Semantic array programming in data-poor environments; assessing the interactions of shallow landslides and soil erosion

by

Claudio Bosco

Doctoral Thesis

Submitted in partial fulfillment of the requirements

for the award of

Doctor of Philosophy of Loughborough University

January 2019

Acknowledgements

I decided to enrol in a Ph.D. programme at the University of Loughborough, when I was well into my working carrier as a researcher in Italy. Embarking in such a venture required support from many, and particularly from my family, my closest colleagues and my supervisors. During this endeavour, I received constant encouragement and support from many people and especially from my wife Stefania. Her unconditioned and relentless assistance and appreciation was indispensable.

I would like to thank everyone who has contributed to this research and provided support and encouragement on this journey, specific thanks are due to: The School of Architecture, Building and Civil Engineering of the Loughborough University, for providing me with funding. My supervisors, Prof. Graham Sander and Dr. Tom Dijkstra. Both Graham and Tom were more than generous with their guidance, time, knowledge and patience. I would like to offer further thanks to them in guiding me through the writing up process. Thanks are also extended to Dr Janusz Wasowski of CNR-IRPI of Bari for his valuable support during the field survey and for sharing with us many valuable information of the study site. I am also in debt with Dr. Daniele de Rigo, for introducing me to statistical modelling and in particular machine learning techniques. Daniele is a sharp scientist and a patient teacher with the rare ability to make difficult problems look easy. He is a good friend.

I dedicate this work to my much loved daughters Beatrice and Rebecca.

Table of Contents

List of acronyms / abbreviations

AGNPS	Agricultural Non-Point Source pollution model (Young et al., 1989)
ANFIS	Adaptive Neuro-Fuzzy Inference Systems
ANN	Artificial Neural Network
ANSWERS	Areal Nonpoint Source Watershed Environmental Response Simulation model (Beasley et al, 1980)
AP	Array Programming
ASTER	Advanced Spaceborne Thermal Emission and Reflection radiometer
Bsp	Bare soil proportion
CI-SLAM	Landslide susceptibility model (Lanni et al., 2012)
CLC	CORINE Land Cover
CNR	National research council
CREAMS	A model for Chemicals, Runoff and Erosion from Agricultural Management Systems (Knisel, 1980)
DBM	Data-Based Mechanistic
DEM	Digital Elevation Model
dSLAM	Slope stability model (Wu and Sidle, 1995)
D-TM	Data-Transformation Model
EGEM	Ephemeral Gully Erosion Model (Woodward, 1999)
EHD	Effective Hydrological Depth
EMSS	Environmental Management Support System (Vertessey et al., 2001; Watson et al., 2001)
E-OBS	Ensembles OBServations gridded data set
ER	Effective Rainfall
EROSION 3D	Soil erosion model (von Werner, 2004)
EUROSEM	EUROpean Soil Erosion Model (Morgan et al., 1998a,b)
FS	Factor of Safety
GC	Ground Cover
GeoSemAP	Geospatial Semantic Array Programming
GEOtop	Hydrological model (Rigon et al., 2006)
GIS	Geographic Information systems
GUEST	Griffith University Erosion System Template (Misra and Rose 1996; Rose et al. 1997)

HSPF	Hydrologic Simulation Program—Fortran (Johanson et al., 1980)
IFFI	Inventory of landslide phenomena in Italy
IHACRES-WQ	Identification of unit Hydrographs and Component flows from Rainfall, Evaporation and Streamflow data (Jakeman et al., 1990, 1994; Dietrich, 1999)
IKONOS	Global High-resolution satellite imagery
INRMM	Integrated Natural Resources Modelling and Management
IRPI	Institute for geo-hydrological protection
KED	Kriging with External Drift
KINEROS 2	KINematic runoff and EROSion model (Smith et al., 1995a,b)
K-NN	K-Nearest Neighbours
LAI	Leaf Area Index
LASCAM	LArge Scale CAtchment Model (Viney and Sivapalan, 1999)
LISEM	Limburg Soil Erosion Model (de Roo et al., 1994)
LR	Logistic Regression
MAE	Mean Absolute Error
MEDRUSH	MEdalus Desertification Response Unit SHe (Kirkby et al., 1998a; Kirkby, 1998b)
MFM	Mass-Flux Method
MIKE-11	A modelling system for rivers and channels (Hanley et al., 1998)
MMF	Morgan-Morgan-Finney soil erosion model
MSE	Mean squared error
MUSLE	Modified Universal Soil Loss Equation (Williams, 1975)
NDVI	Normalized Difference Vegetation Index
NIR	Near Infrared Reflectance
PCHIP	Piecewise Cubic Hermite Interploating Polynomials
PERFECT	Productivity, Erosion, and Runoff, Functions to Evaluate Conservation Techniques (Littleboy et al., 1992)
PESERA	Pan-European Soil Erosion Risk Assessment (Kirkby et al., 2003)
PROBSTAB	PROBability of STABility model (van Beek, 2002)
PSIAC	Pacific Southwest Inter-Agency Committee model (PSIAC, 1968)
RDS	Relative-Distance Similarity
RK	Regression Kriging
RMMF	Revised Morgan-Morgan-Finney soil erosion model (Morgan, 2001)
RMSE	Root Mean Square Error

RUSLE	Revised Universal Soil Loss Equation model (Renard et al., 1997)
RUSLE2	Revised version of the RUSLE model (Foster et al., 2000)
SEDEM	SEdiment DElivery Model (Van Rompaey et al., 2001)
SedNet	Sediment river Network model (Prosser et al., 2001)
SemAP	Semantic Array Programming
SEMMED	Soil Erosion Model for Mediterranean Areas (de Jong et al., 1999)
SGDBE	European Soil Geographical Database
SHETRAN	Système Hydrologique Européen-TRANsport. Soil erosion model (Ewen et al., 2000)
SIBERIA	Lanscape evolution model (Willgoose et al., 1991 a-d)
SIMWE	SIMulation of Water Erosion (Mitas and Mitasova, 1998)
SINMAP	Stability Index MAPping model (Pack et al., 1998)
SMU	Soil Mapping Unit
SRTM	Shuttle Radar Topography Mission
SSE	Sum of Squared Errors
STARWARS	Storage and Redistribution of Water on Agricultural and Re-vegetated Slopes model (Van Beek, 2002)
STREAM	Sealing and Transfer by Runoff and Erosion related to Agricultural Management (Cerdan et al., 2001)
STU	Soil Typological Unit
SWAT	Soil and Water Assessment Tool (Arnold et al., 1998; Arnold and Fohrer, 2005)
SWRRB	Simulator for Water Resources in Rural basins (Arnold et al., 1990)
TC	Transport Capacity
TOPOG	TOPOGraphy. Hydrologic model (O'Loughlin, 1986; CSIRO, 2017)
TRIGRS	Transient Rainfall Infiltration and Grid-based Regional Slope stability model (Baum et al., 2008)
TWI	Topographic Wetness Index
UK	Universal Kriging
USPED	Unit Stream Power Erosion deposition model (Mitasova et al., 1996)
WATEM	WATer and Tillage Erosion Model (Van Oost et al., 2000)
WEPP	Water Erosion Prediction Project (Laflen et al., 1991)

Publications

A paper has been presented at the ISESS 2013 conference for presenting the techniques mentioned in chapter 5.0 (Bosco et al., 2013).
The methodology presented in chapter 6 was published in 'IEEE Earthzine' in a special issue on Geospatial Semantic Array Programming (Bosco and Sander, 2015).

List of papers

Bosco C. and Sander G., 2015. Estimating the effects of water-induced shallow landslides on soil erosion. IEEE Earthzine 7, 910137. (doi:10.1101/011965)

Bosco, C., de Rigo, D., Dijkstra, T., Sander, G., Wasowski, J.: Multi-Scale Robust Modelling of Landslide Susceptibility: Regional Rapid Assessment and Catchment Robust Fuzzy Ensemble. IFIP Advances in Information and Communication Technology, Springer. 2013.

inclusion in future work, and to publish a substantially revised version (at least 30% new content) elsewhere, provided that the original Springer Contribution is properly cited".

The journals of the Institute of Electrical and Electronics Engineers, Incorporated (IEEE) (such as IEEE Earthzine) normate the use of the published material (hereinafter called the Work) by means of a "IEEE Copyright and Consent Form" (available online at http://www.ieee.org/publications_standards/publications/rights/ieeecopyrightform.pdf, archived by WebCite at http://www.webcitation.org/6K9q4dseM). In particular, authors "retain all proprietary rights in any process, procedure, or article of manufacture described in the Work". They "may reproduce or authorize others to reproduce the Work, material extracted verbatim from the Work, or derivative works for the author's personal use or for company use, provided that the source and the IEEE copyright notice are indicated, the copies are not used in any way that implies IEEE endorsement of a product or service of any employer, and the copies themselves are not offered for sale". Furthermore, the online document available at http://www.ieee.org/publications_standards/publications/rights/permissions_faq.pdf (archived by WebCite at http://www.webcitation.org/6K9qkUP6d) clarifies that "IEEE does not require individuals working on a thesis to obtain a formal reuse license" provided proper credit is given to the original source.

1. Introduction

1.1. Introduction and problem statement

Soil erosion is the displacement of soil particles from one location to another by the action of physical forces. The agents of soil erosion are mainly water and wind, each contributing a significant amount of soil loss each year. Soil erosion by rainfall and runoff is one of the main threats to soil sustainability in Europe (Boardman and Poesen, 2006). Approximately 15% of the European territory is affected by significant soil erosion (Bosco et al., 2015; Cerdan et al., 2010). Soil erosion is the result of a complex suite of processes involving both land degradation and soil deformation (ranging from micro-straining involving particle rearrangements to large scale displacements and mass movement of soils in the form of slides, falls or flows) (see chapter 2). The upper part of the soil, which is generally the most fertile layer, is also the most prone to erosion. When this layer is lost through soil erosion, large costs are incurred (Posthumus et al., 2015; Pimentel and Burgess, 2013; Bosco et al., 2015). For example, soil loss can lead to reduced soil quality, including a lower nutrient capacity, that, in turn, will lead to a significant rise in food production costs. A nutrient deficient soil produces up to 30% lower crop yields (Pimentel, 2006; Lal, 1998). The nutrient loss cost in the U.S.A. alone reaches several billion dollars annually (Pimentel and Burgess, 2013).

Soil erosion also has other environmental consequences, research has shown that nutrient and carbon cycling are significantly altered by mobilization and deposition of soil (Quinton et al., 2010; Stockmann et al., 2013; Novara et al., 2016) and an eroded soil may lose 75–80% of its carbon content, with the consequent release of carbon to the atmosphere (Morgan, 2005; Yue et al., 2016).

Soil erosion is linked to several natural hazards, such as floods and landslides (Markantonis et al., 2012). It can also cause water pollution and siltation, loss of organic matter and a reduction in water holding capacity (Boardman and Poesen, 2006). The protection of soil resources has therefore been recognized as an important objective of environmental policy (CEC, 2006). Given the increasing threat

of soil erosion all over the world, and the implications this has on future food security and soil and water quality, an in-depth understanding of the rate and extent of erosion processes is crucial. Despite many years of research already undertaken in this field, to date scientists appear to have inadequately addressed some of the 'big questions' of our discipline such as: where is erosion occurring? How severe is it? (Boardman, 2006; de Vente et al., 2014). This research aims to contribute in decreasing some of the weaknesses in soil erosion modelling, especially in data-poor areas.

It is impractical to measure soil loss across whole landscapes by directly measuring water-induced soil erosion across large areas using experimental plots, soil erosion markers (e.g Caesium 137) or sampling river sediment load. This is technically and logistically difficult, and very expensive. Regional assessments involving techniques such as those based on remote sensing also have limitations, such as, for example, the cost and availability of remotely sensed data with adequate resolution and frequency (especially on arable land where spectral patterns are extremely time-dependent) (Boardman, 2007). Therefore, further research is needed to improve methods and models for the estimation of soil erosion rates, so that appropriate management and mitigation strategies can be assessed, designed and implemented. Several models exist to predict soil erosion rates by water. These differ greatly in terms of complexity, inputs, spatial and temporal scale (see section 2.2). Heterogeneity of the models also affects the modelled processes (sheet-, rill-, ephemeral gully erosion), the manner in which these processes are represented and the types of output information they provide (e.g. mean annual or event-based soil erosion rate) (de Vente et al., 2013). Many efforts have been made to describe soil erosion processes within models to achieve a better predictability and a more effective identification of the parameters involved (Wilken et al., 2017; Rose, 2017; Boardman, 2006; Sander et al., 2002). Unfortunately, input data of sufficient accuracy, granularity and continuity may not always be available (Jones et al., 2003) and application outside the spatial domain in which erosion models have been tested could be problematic (Favis-Mortlock, 1998). Models therefore need to be developed

that recognize these data limitations, and that can still be applied to data-poor regions.

Often soil erosion is closely related with an enhanced susceptibility of a landscape to generating mass movements. Either can be a cause or an effect of the other (Larsen et al., 2010; Burton and Bathurst, 1998). Mass movements and soil erosion are part of a system of multiple interacting processes operating in a complex hierarchy, with both being highly visible expressions of critical instabilities in a landscape. Both can significantly affect sediment budgets, particularly at hillslope and catchment scales (Benda and Dunne, 1997 in Van Beek, 2002). Sediment originates from soil erosion processes in pre-failure conditions, from landslides during failure and again from soil erosion after the failure occurs. As a consequence, a great deal of attention has been targeting these processes (de Vente et al., 2013; Rozos et al., 2013; Bosco and Sander, 2015) and this has led to increasing our understanding of the processes involved and the parameters required. However, approaches combining both these processes into integrated assessments of catchment hillslope dynamics remain few and far between.

It is still a challenge to improve the estimation and prediction of soil erosion by water at the catchment scale, particularly when trying to consider landslides processes. Landslides strongly interact with surface erosion processes, both directly through their capacity to move a considerable quantity of soil directly into a channel network and indirectly by changing the local topography, vegetation cover and soil properties (Acharya et al., 2009; Cochrane and Acharya, 2011). As for soil erosion modelling, the prediction of spatial and temporal probability of landslide occurrence is still an open challenge (Bosco et al., 2013; Van Westen et al., 2006; Wasowski et al., 2011). Furthermore, both processes potentially interact with the local pattern of vegetation composition and health, and with the human management of land cover (for example, forest resources, agriculture or agroforestry resources). Sediments and geomorphic changes (such as gullies and significant mass movements) may have an impact on how each precipitation event affects the water resources at the local and catchment scale. This impact could potentially influence both water quality (sediment transport dynamics) and quantity (runoff patterns, including flooding frequency and intensity). Disturbances to natural resources may further complicate the chain of interactions.

For example, wildfires may drastically change the protective effect of vegetated land cover. Conversely, intense post-fire soil loss or instability may hamper the recovery of sensitive forest ecosystems or crop cultivations. Therefore, soil erosion and landslides may configure non-negligible dependencies between soil, water and forest resources, and agriculture practices. Models specialised in domain-specific components of this integrated problem may have required several years to be developed. The complex causal network entangling different natural resources and processes shows characteristics cyclic dependencies (de Rigo, 2012c; Figure 1.1), also known as 'feedback' in system science and modelling (Koopmans and Stamovlasis, 2016; Hieronymi, 2013; Richardson, 2009). The transdisciplinary modelling integration required to connect domain-specific model components may also require the investment of several years of research (Laniak et al., 2013; Kelly Letcher et al., 2013). Reliable modelling architectures are necessary for supporting an integrated assessment and management of natural resources and processes, especially considering that multiple domains of expertise are typically involved (de Rigo and Bosco, 2011; de Rigo, 2015). This is also the case for studying the interactions between soil erosion and landslides by integrating a multiplicity of computational-science models and techniques.

This research focuses on the development of an integrated modelling architecture for the assessment of soil erosion by water in data-poor regions affected by slope instability (Figure 1.2). The proposed architecture should be sufficiently flexible to enable the design of future transdisciplinary scenario-analyses. In particular, the architecture might contribute as a novel component to simplify future integrated analyses of the potential impact of wildfires or vegetation types and distributions, on sediment transport from water induced landslides and erosion.

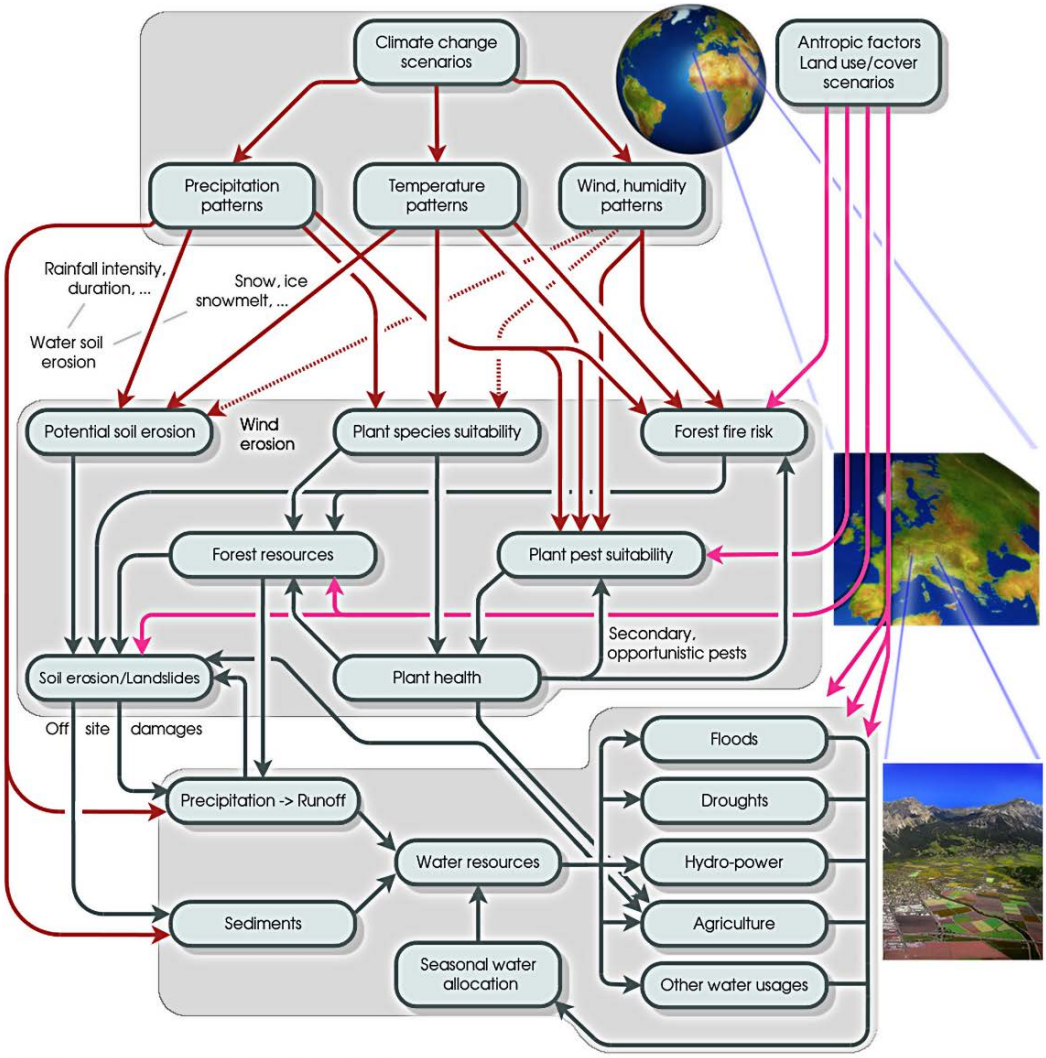


Figure 1.1 - An example of the typical complexity and cyclic dependencies among natural resources relationships. (Credit: Copyright (C) 2010-2015 Daniele de Rigo) (source: de Rigo, 2012c; 2015).

Figure 1.2. – The picture shows some of the soil erosion processes (rills and ephemeral gullies – top picture) and mass movements (shallow landslides – central picture) that characterize the study site, located in Italy within a catchment close to Rocchetta Sant'Antonio. The image also illustrates the heterogeneous land cover patterns that are present in this area (agricultural areas, grassland, shrubs and forest), with uneven patch size and complex connectivity.

1.2. ntegrated natural resources modelling and management (INRMM)

Soil scientists cope with a broad set of problems, each associated with different conceptual and computational tools. For example, the specific theory, modelling methods, and data required for approximating the rate and extent of soil loss differ from the ones needed to study slope instability. Data scarcity and the range of spatial and temporal scales of potential interest may further exacerbate differences, given the variable set of simplifications and assumptions which every specific problem could suggest. This variety of problems may lead to the necessity of a multiplicity of expertise competences which differentiate the sub-domains within soil science. Accordingly, a number of computational models are available, whose normal application is limited to domain experts who master their underpinning semantics. Under normal circumstances and typical domain-specific usage, these semantics can safely remain implicit without a proper cross-disciplinary formalisation. These unexpressed semantics may become an issue when established domain-specific models are considered for integration within a broader context.

As briefly outlined in the previous section, natural resources are part of a complex causal network (Figure 1.1) with complex cyclic dependencies, such as the relationship between soil erosion and land cover. Land cover strongly influences the precipitation-runoff relationship and thus plays a decisive role in mitigating or exacerbating soil erosion and floods. A good level of vegetation cover or good agricultural practices (such as land management reflecting site specific conditions to limit soil loss) can positively reduce soil erosion while a degraded land cover (e.g. caused by wildfires or outbreaks of vegetation pests) or bad agricultural practices (e.g. tilling in the down-slope direction) have negative effects on the soil erosion rate. In turn, the climate can directly or indirectly affect soil erosion by changing the precipitation intensity or driving changes in land cover (de Rigo, 2012c). At the same time, soil erosion has a direct influence on water sediment transport and water resources quality. Vegetation cover, soil, water resources and land use and management are thus intrinsically linked. These few examples illustrate the complex

interaction among soil, forest, and water resources (all instances of natural resources), and man-made resources such as agriculture resources,.

showing the typical multiplicity of dimensions which may guide corresponding modelling-integration efforts. Different models may support different conceptual steps of an integrated analysis. In each specific model, the modelled physical quantity (e.g. tonnes of soil loss per year), or more abstract index (e.g. a dimensionless landslide-susceptibility index), may serve as a decision-making metric to support prioritising the intervention in some more critical areas. Within an integrated perspective, the variety of factors of potential interest is often unsuitable to be assessed using a unique single metric. Multi-dimensional criteria are often needed to assess where the array of interrelated impacts escapes a one-dimensional simplification. An integrated perspective on the modelling and management of natural resources (Integrated Natural Resources Modelling and Management, INRMM, (de Rigo, 2012c) may frequently imply that the set of ecosystem services affected by resource instability or disturbances is better described with a multiplicity of dimensions and criteria (Maes_et al., 2013; Maes_et al., 2016; de Rigo_et al., 2016, Mubareka_et al., 2016).

This general principle may be exemplified in the specific context of this thesis, assessing the interactions of shallow landslides and soil erosion. The slope-instability analysis offered by a single conceptual model may be complementary to a similar analysis perfomed with a statistical model, based on a rather different computational mechanism. Estimates by multiple models may be integrated as different components of a multi-model ensemble. Moreover, a specialised model of soil erosion by water may integrate information produced by the ensemble of landslide-susceptibility models, so that the protection services of different vegetation types may be assessed integrating the dimensions of soil erosion and slope instability.

1.3. Integrating uneven arrays of data and computational components: the role of Semantic Array Programming

This defines the aforementioned wider context, and an inter-domain diversity of computational models and methods which complements the previously discussed intra-domain diversity characterising soil science numerical applications. Again, the multiplicity of existing computational methods is often specific to particular domains (e.g. forest resources, software engineering, signal processing, advanced statistics). Semantic array programming (SemAP) (de Rigo, 2015, 2012a, 2012b) has been introduced as a modelling paradigm to ease the integration of the various conceptual modelling-units by formulating them as data-transformation models (D-TM). D-TM units do not force a user to master their internal details, since they exclusively exchange data (extended to include parameters), with broadly supported formats. An integrated D-TM consists of a chain of D-TM units (whose implementation may also be based on different programming languages) which starting from input data generate a series of intermediate derivative data, up to the final desired output.

SemAP is designed to ease the computational communication between local-contexts, different expertise and disciplines in a simple way – but also a compact and unambiguous one. This is achieved by limiting the potential generality of the exchanged data by means of array-based semantic constraints (de Rigo, 2012d, 2015). The interface of each D-TM unit is formalised so as to annotate the logical pre-conditions required for the D-TM input data to be consistent (input semantic constraints). For example, a certain input may be expected to be a nonnegative column vector, while some of the input arrays may require compatible dimensions for the subsequent computation to be numerically feasible. Analogously, the output arrays of data generated by the D-TM may logically expect some semantic constraints to be respected (so called post-conditions). For the D-TM modules written directly following the SemAP paradigm, also internal semantic constraints (so called invariants) may be periodically checked, to detect anomalies in the algorithm. This semantic annotation of intermediate data is essential at the computational

integration level, so as to prevent inconsistent applications of domain-specific modules.

This research proposes to apply the SemAP paradigm to address the problem of soil erosion and mass movement modelling. This approach provides the support that is required to capture the complexity of the environmental modelling architecture (further information on semantic array programming is discussed in section 2.5), and is an essential premise for the proposed integrated modelling architecture to be easily expanded in future, to include new steps toward a fully integrated environmental analysis.

1.4. Key processes - The interactions between soil erosion by water and mass movements

Following the overview on the computational-science integration challenges and approaches which characterise the novel contribution here proposed, a more specific discussion is needed concerning the core interaction between soil erosion by water and the mass movement of sediments.

Soil erosion by water is particularly high where erodible soils are coupled with a high-energy relief and high intensity rainfalls. Landslides may play an important role within soil erosion process through their capacity to remove and expose large parts of slopes in a relatively short time (Van Beek, 2002). Soil erosion by water and shallow landslides are often the main source of waterway sediment load in hilly catchments (Morgan, 2005; Benda and Dunne, 1997; Acharya et al., 2011). Although spatially and temporally constrained, especially when compared with surface wash, shallow landslides can have a high impact on the sediment budget of a catchment (Van Beek, 2002), for example by displacing large volumes of soil along gully systems created by surface erosion processes (Pla Sentis, 1997), by exposing soils along landslide scars (Valentin et al., 2005; Mazaeva et al., 2013), or by retrogressive movement of a mass movement increasingly exposing soils (Acharya, 2011).

Soil erosion processes can enhance landslide susceptibility as it affects the vegetation cover, changes the local topography and changes the properties of the near surface deposits, altering the hydro-geology and reducing the shearing resistance of these materials (Rozos et al., 2013; Lee, 2004; Popescu, 1994; Cochrane and Acharya, 2011; Acharya et al., 2009). The spatial distribution of the reworked sediments and their particle size composition play an important role in soil erosion processes (Sander et al., 2011) and mass movements, often in the form of mudflows, can form where significant accumulations of eroded slope materials are found (Nearing et al., 2005).

The changes in the rate of soil erosion by water, occurring after a landslide has taken place, can be strong enough to impact on ecosystems services. Ecosystem services consist in the benefits that the natural environment and a properly-functioning ecosystem bring to people (de Rigo et al., 2016; Maes et al., 2016). The magnitude and frequency of landslide-derived sediment yields is also of critical importance for the safety of settlements and infrastructure located in downslope areas (Acharya, 2011). However, further research is needed to achieve a better understanding of the effects and evolution of landslide-altered topographies on soil erosion. In turn, this will assist with the development of improved management of land and water resources (Acharya et al., 2009).

1.5. Aims and objectives of the thesis

In the previous sections, some core components of the proposed research contribution were introduced in their context. A synthesis of aims and objectives is here offered.

The overall aim of this research is to better integrate and quantify the role of shallow landslides within soil erosion process modelling in data-poor regions.

Hillslope processes can be envisaged as a cascade where surface erosion and mass movements are visible expressions of critical instabilities in a complex system of interacting processes that control the downslope movement of material (Van Asch, 1980 in Van Beek, 2002). Landslide events can result in changes in soil erosion rates

that are strong enough to deliver significant cascading impacts on ecosystems. To support the integrated assessment of these processes, the design and development of reliable modelling architectures is required. This research,focused on data-poor regions, proposes a new semi-quantitative method for better estimating the contribution of shallow landslides on soil erosion losses in areas affected by slope instability and with a limted data availability, by combining heuristic, empirical and probabilistic approaches.

The main objectives of this thesis are to:

1) improve, in data-poor regions, the estimation and prediction of soil erosion by water in catchments affected by shallow landslides triggered by water.

2) Develop a robust approach to reduce the uncertainty in shallow landslide susceptibility assessments in data-poor condition.

The main activities of this study are to:

1) Select and apply appropriate soil erosion models to estimate soil erosion by water in data-poor regions.

Despite the efforts of the scientific community, the predictive value of soil erosion models is still limited, especially when the necessary data for running and calibrating the models (e.g. soil structure or temporal high resolution rainfall data) are poor or lacking (Govers, 2011). An evaluation of quantitative soil erosion and sediment yield models resulted in the selection of two approaches for further modification as they have the greatest flexibility for modelling soil erosion in data-poor regions and are suitable for application at a catchment scale.

The first of the selected soil erosion models is an extended version of the Revised Universal Soil Loss Equation (RUSLE) (Renard et al., 1997). The model was applied, within the study area, for estimating the pre- and post-failure rate of soil erosion by water. The other model is the revised version of the Morgan-Morgan-Finney model (MMF) (Morgan et al., 1984). Despite its suitability to run in data-poor areas, due to the lack of some of the input data, we decided to use only the first of these models for estimating the soil loss within the catchment.

2) Develop a robust approach for modelling landslide susceptibility in a data-poor area.

The assessment of spatial probability is one of the main challenges in landslides modelling. Despite the many different approaches (heuristic, deterministic and statistical; van Westen et al., 1997) spatial landslide susceptibility assessment remains a challenge. Different techniques for improving the spatial probability prediction of shallow landslides in data poor regions were investigated.

A new semi-quantitative method, based on an ensemble approach, was used for combining deterministic and probabilistic approaches in order for the uncertainty to be mitigated.

The application of an ensemble approach, especially in data poor regions, could potentially reduce the uncertainty and mitigate local poor performance associated with individual models, by excluding outlier estimations.

3) Estimate the effects of shallow landslides on water-induced soil erosion in a data-poor catchment

An in-depth analysis of the relationship between soil erosion and shallow landslides within the hydrological system (the set of interacting or interdependent component parts forming a complex whole) was carried out. A semi-quantitative modelling methodology to support the integrated assessment of soil erosion, by incorporating rainfall induced shallow landslides processes in data-poor regions, was developed and tested in the study area.

1.6. Assumptions and constraints

This research was designed and undertaken in collaboration with the National Research Council – Institute for Geo-Hydrological Protection (CNR-IRPI) of Bari (Italy). A field survey, with important local support from Dr. Wasowski (CNR-IRPI, Bari) was designed to collect information on soil cohesion, soil texture, soil moisture, bulk density and plant height (mainly necessary to run the revised MMF (RMMF; Morgan,

2001) model and to predict shallow landslide susceptibility applying the infinite slope stability model) on a number of different plots.

The project results were constrained by the quality and coverage of the data obtained during the field survey and the availability of data provided by the CNR-IRPI. Unfortunately, adverse weather condition during the field season (a prolonged dry spell with temperatures well above the historical average) resulted in a reduced set of data (because of the difficulties in sampling) that was interesting in its representation of extreme dry conditions, but far from characteristic of long term conditions.

The thesis is therefore mainly based on freely available datasets (described in section 4.2.3.1) and three datasets provided by Dr. Wasowski (CNR-IRPI, Bari):

i) a vector map of the mass movement which occurred in 2006 within the study area (figure 3.6)
ii) a land cover map of the catchment based on ASTER imagery (figure 3.8)
iii) a digital elevation model (raster) of the study site at a resolution of 5 metres.

Mass movement map

This landslide inventory was made available in a geospatial vector data format (Esri Shapefile) projected in Universal Transverse Mercator coordinate system (WGS1984 UTM Zone 33N). It contains more than 400 landslides that affected the catchment in the year 2006 (more details in chapter 3) ranging from few to thousands of square meters. This dataset is based on high resolution IKONOS satellite imagery and was created by the CNR-IRPI of Bari (Wasowski et al., 2010).

Land cover map

Classified land cover for the year 2000 projected in WGS1984 UTM Zone 33N. The dataset has a resolution of 5x5 metres. The land cover database consists of four different classes (pastures, crops, grass and woods) having a resolution of 5x5 metres, and was produced by CNR-Irpi (Bari) using ASTER imagery (July 2000). Four ASTER bands were analysed and used for classifying this area (Wasowski et al., 2010). Further information is available in section 3.5.

Digital elevation model

The Digital Elevation Model (DEM) has a resolution of 5x5 meters and WGS1984 UTM Zone 33N as coordinate system. It covers the whole study site and was obtained by researchers of the CNR-Irpi of Bari from a 1:5,000 scale topography map of the area (Wasowski et al., 2012).

The lack of detailed data in this case provided an ideal basis for this research to develop innovative modelling solutions that would work in data-poor conditions.

1.7. Thesis structure

This thesis contains seven chapters. This introduction provides the outline of the research and a further six chapters provide greater detail of key aspects of the research.

Chapter two covers an overview of the physical processes of slope stability and soil erosion by water and analyses existing integrated modelling approaches.

Chapter three describes the physiography of the study area near Rocchetta Sant'Antonio and the Daunia region in southern Italy.

Chapter four develops an approach to model soil erosion in data poor regions.

Chapter five addresses the multi-scale robust modelling approach to estimate landslide susceptibility.

Chapter six forms the core of the thesis and outlines the coupled architecture for modelling the effects of shallow landslides triggered by water on soil erosion using a case study in the Rocchetta Sant'Antonio catchment.

Chapter seven presents a summary of the obtained results, and provides suggestions for future research.

2. Literature review

2.1. Soil erosion and slope stability processes: an overview

Land degradation can be regarded as any change or disturbance to the land perceived to be deleterious or undesirable (Johnson et al., 1997). Land degradation is an important issue globally and can be the result of multiple interacting processes, including soil erosion, soil sealing, soil compaction, the decline in organic matter, mass movements, salinization, contamination and biodiversity decline (Montanarella, 2007). All these processes can lead to a reduction of the potential agricultural productivity with a consequent high impact on food security. For example, the productivity of some lands can decline up to 50% due to soil erosion and desertification (Eswaran et al., 2001). During the last 40 years, as a result of soil erosion, about 30% of the world's cropland has become unproductive (Pimentel and Burgess, 2013). Estimates of the extent of land degradation vary, but approximately one third of the world's arable land has been affected by degradation and desertification to date (UNCCD, 2015).

Although for a holistic analysis of land degradation all the involved processes are important, here we will focus on soil erosion by water and mass movements being these processes the foundation of this research that aims to better integrate and quantify the role of landslides in soil erosion.

Soil erosion

According to Huber et al. (2008): "Soil erosion is a natural process that has been largely responsible for shaping the physical landscape we see around us today, through distribution of the weathered materials produced by geomorphic processes".

Soil erosion is the wearing away of the land surface by physical forces such as rainfall, flowing water, wind, ice, temperature change, gravity or other natural or

anthropogenic agents that abrade, detach and remove soil or geological material from one point on the earth's surface to be deposited elsewhere. Soil erosion is a complex phenomenon influenced by very diverse factors such as land cover, climate and topography, and strictly linked to human practices and activities (Guerra et al., 2017; Goudie and Boardman , 2010) that can also exacerbate this process (Bosco et al, 2009). While land cover affects soil erosion either positively (i.e. forests cover and good agricultural practices) or negatively (wildfire-degraded cover and bad agricultural practices (Foley et al., 2005)), climate affects soil erosion, both indirectly by driving land cover changes and directly by varying precipitation intensity and duration. At the same time, soil erosion influences water sediment transport, water resources quality and water storage loss (Hansen and Hellerstein, 2007).

Referring to soil erosion as a threat to soil implicitly means 'accelerated soil erosion'. The threshold above which soil erosion should be regarded as a major problem is controversial. The soil formation processes and rates vary substantially spatially. For example, considering the European continent, in Switzerland, the tolerable soil erosion rate is generally 1 t ha^{-1} yr^{-1}, that can increase to 2 t ha^{-1} yr^{-1} for some soil types (Schaub and Prasuhn, 1998). Verheijen et al. (2009) report a general upper limit of 1.4 t ha^{-1} yr^{-1} in Europe, while 2t ha^{-1} yr^{-1} is the threshold in Norway for considering the soil loss as tolerable (Srebotnjak et al., 2010). For establishing what could be the tolerable soil erosion the soil formation rates were proposed as a basis. In Europe the current scientific knowledge indicate that 0.3 – 1.4 t hs^{-1} yr^{-1} is the rate of a tolerable soil loss (Verheijen et al., 2009). This range depends on the rate of dust deposition and on the driving factors of weathering (e.g. parent material, climate, etc.). Soil erosion by water accounts for the greatest loss of soil in Europe compared to other erosion processes (e.g. wind erosion) (Panagos et al., 2015) and the recent policy developments in the European Commission, as the Soil Thematic Strategy (EC, 2006) and the 7th Environmental Action Programme (EP and Council, 2013), call for quantitative assessments of soil loss rates. We here focused our attention in improving modelling techniques for assessing soil erosion by water, also considering that sheet and rill erosion (see following paragraph) are the dominant types of erosion (Pimentel and Burgess, 2013) all over the world.

Soil erosion by water

Water-induced soil erosion can result from rainfall, snowmelt or artificially by irrigation (Foster, 1982), and there are three main processes involved: detachment, transport and deposition (Figure 2.1). Soil erosion occurs in various forms such as splash, sheet, rill or gully erosion (Morgan, 2005) depending on the stage of progress in the erosion process and the position in the landscape (Figure 2.2). These processes are briefly explained below.

Detachment - The soil detachment essentially involves two processes, the impact of raindrops on soil surface, where it overcomes the interstitial force of soil particles (rainsplash erosion), and the flow traction. Rainsplash action is only effective if the rain falls with sufficient intensity. The kinetic energy of raindrops is able to detach and move soil particles a short distance (Bryan, 2000). Although considerable quantities of soil may be moved by rainsplash, it is generally all redistributed back over the surface of the soil. On steep slopes, there can be a modest net downslope movement of splashed soil due to the effect of gravity and the gradient of the land. The main consequence of rainsplash erosion is to weaken the soil surface for transport by overland flow (Morgan, 2005).

Transport - Transportation of soil particles occurs by surface runoff. It occurs when the amount of water accumulating on the soil surface exceeds the infiltration capacity of the soil and excess water from rain, meltwater or other sources, flows over the land as a sheet (Beven, 2004).

Deposition – "The deposition of soil occurs when the transport capacity of overland flow becomes smaller than the settling velocity of particles owing to gravity. These particles are loosely deposited and can be easily remobilized" (Saavedra, 2005). For example, on the upslope part of erosion plots, where the flow velocity is low because of a small flow-contributing area as well as a short slope length, sediment is easily trapped in depressions or in channel beds. It is remobilized mainly by raindrop impacts. Downslope, an increase in flow velocity enhances the soil particle remobilization and rain-impacted flow transport, mainly depending on slope steepness (Saavedra, 2005).

Sheet erosion - Sheet erosion is the uniform, evenly distributed detachment of soil from the soil surface (Hairsine and Rose, 1992). Sheetwash erosion occurs without any well-defined channel and can manifest itself across entire slopes. As a consequence, the erosion can affect large areas and move significant amounts of soil.

Rill erosion - Rills occur when overland flow begins to develop preferential flow paths. In turn, these flow paths are eroded further which results in small, well-defined concentrations of overland water. In many cases, small rills may disappear over time due to sedimentation. However, persistent micro-rills can develop further to become rills with a subset eventually becoming gullies (Figure 2.2). (Nearing et al., 1997; Saavedra, 2005).

Gully erosion - Gullies are deeper channels, often resulting from unchecked rill erosion. Due to their size, gullies are capable of moving large amounts of soil, into larger channels such as streams and rivers and thus out of the original site. Gully erosion is often the main source of sediments in a catchment (Valentin et al., 2005).

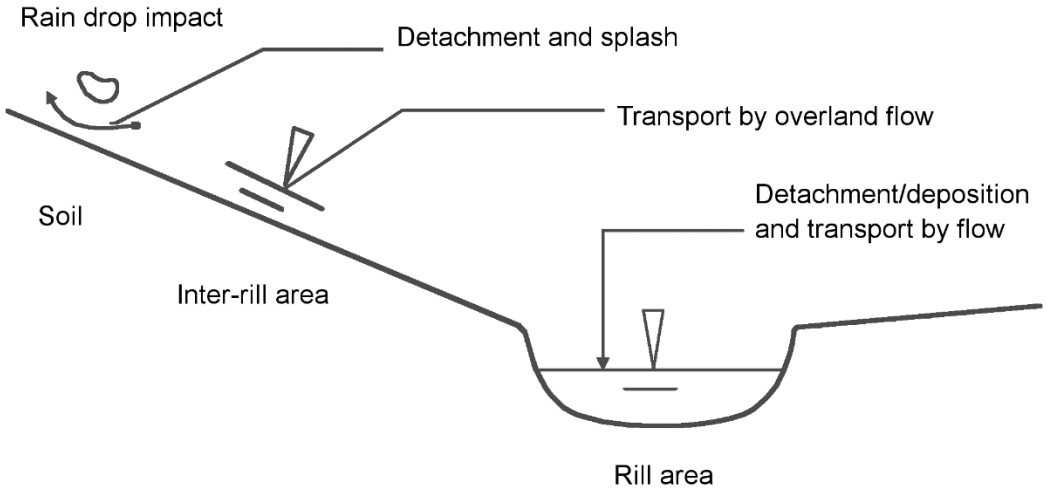


Figure 2.1 - Soil erosion and transport on inter-rill and rill areas (source: Doe and Harmon, 2001).

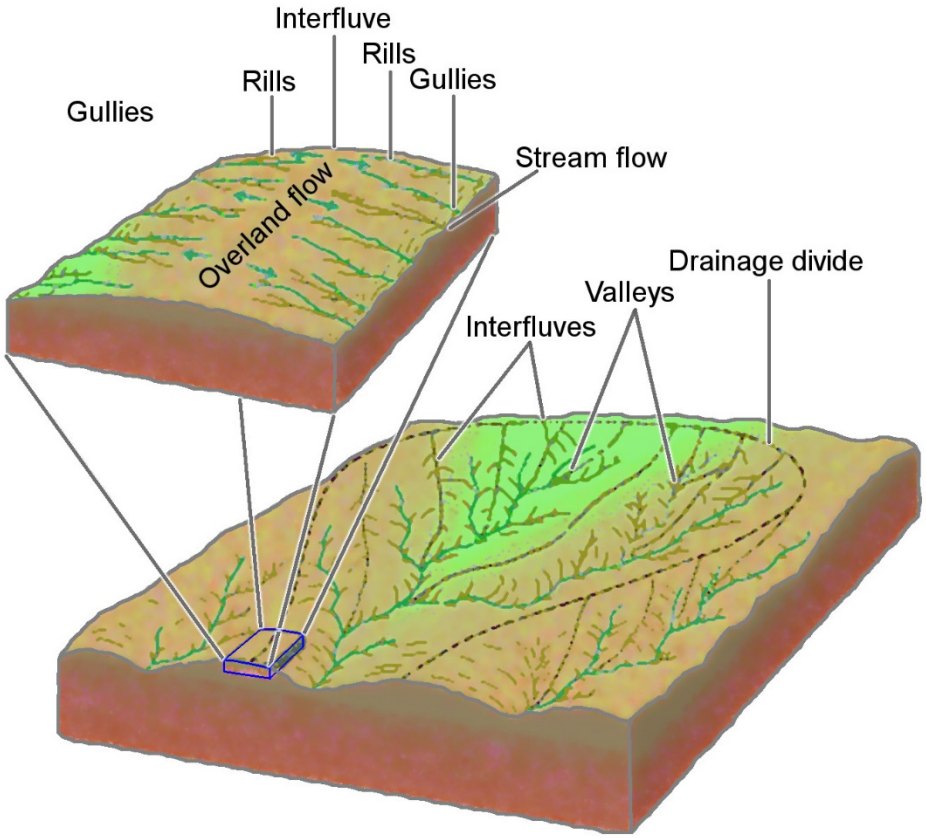


Figure 2.2: Schematic representation of rill, inter-rill areas and gullies in a sub-catchment (source: Saavedra, 2005).

Landslides

The term 'landslide' generally denotes a downslope movement of earth, rock or debris due to the action of one or more external forces acting together. Rainfall, earthquakes, volcanic eruption or anthropogenic activity are only some of the numerous forces capable to generate landslides. Following Varnes (1978) a landslide can be generally classified by two names, the first used to describe the materials forming the landslide (e.g. earth, rock, mud or debris) and the second that represent the type of mass movement (e.g. falls, slides, topples).

The landslides occur when stresses acting on a soil mass on a hillslope exceed the soil strength. It has generally been recognized that these forces are functions of various parameters relating to bedrock geology, lithology, geotechnical properties, rainfall characteristics and duration, groundwater conditions and land-use patterns. As well

as natural factors, in many cases human interferences are also responsible for triggering the landslides and create the same effects on a slope as a range of natural processes. Some of the common examples of human interferences leading to landslides are changes in land-cover, deforestation and cutting of slopes (Lanni, 2012).

The focus is on rainfall triggered shallow landslides involving the downslope movement of soil or rock occurring predominantly on the surface of the rupture or in relatively thin zones of intense shear strain within the soil mantle or weathered bedrock (typically to a depth from a few decimetres to several metres).

It is now widely accepted to consider that water-induced shallow landslides in steep hillslopes are triggered by significant rainfall events which substantially increase the soil pore pressure (Bordoni et al., 2015; Anagnostopoulos et al., 2015; Tohari et al., 2007). An increase in pore pressure reduces the soil's shear strength eventually leading to slope failures. It is typically observed that significant pore pressures are generated in the lower areas of a hillslope (Tohari et al., 2007; Anderson and Sitar, 1995). However, the generation of pore pressure depends on various site specific factors related to hydrology, topography and soil properties.
Water plays a major role not only in the initiation of failure, but also in the way that the earth then flows or slides and the distance that the landslide mass travels. Often, shallow landslides move fast and can be extremely destructive.

The protection of soil has been recognised as one of the main challenges to society, addressing this challenge has therefore formed the focus of many environmental policies (CEC, 2006). As mentioned in Chapter 1, it is well documented that soil erosion leads to a decline in organic matter and carbon cycling, a reduction of crop productivity and water storage capacity, a breakdown of soil structure and a host of other processes such as enhanced siltation of streams and reservoirs, and enhanced flood risk (Pimentel and Burgess, 2013; Quinton et al., 2010; Bosco et al., 2009; Boardman, 2006; Bakker et al., 2004) and it is also closely related with an enhanced susceptibility of a landscape to generating mass movements (Larsen et al., 2010; Burton and Bathurst, 1998). Mass movements are often more confined in both space

and time than soil erosion, essentially a highly disperse manifestation of mass re-distribution in a landscape.

Good qualitative and quantitative data sets and process models are required at a range of scales to enable the evaluation of management strategies that aim to reduce the negative impact of these processes on the economic, social and environmental development of sensitive regions. Although past research has identified the key mechanisms involved (Morgan and Nearing, 2016; Shi et al., 2012; Morgan, 2005; Bryan, 2000; Sidle and Ochiai, 2006; Montrasio and Valentino, 2008), soil erosion and landslide modelling still faces some fundamental problems. These include the lack of high resolution input data, the processes considered within the models (e.g. rill erosion, gully erosion, sediment deposition etc.) and the complex interactions among the involved processes. Several studies documented the large impact of landslides on catchment sediment yield (Figure 2.3), but approaches combining both soil erosion and slope instability into integrated assessments of catchment hillslope processes remain few and far between (de Vente et al., 2013).

The capacity of the existing models to consider, at the same time, the spatial and temporal probability of landslides and soil erosion occurrence will be discussed in the following paragraphs.

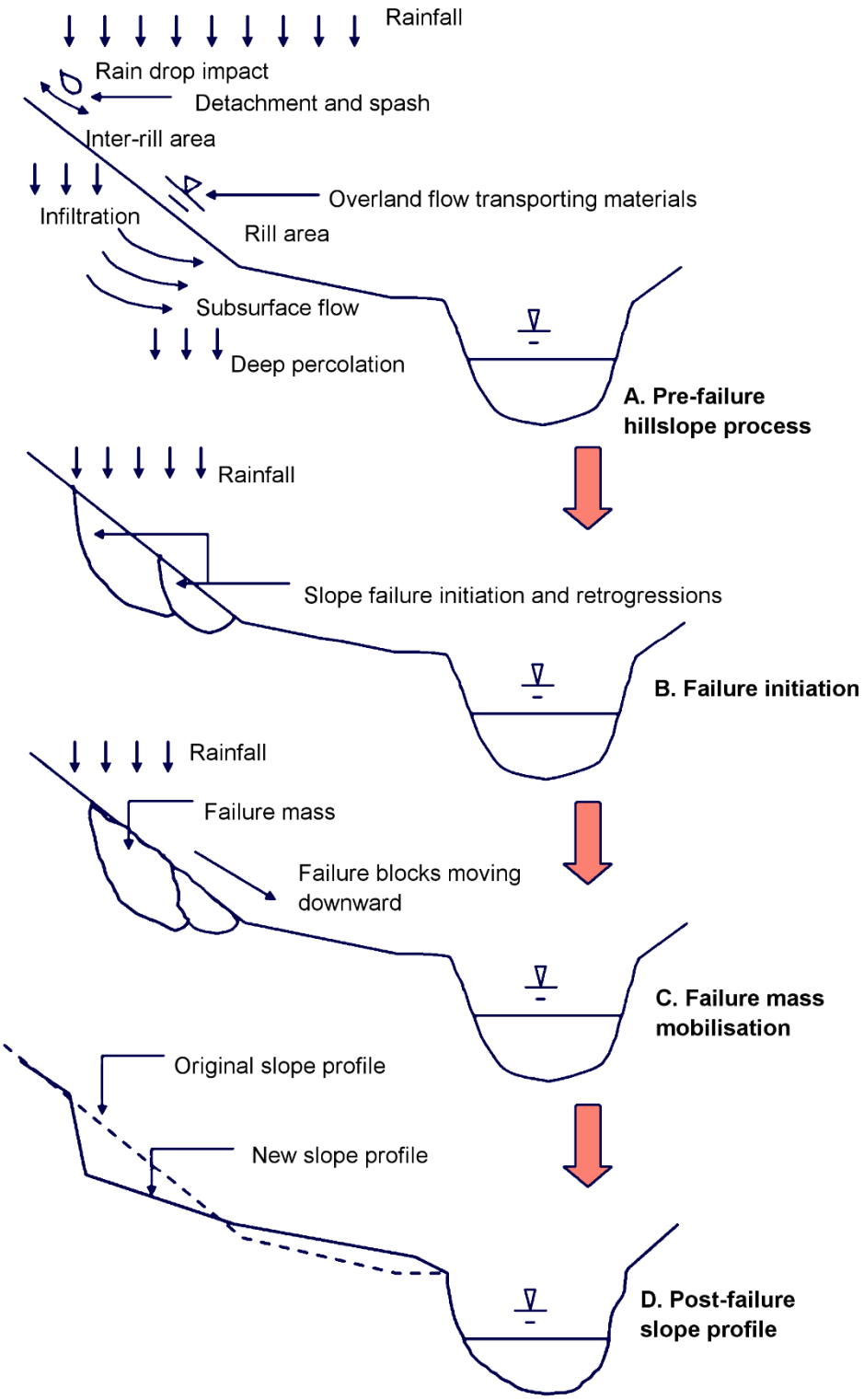


Figure 2.3: Schematic diagram showing different phases of sediment transport in a hillslope: (A) initiation of overland flow and soil erosion, (B) failure initiation and retrogressions, (C) mobilisation of failure materials and (D) changes in hillslope profile following landslides (source: Acharya, 2011).

Both soil erosion and landsliding are manifestations of critical instabilities in catchments and form important indicators of the health of our landscapes. It is impractical to directly measure soil erosion in the field at a landscape or larger scale, and therefore a modelling approach is necessary (Bosco et al., 2015). Many limits in modelling soil erosion still exist; most of the developed models can suffer from a plethora of problems such as over-parameterisation, unrealistic requirement of input parameters or the unsuitability to the local conditions of the modelled processes (Merritt et al., 2003). One of the major limitations in modelling soil erosion is that its interactions with mass movements and riverbanks are often not considered (de Vente et al., 2013). Landslides strongly interact with surface erosion processes, directly and indirectly. As already mentioned, they have a capacity to move a considerable quantity of soil directly into the stream network and change the local topography and soil properties (Acharya et al., 2009). As for soil erosion modelling, the prediction of spatial and temporal probability of landslide occurrence is still an open challenge (Van Westen et al., 2006; Bosco et al., 2013) and its impact is therefore difficult to include.

2.2. Model complexity and accuracy

2.2.1. Physically based, conceptual and empirical modelling

The practical suitability of physically based models is debated; the deterministic verification of any model outcome is largely not possible, to exactly reproduce a feature of nature that is the outcome of a highly non linear system (having, generally, poorly known initial and boundary conditions) it is almost impossible (Bras et al., 2003; Bosco et al., 2015). Theoretically, physically process based models have the greatest potential to be applied in environmental modelling but their optimisation for the local condition of small catchments is a strong limitation. Often they show poor predictive capabilities if applied in different catchments and conditions (de Vente et al., 2013). In addition, the enormous gap between the richness and accuracy of the input parameters required by physically based models and the actual

availability of verifiable large-scale datasets limits their applicability (Stroosnijder, 2005). The input parameters uncertainty is probably the main reason why more complex physically based models do not generally obtain better results than regression-based models, especially if applied at the catchment or larger scale.

More complex models with better process descriptions should, in principle, be capable of better output forecasts. Jetten et al. (2003) found that the introduction of additional parameters into a model introduce an additional error that often outweighs the potential improvement linked with a better process description. Anyway, models are never totally physically based, numerous authors showed that for obtaining reasonable results it is necessary to calibrate the models (Hessel et al., 2003).

The growing desire by decision makers to use models for efficiently capturing and measuring the spatial and temporal aspects of soil erosion and landslide susceptibility feeds an effort to improve the performance of these models. The development of a distributed model with plausible physical basis and relatively low complexity is particularly attractive. However, the environmental complexity due to interactions of physical, biological and chemical processes is very high. Many of the environmental processes are nonlinear, with considerable uncertainty about their nature and their interconnections. Under these conditions stochastic, dynamic models should be the rule rather than the exception; the uncertainty which pervades most environmental systems demands an alternative stochastic approach to the deterministic mathematical equations based on well known scientific laws (Young, 2002). This is discussed in further detail below.

To overcome the limits of both the physically based and empirical approaches it would be interesting to explore the applicability of some core concepts derived from a Data Based Mechanistic (DBM) approach (Young and Lees, 1993; Young, 1998).

Data-Based Mechanistic models constitute a class of models that have intermediate characteristics with respect to physically based and empirical models. Young and Lees (1993) were the first to use the term ‘data-based mechanistic modelling’, but this concept had been around and developed considerably over several decades. The first applications in a hydrological context were published in the early 70s modelling

rainfall and flow processes (Whitehead and Young, 1975; Young, 1974) and water quality (Beck and Young, 1975). These approaches share some of the properties that are typical to empirical models, in particular the inductive approach to identification, in contrast to the hypothetical-deductive approach that is typical of mechanistic models. However, they are similar to mechanistic models in their attempt to provide at least a partial description of the physical behaviour of the system. After an initial black-box modelling stage has been carried out, the model follows a mechanistic approach on the physical laws that are most likely to control the behaviour of the system under study (Young, 2002).

In particular, a recurring DBM methodology relies on identifying some mathematical properties of the process to model, based on the physical nature of the process, and to constrain the numerical model under development to automatically satisfy these properties. For example, knowing that a physical quantity represents the share of terrain (from 0 % to 100 %) with a certain characterisation of interest, an empirical model trained with available measured data is generally unable to guarantee that the estimated quantity will not occasionally have values higher than 100 %.
However, this semantic property might be added into the mathematical formulation of the empirical model. This way, the model becomes slightly less general but physically consistent.

In recent years the hydrological literature has shown an increasing interest in the top-down approach (e.g. Jothityangkoon et al., 2001; Son and Sivapalan, 2007; Willems, 2014; Hrachowitz and Clark, 2017) due to the relative failure of the reductionist approach in this field. It involves starting with the simplest model configuration at a large time scale (i.e. annual), and gradually increasing the complexity of the model with decreasing time scales (annual to monthly and finally to daily), in response to an evaluation of the model predictions at each time scale. To follow the same direction in modelling soil erosion, landslide susceptibility and their integration into one modelling approach is therefore worth investigating.

2.2.2. Soil erosion by water and landslide prediction: an overview on capabilities and limitations

2.2.2.1. Soil erosion by water: models and their limitations

Many different models exist for estimating the process of soil erosion by water. These differ in terms of complexity, data requirement, spatial and temporal scale, physical processes and in the manner that these processes are represented (Table 2.1). In-depth reviews on different soil erosion models are available in many different publications such as: Merritt et al. (2003), Aksoy and Kavvas (2005), de Vente and Poesen (2005), de Vente et al. (2013) and Pandey et al. (2016), with regard to the previously mentioned factors. Table 2.1 summarises 33 different soil erosion models in terms of their classification, scales of application and input data requirements. These models are classified in physically based (yellow-green colour in the table) (based on the solution of fundamental physical equations), conceptual (in salmon) (when exists a non-physical but conceptually meaningful relation between the elements of the process) and empirical (in blue) (based on the analysis of observations). The model classification refers to the over-arching process representation of the model. These models were developed for a wide range of applications, over a range of different scales (from the plot to the regional scale approaches) and, regarding the temporal scale, can be classified as event-based models (that model within-storm runoff and soil erosion processes), continuous simulation models (calculates erosion through the year and over many years) or annual (for models calculating only yearly averages). The data requirement, including the spatial and temporal variation of model input, range from low to high (typical of physically based models).

Table 2.1 – List of the analysed soil erosion and sediment transport models and of their main characteristics (spatial and temporal scale, type of the model and level of the data requirement). The acronyms and associated references in this table are discussed in the main text.

Model	Type	Spatial Scale	Temporal Scale	Data Requirements	Reference
MUSLE	Empirical	Catchment/regional	Event-based	Moderate/high	Williams, 1975
RUSLE	Empirical	Hillslope/catchment/regional	Continuous	Moderate	Renard et al., 1997
RUSLE2	Empirical	Hillslope/catchment/regional	Continuous	Moderate	Foster et al., 2000
IHACRES-WQ	Empirical / conceptual	Catchment	Continuous	Low	Jakeman et al., 1990, 1994; Dietrich, 1999
RMMF	Empirical / conceptual	Hillslope/catchment	Annual	Moderate	Morgan, 2001
SedNet	Empirical / conceptual	Catchment/basin	Annual/continuous	High	Prosser et al., 2001
USPED	Empirical / conceptual	Catchment	Event/annual	Moderate	Mitasova et al., 1996
AGNPS	Conceptual	Small catchment	Event-based	High	Young et al. (1989)
EMSS	Conceptual	Catchment	Continuous	Low	Vertessey et al. (2001)
HSPF	Conceptual	Catchment	Continuous	High	Johanson et al., 1980
LASCAM	Conceptual	Catchment/basin	Continuous	High	Viney and Sivapalan, 1999
SEMMED	Conceptual	Regional	-	Moderate/high	de Jong et al., 1999
SWAT	Conceptual	Catchment/basin	Continuous	High	Arnold et al., 1998; Arnold and Fohrer, 2005
SWRRB	Conceptual	Catchment	Continuous	High	Arnold et al., 1990
WATEM/SEDEM	Conceptual	Catchment	Annual	Moderate	Van Oost et al., 2000; Van Rompaey et al., 2001
ANSWERS	Physical	Small catchment	Event/continuous	High	Beasley et al. (1980)
CREAMS	Physical	Plot/field	Event/continuous	High	Knisel (1980)
EGEM	Physical	-	Continuous	-	Woodward, 1999
EROSION 3D	Physical	Catchment	Event-based	High	von Werner, 2004
EUROSEM	Physical	Small catchment	Event	High	Morgan et al., 1998a,b
GUEST	Physical	Plot/field	Continuous	High	Misra and Rose 1996; Rose et al. 1997
KINEROS2	Physical	Hillslope/small catchment	Event	High	Smith et al., 1995a,b
LISEM	Physical	Small catchment	Event	High	de Roo et al., 1994
MEDRUSH	Physical	Large catchment	Continuous	-	Kirkby et al., 1998a; Kirkby, 1998b
MIKE-11	Physical	Catchment	Continuous	High	Hanley et al. (1998)
PERFECT	Physical	Plot/field	Continuous	High	Littleboy et al. (1992)
PESERA	Physical	Hillslope/regional	Continuous	High	Kirkby et al., 2003
SHETRAN	Physical	Catchment	Event	High	Ewen et al., 2000
SIBERIA	Physical	Catchment	-	Moderate/low	Willgoose et al., 1991 a-d
SIMWE	Physical	Field/small catchment	Event-based	High	Mitas and Mitasova, 1998
STREAM	Physical	Catchment/regional	Continuous	Low	Cerdan et al., 2001
TOPOG	Physical	Hillslope		High	O'Loughlin, 1986; CSIRO, 2017
WEPP	Physical	Hillslope/catchment	Continuous	High	Laflen et al. (1991)

As noted earlier in section 2.1, soil detachment, sediment transport and deposition are the main soil erosion processes. Although the importance of overland flow has now been recognised, in the past the detachment of soil from the surface was considered to be only the result of raindrop impact. The main involved processes, leading to sediment detachment, differ in different conditions (Merritt et al., 2003).

