FP7 2013 Cooperation Work Programme
Theme 6: Environment (Including Climate Change)

Novel indicators for identifying critical
*INFRA*structure at *RISK* from Natural Hazards

Deliverable D6.4

Suitability analysis of data-mining models for stress tests

Primary Author	Tao Cheng, Khaled Taalab/University College London (UCL)
WP	6.4
Submission Date	15/07/2016
Primary Reviewer	Pieter van Gelder/ Probabilistic Solutions Consult and Training (PSCT)
Dissemination Level	PU

This project has received funding from the European Union's Seventh Programme for research, technological development and demonstration under grant agreement No 603960.

Project Information

Project Duration: 1/10/2013 - 30/09/2016

Project Coordinator: Professor Eugene O' Brien
eugene.obrien@rod.ie

Work Programme: 2013 Cooperation Theme 6:
Environment (Including Climate Change).

Call Topic: Env.2013.6.4-4 Towards Stress Testing of Critical Infrastructure Against Natural Hazards-FP7-ENV-2013-two stage.

Project Website: www.infrarisk-fp7.eu

Partners:

Eidgenössische Technische Hochschule Zürich, Switzerland.

Dragados SA, Spain.

Gavin and Doherty Geosolutions Ltd., Ireland.

Probabilistic Solutions Consult and Training BV, Netherlands.

Agencia Estatal Consejo Superior de Investigaciones Científicas, Spain.

University College London, United Kingdom.

PSJ, Netherlands.

Stiftelsen SINTEF, Norway.

Ritchey Consulting AB, Sweden.

University of Southampton (IT Innovation Centre), United Kingdom.

Document Information

Version	Date	Description	Primary Author
Rev01	28/04/2016	Version 1	Tao Cheng & Khaled Taalab
Final	15/07/2016	Version 2	Tao Cheng & Khaled Taalab

Executive Summary

Landslides are typically triggered by intense or prolonged rainfall. A rainfall threshold, as it relates to landslides, is the amount of rainfall required to trigger slope failure. In order to inform early warning systems, manage risk and put mitigation strategies into effect in a timely manner, it is necessary to predict where and when landslide will occur.

Based on a landslide susceptibility map developed in Deliverable 5.2 and a suite of rainfall data, a Random Forest (RF) algorithm was used to develop a model to predict rainfall threshold values for the Piedmont region of Italy. This model can be used to predict the distribution of rainfall required to trigger landslides within the next 24 hours. The promising results of this model suggest that RF can be used as part of an early warning system which indicates the locations of highly susceptible areas based on measured and forecast rainfall.

Based on historical rainfall data, the model can also be used to 'stress-test' the system. By identifying different rainfall return periods it is possible to test how the distribution of areas which are at risk of landslide occurrence changes (and its size increases) with rainfall events of increasing magnitude.

Table of Contents

1.0 INTRODUCTION

In Italy, landslides are typically triggered by intense or prolonged rainfall (Giannecchini et al., 2012). A rainfall threshold, as it relates to landslides, is the amount of rainfall required to trigger slope failure. In order to inform early warning systems, manage risk and put mitigation strategies into effect in a timely manner, it is necessary to predict where and when landslide will occur. There are two primary approaches to predicting landslide occurrence based on rainfall thresholds. The first are physical, process-based models (e.g. Wu et al., 2015) and the second are empirical models (e.g. Vallet et al., 2015). Physical models require in depth geotechnical characterisation of an area, which is both time consuming and expensive. For this reason, the use of physical models is typically restricted to small, high risk areas. Empirical studies use recorded instances of landslide occurrence and measured rainfall. They can generally be applied to much larger regions. The geographic extent of an empirical rainfall threshold model is determined by the geographic extent of the data used to train the model.

Determining empirical rainfall thresholds is a complex process. Typically, empirical models use rainfall duration, intensity and antecedent rainfall to make predictions of threshold values (Giannecchini et al., 2012). Rainfall thresholds are place dependent, as across a given region the relationship between duration, intensity and antecedent rainfall will vary spatially based on a suite of geomorphological attributes (e.g. slope, elevation, land use). Moreover, Dahal, & Hasegawa (2008) state that a threshold is normally the value above which an event occurs 100% of the time. For landslides, however, the minimum amount of rainfall required to trigger an event is also of (possibly more) interest.

