, while $\tau_c$ is the concentration time. Effective rainfall intensity is determined using the well established *SCS Curve Number* infiltration method and the rainfall intensity modeled by multiscaling power law relationship. The critical duration $t^*$ associated with the extreme peak runoff is independent of the return period and of the rainfall intensity; the corresponding rainfall volume is calculated for the rainfall duration $t_p$.

## 5 A Case Study: Distributed Application to a River Basin

The case study we present in this section refers to the Acquarossa Creek in the Blenio Valley, an area located in the North-Eastern part of the Ticino Canton, Southern Switzerland. This area was selected because of the potential hazard caused by debris flows to communication lines and villages. That creek is a small tributary of the Brenno river, characterized by a high altitude range (from $530\mathrm{m}$ up to $2580\mathrm{m}$ a.s.l.) of the Simano Peak. Debris torrents are usually triggered by intense rainfall, following a period of abundant precipitation. Eight historical debris flow events were recorded in that area during the last 150 years. Most of them caused high damages to infrastructures on the alluvial fan, transporting several thousand cubic meters of material. For instance, during the last event in August 2003, a volume of about $15'000\mathrm{m}^3$ were estimated on the alluvial fan, and a similar pattern was observed in 1983 and 1987. That represents a relatively high frequency of debris flow events. Accordingly, the triggering factors appear to be already effective in many parts of the basin with storm events of low and medium return period.

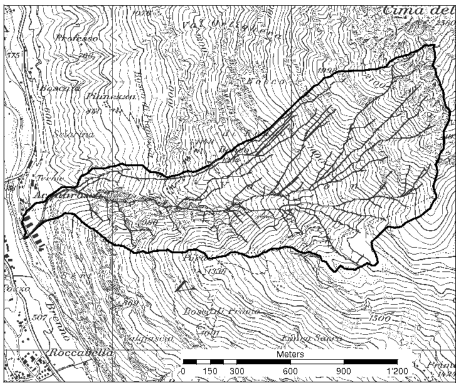


**Figure 2:** Acquarossa Creek Basin (area $1.6\mathrm{Km}^2$, length $3.1\mathrm{Km}$).

In order to gather evidential information about the geomorphological characteristics of the basin, a highly precise DEM based on airborne laser scanning produced by the Swiss Federal Office of Topography has been employed. That offers a spatial resolution of $4$ meters, which is comparable with the typical channel width; that defines a drainage network of $6310$ cells. Most of the morphological data used for our identification analysis (slope, flow-direction and flow-accumulation) were derived from this dataset, and the SPI was calculated as in Eq. (3).

Finally, regarding the observation of the granulometry, a field survey was conducted. The river bed and lateral debris levees were analyzed in order to determine the grain-size distribution of the debris material. A significant difference was observed for the grain-size distributions obtained from several samples. We have therefore decided to split the basin into two sub-regions of "uniform" granulometry, and describe the outcome of the sampling by a *soft evidence* modeled as a new unconditional credal sets for the corresponding node in the credal network, according to the procedure described in Section 3.1.

In order to avoid unnecessary computations, for each point of the basin, we have preliminarily checked whether or not the observed slope and the values of the friction angle compatible with the soft observation of the granulometry were compatible with the constraint in Eq. (2). This deterministic pre-analysis detects 170 pixels where only ordinary sediment transport is possible and 135 pixels that are already unstable without complete soil saturation. For the remaining 6005 pixels, we have computed the posterior lower and upper probabilities for the movable debris thickness corresponding to observed geomorphological factors and rainfall intensity for a return period of 10, 30 and 100 years. These computations have been exactly performed by exhaustive approaches based on the iteration of standard algorithms for Bayesian networks as our credal network is equivalent to about 500 Bayesian networks. The network is thus expected to predict the probability of a debris flow event with the defined frequency level at each point of the drainage network. In this way, we aim at verifying whether the network would have been a valuable tool to predict considerable events of debris flows, which actually happened in the areas under consideration, and, more important, to identify the points where the debris flow is most likely to occur in the future. Figure 3 reports the results of the inference process for respectively 10 and 100 years return period rainfall event.

We observe that, according to the outputs of the credal net, debris flows are more likely to initiate on the main channel, even in the lower part of the basin. In fact, during a field survey conducted immediately after the debris flow event on August 2003, we observed typical evidences of bed erosion and channel-bed failure in the lower part of the main channel, up to an altitude of 700 m a.s.l; this observation is effectively confirmed by our results. Regarding the role of the return period in our tests, we observe an increase of the number of dangerous points, that spread upstream along the drainage network: for higher return periods even a small upstream area is sufficient to produce a peak runoff that can trigger a debris flow. The promising results of the credal net will be further compared and quantitatively evaluated by applying the model to other watersheds, where detailed geomorphological maps are obtainable and field observations of availability and characteristics of debris material along the drainage network have been recently collected.