As previously mentioned, raindrop impact generates a locally intense shear stress at the soil surface, resulting in rainfall detachment (Loch and Silburn, 1996). Likewise, overland flow causes a shear stress on the soil surface which, if it exceeds the cohesive strength of the soil, results in flow detachment. Sediment transport capacity is the maximum sediment flux that can be transported by a flow. Several equations, depending on different hydraulic variables such as shear stress, flow discharge or unit stream power, may be applied (Wainwright and Parsons, 1998). In general, the sediment transport formulas, in soil erosion modelling, belong to the following family of equations:

$$DF = \alpha(\tau - \tau_{cr})(q_c - T_c), \qquad (2.1)$$

where *DF* is flow detachment (kg m^{-2} s^{-1}), α (s^2 kg^{-1}) is a parameter, $\tau - \tau_{cr}$ (kg m^{-1} s^{-2}) represents the excess shear stress and $q_c - T_c$ (kg m^{-1} s^{-1}) the difference between sediment flux and transport capacity.

Physically based models

Soil erosion physics-based models are based on the solution of fundamental physical equations describing streamflow and sediment and associated generation in a catchment. A sediment transport equation is present within every physically based soil erosion model (e.g. WEPP (Laflen et al., 1991), PESERA (Kirkby et al., 2003) and LISEM (de Roo et al., 1994)). Both sediment concentration and sediment load can be used for calculating the sediment transport capacity but the concentration is normally considered as a more fundamental variable (Aksoy and Kavvas, 2005). The majority of the transport capacity equations derive from relationship initially applied in alluvial rivers and were adapted for shallow overland flow (Merritt et al., 2003).

In the rare tests comparing process-based soil erosion models against measured data, major discrepancies are reported (Boardman, 2006; de vente et al., 2013; Pandey et al., 2016). Takken et al. (1999), in a rigorous test applying LISEM with field data from an extreme precipitation event, showed that soil erosion rate is over-predicted for well-vegetated fields. The ANSWERS model has been evaluated in different environments worldwide. Its application reveals that the predicted runoff rate is generally very close to the observed data but the sediment yield is generally underestimated (Pandey et al., 2016). Another process-based model (MIRSED-WEPPP) overestimates erosion and predicts erosion where it doesn't actually occur (Boardman,2006; Mahmoodabadi and Cerdà, 2013). The failure of MIRSED-WEPP in predicting soil erosion is, as suggested by the authors, probably due to the assumption about runoff being generated by Hortonian overland flow while it is probably due to saturation excess (Boardman, 2006). Event-based models, that model within-storm runoff and soil erosion processes, are sensible to initial conditions that are often difficult to specify. The scale at which the majority of soil erosion data was collected (experimental plot) has probably inhibited the development and validation of erosion models suitable for the landscape scale (Boardman, 2006). The plot scale is totally inadequate for exploring the effects of extreme events on soil erosion (Baffaut et al., 1998).

It could also happen that the models well simulate erosion data at the outlet, but fail in reproducing the spatial detail of runoff and soil erosion within the catchment (Jetten et al., 1999; Thapa, 2010), thus providing the right results but for the wrong reason. Therefore, validation of models using only outlet data is not reliable. Only by using spatially distributed data is it possible to validate the behaviour of spatially distributed soil erosion models (Hughes and Croke, 2011).

Empirical models

Empirical models are generally the simplest of all the model types discussed in section 2.2.1. They are based on the analysis of observations and seek to characterise response from the data (Merritt et al., 2003, Pandey et al., 2016). As previously discussed their data requirement is usually lower than physically based and

conceptual models making them able to be supported by coarser data measurements and limited data (Pandey et al., 2016). Empirical models are often ideal for the analysis of data within catchments because many of that are based on the application of stochastic techniques to analyse catchment data (Wheater et al., 1993) (e.g. stochastic regression based primarily on observations).

One of the main critiques to empirical models is to employ unrealistic assumptions about the physiscs of the catchment ignoring the ethereogenity of the catchment characteristics (such as rainfall) and the non-linearity that are present within the system (Wheater et al., 1993). Empirical models are also based on the assumption that underlying conditions remain unchanged during the study period so that tend not to be event-responsive. Nonetheless all these limits, empirical models are frequently used in situations with limited data and parameter inputs for al the reasons that were discussed in section 2.2.1.

Conceptual models

In conceptual models, a catchment is represented as a series of internal Storages (Pandey et al., 2016). This family of models generally include a general description of soil erosion processes within the catchment, without including any specific details on process interactions, which would require detailed input data (Merritt et al., 2003; Pandey et al., 2016). In conceptual models there is a non-physical but conceptually meaningful relation between the elements of the process, these models play an intermediate role between empirical and physically based models (Aksoy and Kavvas, 2005; Pandey et al., 2016). Their architecture allows to this family of models to provide a qualitative and quantitative indication of the effects produced, for example, by land use changes within the catchment, without requiring a large amount of spatial and temporal data (Merritt et al., 2003). Typically, the values of parameters necessary to apply conceptual models are obtained through calibration against observed data (such as, for example, stream discharge and concentration measurements) (Hajigholizadeh et al., 2018). Because of that, conceptual models suffer from problems associated with the identifiability of their parameter values (Hajigholizadeh et al., 2018).

Most of the models capable of modelling erosion only predict a selection of the involved processes (sheet, rill, gully and in-stream erosion) (de Vente et al., 2013; Govers, 2011; Merritt et al., 2003). Sheet erosion together with rill erosion can be classified as overland flow erosion (Merritt et al., 2003). As defined by Loch and Silburn (1996) rills are small erosion channels that can be easily obliterated by tillage whose initiation is controlled by the soil cohesion and shear forces. Gullies are channels too deep to be easily obliterated by cultivation (Loch and Silburn, 1996). In gully erosion process, raindrop impact is no longer a significant factor in terms of particle detachment due to the greater flow depth absorbing the raindrop energy. Finally, the in-stream erosion involves the direct detachment of soil particles from stream banks or bed. All these soil erosion processes do not necessarily occur at different times from one another.

The majority of the available models for estimating soil loss focus on rill and sheet erosion, thereby excluding the effect of different processes. This is a major limitation; permanent gullies, mass movements and in-stream erosion are often not considered (de Vente et al., 2013). Valid tests on the performance of these models can be carried out using data on sheet and rill erosion rates only. Misapplications occur when the total soil redistribution is equated to sheet and rill erosion (Govers, 2011). Some models exist for specifically considering gully and bank erosion. Up to now it has been not possible to develop a model including all the different components and applicable at catchment, or larger scale, with reasonable results (de Vente and Poesen, 2005; de Vente et al., 2013). A combination of natural complexity, lack of available data and spatial heterogeneity makes it problematic (Jakeman et al., 1999; Wasson, 2002). It is therefore required to integrate the work of soil erosion modellers with that of experts assessing soil erosion in the field (Boardman, 2006).

Sediment deposition is another fundamental part of the soil erosion process. A large quantity of the sediment transported by water is normally deposited prior to reaching the catchment outlet. The direct incorporation of sediment deposition in erosion modelling is not only important for avoiding the overestimation of soil loss rate in a plot, field or catchment but also because the spatial distribution of deposited sediments and the particle size composition play a significant role in determining the

response of the land surface to the erosive process (Sander et al., 2011). Many empirical and conceptual models, such as PERFECT (Littleboy et al., 1992), SWRRB (Arnold et al., 1990), RUSLE (Renard et al., 1997) and MUSLE (Williams, 1975) do not consider the deposition process in soil erosion estimation. Other models as for example WATEM/SEDEM (Van Oost et al., 2000; Van Rompaey et al., 2001), WEPP (Laflen et al., 1991), KINEROS2 (Smith et al., 1995a,b) EUROSEM (Morgan et al., 1998a,b) or SIMWE (Mitas and Mitasova, 1998) do explicitly consider sediment transport and deposition within their equations. Unfortunately, most of these models are evaluated using data from erosion plots, without considering, as stated by Nearing (2006) and Licciardello et al. (2009) that within these plots the sediment deposition is negligible. Even fully calibrating the modules for calculating sediment transport and deposition their application at catchment scale would remain problematic. The lack of input data and the complex interactions among the involved processes remain a limitation (de Vente et al., 2013).

2.2.2.2. Shallow landslides: models and their limitations

Landslide modelling is based on many different approaches and techniques. One of the main challenges in landslide modelling is related to the assessment of spatial probability (Van Westen et al., 2006; Bosco et al., 2013). As for soil erosion prediction, many different approaches can be used for obtaining spatial probability maps. Heuristic, inventory-based, deterministic and statistical approach are the most important methods proposed in the literature (Van Westen et al., 2006, 1997; Guzzetti et al., 1999; Aleotti and Chowdury, 1999) but physically based and statistical approaches are the most common for predicting landslide occurrence.

Statistical methods, which are based on establishing relationship among variables correlated with slope instability (Guzzetti et al., 1999), became popular with the spread of Geographic Information Systems (Van Westen et al., 2006). The basic assumption of this technique is that landslides occur with the same condition as they occurred in the past.

The first requirement of this methodology is to identify and map the landslide conditioning factors (geological and geomorphological). Estimations are then made of their relative contribution in slope failure and areas are classified as having a different hazard or susceptibility degree (Suzen and Doyuran, 2004; Pathak and Nilsen, 2004; Acharya, 2011). The most commonly used methodologies for these predictions are the bivariate and multivariate statistical methods such as logistic regression (LR), artificial neural network (ANN) or multiple regression analysis. This kind of analysis is based upon the presence or not of stability phenomena within the classified areas (Van Westen, 2000). One of the main limitation using this approach is the high sensitivity of the results to the set of input data jointly with the changes occurring to environmental parameters (such as slope, land cover or soil thickness) after a mass movement (Van Westen et al., 2006). Furthermore the triggering factors are hardly ever incorporated into statistical methods and never if we consider also the temporal aspect (Van Westen et al., 2006), it is a challenge to add the temporal dimension to a susceptibility map. It is also not easy to derive the probability of occurrence from the susceptibility (Acharya, 2011).

Deterministic methods offer quantitative results that have a stronger physical basis. In landslide modelling, process-based models are normally developed for studying a specific class of landslides (debris flows, rock falls, etc.) or for investigating a specific triggering factor (Guzzetti, 2005). When applied to predict shallow landslides triggered by water, physically based models spatially extend slope stability models usually applied in geotechnical engineering (Guzzetti, 2005). Properties such as soil cohesion or internal friction, and parameters such as pore water pressure, are normally required by this family of models. But these are difficult to measure at catchment or regional scale because of their spatial variability and stratigraphic heterogeneity. A high degree of simplification is necessary to be applied to effectively implement these models (Dai et al., 2002). Even if we consider slopes as relatively simple systems governed by a few key properties, and we approximate parameter characterisation to help populate models at different granularities, the quality and uncertainty of the input data constitute a major constraint reducing the model's

prediction capacities over wide temporal and spatial scales (Bosco et al., 2013; Dai et al., 2002; Van Westen, 2004).

In deterministic methods, a slope stability analysis is used to determine a factor of safety (FS), so that the landslide susceptibility can be estimated. As with soil erosion many different models exist for estimating landslide susceptibility. The main difficulty overcome of the various physically process based landslide models is the coupling of the dynamic hydrology and (un)saturated soil mechanics to obtain spatio-temporal slope stability indications (Van Beek, 2002; Van Westen, 2004). In all these models, the local equilibrium along slip surfaces is estimated to evaluate the stability of a slope. Calculating a ratio between resisting and driving forces (FS) it is possible to express this equilibrium. If the factor of safety is lower than unity, the slope is unstable.

$$FS = \frac{resisting\ forces}{driving\ forces} = \frac{shear\ strenght}{shear\ stress} = \frac{S}{\tau}\,, \tag{2.2}$$

Since initial failures due to rainfall infiltration often have small depth-to-length ratios, and form failure planes parallel to the slope surface, the use of infinite slope stability analysis for the evaluation of landslides induced by rainfall is justified and often preferred for its simplicity.

Within this equation S represents the shear strength mobilised along the sleep surface. The Mohr-Coulomb criterion is commonly applied for describing the state of soil strength (Cernica, 1995; Bourne and Willemse, 2001). It can be represented in its simplest form as:

$$S = c + (\sigma - u)\, tan\, \emptyset\ , \tag{2.3}$$

where c is the soil cohesive resistance (kPa) (that can be enhanced by the presence of roots and other elements), σ represents the total normal stress (kNm^{-2}) and u is the pore pressure (kNm^{-2}), both acting on the slip surface and affecting the amount of frictional resistance $\emptyset$ that can be mobilised.

The driving force is represented by τ the slope parallel component of gravity acting on the soil mass above the slip surface.

$$\tau = \gamma_s D \cos\beta \sin\beta \,, \tag{2.4}$$

Where γ_s is the soil unit weight (kN/m^3), β is the slope angle (°) and *D* is the vertical soil depth (m).

Considering the weight of the soil as the only load acting at the potential shear plane the total normal stress is given by:

$$\sigma = D\gamma_s \cos^2\beta \,, \tag{2.5}$$

The methods used in traditional infinite slope analysis (Skempton and Deloy, 1957) must be modified to take into account the variation of the pore water pressure profile that results from the infiltration process. For example, based on the extended Mohr–Coulomb failure criterion (Fredlund et al., 1978), the safety factor of an unsaturated uniform soil slope can be expressed as (Cho and Lee, 2002):

$$FS = \frac{c_s + (u_a - u_w)\tan\emptyset^b + (\sigma_n - u_a)\tan\emptyset}{\gamma_s D \sin\beta \cos\beta} \,, \tag{2.6}$$

where c_s is the effective cohesion, u_a is the pore air pressure, u_w is the pore water pressure, u_a - u_w is the matrix suction, σ_n is the total normal stress, σ_n - u_a is the net normal stress on the slip surface and $\emptyset^b$ is an angle indicating the rate of increase in shear strength related to matrix suction.

In the hydrological component of physically based models for predicting shallow landslide susceptibility, various approaches have been proposed for modelling the water infiltration process. These range from very simple topographic index models to complex three dimensions models based on Richards' (1931) equation (Lanni, 2012). SHALSTAB (Montgomery and Dietrich, 1994) and SINMAP (Pack et al., 1998) are an example of models implementing a topographically based steady-state hydrology model (based on the work of O'Loughlin (O'Loughlin, 1986)). A topographic index

based on the ratio between the specific upslope contributing area and the local slope has been used for calculating the water table depth (Lanni, 2012; Jakob and Hungr, 2005; Pack et al., 2005). The stability of each analysed element is evaluated by the slope stability component of the model using this topographic index. These models allow uncalibrated predictions and are generally used for a preliminary assessment over large areas obtaining reasonably successful results though with a tendency to over-predict (Lanni, 2012; Dietrich et al., 2001).

Within its more complex approach TRIGRS (Baum et al., 2008) combines a one dimensional analytical transient flow model for vertical infiltration in homogeneous materials for either saturated or unsaturated soils, with a slope stability model. For setting the initial soil moisture conditions a topographic index approach is used to simulate a subsurface flow parallel to the slope.

The development of three dimensional physically based hydrological models has been performed for obtaining better physically based simulations in well characterized study sites. The Van Beek's model (Van Beek, 2002) is a 3D model coupling a hydrological (STARWARS) and a stability (PROBSTAB) module. The hydrological component consists of a module that predicts percolation resolving dynamic equations for saturated and unsaturated conditions and of sub-models that describe specific hydrological processes, such as interception or snow melt (Krzeminska et al., 2012; Malet et al., 2005). The hydrological model outputs are the daily groundwater height and the volumetric moisture content which is then used by the stability component for modelling shallow landslide susceptibility.

In GEOtop (Rigon et al., 2006), a numerical solution of the three dimensional Richard's equation for modelling subsurface flows and suction dynamics is applied. GEOtop-FS (Simoni et al., 2008) combines a basic equation of the Factor of Safety (FS) with the GEOtop model. GEOtop-FS could predict a real landslide which occurred within the Sauris catchment (Eastern Italian Alps) but unfortunately, for being satisfactorily applied, the model requires the solution of large systems of complex equations.

Many other models exist for shallow landslides prediction, from relatively simple one-dimensional models such as dSLAM (Wu and Sidle, 1995) to complex physically based two- or three-dimensional models such as InHm (Vanderkwaak, 1999), CI-SLAM (Lanni et al., 2012) or SHETRAN (Ewen et al., 2000). Because of the difficulties in calibration and parameterization of these more complex models due to the lack of accurate available data (Hilberts, 2006), it is often necessary at watershed scale to make simplified assumptions related to the hydrologic response (Loague et al., 2006). This makes the simpler topographic-index based models often preferred over more complex process-based hydrological models.

Another real challenge in landslide prediction is represented by adding the temporal probability to the susceptibility maps. The temporal probability is the frequency of occurrence of landslides in a predetermined time interval. Compared with spatial probability, less research has been done for establishing the temporal probability of landslides (Guzzetti et al., 2005; Jaiswal and Van Westen, 2009) but recently there was a research trend towards a frequency-magnitude quantification of landslides (Bovolo and Bathurst, 2012). For the majority of landslide types the probability that a similar event occurs once there has been mass movement, decreases with an increase in the degree of change in local conditions (Van Westen et al., 2006). Generally, statistical models do not consider the temporal aspect linked with slope stability processes as they cannot predict the changes in triggering and controlling conditions (e.g. changes in land cover or in water table depth; Van Westen et al., 2006). However, using statistical methods combined with landslide records and the return period of triggering events it is possible to analyse the temporal probability of landslides (Zezere et al., 2004; Dai and Lee, 2003; Guzzetti et al., 2005). If the precipitation event causing landslides and the related rainfall parameters are known, it is possible to assess the spatio-temporal probability of landslide at a given location. Unfortunately the lack of landslide records is the main obstacle in following this approach (Petley, 2012; Van Westen et al., 2006).

If, for a specific location, with the exception of rockfalls and debris flows, there isn't normally a magnitude-frequency relation linked with landslides (Van Westen et al., 2006), some physically based and empirical models are available for elaborating this

relationship over large areas. Combining statistical and deterministic methods, it is possible to perform spatio-temporal analysis from local to wider scale.

It is clear that further efforts are required to determine the spatio-temporal probability of landslides occurrence in order to minimize the hazards, especially in data-poor conditions and under possible future climate change scenarios. The research conducted in this thesis that proposes a statistical approach incorporating a frequency-area landslide distribution model (Malamud et al., 2004) within the framework of landslide susceptibility mapping, is a step forward in this direction.

2.3. Remote sensing and pedometrics for improving the quality of modelling input data

Remote sensing data from satellite imagery and pedometric methodologies offer considerable scope for improving the input quality of soil, landscape hydrology and vegetation cover components that characterize most soil erosion and shallow landslide models. To date, only a few models specifically use remotely sensed data as part of the model process (Jetten et al., 2006).

“Pedometric mapping is generally characterized as a quantitative geo-statistical production of soil geoinformation also referred as digital soil mapping” (McBratney et al., 2003 in Hengl, 2003). Pedometrics is “the application of mathematical and statistical methods for the quantitative modelling of soils, with the purpose of analyzing its distribution, properties and behaviours” (McBratney et al., 2003; Lucà et al., 2018). The use of pedometric techniques and remote sensing could include new technologies such as close-range remote sensing, GPS positioning and advanced computational analysis (McBratney et al., 2003; Lucà et al., 2018). But also can include the use of well-known techniques as the Neural Networks or the application of Fuzzy systems (e.g. the Fuzzy Inference System (FIS) or the Adaptative Neuro-Fuzzy Inference Systems (ANFIS) (Hosseini et al., 2017) is analogous to neural networks and can be used to predict continuous variables.) (McBratney et al., 2003).

With the rapid development of methodologies for deriving auxiliary maps, the remote sensing and the terrain parameters (parameters derived from a Digital Elevation Model (DEM)) play a key role in digital soil mapping (Brevik et al., 2016) quantifying morphology of the terrain and surface characteristics.

Although remote sensing has revolutionized vegetation mapping (for example with the use of the Normalized Difference Vegetation Index (NDVI) to determine the density of green on a patch of land or of the Leaf Area Index (LAI) to characterize plant canopies), our knowledge of how to apply advances in remote sensing to soil properties and terrain mapping is still incomplete (Mulder et al., 2011). It is not possible to directly measure soil characteristics using visible or infrared images in areas covered by vegetation. However a good correlation has been found between compound indices, such as NDVI, and soil carbon and nitrogen content (respectively $r = 0.55$ and $r = 0.52$; $P < 0.01$) (Sumfleth and Duttmann, 2008) or root zone soil moisture during the growing season (r: 0.46 – 0.55) (Wang et al., 2007). A logical further step for improving the prediction capacities in soil mapping was to combine DEM derived data and remote sensing (Hengl, 2003; Mohamed, 2017).

One of the most widespread techniques for estimating the values of soil properties in unvisited locations is spatial prediction or spatial interpolation. Kriging and its derivatives have been recognized as one of the main spatial interpolation techniques from 1970s (Hengl, 2003; Li and Heap, 2011). Simple linear regression models linking terrain attributes and soil parameters formed the basis of the first applications (Gessler et al., 1995 and Moore et al., 1993). The second step was the 'environmental correlation' of McKenzie and Ryan (1999) or the spatial prediction of Odeh et al. (1994, 1995, in Hengl, 2003) that extended the predictors to a set of environmental variables and remote sensing images.

A spatial interpolation methodology that employs correlation with auxiliary maps and spatial correlation is the universal kriging method (UK) (Matheron, 1969). Numerous authors agree in reserving the name 'universal kriging' for the case where the drift (or trend) in the kriging process is modelled as a function of the coordinates only" (Wackernagel, 1998 in Hengl, 2003; Moral, 2010) and Kriging with external drift (KED) when the drift is defined externally using auxiliary variables (Moral, 2010). If drift and

residuals are fitted separately and then summed we have the 'regression kriging' (RK) (Odeh et al, 1994, 1995; Moral, 2010). Universal kriging, kriging with external drift and regression kriging can be considered equivalent methods (Hengl, 2003) giving the same predictions given the same assumptions. The advantage of RK is that it is not subject to the instability that can be present in the KED system (Goovaerts, 1997, in Hengl, 2003).

Another very interesting technique is the K-nearest-neighbour. It is a machine learning technique developed to recognize patterns of data without an exact match to any stored information. The K-Nearest Neighbours (K-NN) algorithm (Keller et al., 1985; Cunningham and Delany, 2007) is a nonparametric method. The proximity of neighbouring input observations in the training data set and their corresponding output values are used to test their predictions against a validation data set.

Between the many different techniques that is possible to apply for producing modelling input data, remote sensing (for vegetation parameters) and kriging or K-nearest neighbours (and its derivatives) (for soil characteristics) appear to be the more promising to support modelling prediction capacity in data-poor condition.

2.4. Modelling soil erosion and landslides interactions

2.4.1. The integrated modelling approaches. An overview

As mentioned, many models exist for predicting shallow landslides and soil erosion by water independently, but relatively few attempt to develop an integrated approach to combine soil erosion and shallow landslide modelling (Burton and Bathurst, 1998; Bathurst et al., 2010).

The sediment transport modelling system (SHETRAN) (Ewen et al., 2000) is such a model. SHETRAN is capable of predicting shallow landslides, soil erosion and sediment yield at a large spatial scale (Burton and Bathurst, 1998). Three main components lie at the core of SHETRAN, one each for water flow, sediment transport, and solute transport (equations are available in Acharya, 2011). It provides an

integrated surface and subsurface representation of water movement in a catchment, incorporating the major factors involved in hydrological cycles (interceptions, evapo -transpiration, snowmelt, overland and channel flow in unsaturated and saturated zones) (Acharya, 2011). In SHETRAN sediment yield as a function of soil erosion and shallow landslides with their driven sediment yields are modelled but, within the model, there is still the necessity to better understand hillslope processes involving shallow landslide generation, chnages in topography (that largely affect erosion) and to better predict changes in soil erosion after landslides occurrence (Cochrane and Acharya, 2011).

SIBERIA is a physically based model for measuring soil erosion rates and for studying the erosional development of basins and their network of channels. It considers many different mass transport processes, such as fluvial sediment transport (applying the Einstein-Brown equation) and a conceptualization of mass movement mechanisms such as soil creep and landslides combining these effects at a diffusive term of the model's equation (Willgoose and Riley, 1998). In SIBERIA the interactions between gullies or channels and the hillslopes are explicitly incorporated into the modelling architecture and the processes acting in the channels and on the hillslopes are specifically differentiated (Willgoose and Riley, 1998). The model developed by Willgoose et al (1989, 1990, 1991a,b,c,d), which can simulate the evolution of landscapes over time, is calibrated to existing hydrogeomorphic data.

SOMORE (Pla Sentis, 1997), modelling soil hydrological processes, provides an index for understanding the conditions where both the potential soil and landslide erosion may be more critical. This model simulates the evolution of soil water balance under different climatic and topographic conditions. SOMORE, compares the soil moisture above the liquid limit for identifying the most favourable conditions for mass movements (Pla Sentis, 1997). It has the advantage, when compared with other physically based models, of having a low data requirement.

Two other models, PSIAC (1968) and the model of Gavrilovic (1976) specifically consider the contribution of landslides to sediment yield. In PSIAC several factors are used to assess the sensitivity of a catchment towards erosion and sediment transport.

A score is attributed to each of these factors and the results are used for calculating an index related with the catchment sediment yield (de Vente et al., 2006; Globevnik et al., 2003). PSIAC uses nine factors related to sediment yield attributing different scores to each one (PSIAC, 1968), these are: geology, soils, climate, runoff, topography, ground cover, land use, upland and channel erosion and sediment transport. A numerical value represents the relative significance of that factor in the yield rating (PSIAC, 1968) and the yield rating is the sum of that values. These values are divided into five classes representing an average annual yield in acre-feet per square mile. The highest scores are attributed to strong signs of erosion such as gullies, rills or landslides (de Vente et al., 2005). The different scores contain a weighting system, for example a good vegetation cover, due to its negative score, results in a decrease in soil loss (de Vente and Poesen, 2005). Being PSIAC developed for the arid and semi-arid areas of southwestern USA, applying the model in other regions, a new relation between the sediment yield and the PSIAC index should be found (de Vente and Poesen, 2005). PSIAC is recommended to be used in catchments no smaller than 25 square kilometers and in broad planning purposes (PSIAC, 1968; de Vente et al., 2005). In Johnson and Gebhart (1982) a modified version of PSIAC introducing empirical relations for assessing the different scores in order to reduce the subjectivity present within the model (de Vente and Poesen, 2005).

The model of Gavrilovic offers a semi-quantitative method for modelling sediment yield. The sediment delivery ratio is estimated by multiplying the calculated sediment retention coefficient and the average annual gross erosion. The model of Gavrilovic uses an approach based on scores only for three of the modelling variables (soil cover, soil resistance, type and extent of erosion) (de Vente and Poesen, 2005), whereas the other variables are quantitative descriptors of the catchment conditions; this makes it one of the most quantitative of the analysed models. Coefficients for soil protection, type and extent of erosion processes and soil resistance are used for calculating the erosion into the model. The landslide erosion is accounted within the coefficient of type and extent of erosion. Unfortunately within the model of Gavrilovic and in PSIAC only observation of landslide occurrence is considered for the sediment yield calculation (de Vente and Poesen, 2005).

SSYIndex is a model for estimating sediment yield at a large spatial resolution, calibrated for 29 large rivers in Europe (Delmas et al., 2009). Four indicators represent different processes considered as sources, sinks or transfers of sediments (de Vente et al., 2013). Mass movement and hillslope erosion are considered as the main sources of sediment. The hillslope erosion indicator is assessed from a pan-European compilation of measured rill and interrill erosion rates in plot studies, and subsequent interpolation as a function of topographical, land use, and soil parameters (Cerdan et al., 2010). The mass movement indicator is defined through an expert assessment as the percentage of catchment area with potential for occurrence of mass movements, based on slope and lithology maps (de Vente et al., 2013). A simple sediment yield index is obtained summing the sediment sources and transfer potential and subtracting the sinks (Delmas et al.,2009).

In TOPOG (O'Loughlin, 1986; CSIRO, 2017) both soil erosion and landslides are integrated into the model (CSIRO, 2017). TOPOG is a hydrological model describing how the water moves through and over the soil and back to the atmosphere. TOPOG predicts the degree of soil saturation in response to a steady state rainfall for topographic elements defined by the intersection of contours and flow tube boundaries, this relative soil saturation is used by the slope stability component to analyze the slope stability of each topographic element (Montgomery and Dietrich, 1994).

TOPOG provides an index of the shallow landslide potential susceptibility (Montgomery and Dietrich, 1994). This model can also be applied for identifying the presence of areas affected by soil erosion problems. It is possible to extend the model's analysis capacity, for calculating the potential to erosion, by supplying additional information describing the relationship between soil cover and particle entrainment (Vertessey et al., 1990). The main limit of TOPOG in analysing the integrated process of soil erosion and landslides is in the lack of a clear link between a mass movement and its effect on soil erosion rate on a slope profile.

It is only in WEPP-SLIP (Cochrane and Acharya, 2011) that post-failure soil erosion is considered. Within WEPP-SLIP (Water Erosion Prediction Project Shallow Landslide Integrated Prediction) the physically based WEPP model (Laflen et al., 1991) is applied

for estimating pre-failure soil erosion rates by water. The infinite slope stability model of Skempton and DeLory (Skempton and DeLory, 1957) is then applied for evaluating the landslide susceptibility of the study area and a simple rule-based soil redistribution model is used for estimating runout distance and the changes in topography after the mass failure (Acharya, 2011). Thereafter the WEPP model is re-applied for measuring post-failure soil erosion rates by water considering the changes in topography and land cover (Cochrane and Acharya, 2011). Landslides can also trigger important processes such as the soil armouring. In WEPP-SLIP the soil armouring has been considered for accounting the changes in cover and soil properties.

For determining the height of the water table, necessary for calculating the slope instability, the WEPP model is applied to each flowpath present in the study area (Acharya, 2011). One of the limits of this approach is that the total soil water content is manually attributed to each grid cell from each simulated flowpath. It makes this technique extremely time consuming (Acharya, 2011). WEPP-SLIP does not also simulate important sediment sources as the channel erosion but the main limitation is probably linked with the modelling of slope stability and runout distance. For assessing the post-failure soil erosion it is necessary to exactly know where and when a landslide will occur and to know the exact dimension of the affected area. Further steps are necessary in this direction.

Two other models (E_s and GLASOD) belonging to a family of models focusing on assessing the land degradation rather than estimating erosion or sediment yield and integrate soil and landslide erosion in their architecture. The erosional susceptibility (E_s) model of de Ploey et al. (1995) includes a headcut retreat model for rills and gullies (de Ploey, 1989) and is used for calculating the erosion susceptibility of a catchment (e.g rill, sheet, gully, landslide or wind erosion). The erosion susceptibility is determined by comparing the volume of removed sediment and the energy input by water and air. The model requires volumetric soil loss estimates and basic pluviometric and aerodynamic data (de Vente and Poesen, 2005). At the basis of the E_s model there is the idea that the majority of the erosion processes have a common operational mode: e.g. ”the retreat of an erosion border, usually a topographic ”cliff”,

in the top layer of the soil" (de Ploey et al., 1995). The efficiency of the different erosional processes incorporated in the model, was shown by applying E_s in different studies around the world (de Vente and Poesen, 2005).

GLASOD is a qualitative model directly based on observation of the involved processes using expert judgment. After delineating physiographic units showing homogeneity of soils, climate, topography, land use and vegetation cover, an evaluation on the degree and extent of human-induced degradation risk is carried out and a score which distinguishes between four different levels of soil degradation is attributed (de Vente and Poesen, 2005). The GLASOD methodology recognizees 4 main types of soil degradation (water erosion, wind erosion, chemical deterioration and physical deterioration), soil erosion by water and landslide erosion are jointly considered and evaluated in the water erosion class (Oldeman et al., 1990). After the preparation of the GLASOD map the majority of the conclusions stated that a more detailed information at the national level is required and more objective ways should be found to prepare a base map and analysis of soil degradation risk (de Vente and Poesen, 2005).

Although the changes in vegetation cover and topography are readily modelled by the majority of the erosion models (Dymond et al., 2006), the changes affecting soil properties are hardly ever considered. Soil input properties remain normally unchanged during the whole simulation period. Topographic changes and evolution of the hillslope profiles are also not well documented (e.g. Cendrero and Dramis, 1996; Hovius et al., 1997) with a consequent lack of data for describing the changes in sediment yield after landslide occurrence.

Not many models exist for integrating soil erosion processes and landslides prediction. The existing qualitative or quantitative models, presented here, are based on completely different approaches showing many different limitations. Most of the models do not consider post-failure scenarios and are only partially sensitive to climate or land use changes hampering their applicability in climate change analysis (de Vente et al., 2013). Within the next section these limits will be further investigated.

2.4.2. Limitations of the present approaches

The main limitation of the majority of the integrated erosion-landslides models is their lack of considering post-failure soil erosion changes. Only in WEPP-SLIP post failure soil erosion is explicitly considered (post failure long-term sediment yield). The soil redistribution within the model needs to be further developed for the simulation of soil water content at catchment scale because of the extremely time-consuming process that consists in manually attribute the maximum value of the total soil water content from each simulated flowpath to each grid-cell (Acharya, 2011).

Physically based approaches could be modified for considering post-failure soil erosion changes. These models use local terrain characteristics and a dynamic hydrological model with rainfall as the main variable (Jaiswal and Van Westen, 2009). This implies, as already mentioned, that landslides temporal probability also can be easily incorporated into the integrated model. Unfortunately these models are less suitable to be applied in data poor regions and at a catchment or larger scale as they usually require a detailed knowledge of local terrain characteristics (e.g. soil properties, high resolution climatological data, shear parameters) and are often optimised for the local conditions of small catchments (de Vente et al., 2013).

Furthermore, due to the limits of reductionist models to adequately incorporate many different soil erosion processes as gully and bank erosion or mass movements, alternative approaches for catchment or wider scale estimates are required (de Vente and Poesen, 2005).

Applying statistical methods it is possible to overcome the lack of detailed input data over large areas. Unfortunately their lack of accounting for the temporal aspect is the main limit of such an approach. For building an integrated system between soil erosion and shallow landslides in addition to the landslide susceptibility it is necessary to estimate the landslides temporal probability or the frequency-area distribution. For improving soil erosion estimation considering landslides within the erosion process it is necessary to estimate not only where, but also when a landslide will occur along with the size of the event.

This research follows this approach of considering the frequency-area distribution of landslides in order to quantify the size and number of landslides associated with precipitation events in data-poor regions.

In chapter 6, a new method is proposed for empirically estimating the importance and extent of landslides on soil erosion losses. This has been achieved by sampling the frequency-size landslide distribution proposed by Malamud et al. (2004), and stochastically distributing (Monte Carlo method) the landslide location across the catchment.

The proposed methodology is based on the geospatial semantic array programming paradigm (see section 2.5 and 6.2.1) and has been implemented on a catchment scale methodology using Geographic Information Systems (GIS) spatial analysis tools and GNU Octave .

The Monte Carlo method was applied for repeating random frequency-size landslide distributions in order to calculate the mean change in soil erosion linked with landslide activities. Using this technique it is not possible to predict the exact time, position and extension of a landslide within the study area, but it is possible to predict the mean change in soil erosion due to landslide activity.

2.5. Introducing the Semantic Array Programming paradigm

Within specialised computational models, the programming environment might be stable with centralised, strictly codified, and often highly customised internal data structures (monolithic models or frameworks, de Rigo, 2015). As a consequence, the various parts of a single monolithic model may exchange information with direct access to the implementation details. For example, object oriented approaches are suited for representing and transforming information within a specific model in sophisticated and flexible modalities. As highlighted by de Rigo (2015), the objects of a specialised “monolithic model are typically straightforward to propagate and very effective in transferring structured information with default behaviours/assumptions”. However, this direct “internal” communication may become more and more complicated when the information is needed from

heterogeneous modelling sources, for example from multiple monolithic models (typically not designed to interact between each other) implemented with different programming languages and tools by many research teams along several years of independent development, “and possibly no single expert able to cope with the overall integration complexity” (de Rigo, 2015). However, a *direct* modelling integration is not always necessary for heterogeneous models to be able to communicate effectively.

As outlined in the introduction, an important subset of computational modelling applications may be formulated by considering conceptual modelling-units as data-transformation models or modules (D-TM) (de Rigo, 2015; 2013; de Rigo et al., 2013b). In this work, this abstraction concept will be exploited with a focus on soil science and its potential integration within some cross-disciplinary aspects of INRMM. Since a D-TM unit *f* exclusively exchanges data (extended to include model parameters), transforming input data into derivative output data, then the internal details of *f* may be separated from the details of other D-TM units which need the output data of *f* as their input. In particular, different D-TM units might be implemented in different programming languages. Furthermore, an asynchrounous exchange of intermediate data is easy even if not all the involved D-TMs share the same computing environment (for example, in case some D-TM physically run in different computational facilities), simplifying collaboration between research teams. This requires data to be expressed in portable formats, and the semantic compatibility among initial or intermediate data from different D-TM units (which may have been originally designed for domain-specific purposes, without any plan to integrate them) to be verified. However, these requirements are far easier to satisfy compared with the potential costs of porting entire models from their original implementation to any given monolithic framework (Mäntylä and Lassenius, 2006; Lehman and Ramil, 2003; Lauder and Kent, 2000; Hatton and Roberts, 1994)

Using portable data formats, and checking for the semantic consistency of multiple data sources to manipulate together are non-intrusive requirements which do not require a monolithic approach. For example, the first requirement may be fulfilled with a post-processing of data formats to translate less portable to more portable

ones; and the second one with a pre-processing of input data sources to check whether their semantic compatibility is satisfied, as a pre-condition to then pass the inputs to the D-TM which will process them. These methodologies belong to the core recommendations of the Semantic Array Programming paradigm (de Rigo, 2012a; 2012b; 2015) and will be applied throughout the computational modelling applications described in the next chapters.

In this work, some of the specialised models which will be discussed are characterised by an array-based structure.

They may be composed by an array of sub-models, or may process an array of datasets. Some layer may be in turn characterised by sub-arrays of data layers and corresponding data-transformations (Bosco et al., 2015; Bosco and Sander, 2015; Bosco et al., 2013). For example, the soil erosion model which will be discussed in Chapter 4 (e-RUSLE model) is based on a multiplicative structure of several specific factors (sub models), each dedicated to describing corresponding aspects of the erosive process. Each factor is represented as a geospatial grid of values, a value per each geographical unit cell *c* in the spatial extent of interest. Therefore, to each factor a spatial matrix of values is associated, and the final model is estimated by aggregating an array of factor-specific matrices.

In data-poor regions, some of these factors may lack part of the necessary data for their estimation, so that a more approximated estimation strategy may be needed. For example, the intensity of precipitation is a key element affecting soil erosion by water (Wischmeier, 1959). A specific, unfortunately data-demanding factor is dedicated to this in the proposed soil erosion model: the erosivity factor (Wischmeier, 1959; Wischmeier and Smith, 1978). In paragraph 4.2.3.2, the topic will be discussed in detail. Here, this factor is mentioned to serve as an example for introducing a modelling procedure which will be applied in different parts of this work, to enable less fragile quantitative estimations to be computed in data-poor areas. The scarcity of accurate datasets for directly assessing the quantity Y_c to be modelled (in the example, soil erosion rates and their erosivity component, for each spatial cell *c*) may motivate the replacement of the original model (or sub-model) with a surrogate model $Y_c^{est} = f(\theta, X_c)$ based on custom D-TMs. These data

transformations, from an array of proxy data X_c (hereinafter referred to as predictors or covariates), and an array of custom parameters θ, approximately estimate the otherwise too data-demanding quantity Y_c.

Typically, it cannot be expected for the surrogate model to perform equivalently to the original model. Therefore, despite the efforts to select a single 'optimal' surrogate model among the many potential candidates, that model might still present important weaknesses. This may be set in the wider context of computational modelling (de Rigo, 2015) by considering the difficulty to isolate a regression or classification algorithm universally superior to the other ones (Wolpert, 1996; Wolpert and Macready, 1997; Koppen et al., 2001; Xu et al., 2012).

However, following a natively array-based modelling strategy, an array of different surrogate models $\{ Y_c^{est,1}, Y_c^{est,2}, \dots, Y_c^{est,n} \}$ may be estimated instead of a single one, and an aggregated ensemble may be derived to mitigate part of the weaknesses of individual approximations.

In the example, the surrogate model will take the form of a climatic-based empirical model to estimate the erosivity of rainfall. These kinds of empirical models are relatively frequent in the literature (for the exemplified erosivity quantity, Bollinne et al., 1979; Rogler and Schwertmann, 1981; Ferro et al., 1999; de Santos Loureiro and de Azevedo Coutinho, 2001), and are derived from more detailed data by means of a regression approach to correlate, over a given study area, the quantity to approximate with easily available predictor information (in the erosivity example, information on monthly precipitation patterns). This approach allows the empirical relationship to be extrapolated even outside its original study area, since the predictor information is often available over much wider areas than the study area. Therefore, from an array of estimates extrapolated from different study areas, an aggregated ensemble may be computed to approximate the missing original quantities over a given area of interest (see paragraph 4.2.3.2). A variant of this approach may be applied even when the surrogate models are not taken from existing literature, but instead are directly tuned to best fit a set of available measures. If a multiplicity of model families is exploited to tune each corresponding model $\{ Y_c^{est,1}, Y_c^{est,2}, \dots, Y_c^{est,n} \}$, then each model will be associated with the

advantages and weaknesses expressed by its model family, and a carefully designed final aggregated ensemble will again be able to partly mitigate some of the individual weaknesses. In chapter 5 examples of this array-based modelling strategy will be discussed.

Following this overview of applications, the Semantic Array Programming paradigm may be summarised in a more formal way. Array Programming (AP) originated for "reducing the gap between mathematical formulation and code implementation" (Iverson, 1980) with the introduction of very concise operators and coding patterns to manipulate variables composed by large number of elements (for example, wide matrices of geospatial raster data). AP considers these variables as atomic by providing abstract operators which do not oblige the computational modeller to explicitly track the detailed shape of each array. Examples of popular AP languages are GNU Octave (https://gnu.org/software/octave/) and MATLAB (http:// mathworks.com/help/matlab/), GNU R (https://gnu.org/software/r/) and Python (http://python.org) with NumPy and SciPy (Eaton et al., 2008; Venables et al., 2009; van Rossum and Drake, 2011; The Scipy community, 2012a; 2012b).

The Semantic Array Programming paradigm (SemAP) complements the generality of abstraction supported by AP approaches with two additional ideas. First, SemAP introduces the systematic use of a rich set of array-based semantic constraints as provided by the Mastrave modelling library (http://mastrave.org), which implements the paradigm (de Rigo, 2012d). Second, SemAP encourages an explicit effort towards a disciplined modularisation of each conceptual modelling unit. SemAP modules are typically associated to corresponding D-TM units, with a precise semantic annotation of the array-based mathematical constraints required for different D-TM input data to be compatible between each other.

In particular, a D-TM semantically-enhanced following the SemAP paradigm explicitly expresses a set of semantic constraints for each of its input data, and optionally for the returned output derivative data. These constraints may be easily annotated even in a not fully formal context, for example within natural language descriptions (such as within scientific articles or reports). They take the form of a sequence of categories between "**::**". The special token :: is used as delimiter "quotation" of the semantic

constraints. For example, if a certain semantic constraint **sem** applies to one of the D-TM input data, that input will be denoted as ::**sem**:: . This “informal” annotation has a precise formal equivalent in the Mastrave implementation of the SemAP paradigm, accessible online for each constraint ::**sem**:: at its corresponding URL http://mastrave.org/doc/mtv_m/check_is#SAP_**sem** . For example, an input data expected to be a matrix of nonnegative values, and limited by the D-TM capabilities in a way that causes missing data to generate an error, can be semantically annotated as a ::**matrix**:: of ::**nonnegative**:: elements, each of them ::**nanless**:: (i.e. without Not-a-Number values, NaN). The formal definition of each constraint is respectively available at the URLs:

::**matrix**:: - http://mastrave.org/doc/mtv_m/check_is#SAP_matrix ;

::**nonnegative**:: - http://mastrave.org/doc/mtv_m/check_is#SAP_nonnegative ;

::**nanless**:: - http://mastrave.org/doc/mtv_m/check_is#SAP_nanless .

In computational science applications, the semantics may be characterised by referring to multiple dimensions. The mentioned set of array-based semantic constraints defines a specific, portable dimension grounded on the mathematics of arrays. These constraints are designed to be very compact, and in several cases easily understandable even without accessing their formal definition and computational implementation. Among the many other dimensions of semantics, a particular role may be highlighted for the geospatial semantics, which is essential in spatially explicit modelling. SemAP and geospatial semantics coexist with the geospatial application of the semantic array programming paradigm (Geospatial Semantic Array Programming, GeoSemAP) formalised by de Rigo et al. (2013b) and de Rigo (2015) and applied, for example, in Bosco and Sander (2015). In chapter 6, GeoSemAP will be exploited.

3. Study area

3.1. Catchment choice

The study area is located in southern Italy, within the Puglia region (Figure 3.1). It covers around 10 km^2 in the municipal territory of Rocchetta Sant'Antonio which is situated in the southern part of the Daunia Appennines. The area is characterized by moderate relief topography, with elevations generally below 800m and modest slope inclination (around 10° on average). Vgetation cover is dominated by agricultural activity (mainly cereals) with grassland-pastures and trees only locally significant. The climate is Mediterranean (sub-humid) with large variation in annual rainfall values. Autumn and winter present a similar precipitation range that account for 60% of the annual total and summers are dry and hot (Wasowski et al, 2014).

The Daunia Appennines are characterised to be highly susceptible to landsliding (Iovine et al., 1996; Magliulo et al., 2008; Zezza et al., 1994) and the study site has a frequency for landslides that exceed 20% for the overall Daunia Appennines (Wasowski et al., 2010). This site has been studied for some years (Wasowski et al., 2007; Mossa et al., 2005) and due to the high number of shallow landslides affecting the local economy (Wasowski et al., 2010), a set of different data has been collected and produced: a map of the lithological units (section 3.2), a landslide inventory related to the year 2006 (section 3.3), a digital elevation model (DEM) with a resolution of 5x5 metres, rainfall data from the pluviometric station located in Rocchetta sant'Antonio (section 3.4), a map of the land cover (section 3.5) and some results from subsurface borehole investigation and piezometer monitoring.

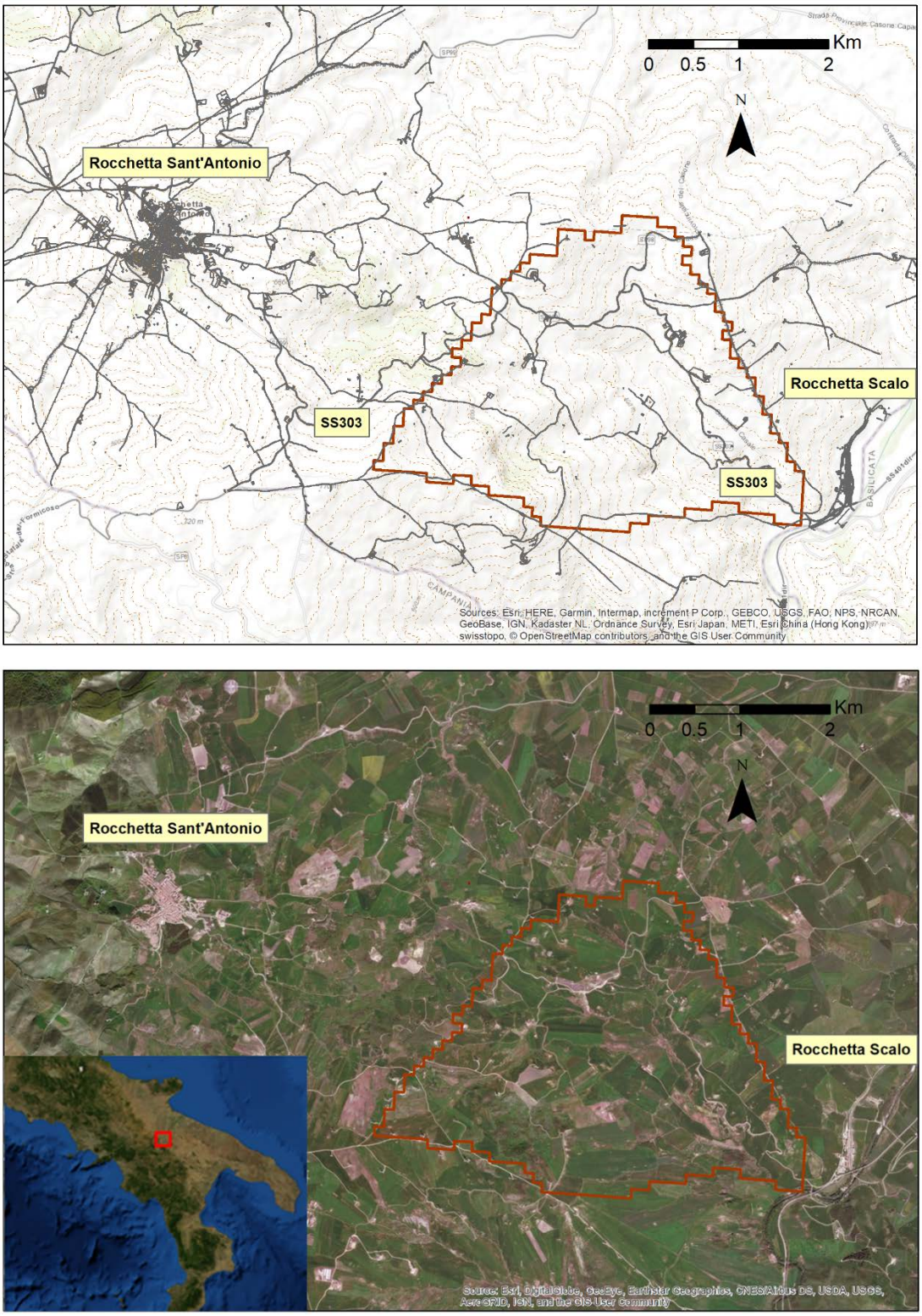


Figure 3.1 – The study area (Rocchetta Sant'Antonio, Italy). Google Earth, © 2013 Google. (source of the Background map: Esri, et al., 2018).

This site was chosen because it has characteristics that make it suitable for our research purposes. These include:

- With its rural characteristics and frequent recurrence of predominantly shallow landslides (up to 10 m depth (Wasowski et al., 2010)), the study site can be considered a representative portion of the Daunia Mountains (Wasowski etal., 2010).
- The site has been studied by local researchers for some years (Wasowski et al., 2007, 2010, 2012), making available for the scientific community a set of data at high spatial resolution.
- The catchment is affected by both mass movements and soil erosion processes (with relevant costs for the local community[1] (Comune di Rocchetta sant'Antonio, 2015)). The contemporary susceptibility of the area to soil erosion and shallow landslides is a precondition for evaluating the effects and interactions between these two processes.
- The spatial size of the catchment is compatible with the field collection of the data neded for both soil erosion and landslide susceptibility models.T

3.2. Geology and soil

The Daunia region is located in the transition area that includes the most advanced part of the frontal thrust of the Southern Apennines and the most westerly area of the foredeep (Dazzaro et al., 1988; Ciarcia et al., 2003; Wasowski et al., 2010). A tectonally deformed lithostratigraphic succession belonging to the period between the Late Cretaceous and the Miocene age, characterises the chain units in this area. The clay-rich flysch formations of the Daunia region and the presence of intensely deformed geological units are factors predisposing to the slope instability (Wasowski et al., 2010).

The outcropping formations within the selected catchment are divisible into three main categories, each dominated by one specific lithology: sandstone, limestone and clay-shales (Figure 3.3). In the study area, the clay-rich lithology, belonging to the Late

[1] Within the municipality of Rocchetta Sant'Antonio the costs for mitigating the hydrogeological risk affecting the S.P. 99 bis (ex S.S. 303) were in the order of one million euro.

Cretaceous-Paleogene sedimentary succession, covers more than 76% of the territory. The 'Complesso Indifferenziato' composed of clay-shales is present in more than 50% of the territory (Wasowski et al., 2010), with Miocene age Flysch and alluvial deposits covering the other portion.

The 'Complesso Indifferenziato' is marked by a lithological variability with clay-shales predominant on limestones, calcarenites, breccias and sandstones and shows an intense deformation at the outcrop (Wasowski et al., 2010). These deposits are known as the 'Argille Varicolori' that in turn are shared in clay-rich and limestone-marlstone-rich members ('Flysch Rosso'; Dazzaro et al., 1988; Wasowski et al., 2010). Sandstones and limestones cover respectively 18% and 5.5% of the study area (Wasowski et al., 2010, 2012).

Shear strength data on the lithological units of this area is available from a series of commercial laboratory geotechnical tests (Wasowski et al, 2012) a subset of which was made available by CNR-IRPI (Bari) (see chapter 4, figure 4.4). These data show a low strength estimate of the clay-shales, resulting in a high landslide susceptibility that is consistent with the observed high frequency of landslides affecting the catchment. The predominance of sheared, scaly clays with weak geotechnical properties (ϕ'_{Γ} varying from 6.9° to 16.5°) and the presence of poorly drained slopes are considered the underlaying causes of landsliding (Wasowski et al., 2010, 2012).

Agricultural soils cover the majority of the catchment (Figure 3.4 and 3.5).

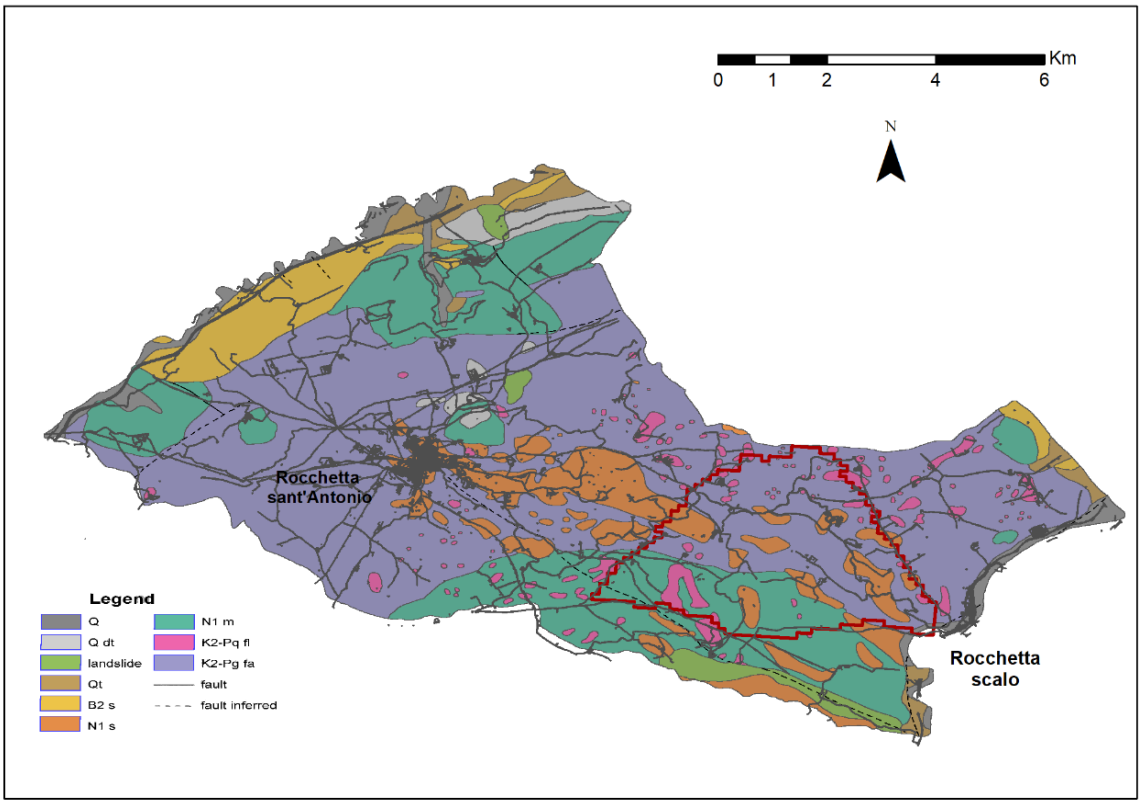


Figure 3.2 Geological map of the municipal territory of Rocchetta S. Antonio[2]. (Source: Wasowski et al., 2010).

[2] Q) recent alluvium; Q dt), slope debris, including landslide deposit (Holocene); Qt), alluvial terrace deposit (sand, gravel and clay) (Pleistocene–Holocene); N2 s), sand, clay, marly clay and conglomerate (Pliocene); N1 s), sandstone including marly clay–shale intercalations (Miocene); N1 m), marly limestone, sandstone and clay–shale (Miocene); K2-Pg fl), Undifferentiated Complex (limestone breccia, calcarenite) (Palaeogene); K2-Pg fa), Undifferentiated Complex (primarily clay–shale and marly clay–shale, and secondarily limestone, calcarenite, breccia, sandstone) (Late Cretaceous–Palaeogene)

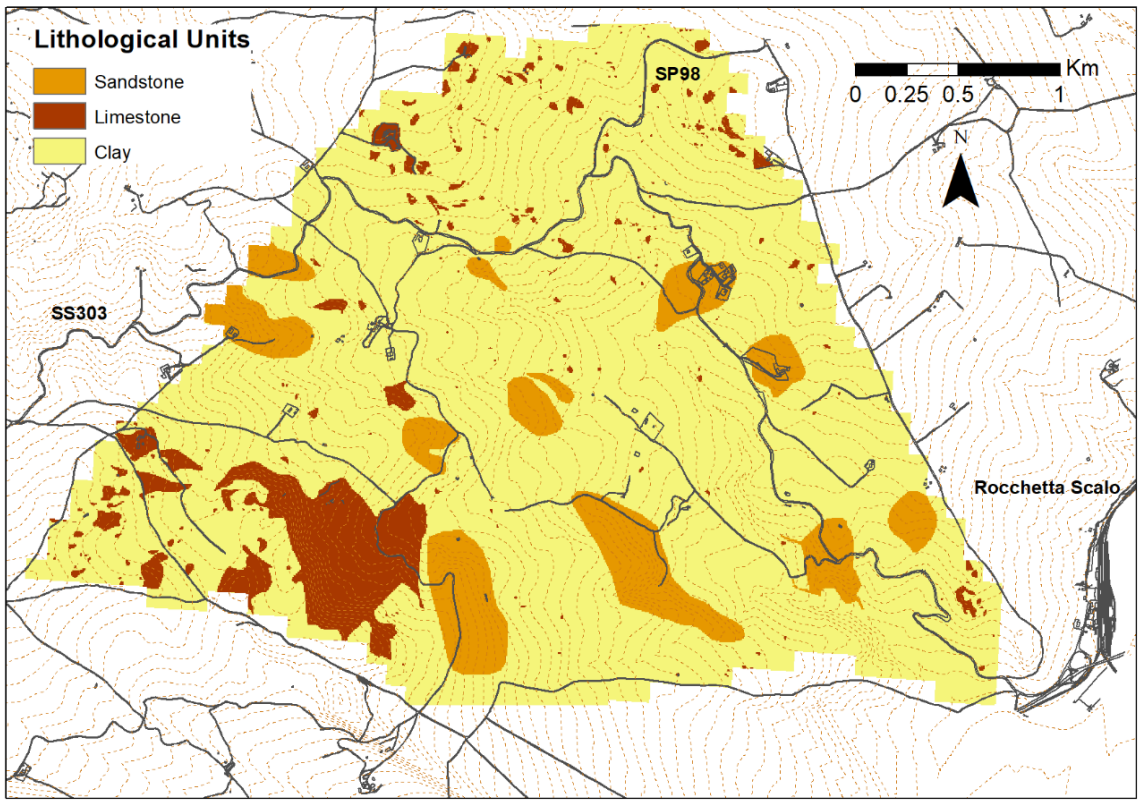


Figure 3.3 – Map of the main lithological units in the study area. This map was obtained by integrating information from the state geological map (Malatesta et al., 1967) and in situ checks (Wasowski et al., 2012).