As stated, empirical thresholds are derived from the complex interactions between rainfall duration, intensity, antecedent rainfall and a host of geomorphological factors. The data used to derive these thresholds (i.e. records of historic landslides) generally covers a relatively small area within a large region for which the model is being developed. As such, even with large amounts of training data, there may be relatively few instances of landslides triggered with under similar geomorphological and rainfall conditions. This makes predicting rainfall thresholds across space problematic; however, it is necessary as without this approach, the model will fail to represent how changes in the landscape will affect rainfall thresholds. The process is further complicated by the issue that there are many different types of landslides and different rainfall patterns can be associated with each type. As a generalisation, shallow landslides or debris flows are triggered by shot bursts of intense rainfall, whereas deep landslides are initiated by prolonged rainfall (Martelloni et al., 2012). A comprehensive overview of the multitude of types and triggering processes which can occur is available from Varnes (1978).

Another issue concerning the use of rainfall thresholds to inform early warning systems is the spatial resolution of prediction. Traditionally, when a threshold was set to be exceeded, a warning would be issued for the entire area from which landslide samples were taken to calibrate the model (Segoni et al., 2015). As the spatial resolution of rainfall data is always much larger than that of the landslide area, this is problematic, as in order to be effective, an early warning system requires a greater degree of spatial accuracy. The challenge is to refine the spatial resolution of landslide threshold predictions and develop spatially distributed thresholds that incorporate changes in geomorphological conditions (Li et al., 2010). For this reason, it is preferable to have a range or mosaic of threshold predictions across a given region. While this does not provide the exact location of future landslides, it substantially narrows the areal extent (Segoni et al., 2014).

The purpose of this Deliverable is to build upon the landslide susceptibility maps developed in Deliverable 5.2 to assess the suitability of a data mining approach for the development of empirical models that predict the spatio-temporal occurrence of landslides based on rainfall conditions. Suitability will be assessed by model cross-validation and independent validation using date from landslides-triggering rainfall events that were not used to train the RF model. The suitability of RF models fro stress testing will be demonstrated using a set a rainfall of scenarios. In keeping with the models developed in Deliverable 5.2, a Random Forest data mining algorithm is used to develop these models. Later on in this deliverable, a useful application of the RF rainfall threshold model which has been developed is to test how certain rainfall scenarios, such as major rainfall events may lead to future landslides. This is essential a landslide occurrence 'stress test' based on rainfall return periods.

2.0 METHODOLOGY

2.1 Random Forest

The Random Forest algorithm is based on an ensemble of decision trees, which aim to classify data by recursive partitioning based on some explanatory variable (Figure 1). Decision trees use a set of binary rules to predict a target value, based upon a set of training data containing all data on the conditioning factors (represented by the root node). In this instance the target value is species of iris taken from Fisher's famous Iris dataset (Fisher, 1936).

The algorithm determines both the conditioning factor which most accurately separates the data into species and the threshold value at which to split data. Using the example in Figure 1, petal length is used to as the conditioning variable used to split data in the root node (all the training data). In this example, the threshold value is a length of 2.45 cm. Below this length all records are classified as 'setosa', above this length further splitting is required to determine species type.

Splitting will stop when an internal nodes contains data of only one class (i.e. all the data has been classified). This becomes a terminal node and is classified as a species. If we wanted to predict new data using this model, there would be a lot of misclassification as the model is so finely tuned to a single dataset that it would not be generalise well. To ensure RF models can predict new data, they use an ensemble of decision trees. If the same data was used at the root node and at each split, each decision tree would be identical, making an ensemble pointless. For this reason, RF uses a different subset of the data to train each tree in the ensemble and limits the number of conditioning factors that can be used for splitting at each node. Predictions are made based on the majority vote of all trees in the forest (Breiman, 2001). Bootstrap sampling is used to generate the random subsets of data, which are drawn from the full dataset with replacement.