**Figure 3:** Spatially distributed identifications for the basin in Figure 2 and rainfall return periods of 10 (left) and 100 (right) years. Points for which the credal net predicts the lower class of risk are depicted in gray, while black refers to points where higher levels of risk cannot be excluded.

## 6 CONCLUSIONS AND OUTLOOKS

We have presented a model for automatic identification of potential source areas of debris flows based on *credal networks*. Our network provides a refined description of the meteorological and hydrological processes contributing to the debris flow initiation, and allows to model also *vague* observations of the triggering factors. The identification procedure can be extensively applied to whole basins, and unnecessary computations are avoided for areas where the geomorphological conditions are not compatible with debris flow initiation. As a spatially distributed case study, we tested our model for a debris flow prone watershed in Southern Switzerland and the obtained results agree with field observations collected after the last debris flow event of 2003. The model is able to detect the areas inside the basin more prone to debris flow initiation and also shows that different rainfall return periods produce different hazard patterns. That makes it possible to determine the return period of the critical rainfall that triggers debris flow as a result of channel-bed failure in a specific point along the drainage network. As a possible development of the present work, we intend to design a post-processing procedure for our simulations that produces integral risk indicators based on neighborhood relations among the detected dangerous points.

## ACKNOWLEDGMENTS

This work was partially supported by the Swiss NSF grants 200021-113820/1 and 200020-116674/1, and by the Hasler Foundation grant 2233.

## REFERENCES

Antonucci, A., A. Salvetti, and M. Zaffalon. Hazard assessment of debris flows by credal network. In *iEMSs 2004 International Congress: Complexity and Integrated Resources Management*, Osnabrück, Germany, 2004. International Environmental Modelling and Software Society.

Armanini, A. and C. Gregoretti. Incipient sediment motion at high slopes in uniform flow condition. *Water Resources Research*, pages W12431, doi:10.1029/2005WR004001, 2005.

Benda, L. and T. Dunne. Stochastic forcing of sediment routing and storage in channel networks. *Water Resources Research*, 33(12):2865–2880, 1997.

Berti, M. and A. Simoni. Experimental evidences and numerical modelling of debris flow initiated by channel runoff. *Landslide*, 2:171–182, 2005.

Costa, J. E. Physical geomorphology of debris flows. In *Costa, J. E. and Fleisher, P. J. Developments and Application of Geomorphology*, Berlin, Germany, 1984. Springer Verlag.

Cozman, F. Credal networks. *Artificial Intelligence*, 120:199–233, 2000.

Dalla Fontana, G. and L. Marchi. Slope-area relationships and sediment dynamics in two alpine streams. *Hydrological Processes*, 17:73–87, 2003.

Dietrich, W., D. Bellugi, and R. Asua. Validation of the shallow landslide model, shalstab, for forest management. In *Land use and watersheds: human influence on hydrology and geomorphology in urban and forest areas edited by M.S. Wigmosta and S.J. Burges*, Washington D.C., 2006. AGU.

Glade, T. Linking debris-flow hazard assessments with geomorphology. *Geomorphology*, 66: 189–213, 2005.

Griffiths, P., R. Webb, and T. Melis. Frequency and initiation of debris flows in grand canyon, arizona. *Journal of Geophysical Research*, 109:4002–4015, 2004.

Prosser, I. and B. Abernethy. Predicting the topographic limits of a gully network using a digital terrain model and process threshold. *Water Resources Research*, 32(12):2289–2298, 1996.

Reichert, P. On the necessity of using imprecise probabilities for modelling environmental systems. *Water Science and Technology*, 36(5):149–156, 1997.

Rodriguez-Iturbe, I. and A. Rinaldo. *Fractal river basins - Chance and Self Organization*. Cambridge University Press, Cambridge, UK, 1997.

Rodriguez-Iturbe, I. and J. Valdes. The geomorphological structure of hydrogeological response. *Water Resources Research*, 15:1409–1420, 1979.

Takahashi, T. *Debris Flow*. A.A. Balkema IAHR Monograph, Rotterdam, 1991.

Walley, P. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, New York, 1991.

Walley, P. Inferences from multinomial data: learning about a bag of marbles. *J. R. Statist. Soc. B*, 58(1):3–57, 1996.