3.3. Geomorphology

The Daunia Appennines are affected by several types of mass movements. Complex landslides, involving both rotational and translational sliding (Figure 3.4), often evolving into debris or earth flows are the most common type of slope instability affecting this region (Andriani et al, 2009). Mud flows are also a common phenomenon, especially in the areas with prevailing clay deposits. In most of the cases, the slope instability affecting this area is linked with re-activation of 'dormant phenomena' triggered by rainfall (Andriani et al., 2009; Cotecchia et al., 2009).

In the study area, active landslides are typically shallow translational landslides of moderate dimension (from tens to hundreds of meters with basal slip plane less than 10 m below ground surface) and mainly occurring on cultivated slopes. Because of

seasonal ploughing the landslide surface signatures are rapidly lost. Two maps of landslides are available for this area (Wasowski et al., 2007, 2010) and both were created by the department of Bari of the Research Institute for Hydrogeological Protection of the National Research Council (CNR-IRPI of Bari). These maps were derived from aerial photography and satellite image interpretation related to the years 1976 and 2006 and subsequent field investigations. By creating these two datasets, for simplicity, a distinction was made by the authors only between active and inactive landslides. They followed the geomorphic criteria typically adopted for the recognition of landslides and their state of activity from air-borne imagery (e.g. Wieczorek, 1984). Firstly, a stereoscopic interpretation of 1976 aerial photos was used to obtain a historical inventory of landslides related to that year, Secondly, a landslide inventory based on the high resolution IKONOS satellite imagery was created for the spring of 2006. To facilitate the interpretation the satellite imagery was first orthorectified and pansharpened.

Comparing slope failure inventories related to 1976 and 2006 (Figure 3.6) the frequency of active landsliding in 2006 is 160% higher than for 1976. This increase coincided with a significant change in land use over time (Wasowski et al, 2010). In particular, the areal extent of sown fields grew passed from 52% (of the total territory) in 1976 to 75% by the year 2000.

Soil erosion by water is also a common process in the Daunia Appennines (Iannetta and Trotta, 2008), although it is much less intensively investigated. Ephemeral gullies and rills are widespread phenomena within the study area (Figure 3.5). The silty-clay slopes, especially where sown fields are present are particularly affected by soil erosion processes.

Within the catchment, rills are very common and usually concentrated on the steepest cultivated slopes, were surface runoff concentrates. Their number is particularly high during the wet season. The lack of a good soil and vegetation cover and the low soil cohesion form the main contributing factors for the formation of rills. Ephemeral gullies also form where overland flow concentrates. They are small channels that can be easily erased by tillage practices but often reappearing again in approximately the same location (Soil Science Society of America, 2001). It is only at

the end of the last century that ephemeral gullies have been recognisd as a consistent part of the soil erosion system on croplands (Evans, 1993). Generally larger than rills and smaller than permanent gullies, ephemeral gullies are the result of rills forming a dendritic pattern of channels. Within the catchment ephemeral gullies are quite common geomorphological elements; as for rills they are mainly concentrated on the steepest cultivated slopes and especially where the vegetation cover is limited or absent.

The study area is generally characterized by a moderate relief that, with a few exceptions, doesn't exceed 1000m in height and generally has an elevation that is below 800 meters. The surface slopes are on average around 10° and peak slope angles rarely exceeds 25-30°.

The catchment has an ephemeral drainage network (active in wet season) that includes a main watercourse and a network of small tributaries. The upper hillslope portion of the catchment is characterized by wet zones (areas with free surface-water, including ponds, disordered migrating surface water, seeps) which are concentrated close to the boundaries of limestone and sandstone (Wasowski et al., 2012).

Figure 3.4 – Examples of translational slides in the Rocchetta Sant'Antonio catchment (figure (a) and (b)) (spring of 2012) and detail of the main scarp and head of a big translational slide (figure c, October 2012) occurred in the same area highlighted in picure b. All the landslides in the pictures have a length not exceeding a few tens of meters.

Figure 3.5 - ephemeral gully (a-c) and rills (b-c) (see section 2.1) in Rocchetta Sant'Antonio. Picture b is related to the fall period (October 2012), pictures a and c were taken in the spring of 2012. The maximum depth of the ephemeral gully in picture (a) is around 80-100 cm, the rills in pictures (a) and (b) are generally of uniform spacing and dimension , have a depth generally below 10 cm and are much more narrow than ephemeral gullies. Geomorphological features having similar dimensions and charactristics are present all over the catchment and especially during the wet season (October-March).

3.3.1 Landslides

The Daunia Appennines are well known for recurrent landslide problems (Zezza et al., 1994, Maglioulo et al., 2008). In a study conducted by Mossa et al in 2005 in an area situated at the north-west of the territory of Rocchetta sant'Antonio, a frequency of landslides exceeding 20% was reported. The study also showed that the slope class 10-15° present the highest landslide frequency, followed by slopes characterized by steepness between 5-10° and 15-20°.

The landslide inventory complied by Lamanna et al. (Lamanna et al., 2009; Wasowski et al., 2007) within the municipal territory of Rocchetta Sant'Antonio (Figure 3.6), revealed a strong impact of slope failure, with areal frequency of active landslides amounting to 2% in 1976 and 5.2% in 2006. The majority of these landslides were small and the density per km^2 ranges from 6 (1976) to 34 (2006).

Frequent field visits conducted in the last several years in the Daunia Apennines (Wasowski et al., 2007) confirm that seasonal remobilisations (mainly in winter and spring time) of pre-existing landslides are common. Nevertheless, also first-time shallow landsliding is widespread in rural areas. In most cases the triggering factors seem to be related to rainfall events.

Short-lived shallow mass movements form the great majority of the inventoried slope failures, but the surface expression of much less common, larger and deeper landslides is more persistent. Field observations suggest that the signs of recent activity of these deep landslides, recognizable on remotely sensed imagery, typically do not persist for more than 2–3 years (Wasowski et al., 2007).

Some studies within the municipality of Rocchetta sant'Antonio focused on the catchment traversed by the SP99bis road. This road was damaged by a series of big landslides between the years 2003 and 2005. Following these events subsurface geotechnical investigations were perforemd, as well as piezometer and inclinometer borehole monitoring. The outcomes of these investigations demonstrated the relationship between the presence of high piezometric levels in winter and early spring months and the occurrence of predominantly shallow slope failures (Wasowski et al., 2010, 2012).

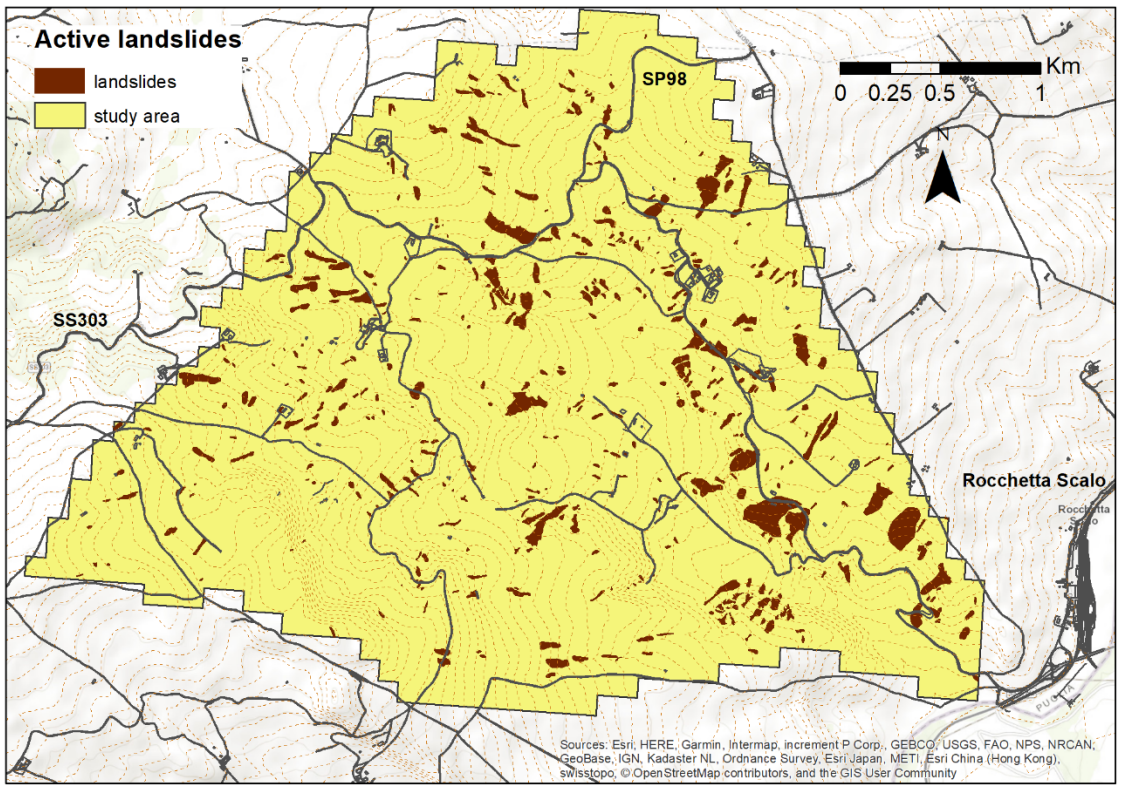


Figure 3.6 - Distribution of active landslides for the year 2006 in the study site of Rocchetta Sant'Antonio, Italy (source of the Background map: Esri et al., 2018).

3.4. Climate

The proximity of the Adriatic Sea directly influences the Rocchetta Sant'Antonio territory resulting in a Mediterranean sub-humid climate. The precipitation range typically varies from 600 to 750 mm with about 60% of the precipitation occurring during the autumn-winter period (Wasowski et al., 2010). The winters are generally mild with limited snow precipitation and the summers are dry (Figure 3.7) and hot, with maximum temperatures easily overpassing 30°. From analysing more than 50 years of rainfall data (1955-2008) Wasoski et al. (2010) found a mean annual precipitation of 667 mm (449 – 1037 mm) with an average of 202 mm in winter (64-470 mm) and 110 mm in summer (19-266 mm). In a mediterranean climate precipitation events are designed as severe if containing at least one hourly reading

exceeding 50 mm (Molini et al., 2011). In the study area relatively intense rainfall events exceed 40 mm (Wasowski et al., 2010).

As is typical in the Mediterranean region, the inter-annual variation of the precipitation is large (Haylock and Goodess, 2004). The precipitation regime characterising this area favours the groundwater recharge over the period October – March (average 403 mm) and overlaps with the maximum instability of the slopes (Wasowski et al., 2012). A part of the climatic data used within the present research comes from a meteorological station located in Rocchetta Sant'Antonio (other climatological data comes from the E-OBS dataset (Haylock et al., 2008)). These data comprise a series of monthly rainfall data covering the period 1955 – 2011 and of three years of daily rainfall data (2010 – 2012) both provided by the CNR- IRPI of Bari.

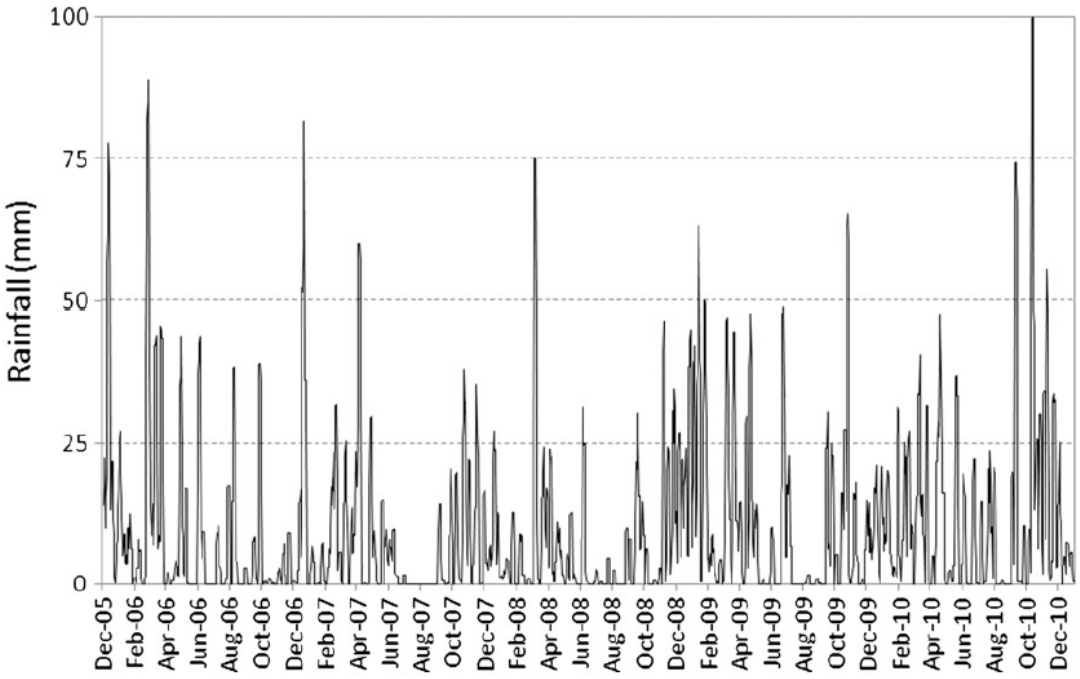


Figure 3.7 – Five-day cumulative rainfall registered in Rocchetta Sant'Antonio (December 2005 – December 2010; source: Wasowski et al., 2012)

3.5. Land use and cover

The study area is intensively cultivated and within the catchment it is possible to distinguish three major classes of land use: i) agricultural soils (covering most of the area), ii) woodland (including shrubs) and iii) grassland and pastures (Figure 3.8). The presence of bare soil is very limited (except for short periods during the year) and mainly corresponds to rock and outcrops (Wasowski et al., 2012). The agricultural land is mainly used for cereal cultivation covering 75.6% of the territory's area. Arborous lands represent 6.3% and are mainly concentrated in the more elevated areas. The remaining part is principally covered by pastures and grassland. As already mentioned, bare lands are very sparse (Wasowski et al., 2012).
The CNR-Irpi (Bari) produced a land use map of the study site using ASTER imagery (July 2000). Four ASTER bands were analysed and used for classifying this area. The procedure has been checked using field data and the results were consistent with the local knowledge of this territory (Wasowski et al., 2010). The limited variability of the land cover characterizing this area also favoured the realiability of the map. The ASTER imagery was preferred over IKONOS because the images were related to the summer period after harvesting and were therefore more suitable for a supervised land-use classification. Furthermore, from analysing both the available images, only minor changes occurred from 2000 to 2006 (Wasowski, 2010).

During the last 30 years the percentage of cultivated areas within the catchment has increased. From about 50% of the land being used as sown fields in 1970 to 75% by the year 2000. This abrupt change is probably related to the introduction of the EU sponsoring wheat cultivation during the mid-late 70s in southern Italy (Wasowski et al., 2012).

The potential geotechnical impact of the changes in land use and vegetation cover is difficult to quantify. These changes are potentially able to modify the equilibrium of the slopes. Cereal cultivation, requiring deep ploughing (depth greater than 50 cm), can alter the soil mechanical and hydraulic properties and strength components (a decreased soil strength and angle of internal friction appear as consequence of soil tillage) (Wasowski et al., 2010). Also, the soil water balance can be altered by deep

ploughing, with direct consequences on the soil water pressure and consequently on the slope stability. An analysis of the landslide trends between the 1955 and 2011 showed a significant increase of landslide activity in this area (Wasowski et al., 2014). The highest landslide susceptibility is focused in the areas that passed from pastures and grazing to cereals. The decrease in effective strength of the soil due to the land cover changes from grasslands to sown fields can be considered as the main cause of the reduced slope stability (Wasowski et al., 2014).

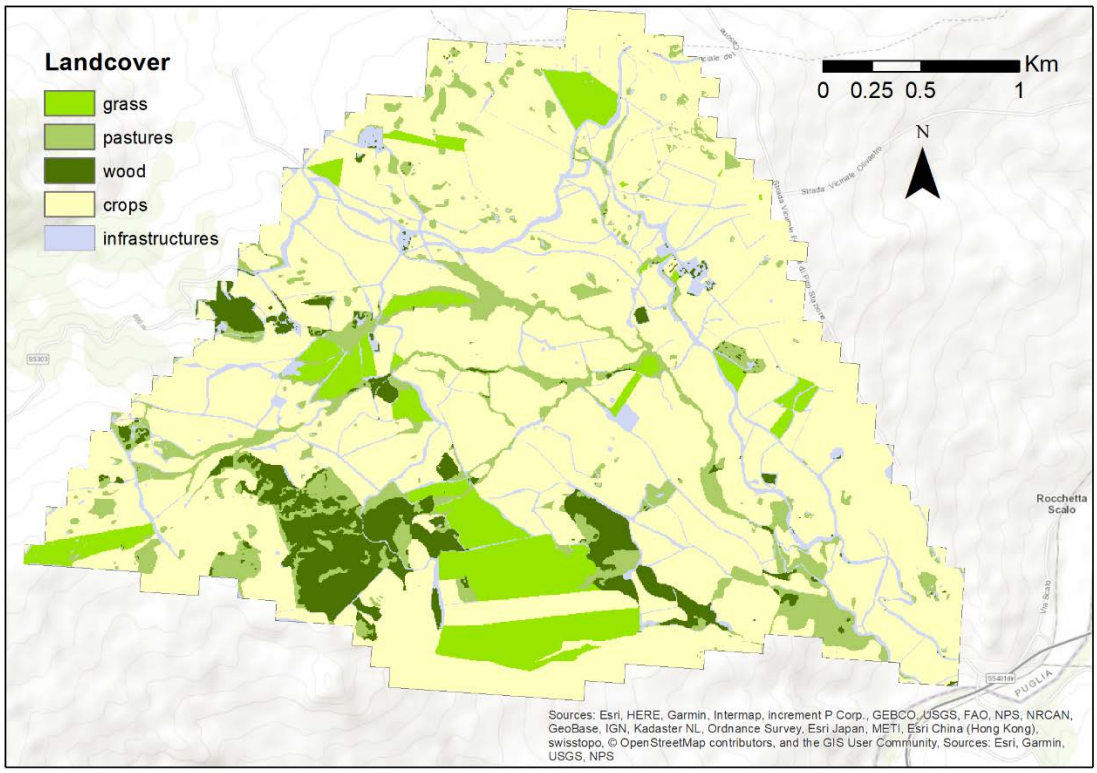


Figure 3.8 – Land cover map of the study catchment. Cereals are the most widespread type of crop and woods are represented by broad-leaved species. Pastures characterize the non-cultivated areas that are mainly localized along the ephemeral drainage network and the main watercourse and grass is present on fallow fields. (source of the Background map: Esri et al., 2018).

3.6. Summary and conclusions

A significant part of the Daunia Appennines is affected by slope instability (Andriani et al., 2015; Magliulo et al., 2008; Cotecchia et al., 2009). Prolonged rainfall and intense storms are the main triggering factors in this region.

Complex landslides form the most common type of landslide affecting this region, particularly occurring in areas where clays-shales outcrop (Andriani et al., 2009, 2015). Soil erosion is also widespread and ephemeral gullies and rills are common features in this region (Figure 3.5). The areas more susceptible to soil erosion process are the cultivated silty-clay slopes, especially after the harvest when the soil is bare.

The study site is particularly susceptible to landsliding (in 2006 more than 400 landslides affected this catchment) (Figure 3.6). Wasowski et al. (2010) describe the most common mass movement affecting this area as seasonal superficial translational slides. These types of landslides are generally characterized by depths not exceeding a few meters (Godt et al., 2008; Meisina and Scvarabelli, 2007). Within the catchment most landslides have slip surfaces that occur within 3 to 4 meters below the ground surface. These typically outnumber the deeper landslides found in the catchment (Wasowski et al., 2012). Widespread shallow landslides mainly occur on the cultivated slopes following very wet winter periods. In most of the cases the landslides were triggered by rainfall but the high groundwater level following prolonged precipitation periods is a causative factor of slope instability (Wasowski et al., 2010), there is a balance between antecedent soil moisture conditions and the level of rainfall required to cause failure. Most of the mass movements occur on slopes between 5 and 20° and the slopes between 10 and 15° have the highest landslide frequency (Lamanna et al., 2009).

Following an analysis of local historical precipitation patterns, landslide occurrence and land use changes, it was found that an increase in landslide susceptibility was the result of land use changes responding to EU economic incentives for the production of durum wheat (Wasowski et al., 2012). This study clearly illustrated the potential of appropriate land management in managing potentially unstable topographies.

4. Modelling soil erosion in data-poor regions

4.1. Introduction

Substantial efforts have been spent on the development of soil erosion models (Nearing et al., 2005). Often, a quantitative assessment is needed to infer on the extent and magnitude of soil erosion problems so that sound management strategies can be developed. Several soil erosion models exist with varying degrees of complexity. While physically based models can in principle offer scientifically sound methods for deriving soil erosion rates from a plethora of detailed input data, their practical suitability at regional/continental scale or at local scale in poor-data conditions is controversial (Bras et al., 2003) (see paragraph 2.2.1). The enormous gap between the type and accuracy of the required input parameters and the actual availability of harmonized, verifiable data sets limits the applicability of such models (Stroosnijder, 2005).

In theory, when working with physically based models, possibly all the requested parameters are measurable and can then be considered as "known". In practice, often the parameters have to be calibrated against observed data (Beck et al., 1995; Wheater et al., 1993). This calibration adds nonnegligible uncertainty in the parameters' values. The heterogeneity, variability and uncertainty associated with input parameter values and their interpolation in spatial or temporal domains outside the observed ones should be considered as key factors (Saltelli et al., 2010; Jetten et al., 2003) which may partially explain why often lumped regression-based models can perform better than more complex physically based models (Bosco et al., 2013; de Vente et al., 2013).

In these conditions, the adaptation of widely used low data demanding empirical models and their application to local conditions could play a meaningful role. Approximations with robust empirical modelling could provide useful – even if necessarily less accurate – support for risk assessors involved in decision-making processes in data-poor regions. The main limit of such an approach is that empirical

models do not necessarily model the right processes and should only be used for the range of conditions they were developed for (Hessel, 2002; de Vente et al., 2013).

4.2. Modelling soil erosion

4.2.1. The model selection process

For selecting the more appropriate soil erosion model to the climatic, geological and geomorphological characteristics of the study area and to the limited amount of available data, an in depth review over thirty-three different soil erosion models was carried out. The soil erosion models we analysed greatly differ in terms of their input requirements, represented processes, application scale and information provided. Following the approach of Merritt et al. (2003) and Aksoy and Kavvas (2005), the processes represented by the models, their temporal and spatial scale of application, the required input, the type of model (physically based, conceptual or empirical), and other additional information such as their integration in a GIS environment were registered (a summary is available in table 2.1).

By considering the characteristics of the Rocchetta Sant'Antonio area and the low input data availability, an in depth discussion on the key aspect of the different factors involved in soil erosion modelling, jointly with the related peculiarities and limits of the analysed models, is reported in this section. The consideration that follow in the text were used for selecting the more suitable models to our conditions.

EMSS (Vertessey et al., 2001; Watson et al., 2001), IHACRES-WQ (Jakeman et al., 1990, 1994; Dietrich, 1999), LASCAM (Viney and Sivapalan, 1999), MIKE-11 (Hanley et al., 1998), SedNet (Prosser et al., 2001), together with SIBERIA (Willgoose et al., 1991 a-d), EGEM (Woodward, 1999), and STREAM (Cerdan et al., 2001) all consider soil erosion process, but they were not specifically developed for estimating soil erosion by water and consequently have limitations over being suitable for our purposes.

IHACRES-WQ, EMSS and MIKE-11 have been mainly developed for the water quality assessment, LASCAM is an hydrological model modified by Viney and Sivapalan (1999) incorporating a conceptualization of the USLE, and SedNet primarily is a sediment transport model (Merritt et al., 2003). EGEM is a model created for estimating ephemeral gully erosion only. SIBERIA despite linking widely accepted hydrological and soil erosion models principally simulates the geomorphic evolution of landforms subjected to fluvial and diffusive erosion and mass transport processes. STREAM is a water balance model that, coupled with other algorithms, is suitable to also measure soil erosion but it is not its primary function. Also HSPF (Johanson et al., 1980) is mainly a model for the simulation of watershed hydrology and water quality, furthermore another limit of the model is that it relies heavily on calibration against field data for parameterisation (Walton and Hunter, 1996).

The MEDRUSH (Kirkby et al., 1998a; Kirkby, 1998b) model also has never been really effective if applied at a small scale (less than 10 km^2). MEDRUSH has been developed for modelling large catchments (100–2500 km^2) in areas dominated by a natural or semi-natural environment (Kirkby and McMahon, 1999). The main model limitation in applying MEDRUSH for our site is imposed by the choice of a minimum 5 km reaches for effective channel routing (Kirkby and McMahon, 1999).

Lumped against distributed models

Many of the factors influencing soil erosion such as soil type, slope and land use, have a strong spatial variability. These factors can not be described using mean values even over areas as small as one field. Spatially lumped models consider an area as a single unit having uniform characteristics. If the spatial variability is a factor that needs to be considered, a dynamic distributed model is the more appropriate option (Morgan et al., 1998). Distributed soil erosion models, with the capability to incorporate a variety of spatially-varying land characteristics and precipitation forcing data, are thought to have great potential for improving soil erosion modelling. However, uncertainty in the high resolution estimates of model parameters may diminish potential gains in prediction accuracy achieved by accounting for the inherent spatial variability.

Climate

The frequency of high intensity precipitations, typical of the Mediterranean climate, is an important selection parameter. The use of models as PERFECT (Littleboy et al., 1992) and PESERA (Kirkby et al., 2003) that do not take into account the intensity of precipitation is not suitable for the Daunia region. Also models such as SWAT (Arnold et al., 1998; Arnold and Fohrer, 2005) or TOPOG (O'Loughlin, 1986, CSIRO, 2017), whose climatic aspects still need further investigation, have to be considered not suitable for our aims.

Scale

The problem of scale' is common to many different disciplines (hydrology, hydrogeology, soil physics, geophysics, and so forth). The values obtained by measuring a physical property (e.g., the saturated soil cohesion) in several points cannot simply be ''averaged'' to get a single value that properly reflects the physics of the process viewed at the watershed or larger scale (Bloschl and Sivapalan 1995). Due to the scale-dependence of erosion models, these models may not be easily used at another spatial scale. The presence of scale-dependence will obfuscate experimental results, particularly if these are carried out under less controlled, semi-natural conditions. This has been explored in various studies (e.g. Lark et al., 2004; Corstanje et al., 2007, 2008) and, for instance, has shown that strong relationships can be observed at some scales even though poor correlations were obtained in the overall experiment.

Models as CREAMS (Knisel, 1980), GUEST (Misra and Rose 1996; Rose et al. 1997) and PERFECT and probably also EUROSEM (Morgan et al., 1998a,b), that are suitable to be applied at the plot or field scale should be used only if the advantage, linked with their use, is so strong to motivate the huge effort in applying the models at a larger scale. Another model that could present some limitations due to the application scale is SEMMED (de Jong et al., 1999). It is a promising model that predicts soil loss at a regional scale with a limited amount of soil data. The model results should also be interpreted with caution due to its sensitivity for rooting depth and initial soil moisture storage capacity (de Jong et al., 1999).

Geomorphology

Soil erosion generally increases with slope steepness and slope length as a result of an increase in the velocity and volume of surface runoff. It is always true until the flow velocity can become independent of the slope, and it happens when the flows have a Froude number around one, with the bed morphology evolving with the flow (Gimenez and Govers, 2001). Dunne (1977) observed that topographic steepness is a significant factor affecting sediment yields. Zingg (1940) was one of the early researchers to relate erosion to slope steepness. He analysed numerous simulated rainfall data finding that the relationship between steepness of slope and soil erosion is represented by a power law equation. The CREAMS model is lacking in this aspect, the outputs are provided for a catchment that is assumed to have a uniform soil topography (Merritt et al., 2003). Also LISEM (de Roo et al., 1994) does not specifically take into account the effect of steep slopes and the presence of gullies (Hessel, 2002).

Unfortunately, only few of the revised models e.g. CREAMS, HSPF, WEPP and WATEM/SEDEM (Van Oost et al., 2000; Van Rompaey et al., 2001) are able to consider the gully erosion process. Due to the small number of ephemeral gullies in our study site, gully erosion was not considered as a fundamental prerequisite in our selection process.

Anthropic factor

Another very important aspect of every soil erosion model relates to the impact of human activities on soil erosion. It is mainly through modifying the flow pattern, the direction of the surface runoff and reducing the amount and rate of runoff, that human activities affect soil water erosion (Renard and Foster, 1983). The anthropic factor can include tillage practices, terracing, and subsurface drainage as well as soil slope stabilization with geogrids or cellular faces. The anthropic factor can lead to a significant increase (i.e. between 2 and up-to 200 times) in total sediment yield (e.g. Einsele and Hinderer, 1997), or in a reduction after implementation of conservation measures (e.g. McIntyre, 1993; Renwick and Andereck, 2006). In catchments having an important human impact “different erosion processes are responsible for the main part of sediment yield than in catchments with limited human impact”. (De Vente,

2009). The collected information related to the reviewed models shows that many models don't implement or only partially consider this aspect. A few of the models (as for example ANSWERS (Beasley et al., 1980), WEPP, LISEM and RUSLE) include a factor for anthropogenic soil disturbance or for conservation measures.

Data Requirement

A considerable effort is still required for improving the prediction capacity of soil erosion models and this should be undertaken in conjunction with methods for improving the quality of the input data: "The quality of model predictions strictly depends on the input data. Judicious data collection is required in order to achieve the maximum benefit in terms of model performance". (Merrit et al., 2003).

The input data required by a model is one of the key factors to be considered for selecting the more suitable model for application. Due to our aim to apply the model at catchment scale in data-poor conditions, a too high data requirement would be a limitation. Although at the local scale, organizing field activities is possible to collect the necessary data, it requires a strong economic and time investment. Data requirement and availability need to be carefully evaluated.

Physically based models generally have a high data requirement, and are very complex to apply. Extensive spatially-distributed data are required to develop inputs for KINEROS2 (Smith et al., 1995a,b), and the subdivision of watersheds into model elements and the assignation of appropriate parameters is both time consuming and computationally complex (Semmens et al, 2008). Also SHETRAN (Ewen et al., 2000) has too high a data requirement and complexity for our scope. Its high data requirement and the components that model physical processes, mainly represented by partial differential equations, require a strong effort for collecting the data and for the calibration of the model. Extensive data sets are required for model parameterization as for example initial overland and channel flow depths, canopy drainage parameters and storage capacities, vegetation root density distribution over depth or the sediment porosities and particle size distributions (Ewen et al., 2000).
The climate component of WEPP (Nicks, 1985) generates mean daily precipitation, daily maximum and minimum temperature, mean daily solar radiation, and mean

daily wind direction and speed (Flanagan et al., 1995). Developing these climatic related files could be very time demanding and expensive.

Sediment deposition

Sediment deposition is a fundamental part of the soil erosion process. Its direct incorporation in soil erosion modelling is not only important for avoiding the overestimation of the soil erosion rate in a plot, field or catchment but also because the spatial distribution of deposited sediments plays a significant role in determining the response of the land surface to the erosive process (Sander et al., 2011).

Many of the reviewed models, such as PERFECT (Littleboy et al., 1992), SWRRB (Arnold et al., 1990), RUSLE (Renard et al., 1997) and MUSLE (Williams, 1975) do not consider the deposition process in estimating soil erosion losses. Anyway, it is always possible to improve a soil erosion model by adding new tools and characteristics. For example, the RUSLE2 (Foster et al., 2000) model (an improved version of the RUSLE) considers two types of deposition, local and remote (Foster et al., 2003).

In this work, an integration technique will be proposed to couple the dominant effects of shallow landslide and soil erosion at the catchment scale, in data-poor areas. At this scale, landslides may affect a relatively small area of the catchment. However, their local effects on the pre-existing land cover may include sharp discontinuities with areas suddenly exposed to the consequences of a significant component of bare soil. Among the dominant effects, an increased rate of soil loss is essential to be considered. Secondary redeposition of sediments may also be present, with local dynamics which may potentially be very complex to model given the extreme conditions characterising the landslide terrain. However, the mitigating effect of this natural process may be unlikely to fully compensate the additional soil loss due to land-cover changes and the newly exposed bare soil component. Therefore, as a first approximation, the overestimation of soil loss without accounting for potential redeposition may be considered as an acceptable simplification.

Land use/cover

Only a few of the reviewed models do not take into account the land use or the land cover for reflecting the effect of the use of management practices or the presence of different land cover in the calculation of the water erosion rate (CREAMS, GUEST and partially SIBERIA). Since this is a fundamental parameter in soil erosion modelling, any model not considering land use/cover measuring soil loss should not be considered as suitable for soil erosion estimation.

Final consideration in modelling selection

By evaluating all the considerations raised from the modelling review, only a few of the 33 models showed the necessary characteristics to be applied in a data-poor area having climatical, geological and geomorphological characteristics as described in chapter 3.

Even if models such as TOPOG, EROSION 3D (von Werner, 2004) or SWAT can be considered as a good option for our aims, they are too complex to be easily implemented and adapted to our scope.

The empirical models: RUSLE2, USPED (Mitasova et al., 1996) and WATEM/SEDEM, are suitable to be applied such as the totally physically based model SIMWE (Mitas and Mitasova, 1998) and three conceptual models: RMMF, AGNPS (Young et al., 1989) and ANSWERS. All these models have the characteristics we highlighted as fundamental in the previous sections. By following the considerations presented in the paragraph 2.2.1. RMMF looks as the more suitable model to be applied in this project.

The RMMF model is simple, flexible and easy to modify, it retains some empiricism but has a strong physical basis. The Morgan–Morgan–Finney model (Morgan et al., 1984; Morgan, 2001) has been used successfully at plot, hillslope and catchment scales in many different environments such as Africa (Vigiak et al., 2005), the Rocky Mountains of the USA (Morgan, 1985), Nepal (Morgan, 2001) and Spain (Lopez-Vicente et al., 2011). Another adavantage of this model is that it is also easy to integrate within a GIS environment.

Despite all these positive considerations on the application of the RMMF model, because of the difficulties we had during the field survey that we planned for collecting some of the data necessary to apply this model, we decided to use a different soil erosion model to estimate soil loss within the catchment. The limiting environmental conditions made not possible to collect information as for example the soil moisture or the effective hydrological depth that are required to properly apply the RMMF model and we also found strong discrepancies between the few data we measured on filed and the data available in the literature (Morgan, 2001; Morgan and Duzant, 2008) (see section 4.2.2.3). To avoid to add further uncertainty to the modelling results by exploiting other models to estimate the lacking modelling data, we decided to apply a less data demanding model more suitable to be applied in very data-poor conditions but still retaining the main characteristics highlighted in our review. Thus, in order to limit the high uncertainty due to the lack or low quality of some of the input data, we applied a modified version of the RUSLE model (e-RUSLE, Bosco et al., 2015).

Before starting to describe the e-RUSLE model and its application within our study, the following paragraph highlight the architecture, potentiality and limits of the RMMF and also the work done to implement this model within our modelling architecture.

4.2.2. RMMF, a soil erosion model suitable for data-poor regions

The Morgan-Morgan-Finney model (MMF) (Morgan et al., 1984) is based on the concepts developed by Meyer and Wischmeier (1969) and Kirkby (1976). Despite being based on an empirical approach it provides a stronger physical basis than the Universal Soil Loss Equation (Wischmeier and Smith, 1978), it retains the advantages of having a low data demand and ease of understanding and applicability. Because of its stronger physical basis many different authors (e.g. de Vente and Poesen, 2005) consider MMF as a conceptual model, and has been used successfully all over the world in a wide range of different environments (Morgan, 1985; Besler, 1987; Shrestha, 1997).

The MMF model retains some simplifications (Morgan, 2001). The model assumes that the rainfall intercepted by vegetation cover does not contribute to soil detachment whereas the proportion which reaches the ground as leaf drainage has the capacity of detaching soil particles, depending on the height of fall (Finney, 1984; Morgan, 2001). An improving in calculating soil detachment by raindrop impact including the leaf drainage is therefore required. Another consistent limit of the MMF model is in not considering the capacity of runoff to detach soil particles. It is difficult to sustain that it doesn't affect the predictive capacity of the model, especially on steep slopes and in presence of rills (Morgan, 2001). Therefore, it seems necessary to modify the MMF model for including soil detachment by runoff also.

Although the MMF model can be considered a simple conceptual model to estimate annual soil erosion and runoff, some of its input parameters are difficult to determine. For example, the top soil rooting depth (D_r) is a component of the model difficult to estimate that also give problems of definition (Morgan et al., 1984, Morgan, 2001). Although rots clearly affect the soil water dynamics it also depends upon the depth of the horizons (especialy the A-horizon).
By answering to all these limits, in 2001, a revised version of the Morgan-Morgan_Finney model (RMMF) was presented by Morgan (2001). This new version of the model, tested with the same dataset used for validating MMF, improved the soil erosion processes description and provided a better support to the users for selecting input parameter values (Morgan and Duzant, 2008; Morgan, 2001).

4.2.2.1. The modelling architecture

The RMMF model and its predecessor the MMF model, requires a moderate number of inputs and has been applied under numerous different land-use and climatic scenarios (López-Vicente and Navas, 2010). It calculates the annual soil erosion rate (E_i, Mg ha^{-1} yr^{-1}) by comparing the total soil detachment ($F+H$) (F is the detachment by raindrop impact and H is the detachment by runoff) and the sediment transport capacity (TC) and taking the lower value (see equation 4.1).

$$Ei = min[\,(F+H),\, TC\,], \tag{4.1}$$

Rainfall Energy

Within the RMMF the calculation of the rainfall energy has been revised from the MMF model by including the effect of leaf drainage.

The procedure for calculating the energy of rainfall starts from the estimation of the effective rainfall (ER, mm). ER is the quantity of the total annual precipitation (R, mm) that directly reaches the ground surface after allowing for rainfall interception by vegetation cover (A, %).

$$ER = RA\ , \tag{4.2}$$

ER is then split into two components: LD and DT. LD represents the rainfall intercepted by plant canopy that reaches the terrain as leaf drainage, DT is the precipitation that reaches the ground surface as direct throughfall.

$$LD = ER \cdot CC \tag{4.3}$$

$$DT = ER - LD\ , \tag{4.4}$$

where CC (%) is the percentage canopy cover (expressed as a proportion between 0 and 1) representing the percentage of soil covered by canopy.

The kinetic energy of the rainfall that directly impacts the ground surface ($KE(DT)$; J/m^2) is a function of the rainfall intensity (I; mm/h). Typical values for the intensity of the erosive rain are: 10 (mm/h) for temperate climates, 25 (mm/h) for tropical climates and 30 (mm/h) for strongly seasonal climates (e.g. Mediterranean area). Within the MMF model the relationship of Wischmeier and Smith (1978) as used in the USLE and applicable to the majority of the United States east of the Rocky Mountains, is used:

$$KE(DT) = DT(11.9 + 8.7\ log\ I), \tag{4.5}$$

Additional equations based on local relationships of the rainfall energy–intensity are available. For example, in central Italy Zanchi and Torri (1980) developed a new equation suitable for the Mediterranean climate (applied within our GIS modelling architecture):.

$$KE(DT) = DT(9.81 + 11.25\, log_{10}\, I)\,, \tag{4.6}$$

$KE(LD)$ (J/m^2) represents the kinetic energy of the leaf drainage. It is the energy of rainfall reaching the soil from leaves and branches of the vegetation cover and, as proposed by Brandt (1990), depends upon the plant canopy height (PH; m):

$$KE(LD) = (15.8 \cdot PH^{0.5}) - 5.87\,, \tag{4.7}$$

When the value of Equation 7 is negative, $KE(LD)$ is assumed to be equal to zero. The total energy of the effective rainfall (KE) is obtained by adding the kinetic energy of the direct throughfall to the kinetic energy of the leaf drainage:

$$KE = KE(DT) + KE(LD)\,, \tag{4.8}$$

Runoff

The procedure for estimating the annual runoff Q(mm) is the same as that applied in the MMF model. The methodology was proposed by Kirkby in 1976 and assumes that runoff occurs when the daily total rainfall exceeds the soil moisture storage capacity (R; mm), and that daily runoff amounts approximate an exponential frequency distribution.

$$Q = R\; exp\left(-\frac{R_c}{R_o}\right)\,, \tag{4.9}$$

where R_o (mm) is the mean rainfall per rain day (mm) and R_c is the soil moisture storage capacity given by:.

$$R_c = 1000\, MS \cdot BD \cdot EHD\left(\frac{E_t}{E_0}\right)\,, \tag{4.10}$$

In the above formula for $R_c\,MS$ is the soil moisture content at field capacity (% w/w), BD is the bulk density of the soil (Mg/m^3), EHD is the effective hydrological depth of the soil (m) and E_t/E_0 is the ratio of actual to potential evapotranspiration. The term, EHD, replaces the rooting depth used in the original model and indicates the depth of soil within which the moisture storage capacity controls the generation of runoff. It is a function of the plant cover, which influences the depth and density of roots, and, in some instances, the effective soil depth, for example on soils shallower than 0.1 m or where a surface seal or crust has formed.

Soil particle detachment by raindrop impact

In the revised MMF model, rainfall interception is allowed when estimating the rainfall energy. It is therefore removed from the equation used to describe soil particle detachment by raindrop impact (F; kg/m^2) which then simplifies to:

$$F = K \cdot KE\ 10^{-3}\,, \tag{4.11}$$

where K is the erodibility of the soil (g/J). Morgan (2001) revised K values in order to cover a wider range of soil textures. The values range from 0.05 for clay to 1.2 for sand passing through Loamy sand (0.3), silty clay (0.5) or Loam (0.8) with values for a total of 12 different soil types.

Soil particle detachment by runoff

The revised model includes a new component to estimate the detachment of soil particles by runoff and is based on experimental work by Quansah (1982). The runoff detachment (H; kg/m2) is considered as a function of runoff (Q), slope steepness (S) and the resistance of the soil (Z) and is estimated from:

$$H = ZQ^{1.5}\,sin\,S\,(1 - GC)10^{-3}\,, \tag{4.12}$$

where GC = percentage ground cover. The equation assumes that soil particle detachment by runoff occurs only where the soil is not protected by ground cover. As a first approximation, this seems reasonable since, where a vegetation cover is

present, the shear velocity of the flow is imparted to the plants and not to the soil. Of course if plants are for a part the effect will be minimal.

For loose, non-cohesive soils, Z=1.0 but based on the work of Rauws and Govers (1988) the dependence of Z on the cohesion of the soil (*COH*, kPa) is given through:

$$Z = \frac{1}{(0.5\ COH)} \tag{4.13}$$

Some guide values on soil cohesion are available in literature (Morgan, 2001) and are based on those used in EUROSEM (Morgan et al., 1993). These range from 2 kPa for sand to 12 kPa for clay.

Transport capacity of runoff

The method for estimating the transport capacity of the runoff (*TC*; kg/m^2) remains unchanged from that used in the original version of the model, so that:

$$TC = CQ^2\ sin\ S \cdot 10^{-3}, \tag{4.14}$$

where C = the crop or plant cover factor, taken as equal to the product of the *C* and *P* factors of the Universal Soil Loss Equation, and S is the slope angle (°).The cover-management factor (*C* factor of the USLE model) represents the influence of land cover, cropping and management practices on soil erosion by water. The human practices factor (*P*) reflects the effects of practices that will reduce the amount and rate of the water runoff and thus reduce the amount of erosion. The C factor can be adjusted to take account of different tillage practices and levels of crop residue retention (Morgan et al., 2001).

4.2.2.2. RMMF in a GIS environment

Geographic Information Systems have been used in various environmental applications since the 1970s, however, it was necessary to wait until the early 1990s to see an extensive application of GIS to hydrologic and hydraulic modelling and mapping (Moore et al., 1991; Maidment and Djokic, 2000). Soil erosion by water is affected by numerous different factors: topography, vegetation, soil characteristics and land use, a Geographic Information System is a very useful tool for managing the large number of spatial data and the complex relationships present within the erosion modelling process.

Despite the implementation of a variety of models to estimate soil erosion, there is still a lack of harmonization of assessment methodologies. Often, distinct approaches lead to significantly different soil erosion rates and even when the same model is applied to the same region the results may differ (Bosco et al., 2011). This can be due to the way the model is implemented (i.e. with the selection of different algorithms either for model's equations or for GIS implementation) and/or to the use of datasets having distinct resolution or accuracy. Scientific computation is one of the central topics within environmental modelling (Casagrandi and Guariso, 2009), to overcome these problems there is thus the need to contemplate the development and implementation of reproducible computational methods during research activities.

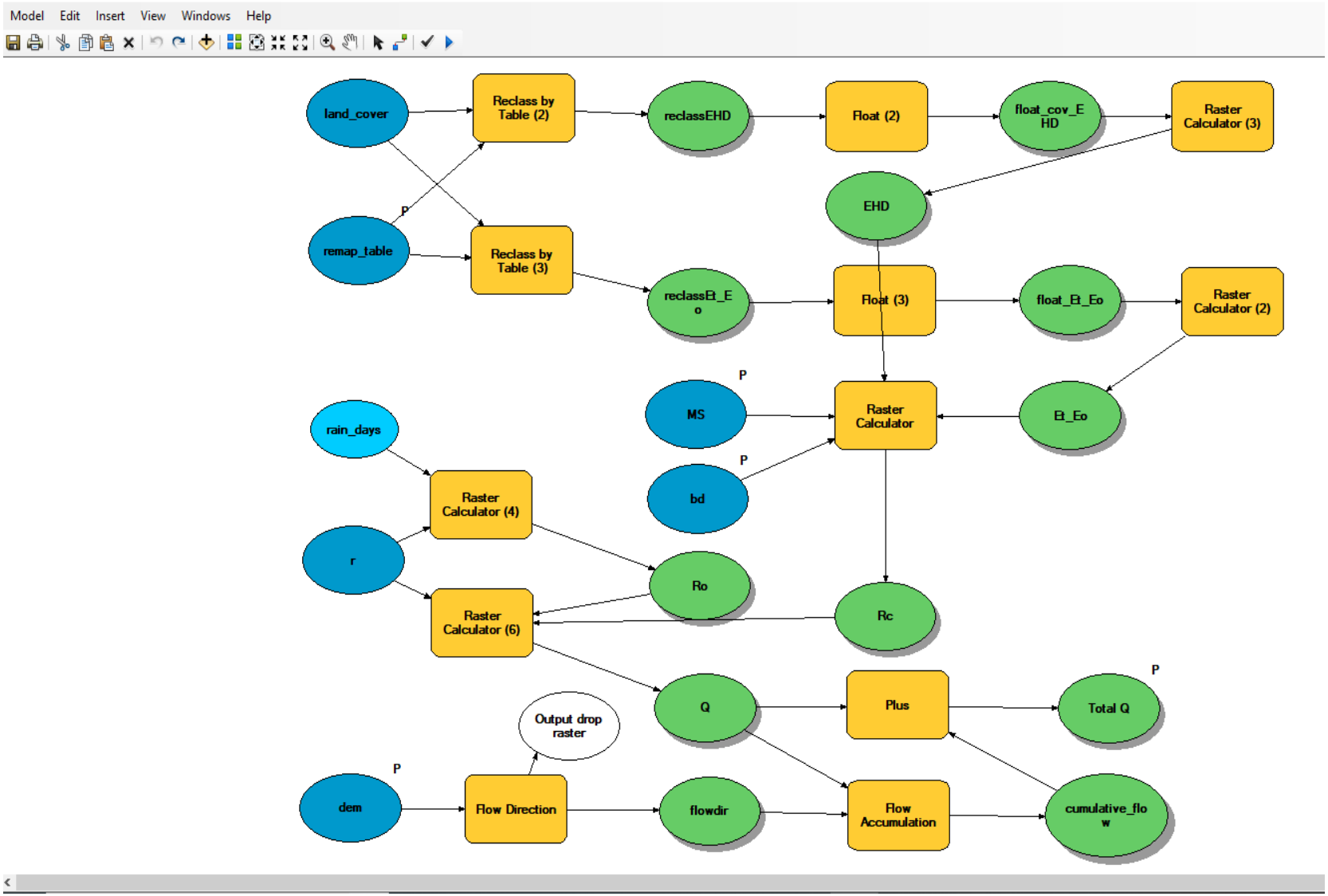
The RMMF model was implemented in ArcGis (using ESRI ModelBuilder) to support the reproducibility of the applied methodology. ModelBuilder is an application for creating, editing and managing models that is part of the ArcGis geoprocessing framework. It is a work flow tool enabling the creation and execution of consistent, repeatable models comprised of one or more processing steps. The use of ModelBuilder ensures the integrity of a particular model or set of analytical processes through modelling, storing, and publishing complex operations and workflows. Within ModelBuilder, a model consists of processes and the connections between them. Most of the geoprocessing tools available in ArcGIS can be used as processes within ModelBuilder. Once implemented, a model is available as a collection of lines

of code written in Python and suitable to be modified or personalized. Parameters can be defined that will be filled into a pop-up form at runtime enabling to run the model with different data for evaluating scenarios.

The Revised Morgan-Morgan-Finney model was implemented as a collection of three different sub-models (scripts). The designed modularization is essential to ease future interactions with third-party sub-models. Ideally, each of the modules might easily be replaced by different arrays of sub-modules and data, without implying a major change in the modelling architecture. The RMMF model considers the soil erosion process as a combination of a water phase and a sediment phase. The water phase is linked with the energy of the rainfall and the volume of the runoff, while the sediment phase considers the soil particle detachment along with the transporting capacity of runoff (Morgan, 2001).

The first of the implemented sub-models (Figure 4.1) calculates the runoff as given by equation 4.9. The second sub model (Figure 4.2) calculates the rainfall energy as given by equation 4.8, and the third sub-model calculates the soil particle detachment and the transport capacity of runoff. The union of the three sub-models then permits the soil water erosion to be calculated.

Figure 4.1 – The picture shows the architecture created for running the Runoff sub-model of the RMMF model calculated by applying the D∞ algorithm of Tarboton (1997) (see section 4.2.3.3). The input parameters (in blue) used in the runoff sub-model (Soil moisture content (MS), Bulk density (BD), mean annual rainfall (R), etc.) are detailed in table 4.1. The equation of the Runoff sub-model is explained in section 4.2.2.1 (Runoff). The land cover map of the area (land cover raster) was used to calculate the values of the effective hydrological depth (EHD) and the ratio of actual to potential evapotranspiration (E_t/E_0) using data from literature (Morgan and Duzant, 2008; Morgan, 2001). The values of Soil moisture content and bulk density were also derived from literature (Morgan and Duzant, 2008; Morgan, 2001) on the basis of the soil types present within the catchment.


Model
Edit
Insert
View
Windows
Help
land_cover
Reclass by Table (2)
reclassEHD
Float (2)
float_cov_EHD
Raster Calculator (3)
EHD
P
remap_table
Reclass by Table (3)
reclassEt_Eo
Float (3)
float_Et_Eo
Raster Calculator (2)
P
MS
Raster Calculator
Et_Eo
rain_days
P
bd
Raster Calculator (4)
r
Ro
Raster Calculator (6)
Rc
P
Q
Plus
Total Q
Output drop raster
P
dem
Flow Direction
flowdir
Flow Accumulation
cumulative_flow

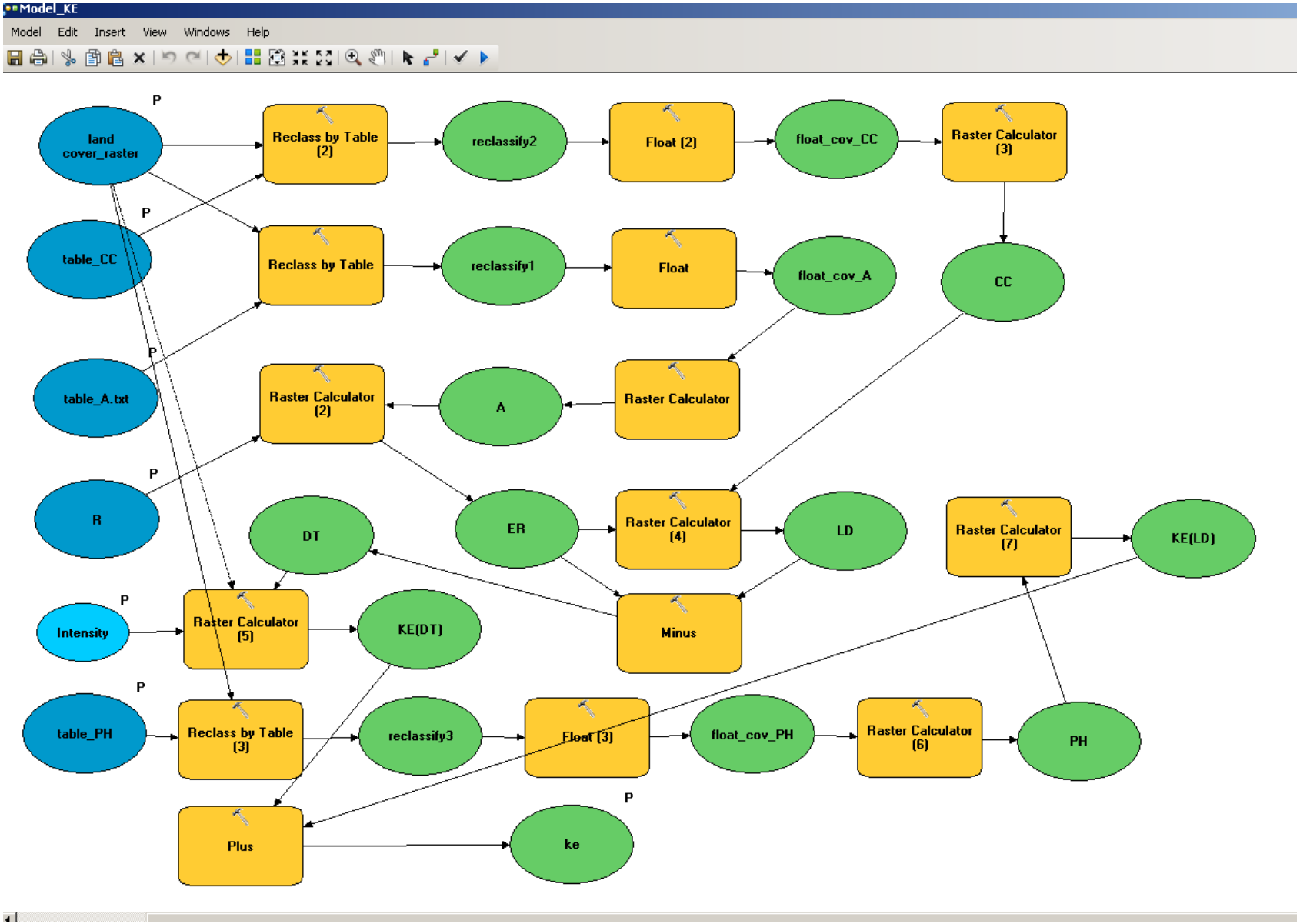

Model_KE
Model
Edit
Insert
View
Windows
Help
P
land cover_raster
Reclass by Table (2)
reclassify2
Float (2)
float_cov_CC
Raster Calculator (3)
table_CC
Reclass by Table
reclassify1
Float
float_cov_A
CC
table_A.txt
Raster Calculator (2)
A
Raster Calculator
R
DT
ER
Raster Calculator (4)
LD
Raster Calculator (7)
KE(LD)
Intensity
Raster Calculator (5)
KE(DT)
Minus
table_PH
Reclass by Table (3)
reclassify3
Float (3)
float_cov_PH
Raster Calculator (6)
PH
Plus
ke

Figure 4.2 –Part of the RMMF model, specifically the Rainfall energy estimation sub-model as builded in Modelbuilder (ArcGis). The equations for estimating the rainfall energy are described in section 4.2.2.1 (Rainfall energy). The main modelling input are the mean annual rainfall (R) and the map of the land cover, jointly with data from Literature (Morgan and Duzant, 2008) used to derive the percentage canopy cover (CC), the percentage of the rainfall intercepted by the vegetation (A) and the plant height (PH) (see table 4.1).

4.2.2.3. The data set and the limits in applying RMMF

Modelling is not simply running a model using a basis of data as input. The construction of the input dataset requires a strong effort. Often the input data is derived from a few basic variables available as raw data. There are numerous choices the modellers have to take regarding input data. For example, how to derive the input variables from available datasets, the way to create continuous maps (interpolation methods have a high degree of subjectivity), and the spatial discretization of the study area (number, shape and size of selected spatial units). These choices are based either on the modeler's experience or on practical considerations (e.g. model limitations).

The RMMF model, requires several land cover and soil data as shown in table 4.1 . Since the majority of these data are not available in Rocchetta Sant'Antonio, we planned a field survey (between September and October 2012) for collecting data on soil cohesion, soil texture, bulk density, soil moisture, soil depth and plant height on a number of fields considered representative for the study catchment. Our decision to collect field data rather than from field-based soil erosion plots, is because in oreder to validate any model at catchment scale or lower, it needs to be tested against data collected in the field and not from plot experiments (Poesen et al., 1996).

The RMFF model requires the cohesion for a saturated soil for calculating the soil particle detachment by runoff; the height of the plants is used for calculating the kinetic energy of the leaf drainage and the soil texture is required by the model for

calculating the soil particle detachment by raindrop impact. Both the bulk density and the soil moisture at field capacity are required for calculating the runoff. Unfortunately, due to the limited amount of available data collected during the field survey it was not possible to fully implement the model as supposed. The prolonged dry season and the unusual very high temperature hampered the collection of sufficient data for mapping these parameters over the whole catchment, even when applying advanced multivariate statistical methods.

Table 4.1 - Input parameters of the RMMF model (Morgan, 2001)

Factor	**Parameter**	**Definition and remarks**
Rainfall	*R*	Annual or mean annual rainfall (mm)
	R_n	Number of rain days per year
	I	Typical value for intensity of erosive rain (mm/h); use 10 for temperate climates, 25 for tropical climates and 30 for strongly seasonal climates (e.g. Mediterranean type and monsoon)
Soil	*MS*	Soil moisture content at field capacity or 1/3 bar tension (% w/w).
	BD	Bulk density of the top soil layer (Mg/m^3)
	EHD	Effective hydrological depth of soil (m); will depend on vegetation/ crop cover, presence or absence of surface crust, presence of impermeable layer within 0.15 m of the surface
	K	Soil detachability index (g/J) defined as the weight of soil detached from the soil mass per unit of rainfall energy
	COH	Cohesion of the surface soil (kPa) as measured with a torvane under saturated conditions
Landform	*S*	Slope steepness (°)
Land cover	*A*	Proportion (between 0 and 1) of the rainfall intercepted by the vegetation or crop cover
	E_t/E_o	Ratio of actual (E_t) to potential (E_o) evapotranspiration
	C	Crop cover management factor; combines the C and P factors of the Universal Soil Loss Equation
	CC	Percentage canopy cover, expressed as a proportion between 0 and 1
	GC	Percentage ground cover, expressed as a proportion between 0 and 1
	PH	Plant height (m), representing the height from which raindrops fall from the crop or vegetation cover to the ground surface

The database to run the RMMF model was compiled using literature (Morgan, 2001; Morgan and Duzant, 2008). Two dataset representing the land cover and lithology of the catchment were made available and published by Dr Janusz Wasowski of CNR-IRPI of Bari (Wasowski et al., 2010, 2012) and a limited amount of data on the lithological and soil characteristics of specific areas of the catchment were obtained as result of commercial laboratory geotechnical tests (Wasowski et al., 2012) (Figure 4.3, 4.4).