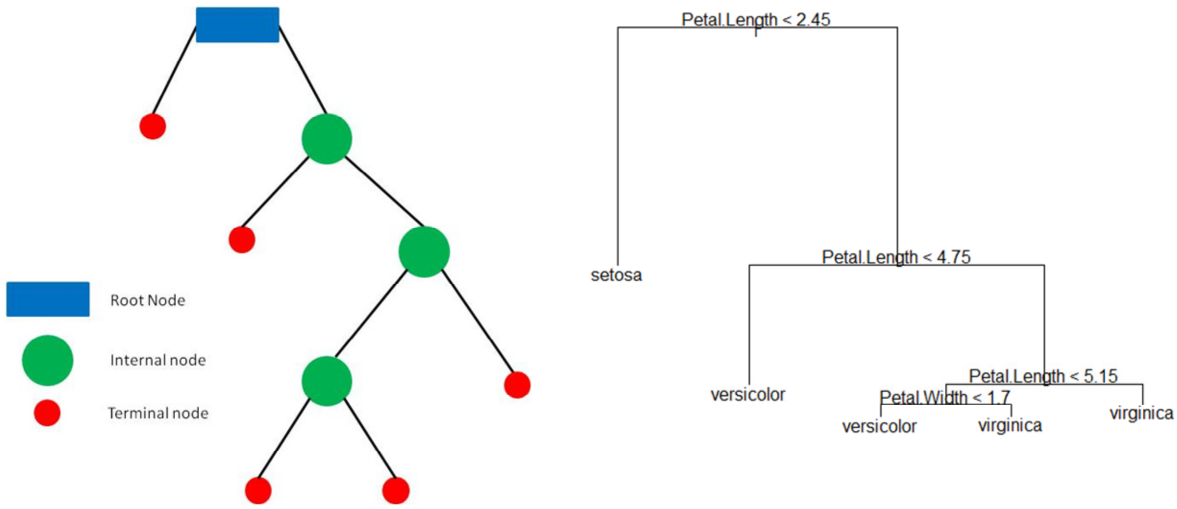


Figure 1: An example decision tree used for classification of Fisher's Iris dataset (Fisher, 1936)

More formally, we have a training dataset T comprising R records. T contains all the geo-environmental data (X) and all the landslide data (Y). The landslide data $Y_1, \mathrm{Y}_2, \ldots \mathrm{Y}_\mathrm{R}$ are binary, indicating the presence or absence of landslides (it should be noted that if a model was trained for a different purpose, for example to predict the size or volume of a landslide, it is possible to model the target variable as continuous rather than nominal). For each landslide sample, there is a corresponding list of environmental variables taken from the same location. If $X_{i1}, X_{i2} \ldots \mathrm{X}_{\mathrm{iR}}$ represent slopes, then X_{i1} is the slope at Y_1.