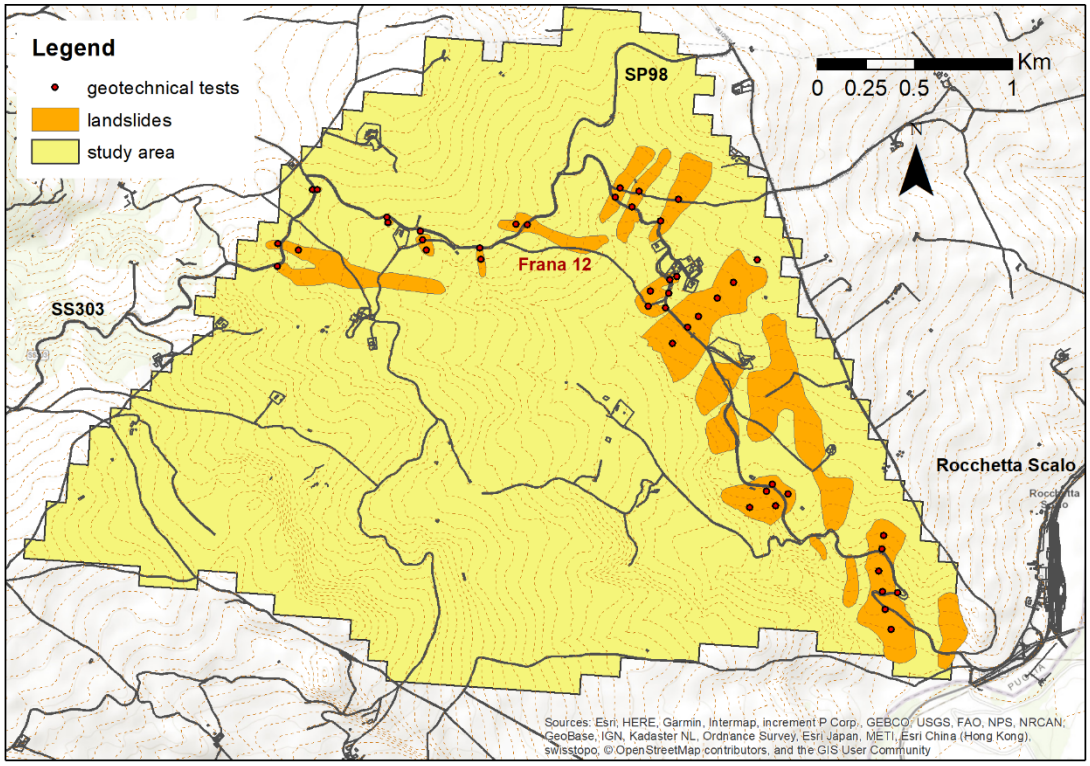


Figure 4.3 – Map of Geotechnical test sites commissioned in the Rocchetta Sant'Antonio catchment by CNR Irpi of Bari. Following a series of landslides that between the years 2003 and 2005 damaged the main road crossing this area,subsurface geotechnical investigations were perforemd. (source of the Background map: Esri, et al., 2018).

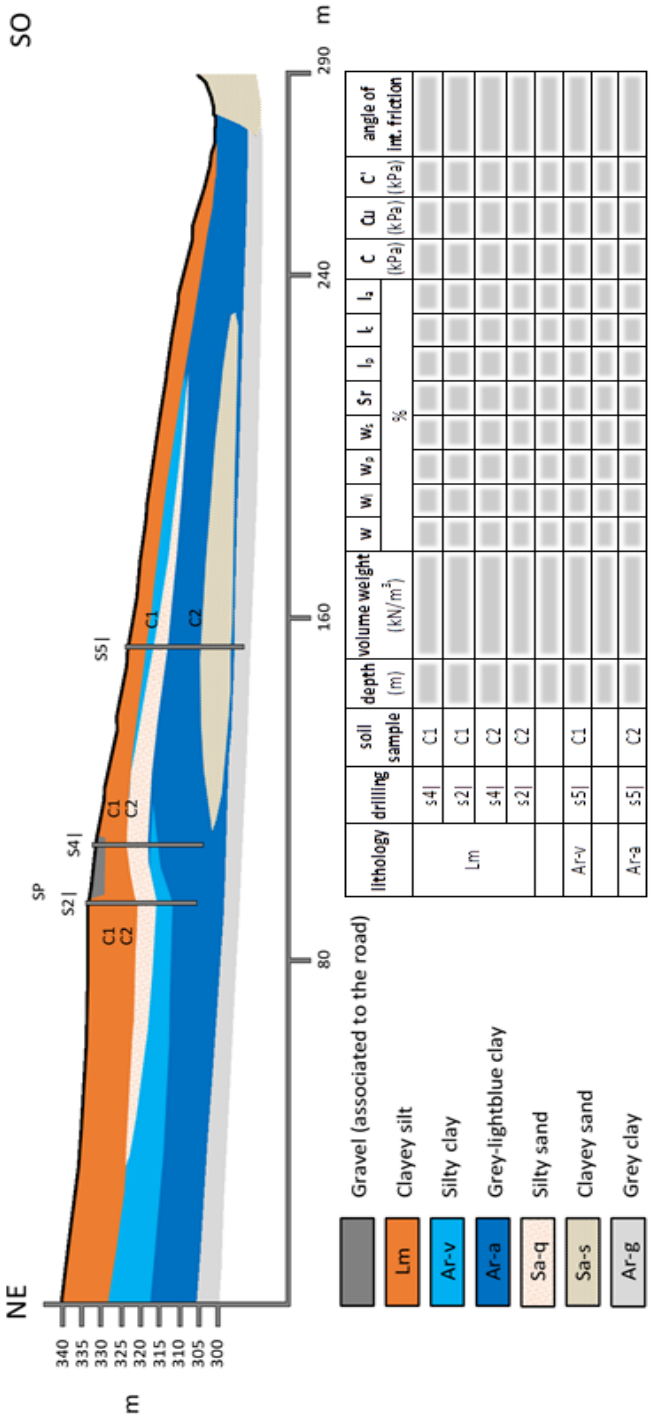


Figure 4.4 - Example of structured information and data collected for a typical set of geotechnical tests associated to a specific landslide in the Rocchetta Sant'Antonio catchment (see Figure 4.3). For copyright reasons, the values were removed from the table (an example of the values related to the tests performed on 'Frana 12' (see figure 4.3) is reported in table 4.8).

The soil related parameters of the RMMF model were also derived, using literature data for the type of soils characterizing the study site (table 4.2). In turn, the information on the different soils type that characterize the catchment were obtained by correlating the lithological map of the area and the geotechnical tests provided by the CNR-IRPI of Bari, an example of which in figure 4.4. We analysed the lithological characteristics of all the test sites to find a relationship with the type of soils characterizing that areas and validationg our results through the collection of some soil samples within the catachment.

Table 4.2. Typical values for soil parameters for the RMMF model (Morgan and Duzant, 2008). MS is the soil moisture content at field capacity and BD is the bulk density of the top soil layer (additional information are reported in table 4.1).

Soil type	**%clay**	**%silt**	**%sand**	**MS**	**BD**
Sand	4	4	92	0·08	1·5
Loamy	6	11	83	0·15	1·4
Sandy	10	25	65	0·28	1·2
Loam	20	35	45	0·20	1·3
Silt	5	89	6	0·15	1·3
Silt	15	66	19	0·35	1·3
Sandy	28	14	58	0·38	1·4
Clay	36	35	29	0·40	1·3
Silty	36	55	7	0·42	1·3
Sandy	42	5	53	0·28	1·4
Silty	48	45	7	0·30	1·3
Clay	64	18	18	0·45	1·1

Table 4.3 - Example of land cover parameters for the RMMF Model (Morgan and Duzant, 2008). CC and GC respectively represent the percentage canopy and ground cover and PH is the height of the plants.

Cover	EHD	E_t/E_0	CC	GC	PH
Woodland	0·20	0·95	0·98	1·0	30·0
Woodland	0·20	0·95	0·95	0·95	25·0
Moorland	0·12	0·90	0·98	1·0	0·5
Moorland	0·12	0·90	0·75	0·30	0·5
Moorland	0·12	0·90	0·95	0·8	0·2
Lowland grass	0·12	0·86	0·90	0·6	0·1
Lowland grass	0·12	0·86	0·80	0·5	0·1
Silage (grass	0·12	0·86	0·90	0·6	0.07
Spring cereals	0·12	0·58	0·80	0·3	1·0
Winter cereals	0·12	0·60	0·80	0·3	1·5
Forage crops	0·12	0·65	0·60	0·6	1·0
Orchards	0·15	0·70	0·98	0·4	4·0
Carrot	0·12	0·70	0·60	0·2	0·3
Maize	0·12	0·68	0·65	0·5	2·0
Vineyards	0·12	0·30	0·80	0·80	1·5
Bare soil (no	0·09	0·05	0·0	0·0	0·0

For the land cover related parameters of the RMMF model (e.g. percentage canopy cover, percentage ground cover, plant height, etc.), these were derived from the work of Morgan and Duzant (2008) using the land cover map of the study catchment. Despite the limiting environmental conditions that occurred during the field survey allowed us to collect a limited amount of data, by comparing the collected information on soil cohesion, bulk density, soil moisture and plant height with the layer determined from literature data, we noticed numerous significant differences. For example, by analysing the layer related to the plant height, we noticed a strong discrepancy among the observed and derived plant height. The information collected during the field survey show a vegetation height of the wooded areas generally among three and six meters (points 41, 19, 107, 0 and 51 in figure 4.5), showing high discrepancies with the values of Table 4.3. Due to the significant variability of this parameter that highly depends from local climatic conditions (water availability, mean annual temperatures, windiness, etc.) and anthropic influence (Moles et al., 2009; de Bello et al., 2005) data from the literature are not suitable for calculating this model input parameter.

Also the values of soil cohesion we estimated measuring the shear strength in different points within the catchment (Figure 4.5) are not in line with data we obtained from the literature (Morgan, 2001). Due to the extreme dry conditions it was not possible to estimate soil cohesion under satured conditions. The soil cohesion is a parameter that normally has a high variability in the field and consequently cannot be adequately represented by a single number (Morgan and Quinton, 2001). The single value of 12 kPa suggested by Morgan (2001) for clay is indeed also not in line with the high variation of this parameter that we measured on field. In figure 4.6 is reported an example of the values we obtained using a shear vane (Figure 4.7) and that we converted to get the values of soil cohesion on a clay soil, this parameter showed a high variability such as in almost all the areas we analysed (see Figure 4.5).

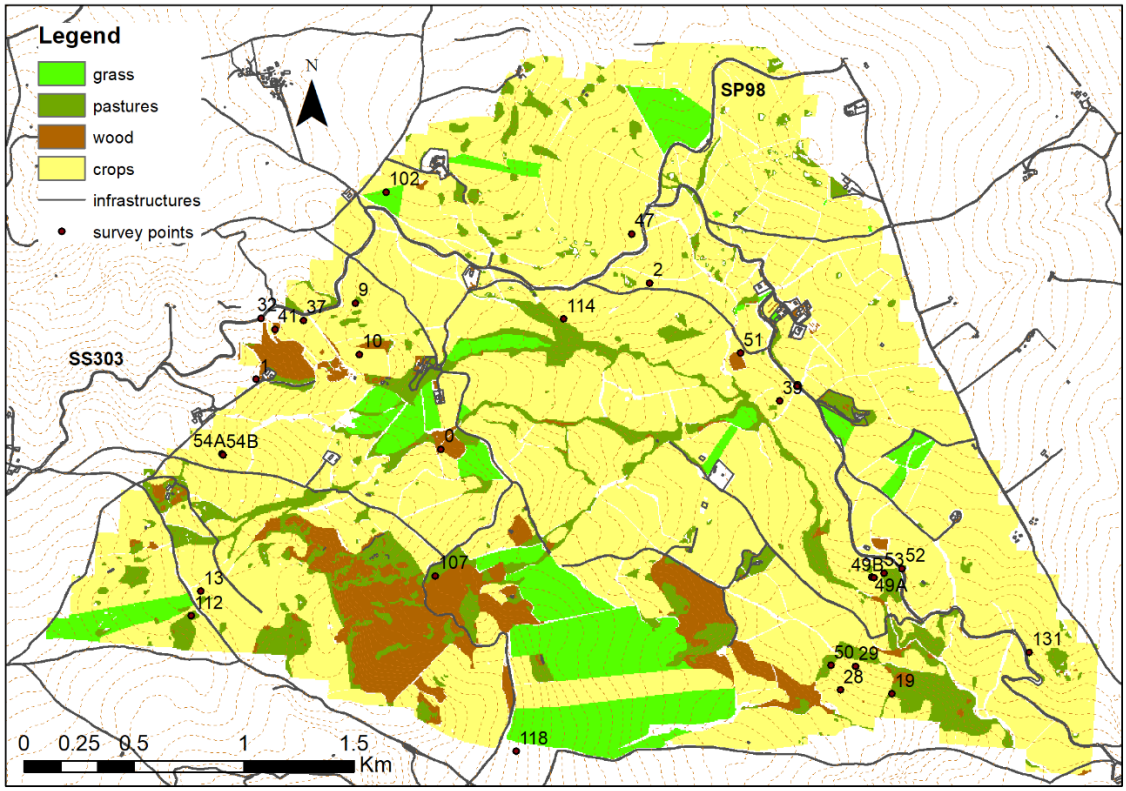


Figure 4.5 – Map of the land cover and of the survey points related to the field survey carried on within the study site between September and October 2012 (Wasowski et al., 2010).

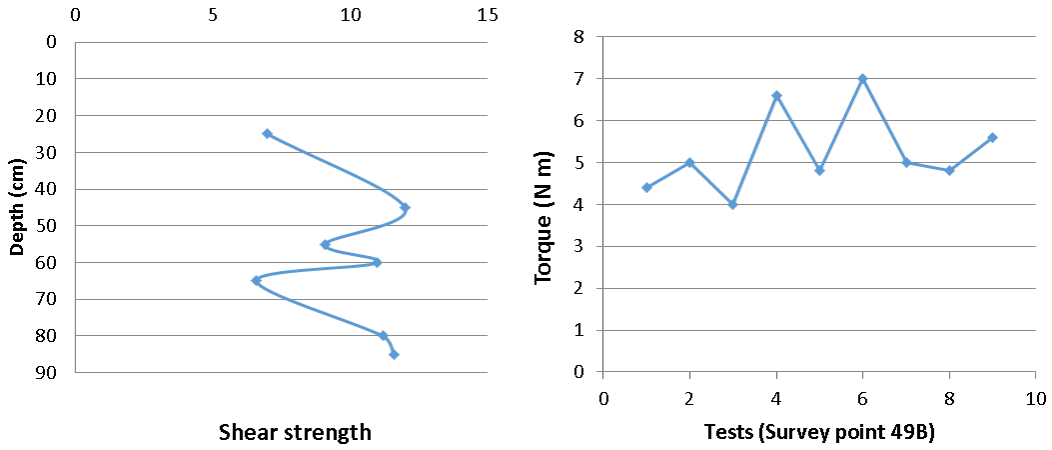


Figure 4.6 – Results of one of the field tests performed using a shear vane for measuring the torque and undrained shear strength (kPa) in the study site. The undrained shear strengths appear to be very low. Results refer to the values measured in the survey point 49B. The picture on the right shows the values related to different superficial tests performed in the surrounding area of point 49B (max distance 50 cm).

Figure 4.7 – In-situ test with a shear vane for determining soil cohesion (see picture 4.6).

Thus, as already mentioned at the beginning of this chapter, to limit the modelling uncertainty due to the lack of detailed input data, the e-RUSLE model is now considered.

4.2.3. e-RUSLE, a soil erosion model for data really-poor regions and large spatial extents

A modelling architecture based on an extended version of the RUSLE model (e-RUSLE) (Bosco et al., 2015) is now presented and adapted to be applied at our study site.

The e-RUSLE soil erosion model is based on an architecture designed for easing the integration of erosion-related natural resources models. The semantic array programming paradigm (de Rig, 2012a, 2012b) and the computational reproducibility (Bosco et al., 2011; de Rigo and Bosco, 2011) are at the basis of the applied modelling architecture. Its flexibility, low data demanding and architecture make this model as one of the best options for modelling soil erosion in data-poor regions and at large spatial extents. Furthermore, the family of models based on the USLE provides long-term average soil loss estimates and has been applied all over the world in different environments and various climatic conditions (e.g. Kinnell, 2010; Lu et al., 2004; Angima et al., 2003; 165 Bosco et al., 2009).

We have already discussed the (still open) challenges to obtain an appropriate parameterisation for physically based models to run in data-poor areas (see section 2.2.1). Their failure to produce better results than achieved using the USLE/RUSLE family of models (Tiwari et al., 2000), jointly with the difficulties we also met in applying conceptual models as RMMF, encourage the use of the USLE/RUSLE model in applications for which it was not designed (Kinnell, 2010). Although even the e-RUSLE parameterisation is not trivial, in particular considering the lack of data affecting the study site, the e-RUSLE array-based structure will be shown to offer a potential strategy to integrate uneven but locally accurate spatial information on key quantities – here illustrated for the case of erosivity.

4.2.3.1. Data set

Electronic archives are an important data source for the scientific community. The added value and criteria for the selection of electronic archives are the accessibility of large volumes of data, their spatial coverage and their ability to preserve historical data (Panagos et al., 2011) and often their free availability.

The data needed for running the model was compiled using the literature, public available data sets, and few detailed set of data coming from the National Research Council (CNR) of Bari:

- Map of the main lithological units (figure 3.3) in the study area (Wasowski et al., 2012).
- ENSEMBLES Observations gridded data set (E-OBS) (Haylock et al., 2008). https://www.ecad.eu/download/ensembles/download.php
- Precipitation data from a meteorological station located in Rocchetta Sant'Antonio (figure 3.7) (Wasowski et al., 2012).
- Map of land cover produced by CNR-IRPI of Bari (figure 3.8 and 4.5) (Wasowski et al, 2010).
- European Soil Geographical Database (SGDBE) (Heineke et al., 1998). https://esdac.jrc.ec.europa.eu/ESDB_Archive/ESDBv2/fr_intro.htm
- A digital elevation model (DEM) with 5 m grid size (Wasowski et al., 2012).

The map of the main lithological units was derived from the map published by Wasowski et al. in 2012. It has a spatial resolution of 5x5 meters and was georeferenced in WGS1984 UTM zone 33N. The original map was produced by the CNR-IRPI (Bari) and was obtained by integrating information from the state geological map (Malatesta et al., 1967) and in situ checks (Wasowski et al., 2012).
A part of the climatic data used within the present research come from a meteorological station located in Rocchetta Sant'Antonio. These data comprise a series of monthly rainfall (mm) covering the period 1955 – 2011 and of three years of daily rainfall (2010 – 2012) coming from a pluviometric station located in Rocchetta

sant'Antonio, both kindly provided by the CNR- IRPI of Bari (Wasowski et al., 2012). Because daily rainfall data covers only a few years, these datasets were not used to directly calculate the rainfall erosivity within the e-RUSLE model but were used to test the reliability of the E-OBS dataset in the region.

E-OBS is a European daily gridded observational data set for precipitation and air temperature that covers the period 1950–2018. The database contains gridded data for 5 elements (daily mean, minimum and maximum temperature, daily precipitation sum and daily averaged sea level pressure). The dataset is available in compressed NetCDF format. E-OBS is based on the largest available pan-European precipitation data set, and its interpolation methods were chosen after careful evaluation of a number of alternatives (Haylock et al., 2008). The gridded data are delivered on four spatial resolutions, the 0.25˚ regular lat–long grid resolution has been used for our site. An added value of the E-OBS data set is the daily estimates of interpolation uncertainty, provided as standard error.

The CNR-IRPI of Bari produced a 5x5 metres resolution land cover map of the study site using ASTER satellite multi-spectral imagery (July 2000) (projected in WGS1984 UTM Zone 33N). To produce this map, four ASTER bands (three visible bands and one VNIR band) were analysed using standard commercial software (Wasowski et al., 2010) and used for classifying this area. The applied procedure was checked using field data and the results were consistent with the local knowledge of this territory (Wasowski et al., 2010). The reliability of the results was also favoured by the limited number of land-cover types considered (all characterized by a distinctive spectral signatures) (Wasowski et al., 2010). Three main classes were distinguished: (1) agricultural land (with mainly cereal cultivation and that represents the predominant class), (2) wooded landv (3) other (that includes uncultivated and bare land, grassland, pasture and infrastuctures).

Information on soil rock fragment content was determined from the SGDBEdataset (Heineke et al., 1998), all the data present within this dataset are available in Lambert Azimuth coordinate system. The 1:1,000,000 SGDBE data set contains a list of Soil Typological Units (STU) representing the properties of European soils. In the STU table, each STU has a number of attributes (e.g. dominant parent material, dominant

surface textural class, etc.). I The STUs are grouped in Soil Mapping Units (SMU) to form soil associations because of the difficulty in delineating the STUs at the database scale. Lastly, the digital elevation model (DEM) (coordinate system: WGS1984 UTM zone 33N) with a 5 m grid size of this area was obtained from 1:5,000 scale topography map (Wasowski et al., 2012).

4.2.3.2. The modelling architecture

The e-RUSLE architecture inherits from the RUSLE the ability to be easily linked to other related natural resources. For example, some effects of forest resources and generally of the vegetation component within land-cover are straightforward to assess (de Rigo and Bosco, 2011). Furthermore, approximated rapid assessments of the impact of disturbances (e.g. wildfires, de Rigo et al., 2013a; Di Leo et al., 2013) may be performed by exploiting the RUSLE modular architecture which easily allows potential and actual erosion rates to be estimated for different environmental conditions by simply considering different arrays of layers. The further seamless integration of multiple estimates in the rainfall erosivity component supports a more robust adaptability to heterogeneous climatic conditions typical of the Southern Italy. The RUSLE retains from the USLE (based on empirical correlation) some limitations. Within the model there are no factors directly representing physical processes (i.e, runoff, infiltration) and RUSLE only predicts soil losses caused by sheet and rill erosion, not by (ephemeral) gully erosion. Another fundamental lack is linked to the absence of estimating sediment deposition which can lead to overestimating soil erosion rates. However, the USLE multiplicative structure (Ferro, 2010) is well suited for transforming the modelled quantities into other correlated ones by simply adding custom factors. As an example, for overcoming the absence of sediment deposition calculations, Mitasova et al. (1996) replaced the LS factor with a new index considering the spatial distribution of areas with topographic potential for soil erosion and sediment deposition.

The e-RUSLE preserves the structure of the RUSLE adding to its array of multiplicative factors one more factor for better considering the effect of stoniness on soil

erodibility. An array of local estimations of rain erosivity has also been introduced for mitigating the extrapolation uncertainty associated to each single rainfall erosivity equation. The array-based estimation of rainfall erosivity is proposed to be an ensemble of multiple estimations from partly independent modules (empirical equations) aggregated by a similarity analysis so as to also increase the design diversity (de Rigo, 2013).

The methodology relies on the paradigm of Semantic Array Programming (de Rigo, 2012a,b) (see section 2.5) which allows the multi-dimensional structure of the mathematical and computational model to be explicitly and concisely exploited. This is achieved by semantically enhancing the chain of involved data-transformation modelling (D-TM) modules so as to better focus on a compact, modular integration of the arrays of data and geospatial layers.

Applying the semantic array programming paradigm

Although the impact of the computational aspects in environmental modelling is steadily growing (Casagrandi and Guariso, 2009), they may be undervalued (Merali, 2010) and the mitigation of the software-driven component of uncertainty in complex modelling might be understated while focusing on more traditional sources of uncertainty (Cerf, 2012; de Rigo, 2013).

Part of the complication in computational models (affecting even their maintainability and readiness to constantly evolve) may be mitigated (McGregor, 2006). Compared to other computational approaches, array programming (AP) understands large arrays of data as if they were a single logical piece of information. For example, a continental-scale gridded layer (such as the layers in the e-RUSLE application at the European scale, Bosco et al, 2015) may be managed by AP languages as if it were a single variable instead of a large matrix of elements. As a consequence, a disciplined use of AP (Iverson, 1980) may allow nontrivial computational workflow to be espressed with more compact data-processing operators (Taylor, 2003). This level of conciseness reduces the number of lines of code and may simplify the control flow in the model. For example, assigning all the values of a matrix M greater than a threshold T to a new value N may typically require a nested loop for each row and each column of the matrix, with an if statement to

check whether the element currently under examination exceeds the threshold. With AP languages such as GNU Octave, the list of elements exceeding the threshold T is simply assigned to the new value N with the instruction M(M > T) = N, which is a greatly more compact notation.

However, the powerful abstraction and conciseness of AP – without the additional use of a disciplined semantics-aware implementation – might offer a weak support for checking the correctness of some extremely compact algorithms (de Rigo, 2012b). For example, in GNU Octave and MATLAB languages computing the square root of negative values is a prefectly legal operation which leads to a complex-valued result. This kind of result may be a physical nonsense for measured quantities, and a semantic check (e.g. to simply detect negative values where only nonnegative ones should be present) may suffice to avoid it.

This is why the e-RUSLE computational modelling methodology (de Rigo and Bosco, 2011; Bosco and de Rigo, 2013) follows the paradigm of Semantic Array Programming (as introduced in section 2.5) by combining concise implementation of the model with its conceptual subdivision in semantically enhanced abstract modules.

It is worthy recalling two main aspects which characterise SemAP as a specialisation of AP:

i) the modularisation of sub-models and autonomous tasks, paying attention to their concise generalization and the potential reusability in other contexts;and

ii) the use of terse array-based constraints (SemAP semantic checks, de Rigo, 2012d) to emphasize the focus on the coherent flow of the information and data among modules – which are often nontrivial in computational science.

The SemAP semantic constraints apply to AP variables irrespective of their size (e.g. large arrays such as continental-scale geospatial layers). The semantic coherence of the information entered in and returned by each D-TM module (D-TM inputs and ouputs) is checked locally instead of relying on external assumptions. This may be essential especially when different modules rely on different expertise. This way, even the essential implementation details within each module (for example, the

implementation of the erosivity layer in the e-RUSLE as a climatic-driven composition of an array of local empirical relationships) may be at least partially decoupled from the overall modelling architecture.

Ideally, modules might easily be replaced by more complex compositions of arrays of sub modules and data, without implying a major change in the modelling architecture. For example, the same methodology exploited for the erosivity layer was also exploited in Bosco et al. (2013) for estimating landslide susceptibility.

SemAP array-based semantic constraints (de Rigo, 2012d) have been exploited in the model implementation. Some of them are exemplified hereinafter as active links ::**sem**::[3] following the notation introduced in section 2.5..

The Extended RUSLE model (e-RUSLE)

The e-RUSLE model is designed to predict only soil loss by sheet and rill erosion. As previously mentioned, sediment deposition processes or concentrated overland flow erosion (ephemeral gully erosion) are not considered in the equation. The model uses different factors representing the effect of topography, land cover, climatic erosivity, management practice and soil erodibility.

The basic equation of the extended RUSLE is as follows:

$$Er_{c,Y} = R_{c,Y} K_{c,Y} L_{c,Y} S_{c,Y} C_{c,Y} St_{c,Y} P_{c,Y} \tag{4.15}$$

where all the factors refer to a given spatial grid cell c and are the average within a certain set of years $Y = y_1, \cdots, y_i, \cdots, y_{nY}$ of the corresponding yearly values:

$Er_{c,Y}$ = average annual soil loss (t ha^{-1} yr^{-1}).

$R_{c,Y}$ = rainfall erosivity factor (MJ mm ha^{-1} h^{-1} yr^{-1}).

$K_{c,Y}$ = soil erodibility factor (t ha h ha^{-1}MJ^{-1}mm^{-1}).

$L_{c,Y}$ = slope length factor (dimensionless).

[3] The mathematical notation ::**sem**:: refers to the online taxonomy of array-based semantic constraints which defines the Semantic Array Programming paradigm (http://mastrave.org/doc/ mtv_m/check_is, de Rigo, 2012d).

$S_{c,Y}$= slope steepness factor (dimensionless).

$C_{c,Y}$= cover management factor (dimensionless).

$St_{c,Y}$= stoniness correction factor (dimensionless).

$P_{c,Y}$= support practice aimed at erosion control (dimensionless).

Given the multiplicative structure, all layers are expected to be defined in a given grid cell *c* without missing values (::**nanless**::[4]) in order for the soil loss to be computable in c.

Rainfall Erosivity Factor

The intensity of precipitation is one of the main factors affecting soil water erosion processes. The Rainfall Erosivity Factor (R) is a measure of precipitation's erosivity. Wischmeier (1959) identified a composite parameter, EI30, as the best indicator of rain erosivity.

The rainfall erosivity factor has been implemented in numerous soil erosion models: AGNPS (Young et al., 1989), WATEM (Van Oost et al., 2000), USPED (Mitasova et al., 1996), SEMMED (De Jong et al., 1999) and MMF (Morgan et al., 1984). The rainfall erosivity factor has been widely applied all over the world and it is considered as an important factor for soil erosion assessment under climate change scenarios. Despite its frequent use, it retains some limitations. The main weakness of the R factor is in not explicitly considering runoff and this highly influences the capacity of the model to account for event erosion (Kinnell, 2010) and seasonal effects.

The scarcity of accurate datasets for assessing soil water erosion rates in data-poor conditions motivated the introduction of a climatic-based ensemble model to estimate erosivity of rainfall. The climatic layers have been computed using GNU R (R Development Core Team, 2014) and GNU Octave (Eaton et al., 2008) with the Mastrave modelling library (de Rigo, 2012a; 2012b). The ensemble is an unsupervised data-transformation model applied to climatic data to reconstruct erosivity.

Due to the short period over which daily rainfall data from the meteorological station located in Rocchetta Sant'Antonio are available, the R factor has been

[4] http://mastrave.org/doc/mtv_m/check_is#SAP_nanless

computedusing the E-OBS database for calculating the mean rainfall erosivity over a period of 30 years.

The same seven empirical equations (Table 4.6) that have been selected from the literature in order for the erosivity to be correlated with climatic information for large scale applications (Bosco et al., 2015), were exploited. Due to the high seasonal variability characterizing the precipitation pattern of the study site, we considered the same equations (Table 4.6) we tested in our application of the e-RUSLE at European scale (Bosco et al., 2015), to be also able to characterize the large inter-annual variation of the precipitations affecting this area.

Spatially distributed climatic information (such as average annual precipitation, Fournier modified index, monthly rainfall for days with ≥ 10.0 mm, (see Table 4.5) has been computed from the daily reconstructed (E-OBS) patterns of precipitation in Europe (years 1980-2009).

Table 4.4 - Climatic information: auxiliary variables based on precipitation patterns $P\cdot,c$ in a given spatial grid cell c. $P_{day,c}$ and $P_{m,c}$ respectively refer to the precipitation in c for the day day and the month m. The values are computed considering years y in a set of n_Y years.

Climatic information	Definition	Number of variables
Average monthly number of days with daily rain ≥ 10 mm [*dimensionless*]	$D^{10}_{m,c} = \frac{1}{n_Y} \sum_{\text{month}(day)\equiv m} \left[P_{day\,c} \geq 10\text{mm}\right]$	12
Annual Modified Fournier index [*mm*]	$F_{y,c} = \sum_{m=1}^{12} \frac{\left(\sum_{\substack{\text{month}(day)\equiv m \\ \text{year}(day)\equiv y}} P_{day,c} \right)^2}{\sum_{\text{month}(day)\equiv m} P_{day,c}}$	n_Y

Table 4.5 - Climatic information: covariates based on precipitation patterns in a given spatial grid cell c. P refers to the precipitation in c (for the day day and the month m). The values are computed considering a specific year y in a set of n_Y years.

Climatic information	Definition	Number of covariates
Average monthly precipitation [*mm*]	$P^0_{m,c} = \frac{1}{n_Y} \sum_{\text{month}(day)\equiv m} P_{day,c}$	12
Average monthly precipitation of days with daily rain ≥ 10 mm [*mm*]	$P^{10}_{m,c} = \frac{1}{n_Y} \sum_{\text{month}(day)\equiv m} P_{day,c} \cdot \left[P_{day,c} \geq 10\text{mm}\right]$	12
Modified Fournier index [*mm*]	$F_c = \frac{1}{n_Y} \sum_{m=1}^{12} \frac{\left(P^0_{m,c}\right)^2}{\sum_{m=1}^{12} P^0_{m,c}}$	1
Fournier-Ferro index [*mm*]	$F_{F,c} = \frac{1}{n_Y} \sum_{y=y_1}^{y_{n_Y}} F_{y,c}$	1

Table 4.6 - List of empirical equations for estimating the rainfall erosivity (EI_{30}) in a given spatial grid cell c. α and β are coefficients with α_{b1} and α_{b2} in MJmmha^{-1} h^{-1} yr^{-1}, β_{b1} and β_{b2} in mm^{-1}. β_{r1} and β_{r2} are expresses in mm^{-1} · MJmmha^{-1} h^{-1} yr^{-1} and $\alpha_{r1,2}$ in MJmmha^{-1} h^{-1} yr^{-1}. α_{d1} and β_{d1} were respectively in mm^{-1} · MJmmha^{-1} h^{-1} yr^{-1} and MJmmha^{-1} h^{-1} yr^{-1}. β_{fi} is dimensionless and α_{fi} in mm$^{-\beta fi}$· MJmmha^{-1} h^{-1} yr^{-1}.

Country/ region	ISO 316 6	Definition	Reference
A_1 Belgium	BE	$\begin{cases} R_{b1,c} = \alpha_{b1} \exp\left(\beta_{b1} \cdot \sum_{m=1}^{12} P^0_{m,c}\right) \\ R_{b2,c} = \alpha_{b2} \exp\left(\beta_{b2} \cdot F_c\right) \end{cases}$	Bollinne *et al.,* 1979
A_2 Bavaria, Germany	DE-BY	$\begin{cases} R_{r1,c} = \alpha_{r1} + \beta_{r1} \cdot \sum_{m=1}^{12} P^0_{m,c} \\ R_{r2,c} = \alpha_{r2} + \beta_{r2} \cdot \sum_{m=5}^{10} P^0_{m,c} \end{cases}$	Rogler and Schwertmann, 1981
A_3 Algarve, Portugal	PT-08	$R_{d1,c} = \alpha_{d1} \cdot \sum_{m=1}^{12} P^{10}_{m,c} + \beta_{d1} \cdot \sum_{m=1}^{12} D^{10}_{m,c}$	de Santos Loureiro and de Azevedo Coutinho, 2001
A_4 Sicily, Italy	IT-82	$R_{fi,c} = \alpha_{fi} \cdot F_{F,c}^{\beta_{fi}}, \quad i = \{1,2\}$	Ferro *et al.,* 1999

The equations in table 4.6 refer to climate-erosivity regressions, which have been validated in 4 geographical areas. Many other elementary relationships exist between climate and erosivity (Bosco et al., 2009; Bosco et al., 2015), the selected ones fulfil a series of expert-based criteria such as their reproducibility using the available datasets, a solid international literature, the climatic coverage of heterogeneous areas ranging from arid to wet conditions and the good regression performance validated in their spatial extent (see figure 4.8 and the discussion in section 2.5). As an example, although linear in the parameters' regression, the empirical approach (eq. Rd1, region A3 in Table 4.6) proposed by de Santos Loureiro and de Azevedo Coutinho (2001) received wide acceptance (Onyando et al., 2005; Taveira-Pinto et al., 2009; Ranzi et al., 2012). The relationship has been tested in Italy (Diodato, 2004) where it provided estimates more stable (lower error) than the ones provided by other widely used empirical equations (in the limited validation set, the estimates of Rd1 did not show rank reversals when compared to the measured erosivity). This equation as also the advantage to be tested in an area with climatic conditions that could be similar to the study site (e.g. summers characterized by prolonged hot and dry conditions).

The rationale for not limiting the estimation of the R factor to the use of one preferred regression based equation lies on the strengths and limitations that the empirical nature of those simplified equations show in different geographic and climatic conditions. As discussed in section 2.5, this motivates the use of multiple empirical equation families with parameterisations covering diverse climate areas (Table 4.6).

The required integration exploited the array structure of the aforementioned quantities (semantic array programming). In particular, the array of regressors ($R_{i,}$ 7 dimensions, Table 4.6) and corresponding validated areas (A_i, 4 dimensions, Table 4.6), as well as the array of covariates (C^j, 26 dimensions, Table 4.5) have been used.

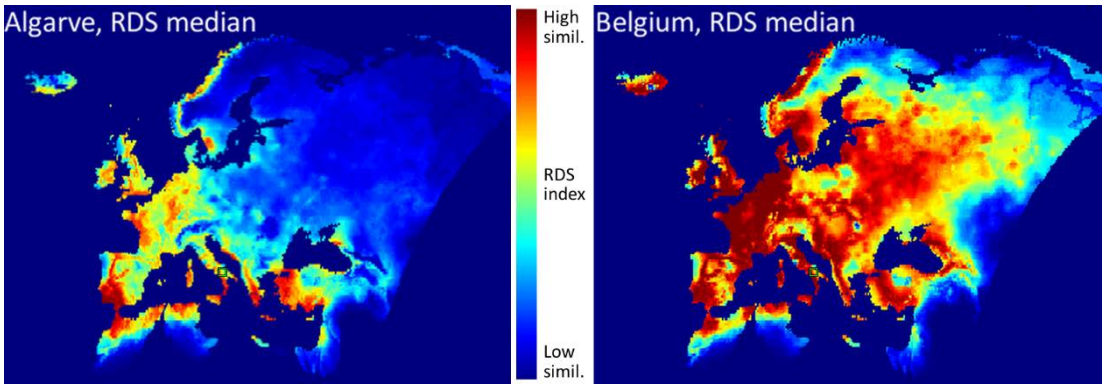


Figure 4.8 - Climatic similarity estimated applying the Relative Distance Similarity (RDS) to the Bollinne equation (Belgium) and to the equation of de Santos Loureiro and de Azevedo Coutinho (Algarve) for rainfall erosivity. The similarity of 26 climatic indicators over the whole Europe is shown (red: maximum similarity; blue: maximum dissimilarity) and aggregated computing the median (Bosco et al., 2011).

The proposed approach considers each estimate R_i as covering an area larger than the study site with a spatially varying degree of reliability (many-to-one approach). This way, for each equation R_i a corresponding map of estimates is computed over the entire climatic extent (feature or covariate space) and transferred in the corresponding spatial extent (geographic space). As a consequence, for each pixel of the spatial extent a multiplicity of estimates becomes available. Depending on the climatic similarity of that pixel with the climate for which the equation R_i was originally designed (i.e. the climate charaterizing the original validated area A_i), a variable reliability is associated to the corresponding estimate. Hence, a weighting of the many estimates is performed on the basis of the local reliability of each estimate, generating a final aggregated estimate. The reliability is based on the Relative Distance Similarity (RDS) algorithm as implemented by the Mastrave modelling library (de Rigo, 2012 a; 2012 b) and is applied for each equation R_i to compare the climatic spatial information of each cell with the corresponding values in A_i. The RDS index has been successfully used in environmental fuzzy ensemble applications (de Rigo et al., 2017, 2016b, 2013a; Bosco et al., 2013). It defines the relative distance between two values C^j_1 and C^j_2 of a given nonnegative covariate. The relative distance is a dimensionless number between 0 (maximum dissimilarity) and 1 (maximum

similarity) and is simply the ratio between the minimum and the maximum value: $min(C^j_1, C^j_2)$ / $max(C^j_1, C^j_2)$. The behavior of each empirical equation outside its definition domain was also assessed to prevent meaningless out-of-range values to degrade the ensemble estimation. Therefore, for both the inputs (covariates) and the output (erosivity estimates) of the regressors R_i the RDS index has been computed and then aggregated cautiously considering the minimum index. This may be defined here as:

$$\begin{aligned} RDS_c^{i,\text{input}} &= \max_{\alpha \in A^i} \left(\Omega_{j=1}^{26} \left(\frac{\max\left(\min\left(C_c^j, C_\alpha^j\right), \delta C^j\right)}{\max\left(\max\left(C_c^j, C_\alpha^j\right), \delta C^j\right)} \right) \right) \\ RDS_c^{i,\text{output}} &= \max_{\alpha \in A^i} \left(\frac{\max\left(\min\left(R_{i,c}, R_{i,\alpha}\right), \delta R_i\right)}{\max\left(\max\left(R_{i,c}, R_{i,\alpha}\right), \delta R_i\right)} \right) \\ RDS_c^{i} &= \min\left(RDS_c^{i,\text{input}}, RDS_c^{i,\text{output}}\right) \end{aligned} \tag{4.16}$$

where δC^j is half of the measurement accuracy of the covariates and δR_i is the half of the tolerance of the erosivity estimates. Ω is a statistical operator with which the relative distances along each dimension of the covariates are aggregated in the RDS index. Among the many possibilities, a simple median has been selected here. The median is also a typical robust statistical operator frequently used for ensemble models.

The weighted median (de Rigo, 2012c) of the 7 empirical models has here been used (using RDS_c^i as weights) for calculating the final R factor map. The values of R factor characterizing the study site have been extracted from this layer.

The weighted median of a vector $R_{1\cdots n} = [\, R_1 \;,\; R_2 \;\cdots R_n \,]$ with integer weights $w_{1\cdots n} = [\, w_1 \,, w_2 \cdots w_n \,]$ is equivalent to the median of the vector $[\, R(1) \diamondsuit w(1) \;;\; R(2) \diamondsuit w(2) \;;\; \ldots \,]$, where the operator $\diamondsuit$ denotes duplications (de Rigo, 2012c, Yin et al., 1996). Besides integer weights, the weighted median may be generalised to consider any nonnegative real weights. The weighted median R^{wmed} of the vector of $R_{1\cdots n}$ with nonnegative weights $\bar{w}_{1\cdots n} = [\, \bar{w}_1 \,, \bar{w}_2 \cdots \bar{w}_n \,]$ is defined as

$$R^{wmed} = \arg\min\left(\bar{w}_{1\cdots n} \circ \operatorname{abs}\left(R_{1\cdots n} - R^{wmed} \right) \right)$$

where ∘ is the operator representing the scalar product between two vectors.

Therefore, in a given spatial cell *c* the weighted median of the vector of erosivity estimates by the 7 empirical models $R_{c,1\cdots7}$ = [$R_{c,1}$, $R_{c,2}$ ⋯ $R_{c,7}$], with weights $RDS_{c,1\cdots7}$ = [RDS^1_c , RDS^2_c ⋯ RDS^7_c] is

$$R_c^{wmed} = \arg\min(\ RDS_{c,1\cdots7} \circ \mathrm{abs}(\ R_{c,1\cdots n} - R_c^{wmed}\ )\ )$$

as computed by the corresponding Mastrave function (de Rigo, 2012c).

Soil erodibility factor

The soil erodibility factor (K) "represents the effects of soil properties and soil profile characteristics on soil loss" (Renard et al., 1997). Soil erodibility is related to the integrated effect of rainfall, runoff, and infiltration on soil erosion. The K factor is affected by many different soil properties (soil texture, permeability, organic matter, etc.) and therefore quantifying the natural susceptibility of a soil is problematic. For this reason, *K* is usually estimated using the soil erodibility nomograph (Wischmeier and Smith, 1978) (Figure 4.9).

The K factor is commonly included in soil erosion models (e.g. PERFECT (Littleboy et al., 1992), AGNPS and USPED) and it is usually determined experimentally using runoff plots. Determining the K factor using the nomograph requires a range of soil properties (soil texture, structure, permeability and percentage of organic matter) but not all of these are available for the Rocchetta Sant'Antonio catchment. A simplified equation was therefore applied. This equation was calibrated using a world-wide dataset of measured K-values (Romkens et al., 1986; Renard et al. 1997) and has already been applied in Bosco et al. (2015):

$$K = 0.0034 + 0.0405 \cdot exp\left(-0.5\left(\frac{logD_g + 1.659}{0.7101}\right)^2\right), \qquad (4.17)$$

where D_g is:

$$D_g = exp\left(\sum f_i \cdot ln\left(\frac{d_i + d_{i-1}}{2}\right)\right), \qquad (4.18)$$

D_g is the geometric mean particle size (mm), for each particle size class (clay, silt and sand), d_i is the maximum diameter (mm), d_{i-1} is the minimum diameter and f_i is the corresponding mass fraction.

The equation for the calculation of the K factor was applied using the percentage (::**proportion**::[5]) of sand, silt and clay present in the sandstone, clay-shale and limestone characterizing this area (Table 4.7). Information on these proportions were derived both form geotechnical tests carried on in this area (see Table 4.8) and literature (Shirazi and Boersma, 1984; Leone and Sommer, 2000).

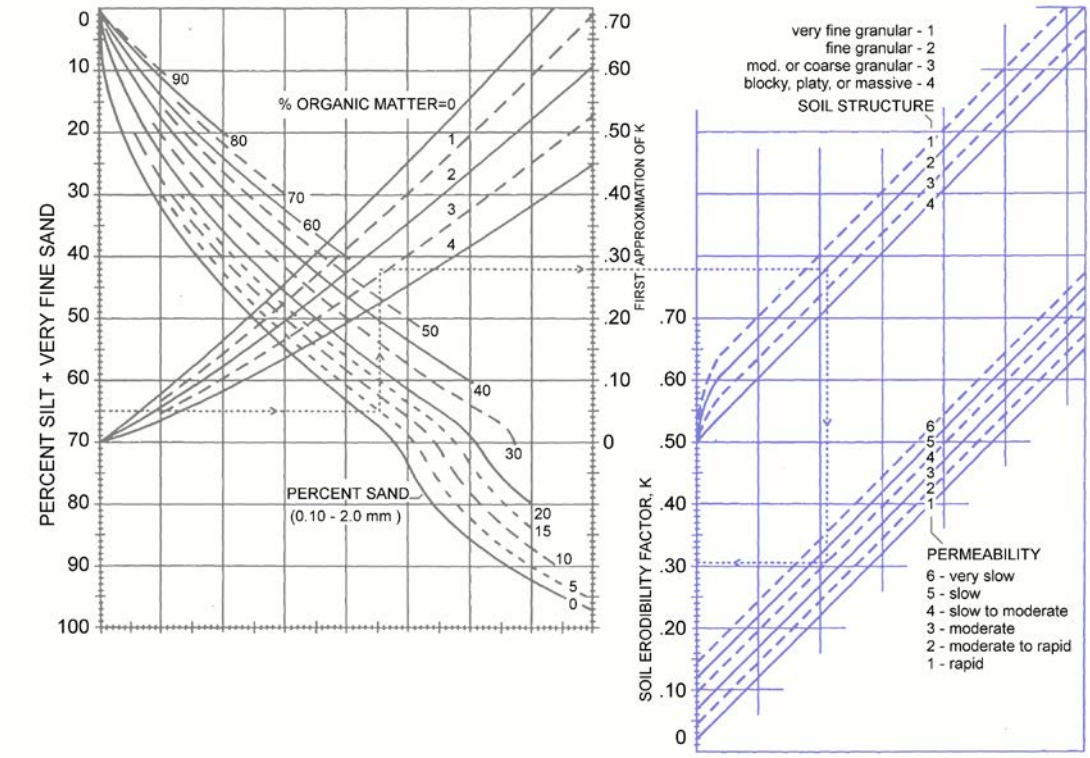


Figure 4.9 – The soil-erodibility nomograph. (source: Wischmeier and Smith, 1978)

Only the effects of rock fragment within the soil profile are considered in the estimation of the K value. Furthermore, the equation applied for calculating the e-RUSLE soil erosion map considers soils with less than 10% of rock fragment by weight (> 2mm) (Renard et al., 1997).

[5] 5http://mastrave.org/doc/mtv_m/check_is#SAP_proportion

Rock fragments have a major effect on soil erosion as they alter soil properties such as water holding capacity, soil erodibility, rooting volume or bulk density, influencing the hydrological response of a soil as well as its degradation and productivity (Poesen et al, 1994a). In estimating the *K* factor only the effects of rock fragment within the soil profile are considered (Renard et al., 1997). The possible presence of rock fragments on the soil surface and within the soil profile require special consideration that led to the introduction of the stoniness correction factor within the model.

Table 4.7 – estimation of the mean percentage of sand, silt and clay present in soils derived by Sandstone, Clay-shales and Limestone that characterize the study site. These values have been selected by analysing data from the literature, geotechnical tests (see examples in figure 4.4 and table 4.8) and personal communications with local experts (CNR-IRPI (Bari)).

	% of sand	**% of silt**	**% of clay**
Clay-shales	10	25	65
Sandstone	58	32	10
Limestone	20	50	30

Table 4.8 – Example of some of the main structured information collected for a typical geotechnical test (results are related to Frana 12) associated to a specific landslide in the Rocchetta Sant'Antonio catchment (see Figure 4.3, 4.4). The parameter *soil depth* is related to the depth of soil from the ground surface to the parent material. The *weight density* is the weight of soil in a given volume. *Specific weight of particles* is the weight density of the solid particles. *Dry weight* is the weight of the soil when the degree of saturation is zero. *Satured weight* is the weight of the soil when the pore are fully filled with water. The *moisture content* of a soil, expressed by volume, is defined as the ratio of the volume of water present in a soil sample to the dry volume of the soil sample. *Porosity* is a measure of the void spaces in the soil, and is calculated as the fraction of the volume of voids over the total volume. *Particle size distribution* represents the relative proportions of soil mineral particles (with major size classes clay, silt, sand and gravel).

Parameters		unit of measure	min and max values
soil depth		m	2.5 - 5
weight density		kN/m^3	19.2 - 22
specific weight of particles		kN/m^3	26 - 27
dry weight		kN/m^3	17.1 – 20.1
satured weight		kN/m^3	20.7 – 22.3
moisture content		%	9.5 – 14.8
saturation			0.5 – 0.8
porosity			0.2 – 0.3
particle size	gravel	%	2.4 – 31.2
	sand		26.6 – 56.2
	silt		23.7 – 30.5
	clay		13.6 – 18.4
definition (AGI)			

Topographic factor

The effect of topography within the model is accounted for by the *L* and *S* factors. Either slope length or slope steepness substantially affect sheet and rill erosion estimated by the model. The LS factor of the RUSLE model ($L \cdot S$), as are the K and R factor, is present within the architecture of many different soil erosion models (AGNPS, PERFECT, MUSLE (Sadeghi et al., 2014; Williams, 1975). L and S factors have been determined through GIS procedures carried out using the Moore and Burch (1986) equation (4.19) and the Nearing's (1997) formula (4.20).

$$LS = \left(\frac{A}{22.13}\right)^{m} \times \left(\frac{sin\,\alpha}{0.0896}\right)^{n} \quad (4.19)$$

$$S = -1.5 + \frac{17}{(1 + e^{(2.3 - 6.1\ sin\,\alpha)})} \quad (4.20)$$

where:

A is the drainage area of a point belonging to a certain cell of the grid, α is the slope ::**angle**::[6] and m and n are parameters. The values of m and n were considered respectively as 0.4 and 1.3. as reported in Bosco et al. (2008).

The formula of Moore and Burch (1986) considers, within the calculation process, the concept of a specific catchment area A accounting for flow convergence and divergence through this term of the equation (Moore et al., 1991). Specific catchment area is one of the most commonly used terrain attributes in hydrological modelling (Erskine et al., 2006). It represents the area that can potentially produce runoff to the location of interest per unit length of contour (Bosco et al., 2015) and it gives to the LS factor stronger physical basis making it suitable for soil erosion modelling.

The approach proposed in Bosco et al. (2015) for calculating the LS factor within the e-RUSLE model, was slightly modified. This was based on the use of the Nearing's (1997) equation for calculating the S factor because of its better performance for steep slopes (up to 50%) (Bosco et al., 2008). However the slope steepness component of the Moore and Burch (1986) formula is more appropriate for slopes lower than 12.73 degrees because it gives the correct limiting value of zero in absence of any steepness (see equations 4.19 and 4.20). A comparison of both formulas is presented in figure 4.10, where a close matching trend is observed between 0 and 12.73 degrees (or 0 - 0.22 rad). Consequently we applied a merged formula obtained by using the Moore and Burch equation for slopes less than 12.73 degrees and then the Nearing formula for higher slopes.

To calculate the slope length factor required in e-RUSLE, the D-infinity (D∞) algorithm of Tarboton (1997) was first used to calculate the flow direction and then the flow length. Due to the geomorphological characteristics of the study area, a multiple-neighbour flow algorithm was required with the D∞ algorithm being one of the most suitable (Gruber and Peckham, 2009; Chirico et al., 2005; Erskine et al., 2006) (see paragraph 4.2.3.3 for further information).

[6] http://mastrave.org/doc/mtv_m/check_is#SAP_angle

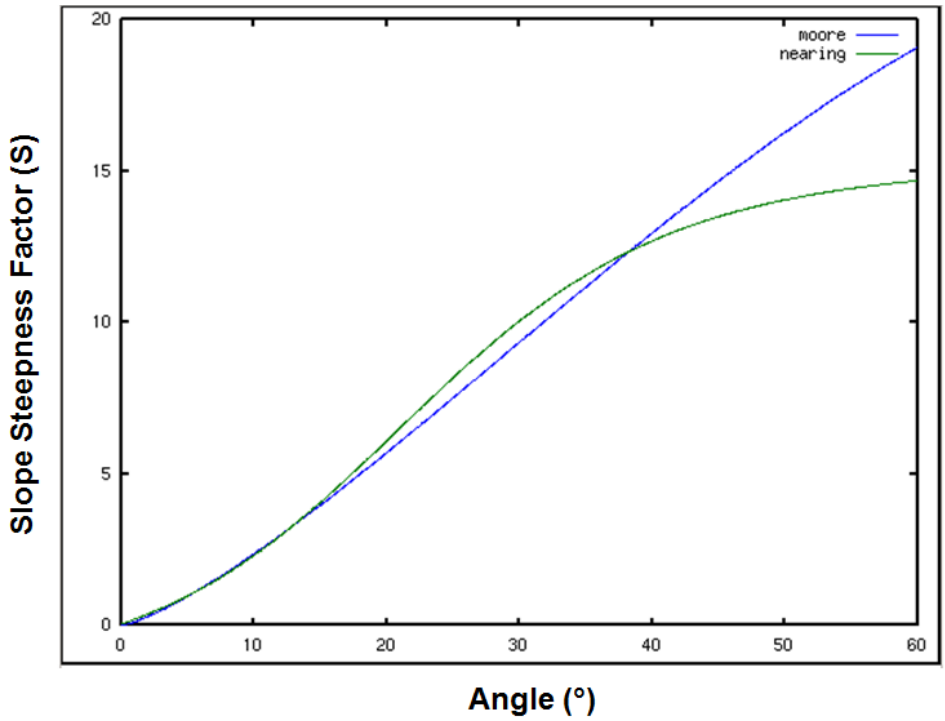


Figure 4.10 - Comparison between the Moore and Burch (1986) relation and the Nearing's (1997) formula applied for calculating the S factor of the e-RUSLE model.

Cover and Management Factor

The cover-management factor represents the influence of land cover, cropping and management practices on soil erosion. A vegetative cover, changing the impact and intensity of rainfall, the resistance to water flow or the sediment transport, can influence soil and water losses (De Ploey, 1982, 1984). As with most of the soil erosion factors within the RUSLE, the C factor is based on the concept of deviation from a standard. As standard, an area under clean-tilled continuous fallow conditions was selected. The C factor incorporates the effects of plants, soil cover, soil biomass, and soil disturbing activities on soil erosion. For calculating the cover-management factor, a subfactor method for computing soil erosion ratios is normally used. The C factor value is an average soil loss ratio weighted according to the distribution of R during the year. The subfactors used for computing these ratio values are: canopy, surface cover, surface roughness, prior land use and antecedent soil moisture. Each subfactor

contains cropping and management variables affecting soil erosion by water and includes one or more variables (e.g. residue cover, canopy cover, canopy height, etc.).

The dependence of the cover-management factor from many different parameters (as already mentioned) makes it difficult to calculate. Because of the difficulties in processing all the parameters due to the lack of data affecting the study site, values from literature were applied for calculating the C factor.

The initial approach was to apply, for all the land cover classes present in the catchment, the worst value from the extensive literature cited in Bosco et al. (2015). Such an approach led to an overestimation of soil erosion phenomenon due to the strong difference that the C factor can present (e.g. the natural grassland present literature values from 0.001 to 0.1). After discussion with local experts, an approach based on our judgment was then used to select the most representative values for the local average conditions. We respectively applied the values of 0.05 for grassland, 0.02 for pastures, 0.0025 for woods and finally 0.3 for the areas covered by crops, such as reported in figure 4.11.

For decreasing the influence of classification errors, to account for within-class variability and for the temporal variation, the use of innovative techniques is the most effective method to obtain wide range information on land cover. Unfortunately, although the studies exploiting remote sensing techniques have reached good achievements they still need improvements (Zhang et al., 2011).

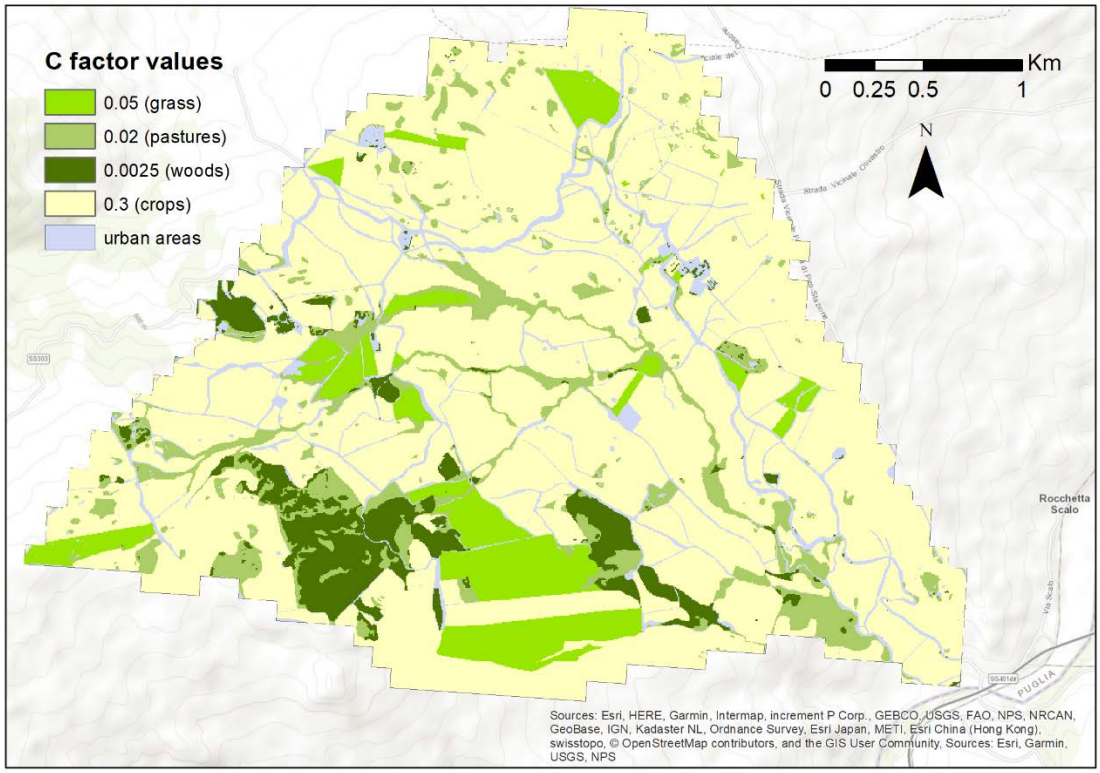


Figure 4.11 – Map of the e-RUSLE C factor in the study area. The values reported in the picture (dimensionless) represent the mean C factor for pastures, grass, broad-leaved forests and cereals. These values are based on data from literature and from information collected during the field survey.

Stoniness Correction factor

Soil stoniness is known to have a strong influence on erosion rates (Poesen et al., 1994). Rock fragments in the soil top layers affect soil water erosion processes in various ways, both directly and indirectly. Over the last years, there was a growing interest in soils containing considerable amounts of rock fragments (Cerdan et al., 2010). These soils are widespread and in particular are present in the Mediterranean where they can occupy more than 60% of the land (Poesen and Lavee, 1994).

The RUSLE model considers stoniness indirectly within the K and the C factor. Regarding the K factor, only the effects of rock fragments within the soil profile are considered. For the C factor stoniness is taken into account in calculating the surface cover sub-factor. Due to difficulties in calculating the C factor for the whole

catchment applying the original equations, the contribution of stoniness was not really considered. To avoid possible overestimation, the application of a new factor for calculating the contribution of stoniness in mitigating soil erosion by water (stoniness correction factor) has been analysed. Poesen and Ingelmo-Sanchez (1992) have given a decreasing relation between rock fragment cover (Rc) and relative interrill sediment yield (s):

$$IR = e^{-b(Rc)}, \qquad (4.21)$$

where b is a coefficient indicating the effectiveness of the rock cover (R_c, ::**proportion**:: ∈ [0; 1]) in reducing interrill soil loss.

They found an experimental value for the coefficient b of 0.02 if the rock fragments are partly embedded in the sealed topsoil, and a value of 0.04 if the fragments are placed on the soil surface. These values are close to those reported by Box, (1981) and Collinet and Valentin (1984) ranging from 0.0256 to 0.058.

Unfortunately detailed information on soil stoniness is not available for the catchment. The only available information is the volumetric rock fragment content of the soils contained in the ESGDB database. The volumetric content percentage of rock fragments in the top soil and the cover percentage of rock fragments at the soil surface are two different parameters (Poesen and Lavee, 1994). As a first aproximation and due to the limited available data on soil stoniness, we assumed the rock fragment cover equals the volumetric rock fragment content as suggested in (Govers et al., 2006).

Due to the scale of the ESGDB dataset, this catchment is covered by a single volumetric rock fragment content that corresponds to a very low percentage of rock fragments in the soil. Despite the use of a single correction factor for the whole catchment is not corresponding to reality, the observation conducted during the field survey confirms that in general the soil is characterized by a low or very low content of rock fragments.

Human Practices factor

P is the support or land management practice factor. By definition the P factor is the ratio of soil loss with a specific support practice to the corresponding loss with upslope and downslope tillage (Renard et al., 1997). It represents how surface and management practices like terracing, stripcropping or contouring affect erosion phenomenon. For areas where there is not support practices or without any data, the *P* factor is set equal to 1.0. Within the study catchment we have limited information on land management practices. These were mainly collected during the field survey and are limited to ploughing practices applied in a small number of fields to contrast soil erosion, therefore it was decided to consider the P factor equal to 1 everywhere.

4.2.3.3. Geomorphometrical considerations for calculating the LS factor

Considering that errors in slope computation could be exaggerated in soil erosion models because of the exponential relation between slope and soil erosion (Warren et al., 2004), variation in land surface computation can result in significantly different values. Prior to utilizing the available DEM for evaluating land-surface parameters, it is necessary to prepare the DEM. Some of the most common errors present within a DEM are due to local outliers, padi terraces (areas having pixels shown the same value typical of closed contours) or sinks (an erratic feature not corresponding to the actual feature of the terrain). Numerous statistical approaches or tools are available for correcting these errors (Hengl et al.,2004), for filling the sinks present in the DEM we exploited the tool available within ArcGis.

Very important is also the selection of the more appropriate flow algorithm. One of the most applied flow algorithms is the 'D8' (O'Callaghan and Mark, 1984) (Figure 4.12), a single neighbour flow algorithm. The main limit of the D8 algorithm is that it can model flow convergence but not divergence and in case of uncertainty this assignment is arbitrary. For cells far enough downstream to be in fully convergent channelized portion of the area the results are very good. In case of hillslope or near

peaks the values can present errors which can be of an order of magnitude (Gruber and Peckham, 2009). However using a multiple-neighbour flow algorithm it is possible to consider the effect of divergent flow (Gruber and Peckham, 2009).

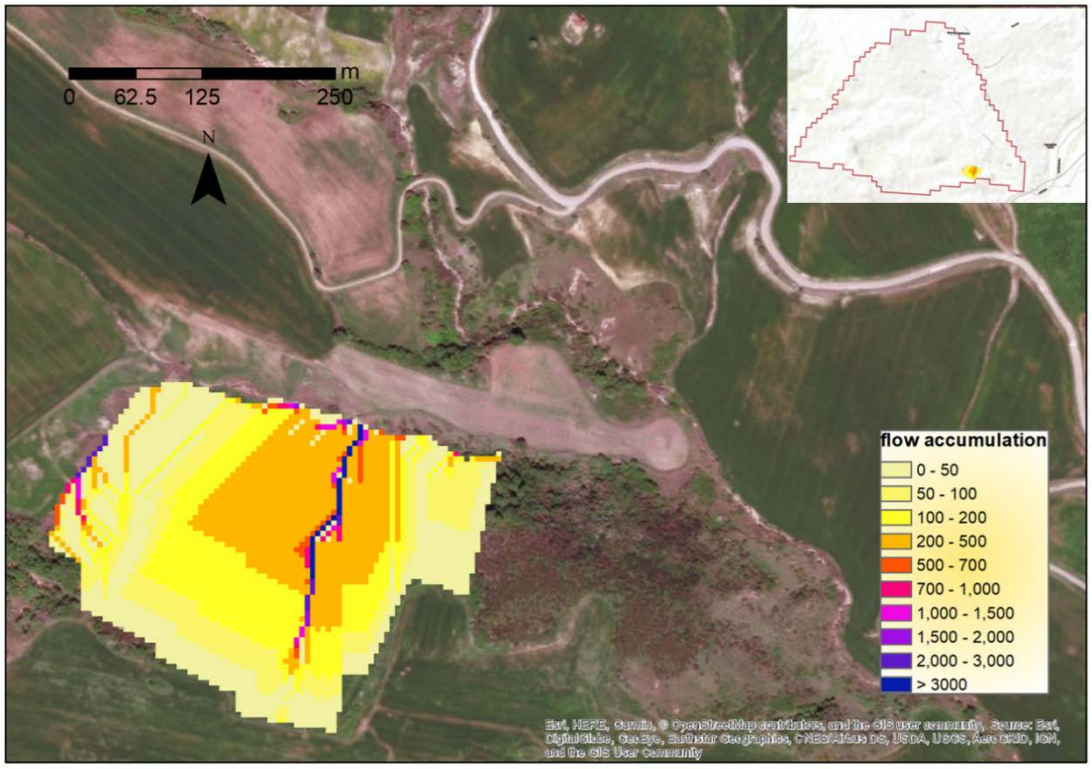


Figure 4.12 - Application of the D8 algorithm to calculate the flow accumulation in a little plot located on a steep slope within the study area.

Other numerous algorithms also exist for handling not convergent flow as the TOPMODEL approach (Quinn et al., 1991) or the algorithms of Freeman (1991, in Gruber and Peckham, 2009) and Holmgreen (1994; in Gruber and Peckham, 2009). The main limitation of these algorithms is that they can produce over-dispersion (Costa Cabral and Burges, 1994; Tarboton, 1997). Three further algorithms has been proposed to overcome this limit, D^{∞} (Tarboton, 1997), DEMON (Costa-Cabral and Burges, 1994) and the Mass-Flux Method (MFM) algorithm of Peckham (Gruber and Peckham, 2009).

Due to the importance that the extraction of parameters from a DEM has for the implementation of a model in a GIS environment, it was necessary to analyse this

aspect with the aim of reducing any error in the geomorphometric analysis. Applying the D8 algorithm for the calculation of the flow direction (Figure 4.12), it was immediately evident that this way was not suitable for our objectives. Using an artificial DEM (10x6) as a test, the calculated flow direction layer has shown its limit in representing not fully convergent situations. All flow is passed to the neighbour with the steepest downslope gradient (resulting in 8 possible drainage directions). It can model convergence (several cells draining into one), but not divergence (one cell draining into several cells). Ambiguous flow directions (as in Figure 4.13c where the same minimum downslope gradient is found in two cells) are resolved with an arbitrary assignment.

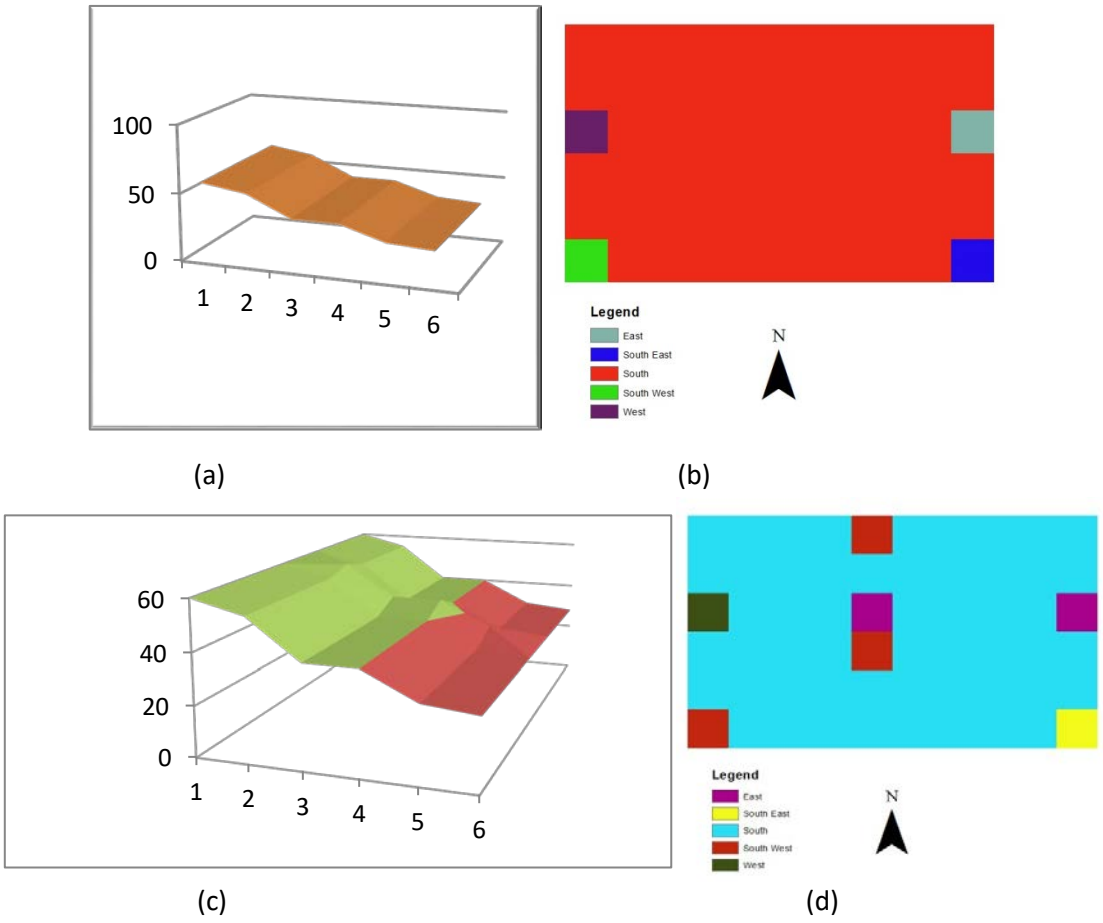


Figure 4.13 – Slope (a, c) and Flow direction (b, d) for a 10x6 artificial grid using D8 algorithm implemented in ArcGis. Figure (d) shows as using the D8 algorithm in divergent condition there is an arbitrary assignment of the flow direction. The east/west or south west/east assignement to the cell on the ridge is totally aribitrary.

The D8 algorithm provides a good estimate of the catchment area for grid cells that are far enough downstream to be in the fully convergent, channelised portion of the landscape. Only multiple-neighbour flow methods can accommodate the effects of divergent flow (from one cell to several downhill cells) and this is especially important on hill slopes. The geomorphological characteristics of the study area (mainly hill slopes) indeed require a multiple-neighbour flow algorithm for the calculation of the flow, with D∞, DEMON and MFM being the most suitable algorithms. The D∞ algorithm of Tarboton (1997) was selected to calculate the flow accumulation within the catchment (Figure 4.14) because DEMON is not available in the language required and its implementation would be difficult, and D∞ was succesfully applied in many different fields (Hamel et al., 2017; Regalado and Kelting, 2015; Lucieer et al., 2014).

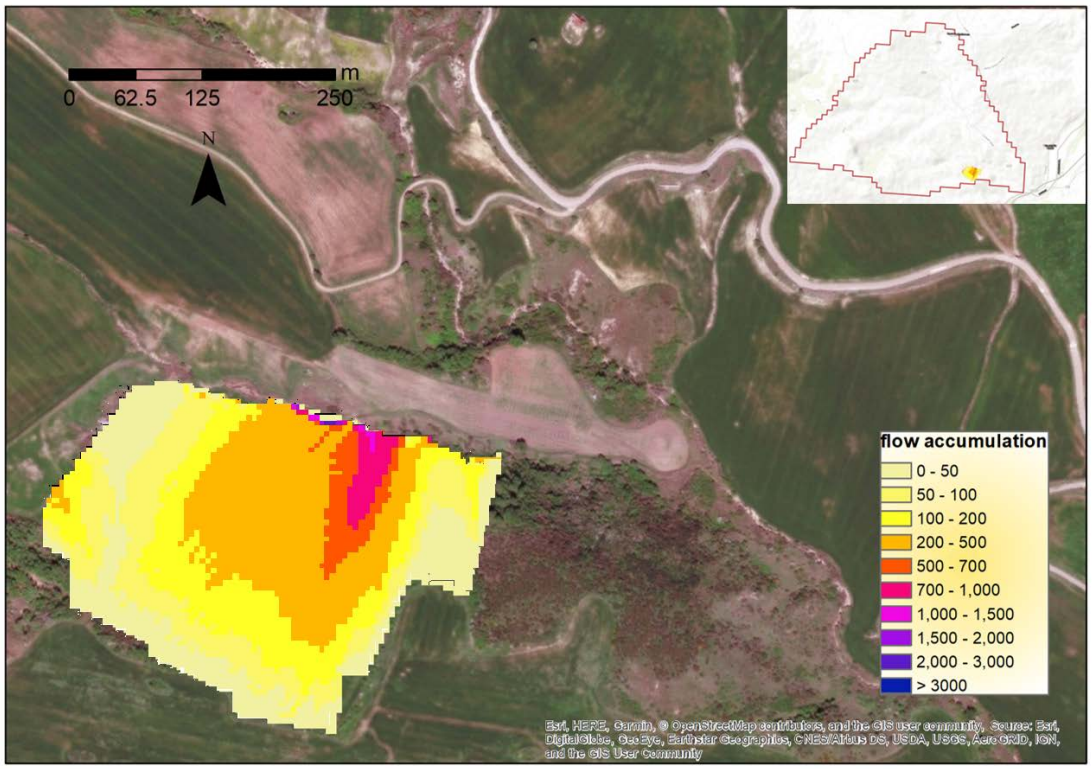


Figure 4.14 – D∞ algorithm applied in a small plot within the study site to calculate the flow accumulation. Comparing this picture with figure 4.12 it is possible to see how the D∞ algorithm consider the flow along the slope as less concentrated.

4.3. Results and discussion

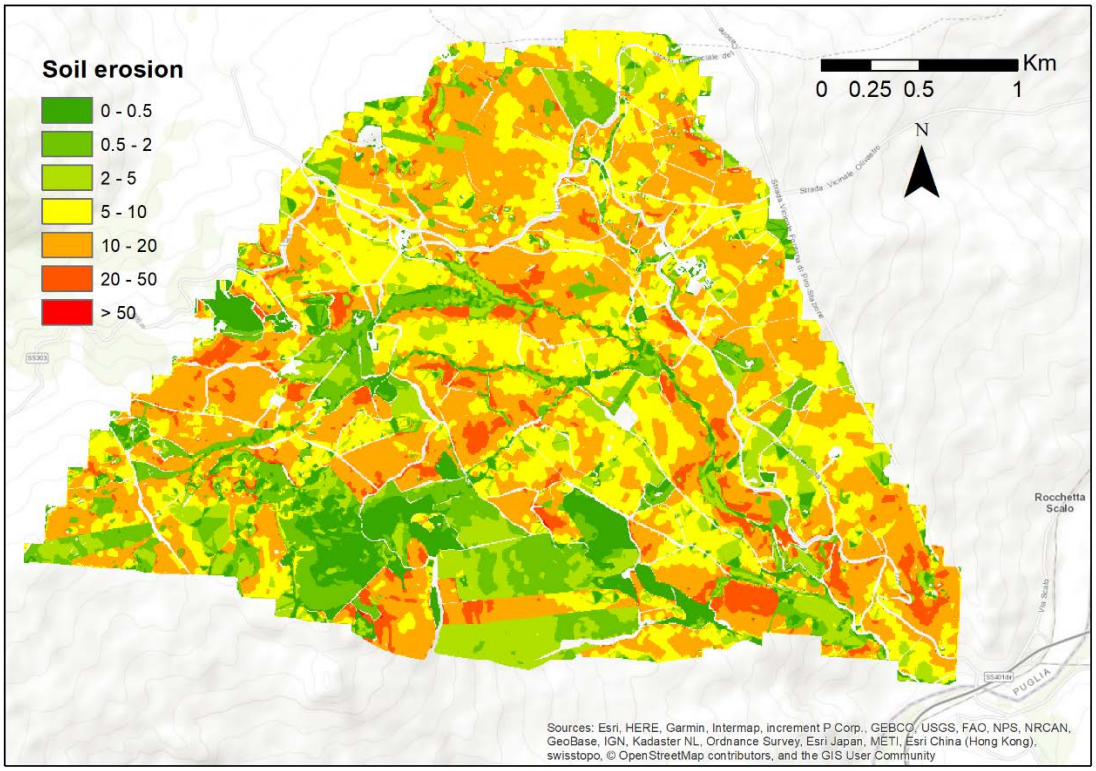


Figure 4.15 – Map of the soil erosion by water in the study area as calculated by applying the e-RUSLE model. Soil erosion is here reported in tonnes per hectare per year.

The resulting soil erosion map is shown in figure 4.15. The well-known role of natural vegetation in mitigating soil erosion (Cerdan et al., 2010; de Rigo and Bosco, 2011; Maetens et al., 2012) may be observed by comparing the map with the land cover map of the catchment (Figure 3.8). Areas covered by grassland and woods present, as expected, a considerably lower soil erosion rate than cultivated areas.

Since some essential factors in the e-RUSLE (C, K factors and stoniness) are derived from categorical information, the uncertainty associated with the corresponding classification may be propagated in the final erosion map (Figure 4.15).

For improving the estimation of the C factor, that is still a weakness within the model, it may be necessary to develop new techniques or equations improving the collaboration between soil erosion scientists and remote sensing experts. Although good relationships were obtained since the 80s between the C factor and band ratios

of Near Infrared Reflectance (NIR) to red reflection (Cihlar, 1987; Stephens and Cihlar, 1982) and the studies of the C factor estimation using remote sensing techniques have reached good achievements, they still need improvement (Zhang et al., 2011). Further analysis with detailed forest types and tree species distribution maps seem to be necessary for increasing the accuracy of the C factor (de Rigo and Bosco, 2011; Geißler et al., 2012).

The RUSLE model generally tends to overestimate soil loss, probably because it does not account for deposition in local depressions (De Jong et al., 1986).
Because of the weakness of data for some of the model's parameters and the use of coarse spatial data (e.g. E-OBS and ESGDB) along with data having sub-optimal resolution (e.g. SRTM), the application of the model can lead to noticeably uncertain soil erosion rates in certain areas. For example, the lack of appropriate datasets for soil stoniness could locally lead to an over- or under-estimation of the erosion rate. However, the precise delimitation of such issues is very difficult as field investigations for validation are required.

Because the main aim of this PhD thesis is to better integrate and quantify the role of landslides within soil erosion processes, the lack of an accurate quantitative prediction of the local soil erosion rate does not affect the scientific result of this work. A slight over- or under-estimation of soil erosion does not affect our effort in measuring the relative changes in soil erosion due to mass movements occurrence. However, even if the results provide an overview of the soil erosion susceptibility in the landscape rather than an accurate quantitative estimation for a specific location, they can be considered robust enough for our scope.

The classification scheme used for measuring the soil erosion rates (Figure 4.15) is based on the one applied in Bosco et al. (2015). The thresholds above which soil erosion should be regarded as a serious problem is controversial, the soil formation processes and rates can substantially differ in different areas (see paragraph 2.1). By analysing the high resolution soil erosion map we produced (Figure 4.15), more than 60% of the Rocchetta Sant'Antonio catchment is affected by significant soil erosion (over 5 t ha^{-1} yr^{-1}, moderate – high level). The numerous rills and ephemeral gullies

present in this area are tangible evidence of this process. This also demonstrates the importance of maintaining permanent vegetation cover as a mechanism to combat soil erosion, were possible, and to adopt appropriate crop managing practices. As already mentioned, there is a probability for some of the model results to be over-estimated. The R and C factor uncertainty and the possible presence of areas having a stoniness values much higher than the value indicated by the underlying soil database, could be at the basis of many of the possible uncertain estimations of sheet and rill erosion rates. Another limit of the proposed approach is that the model does not consider erosion processes such as channel or gully erosion, that locally may cause very high soil losses (Poesen et al., 2003; Mathys et al., 2003; Collinet and Zante, 2005), or tillage erosion that in this area may have a similar rate as soil erosion by water (van Oost et al., 2009). Anyway, the proposed architecture is designed not to be data demanding while still being able to scale up to the continental scale, such as presented in Bosco et al. (2015).

The common validation procedures were not technically and financially applicable for the present work. Nonetheless, some validation options are still applicable. To validate the map of soil erosion by water (figure 4.15) we applied a qualitative approach based on visual interpretation applying the same methodology used by Bosco et al. (2015, 2014). This methodology is based on a visual and categorical comparison between modelled and observed erosion rates (Bosco et al., 2014). A procedure employing high-resolution Google Earth (Google Earth. Mountain View, CA: Google Inc.) images and pictures as data for a plausibility check was applied.

The good resolution of Google Earth images jointly with a set of pictures collected during the field survey, allow for a visual qualitative estimation of soil erosion phenomena. By overlaying the map of soil erosion and the selected validation points (that corresponds to the same points were the field survey was carried out), a visual plausibility check, inspired on the erosion/deposition categories for field validation of Warren et al. (2005), and also of Berry et al. (2003) and Kapalanga (2008), was carried out. A buffer of 25x25 metres around the selected points (points are shown

in figure 4.5) was analysed, with over 700 cells at 5 m x 5 m. For each cell, a visual assessment relied on high-resolution images.

As a result of the validation exercise, the soil erosion rates we estimated, despite the above mentioned limits, seems to be reliable enough for helping to identify areas in the Rocchetta sant'Antonio catchment where to concentrate the effort for preventing soil degradation.

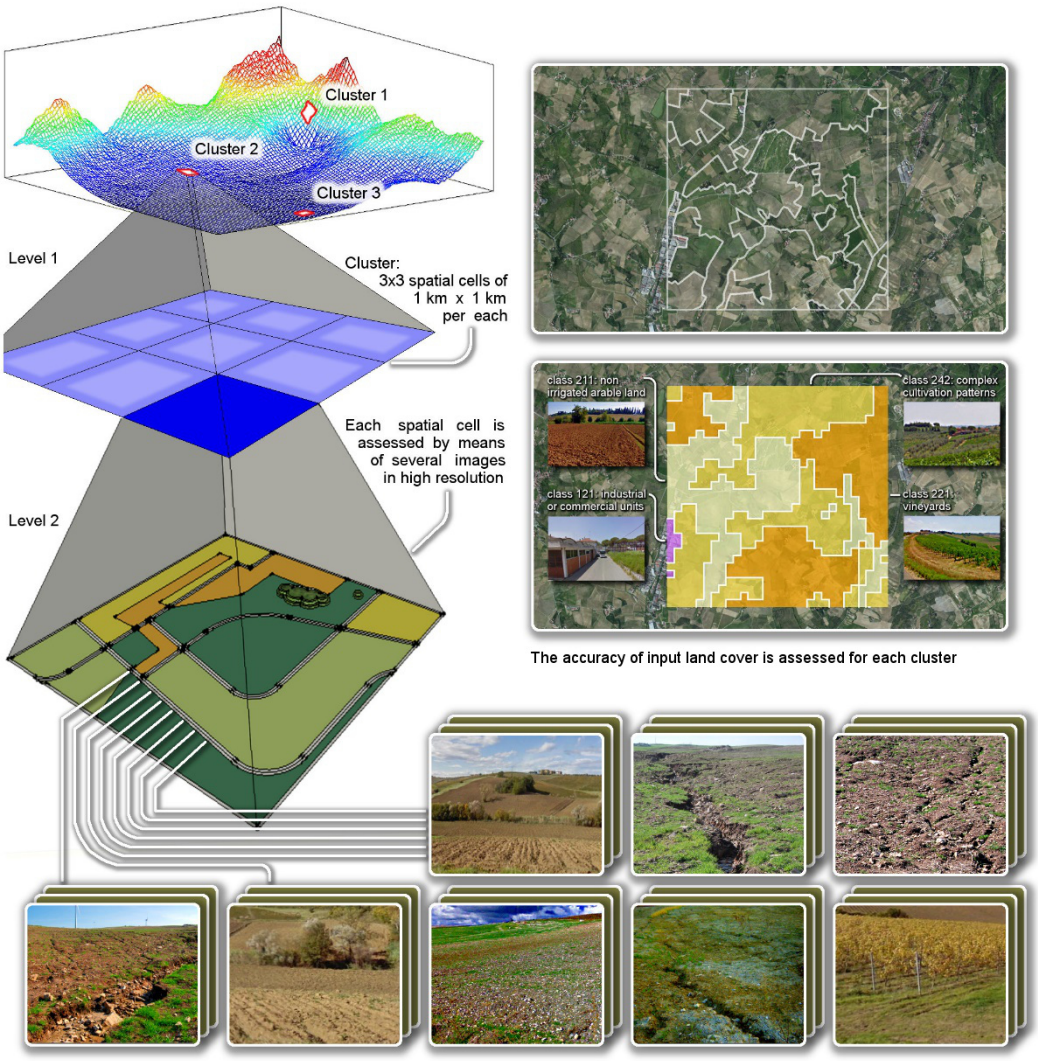


Figure 4.16 – Example of the plausibility check such as performed in Bosco et al. (2015). The check is over grid cells (Level 1) by integrating Google Earth and Bing higher-resolution information also including high resolution images (Google street view and georeferenced crowd-sourced pictures) (Level 2).
Image from Bosco et al. (2015); Bing Maps, © 2013 Microsoft Corporation; Google Street View, © 2013 Google Inc., Mountain View, CA.

4.4. Conclusions

An estimation of soil erosion using a modified version of the RUSLE model has been carried out by merging existing empirical rainfall-erosivity equations within a climatic ensemble model based on the relative-distance similarity and by adding a new factor for better considering soil stoniness. The lack of high-resolution datasets to calculate rainfall erosivity and stoniness and the limitations inherited from the RUSLE architecture lead to a considerable level of uncertainty.

Some of the individual factors could also be interdependent, which results in an even greater impact on the model results (van der Knijff et al., 1999). As a consequence, quantitative assessment using the model should not be undertaken without the right awareness.

The provided estimates cautiously model the erosion rates in the absence of mitigating management practices – which, in an agricultural area as the Rocchetta sant'Antonio catchment, should be regarded as a main factor for limiting the impact of erosion. It is necessary to have in mind that the main objective of the present thesis is not the production of a new accurate soil erosion map of this area but to contribute to soil erosion research better integrating and quantifying the effect of landslides in soil erosion modelling for improving soil erosion estimation at local and regional scale.

Overall, our spatially distributed assessment of soil erosion, carried out using the e-RUSLE model, even considering all the limits of our approach, can help in identifying areas within this catchment where to concentrate efforts for preventing soil degradation.

Improvement in these erosion estimates lies in better climate and soil data potentially available from national archives. Land cover requires frequent updating, because changes in land use have a major impact on erosion rates. There is the potential to do this through the analysis of remotely sensed images.

5. Modelling shallow landslides triggered by water. A multi-scale robust modelling approach for estimating landslide susceptibility

Estimating the landslide susceptibility of a territory may be supported by many different analytical approaches: heuristic, deterministic and statistical (van Westen et al., 1997). Statistical landslide susceptibility methods are based on establishing relationships among measurable variables whose combination is empirically found to correlate with observed landslide occurrences. Theoretical insights guide the selection and analysis of landslide triggering factors (geological, hydrogeological and geomorphological) to determine the most appropriate input - the best suited set of predictors or covariates (Brenning, 2005) - to use for statistically reconstructing landslide susceptibility.

In deterministic approaches, the landslide susceptibility is evaluated by carrying out a slope stability analysis. It results, for example, in the calculation of a factor of safety distribution across the study area. The deterministic approaches should be able, in theory, to provide more reliable results (especially where no field measurements are available on landslides and stable areas) but require detailed datasets describing conditioning and triggering factors. Many different models, which are usually composed of coupled dynamic hydrological and slope stability models, have been developed by several authors (van Beek, 2002; Chen and Lee, 2003; van Westen, 2004) (see section 2.2.2.2).

5.1. Susceptibility forecast

Within the study area, precipitation is the main triggering factor for landslide occurrence (Wasowski et al., 2010). In order to improve the spatial prediction of landslides where water is the triggering factor, a combined total of five different deterministic and statistical models have been applied. In order to enhance the determination of landslide susceptibility, a new method based on an ensemble

approach has been used for aggregating the modelling results. The ensemble approach is a reproducible D-TM (see equation 5.1, and section 2.5) applied to the results of the array of models and is based on relative-distance similarity (RDS). The application of an ensemble approach, especially in data poor regions, could potentially reduce the uncertainty and mitigate local poor performance associated with individual models, by excluding outlier estimations.

Uncertainty may affect these models from the inaccuracy of required input data layers X and parameters θ to the approximation of their reconstruction (e.g. by means of other specialised D-TMs) when not directly accessible as available datasets. Uncertainty may also be exacerbated by modelling simplifications or overcomplication. In the latter case, a perhaps theoretically accurate approach might sometimes result in a poorly performing D-TM implementation due to site-specific information gaps and possible impacts of site complexity where multiple conceptual mechanisms coexist as landslide drivers. Furthermore, nontrivial computational models may be affected by software uncertainty (de Rigo, 2013; de Rigo et al., 2013b) (see equation 5.1), namely the distance from the theoretical mathematical formulation and the actual model implementation in one or more artificial programming languages.

$$
\begin{aligned}
& Y = f^{*}(X) = f(\theta^{*}, X) && \text{Theoretic D-TM: instance of a model family } f \text{ by means of selected parameters } \theta^{*} \\
& Y = f^{\zeta}(\theta^{*,\zeta_\theta}, X^{\zeta_X}) && \text{Real D-TM with uncertainty affecting input data, parameter calibration and the model itself (software uncertainty)} \\
& Y = \bigcup_{i=1}^{n} :: \left| f_i^{\zeta}(\theta^{*,\zeta_\theta}, X^{\zeta_X}) \right|_{::}^{sem_i} && \text{Mitigating uncertainty by ensembling an array of } i \in \{1, \cdots n\} \text{ heterogeneous models (design diversity) with semantic checks } sem_i
\end{aligned}
\tag{5.1}
$$

Silent faults (Hook and Kelly, 2009) are a class of software errors that can alter computational output without any evident symptom (such as, for example, premature interruption or unrealistic results). Because of these silent faults, “many

scientific results are corrupted […] by undiscovered mistakes in the software used to calculate and present those results" (Hatton, 2007).

In order for the uncertainty to be mitigated, a robust fuzzy ensemble model is proposed to aggregate an array of different susceptibility zonation maps. Each susceptibility zonation has been obtained by applying heterogeneous models (physically based and statistical methods), to increase design diversity (de Rigo, 2013). The technique is designed to scale to different arrays of models. Each model is adapted to fit the ensemble array by wrapping its interface to behave as a semantically enhanced module. In the computational science domain, a wrapper unit is a computational module which modifies the input or the output of a pre-exisiting module. The pre-exisiting module is often unable to perform satisfactory checks concerning the semantic consistency of its set of input and output arguments.

The wrapper module is designed to expand the behaviour of the pre-existing module by adding the missing semantics so that the input arguments are appropriately verified before they are passed to the module and for the output arguments to be verified by the semantic wrapper module after the computing of the pre-exisiting module, and before the output is passed to other modules. Semantic checks further mitigate inconsitencies between input data, parameters and outputs, following the paradigm of semantic array programming (de Rigo et al., 2013b, de Rigo, 2012a, 2012b) (see section 4.2.3.2) (see equation 5.1).

As discussed in section 2.5, SemAP complements the compactness of array programming notation with an effort towards the most concise generalisation of autonomous tasks as modules which are subject to array-based semantic checks. Each model is considered as a semantically-enhanced module of the ensemble. As described in chapter 2, a few straightforward semantic checks sem_i are exemplified in the following with the notation ::**sem**:: with link to the corresponding online description.

Exploiting the availability of landslide maps and the environmental information (DEM, land cover and lithological map, geomorphology) two deterministic models based on the infinite slope equation (SINMAP (Pack et al., 1998, 2005) and a simple slope stability model derived by van Beek's PROBSTAB model (van Beek, 2002)), as

well as three statistical models based on Artificial Neural Network (ANN), Logistic Regression (LR) and RDS, were calibrated and validated.

5.1.1. The deterministic approach

The evaluation of slope stability conditions in a landscape can be performed by considering the local equilibria (FS) of forces along pre-determined, shallow slip surfaces representative of translational slide mechanisms. As previously mentioned, two deterministic models based on the infinite slope equation (Stability INdex MAPping (SINMAP) and a simple slope stability model hereafter named TransSlide) were applied.

The SINMAP model

The theoretical basis of SINMAP involves a mechanistic infinite slope stability model (e.g. Hammond et al., 1992; Montgomery and Dietrich, 1994) linked with a topographically based steady-state hydrology model. The SINMAP approach is similar to that of Montgomery and Dietrich (1994), both combine the infinite slope stability model with the steady-state hydrologic concepts. The slope stability model balances the destabilizing components of gravity and the restoring components of friction and cohesion on a failure plane parallel to the ground surface with edge effects neglected (Pack et al., 2005). The theory at the basis of the model applies to translational slides where fluctuating pore pressures form the dominant trigger factor. The landslide susceptibility distribution is governed within the model by calculating slope and specific catchment areas starting from a DEM. The model parameters are allowed to be flexible following a uniform distribution between an upper and lower limit. The parameters may be calibrated using geographic calibration regions based on lithological, land cover or soil characteristics (Pack et al, 1998, 2005).

SINMAP is based on the infinite-slope form of the Mohr-Coulomb failure law (Hammond et al., 1992; Pack et al, 1998):

$$FS = \frac{c_s + c_r + (\gamma_s D - \gamma_w D_w)cos^2\beta \, tan\,\emptyset}{\gamma_s D \, sin\,\beta \, cos\,\beta} \,, \tag{5.2}$$

where c_s is the soil cohesion (kPa), c_r is the root cohesion (kPa), γ_s is the soil unit weight (kN/m^3), γ_w is the water unit weight (kN/m^3), D is the vertical soil depth (m), D_w is the vertical water depth (m), β is slope angle (°) and ϕ is the internal friction angle (°).

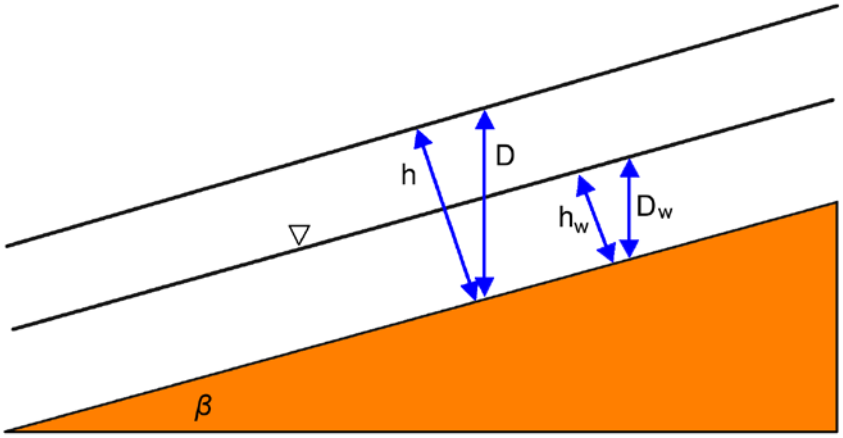


Figure 5.1 – SINMAP Infinite slope stability scheme. S is the slope expressed as a decimal drop per unit horizontal distance (source: Pack et al., 2005).

Within the SINMAP model the soil thickness is interpret as perpendicular to the slope profile (Figure 5.1). Introducing the variables:

$$C' = \frac{c_s + c_r}{h\gamma_s} \,, \qquad r = \frac{\gamma_w}{\gamma_s} \,, \qquad w = \frac{D_w}{D} = \frac{h_w}{h} \,, \tag{5.3}$$

where h=$D\,cos\beta$ is the soil thickness, C' is the combined cohesion, w is the relative wetness and r is the water to soil density ratio, equation (5.2) can be written as:

$$FS = \frac{C' + cos\,\beta\,(1 - wr)\,tan\,\emptyset}{sin\,\beta} \,, \tag{5.4}$$

Equation (5.4) is the dimensionless form of the infinite slope stability model. This is convenient because it directly combines cohesion with the soil density and thickness into a dimensionless cohesion factor, C'(equation 5.3). This may be thought of as the ratio of the cohesive strength relative to the weight of the soil, or the relative contribution to slope stability of the cohesive forces (Pack et al, 2005) (Figure 5.2).

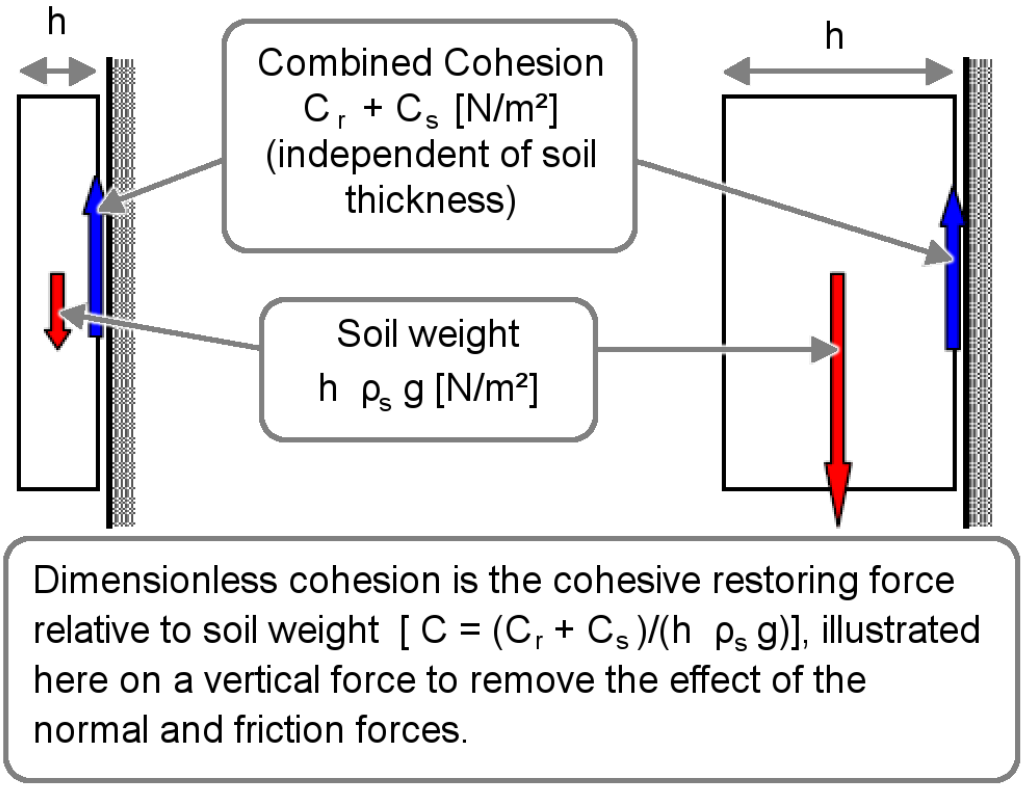


Figure 5.2 - Illustration of dimensionless cohesion factor concept, where $\rho_s g = \gamma_s$ (source: Pack et al, 2005).

Practically, the model works by computing slope and wetness at each grid point, but assuming other parameters are constant (or have constant probability distributions) over larger areas. With the form of equation (5.4) this amounts to implicitly assuming that the soil thickness (perpendicular to the slope) is constant.

Adopting a modified version of the TOPMODEL approach (Beven and Kirkby, 1979), the relative wetness (w) can be written as:

$$w = min\left(\frac{R'a}{T\,sin\,\beta}, 1\right), \tag{5.5}$$

where T is the soil transmissivity (m^2/h), R' is the steady state recharge that is an estimation of the lateral discharge (m/h), a is the upslope drained area per unit contour length (m^2/m).

One of the assumptions SINMAP is based on is that the capacity for a lateral flux at each point is $Tsin\,\beta$. This assumption differs from the TOPMODEL of Beven and Kirkby (1979) because the hydraulic conductivity is not assumed to decrease with depth. Here a uniform conductivity of a soil mantle overlying relatively impermeable bedrock is assumed. In addition, sin β is used rather than tan β because the flow distance is along the slope (Pack et al., 2005).

For implementing the stability index in SINMAP, the wetness index from equation (5.5) is incorporated into equation (5.4), which becomes:

$$FS = \frac{C' + cos\,\beta\,[1 - min\left(\frac{R'}{T}\,\frac{a}{sin\,\beta}, 1\right)r]\,tan\,\emptyset}{sin\,\beta}\ , \tag{5.6}$$

For areas where the minimum factor of safety is less than 1, then there is a possibility (probability) of failure.

The TransSlide model

TransSlide (Bosco et al., 2013) is based on a translational slope stability function calculating the factor of safety at a potential shear plane based on variations of the groundwater level and volumetric moisture content. TranSlide's basis is also in the equation (2.2) where, as already mentioned in chapter 2:

$$S = (c_s + c_r) + (\sigma - u)\,tan\,\emptyset \qquad \tau = \gamma_s D\,cos\,\beta\,sin\,\beta \tag{5.7}$$

and σ is calculated following equation (2.5).

To determine the pore pressure that is necessary for calculating shear strength the following equation was used (van Beek, 2002):

$$u = \gamma_w D_w \cos^2\beta = rw\sigma , \quad (5.8)$$

where r and w are calculated as in the SINMAP model. The static inputs include soil shear strength (cohesion and friction) and an additional root cohesion function representing land use. The model calculates the local instability on the balance between resisting and driving forces. The model was implemented in a raster based environment and calculates the local instability on the balance between resisting and driving forces. Due to the paucity of available data in the study area, the full functionality of TransSlide could not be mobilised. It affected the reliability of the modelling outputs, that results in underestimating the areas susceptible to landslide within the catchment (see text and figures in section 5.5)

5.1.2. The statistical approach

In statistical landslide susceptibility methods, semi-automated computational methods may benefit from meaningful interpretation which domain experts can supplement, for example on the relative contribution and emerging limitations of different semi-automated methods in estimating slope failure and classifying areas as having different hazard or susceptibility degree (Aleotti and Chowdhury, 1999; Suzen and Doyuran, 2004; Acharya, 2011). Multivariate statistical methods are common methodologies for these analyses, which are based upon the presence or not of stability phenomena within the classified areas (van Westen, 2000). A key problem using this approach is the high sensitivity of the results to the input data and the difficulty in deriving the probability of occurrence from the susceptibility (Acharya, 2011). In this work, three statistical models based on ANN, LR and RDS were calibrated and validated for susceptibility.

Artificial Neural Networks

An ANN is a data-transformation model able to derive from a set of input data a corresponding set of outputs. Neural networks resemble a human brain because of the acquiring of knowledge through learning and storing the acquired knowledge within inter-neuron connection strengths (synaptic weights) (Figure. 5.3, 5.4).

A peculiarity of ANNs is that the number of parameters (also known as weights of the ANN) may be limited to grow linearly with the number of input dimensions. The same applies to the number of output dimensions. An ANN is implemented by a system of interconnected nodes. Information propagates through nodes transforming the inputs in intermediate derived signals up to generate the final outputs. The internal nodes are called neurons and define the ANN hidden layers. Each node is a processing element propagating weighted inputs received from other nodes (Pradhan and Lee, 2009) (Figure 5.4).

Depending on the specific ANN architecture, the inputs of a given node may include or exclusively be constituted by intermediate derived signals. The learning process comes from adjusting the weights between neurons analysing the error between the predicted and target output. The output of a neural network, after the training, is a model that starting from an input dataset is capable of predicting a target value (Lee et al., 2007).

The power and main advantage of using ANN lies in their capacity to model both linear and non-linear relationships and to learn these relations directly from the data. Because many complex problems are characterized by having a non-linear behaviour, traditional linear models are often inadequate.

Most papers on the use of ANNs apply a multilayer feed-forward network (Maimon and Rokach, 2005). A feed-forward neural network is an ANN having connections between the different units not forming a cycle or loop. In this architecture information moves in only one direction, from the input nodes, through the hidden layers (if any) to the output nodes. The main reason for the use of this type of ANN is the simplicity of its theory, ease of programming and good results (Figure 5.5 shows the scheme of this kind of ANN).

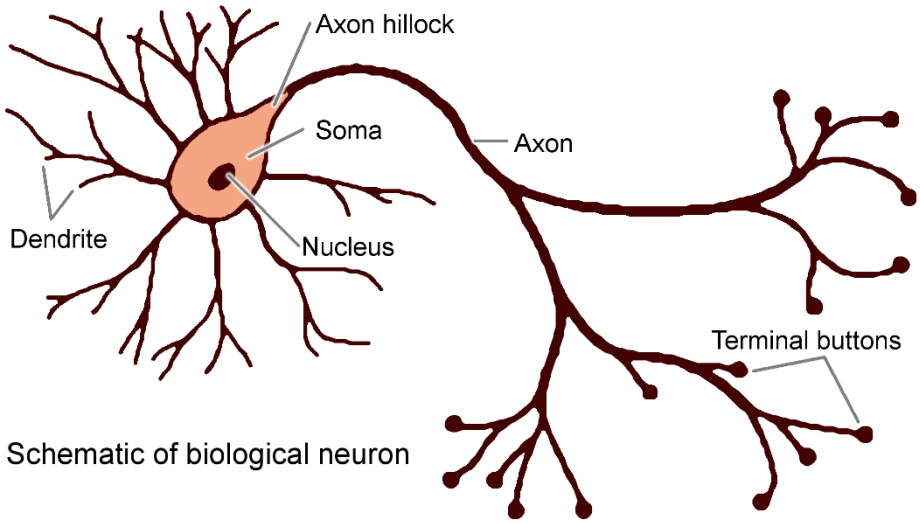


Figure 5.3 – The biological neuron. A typical nerve cell conists of four parts: dendrites (accept inputs), Soma (process the inputs), Axon (turns processed inputs into outputs), Synapses (the electrochemical contacts between neurons). (Derived after: http://rslab.movsom.com/paper/somrs/html/chapter3.php.).

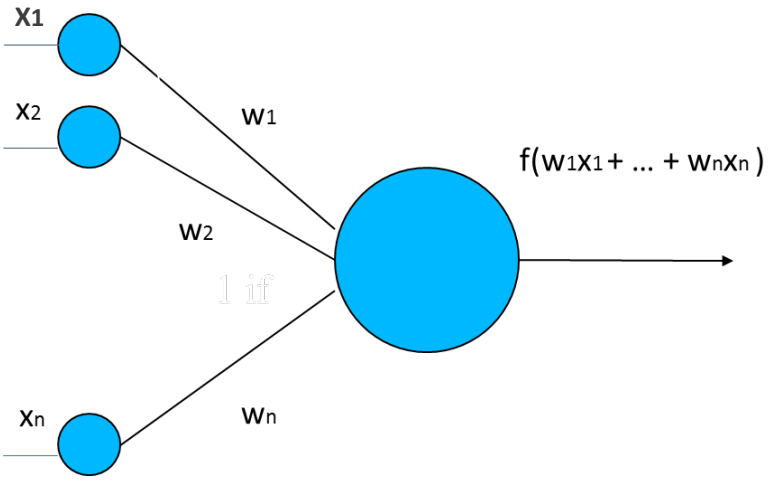


Figure 5.4 – Artificial neuron model. Inputs to the network are represented with the symbol x_n, each of thse inputs are multiplied by a connection weight w_n, summed and fed through the transfer function *f()* to generate a result and the output.

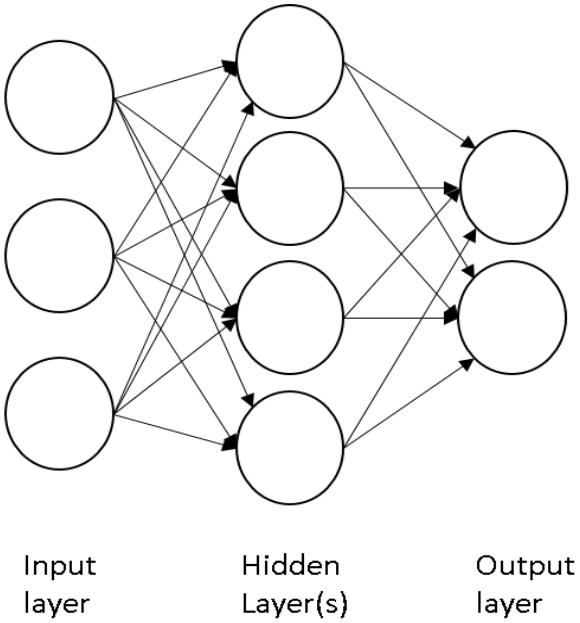


Figure 5.5: A generic single layer feed-forward neural network.

Each of the processing neurons calculates the weighted sum of all interconnected signals from the previous layer plus a bias term and then produces an output through the activation function. The activation function associating individual nodes have typically a sigmoid shape (Figure 5.6). Other transfer functions can also be applied. The adjustment of the ANN function to experimental data (training of the network) is based on a non-linear regression procedure (Fraser, 2000). Random weights are assigned to each neuron, the output of the network is evaluated and the error between the output of the network and the training dataset is calculated. If the error is large, the weights are adjusted and the process goes back to evaluate the network's output. This cycle is repeated until the error is small or a stopping criterion is satisfied.

During the training of a neural network, the prediction error is evaluated for each iteration. The use of a ANN with too many neurons allows an excess of degrees of freedom and can cause overfitting of the data. A test dataset can be kept separated from the training phase and exploited to check how good the prediction capacity of the ANN is, on the basis of the sum of squared prediction errors. For obtaining the optimal intensity of training, a possibility is to explore the ANN performance in order to minimize the sum of the training plus validation (or cross-validation) errors (Figure

5.6), paying attention to not to stop the training process at the first point of minimum. Given the stochasticity of the aggregated error index, waviness in its value as a function of the training intensity may appear (which may lead to wrongly identify a local minimum with a too early stopping of the procedure). As a simple heuristics to mitigate this problem, the training should be allowed to proceed further following the first detected minimum in the aggregated error index, to check whether or not that training intensity is associated with a point of local minimum (de Rigo et al, 2005).

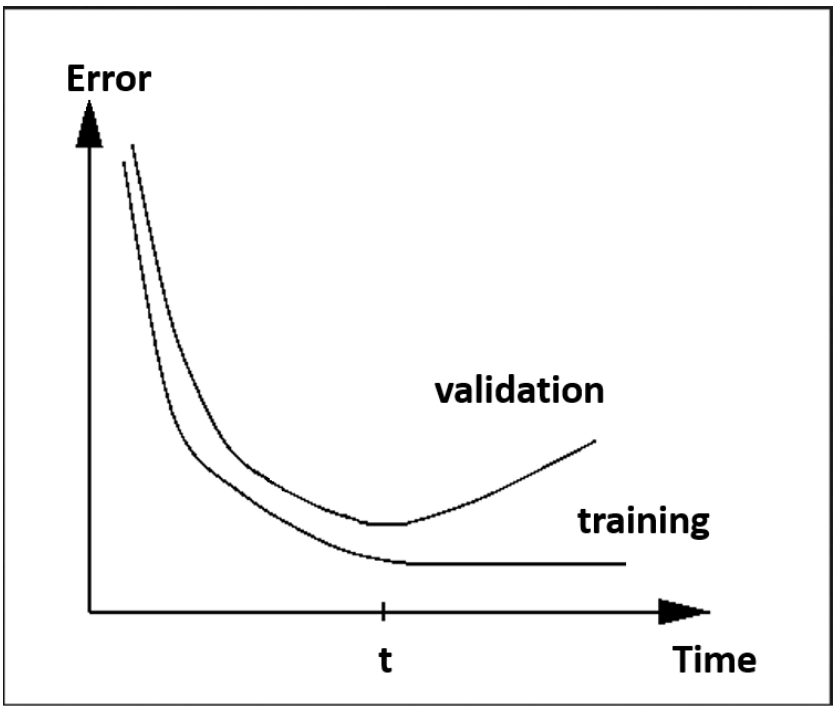


Figure 5.6 – Profiles for training and validation errors

Within the present study we explored the use of a feed-forward neural network through the package nnet (Ripley, 1996; Venables and Ripley, 2002) in GNU R. The functions in this package allow to develop the most common type of neural network model (the feed-forward multi-layer perceptron). This package is widely exploited in the scientific literature for multiple applications (Ashtawy and Mahapatra, 2015; Herrera et al., 2010; Lawler et al., 2006). The functions have enough flexibility to allow the user to develop the best or most optimal models by varying parameters during the training process. Feed-forward neural networks provide a flexible way also to

generalize linear regression functions. They are non-linear regression models but with so many parameters that they are extremely flexible, flexible enough to approximate any smooth function. (Venables and Ripley, 2002).

Nnet implements a feed-forward neural network with single hidden layer (Figure 5.5) allowing 'skip-layer' connections from input to output (some input signals have a direct connection to the output layer) having:

$$y_k = \sigma_o \left(\alpha_k + \sum_h w_{hk}\, \sigma_h \left(\alpha_h + \sum_i w_{ih}\, x_i \right) \right) \tag{5.9}$$

allowing the non-linear units to perturb a linear functional form. The units of input distribute the inputs (x_i) to the hidden layer that sum the inputs, add a constant (α_h and α_k) and take a function σ_h (continuous and limited) of the results. The outputs (y_k) have the same form but with output function σ_o. w_{ih} and w_{hk} represents a set of weights that can also assume negative values [$-\infty, \infty$].

The activation function σ_h of the hidden layer units is frequently a logistic function (sigmoid curve):

$$\sigma_h(x) = \frac{1}{1 + e^{-x}} \tag{5.10}$$

the output units can be linear, logistic or an activating threshold α. Within the nnet package the default is logistic output units.

The ANN architecture coming from the calibration of the model (section 5.3) implemented exploiting the nnet package, use a logistic activation function, no skip-layer connections, 5 neurons and a weight decay of 1e-04 (weight decay specifies regularization in the neural network, it is a regularization term that penalizes big weights).

Logistic Regression

Logistic regression analysis, introduced by Cox in 1958, is one of the more commonly used statistical methods in earth science (Yilmaz, 2009).

Logistic regression can be seen as a special case of the generalized linear model and thus analogous to linear regression. The model of logistic regression, however, is based on quite different assumptions (about the relationship between dependent and independent variables) from those of linear regression. In particular the key differences of these two models can be seen in the following two features of logistic regression. First, the conditional distribution $y \mid x$ is a Bernoulli distribution rather than a Gaussian distribution . Second, the predicted values are probabilities and are therefore restricted to (0,1) through the logistic distribution function because logistic regression predicts the probability of particular outcomes.

Many authors found logistic regression to be a better predictor than bivariate methods (Ayalew and Yamagashi, 2005; Nandi and Shakoor, 2009) and to compare well with artificial neural networks in its predictive performance (Yilmaz, 2009; Rossi et al., 2010). Furthermore, Brenning (2005) showed that logistic regression is less prone to over-fitting the data than support vector machines.

LR is adopted for finding the best-fitting model describing the relationship between a dependent variable (y) (assuming a distribution between presence [1] and absence [0] of landslides) and n explanatory variables (the covariates $x_1, x_2, \ldots, x_n$).

It follows that logistic regression involves fitting an equation of the following form to the data:

$$z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_j x_j , \qquad (5.11)$$

where β_0 is the intercept of the model, the β_j (j = 0, 1,2, ... ,n) are the slope coefficients of the logistic regression model, and the x_j (j = 0, 1, 2, ... , n) are the independent variables.

In our study, the results of the LR can be interpreted as the probability of occurrence of shallow landslides, it does not predict presence or absence of landslides (Brenning, 2005).

For the probability of occurrence of shallow water-induced landslides, given the selected independent variables (listed in section 5.2), the logistic response function is:

$$P(y = 1) = \pi = \frac{1}{1+e^{-z}}, \tag{5.12}$$

where π is the probability of landslide occurrence or susceptibility.

The maximum-likelihood method is used for estimating the coefficients of the logistic multiple regression model. Because of the non-linearity between independent variables and probability, parameter estimation requires the application of an iterative algorithm [5.13]. In order to model the probability π, equation 5.12 is linearized using the logit transformation:

$$logit(\pi) = log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \sum_{j=1}^{k} \beta_j x_j , \tag{5.13}$$

The advantage of logistic regression is that, through the addition of an appropriate link function that generalizes linear regression by allowing the linear model to be related to the response variable, the variables may be either continuous or discrete, or any combination of both types, and they do not necessarily have normal distributions. In the present situation, the dependent variable is a binary variable representing the presence or absence of landslides.

We fitted a logistic regression model in GNU R calling the function glm (), the results are reported in table 5.1. The fitting process is not so different from the one used in linear regression. Using the logistic regression model, the spatial relationship between landslide occurrence and factors influencing landslides was assessed.

Table 5.1 – Values of coefficients $\beta_0, \beta_1, \ldots, \beta_j$ associated to the covariates $x_1, x_2, \ldots, x_n$ as coming from the training of the logistic regression applied exploiting the function glm() in R. Landocover, TWI, slope angle, aspect, elevation, profile- and plan-curvature are the covariates selected to running the model. More information are available in section 5.2.

Coefficients:	Estimate	Std. Error	Z value	Pr (> | Z |)
(Intercept)	1.234e-01	4.958e-01	0.249	0.80352
elevation	-1.951e-04	4.461e-04	-0.437	0.66188
aspect	1.926e-03	5.706e-04	3.374	0.00074
landcover2	-2.542e+00	3.411e-01	-7.453	9.15e-14
landcover3	-5.515e+00	3.782e-01	-14.581	< 2e-16
landcover4	-2.210e+00	3.362e-01	-6.573	4.94e-11
slope	5.028e-02	7.664e-03	6.561	5.33e-11
twi	2.296e-01	2.670e-02	8.598	< 2e-16
plan_curv	-1.980e-10	3.819e-08	-0.005	0.99586
profile_curv	-1.941e-09	1.304e-08	-0.149	0.88171

Relative Distance Similarity

RDS (de Rigo, 2015; Bosco et al., 2015; de Rigo, in prep), already introduced in chapter 4, is a machine learning approach inspired by the architecture of a perceptron (Rosenblatt, 1962). The RDS can be seen as a neural network with a single hidden layer where each of the neurons is linked with a single point of training. In a perceptron each of the training points is linked with all the neurons. In the Relative Distance Similarity the single link among a neuron and a point of training has the advantage to avoid overfitting.

The RDS index of a given multi-dimensional point *c* with respect to a set *A* of reference points involves the relative distance among the pairs $\{C_c, C_\alpha\}$ for each $\alpha \in A$ and each dimension *j* of the N^C covariates

$$RDS_c = \max_{\alpha \in A} \left(\prod_{j=1}^{N^C} \left(\frac{max\left(min\left(C_c^j, C_\alpha^j\right), \delta C^j\right)}{max\left(max\left(C_c^j, C_\alpha^j\right), \delta C^j\right)} \right) \right), \tag{5.14}$$

where δC^j (as in the case of equation 4.16) is half of the measurement accuracy of the covariates and Ω (de Rigo, in prep) is a statistical operator with which the relative distances along each dimension of the covariates are aggregated in the RDS index. The aggregated RDS index is mathematically constrained to vary between 0 and 1. For most of the families of Ω, the aggregated index has value 1 only if all the relative distances along each dimension of the covariates have value 1 (de Rigo, in prep). As a consequence, in RDS when the values of the covariates in a point have exactly the same value of the covariates in one of the points of training, then the output value of the neuron linked with that point is 1. The higher is the distance between the covariates, the more the output of the neuron approaches zero. The closer are the values of the covariates to the values of a training point, the more the neuron linked with that point of training is activated.

5.2. Data and explanatory variables

A dataset of more than 400 reported landslides that affected the catchment in 2006 and described in chapter 3, was used. This dataset is based on high resolution IKONOS satellite imagery. To make the interpretation easier, the satellite images were orthorectified and pansharpened. For running the statistical models, a set of calibration parameters was selected from international literature (Yilmaz, 2009; Pradhan and Lee, 2009; Rossi et al., 2010). Seven parameters, commonly assumed as directly or indirectly related to landslide occurrence (Yilmaz, 2009; Pradhan and Lee, 2009), have been used for calculating the ANN, LR and RDS models. The Topographic Wetness Index (TWI) (Beven and Kirkby, 1979), the slope angle and aspect, the profile and plan curvature have been calculated starting from a digital elevation model with a resolution of 5 meters by exploiting the tools available in ArcGis. The remaining models covariates are the land cover and the elevation.

The input parameters for running the deterministic models SINMAP and Translide (root and soil cohesion, internal friction angle, bulk density of the soil, groundwater height, soil depth and the effective recharge rate (Table 5.2)), have been determined using different methods. Where data was not measured directly, estimated values

from the literature were applied (Witt, 2005; Horn and Fleige, 2003; Morgan et al., 1998; Cotecchia et al., 2006), if available, otherwise physically sensible values based on our judgment were used.

The effective recharge rate used in SINMAP (50 mm/d) was derived comparing the available climatological data (Rocchetta SantAntonio and Rocchetta scalo meteorological stations) and values and approaches from literature (e.g. for calculating the transmissivity rate) (Witt, 2005). The hydraulic conductivity as the soil cohesion and the internal friction angle (used both in SINMAP and TransSlide) were informed by field data and the distribution of general soil characteristics of the catchment derived from the lithological map of the study area (Wasowski et al., 2012). These values compare well with typical soil properties values from the literature (Morgan et al., 1998; Horn and Fleige, 2003). The model input values linked with soil characteristics are shown in the table below.

Table 5.2 - Modelling input values, used in SINMAP and Translide, and linked with soil characteristics (the values reported for the bulk density are on moist conditions). These values comes from literature or from data collected in a few spots during the field survey. Because of the limited amaunt of available information, average values for the whole study site, were applied within these deterministic models.

Parameters	**Soils over sandstone**	**Soils over limestone**	**Soils over clay-shales**
Soil cohesion	0-10 (kPa)	0-8 (kPa)	0-10 (kPa)
Int. friction angle	25-32 º	31-34 º	10-22 º
Soil depth	1.5 (m)	1.5 (m)	1.5 (m)
Bulk density	1550 (kg/m^3)	1600 (kg/m^3)	1450 (kg/m^3)
Groundwater height	1 - 1.5 (m)	1 – 1.5 (m)	1 – 1.5 (m)
Effect. recharge rate	50 (mm/d)	50 (mm/d)	50 (mm/d)

The input values for cohesion (c_s+c_r) used within the model were in the interval 0.1 - 0.4 in areas covered by crops and pastures and between 0.6 and 0.8 in areas covered by bush or forest. Due to the difficulties in measuring the soil depth we had during the field survey, a constant value of 1.5 meters was used, it derives from the few information collected on field and from the judgment of local experts.

5.3. Calibration of the models

An initial set of data was partitioned into two subsets, with the calibration carried out on a training subset (2091 pixels) and the validation performed on the testing subset (682 pixels). The three statistical models were executed using both the same training set and the same set of covariates. The selected areas were subject to **::binary::**[7] classification as stable (0) or not stable (1) (Figure 5.7, 5.8) on the basis of a dataset presented by (Wasowski et al, 2012). The calibration points represent the 0.6% of the catchment area. These have been selected by applying the RDS ensemble technique for analysing the relative distance between the model's covariates within the catchment. Applying this method, for selecting the training and validation points, it is possible to minimize the presence in the training set of areas having similar characterisics.

The multicollinearity of the dataset of **::nonnegative::**[8] **::finite::**[9] predictors was also analysed in order to avoid strong correlation between different predictors. This was done for preventing the possibility that small changes in the data can cause an erratic change of the coefficient estimates.

For estimating the more appropriate model parameters we applied a repeated random sub-sampling cross-validation to the training set of data. Using cross-validation it is possible to estimate how accurately the models perform.

For measuring modelling performance during calibration process, mean absolute error (MAE) and root mean square error (RMSE) against validation data were calculated. Although some authors suggest inter-comparisons of average model performance should be based on MAE (Willmott and Matsuura, 2005), RMSE was also calculated here because of its greater sensitivity to occasional large error compared to other measures.

The classification criteria used for measuring the landslide susceptibility fluctuate in a range between 0 (stable conditions) and 1 (unstable conditions). The adopted

[7] http://mastrave.org/doc/mtv_m/check_is#SAP_binary
[8] http://mastrave.org/doc/mtv_m/check_is#SAP_nonnegative
[9] http://mastrave.org/doc/mtv_m/check_is#SAP_finite

classification scheme is as follows: 0-0.4 (stable), 0.4-0.6 (area of model uncertainty) and 0.6-1 (unstable).

Heterogeneous quantities provided as model outputs need to be transformed so as to be homogeneous in order for the models' performance to be comparable. In the output of the physically based models values minor than 1 represents unstable conditions, hence, it should be remapped as 1 and high output values represent stable conditions, which should remapped as 0.

The ::**nonnegative**:: output of the two deterministic models has been remapped to the corresponding ::**possibility**::[10] values ∈ [0 1] by means of Piecewise Cubic Hermite Interpolating Polynomials (PCHIP) (Figure 5.9) with the codelet (MATLAB language): pchip([0 0.5 1 1.25 1.5 10 inf], [1 0.8 0.6 0.5 0.4 0 0], output). PCHIP has been chosen because ensuring monotonicity, continuity and derivability.

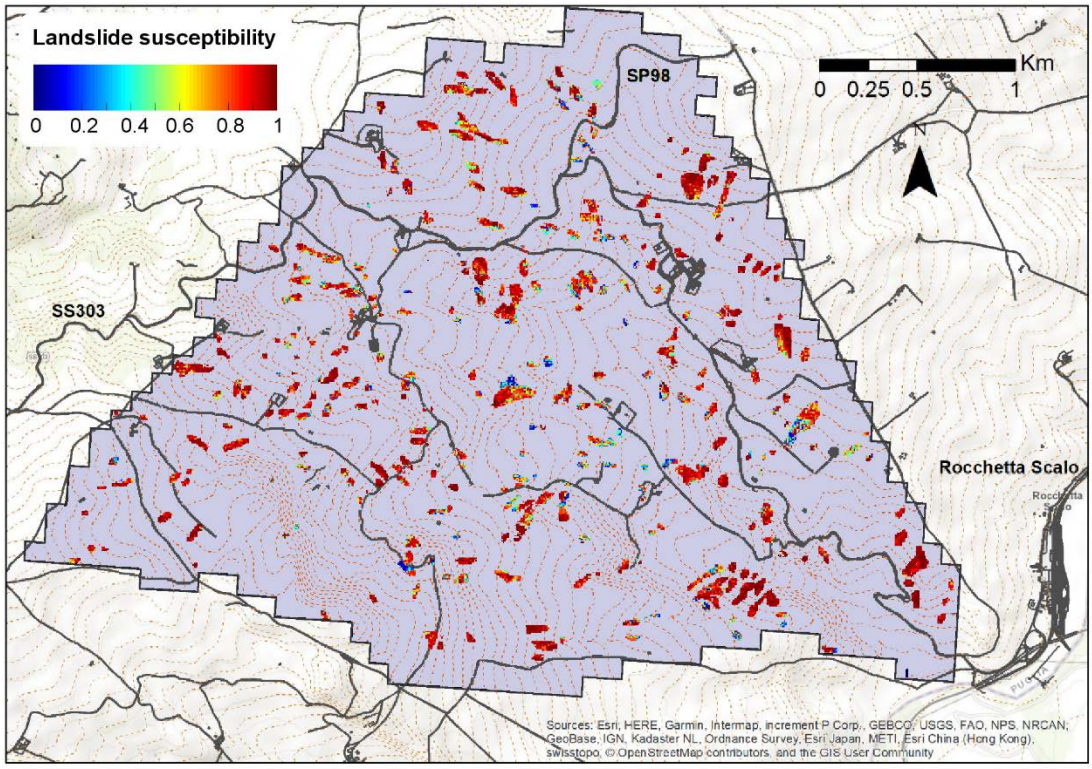


Figure 5.7 – This map represents the landslides areas that were used for selecting the not stable points suitable to populate the modelling set of data with the associated degree of instability coming from the application of the RDS model.

[10] http://mastrave.org/doc/mtv_m/check_is#SAP_possibility

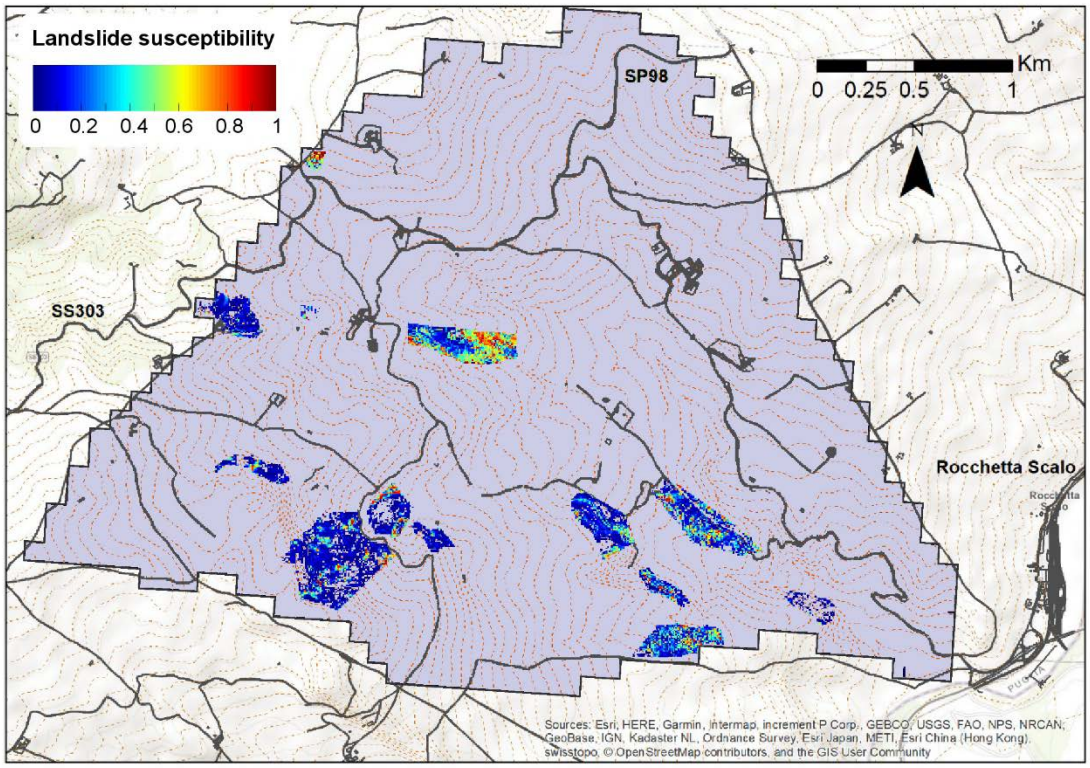


Figure 5.8 - Map of the stable areas that were used for populating the set of points representing a condition of slope stability within the catchment with associated the degree of stability predicted by the RDS model . These areas of stability were selected during our field survey both with direct observation and collecting historical information from local people.

In calibrating ANN the RMSE was determined for every combination of weight decay and number of neurons calculated by the ANN model during the network training. The final architecture of the ANN comes from the selection of weight decay and number of nodes minimizing the RMSE. The calculation of RMSE was then used for evaluating the fitting performance of the models.

Calibration of the deterministic models was performed using data and information from the literature (Witt, 2005; Horn and Fleige, 2003; Morgan et al., 1998), from expert judgement or collected during the field survey. Due to the paucity of high resolution data for calculating the parameters required by SINMAP and TransSlide only a basic calibration was possible. We tested the model using different parameter values, the data coming from the field survey or derived from literature were used for example to calculate in SINMAP lower and upper bound of the ratio of

transmissivity to the effective recharge rate (T/R), of the soil friction angle or of the dimensionless cohesion (that takes into account both root and soil cohesion). The low resolution of the available data unfortunately does not allow the deterministic models to catch local variations within the catchment. The adjustments coming from the calibration process were necessary for capturing a major proportion of landslides in areas having a low stability index.

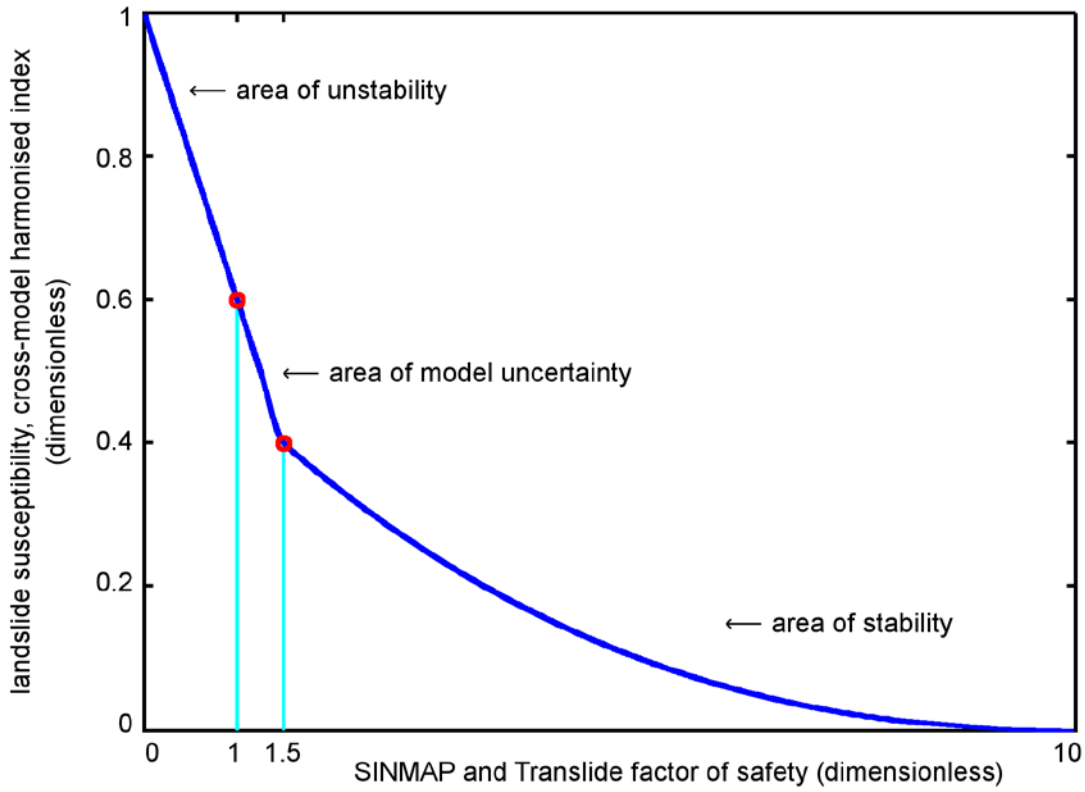


Figure 5.9 – Use of the Piecewise Cubic Hermite Interpolating Polynomials (PCHIP) for remapping the deterministic models to the corresponding ::**possibility**:: values ∈ [0 1].

5.4. The fuzzy ensemble approach

A semi-quantitative method, based on an ensemble approach, has been used for combining deterministic and probabilistic approaches. The ensemble approach is a reproducible Data Transformation Model applied to the results of the array of models of landslide susceptibility and is based on the relative-distance similarity

method. The application of an ensemble approach, especially in data poor regions, could potentially reduce the uncertainty and mitigate local poor performance associated with individual models, by excluding outlier estimations.

Uncertainty may affect these models from the inaccuracy of required input data and parameters to the approximation of their reconstruction.

In this application, the set A of reference points, mentioned in eq. 5.14, is instantiated for stable areas (S^S) and unstable ones (i.e. the areas which might be subject to landslide phenomena, S^L).

In the landslide application, the indices RDS_c^L and RDS_c^S express the possibility [0,1] for *c* to respectively belong to L (unstable areas) or S (stable areas)

$$\begin{cases} RDS_c^L = \max\limits_{\alpha \in S^L} \left(\prod\limits_{j=1}^{N^C} \left(\frac{max\left(min\left(C_c^j, C_\alpha^j\right), \delta C^j\right)}{max\left(max\left(C_c^j, C_\alpha^j\right), \delta C^j\right)} \right) \right) \\ RDS_c^S = \max\limits_{\alpha \in S^S} \left(\prod\limits_{j=1}^{N^C} \left(\frac{max\left(min\left(C_c^j, C_\alpha^j\right), \delta C^j\right)}{max\left(max\left(C_c^j, C_\alpha^j\right), \delta C^j\right)} \right) \right) \end{cases} \quad (5.15)$$

A final ensemble takes into account both indices RDS_c^L and RDS_c^S so as to derive a harmonised RDS model for the landslide susceptibility.

5.5. Validation and analysis of the models performance

The performances of the different approaches have been estimated with an independent set of data (682 points).

These data were used for measuring the modelling performance by calculating MAE, RMSE and the explained variance of the model (expressed in proportional terms).

For calculating the explained variance, we used the pseudo – R^2 reported in equation 5.16:

$$pseudo - R^2 = 1 - \frac{MSE}{var(obs)} \quad (5.16)$$

where *var(obs)* is the variance of the observed data and *MSE* is the mean square error.

The results of validation are summarized in table 5.3. The table shows a better performance of the statistical methods when compared with deterministic approaches. A spatial analysis of the predicted landslide susceptibility (Figure 5.10) reveals the difficulties of physically based models to identify instabilities in some areas of the catchment. For example, the northeastern part of the catchment, largely affected by slope instability, is considered as stable by these models. This could be linked to the cumulated bias in the parameters required by physically based models, which are weakly approximated due to the lack of required information at the appropriate spatial resolution. Concerning data-driven statistical models, a key difference should be highlighted related to the performance which may be obtained with the use of out-of-the-box tools and custom designed machine learning models. The first category of tools is easily accessible by researchers even when their background does not include advanced modelling training. These tools (for example, the tool here applied for estimating ANN statistical method) are relatively easy to run with minimal expertise on machine learning modelling. As a trade-off, the performance achievable by non-experts is often not comparable with the performance obtainable by experienced modellers with custom designed machine learning approaches (here the RDS approach exemplifies a custom designed modelling approach).

Both ANN and LR show some difficulties in predicting stable and unstable areas within the study area. Their explained variance of around 0.3 is in line with many other works (Ermini at al., 2005; Ayalew et al., 2005; Costanzo et al.,2014). Both these models tend to overestimate slope instability in areas where landslide activity is not present. The unusual slightly lower performance of the ANN when compared with LR (Lee et al., 2016) is probably due to the out-of-the-box application of the first model. One of the main objectives of our work was to test the potential to apply ensemble modelling methods to predict landslide susceptibility in data-poor regions. The low prediction capacity of some of the applied techniques highlight the potential of

ensemble methods in reducing locally-poor model performance as it allows the mitigation of outlier predictions.

Between the applied models, the ensemble and RDS give the lowest errors and have the highest explained variance with values (both over 0.9) that are similar to other machine learning applications for measuring landslide susceptibility (Bui et al., 2016). Although the high error rate of some models, the simple models' median (table 5.32), with an explained variance that reaches 0.6 in predicting stable areas within the catchment, is the next best result. Its application as a straightforward unsupervised ensemble might prove useful even where no additional information is available (black box output data).

The proposed ensemble, being a supervised method that infers a function from training data, slightly improves the best model in the array of outputs (in the worste case, the ensemble would have been equal to the best model in the array). The high performance showed by the RDS approach could be linked with the criterion used for the selection of the training and testing set of data. The possible presence of bias in using a similar technique for selecting the data and calculating the landslide susceptibility need to be further investigated. Because the quality of spatial landslide forecasts is largely dependent on the quality of the available datasets, the good performance of the combined model broadens the possibility of applying a quantitative assessment in data-poor regions.

Anyway, the good performance of the ensemble method (as reported in table 5.3) confirms the potential to apply ensemble modelling methods to predict landslide susceptibility even in data-poor regions, where the best available models would simply be impossible to apply due to the lack of detailed information. The work in this thesis contributes to corroborate the hypothesis that where a single state-of-art (but too data-demanding) model cannot be applied, multiple simpler models may be aggregated to improve their performance.

Table 5.3- Values of the MAE and RMSE calculated on the validation set of data related to singles and combined applied models. U and S refer on MAE and RMSE calculated respectively on unstable on unstable (U) and stable (S) areas only. With lower MAE and RMSE values, statistical methods show better performance when compared with deterministic models. The best performance, linked with the lowest values of MAE and RMSE both in Stable and unstable areas, are related to the application of the Relative Distance Similarity and of the ensemble method. Because of the lack of detailed input data to calibrate the deterministic models (SINMAP and TRANSSL) these show the highest RMSE and MAE values.

	RDS	ANN	LR	SINMAP	TRANSSL	MEDIAN	ENSEMB.
MAE	0.003	0.44	0.37	0.45	0.51	0.35	0.001
MAE U.	0.002	0.42	0.36	0.61	0.68	0.45	0
MAE S.	0.003	0.45	0.38	0.3	0.34	0.25	0.001
RMSE	0.02	0.47	0.43	0.54	0.58	0.4	0.019
RMSE U.	0.01	0.45	0.4	0.65	0.7	0.47	0
RMSE S.	0.03	0.48	0.46	0.4	0.42	0.32	0.026

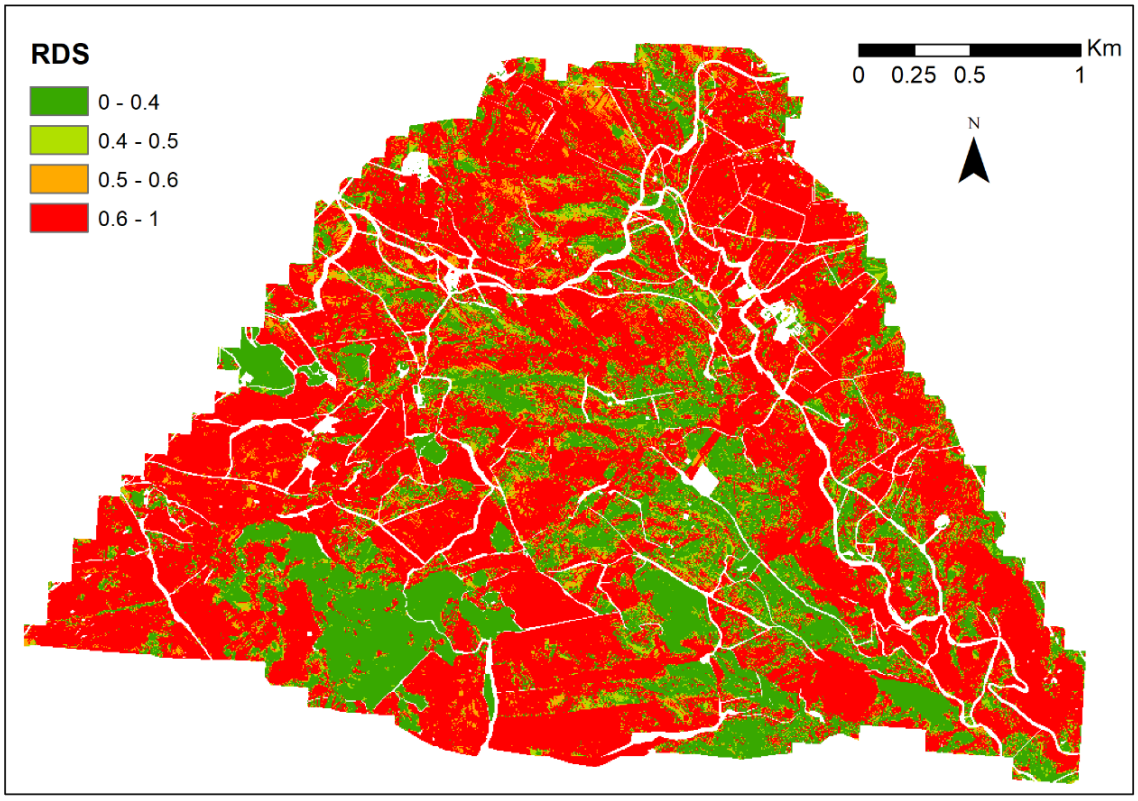


Figure 5.10– Landslide susceptibility map of the study area produced by applying the RDS model (GNU Octave with Mastrave modelling library). The fuzzy classification index for measuring the landslide susceptibility estimates the possibility of instability, and fluctuates in a range between 0 (stable conditions) and 1 (unstable conditions). In red, the areas estimated as more susceptible to landslide occurrence.

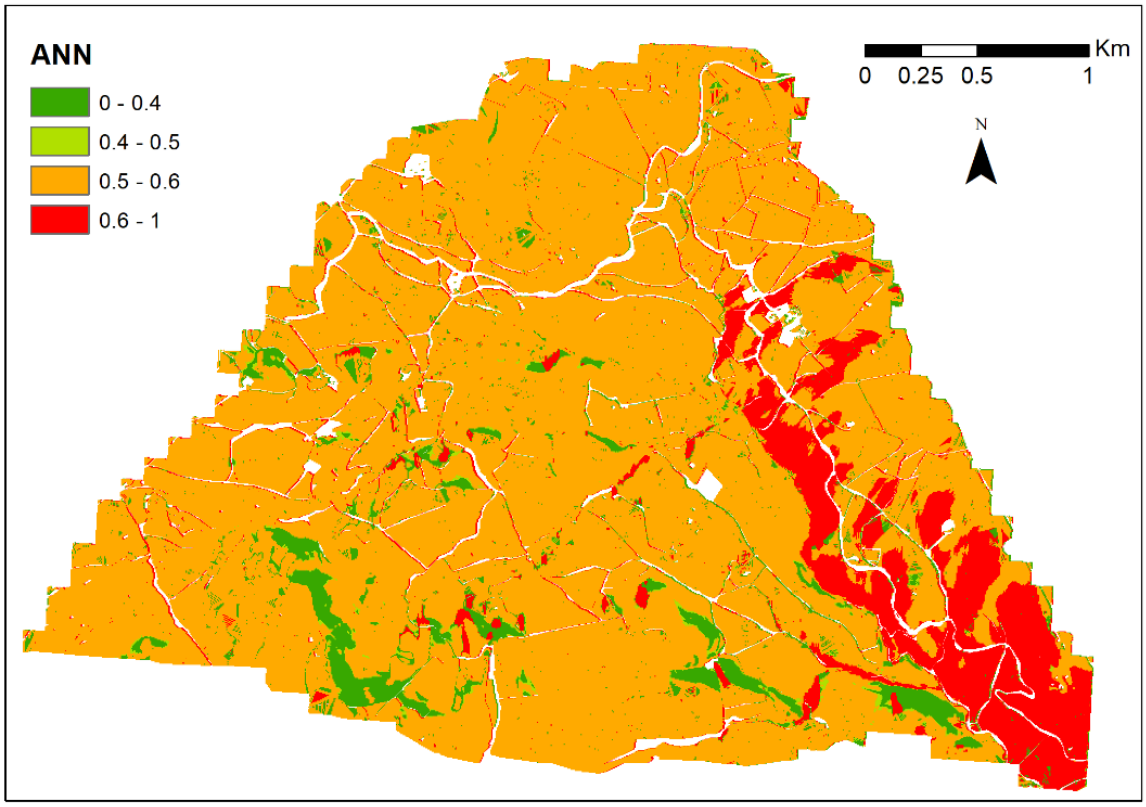


Figure 5.11– Landslide susceptibility map of the study area, produced by applying an artificial neural network (multilayer perceptron, as implemented by the "nnet" package of GNU R). The map ranges from stable conditions (0) to unstable conditions (1). The areas with values between 0.4 and 0.6 are characterised by high modelling uncertainty.

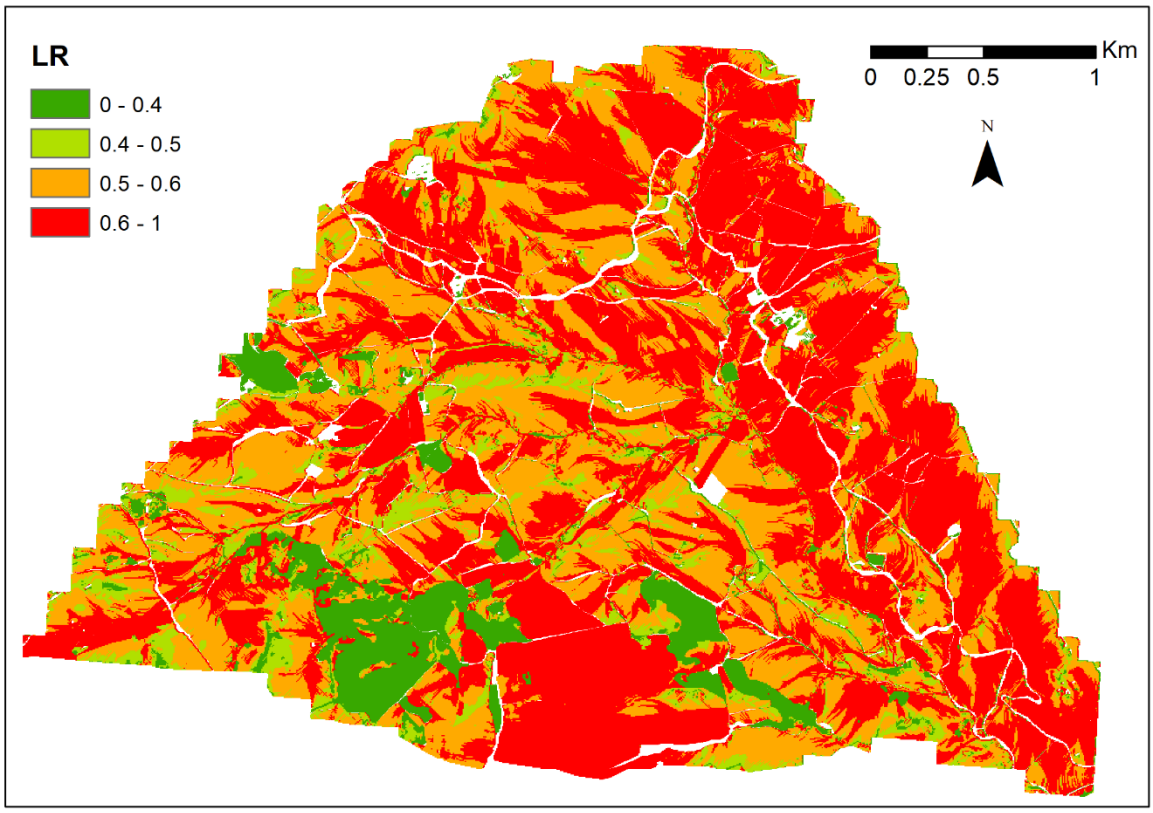


Figure 5.12– Landslide susceptibility map of the study site produced by applying a logistic regression model. In red, the areas estimated as more susceptible to landslide occurrence (values from 0.6 to 1). In green, the areas presenting a low susceptibility to landslides (values below 0.4). In light green and orange, the areas where the uncertainty of the model is high.

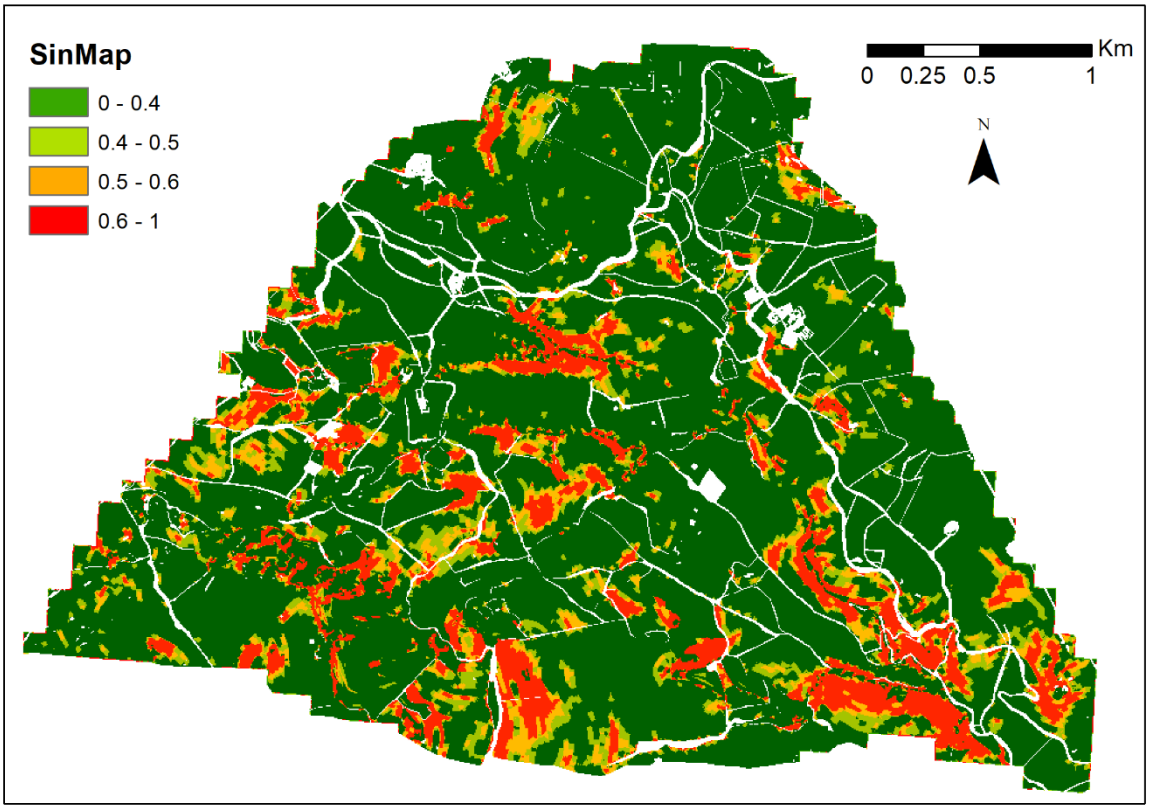


Figure 5.13– Landslide susceptibility map of the study area, produced by applying the SinMap model. The classification index estimating the susceptibility to landlides ranges between 0 (stable) and 1 (unstable).

In the output of the SinMap model, values less than 1 represent unstable conditions, while the higher output values represent stable conditions. To harmonise the susceptibility scale, the model output were here remapped by means of Piecewise Cubic Hermite Interpolating Polynomials (PCHIP) to values $\in [0\ 1]$.

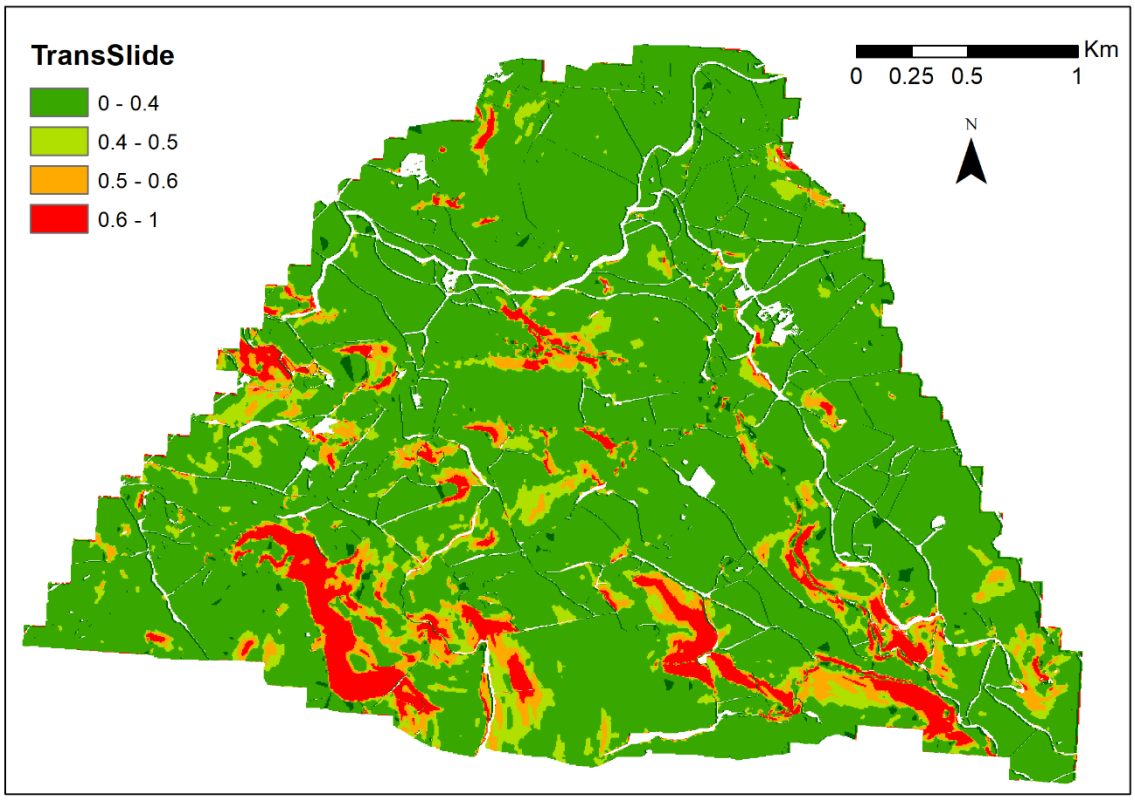


Figure 5.14– Landslide susceptibility map of the study area produced by applying the TransSlide model. The classification index for measuring the landslide susceptibility fluctuates in a range between 0 (stable conditions) and 1 (unstable conditions). In red, the areas more susceptible to landlide occurrence are reported.

As for the Sinmap model, the original classification scale was remapped to harmonise it with the one of the other models. Originally, values less than 1 represented unstable conditions, and high output values represented stable conditions. The model output were remapped by means of Piecewise Cubic Hermite Interpolating Polynomials (PCHIP) to range between 0 and 1.

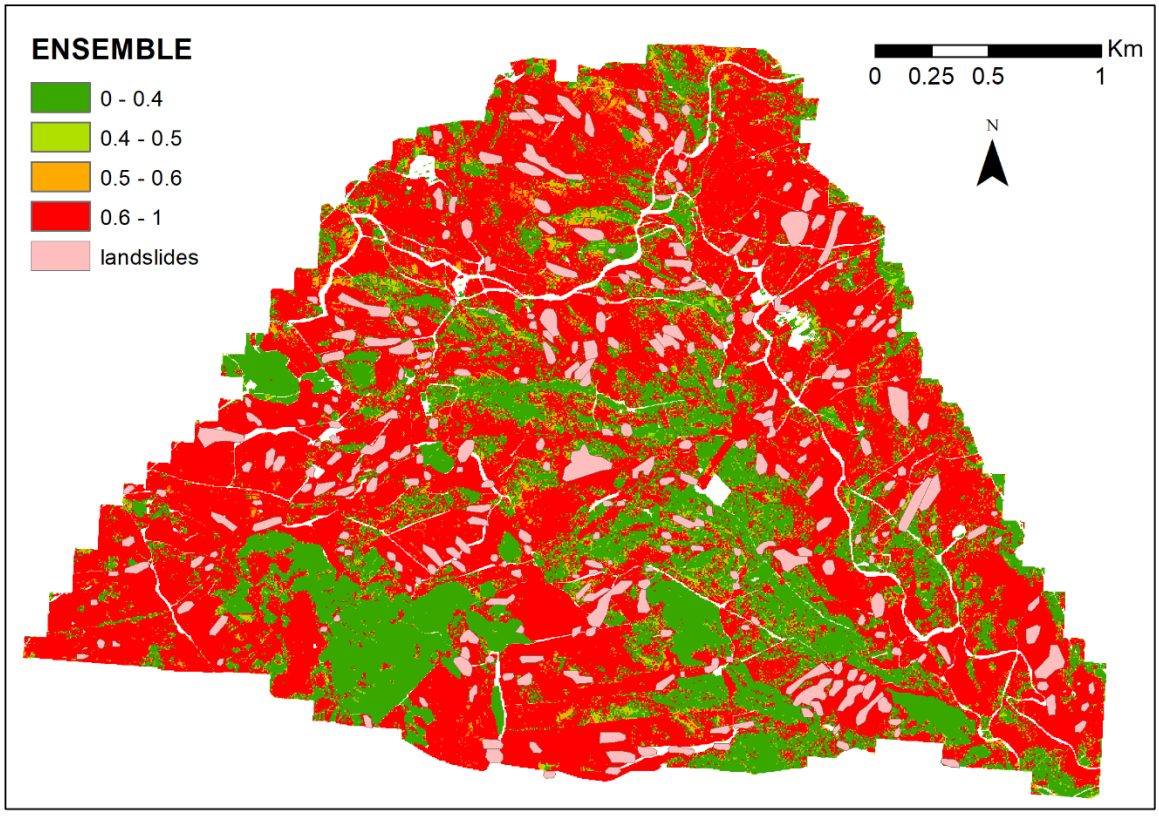


Figure 5.15– Landslide susceptibility map of the study area. This map represents the output of the emsembling model and reports the landslides observed within the study site. In the area, the ensemble estimates fluctuate from stable conditions (0) to unstable conditions (1). The few areas with values between 0.4 and 0.6 are characterised by high modelling uncertainty.

5.6. Conclusions

Landslide susceptibility assessment is a fundamental component of effective landslide prevention. One of the main challenges in landslide forecasting is the assessment of spatial distribution of landslide susceptibility. Despite the many different existing approaches, landslide susceptibility assessment still remains a challenge. A semi-quantitative method was here proposed combining heuristic, deterministic and probabilistic approaches for a robust catchment scale assessment.

Five different techniques for modelling shallow landslide susceptibility were applied to a catchment located in Southern Italy (Rocchetta Sant'Antonio, FG). Each susceptibility zonation has been obtained by applying heterogeneous statistical techniques as logistic regression , relative distance similarity , artificial neural network and two different landslide susceptibility techniques based on the infinite slope stability model (SINMAP and TransSlide). A fuzzy ensemble model has been exploited for aggregating the array of different susceptibility zonation maps.

The performance of the models was evaluated against a landslide inventory of the year 2006 by calculating RMSE, MAE and explained variance. The good results of the ensemble model, when compared with the single techniques, make this method suitable to be applied in data poor regions with a lack of calibration and validation data. Because of the uncertainty in selecting a single suitable method for modelling spatial landslide susceptibility in areas characterized by data weakness, the applied ensemble method can potentially result in a less uncertain zonation. Alhough these preliminary results are promising, further research is required before this method can be used to communicate the findings with relevant authorities. The landslide susceptibility maps (Figure 5.10) calculated applying the statistical methods were obtained using three different scripts implemented using MATLAB and R languages respectively in GNU Octave and GNU R free software along with the modelling library Mastrave which implements the semantic array programming paradigm. The script applied for calculating the ANN is based on the work of Rossi (Rossi et al., 2010). Also the scripts used for selecting the training and testing points and for calculating the combined model were written following the SemAP paradigm in MATLAB language, within GNU Octave as computing environment.

6. A coupled architecture for modelling soil erosion and shallow landslides in data poor regions (the case of Rocchetta Sant'Antonio)

6.1. Introduction

The changes in soil erosion rates that follow landslide events can deliver significant cascading impacts on ecosystems, for example due to an increased sediment yield to a stream network.

This may potentially be of ecological and economical relevance both close to where the landslide events are located (so called "on site" impacts) and at a wider scale (so called "off site" impacts). Local effects may potentially drive complex changes even at the landscape-scale (Bakker et al., 2005; Geertsema and Pojar, 2007). Furthermore, the ecosystem services provided by the areas affected by landslide events may be important for remote service benefit areas connected through so called "service connecting areas" such as stream networks (Syrbe and Walz, 2012).

Natural resources are intrinsically entangled in complex networks whose management is increasingly complicated by climate change. There is indeed a growing awareness of the importance of modelling these cascades and the potential influence of climate change on these processes, and assess the resultant economic and societal consequences (de Rigo, 2012).

Landslide events will result in changes in topography and vegetation cover which in turn will alter surface erosion rates and sediment yields. As stated in Chapter 2, there are a number of relevant models that use an integrated approach to soil erosion and landslide processes, including SHETRAN) (Ewen et al., 2000), TOPOG (a physically-based, distributed parameter, catchment hydrological model) (O'Loughlin, 1986; CSIRO, 2017) , PSIAC (Pacific Southwest Inter-Agency Committee) (PSIAC, 1968) or SIBERIA (also known as the Willgoose Catchment Evolution Model) (Willgoose and Riley, 1998). But it is only in WEPP-SLIP (Water Erosion Prediction project - Shallow

Landslide Integrated Prediction) (Cochrane, and Acharya, 2011) that post-failure sediment yield is explicitly considered. WEPP-SLIP is able to consider the post-failure changes in soil erosion rate through the changes in topography and land cover.

However, there is still room to improve the modelling of the interactions of these processes, for example through assessments of the changes in surface area made more susceptible to soil erosion following landslide events.

To quantify the potential changes in soil erosion due to landslide occurrence it is necessary to know where and when on the slope a landslide initiates and how it evolves. This chapter aims to present a new modelling approach for data-poor regions in an attempt to improve the estimation of sediment budgets derived from rainfall induced landsliding and soil erosion. A statistical approach is proposed that incorporates the frequency-area landslide distribution model of Malamud et al. (2004) within the framework of a spatially distributed empirical soil erosion model.

6.2. A new architecture for coupling of the effects of rainfall-induced shallow landslides and soil erosion

6.2.1. Geospatial semantic array programming

Semantic array programming (see paragraph 2.5) has been used for building the architecture for our modelling approach. The proposed architecture (Figure 6.1) also exploits the geospatial capacities of GIS in order to estimate soil erosion yield (e-RUSLE model). In our modelling approach we integrated SemAP and geospatial tools (ArcGis and GRASS GIS) through the Geospatial Semantic Array Programming paradigm (GeoSemAP). GeoSemAP exploits geospatial tools and Semantic Array Programming for splitting a complex D-TM into logical blocks whose reliability can more easily be checked by applying geospatial and mathematical constraints. Those constraints take the form of precondition, invariant and postcondition semantic

checks. This way, even complex wide-scale transdisciplinary models may be described as the composition of simpler GeoSemAP blocks.

Semantic checks, within and between the different blocks as showed in figure 6.1, are exemplified in the following paragraphs with the already adopted notation ::**sem**::. The semantic constraints were implemented within the code with a specialised module (de Rigo, 2012c) of the Mastrave modelling library. A hyperlink to the corresponding online description is provided.

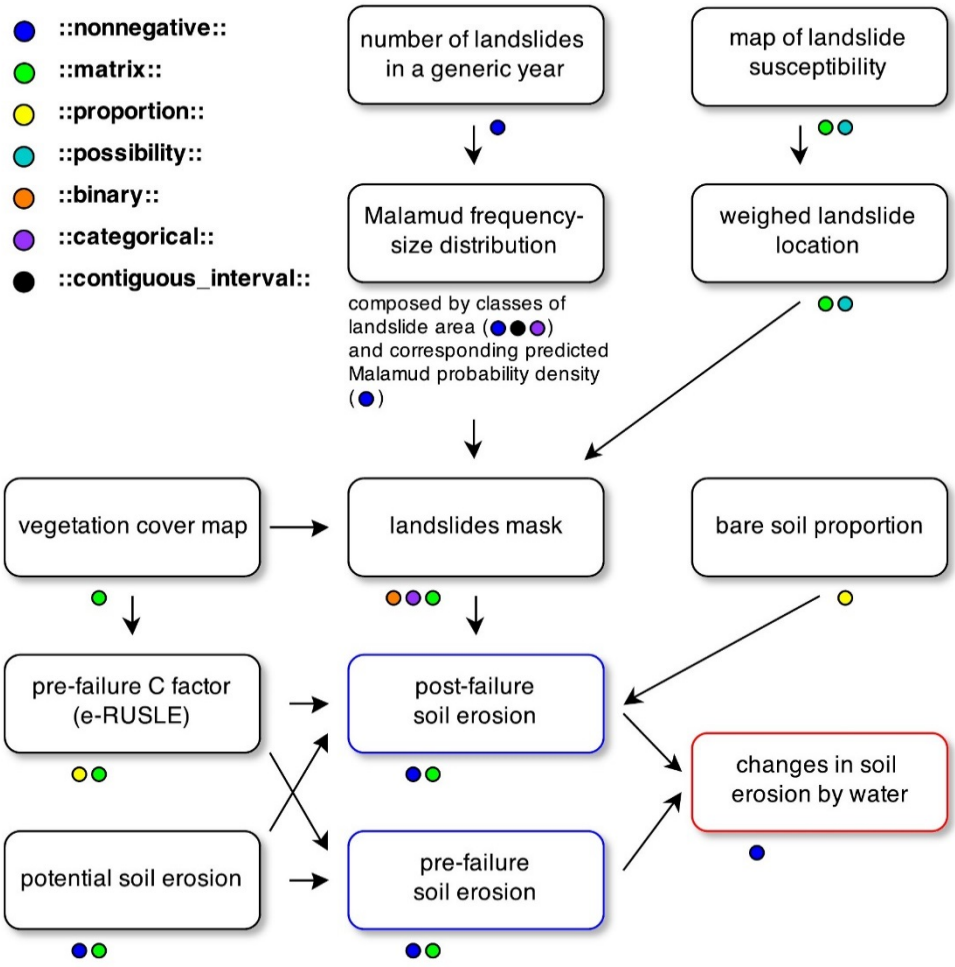


Figure 6.1: Flowchart of the model. The proposed architecture splits a single data-transformation-model (D-TM) for measuring soil erosion, in areas affected by slope instability, into logical blocks whose reliability is checked by applying semantic constraints. The semantic aspects of the data-transformations among model components are highlighted within the workflow with the notation ::sem::.

The decomposition of a complex model in logical components (blocks), and the use of semantic checks within each of them to ensure consistency, help to isolate, mitigate and correct the effects of data inputs when they are occasionally inadequate for some component (this is a typical case when an extensive sequence of numerical runs is required, which otherwise might exceed the ability of computational scientist to verify each single run).

6.2.2. The modelling architecture

The pre- and post-failure soil loss rate was calculated by applying the low data demanding model e-RUSLE (Bosco et al., 2015) that has been presented in chapter 4. Due to the flexibility of the modelling architecture that e-RUSLE is based on, it is possible to calibrate the model for application at different scales (Bosco et al., 2015). e-RUSLE was implemented using the ArcGIS software to first estimate the **::nonnegative:: ::matrix::**[11] representing the soil erosion rates within the catchment without considering the influence of mass movement. The scripts applied for calculating the soil erosion losses was implemented in ESRI ArcGis but can also be easily carried out using an Open Source Free Software such as GRASS GIS or Quantum GIS.

For quantifying the effect of size, position and number of landslides affecting this catchment the frequency-size distribution model proposed by Malamud et al. (2004) was adopted. They found that landslide data from well-documented and substantially complete landslide-event inventories from three quite different locations around the world (Italy, Guatemala and the United States), each with different triggering mechanisms, could be described quite well with the inverse gamma distribution (Figure 6.2):

$$p(A_L, \rho, a, s) = \frac{1}{a\Gamma(\rho)}\left[\frac{a}{A_L - s}\right]^{\rho+1} exp\left[\frac{-a}{A_L - s}\right], \qquad (6.1)$$

In (6.1), p is the probability density (km^{-2}), Γ is the gamma function, A_L is the landslide area (km^2), ρ *(-)* is a parameter which controls the power law decay for medium and

[11] http://mastrave.org/doc/mtv_m/check_is#SAP_matrix

large landslide areas, a (km^2) determines the position of the maximum in the probability distribution and s (km^2) is a parameter which fits the exponential decay behaviour for small landslide areas. Parameter values of ρ = 1.4, a = 1.28 10^{-3} km^2 and s = -1.32 10^{-4} km^2 , such as from the work of Malamud (2004), were shown to provide a good fit to the measured data. The same dataset of over 400 reported landslides described in paragraph 5.2 was used. Unfortunately this dataset is not freely available but the IFFI (Inventario dei Fenomeni Franosi in Italia) database (Agnesi et al., 2007) (a national project that aims at identifying and mapping landslides over the whole Italian territory) is a valuable alternative to apply our modelling approach.

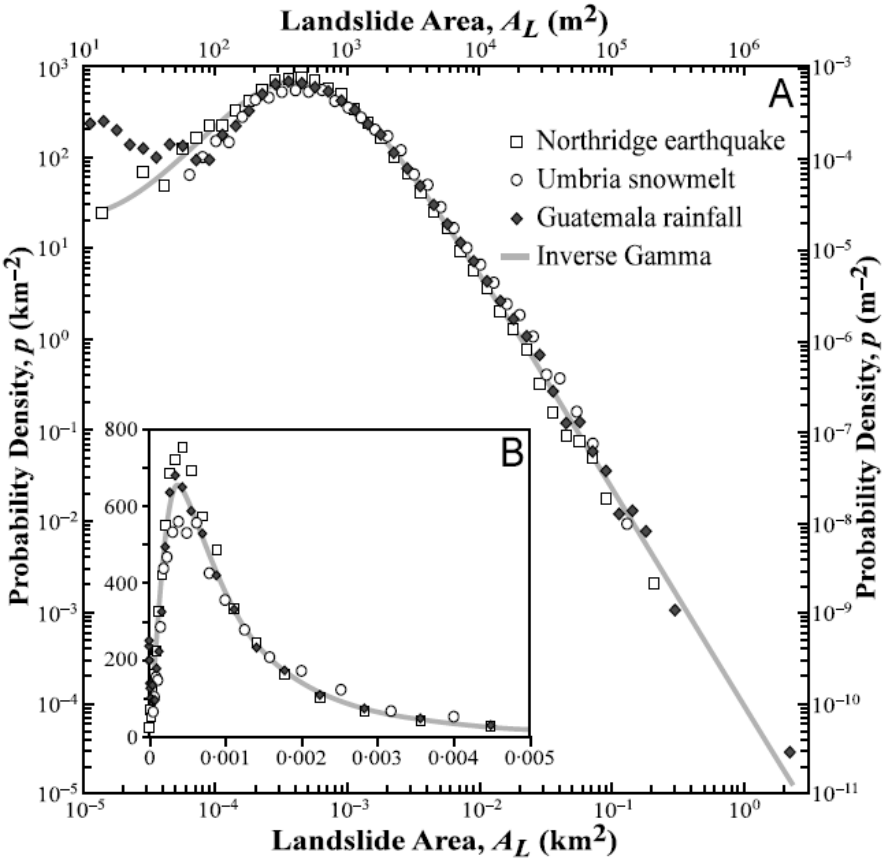


Figure 6.2 - Dependence of landslide probability densities p on landslide area A_L, for three landslide inventories: 1994, earthquake in California, USA (Harp and Jibson, 1995, 1996); 1997, snowmelt event in the Umbria region (Italy) (Cardinali et al., 2000); 1998, landslides triggered by Hurricane Mitch in Guatemala (Bucknam et al., 2001). (Source: Malamud et al., 2004).

Overall, a reasonable correlation between the inverse-gamma distribution of Malamud et al. (Malamud et al., 2004) with the above parameter values and the frequency-size distribution of the landslide database was found (Figure 6.3). The fit is very good for landslide areas greater than or equal to the peak in the distribution. For smaller landslide areas to the left of the peak the agreement is not as good, though modifications to parameters a and s could be made to improve this section.

However the distribution of Malamud et al. (2004) and parameter values they used, were shown to work over a wide range of landslide sizes from various countries around the world. It was found that these same parameter values also provided a similar fit to the data from our field site suggesting the possibility of universality in the parameter values and therefore removing the need for calibrating the distribution for local applications. On this basis we wanted to see how well this would perform against data from the Rocchetta catchment and kept the original Malamud parameter values. The data for the smaller landslides does have a greater degree of uncertainty as its collection could easily have led to either an over or underestimation of the landslide number. This could occur through either medium landslides being classified as smaller due to being covered by larger landslides, or though the smaller landslides being covered by larger ones and therefore missed completely. The main point of this exercise wasn't to match exactly the landslide-area probability distribution, but to have a physically realistic distribution on which to base our modelling. To predict when and where a landslide will occur is one of the main challenges for calculating post-failure soil loss in data-poor regions. We exploited the correlation between the measured data and Malamud's distribution through combination with Monte Carlo simulation to analyse the effects of mass movements on soil erosion by water.

Assuming the validity of the proposed inverse-gamma function for calculating the probability distribution of landslide areas we implemented a simple script (based on SemAP) in MATLAB language. Starting from a **::scalar positive::** [12] number to represent the number of landslides that occurred in the catchment, we then calculate the number of landslides $\delta\ N_L(h)$ in the h-th class of landslides. Each class is a

[12] http://mastrave.org/doc/mtv_m/check_is#SAP_scalar_positive

::categorical-interval::[13] which includes all the landslides with an area from $A_L(h)$ to $A_L(h+1)$. The classes thus form a partition of **::contiguous - interval::**[14] s in [0;A_L(*hmax*)] whose values are found from:

$$\delta N_L(h) = \int_{A_L(h)}^{A_L(h+1)} p(A_L) dA_L \ , \tag{6.2}$$

In order to evaluate the effect of the post-failure changes on the soil erosion rates in the catchment, we applied the Monte Carlo method twice. Once to randomly determine the location of a landslide, and a second time to sample the Malamud distribution to assign its size. The Monte Carlo simulation was also implemented in the MATLAB language following the SemAP paradigm and exploiting the potentiality offered by the Mastrave Library (de Rigo, 2012a) whose tools were largely used within the code.

To be more explicit: considering Y as a random variable distributed according to a given probability distribution, it is possible to generate n pseudo-random instances $Y_1,\ldots, Y_n$ with the same distribution . This may be accomplished with a classical Monte Carlo extraction. Let us define $f(\cdot)$ as a certain function of Y which is implemented, within the SemAP paradigm, as a D-TM transforming an instance of Y into the desired output data. Suppose we are interested in computing the integral A of $f(\cdot)$ over a given domain . This implies considering the probability density function $\pi(\cdot)$ of Y over :

$$A = \int_{\Omega} f(Y) \cdot \pi(Y) dY, \qquad \begin{matrix} Y \in \Omega \\ Y \sim \Phi \\ \pi(Y) density\ function\ of\ \Phi\ in\ Y \\ such\ that\ \int_{\Omega} \pi(Y) d(Y) = 1 \end{matrix} \tag{6.3}$$

Numerically, it is possible to approximately estimate A by exploiting the n Monte Carlo instances $Y_1,\ldots, Y_n$ as

[13] http://mastrave.org/doc/mtv_m/check_is#SAP_categorical-interval
[14] http://mastrave.org/doc/mtv_m/check_is#SAP_contiguous_interval

$$A \approx \hat{A}_n = \frac{1}{n} \sum_{run=1}^{n} f(Y_{run}), \quad \forall\, run,\, Y_{run} \sim \Phi \ , \tag{6.4}$$

where Y_{run} is the run-th instance of Y corresponding to the run-th Monte Carlo iteration. From the law of large numbers, if $n \to \infty$, $\hat{A}_n \to A$. In our particular application, $\hat{A}_n$ is the average over n runs of simulated landslides; in each of them the total erosion by water $f(\cdot)$ is computed for the particular array of landslides Y_{run}. The n arrays of simulated landslides are the basis for $f(\cdot)$ to estimate the corresponding post-landslide soil erosion. Each landslide occurring in the run-th simulation has an area distributed according to $\bar{p}(\cdot)$. This defines $\pi(\cdot)$ as the probability density function with which each *run*-th array of landslides is distributed. The Monte Carlo simulation was iterated 1,000 times. A more robust approach would have been based on 10000 iterations but due to the very high computational time require by the script we decided to reduce the number of iterations to 1000. For each of the iterations the post-failure changes in soil erosion were calculated and compared with the pre-failure estimates.

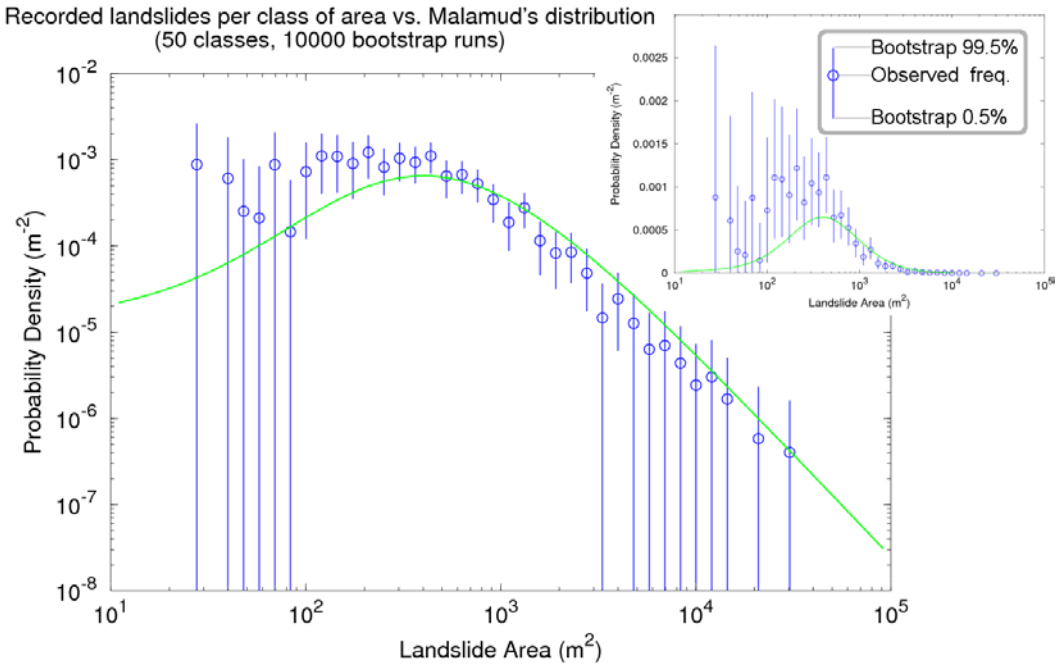


Figure 6.3: Dependence of the landslide probability densities on landslide area for the measured set of data (blue) and for Malamud's distribution (green). The probability density is given on logarithmic and semi-logarithmic scale. A bootstrap analysis was performed to assess the uncertainty of the measured data.

The **::matrix::** representing the cover management factor of the e-RUSLE model was calculated using the 5x5 metre resolution land cover map of the study site, produced by CNR-IRPI of Bari using ASTER satellite multi-spectral imagery and published in (Wasowski et al., 2010) (see paragraph 3.5). The map is not freely available but the CLC (EEA, 2006) is a valid open access alternative. The post-failure changes in vegetation cover were used within the model for estimating the effect of mass movement on soil erosion.

Important processes, as the soil armouring, the spatial distribution of the reworked sediments and the changes in local topography play an important role in soil erosion processes (see paragraph 1.4). Unfortunately, the few available researches for better understanding the evolution of a post-failure slope profile (Acharya, 2011) jointly with the lack of detailed data related to the soil characteristics of the study site led to consider, within the proposed approach, only the post- failure changes affecting the vegetation cover. Because of the use of simplified equations to calculate the *K* factor, it would be impossible to properly consider the effect of local changes of this factor on post-failure soil erosion. However, it would be interesting to estimate the effect of changes in *K* factor by simulating a range of different possible fluctuations of the *K* factor values between pre- and post-failure conditions. This is something that will need further investigation.

Because of the modular modelling architecture (Figure 6.1), the module that calculates the pre-failure C factor can be used as a link between our model and other approaches for measuring different land disturbance effects on soil erosion. The post-failure vegetation cover results were only partially altered by the slow mass movements that characterize this catchment (see Figures 1.2 and 3.4). As locally the slide surface may also remain unchanged, we introduced into the model a value representing the post-failure percentage of bare soil. By analysing the landslide dataset, the available pictures, satellite images and accounting for all the information collected during the field survey carried out within the study area, the percentage of the post-failure bare soil cover was estimated to be not less than 20% of the landslide area. For each of the pixels of the modelled landslides in each of the 1,000 Monte Carlo iterations, the **::scalar positive:: ::proportion::** of bare soil (Bsp) was therefore

randomly determined in the range 0.2 - 1. BSp is selected to follow a maximum entropy distribution in the aforementioned range. If X is a continuous random variable with probability density $p(x)$, then the differential entropy of X is defined as:

$$H(X) = -\int_{0.2}^{1} p(x)\ log\, p(x) dx \tag{6.5}$$

This is simply achieved by sampling the BSp value for each pixel with a Monte Carlo random sampling from a uniform distribution in [0.2 1].

6.3. Results and discussion

Table 6.1 shows the results of the Monte Carlo simulations. We replaced the mean values obtained by applying equation 6.4, with the median, because it is more stable in that it is only marginally affected by extreme values (Hampel et al., 1986). By analysing the median on 1,000 simulations of the cumulated pre-failure and post-failure soil erosion, an increase of 20% of the within catchment total soil loss was estimated. The total soil erosion predicted for the catchment by applying the e-RUSLE model was of $\sim$8000 tons per year. The post-failure soil erosion estimated by applying the presented modelling architecture was over 9500 tons per year.

The post-failure soil erosion rate in areas where landslides occurred is, on average, around 3.5 times the pre-failure value, passing from around 800 to more than 2700 tons per year over an area covering 9.4% of the study site.

A bootstrap analysis (Efron and Tibshirani, 1993; Efron, 1982) based on 10,000 runs was performed in order to assess uncertainty. This analysis was performed exploiting the module 'mbootstrap_idx' of the Mastrave modelling library (de Rigo, 2012a). The analysis of the changes in the rate of soil erosion due to the landslide occurrence predicted by the model shows post-failure increases in soil loss of approximately 1700 tons per year (bootstrap p <= 0.05). This corresponds to an increase of around 22% of the total soil erosion. We also analysed the extension of the area affected by slope instability. The bootstrap analysis shows that in each simulation at least 76 hectares, corresponding to around 8.5% of the catchment, are affected by landslide

activity (bootstrap ≤ 0.05). By comparing this value with the area that presented slope instability in 2006 (around 55 hectares), the applied methodology seems to result in a slight overestimate. The graph in figures 6.2 and 6.3 shows that Malamud's distribution seems to underestimate the number of small landslides (< 300 m^2). Nevertheless, the probability density distribution for the Rocchetta landslides from 2006 is in line with those reported by Malamud et al. (2004). The Malamud distribution was here purposedly applied with an unsupervised model, hence without any custom fine tuning to adapt the original distribution (Malamud et al., 2004) to the specific peculiarities of the catchment (for example, to better fit the distribution to small landslides). Although in the study area the data might allow some better tuning to be achieved in future refinements of the application, this approach is general and reusable even in more severely data-poor regions. The model is in its early developmental phase and fine-tuning the fit of the Malamud distribution to small landslides should help to improve the model predictions. However, for better evaluating the limits or the robustness of the proposed inverse-gamma distribution or of a modified version, further data would be necessary. The bootstrap analysis, with 10,000 runs, performed on the measured data (Figure 6.3) shows the uncertainty associated with a single year landslide dataset is too high to extrapolate different parameter values. A more detailed analysis based on datasets covering a longer time interval would help to improve the applied methodology. An additional source of error contributing to the predictions, which needs further investigation, arises from the selection of the model for estimating soil erosion and its running with limited data: thus, there is considerable scope for errors in the predictions to be strongly linked to this simplification.

Because the capacity to estimate the changes in soil erosion from landslide activity is largely dependent on the quality of the available datasets, the applied methodology broadens the possibility of a quantitative assessment of these effects in data-poor regions. The obtained results, even considering a possible overestimation, confirm the important role of mass movements on soil erosion and the consequent necessity to better integrate these processes into soil erosion modelling.

Table 6.1: Bootstrap analysis of the modelling results. The bootstrap analysis, based on 10000 runs, shows the bootstrap cumulated distribution of the pre-and post-failure soil erosion within the area affected by landslide activity.

Quantile	Pre-failure soil loss (t)	Post-failure soil loss (t)	Estimated landslide activity area (ha)
5%	744.7	2530.3	76.6 (8.4%)
25%	799.2	2762.3	84.4 (9.2%)
50%	828.7	2773.3	85.5 (9.4%)
75%	843.4	2896	87.1 (9.6%)
95%	854.6	3005	88.9 (9.8%)

6.4. Conclusions

A new method for empirically estimating the importance and extent of landslides on soil erosion losses in data-poor regions has been developed. This has been achieved by sampling the frequency-size landslide distribution proposed by Malamud et al. (2004), and stochastically distributing the landslide location across the catchment. Given the increasing threat of soil erosion all over the world and the implications this has on future food security and soil and water quality, an in-depth understanding of the rate and extent of soil erosion processes is crucial. Each year, on average, between 8.5 and 10% of the catchment shows evidence of landslide activity that is responsible for a mean increase in the total soil erosion rate between 22 and 26% above the pre-failure estimate. These results confirm the potential importance of integrating the landslide contribution into soil erosion modelling. While this approach clearly has limitations, the proposed approach can be seen as a first attempt to assess the landslide-erosion interaction in areas with limited data.

The proposed modelling approach is also suitable in applications having a wider spatial extent and to be potentially implemented in a transdisciplinary context. For example, the relevant effect of wild fires on soil erosion and landslide susceptibility (Di leo et al., 2013; de Rigo et al., 2013) could be modelled with a higher reliability

integrating the proposed approach. As stated in de Rigo et al. (2013), wildfires can considerably increase soil erosion by water and landslide susceptibility. The changes in landslide susceptibility may in turn affect soil erosion. In general, considering the modelling architecture (Figure 6.1), if the module that calculates the pre-failure C factor value would provide the layer altered by a different disturbance (e.g. wild_fires or outbreak of pests), the presented modelling architecture could then be applied for estimating the indirect effect of these disturbances on soil erosion, provided a new landslide susceptibility map, that considers the altered vegetation cover, is produced. Despite the promising results, further research is still required to fully assess the reliability and therefore applicability of this method in coupled landslide soil erosion modelling.

7. Summary and recommendations

The methods presented and developed throughout this PhD thesis and the corresponding results are summarized in this final chapter. Here, the most important conclusions and recommendations, the main limits of the approaches used and a focus on further research are presented.

7.1. Summary

This research was conducted with the main objective to better integrate and quantify the role of water-induced shallow landslides within soil erosion processes with a particular focus on data-poor conditions. To fulfill the objectives, catchment-scale studies on soil erosion by water and shallow landslides were conducted.

A new semi-quantitative method, based on an ensemble approach, to predict landslide susceptivility by combining deterministic and probabilistic approaches was proposed and validated jointly with an integrated shallow landslide soil erosion modelling approach.

Problem definition, aim and objectives

Soil erosion and mass movements are part of a system of multiple interacting processes, with both being visible expressions of critical instabilities affecting a territory. Soil erosion by rainfall and runoff is one of the main soil threats in Europe (section 1.1.). Rainfall-induced landslides contribute directly to soil erosion by the displacement of material and indirectly by destroying the vegetation cover that reduces surface flow velocities.

In most of the cases, the slope instability affecting the study area is linked with re-activation of dormant phenomena triggered by rainfall (section 3.3). Because natural resources are intrinsically entangled in complex networks there is a growing awareness of the importance to better quantify and understand their connections in order to develop appropriate management policies.

The main objectives of this study were to:

1) improve the estimation and prediction of soil erosion by water in catchments affected by shallow landslides triggered by water.
2) Develop a robust approach to reduce the uncertainty in shallow landslide susceptibility assessments in data-poor regions.

Because it is impractical to measure soil loss across whole landscapes using experimental plots, soil erosion markers or sampling river sediment load, a strong emphasis has been placed on modelling as a tool to assess soil erosion and the effect of mass movements triggered by water.

The existing integrated modelling approaches and their limits

Numerous models exist for predicting shallow landslides and soil erosion by water, but relatively few attempts were done to model soil erosion by water and shallow landslides using an integrated approach (section 2.4.1). The existing qualitative or quantitative models presented in paragraph 2.4 (e.g SHETRAN, SIBERIA, PSIAC and WEPP-SLIP), are based on completely different approaches and show many different limits. In most of these models post-failure scenarios are not considered. Only in WEPP-SLIP post failure soil erosion is explicitly considered (post failure long-term sediment yield) but the soil redistribution within the model needs to be further developed (section 2.4.2).

Physically based models can be easily modified to consider post-failure soil erosion changes because of their use of local terrain characteristics and dynamic hydrological models. They also have the possibility to easily consider landslides temporal probability of occurrence. Unfortunately these models are also less suitable to be applied in data-poor conditions.

With the use of statistical methods it is possible to overcome some of the lack of detailed input data. Unfortunately this approach does not account for the temporal aspect of mass movements. For improving soil erosion estimation considering mass movements within the erosion process it is necessary to estimate not only where but also when a landslide will occur along with the size of the event.

In chapter 6 a new approach is presented to consider the frequency-area distribution of landslides in order to quantify the size and number of landslides associated to precipitation events in data-poor regions and to quantify their contribution on soil erosion. This was achieved by sampling the frequency-size landslide distribution proposed by Malamud et al. (2004), and stochastically distributing (using a Monte Carlo simulation) the location of landslides across the catchment.

The study site

The study area is situated in southern Italy in the Daunia Appennines of the Puglia region, within the municipal territory of Rocchetta Sant'Antonio. It covers an area of almost 10 km2 (section 3.1). This area is highly susceptible to landslide activity (Iovine et al., 1996; Magliulo et al., 2008) with a consequent negative impact on the local economy (Wasowski et al., 2010). The area neighbouring to the north-west of the Rocchetta Sant'Antonio territory presents a landslide frequency exceeding 20% for the overall area (Mossa et al., 2005; Wasowski et al., 2007, 2010, 2012) (section 3.3). Soil erosion is also widespread and the severity is largely determined by the combination of tillage practices and the high erodibility of the clay-rich units from which some of the local soils are derived (Lamanna et al., 2009).

Within the catchment it is possible to distinguish four major classes of land use (agricultural soils, woodland, pastures and grassland) and three dominant lithologies (limestone, sandstone and clay-shales) (section 3.2, 3.5). Slope angles are on average approximately 10 degrees with peak slope angles rarely exceeding 25 to 30 degrees. An ephemeral drainage network is fed by precipitation during the autumn-winter period when some 600 to 750 mm of rainfall is common (Wasowski et al., 2010). The area is characterized by a Mediterranean sub-humid climate (section 3.4).

This area has been studied for some years resulting in a database of information (including topography and Digital Elevation Model (DEM), site investigation and geotechnical test data, soil distribution and land use maps) (Wasowski et al., 2007, 2010; Mossa et al., 2005).

Modelling soil erosion in data poor regions

Despite numerous efforts, the prediction value of existing models is still limited, especially at regional scale or in data-poor areas, because a systematic knowledge of local climatological and soil parameters is often unavailable. After a first attempt in applying the RMMF model in the study site (section 4.2.2) for measuring soil loss, a new approach for modelling soil erosion in data-poor conditions was proposed. It is based on the joint use of a low-data-demanding models and innovative techniques for better estimating model inputs. This modelling architecture is based on semantic array programming paradigm with a strong effort towards computational reproducibility (section 2.5). An extended version of the RUSLE model was implemented merging different empirical rainfall-erosivity equations within a climatic ensemble model and adding a new factor for a better consideration of soil stoniness (section 4.2.3.2). The map of the soil erosion rates affecting the Rocchetta Sant'Antonio catchment was produced through the use of publicly available data sets and empirical relationships.

A multi-scale robust modelling approach for estimating landslide susceptibility

Landslide susceptibility assessment of a territory is fundamental to prevent landslide occurrence, its spatial distribution is also one of the main challenges in landslides forecasting. Despite the many different approaches that have been tested and developed (van Westen et al., 1997), landslide susceptibility assessment still remains a challenge. A semi-quantitative method that combines heuristic, deterministic and probabilistic approaches is here proposed for a robust catchment scale assessment in data-poor conditions (section 5.4). A set of different susceptibility zonation maps (Figure 5.12) was aggregated exploiting a modelling ensemble. Each susceptibility zonation has been obtained by applying heterogeneous statistical techniques as logistic regression (LR) (Cox, 1958), relative distance similarity (RDS) (de Rigo, 2015; de Rigo et al., 2013a; Bosco et al., 2015), artificial neural network (ANN) (section 5.1.2) and two different landslide susceptibility techniques (SINMAP (Pack et al., 1998, 2005) and TransSlide (Bosco et al., 2013)) based on the infinite slope stability model (section 5.1.1).

The ensemble approach is a reproducible data-transformation model applied to the results of the array of models and is based on RDS. The sequence of data-transformation models has been enhanced following the semantic array programming paradigm. The good performance of the ensemble model (section 5.5), when compared with the single techniques, make this method suitable to be applied in data-poor conditions where the lack of proper calibration and validation data can affect the application of physically based or conceptual models (Bosco et al., 2013). Because of the uncertainty in selecting a single suitable method for modelling spatial landslide susceptibility in areas characterized by data weakness, the applied ensemble method can potentially result in a less uncertain zonation (Bosco et al., 2013). This catchment scale methodology may be exploited for analysing the potential impact of landscape disturbances.

A coupled architecture of soil erosion by water and water induced shallow landslides

Given the aim and constraints of the study area, a robust model that couples hydrology with stability has been developed (section 6.2.2). This chapter proposes a new integrated methodology for a robust assessment of soil erosion rates in data-poor areas affected by landslide activity by combining heuristic, empirical and probabilistic approaches. This proposed methodology is based on the geospatial semantic array programming paradigm (section 6.2.1) and has been implemented on a catchment scale methodology using GIS, spatial analysis tools and GNU Octave. The integrated data-transformation model relies on a modular architecture (Figure 6.1), where the information flow among modules is constrained by semantic checks. In order to improve computational reproducibility. The proposed modelling architecture is flexible enough for future transdisciplinary scenario analysis to be more easily designed.

By analysing modelling results within the study catchment, each year, on average, mass movements, are responsible for a mean increase in the total soil erosion rate between 22 and 26% over the pre-failure estimate. The post-failure soil erosion rate in areas where landslides occurred is, on average, around 3.5 times the pre-failure value (section 6.3). These results confirm the importance to integrate landslide

contribution into soil erosion modelling. The model is in its early developmental phase and fine tuning the fit of the model to small landslides should help to improve the model prediction capacity. Because the estimation of the changes in soil erosion from landslide activity is largely dependent on the quality of available datasets, this methodology broadens the possibility of a quantitative assessment of these effects in data-poor regions.

The approach here proposed is also suitable to be applied on a wider spatial extent and in a transdisciplinary context. For example, due to the relevant effect of wildfires on soil erosion and mass movements (Di leo et al., 2013; de Rigo et al., 2013) (section 6.4) this modelling architecture could simplify future integrated analysis of the potential impact of wildfires on sediment transport from erosion and landslides triggered by water.

7.2. Recommendations for further research and investigations

The research conducted in this thesis has resulted in new and useful modelling techniques , which produced interesting results and conclusions on water-induced soil erosionand shallow landslides. Although the preliminary results obtained applying the new architecture for considering the impact of mass movement on soil erosion in data-poor regions (chapter 6) are promising, further research is required before this method can be applied by the scientific community and relevant authorities with any level of confidence.

Consideration of, and integrating within the model, post-failure changes in topography and soil characteristics (e.g. soil armouring (Acharya, and T.A. Cochrane, 2008) is fundamental for increasing the predictive capacity of the model. Also a better estimation of the bare soil exposed within a landslide is also fundamental for improving our model. It would also be worthwhile to fine tuning the Malamud distribution (Malamud et al., 2004) to the data to better fit the distribution to small landslides as suggested in section 6.3. For obtaining more reliable results, and more robust estimates of the effects of landslides on soil and vegetation cover, it will be also necessary to focus attention on producing a less uncertain zonation of the spatial

probability of the landslide susceptibility in areas characterized by low data availability (Bosco et al., 2013). This highlights the need to fully quantify model uncertainty to communicate where poor model fits exist (Bosco et al., 2017).
The ensemble method proposed in chapter 5, being a supervised method, slightly improves the best model in the array of outputs. The high performance showed by the RDS approach could be linked with the criterion used for the selection of the training and testing set of data. The possible presence of bias in using a similar technique for selecting the data and calculating the landslide susceptibility needs to be further investigated.

One of the main limits of the proposed approach is that the e-RUSLE (chapter 4) does not consider erosion processes such as gully erosion, that locally may cause very high soil losses (Poesen et al., 2003; Mathys et al., 2003; Collinet and Zante, 2005). The possibility to incorporate gully erosion within the modelling architecture should be investigated. In applying the e-RUSLE model, there is also a high probability for some of the model results to be overestimated. The R factor uncertainty and the presence of areas having a stoniness value much higher than reported can be at the basis of many of the uncertain estimations. The rainfall erosivity component of the models was estimated by ensembling an array of erosivity maps based on seven emipircal equations from literature. These equations were selected from the many available in the literature for their reasonable mathematical structure and set of covariate variables. Additional equations may be analysed and tested to complement the array of rainfall erosivity estimates at the basis of the final aggregated ensemble erosivity. Further investigations may also be recommendable on the key role of land cover changes and misclassifications (Bosco et al., 2015), inherited from the available land cover products, since these components of uncertainty can locally have an high impact on the estimated soil erosion rate.

References

Acharya, G., Cochrane, T.A., 2008. Rainfall induced shallow landslides on sandy soil and impacts on sediment discharge: a ume based investigation. The 12th International Conference of International Association for Computer Methods and Advances in Geomechanics (IACMAG), Goa, India. [Online]. Available: ir.canterbury.ac.nz/bitstream/10092/3134/1/12612367_Paper-IACMAG2008.pdf .

Acharya, G., Cochrane, T.A., Davies, T., Bowman, E., 2009. The influence of shallow landslides on sediment supply: a flume-based investigation using sandy soil. Engineering Geology 109, 161-169.

Acharya, G., 2011. Analysing the interactions between water-induced soil erosion and shallow landslides. Phd Thesis, University of Canterbury.

Agnesi, V., Arziello, G., Aucelli, P., Baglioni, A., Bettucci, C., et al., 2007. Rapporto sulle frane in Italia. Il Progetto IFFI: metodologia, risultati e rapporti regionali. APAT, Rapporti 782007," p. 681.

Aksoy, H., Kavvas, M. L., 2005. A review of hillslope and watershed scale erosion and sediment transport models. Catena 64(2), 247–271.

Aleotti, P., Chowdhury, R., 1999. Landslide hazard assessment: summary review and new perspectives. Bulletin of Engineering Geology and the Environment 58, 21–44.

Anagnostopoulos, G.G., Fatichi, S., Burlando, P., 2015. An advanced process-based distributed model for the investigation of rainfall-induced landslides: the effect of process representation and boundary conditions. Water Resour. Res., 51 (9), 7501-7523.

Anderson, S.A., Sitar, N., 1995. Analysis of rainfall-induced debris flows. Journal of Geotechical Engineering ASCE 121(7), 544–552.

Andriani, G.F., Parise, M., Spagnoletta, A., Walsh, N., 2009. Evaluating landslide susceptibility in hillslopes of the Daunia Apennines (Apulia, Italy). Geophysical Research Abstract 11, EGU 2009-4896.

Andriani, F.A., Diprizio, G., Pellegrini, V., 2015. Landslide susceptibility of the La Catola torrent catchment area (Daunia Apennines, Southern Italy): a new complex multi-step approach. Engineering Geology for Society and Territory 5, 387-392.

Angima, S. D., Stott, D. E., O'Neill, M. K., Ong, C. K., Weesies, G. A., 2003. Soil erosion prediction using RUSLE for central Kenyan highland conditions. Agr. Ecosyst. Environ. 97, 295–308.

Arnold, J.G., Williams, J.R., Griggs, R.H., Sammons, N.B., 1990. SWRRB a basin scale simulation model for soil and water resources management. A&M Press, Texas, 142p.

Arnold, J. G., R., Srinivasan, R. S., Muttiah, J. R. Williams., 1998. Large-area hydrologic modeling and assessment: part I., model development. American Water Resour. Assoc. 34(1), 73-89.

Arnold, J.G., Fohrer, N., 2005. SWAT2000: current capabilities and research opportunities in applied watershed modeling. Hydrological processes, 19(3), 563-572.

Ashtawy, H.M., Mahapatra, N.R., 2015. BgN-Score and BsN-Score: bagging and boosting based ensemble neural networks scoring functions for accurate binding affinity prediction of protein-ligand complexes. BMC bioinformatics 16(4), S8.

Ayalew, L., Yamagishi, H., 2005. The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, central Japan. Geomorphology 65, 15–31.

Baffaut, C., Nearing, M.A., Govers, G., 1998. Statistical distributions of soil loss from runoff plots and the WEPP model simulations. Journal of Soil Science Society America 62, 756-763.

Bakker, M.M., Govers, G., Rounsevell, M.D.A., 2004. The crop productivity-erosion relationship: an analysis based on experimental work. Catena 57, 55-76.

Bakker, M.M., Govers, G.,, Kosmas, C., Vanacker, V., Oost, K., et al., 2005. Soil erosion as a driver of land-use change. Agriculture, Ecosystems & Environment 105(3), , 467-481. doi: 10.1016/j.agee.2004.07.009 .

Bathurst, J.C., Bovolo, C.I., Cisneros, F., 2010. Modelling the effect of forest cover on shallow landslides at the river basin scale. Ecological Engineering 36, 317-327.

Baum R.L., Savage W.Z., Godt J.W., 2008. TRIGRSA FORTRAN: program for tran- sient rainfall infiltration and grid based regional slope stability analysis. Version 2.0. U.S. Geol. Surv., Open File Rep. 2008-1159, 74 p.

Beasley, D.B., Huggins, L.F., Monke, E.J., 1980. Answers a model for watershed planning. T. ASABE 23(4), 938–944.

Beck, M.B., Young, P.C., 1975. A dynamic model for BOD-DO relationships in a non-tidal stream. Water Research 9, 769-776.

Beck, M. B., Jakeman, A. J., McAleer, M. J., 1995. Construction and evaluation of models of environmental systems. In: Modelling Change in Environmental Systems, edited by Beck, M.B, McAleer, M. J., John Wiley and Sons, 3–35.

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

Besler, H., 1987. Slope properties, slope processes and soil erosion risk in the tropical rain forest of Kalimantan Timur ŽIndonesian Borneo. Earth Surface Processes and Landforms 12, 195–204.

Beven, K.J., Kirkby, M.J., 1979. A physically based variable contributing area model of basin hydrology. Hydrol. Sci. Bull. 24(1), 43-69.

Beven, K., Robert E., 2004. Horton's perceptual model of infiltration processes. Hydrological processes 18(17), 3447-3460.

Bingner, R.L., Theurer, F.D., 2001. AGNPS 98: a suite of water quality models for watershed use. Proceedings of the 7th Federal Interagency Sedimentation Conference, Reno, NV, 25-29 March 2001, VII-1, VII-8.

Blo¨schl, G., Sivapalan, M., 1995. Scale issues in hydrological modelling: a review. Hydrological Processes 9, 251–290.

Boardman, J., Poesen, J., 2006. Soil erosion in Europe: major processes, causes and consequences. In: Soil Erosion in Europe, edited by Boardman, J., Poesen, J., Wiley, C., 479–487.

Boardman, J., 2006. Soil erosion science: reflections on the limitations of current approaches. Catena 68(2), 73-86.

Boardman, J., 2007. Soil erosion: the challenge of assessing variation through space and time. In: Goudie, A.S., Kalvoda, J. (eds.) Soil erosion: the challenge of assessing variation through space and time. Nakladatelsti P3K, Prague, 205-220. ISBN: 978-80-903584-6-1.

Bordoni, M., et al., 2015. Hydrological factors affecting rainfall-induced shallow landslides: from the field monitoring to a simplified slope stability analysis. Eng. Geol. 193, 19–37.

Bosco, C., Rusco, E., Montanarella, L., Oliveri, S., 2008. Soil erosion risk assessment in the alpine area according to the IPCC scenarios. In: Threats to Soil Quality in Europe, edited by Toth, G., Montanarella, L., Rusco, E., EUR 23438 EN, 47–58.

Bosco, C., Rusco, E., Montanarella, L., Panagos, P., 2009. Soil erosion in the Alpine area: risk assessment and climate change. Studi Trent. Sci. Nat. 85, 119–125.

Bosco, C., de Rigo, D., Dewitte, O., Montanarella, L., 2011. Towards the reproducibility in soil erosion modelling: a new Pan-European soil erosion map. Conference on Applied Soil Science “Soil Science in a Changing World”, 18-22

September 2011, Wageningen, The Netherlands. Author's version DOI:10.6084/m9.figshare.936872 arXiv:1402.3847.

Bosco, C., de Rigo, D., 2013. Land cover and soil erodibility within the e-RUSLE model, Sci. Top. Focus, MRI-11b13, Maieutike Research Initiative, doi:10.6084/m9.figshare.856670.

Bosco, C., de Rigo, D., Dijkstra, T., Sander, G., Wasowski, J., 2013. Multi-Scale robust modelling of landslide susceptibility: regional rapid assessment and catchment robust fuzzy ensemble, IFIP Adv. Inf. Commun. Technol. 413, 321–335, doi:10.1007/978-3-642-41151-9 31.

Bosco, C., de Rigo, D., Dewitte, O., 2014. Visual Validation of the e-RUSLE Model Applied at the Pan-European Scale, Sci. Top. Focus, MRI-11a13, Maieutike Research Initiative, doi:10.6084/m9.figshare.844627.

Bosco C., Sander G ., 2015. Estimating the effects of water-induced shallow landslides on soil erosion. IEEE Earthzine 7, 910137. (doi:10.1101/011965).

Bosco, C., de Rigo, D., Dewitte, O., Poesen, J., Panagos, P., 2015. Modelling soil erosion at european scale: towards harmonization and reproducibility. Natural Hazards and Earth System Sciences 15(2), 225-245. doi:10.5194/nhess-15-225-.

Bosco C., Alegana, V., Bird, T. et al., 2017. Exploring the high-resolution mapping of gender-disaggregated development indicators. J. R. Soc. Interface 14: 20160825. http://dx.doi.org/10.1098/rsif.2016.0825.

Bourne S. J., Willemse E. J. M., 2001. Elastic stress control on the pattern of tensile fracturing around a small fault network at Nash Point, UK. Journal of Structural Geology 23, 1753–1770.

Bovolo, C.I, Bathurst, J.C., 2012. Modelling catchment-scale shallow landslide occurrence and sediment yield as a function of rainfall return period. Hydrol. Process., 579-596, doi: 10.1002/hyp.8158.

Box, J. E., 1981. The effects of surface slaty fragments on soil erosion by water. Soil Sci. Soc. Am. J. 45, 111–116.

Brandt, C.J., 1990. Simulation of the size distribution and erosivity of raindrops and throughfall drops. Earth Surface Processes and Landforms 15, 687–698.

Bras, R. L., Tucker, G. E., Teles, V., 2003. Six myths about mathematical modeling in geomorphology, In: Prediction in geomorphology, edited by Wilcock, P.R. and Iverson, R. M., American Geophysical Union, Washington D.C. , 63–79.

Brenning, A., 2005. Spatial prediction models for landslide hazards: review, comparison and evaluation. Nat. Hazards Earth System Sci. 5(6), 853-862.

Brevik, E.C., Calzolari, C., Miller, B.A., Pereira, P., Kabala, C., et al., 2016. Soil mapping, classification, and pedologic modeling: History and future directions. Geoderma, 264, 256-274.

Bryan R.B, 2000. Soil erodibility and processes of water erosion on hillslope. Geomorphology 32, 385-415.

Bucknam, R.C., Coe, J.A., Chavarria, M.M., Godt, J.W., Tarr, A.C., et al., 2001. Landslides triggered by hurricane Mitch in Guatemala – Inventory and Discussion. US Geological Survey Open File Report, 01– 443.

Bui, D.T, Tuan, T.A., Klempe, H. et al., 2016. Spatial prediction models for shallow landslide hazards: a comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree Landslides, 13: 361. https://doi.org/10.1007/s10346-015-0557-6.

Burton, A., Bathurst, J.C., 1998. Physically based modelling of shallow landslide sediment yield at a catchment scale. Environ Geol. 35(2/3), 89-99.

Cardinali, M, Ardizzone, F, Galli, M, Guzzetti, F, Reichenbach, P. 2000. Landslides triggered by rapid snow melting: the December 1996–January 1997 event in central Italy. Proceedings 1st Plinius Conference on Mediterranean Storms. Claps, P, Siccardi, F (eds). Bios: Cosenza, 439 – 448.

Casagrandi, R., Guariso, G., 2009. Impact of ICT in Environmental Sciences: a citation analysis 1990-2007. Envir. Modelling & Software 24(7), 865-871.

CEC, 2006. Proposal for a directive of the European Parliement and of the Council establishing a framework for the protection of soil and amending Directive 2004/ 35/EC. Brussels, 22.9.2006, COM(2006) 232 final, 2006/0086 .

Cerdan, O., Souchere, V., Lecomte,V., Courturier,A., Le Bissonnais, Y., 2001. Incorporating soil surface crusting processes in an expert-based runoff model: sealing and transfer by runoff and rosion related to agricultural management. Catena 46, 189–205.

Cerdan, O., Govers, G., Le Bissonnais, Y., Van Oost, K., Poesen, J., et al., 2010. Rates and spatial variations of soil erosion in Europe: a study based on erosion plot data. Geomorphology 122, 167–177. doi:10.1016/j.geomorph.2010.06.011.

Cerf, V. G., 2012. Where is the science in computer science? Commun. ACM 55(5). doi:10.1145/2347736.2347737.

Cernica, J.N., 1995. Geotechnical engineering: soil mechanics. John Wiley and Sons, New York, 453 p.

Chen H., Lee C.F., 2003. A dynamic model for rainfall-induced landslides on natural slopes. Geomorphology 51(4), 269-288.

Chirico, G.B., Western, A.W., Grayson, R.B., Bloschl, G., 2005. On the definition of the flow width for calculating specific catchment area patterns from gridded elevation data, Hydrol.Process. 19, 2539-2556. doi: 10.1002/hyp.5730 .

Cho S.E., Lee S.R., 2002. Evaluation of surficial stability for homogeneous slopes considering rainfall characteristics. Journal of geotechnical and geoenvironmental engineering 128(9), 756-763.

Ciarcia, S., Di Nocera, S., Matano, F., Torre, M. 2003. Evoluzione tettono-sedimentaria e paleogeografica dei depocentri 'wedge-top' nell'ambito del 'foreland basin system' pliocenico dell'Appennino meridionale (settore Irpino–Dauno). Bollettino della Società Geologica Italiana 122, 117–137.

Cihlar, J., 1987. A methodology for mapping and monitoring cropland soil erosion. Can. J. Soil Sci. 67, 433–444.

CSIRO: TOPOG Home page, http://www-data.wron.csiro.au/topog/, 2017.

Cochrane, T.A., Acharya, G., 2011. Changes in sediment delivery from hillslopes a_ected by shallow landslides and soil armouring. Journal of Hydrology (New Zealand) 50(1), 5-18.

Collinet, J., Valentin, C., 1984. Evaluation of factors influencing water erosion in West Africa using rainfall simulation. AHS-AISH P. 144, 451–461.

Collinet, J. Zante, P., 2005. Analyse du ravinement de bassin versant à retenue collinaire sur sols à fortes dynamiques structurales (Tunisie). Geomorphologie 1, 61–74.

Comune di Rocchetta sant'Antonio, 2015. Dichiarazione Ambientale. Rev 01 del 27/02/2015.file:///C:/Users/claudiobosco/Downloads/Alleg_Delib.GC_n.20_del_04-05-2015-Dichiarazione%20ambientale%202015%20(1).pdf

Corstanje, R., et al. 2007. Scale dependent relationships between urease activity and soil organic carbon. European Journal of Soil Science 58, 1087-1095.

Corstanje, G.J.D., et al., 2008. Spatial variation of ammonia volatilization from soil and its scale dependent correlation with soil properties. Eur. J. Soil Sci. 59, 1260-1270.

Costa-Cabral, M., Burges, S.J., 1994. Digitale Elevation Model Network (DEMON): a model of flow over hillslopes for computation of contributing and dispersal areas. Water Resources Research 30(6), 1681-1692.

Costanzo, D., Chaco´n, J., Conoscenti, C, Irigaray, C., Rotigliano, E., 2014. Forward logistic regression for earth- flow landslide susceptibility assessment in the Platani river basin (southern Sicily, Italy). Landslides 11, 639–653. doi:10.1007/s10346-013-0415-3.

Cotecchia, F., Vitone, C., Cafaro, F., & Santaloia, F., 2006. The mechanical behaviour of intensely fissured high plasticity clays from Daunia. In Invited paper. In: Second international workshop on characterisation and engineering properties on natural soils, Singapore, 1975-2003 p.

Cotecchia, F., Santaloia, F., Lollino, P., Vitone, C., Mitaritonna, G., 2009. A research project for deterministic landslides risk assessment in southern Italy: methodological approach and preliminary results. In: Honjo et al. (eds) Geotechnical risk and safety, Taylor & Francis Group, London, 363–370.

Commission of the European Communities, 2006. Proposal for a directive of the European Parliament and of the Council establishing a framework for the protection of soil and amending directive 2004/35/EC, Brussels, COM(2006) 232 final.

CSIRO: TOPOG Home page, http://www-data.wron.csiro.au/topog/, 2017.

Cunningham, P., Delany, S.J., 2007. k-Nearest neighbour classifiers. Multiple Classifier Systems, 34(8), 1-17.

Dai, F.C., Lee, C.F., Ngai, Y.Y., 2002. Landslide risk assessment and management: an overview. Eng. Geol. 64(1), 65–87.

Dai, F.C., Lee, C.F., 2003. A spatiotemporal probabilistic modelling of storm-induce shallow landsliding using aerial photographs and logistic regression. Earth Surf. Process. Landforms 28, 527545, doi:10.1002/esp.456.

Dazzaro, L., Di Nocera, S., Pescatore, T., et al., 1988. Geologia del margine della catena appenninica tra il F. Fortore ed il T. Calaggio (Monti della Daunia–Appennino Meridionale). Memorie della Società Geologica Italiana 41, 411–422.

De Bello, F., LepŠ J. A. N., Sebastia, M.T., 2005. Predictive value of plant traits to grazing along a climatic gradient in the Mediterranean. Journal of Applied Ecology42(5), 824-833.

De Jong, E., Wang, C., Rees, H.W., 1986. Soil redistribution on three cultivated New Brunswick hillslopes calculated from 137Cs measurements, solum data and the USLE, Can. J. Soil Sci. 66, 721–730. doi:10.4141/cjss86-071.

De Jong, S. M., Paracchini, M. L., Bertolo, F., Folving, S., Megier, J.,de Roo, A.P.J., 1999. Regional assessment of soil erosion using the distributed model SEMMED and Remotely Sensed Data. Catena 37, 291-308.

Delmas, M., Cerdan, O., Mouchel, J.M., Garcin, M., 2009. A method for developing a large-scale sediment yield index for European river basins. Journal of Soils and Sediments 9(6), 613–626.

De Ploey, J., 1982. A Stemfow equation for grasses and similar vegetation. Catena 9, 139–152.

De Ploey, J., 1984. Stemflow and Colluviation: modelling and implications. Pedologie 34 ,135–146.

De Ploey, J., 1989. Erosional systems and perspectives for erosion control in European loess areas. Soil Technol. Ser. 1, 93-102.

De Ploey, J., Moeyersons, J., Goossens, D., 1995. The De Ploey erosional susceptibility model for catchments, ES . Catena 25(14), 269-314.

de Rigo, D., Castelletti, A., Rizzoli, A. E., Soncini-Sessa, R., Weber, E., 2005. A selective improvement technique for fastening neuro-dynamic programming in water resources network management. In: Zítek, P. (Ed.), Proceedings of the 16th IFAC World Congress. 38. International Federation of Automatic Control (IFAC), 7-12, special issue: "Proceedings of the 16th IFAC World Congress".

de Rigo, D., Bosco, C., 2011. Architecture of a Pan-European Framework for Integrated Soil Water Erosion Assessment, IFIP Adv. Inf. Commun. Technol. 359(34), 310–318, doi:10.1007/978-3-642-22285-6 34.

de Rigo, D., 2012a. Semantic Array Programming for Environmental Modelling: Application of the Mastrave Library. Int. Congress on Environmental Modelling and Software. Managing Resources of a Limited Plant, Pathways and Visions under Uncertainty, Sixth Biennial Meeting, 1167_1176.
http://www.iemss.org/iemss2012/proceedings/D3_1_0715_deRigo.pdf

de Rigo, D., 2012b. Semantic array programming with Mastrave – introduction to semantic computational modelling. http://mastrave.org/doc/MTV-1.012-1

de Rigo, D., 2012c. Integrated Natural Resources Modelling and Management: minimal redefinition of a known challenge for environmental modelling. Excerpt from the Call for a shared research agenda toward scientific knowledge freedom, Maieutike Research Initiative.

de Rigo, D., 2012d. Applying semantic constraints to array programming: the module “check_is” of the Mastrave modelling library, Mastrave project technical report, 2012, available at: http://mastrave.org/doc/mtv_m/check_is, last access: November 2014.

de Rigo, D., 2013. Software Uncertainty in Integrated Environmental Modelling: the role of Semantics and Open Science .Geophys. Res. Abstr. 15,13292+ doi:10.6084/m9.figshare.155701.

de Rigo, D., Barredo, J.I., Busetto, L., Caudullo, G., San-Miguel-Ayanz, J., 2013a. Continental-Scale Living Forest Biomass and Carbon Stock: a Robust Fuzzy Ensemble of IPCC Tier 1 Maps for Europe . IFIP Adv. Inf. Commun. Technol. 413.

de Rigo, D., Corti, P., Caudullo, G., McInerney, D., Di Leo, M., San Miguel-Ayanz, J., 2013b. Toward Open Science at the European Scale: Geospatial Semantic Array Programming for Integrated Environmental Modelling.Geophys. Res. Abstr. 15, 13245 + (2013b) doi:10.6084/m9. gshare.155703

de Rigo, D., Rodriguez-Aseretto, D., Bosco, C., Di Leo, M., San-Miguel-Ayanz, J., 2013c. An architecture for adaptive robust modelling of wildfire behaviour under deep uncertainty, IFIP Adv. Inf. Commun. Technol. 413, 367–380, doi:10.1007/978-3-642-41151-9_35.

de Rigo, D., 2015. Study of a collaborative repository of semantic metadata and models for regional environmental datasets' multivariate transformations. Ph.D. thesis, Politecnico di Milano, Milano, Italy. http://hdl.handle.net/ 10589/101044 , http://mfkp.org/INRMM/article/13769492

de Rigo, D., Bosco, C., San-Miguel-Ayanz, J., Houston Durrant, T., Barredo, J. I. et al., 2016a. Forest resources in Europe: an integrated perspective on ecosystem services, disturbances and threats. In: San-Miguel-Ayanz, J., de Rigo, D., Caudullo, G., Houston Durrant, T., Mauri, A. (Eds.), European Atlas of Forest Tree Species. Publications office of the European Union, Luxembourg, e015b50+. https://w3id.org/mtv/FISE-Comm/v01/e015b50

de Rigo, D., Caudullo, G., Houston Durrant, T., San-Miguel-Ayanz, J., 2016b. The European Atlas of Forest Tree Species: modelling, data and information on forest tree species. In: San-Miguel-Ayanz, J., de Rigo, D., Caudullo, G., Houston Durrant, T., Mauri, A. (Eds.), European Atlas of Forest Tree Species. Publ. Off. EU, Luxembourg,. e01aa69+

de Rigo, D., Caudullo, G., San-Miguel-Ayanz, J, Barredo, J.I., 2017. Robust modelling of the impacts of climate change on the habitat suitability of forest tree species. Publication Office of the European Union, Luxembourg. 58 pp. ISBN:978-92-79-66704-6 , https://doi.org/10.2760/296501

de Rigo, D.: Relative distance similarity as multivariate supervised or unsupervised ensemble interpolation. In prep. (exp. 2018).

de Roo, A. P. J., Wesseling, C. G., Cremers, N. H. D. T., Offermans, R. J. E., Ritsema, C. J., Oostindie, K., 1994. LISEM: a new physically- basedhydrologicaland soil erosion model in a GIS-environment: theory and implementation. In: Variability in Stream Erosion and Sediment Transport ta. by L. J. Olive, R. J. Loughran and J., A. Kesby) (Proc. Canberra Symp., December 1994), 439–448. IAHS Publ. no. 224.

de Santos Loureiro, N. S., de Azevedo Coutinho, M.,2001. A new procedure to estimate the RUSLE EI30 index, based on monthly rainfall data and applied to the Algarve region, Portugal, J. Hydrol. 250, 12–18.

R Development Core Team, 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

de Vente, J., Poesen, J., 2005. Predicting soil erosion and sediment yield at the basin scale: scale issues and semi-quantitative models. Earth-Science Reviews 71, 95-125.

de Vente, J., Poesen, J., Verstraeten, G., 2005. The application of semi-quantitative methods and reservoir sedimentation rates for the prediction of basin sediment yield in Spain. Journal of Hydrology 305(14), 63-86.

de Vente, J., Poesen, J., Bazzoffi, P., Van Rompaey, A., Verstraeten, G., 2006. Predicting catchment sediment yield in Mediterranean environments: the importance of sediment sources and connectivity in italian drainage basins. Earth Surface Processes and landforms 31, 1017–1034.

de Vente J., 2009. Soil erosion and sediment yield in Mediterranean geoecosystems. Scale issues, modelling and understanding. PhD thesis, K.U.Leuven, Belgium.

de Vente, J., Poesen, J. Verstraeten, G., Govers, G., Vanmaercke, M., et al., 2013. Predicting soil erosion and sediment yield at regional scales: where do we stand? Earth-Sci. Rev. 127, 16–29.

Dewitte, O., Daoudi, M., Bosco, C., Van Den Eeckhaut, M., 2014. Predicting the susceptibility to gully initiation in data-poor regions. Geomorphology. doi: 10.1016/j.geomorph. 2014.08.010.

Dietrich, C., Green, T.R., Jakeman, A.J., 1999. An analytical model for stream sediment transport: application to Murray and Murrumbidgee reaches, Australia. Hydrological Processes 13(5), 763–776.

Dietrich, W.E., Bellugi, D., Real de Asua, R., 2001. Validation of the shallow landslide model SHALSTAB for forest management. In: Wigmosta, M.S., Burges, S.J. (Eds.), Land Use and Watersheds: human influence on hydrology and geo- morphology in urban and forest areas. American Geophysical Union Water Science and Application 2, 195–227.

Di Leo, M., de Rigo, D., Rodriguez-Aseretto, D., Bosco, C., Petroliagkis, T., et al., 2013. Dynamic data driven ensemble for wild_re behaviour assessment: a case study. IFIP Adv. Inf. Commun. Technol. 413, 11-22. doi: 10.1007/978-3-642-41151-9 2 .

Diodato, N., 2004. Estimating RUSLE's rainfall factor in the part of Italy with a Mediterranean rainfall regime. Hydrol. Earth Syst. Sci. 8, 103–107, doi:10.5194/hess-8-103-2004.

Dobos, E., Micheli, E., Baumgardner, M.F., Biehl, L., Helt, T. 2000. Use of combined digital elevation model and satellite radiometric data for regional soil mapping. Geoderma 97(3-4), 367–391.

Doe, W.W., Harmon, R.S., 2001. Introduction to soil erosion and landscape evolution modeling. In: R.S. Harmon and W.W. Doe (Editors), Landscape Erosion and Evolution Modeling. Kluwer Academic, Plenum Publishers, New York, 1-14.

Dunne, T., 1977. Evaluation of erosion conditions and trends. In Kunkle, S.H. (ed). Guidelines for Watershed Management. Rome: FAO Conservation Guide 1, 53-83.

Dymond J.R., Ausseil A., Shepher J.D., Buettner L., 2006. Validation of a region-wide model of landslide susceptibility in the Manawatu-Wanganui region of New Zealand. Geomorphology 74, 70–79.

Eaton, J. W., Bateman, D., Hauberg, S., 2008. GNU Octave Manual Version 3. A high-level interactive language for numerical computations. Network Theory Limited, ISBN: 0-9546120-6-X.

Efron, B., Tibshirani, R., 1993. An introduction to the Bootstrap. New York, Chapman and Hall.

Efron, B., 1982. The Jackknife, the Bootstrap, and Other Resampling Plans. NSF-CBMS. Monograph 38, Philadelphia, SIAM.

Einsele, G. Hinderer, M., 1997. Terrestrial sediment yield and the lifetime of reservoirs, lakes, and larger basins. Geologische Rundschau 86, 288-310.

Endrero, A., Dramis, F., 1996. The contribution of landslides to landscape evolution in Europe. Geomorphology 15, 191–211.

Ermini, L., Catani, F., Casagli, N., 2005. Artificial neural networks applied to landslide susceptibility assessment. Geomorphology 66(1-4), 327–343.

Erskine, R.H., Green, T.R., Ramirez, J.A., MacDonald, L.H., 2006. Comparison of grid-based algorithms for computing upslope contributing area. Water Resour. Res. 42, W09416, 2006. doi: 10.1029/2005WR00464 .

Esri, HERE, Garmin, Intermap P Corp., GEBCO, USGS, FAO, NPS, NRCAN, GeoBase,IGN, Kadaster NL, Ordnance Survey, Esri japan, METI, Esri China (Hong Kong), swisstopo, © OpenStreetMap contributors, and the GIS Users Community, sources: Esri, Garmin, USGS, NPS, 2018.

Eswaran, H., Lal, R., Reich, P.F., 2001. Land degradation: an overview. In: Bridges, E.M., I.D. Hannam, L.R. Oldeman, F.W.T. Pening de Vries, S.J. Scherr, and S. Sompatpanit (eds.). Responses to Land Degradation. Proc. 2nd. International

Conference on Land Degradation and Desertification, Khon Kaen, Thailand. Oxford Press, New Delhi, India.

European Commission (EC), 2006. Proposal for a Directive of the European Parliament and of the Council establishing a framework for the protection of soil and amending. Directive 2004/35/EC (COM(2006)232). Avaliable online: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2006:0232:FIN:EN:PDF

European Environment Agency, 2011. Corine Land Cover 2006 raster data version 15 (08/2011). [Online]. Available: http://www.eea.europa.eu/data-and-maps/data/corine-land-cover-2006-raster-1 .

EP and Council, 2013. Decision No 1386/2013/EU of 20 November 2013 on aGeneral Union Environment Action Programme to 2020 'Living Well, Withinthe Limits of our Planet'. OJ L 354/171, Brussels, 171–200.

Evans, R., 1993. On assessing accelerated erosion of arable land by water. Soils and Fertilizers 56 (11), 1285–1293.

Evans, R., Brazier, R., 2005. Evaluation of modelled spatially distributed predictions of soil erosion by water versus field-based assessments. Environmental Science and Policy 8, 493 – 501.

Ewen, J., Parkin, G., OConnell, P.E., 2000. SHETRAN: distributed river basin flow and transport modeling system. Journal of Hydrologic Engineering 5(3), 250–258.

Favis-Mortlock, D., 1998. Validation of field scale soil erosion models using common datasets. In: Modelling Soil Erosion by Water, NATO ASI Series I, 55, edited by: Boardman, J. and Favis-Mortlock, D., Springer, Berlin, 89–127.

Ferro, V., 2010. Deducing the USLE mathematical structure by dimensional analysis and self-similarity theory. Biosyst. Eng. 106, 216–220.

Finney, H.J., 1984. The effect of crop covers on rainfall characteristics and splash detachment. Journal of Agricultural Engineering Research 29, 337–343.

Flanagan, D.C., Ascough II, J.C., Nicks, A.D., Nearing, M.A., Laflen, J.M., 1995. Overview of the wepp erosion prediction model. NSERL report.

Foley, J.A., DeFries, R., Asner, G.P., Barford, C., Bonan, et. al., 2005. Global consequences of land use. Science 309, 570–574. doi: 10.1126/science.1111772 , INRMM:764927.

Foster, G.R., 1982. Modelling the erosion processes. In: C.T. Hann (Editor), Hydrologic Modelling of Small Watersheds. ASAE Monograph, 297-380.

Foster, G.R., Yoder, D.C., McCool, D.K., G.A. Weesies, G.A., Toy, T.J., et al., 2000. Improvements in science in RUSLE2. ASAE paper 002147. ASAE, St. Joseph, MI.

Foster, G.R., Yoder, D.C., Weesies, G.A., McCool, D.K., McGregor K.C., et.al., 2003. Revised Universal Soil Loss Equation version 2 – User's Guide. USDA-Agricultural Research Service.Washington, D.C.

Francis, C., Thornes, J. B., 1990. Runoff hydrographs from three Mediterranean vegetation cover types. In: J. B. Thornes (Ed.), Vegetation and erosion. Chischester, Wiley, 363-384.

Freeman, T.G., 1991. Calculating catchment area with divergent flow based on a regular grid. Computer and Geosciences 17(3), 413-422.

Fredlund D.G., Morgenstern, N.R., Widger R.A., 1978. The shear strength of unsatured soils. Canadian Geotechnical Journal 15(3), 313-321.

García-Ruiz, J. M., 2010. The effects of land uses on soil erosion in Spain: a review. Catena 81, 1-11.

Geertsema, M., Pojar, J.J., 2007. Influence of landslides on biophysical diversity: a perspective from britishColumbia. Geomorphology 89(1-2), 55-69. doi: 10.1016/j.geomorph.2006.07.019 .

Geißler, C., Lang, A. C., von Oheimb, G., Härdtle,W., Baruffol, M., et.al., 2012. Impact of tree saplings on the kinetic energy of rainfall – The importance of stand density, species identity and tree architecture in subtropical forests in China, Agr. Forest Meteorol. 156, 31–40.

Gessler, P. E., Moore, I. D., McKenzie, N. J., Ryan, P. J. 1995. Soil-landscape modeling and spatial prediction of soil attributes. Inter. J. Geogr. Infor. Syst. 9(4), 421–432.

Giménez, R., and Govers, G., 2001. Interaction between bed roughness and flow hydraulics in eroding rills. Water Resources Research 37(3), 791-799.

Globevnik, L., Holjevic, D., Petkovsek, G., Rubinic, J., 2003. Applicability of the Gavrilovic method in erosion calculation using spatial data manipulation tech- niques. In: De Boer, D., Froehlich, W., Mizuyama, T., Pietroniro, A. (Eds.), Erosion Prediction in Ungauged Basins: Integrating Methods and Techniques. IAHS Publication 279, 224-233.

Godt, J. W., Baum, R. L., Lu, N. 2008. Landsliding in partially saturated materials. Geophysical Research Letters. http://dx.doi.org/10.1029/2008GL035996.

Goovaerts, P. 1997. Geostatistics for Natural Resources Evaluation. New York, Oxford University Press, 483 p.

Goudie A S, Boardman J., 2010. Soil erosion. In: Alćantara-Ayala I, Goudie A (eds.) Geomorphological Hazards and Disaster Prevention. Cambridge University Press, Cambridge, 177–188.

Govers, G., Van Oost, K., and Poesen, J., 2006. Responses of a semi-arid landscape to human disturbance: a simulation study of the interaction between rock fragment cover, soil erosion and land use change, Geoderma 133, 19–31.

Govers, G., 2011. Misapplications and misconception s of erosion models. In: Morgan, R.P.C., Nearing, M.A. (Eds.), Handbook of erosion modelling , 1st edition. Blackwell Publishing Ltd, Chichester, UK, 117-134.

Gruber, S., Peckham, S., 2009. Land-Surface Parameters and Objects in Hydrology. In: Geo-morphometry, Concepts, Software, Applications, T. Hengl and H.I. Reuter, Eds., Developments in Soil Science, Elsevier 33, 171-194. doi: 10.1016/S0166-2481(08)00007-X .

Guerra, A. J. T., Fullen, M. A., Jorge, M. D. C. O., Bezerra, J. F. R., Shokr, M. S., 2017. Slope processes, mass movement and soil erosion: a review. Pedosphere 27(1), 27-41.

Guzzetti, F., Carrara, A., Cardinali, M., Reichenbach, P., 1999. Landslide hazard evaluation: a review of current techniques and their application in a multiscale study, Central Italy. Geomorphology 31(14), 181-216.

Guzzetti, F., 2005. Landslide Hazard and Risk Assessment. PhD Thesis, University of Bonn, 389 p.

Guzzetti, F., Reichenbach, P., Cardinali, M., Galli, M., Ardizzone, F., 2005. Probabilistic landslide hazard assessment at the basin scale. Geomorphology 72, 272-299.

Hairsine, P., Rose, C., 1992. Modelling water erosion due to overland flow using physical principles: 2. Rill flow. Water Resources Research 28 (1), 245–250.

Hajigholizadeh, M., Melesse, A., Fuentes, H., 2018. Erosion and Sediment Transport Modelling in Shallow Waters: A Review on Approaches, Models and Applications. International journal of environmental research and public health, 15(3), 518.

Hamel, P., Falinski, K., Sharp, R., Auerbach, D.A., et al., 2017. Sediment delivery modeling in practice: Comparing the effects of watershed characteristics and data resolution across ydroclimatic regions. Science of The Total Environment 580, 1381 - 1388.

Hammond, C., Hall, D., Miller, S., Swetik, P., 1992. Level I Stability Analysis (LISA) Documentation for Version 2.0, General Technical Report INT-285, USDA Forest Service Intermountain Research Station.

Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., Stahel, W.A., 1986. Robust statistics, Wiley Series. In: Probability and Mathematical Statistics, New York, John Wiley & Sons, Inc., ISBN 0-471-82921-8, MR 0829458. Republished in paperback, 2005.

Hanley, N., Faichney, R., Munro, A., Shortle, J.S., 1998. Economic and environmental modelling for pollution control in an estuary. Journal of Environmental Management 52, 211–225.

Hansen, L., Hellerstein, D., 2007. The Value of the Reservoir Services Gained With Soil Conservation. Land Economics 83(3), 285–301.

Harp EL, Jibson RL., 1995. Inventory of landslides triggered by the 1994 Northridge, California earthquake. US Geological Survey Open File Report, 95 – 213.

Harp EL, Jibson RL., 1996. Landslides triggered by the 1994 Northridge, California earthquake. Seismological Society of America Bulletin 86, 319 –332.

Hatton, L., Roberts, A., 1994. How accurate is scientific software? Software Engineering, IEEE Transactions on 20 (10), 785-797.
https://doi.org/10.1109/32.328993.

Hatton, L., 2007. The chimera of software quality. Computer 40(8), 104-103. doi:10.1109/MC.2007.292.

Haylock, M. R., Goodess, C. M., 2004. Interannual variability of European extreme winter rainfall and links with mean large-scale circulation. International Journal of Climatology 24, 759–776.

Haylock, M. R., Hofstra, N., Klein Tank, A. M. G., Klok, E. J., Jones, P. D., et. al., 2008. A European daily high-resolution gridded dataset of surface temperature and precipitation, J. Geophys. Res.-Atmos. 113, D20119. doi:10.1029/2008JD010201.

Heineke, H. J., Eckelmann, W., Thomasson, A. J., Jones, R. J. A., Montanarella, L., et.al., 1998. Land Information Systems: Developments for planning the sustainable use of land resources, Office for Official Publications of the European Communities, Luxembourg, EUR 17729 EN.

Hengl, T., 2003. Pedometric mapping: bridging the gaps between conventional and pedometrical approaches. PhD Thesis, The Wageningen University.

Hengl, T., Gruber, S., Shrestha, D.P., 2004. Reduction of errors in digital terrain parameters used in soil-landscape modelling. International Journal of Applied Earth Observation and Geoinformation 5(2), 97-112.

Herrera, M., Torgo, L., Izquierdo, J., Pérez-García, R., 2010. Predictive models for forecasting hourly urban water demand. Journal of hydrology 387(1), 141-150.

Hessel, R., 2002. Modelling soil erosion in a small catchment on the Chinese Loess Plateau, PhD thesis, Utrecht University, Netherlands, 307 p.

Hessel, R., Jetten, V., Liu, B., Zhang, Y., Stolte, J., 2003. Calibration of the LISEM model for a small Loess Plateau catchment. Catena, 54 (1–2), 235-254.

Hieronymi, A., 2013. Understanding systems science: a visual and integrative approach. Systems Research and Behavioral Science 30(5), 580-595. https://doi.org/10.1002/sres.2215.

Hilberts, A.G.J., 2006. Low-dimensional modeling of hillslope subsurface flow processes. Developing and testing the hillslope-storage Boussinesq model. Dissertation Wageningen Universiteit, Wageningen, 2006. ISBN 90-8504-516-9.

Holmgren, P., 1994. Multiple flow direction algorithms for runoff modelling in grid based elevation models-an empirical evaluation. Hydrological processes 8(4), 327-334.

Hook, D., Kelly, D., 2009. Testing for trustworthiness in scientific software. In: Software Engineering for Computational Science and Engineering, 2009. SECSE '09. ICSEWorkshop on. IEEE, WashingtonDC, USA, 59-64.
doi:10.1109/SECSE.2009.5069163.

Horn, R., Fleige, H., 2003. A method for assessing the impact of load on mechanical stability and on physical properties of soils. Soil. Till. Res. 73 (1-2), 89-99.

Hosseini, M., Agereh, S.R., Khaledian, Y., Zoghalchali, H.J., Brevik, E.C., Naeini, S.A.R.M., 2017. Comparison of multiple statistical techniques to predict soil phosphorus. Applied Soil Ecology, 114, 123-131.

Hovius, N., Stark, C.P., Allen, P.A., 1997. Sediment flux from a mountain belt derived by landslide mapping. Geology 25, 231-234.

Hrachowitz, M., Clark, M.P., 2017. HESS Opinions: The complementary merits of competing modelling philosophies in hydrology. Hydrology and Earth System Sciences, 21(8), 3953-3973.

Hu, Z, Shafiqul, I., Cheng, I., Hu, Z.L., Cheng, Y.Z., et al., 1997. Statistical characterization of remotely sensed soil moisture images. Remote Sensing of Environment 61(2), 310-318.

Huber, S., Prokop, G., Arrouays, D., Banko, G., Bispo, A., et.al., 2008. Environmental Assessment of Soil for Monitoring: Volume I Indicators & Criteria. EUR 23490 EN/1. Office for the Official Publications of the European Communities, Luxembourg, 339 p.

Hughes, A.O., Croke, J.C., 2011. Validation of a spatially distributed erosion and sediment yield model (SedNet) with empirically derived data from a catchment adjacent to the Great Barrier Reef Lagoon. Marine and Freshwater Research, 62(8), 962-973.

Iannetta, M., Trotta, C., 2008. Piano di Azione Locale (PAL) per la lotta alla Siccità e alla Desertificazione della Regione Puglia. ENEA Dipartimento BAS, Gruppo 'Lotta alla Desertificazione'.

Iovine, G., Parise, M., Crescenzi,E., 1996. Analisi della franosita' nel settore centrale dell'Appennino Dauno. Memorie della societa' Geologica Italiana 51, 633-641.

Iverson, K. E., 1980. Notation as a tool of thought, Commun. ACM 23, 444–465.

Jaiswal, P., van Westen, C.J., 2009. Estimating temporal probability for landslide initiation along transportation routes based on rainfall thresholds. Geomorphology 112, 96-105.

Jakeman, A., Littlewood, I., Whitehead, P., 1990. Computation of the instantaneous unit hydrograph and identifiable component flows with application to two small upland catchments. Journal of Hydrology 117, 275–300.

Jakeman, A., Post, D., Schreider, S., Yu, Y.W., 1994. Modelling environmental systems: partitioning the water balance at different catchment scales. In: Zannetti, P. (Ed.), Computer Techniques, Environmental Studies V. Computational Mechanics Publications, Southampton, 157–170.

Jakeman, A.J., Green, T.R., Beavis, S.G., Zhang, L., Dietrich, C.R., et.al., 1999. Modelling upland and in-stream erosion, sediment and phosphorus trans- port in a large catchment. Hydrological Processes 13(5), 745-752.

Jakob, M. Hungr, O., 2005. Introduction. In: Debris-flow Hazards and Related Phenomena, edited by: Jacob, M. and Hungr, O., Praxis Publishing House and Springer, Chichester, UK, p. 1–7.

Jetten, V., de Roo, A., Favis-Mortlock, D., 1999. Evaluation of field-scale and catchment-scale soil erosion models. Catena 37, 521–541.

Jetten, V., Govers, G., Hessel, R., 2003. Erosion models: quality of spatial predictions. Hydrological Processes 17, 887-900. Published online in Wiley InterScience (www.interscience.wiley.com). doi: 10.1002/hyp.1168.

Johanson, R.C., Imhoff, J.C., Davis, H.H., 1980. Users Manual for the Hydrologic Simulation Program-Fortran (HSPF), version5.0, EPA 600(9), 80-105. US EPA Environmental Research Laboratory, Athens, GA.

Jones, R. J. A., Grim, M., Montanarella, L., 2003. Use of Meteorological data sets at European level for input to the PESERA Grid Model, PESERA (Contract QLKS-CT- 1999-01323) final report.

Jothityangkoon, C., Sivapalan, M., Farmer, D. L., 2001. Process controls of water balance variability in a large semi-arid catchment: downward approach to hydrological model development. Journal of Hydrology 1(4), 174-198 .

Keller, J,M., Gray, M.R., Givens, J.A., 1985. A fuzzy K-nearest neighbour algorithm. IEEE Transactions on systems, man and cybernetcis 15(4), 580-585.

Kelly Letcher, R. A., Jakeman, A. J., Barreteau, O., Borsuk, M. E., ElSawah, S., et al., 2013. Selecting among five common modelling approaches for integrated environmental assessment and management. Environmental Modelling & Software 47, 159-181. https://doi.org/10.1016/j.envsoft.2013.05.005.

Kinnell, P. I. A., 2010. Event soil loss, runoff and the Universal Soil Loss Equation family of models: a review. J. Hydrol. 385, 384–397. doi:10.1016/j.jhydrol.2010.01.024.

Kirkby, M.J., 1976. Hydrological slope models: the influence of climate. In: Geomorphology and Climate (Derbyshire, E., Ed.), London, Wiley, 247-267.

Kirkby, M.J., Abrahart, R.J., McMahon, M.D., Shao, J., Thornes, J.B., 1998a. MEDALUS soil erosion models for global change. Geomorphology 24, 35–49.

Kirkby, M.J., 1998b. Modelling across scales: the MEDALUS family of models. In: Boardman, J. and Favis-Mortlock (Eds.). Modelling Soil Erosion by Water. Springer, Berlin, 161-173.

Kirkby , M.J., McMahon, M.L., 1999. MEDRUSH and the Catsop basin-the lessons Learned. Catena 37, 495–506.

Kirkby, M., Robert, J., Irvine, B., Gobin, A., Govers, G., et al., 2003. Pan-European soil erosion risk assesment. EUR 21176 EN.

Koopmans, M., Stamovlasis, D., 2016. Glossary. In: Koopmans, M., Stamovlasis, D. (Eds.), Complex dynamical systems in education - Concepts, methods and applications. Springer, Cham, Switzerland, 395-404. https://doi.org/10.1007/978-3-319-27577-2.

Koppen, M., Wolpert, D. H., Macready, W. G., 2001. Remarks on a recent paper on the "no free lunch" theorems. Evolutionary Computation, IEEE Transactions 5(3), 295-296. https://doi.org/10.1109/4235.930318

Kosmas, C., Danalatos, N., Cammeraat, L. H., Chabart, M., Diamantopoulos, J., et al., 1997. The effect of land use on runoff and soil erosion rates under Mediterranean conditions. Catena 29, 45-59.

Knisel, W.G., 1980. CREAMS: a field scale model for chemicals, runoff and erosion from agricultural management systems. USDA. United States. Dept. of Agriculture. Conservation research report.

Krzeminska, D.M., Bogaard, T.A., van Asch, Th.W.J, van Beek, L.P.H., 2012. A conceptual model of the hydrological influence of fissures on landslide activity. Hydrology and Earth System Sciences 16(6), 1561–1576.

Laflen, J.M., Lane, L.J., Foster, G.R., 1991. The water erosion prediction project: a new generation of erosion prediction technology. Journal of Soil and Water Conservation 46, 34-38.

Lal, R., 1998.: Soil erosion impact on agronomic productivity and environment quality, Crit., ev. Plant Sci. 17, 319–464.

Lamanna, C., Casarano, D., & Wasowski, J., 2009. Cambiamenti dell'uso del suolo e aumento dell'attività franosa nel territorio di Rocchetta Sant'Antonio (Appennino dauno). Il quaternario - Italian Journal of Quaternary Sciences 22(2), 139–156.

Lamanna, C., Casarano, D., Gigante, G., Wasowski, J., 2009. Mappatura e studio dei fenomeni franosi nel Subappennino dauno con immagini satellitari ad alta risoluzione. In: Proc. 13a conferenza Nazionale ASITA, Bari.

Laniak, G. F., Olchin, G., Goodall, J., Voinov, A., Hill, M., et al., 2013. Integrated environmental modeling: a vision and roadmap for the future. Environmental Modelling & Software 39, 3-23. https://doi.org/10.1016/j.envsoft.- 2012.09.006.

Lanni C., 2012. Hydrological controls on the triggering of shallow landslides: from local to landscape scale. PhD Thesis, University of Trento.

Lark, R.M., Milne, A.E., Addiscott, T.M., Goulding, K.W.T., Webster, C.P., et al., 2004. Scale and location-dependent correlation of nitrous oxide emissions with soil properties: an analysis using wavelets. European Journal of Soil Science 55, 611-627.

Larsen, J.L., Montgomery, D.R., Korup, O., 2010. Landslide erosion controlled by hill slope material. Nature Geoscience 3, 247-251.

Lauder, A., Kent, S., 2000. Legacy system anti-patterns and a pattern-oriented migration response. In: Henderson, P. (Ed.), Systems Engineering for Business Process Change. Springer London, 239-250. https://doi.org/10.1007/978-1-4471-0457-5_19.

Lawler, J.J., White, D., Neilson, R.P., Blaustein, A.R., 2006. Predicting climate-induced range shifts: model differences and model reliability. Global Change Biology, 12(8), 1568-1584.

Lee, C. R., Skogerboe, J. G., 1985. Quantification of erosion control by vegetation on problem soils. In: Swaify, A., Moldenhauer, W.C., Lo, A., (Eds.), Soil erosion and conservation. Soil Conservation Soc. of America, 437-444.

Lee, S., 2004. Soil erosion assessment and its verification using the universal soil loss equation and geographic information system: a case study at Boun, Korea. Environ Geol 45, 457–465.

Lee, S., Ryu, J.H., Kim, I.S., 2007. Landslide susceptibility analysis and its verification using likelihood ratio, logistic regression, and artificial neural network models: case study of Youngin, Korea. Landslide 4, 327-338.

Lee, S., Jeon, S.W., Oh, K.-Y., Lee, M. J., 2016. The Spatial Prediction of Landslide Susceptibility Applying Artificial Neural Network and Logistic Regression Models: a Case Study of Inje, Korea. Open Geosciences 8, 117-132. https://doi.org/10.1515/geo-2016-0010.

Lehman, M.M., Ramil, J.F., 2003. Software evolution - background, theory, practice. Information Processing Letters 88(1-2), 33-44. https://doi.org/10.1016/s0020-0190(03)00382-x.

Leone, A.P., Sommer, S., 2000. Multivariate Analysis of Laboratory Spectra for the Assessment of Soil Development and Soil Degradation in the Southern Apennines (Italy). Remote Sensing of Environment 72(3), 346-359. ISSN 0034-4257, https://doi.org/10.1016/S0034-4257(99)00110-8.

Li, J., Heap, A.D., 2011. A review of comparative studies of spatial interpolation methods in environmental sciences: performance and impact factors. Ecological Informatics, 6(3-4), 228-241.

Lucà, F., Buttafuoco, G., Terranova, O., 2018. GIS and Soil. In: Huang, B. (Eds.), Comprehensive Geographic Information Systems. Vol. 2, 37–50. Oxford: Elsevier. http://dx.doi.org/10.1016/B978-0-12-409548-9.09634-2

Licciardello, F., Govers, G., Cerdan, O., Kirkby, M.J., Vacca, A., et al., 2009. Evaluation of the PESERA model in two contrasting environments. Earth Surf. Process. Landforms 34(5), 629-640.

Littleboy, M., Silburn, M.D., Freebairn, D.M., Woodruff, D.R., Hammer, G.L., et al., 1992. Impact of soil erosion on production in cropping systems. I. Development and validation of a simulation model. Australian Journal of Soil Research 30, 757-774.

Loague, K, Heppner, C.S., Mirus, B.B., Ebe,l B.A., Ran Q, et al., 2006. Physics-based hydrologic-response simulation: foundation for hydroecology and hydrogeomorphology. Hydrological Processes 20, 1231-1237.

Loch, R.J., Silburn, D.M., 1996. Constraints to sustainability-soil erosion. In: Clarke, L., Wylie, P.B. (Eds.), Sustainable Crop Production in the Sub-tropics: an australian perspective. QDPI.

López-Vicente, M., Navas, A., 2010. Relating soil erosion and sediment yield to geomorphic features and erosion processes at the catchment scale in the spanish pre-Pyrenees. Environmental Earth Sciences 61(1), 143–158. doi: 10.1007/s 12665-009-0332-x.

López-Vicente, M., Poesen, J., Navas, A., Gaspar, L., 2011. Predicting runoff and sediment connectivity and soil erosion by water for different land-use scenarios in the spanish pre-Pyrenees. Catena 102, 62-73. doi:10.1016/j.catena.2011.01.001.

Lu, D., Li, G., Valladares, G.S., Batistella, M., 2004. Mapping soil erosion risk in Rondonia, brazilian Amazonia: using RUSLE, remote sensing and GIS. Land Degrad. Dev. 15, 499–512.

Lucieer, A., Turner, D., King, D.H., et al., 2014. Using an Unmanned Aerial Vehicle (UAV) to capture micro-topography of Antarctic moss beds. International Journal of Applied Earth Observation and Geoinformation 27(Part A), 53–62.

Maes, J., Teller, A., Erhard, M., Liquete, C., Braat, L., et al., 2013. Mapping and Assessment of Ecosystems and their Services - An analytical framework for ecosystem assessments under action 5 of the EU biodiversity strategy to 2020. Publications office of the European Union, Luxembourg, 57p. ISBN: 978-92-79-29369-6 , https://doi.org/10.2779/12398.

Maes, J., Liquete, C., Teller, A., Erhard, M., Paracchini, et al., 2016. An indicator framework for assessing ecosystem services in support of the EU biodiversity strategy to 2020. Ecosystem Services 17, 14-23. https://doi.org/10.1016/ j.ecoser.2015.10.023.

Maetens, W., Vanmaercke, M., Poesen, J., Jankauskas, B., Jankauskien, G., et al., 2012. Effects of land use on annual runoff and soil loss in Europe and the Mediterranean: a metaanalysis of plot data, Prog. Phys. Geog. 36, 599–653.

Magliulo, P., Di Lisio, A., Russo, F., Zelano, A., 2008. Geomorphology and landslide susceptibility assessment using GIS and bivariate statistics: a case study in southernItaly. Natural Hazards 47(3), 411-435.

Mahmoodabadi, M., Cerdà, A., 2013. WEPP calibration for improved predictions of interrill erosion in semi-arid to arid environments. Geoderma, 204, 75-83.

Maidment, D.R., Djokic, D., 2000. Hydrologic and hydraulic modeling support with geographic information systems. ESRI Press, Redlands, CA.

Maimon, O., L. Rokach, L., 2005. The Data Mining and Knowledge Discovery Handbook, Springer, Heidelberg, 1285 p. doi: 10.1007/978-0-387-09823-4.

Malamud, B.D., Turcotte, D.L., Guzzetti, F., Reichenbach, P., 2004. Landslide inventories and their statistical properties. Earth Surf Processes Landforms 29, 687-711.

Malatesta, A., Perno, U., Stampanoni, G., 1967. Note illustrative del foglio 175, Cerignola. Roma, Servizio Geologico d'Italia.

Malet, J. P., Van Asch, T. W., Van Beek, R., Maquaire, O., 2005. Forecasting the behaviour of complex landslides with a spatially distributed hydrological model. Natural Hazards and Earth System Science 5(1), 71–85.

Mäntylä, M., Lassenius, C., 2006. Subjective evaluation of software evolvability using code smells: an empirical study. Empirical Software Engineering 11 (3), 395-431. https://doi.org/10.1007/s10664-006-9002-8.

Markantonis, V., Meyer, V., Schwarze, R., 2012. Review article "Valuating the intangible effects of natural hazards – review and analysis of the costing methods". Nat. Hazards Earth Syst. Sci. 12, 1633–1640. doi:10.5194/nhess-12-1633-2012.

Matheron, G., 1969. Le krigeage universel. Vol. 1. Fontainebleau, Cahiers du Centre de Morphologie Mathematique, Ecole des Mines de Paris.

Mathys, N., Brochot, S., Meunier, M., Richard, D., 2003. Erosion quantification in the small marly experimental catchments of Draix (Alpes de Haute Provence, France). Calibration of the ETC rainfall-runoff-erosion model. Catena 50, 527–548.

Mazaeva, O., Khak, V., Kozyreva, E., 2013. Model of erosion-landslide interaction in the context of the reservoir water level variations (East Siberia, Russia): factors, environment and mechanisms. J. Earth Syst. Sci. 122(6), 1515–1531.

McBratney, A. B., Mendoça Santos, M. L., Minasny, B., 2003. On digital soil mapping. Geoderma 117 (1-2), 3–52.

McGregor, J. D., 2006. Complexity, its in the mind of the beholder. The J. Object Technol. 5, 31–37.

McIntyre, S., 1993. Reservoir sedimentation rates linked to long term changes in agricultural land use. Water resources bulletin 29(3), 487-495.

McKenzie, N.J., Ryan, P.J., 1999. Spatial prediction of soil properties using environmental correlation. Geoderma 89(1-2), 67-94.
Meisina, C., Scarabelli, S., 2007. A comparative analysis of terrain stability models for predicting shallow landslides in colluvial soils. Geomorphology 87, 207–223.

Merali, Z., 2010. Computational science: error, why scientific programming does not compute. Nature 467(7317), 775-777. doi:10.1038/467775a.

Merritt, W.S., Letcher, R.A., Jakeman, A.J., 2003. A review of erosion and sediment transport models. Environmental Modelling and Software 18(8), 761–799.

Meyer, L.D., Wischmeier, W.H., 1969. Mathematical simulation of the process of soil erosion by water. Trans, ASAE 12, 754-758.

Misra, R.K., Rose, C.W., 1996. Application and sensitivity analysis of process-based erosion model GUEST. European Journal of Soil Science 47, 593–604.

Mitasova, H., Hofierka, J., Zlocha, M., Iverson, R. L., 1996. Modeling topographic potential for erosion and deposition using GIS, Int. J. Geogr. Inf. Sci. 10, 629–641.

Mitchell, D. J., 1990. The use of vegetation and land use parameters in modelling catchment sediment yields. In: J. B. Thornes (Ed.), Vegetation and erosion,processes and environments. Wiley & Chichester, 289-314.

Mitas, L., Mitasova, H., 1998. Distributed soil erosion simulation for effective erosion prevention. Water Resources Research 34(3), 505–516.

Mohamed, M.A., 2017. Analysis of Digital Elevation Model and LNDSAT Data Using Geographic Information System for Soil Mapping in Urban Areas. Natural Resources, 8, 767-787. https://doi.org/10.4236/nr.2017.812047

Moles, A. T., Warton, D. I., Warman, L., Swenson, N. G., Laffan, S. W., et al. 2009. Global patterns in plant height. Journal of Ecology 97(5), 923-932.

Montanarella, L., 2007. Trends in land degradation in Europe. In: Climate and land degradation. Springer Berlin Heidelberg, 83-104.

Montgomery, D.R., Dietrich, W.E., 1994. A physically based model for the topographic control on shallow landsliding. Water Resour. Res. 30(4), 1153–1171. doi: 10.1029/93WR02979.

Montrasio, L., Valentino, R., 2008. A model for triggering mechanisms of shallow landslides, Nat. Hazards Earth Syst. Sci., 8, 1149–1159, doi:10.5194/nhess-8-1149-2008.

Moore I., Burch G., 1986. Physical basis of the length-slope factor in the universal soil loss equation, Soil Sci. Soc. Am. J. 50, 1294–1298.

Moore, I.D., Grayson, R.B., Ladson, A.R., 1991. Digital terrain modelling: a review of hydrological, geomorphological and biological applications. Hydrological processes 5, 3-30.

Moore, I.D., Gessler, P.E., Nielsen, G.A., Peterson, G.A., 1993. Soil attribute prediction using terrain analysis. Soil Science Society of America Journal 57(2), 443-452.

Moral, F.J., 2010. Comparison of different geostatistical approaches to map climate variables: application to precipitation. International Journal of Climatology: A Journal of the Royal Meteorological Society, 30(4), 620-631.

Morgan, R.P.C., Morgan, D.D.V., Finney, H.J., 1984. A predictive model for the assessment of erosion risk. Journal of Agricultural Engineering Research 30, 245–253.

Morgan, R.P.C., 1985. The impact of recreation on mountain soils: towards a predictive model for soil erosion. In: The Ecological Impacts of Outdoor Recreation on Mountain Areas in Europe and North America. Rural Ecology Research Group Report 9, 112-121, , Bayfield NG, Barrow GC (eds).

Morgan, R.P.C., Quinton, J.N., Rickson, R.J., 1993. EUROSEM: A User Guide. Silsoe College, Cranfield University.

Morgan, R.P.C., Quinton, J.N., Smith, R.E., Govers, G., Poesen, et al., 1998a. The European Soil Erosion Model (EUROSEM): a dynamic approach for predicting sediment transport from fields and small catchments. Earth Surf. Process. Landforms, 23, 527-544.

Morgan, R.P.C., Quinton J.N., Smith R.E., Govers G., Poesen J.W.A. et al., 1998b. The European Soil Erosion Model (EUROSEM): Documentation and User Guide. Silsoe College, Cranfield University.

Morgan, R.P.C., 2001. A simple approach to soil loss prediction: a revised Morgan–Morgan–Finney model. Catena 44, 305–322.

Morgan, R.P.C., Quinton, J.N., 2001. Erosion modelling. In: Landscape erosion and evolution modelling. Eds. Russell S., Harmon and William W. Doe III, Springer US, 117-143.

Morgan, R.P.C., 2005. Soil Erosion and Conservation, 3rd edn, Blackwell Publ., Oxford. 400 p.

Morgan, R.P.C., Duzant, J.H., 2008. Modified MMF (Morgan–Morgan–Finney) model for evaluating effects of crops and vegetation cover on soil erosion. Earth Surface Processes and Landforms 32, 90–106.

Morgan, R.P.C., Nearing, M., 2016. Handbook of erosion modelling. John Wiley & Sons.

Mossa S., Capolongo D., Pennetta L., Wasowski J., 2005. A GIS-based assessment of landsliding in the Daunia Apennines, southern Italy. In: Graniczny, M. Czarnogorska, M., et al., (Eds). Proceedings of the International Conference `Mass Movement Hazard in Various Environments', Warsaw, Poland. Polish Geological Institute, 86-91.

Mubareka, S., Jonsson, R., Rinaldi, F., Azevedo, J. C., de Rigo, et al., 2016. Forest bio-based economy in Europe. In: San-Miguel-Ayanz, J., de Rigo, D., Caudullo, G., Houston Durrant, T., Mauri, A. (Eds.), European Atlas of Forest Tree Species. Publications office of the European Union, Luxembourg, e01a52d+. https://w3id.org/mtv/FISE-Comm/v01/e01a52d.

Mulder, V.L., De Bruin, S., Schaepman, M.E., Mayr, T.R., 2011. The use of remote sensing in soil and terrain mapping—A review. Geoderma, 162(1-2), 1-19.
Nandi, A., Shakoor, A., 2009. A GIS-based landslide susceptibility evaluation using bivariate and multivariate statistical analyses. Engineering Geology 110, 11–20.

Nearing, M. A., 1997. A single, continuous function for slope steepness influence on soil loss. Soil Sci. Soc. Am. J. 61, 917–919.

Nearing, M. A., Norton, L.D., Bulgakov, D. A., Larionov, G. A., West, L. T., et al., 1997. Hydraulics and erosion in eroding rills. Water Resources Research 33(4), 865-876.

Nearing, M.A., Jetten, V., Baffaut, C., Cerdan, O., Couturier, A., et al., 2005. Modeling response of soil erosion and runoff to changes in precipitation and cover. Catena 61, 131–154.

Nearing, M.A., 2006. Can soil erosion be predicted? In: Owens, P.N., Collins, A.J. (Eds.), Oil Erosion and Sediment Redistribution in River Catchments. CAB International, 145-166.

Nicks, A.D., 1985. Generation of climate data. In: D.G. DeCoursey (editor), Proc. of the Natural Resources Modeling Symp., Pingree Park, CO, October 16-21, 1983, USDA-ARS, ARS-30.

Novara, A., Keesstra, S., Cerdà, A., Pereira, P., Gristina, L. , 2016. Understanding the role of soil erosion on CO 2-C loss using 13 C isotopic signatures in abandoned Mediterranean agricultural land. Science of the Total Environment 550, 330-336.

O'Callaghan, J.F., Mark, D.M., 1984. The exytaction of drainage networks from digital elevation data. Computer vision, graphics and image processing 28, 323-344.

Odeh, I.O.A., McBratney, A.B., Chittleborough, D.J., 1994. Spatial prediction of soil propertifrom landform attributes derived from a digital elevation model. Geoderma 63(3-4), 197–214.

Odeh, I.O.A., McBratney, A.B., Chittleborough, D.J., 1995. Further results on prediction osoil properties from terrain attributes: heterotopic cokriging and regression-kriging. Geoderma 67(3-4), 215–226.

Oldeman, L.R., Hakkeling, R.T.A., Sombroek, W.G., 1990. World map of the status of human induced soil degradation: an explanatory note. International Soil Reference and Information Centre.

O'Loughlin, E.M., 1986. Prediction of surface saturation zones in natural catchments by topographic analysis. Water Resour. Res. 22(5), 794–804.

Onyando, J. O., Kisoyan, P., Chemelil, M. C., 2005. Estimation of potential soil erosion for river Perkerra catchment in Kenya, Water Resour. Manag. 19, 133–143.

Pack R.T., Tarboton D.G., Goodwin C.N., 1998. The SINMAP. Approach to terrain stability mapping. In: Moore D.P., Hungr O. (eds) Proceedings international congress of the International Association for Engineering Geology and the Environment 8(2), 1157–1165, Rotterdam, Netherlands.

Pack, R.T., Tarboton, D.G.,Goodwin, C.N., Prasad, A., 2005. The SINMAP 2. A stability index approach to terrain stability hazard mapping. Technical description and users guide for version 2.0. Utah State University.

Panagos, P., Jones, A., Bosco, C., Senthil Kumar, P.S., 2011. European digital archive on soil maps (EuDASM): preserving important soil data for public free access. Int. J. Dig. Earth, 4, 434–443.

Panagos, P., Borrelli, P., Poesen, J., Ballabio, C., Lugato, et al., 2015. The new assessment of soil loss by water erosion in Europe. Environmental Science & Policy 54, 438-447.

Pandey, A., Himanshu, S.K., Mishra, S.K., Singh, V.P., 2016. Physically based soil erosion and sediment yield models revisited. Catena, 147, 595-620.

Pathak, S., Nilsen, B., 2004. Probabilistic rock slope stability analysis for Himalayan condition. Bulletin of Engineering Geology and the Environment 63, 25-32.

Petley, D., 2012. Global patterns of loss of life from landslides. Geology 40, 927–930. doi: 10.1130/G33217.1.

Pimentel, D., 2006. Soil erosion: a food and environmental threat. Environment Development and Sustainability 8, 119–137. doi:10.1007/s10668-005-1262-8.

Pimentel, D. Burgess, M., 2013. Soil erosion threatens food production. Agriculture 3(3), 443-463.

Pla Sentis, I., 1997. A soil water balance model for monitoring soil erosion processes and effects on steep lands in the tropics. Soil Technology 11, 17-30.

Poesen, J., Ingelmo-Sanchez, F., 1992. Runoff and sediment yield from topsoils with different porosity as affected by rock fragment cover and position. Catena 19, 451–474.

Poesen, J., Lavee, H., 1994. Rock fragments in top soils: significance and processes. Catena 23, 1–28.

Poesen, J., Torri, D., Bunte, K., 1994. Effects of rock fragments on soil erosion by water at different spatial scales: a review. Catena 23, 141–166.

Poesen. J., Boardman, J., Wilcox, B., Valentin, C., 1996. Water erosion monitoring and experimentation for global change studies. Journal of Soil and water Conservation 51, 386-390.

Poesen, J., Nachtergaele, J., Verstraeten, G., Valentin, C., 2003. Gully erosion and environmental change: importance and research needs. Catena 50, 91–133.

Popescu, M.E., 1994. A suggested method for reporting landslides causes. Bull IAEG 50, 71–74.

Posthumus, H., Deeks, L.K., Rickson, R.J., Quinton, J.N., 2015. Costs and benefits of erosion control measures in the UK. Soil use and management 31(S1), 16-33.

Pradhan, B., Lee, S., 2009. Landslide risk analysis using artificial neural network model focusig on different training sites. International Journal of Physicical Sciences 4(1), 1-15.

Prosser, I.P., Young, B., Rustomji, P., Hughes, A., Moran, C., 2001. A model of river sediment budgets as an element of river health assessment. In: Proceedings of the International Congress on Modelling and Simulation (MODSIM'2001), December 10–13, 861–866.

PSIAC, 1968. Pacific Southwest Inter-Agency Committee. Report of the water management subcommittee on factors affecting sediment yield in the Pacific southwest area and selection and evaluation of measures for reduction of erosion and sediment yield.

Quansah, C., 1982. Laboratory experimentation for the statistical derivation of equations for soil erosion modelling and soil conservation design. PhD Thesis, Cranfield Institute of Technology.

Quinn, P., Beven, K., Chevallier, P., Planchon, O., 1991. The prediction of hillslope paths for distributed hydrological modeling using digital terrain models. Hydrological Processes 5, 59–79.
Quinton, J. N., Govers, G., Van Oost, K., Bardgett, R. D., 2010. The impact of agricultural soil erosion on biogeochemical cycling, Nat. Geosci. 3, 311–314.

Ranzi, R., Le, T. H., Rulli, M. C., 2012. A RUSLE approach to model suspended sediment load in the lo river (Vietnam). Effects of reservoirs and land use changes. Hydrol. 422, 17–29. doi:10.1016/j.jhydrol.2011.12.009.

Rauws, G., Govers, G., 1988. Hydraulic and soil mechanical aspects of rill generation on agricultural soils. Journal of Soil Science 39, 111–124.

Regalado, S.A. and Kelting, D.L., 2015. Landscape level estimate of lands and waters impacted by road runoff in the Adirondack Park of New York State. Environmental monitoring and assessment 187(8), 510.

Renard, K.G., G.R. Foster, 1983. Soil conservation: principles of erosion by water. In: H.E. Dregne and W.O. Willis, (eds.), Dryland Agriculture, 155-176. Agronomy Monogr. 23, Am. SOC. Agron., Crop Sci. SOC. Am., and Soil Sci. SOC. Am., Madison, Wisconsin.

Renard, K. G., Foster, G. R., Weesies, G. A., McCool, D. K., Yoder, D. C., 1997. Predicting Soil Erosion by Water: a guide to conservation planning with the revised universal soil Loss equation (RUSLE), US Dept Agric., Agr. Handbook 703.

Renwick, W.H., Andereck, Z.D., 2006. Reservoir sedimentation trends in Ohio, USA: sediment delivery and response to land-use change. In: Rowen, J.S., Duck, R.W. and Werritty, S. (eds), Sediment Dynamics and the Hydromorphology of Fluvial Systems. IAHS publication 306, Dundee, UK, 341-347.

Richardson, G. P., 2009. System dynamics, the basic elements of. In: Meyers, R. A. (eds.), Encyclopedia of Complexity and Systems Science. Springer New York, 856-862. https://doi.org/10.1007/978-0-387-30440-3_536.

Rigon, R., Bertoldi, G., Over, T.M., 2006. GEOtop: a distributed jydrological model with coupled water and energy budgets. Jour. of Hydromet. 7(3), 371–388.

Rozos, D., Skilodimou, H. D., Loupasakis, C., Bathrellos, G. D., 2013. Application of the revised universal soil loss equation model on landslide prevention. An example from N. Euboea (Evia) Island, Greece. Environmental Earth Sciences 70(7), 3255-3266.

Ripley, B.D., 1996. Pattern Recognition and Neural Networks, Cambridge University press, 406 p.

Römkens, M. J. M, Prased, S. N., Poesen, J. W. A., 1986. Soil erodibility and properties. Trans. 13th congress of the Int. Soc. Of Soil Sci., Hamburg, Germany, 5, 492–504.

Rose, C.W., Coughlan, K.J., Ciesiolka, L.A.A., Fentie, B., 1997. Program GUEST (Griffith University Erosion System Template), a new soil conservation methodology and application to cropping systems in tropical steeplands. ACIAR Technical Reports 40, 34–58.

Rose, C.W., 2017. Research progress on soil erosion processes and a basis for soil conservation practices. In Soil erosion research methods, 159-180. Routledge.

Rosenblatt, F., 1962. Principles of Neurodynamics: perceptrons and the theory of brain mechanisms. Spartan, Washington DC.

Rossi, M., Guzzetti, F., Reichenbach, P., Mondini, A., Peruccacci, S., 2010. Optimal landslide susceptibility zonation based on multiple forecasts. Geomorphology 114, 129–142.

Saavedra, C., 2005. Estimating spatial patterns of soil erosion and deposition of the Andean region using geo-information techniques: a case study in Cochabamba, Bolivia. PhD Thesis, Wagenongen University, ISBN: 90-8504-289-5

Sadeghi, S.H.R., Gholami, L., Khaledi Darvishan, A., Saeidi, P., 2014. A Review of the Application of the MUSLE Model Worl-Wide. Hydrological Sciences Journal 59(1-2), 365-375.

Saltelli, A., Annoni, P., 2010. How to avoid a perfunctory sensitivity analysis. Environ. Modell. Softw. 25, 1508–1517.

Sander, G.C., Rose, C.W., Hogarth, W.L., Parlange J.Y, Lisie, I.G., 2002. Mathematical soil erosion modeling. Water interactions with energy environment, food and agriculture. Encyclopedia of Life Support Systems 2, 1-3.

Sander, G.C., Zheng,T., Heng, P., Zhong, Y. Barry D.A., 2011. Sustainable soil and water resources: modelling soil erosion and its impact on the environment. 19th International Congress on Modelling and Simulation, Perth, Australia. http://mssanz.org.au/modsim2011.

Schaub, D., Prasuhn, V., 1998. A map on soil erosion on arable land as a planning tool for sustainable land use in Switzerland, Adv. Geoecol. 31, 161–168.

Semmens, D.J., Goodrich, D.C., Unkrich, C.L., Smith, R.E., Woolhiser, D.A., Miller, S.N. 2008. KINEROS2 and the AGWA Modeling Framework. Chapter 5. In: Wheater, H., Sorooshian, S., Sharma, K.D. (eds.), Hydrological Modelling in Arid and Semi-Arid Areas, Cambridge University Press, London, 49-69.

Shi, Z.H., Fang, N.F., Wu, F.Z., Wang, L., Yue, B.J. and Wu, G.L., 2012. Soil erosion processes and sediment sorting associated with transport mechanisms on steep slopes. Journal of Hydrology, 454, 123-130.

Shirazi, M.A., L. Boersma., 1984. A unifying quantitative analysis of soil texture. Soil Sci. Soc. Am. J. 48, 142–147.

Shrestha, D.P., 1997. Assessment of soil erosion in the Nepalese Himalaya: a case study in Likhu Khola Valley, Middle Mountain Region. Land Husbandry 2, 59–80.

Sidle, R.C., Ochiai, H., 2006. Landslides: processes, prediction, and land use. Water Resources Monograph 18, 312.

Simoni S, Zanotti F, Bertoldi G, Rigon R., 2008. Modelling the probability of occurrence of shallow landslides and channelized debris flows using GEOtop-FS. Hydrological Processes 22(4), 532–545.

Skempton, A.W., Deloy, F.A., 1957. Stability of natural slopes in London Clay. Proceedings of the 4th International Conference on Soil Mechanics and Foundation Engineering 2, 378–381.

Smith, R.E., Goodrich, D.C., Quinton, J.N., 1995a. Dynamic, distributed simulation of watershed erosion: the KINEROS2 and EUROSEM models. Journal of Soil and Water Conservation 50(5), 517-520.

Smith, R.E., Goodrich, D.C., Woolhiser, D.A., Unkrich, C.L., 1995b. KINEROSa kinematic runoff and erosion model. In: Singh, V.P. (Eds.), Computer Models of Watershed Hydrology. Water Resources Publications, Littleton CO, 697-732.

Soil Science Society of America, 2001. Glossary of Soil Science Terms. Soil Science Society of America, Madison, WI. http://www.soils.org/sssagloss/

Son, K., Sivapalan, M., 2007. Improving model structure and reducing parameter uncertainty in conceptual water balance models through the use of auxiliary data. Water Resour. Res., 43, W01415, https://doi.org/10.1029/2006WR005032.

Srebotnjak, T., Polzin, C., Giljum, S., Herbert, S., Lutter, S., 2010. Establishing Environmental Sustainability Thresholds and Indicators. Final report of the project: Establishing thresholds and indicators for environmental sustainability, 138 p.

Stephens, P. R., Cihlar, J., 1982. Mapping erosion in New Zealand and Canada. In: Johannsen Jr., C.J., Sanders, J.L. (eds). Remote Sensing for Resource Management. Soil Conservation Society of America, Ankeny IA, 232–242.

Stockmann, U., Adams, M. A., Crawford, J. W., Field, D. J., Henakaarchchi, N., et al., 2013. The knowns, known unknowns and unknowns of sequestration of soil organic carbon. Agriculture, Ecosystems & Environment 164, 80-99.

Stroosnijder, L., 2005. Measurement of erosion: is it possible? Catena 64(2-3), 162–173. doi:10.1016/j.catena.2005.08.004.

Sumfleth, K., Duttmann, R., 2008. Prediction of soil property distribution in paddy soil landscapes using terrain data and satellite information as indicators Ecol. Indic., 8 (5), 485-501

Su¨zen, M.L., Doyuran, V., 2004. A comparison of the GIS based landslide susceptibility assessment methods: multivariate versus bivariate. Environmental Geology 45, 665-679.

Syrbe, R.U., Walz, U., 2012. Spatial indicators for the assessment of ecosystem services: providing, benefiting and connecting areas and landscape metrics. Ecological Indicators 21, 80-88. doi: 10.1016/j.ecolind.2012.02.013 .

Takken, I., Beuselinck, L., Nachtergaele, J., Govers, G., Poesen, et al., 1999. Spatial evaluation of a physically-based distributed erosion model (LISEM). Catena 37(34), 431-447.

Tarboton, D.G., 1997. A new method for the determination of flow directions and upslope areas in grid digital elevation models. Water Resources Research 33 (2), 309-319.

Taylor, S., 2003. Extreme terseness: some languages are more agile than others. Lecture Notes in Computer Science 2675, 334–336. doi:10.1007/3-540-44870-5_44.

Taveira-Pinto, F., Petan, S., Mikos, M., Pais-Barbosa, J., 2009. Application of GIS tools for Leç a river basin soil erosion (northern Portugal) evaluation, WIT Trans. Ecol. Envir. 124, 267–278. doi:10.2495/rm090251.

Thapa, P.K., 2010. Physically based spatially distributed rainfall runoff modelling for soil erosion estimation. PhD Thesis, Stuttgart University

The Scipy community, 2012a. NumPy Reference Guide. SciPy.org. http://docs.scipy.org/doc/numpy/reference/

The Scipy community, 2012b. SciPy Reference Guide. SciPy.org. http://docs.scipy.org/doc/scipy/reference/

Tiwari, A. K., Risse, L. M., Nearing, M., 2000. Evaluation of WEPP and its comparison with USLE and RUSLE, Trans. Am. Soc.Agr. Eng. 43, 1129–1135.

Tohari, A., Nishigaki, M., Komatsu, M., 2007. Laboratory rainfall-induced slope failure with moisture content measurement. J. of Geotech. and Geoenvironment. Eng., 133(5), 575-587.

United Nations Convention to Combat Desertification (UNCCD), 2015. Reaping the rewards: financing land degradation neutrality . Bonn, Germany, UNCCD.

Valentin, C., Poesen, J., Li, Y., 2005. Gully erosion: impacts, factors and control. Catena 63(2), 132-153.

Van Asch, Th.W.J., 1980. Water erosion on slopes and landsliding in a Mediterranean landscape. Doct. Thesis, Utrecht University.

Van Beek, R., 2002. Assessment of the influence of changes in climate and land use on landslide activity in a Mediterranean environment. Netherlands Geographical Studies 294, KNAG, Faculty of Geosciences, Utrecht University, 366 p.

Van der Knijff, J. M., Jones, R.J.A., Montanarella, L., 1999. Soil erosion risk assessment in Italy, EUR – Sci. Tech. Res. Rep., EUR 19022 EN, 52 p.

VanderKwaak J.E., 1999. Numerical simulation of flow and chemical transport in integrated surface-subsurface hydrologic systems. PhD dissertation, University of Waterloo, Ontario, Canada.

Van Oost, K., Govers, G., Desmet, P.J.J., 2000. Evaluating the effects of changes in landscape structure on soil erosion by water and tillage. Landscape Ecology 15(6), 579- 591.

Van Oost, K., Cerdan, O., Quine, T. A., 2009. Accelerated fluxes by water and tillage erosion on european agricultural land, Earth Surf. Proc. Land. 34, 1625–1634.

Van Rompaey, A., Verstraeten, G., Van Oost, K., Govers, G. and Poesen, J., 2001. Modelling mean annual sediment yield using a distributed approach. Earth Surf. Proc. Land. 26, 1221–1236.

Van Rossum, G., Drake, F.L., 2011. The Python language reference manual: for Python version 3.2. Network theory Ltd. ISBN: 978-1906966140.

Van Westen, C.J., Rengers, N., Terlien, M.T.J., Soeters, R., 1997. Prediction of the occurrence of slope instability phenomena through GIS-based hazard zonation. Geologische Rundschau 86, 404-414.

Van Westen, C.J., 2000. The modelling of landslide hazards using GIS. Surveys in Geophysics 21(2-3), 241–255.

Van Westen, C.J., 2004. Geo-information tools for landslide risk assessment an overview of recent developments. In: Landslides, Evaluation and Stabilization. Proceed. of the 9th Int. Symposium on Landslides, Rio de Janeiro, 39–56.

Van Westen, C.J., Van Asch, T.W.J., Soeters, R., 2006. Landslide hazard and risk zonation; why is it still so difficult. Bulletin of Engineering geology and the Environment 65(2), 167–184.

Varnes D. J., 1978. Slope movement types and processes. In: Schuster R. L., Krizek R. J. (Eds). Landslides, analysis and control. Transportation Research Board Sp. Rep. 176, Nat. Acad. oi Sciences, 11–33.

Venables, W.N. and Ripley, B.D., 2002. Modern Applied Statistics with S. Fourth edition, Springer.

Venables, W. N., Smith, D. M., R Development Core Team, 2009. An introduction to R: a programming environment for data analysis and graphics. Network Theory. ISBN: 978-0954612085

Verheijen, F. G. A., Jones, R. J. A., Rickson, R. J., Smith, C. J., 2009. Tolerable versus actual soil erosion rates in Europe. Earth-Sci. Rev. 94, 23–38.

Vertessey, R.A., Watson, F.G.R., Rahman, J.M., Cuddy, S.D., Seaton, S.P., et al., 2001. New software to aid water quality management in the catchments and waterways of the south-east Queensland region. In: Proceedings of the Third Australian Stream Management Conference, August 27–29, 611–616.

Vertessy, R.A., Wilson, C.J., Silburn, D.M., Connolly, R.D. Ciesiolka, C.A., 1990. Predicting erosion hazard areas using digital terrain analysis. IAHS AISH Publication 192, 298–308.

Vigiak, O., Okoba, B.O., Sterk, G., Groenenberg, S., 2005. Modelling catchment-scale erosion patterns in the East African Highlands. Earth Surface Processes and Landforms 30, 183–196.

Viney, N.R., Sivapalan, M., 1999. A conceptual model of sediment transport: application to the Avon river Basin in Western Australia. Hydrological Processes 13, 727–743.

von Werner, M., Schröder, D. G. A., Schmidt, J., 2004. Abschätzung des Oberflächenabflusses und der Wasserinfiltration auf landwirtschaftlich genutzten Flächen mit Hilfe des Modells EROSION-3D. Endbericht des FuEVorhabens. Sächsische Landesanstalt für Landwirtschaft, Fa. GeoGnostics, Dresden/Berlin.

Wackernagel, H., 1998. Multivariate geostatistics: an introduction with applications. 2nd edition edn. Springer-Verlag.

Wainwright, J., Parsons, A.J., 1998. Sensitivity of sediment-transport equations to errors in hydraulic models of overland flows In: Boardman, J., Favis-Mortlock, D., Modelling Soil Erosion by Water. NATO ASI Series 55, 271–284.

Walton, R., Hunter, H., 1996. Modelling water quality and nutrient fluxes in the Johnstone River Catchment, North Queensland. In: 23rd Hydrology and Resources Symposium, Sydney.

Wang, X., Xie, H., Guan, H., Zhou, X., 2007. Different responses of MODIS-derived NDVI to root-zone soil moisture in semi-arid and humid regions. J. Hydrol., 340 (1–2), 12-24

Warren, S.D., Hohmann, M.G., Auerswald, K., Mitasova, H., 2004. An evaluation of methods to determine slope using digital elevation data. Catena 58, 215-233.

Wasowski J., Casarano D., Lamanna C., 2007. Is the current landslide activity in the Daunia region (Italy) controlled by climate or land use change. In: McInnes R., Jakeways J., Fairbank H., Mathie E. (eds). Proceedings of the International Conference

on Landslides and Climate Change. Ventor, Isle of Wight, UK, 21–24 May 2007. Taylor & Francis, London, Balkema, Rotterdam, 41–49.

Wasowski, J., Lamanna, C., Casarano, D., 2010. Influence of land-use change and precipitation patterns on landslide activity in the Daunia Appennines, Italy. Quarterly Journal of Engineering Geology and Hydrogeology 43, 387-401.

Wasowski, J., Keefer, D. K., & Lee, C. T., 2011. Toward the next generation of research on earthquake-induced landslides: current issues and future challenges. Engineering Geology 122(1), 1-8.

Wasowski, J., Lamanna, C., Gigante, G., Casarano, D., 2012. High resolution satellite imagery analysis for inferring surface-subsurface water relationship in unstable slopes. Remote sensing of Environment 124, 135-148.

Wasowski, J., Dipalma Lagreca, M., Lamanna, C., 2014. Land-Use change and Shallow Landsliding: a Case History from the Apennine Mountains, Italy. Landslide Science for a Safer Geoenvironment, 267-272.

Wasson, R.J., 2002. What approach to the modelling of catchment scale erosion and sediment transport should be adopted? In: Summer, W., Walling, D.E. (Eds.), Modelling Erosion, Sediment Transport and Sediment Yield. IHP-VI Technical Documents in Hydrology. UNESCO, Paris, 1-11.

Watson, F., Rahman, J., Seaton, S., 2001. Deploying environmental software using the Tarsier modelling framework. In: Proceedings of the Third Australian Stream Management Conference, August 27–29, 631–638.

Wheater, H. S., Jakeman, A. J., Beven, K. J., 1993. Progress and directions in rainfall-runoff modelling. In: Jakeman, A. J., Beck, M. B., and McAleer, M.J. (eds). Modelling Change in Environmental Systems. John Wiley and Sons, Chichester, 101–132.

Whitehead, P.G., Young, P.C. , 1975. A dynamic-stochastic model for water quality in part of the Bedford-Ouse River system. In: Vansteenkiste, G.C. (Eds.), Computer Simulation of Water Resources Systems, North Holland, Amsterdam, 417-438.

Wieczorek, G.F., 1984. Preparing a detailed landslideinventory map for hazard evaluation and reduction, Bull. Int. Ass. Geol., 3, 337–342.

Wilken, F., Fiener, P., Van Oost, K., 2017. Modelling a century of soil redistribution processes and carbon delivery from small watersheds using a multi-class sediment transport model. Earth Surface Dynamics, 5(1), 113-124.

Willems, P., 2014. Parsimonious rainfall–runoff model construction supported by time series processing and validation of hydrological extremes – Part 1: Step-wise model-structure identification and calibration approach, J. Hydrol., 510, 578–590.

Willgoose, G.R., Bras, R.L., Rodriguez-Iturbe, I., 1989. A physically based channel network and catchment evolution model. Technical Report 322, Ralph M. Parsons Laboratory, Department of Civil Engineering, MIT, Boston, MA.

Willgoose, G.R., Bras, R.L., Rodriguez-Iturbe, I., 1990. A model of river basin evolution. Transactions of American Geophysical Union 71 (47), 1806−1807.

Willgoose, G.R., Bras, R.L., Rodriguez-Iturbe, I., 1991a. A physically based coupled network growth and hillslope evolution model: 1 Theory. Water Resources Research 27 (7), 1671−1684.

Willgoose, G.R., Bras, R.L., Rodriguez-Iturbe, I., 1991b. A physically based coupled network growth and hillslope evolution model: 2 Applications. Water Resources Research 27 (7), 1685−1696.

Willgoose, G.R., Bras, R.L., Rodriguez-Iturbe, I., 1991c. A physical explanation of an observed link area-slope relationship. Water Resources Research 27 (7), 1697−1702.

Willgoose, G.R, Bras, R.L., Rodriguez-Iturbe, I., 1991d. Results from a new model of river basin evolution. Earth Surface Processes and Landforms 16, 237−254.

Willgoose, G., Riley, S.R., 1998. Application of a catchment evolution model to the prediction of long-term erosion on the spoil heap at Ranger uranium mine: initial analysis. Supervising Scientist Report 132, Supervising Scientist, Canberra.
Available: http://www.environment.gov.au/resource/application-catchment-evolution -model-production-long-term-erosion-spoil-heap-ranger (archived at: http://www.webcitation.org/6UNHHqeOk).

Williams, J. R., 1975. Sediment yield prediction with universal equation using runoff energy factor. Proceedings of the sediment- Yield Workshop, USDA Sedimentation Laboratory, Oxford, Mississippi.

Willmott, C., Matsuura, K., 2005. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Clim. Res. 30, 79–82. doi:10.3354/ cr030079.

Wischmeier, W. H., Smith, D. D., 1978. Predicting rainfall erosion losses. A guide for conservation planning. US Dept Agric., Agr. Handbook, 537 p.

Witt, A.C., 2005. Using a GIS (Geographic Information System) to Model Slope Instability and Debris Flow Hazards in the French Broad River Watershed. Carolina State University, 165 p.

Wolpert, D. H., 1996. The lack of a priori distinctions between learning algorithms. Neural Computation 8 (7), 1341-1390. https://doi.org/10.1162/neco.1996.8.7.1341 , INRMM-MiD:4301665.

Wolpert, D.H., Macready, W.G., 1997. No free lunch theorems for optimization. Evolutionary Computation, IEEE Transactions 1 (1), 67-82. https://doi.org/10.1109/4235.585893 , INRMM-MiD:404286.

Woodward, D.E., 1999. Method to ptredict cropland ephemeral gully erosion. Catena 37, 393-399.

Wu, W., Sidle, R., 1995. A distributed slope stability model for steep forested basins. Water Resources Research 31, 2097-2110.

Xu, H., Caramanis, C., Mannor, S., 2012. Sparse algorithms are not stable: a no-free-lunch theorem. Pattern Analysis and Machine Intelligence. IEEE Transactions 34 (1), 187-193. https://doi.org/10.1109/tpami.2011.177 , INRMM-MiD:9685435

Yilmaz, I., 2009. Landslide susceptibility mapping using frequency ratio, logistic regression, artificial neural networks and their comparison: a case study from Kat landslides (TokatTurkey). Comput. Geosci. 35 (6), 1125-1138.

Yin, L., Yang, R., Gabbouj, M., Neuvo, M., 1996. Weighted median filters: a tutorial. IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing, 43(3), 157-192. https://doi.org/10.1109/82.486465.

Young, P.C., 1974. Recursive approaches to time-series analysis, Bull. of Inst. Maths and its Applications 10, 209-224.

Young, R.A., Onstad, C.A., Bosch, D.D., Anderson, W.P., 1989. AGNPS: a nonpoint-source pollution model for evaluating agricultural watersheds. Journal of Soil and Water Conservation 44(2), 168–173.

Young, P.C., Lees, M. J., 1993. The active mixing volume: a new concept in modelling environmental systems. Chapter 1 in: Barnett, V., Turkman, K.F. (eds.), Statistics for the Environment, Chichester, 3-43.

Young, P.C., 1998. Data-based mechanistic modelling of environmental, ecological, eco- nomic and engineering systems. Environmental Modelling and Software 13,105-122.

Young, P.C., 2002. Data-based mechanistic and top-down modelling. In: Rizzoli, A.E., Jakeman, A.J. (eds.), IEMSs 2002. International Environmental Modelling and Software Society, Lugano, Switzerland 1, 363–374.

Yue, Y., Ni, J., Ciais, P., Piao, S., Wang, T., et al., 2016. Lateral transport of soil carbon and land−atmosphere CO2 flux induced by water erosion in China. Proceedings of the National Academy of Sciences, 113(24), 6617-22. doi: 10.1073/pnas.1523358113.

Zanchi, C., Torri, D., 1980. Evaluation of rainfall energy in central Italy. In: De Boodt, M., Gabriels, D. (eds), Assessment of Erosion. Wiley, Toronto, 133-142.

Zhang, W., Zhang, Z., Liu, F., Qiao, Z., Hu, S., 2011. Estimation of the USLE cover and management factor C using satellite remote sensing: a review. In: Geoinformatics, 19th International Conference on. IEEE, Shanghai, China, 1–5. doi:10.1109/geoinformatics.2011.5980735.

Zezere, J.L., Rodrigues, M.L., Reis, E., Garcia, R., Oliveira, et al. 2004. Spatial and temporal data management for the probabilistic landslide hazard assessment considering landslide typology. In: Lacerda WA et al. (eds), Landslides, evaluation and stabilization. Proceedings of the 9th international symposium on Landslides, Rio de Janeiro 1, 117-125.

Zezza, F., Merenda, L., Bruno, G., Crescenzi, E., Iovine, G., 1994. Condizioni di instabilita' e rischio da frana nei comuni dell'Appennino Dauno Pugliese. Geologia Applicata e Idrogeologia 29, 77-141.

Zingg, A.W., 1940. Degree and length of land slope as its affects soil loss in runoff. Jour.of Agric Engineering 21, 56-64.

Appendix A - field images

This Appendix is intended to serve as a quick reference to the images captured in the field.

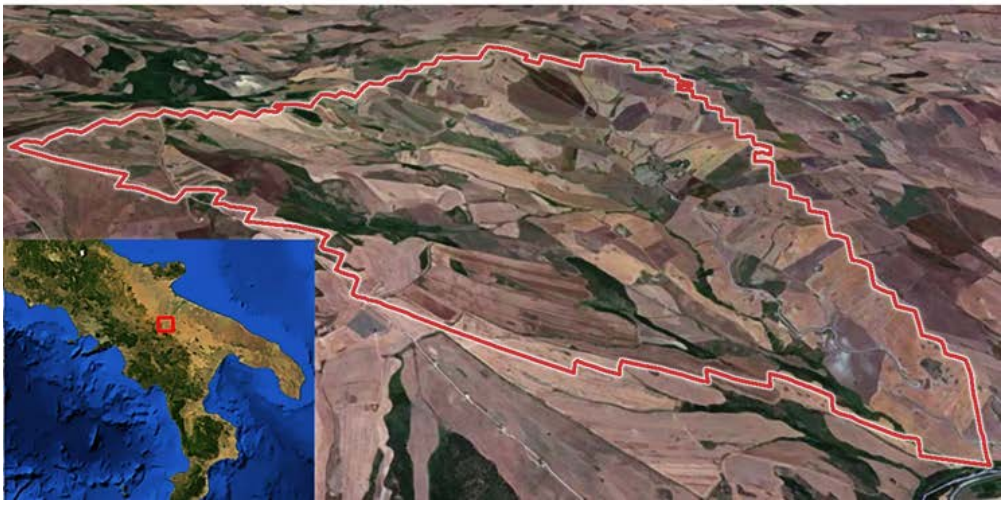


Figure 1.2. (Chapter1) – The picture shows some of the soil erosion processes (rills and ephemeral gullies – top picture) and mass movements (shallow landslides – picture in the middle) that characterize the study site, located in Italy within a catchment close to Rocchetta Sant'Antonio. The image also illustrates the heterogeneous land cover patterns that are present in this area (agricultural areas, grassland, shrubs and forest), with uneven patch size and complex connectivity. These picture were taken in Spring of 2012. (See page 6 of this thesis).

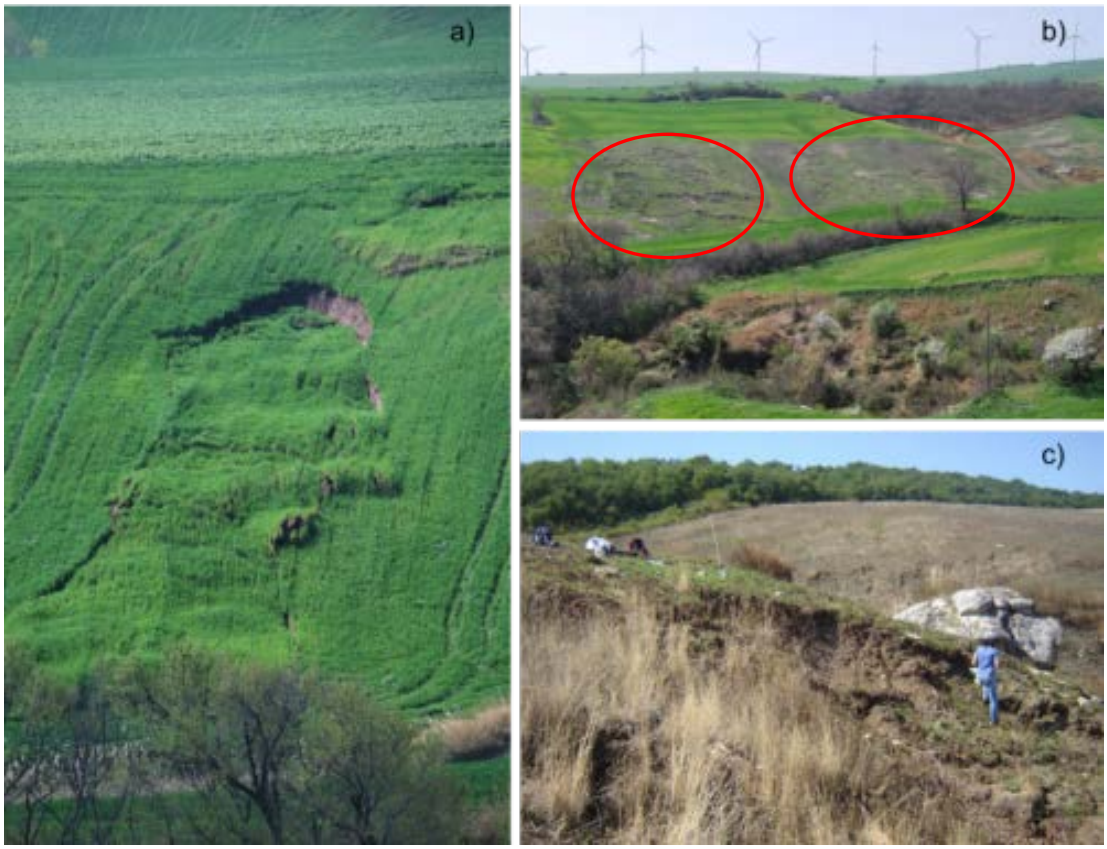


Figure 3.4 (Chapter 3) – Examples of translational slides in the Rocchetta Sant'Antonio catchment (figure (a) and (b)) (spring of 2012) and detail of the main scarp and head of a big translational slide (figure c, October 2012) occurred in the same area highlighted in picure b. All the landslides in the pictures have a length not exceeding a few tens of meters. (See page 62 of this thesis).

Figure 3.5 (Chapter 3) - ephemeral gully (a-c) and rills (b-c) (see section 2.1) in Rocchetta Sant'Antonio. Picture b is related to the fall period (October 2012), pictures a and c were taken in the spring of 2012. The maximum depth of the ephemeral gully in picture (a) is around 80-100 cm, the rills in pictures (a) and (b) are generally of uniform spacing and dimension , have a depth generally below 10 cm and are much more narrow than ephemeral gullies. Geomorphological features having similar dimensions and charactristics are present all over the catchment and especially during the wet season (October-March). (See page 63 of this thesis).

Figure 4.7 (Chapter 4) – In-situ test with a shear vane for determining soil cohesion (see picture 4.6). This picture was taken October of 2012. (See page 95 of this thesis).