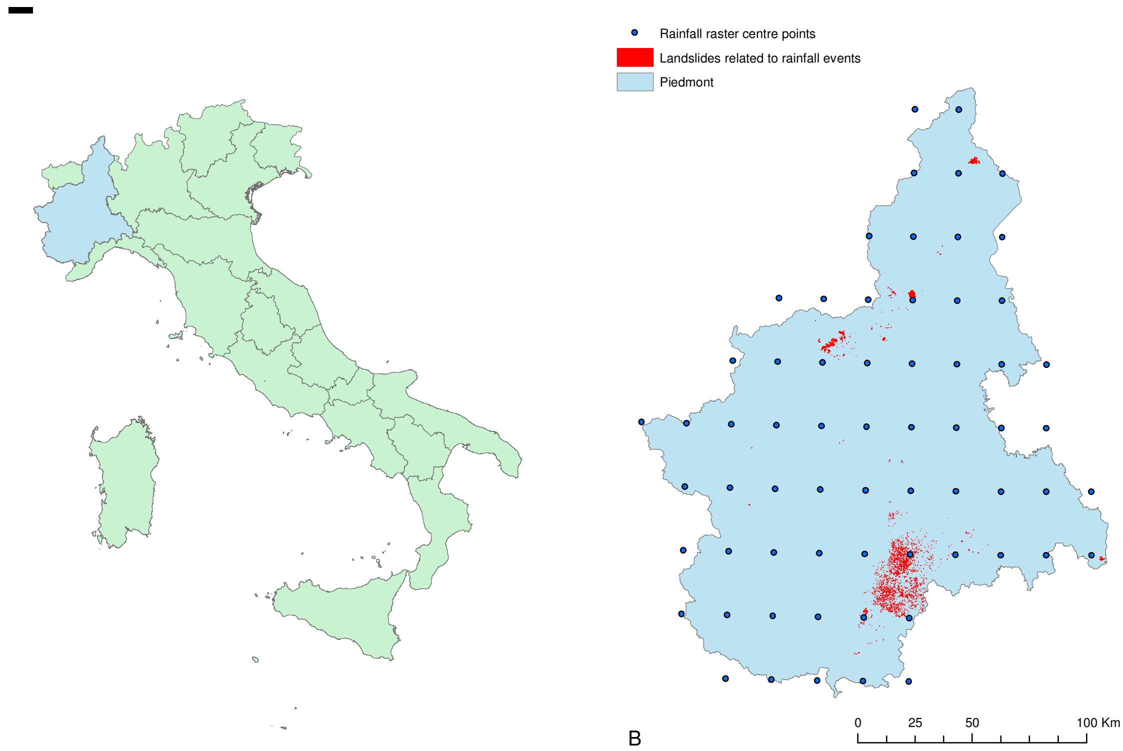
To fit a Random Forest made up of N decision trees, we first generate N equally sized datasets using bootstrap sampling ($\mathrm{T}_1, \mathrm{T}_2 \ldots \mathrm{T}_\mathrm{N}$). This method draws samples with replacement, meaning each dataset T_x will contain many duplicate data records and also have many missing records in comparison to the training data. A decision tree is trained using each of the sample datasets. The number of variables that might be used to split each node is limited to a subset of the total number of variables available (e.g. at every node, splits are made using the best of a limited selection of variables). Classes are predicted using a majority vote based on all trees in the forest. Training many trees using subsets of the entire training dataset and using only a limited number of variables at each node prevents overfitting. Each dataset ($\mathrm{T}_1, \mathrm{T}_2 \ldots \mathrm{T}_\mathrm{N}$), trains a corresponding decision tree ($\mathrm{D}_1, \mathrm{D}_2 \ldots \mathrm{D}_\mathrm{N}$). When classifying data, it passes through each tree ($\mathrm{D}_1, \mathrm{D}_2 \ldots \mathrm{D}_\mathrm{N}$) and its class is determined by a majority vote of all trees. The proportion of votes that a class receives is also used to determine the probability of class membership (Boström, 2007). This is useful for Landslide Susceptibility Mapping (LSM) as it allows predations for be visualised on a continuous scale (e.g. probability of landslides between 0-1) rather than just visualising as susceptible or non-susceptible.

In the model, two parameters need to be determined. These are the number of variables considered for splitting each node (mtry) and the number of trees in the forest (N). It is important to set ntree to a number large enough that the algorithm is stable, but not so large as to become computationally expensive. This parameter can be tuned. In this study ntree was set at 200 based on experimental results which echoed the findings of of Latinne et al. (2001). The mtry parameter is the primary control on the classification accuracy of the model, as it determines both the correlation between any two trees in the forest and the predictive power of each individual tree.

Some variables are better predictors of susceptibility than others, and the best predictors will be selected as a splitting variable at nodes more frequently than worse predictors. For example, if elevation is the key determinate of landslide susceptibility it will be selected for splitting most frequently. The more variables tested at each node, the greater chance that the most influential variables will be selected. This makes an individual tree a stronger predictor, however, it makes individual trees more likely to be highly similar to one another. This means that when averaging across the forest the variance of the model will be high, leading to errors in predicting new data. Increasing mtry will increase the predictive accuracy of individual trees (increasing overall predictive accuracy of the RF) and increase correlation between trees (decreasing overall predictive accuracy of the RF). To balance these effects and find an optimal value, it is necessary to test a number of mtry values. Liaw & Wiener (2002) suggest that using the square root of the number of predictor variables in the model for an initial value of mtry is a good place to begin parameter testing, a 70:30 split between training and validation data is also suggested.

2.2 Study area

To demonstrate the efficacy of RF for spatio-temporal landslide prediction based on rainfall thresholds, the method has been applied to the case study region of Piedmont, a 25402 km^2 region in northwest Italy (Figure 2). Since 1950, in Italy the economic cost of landslides has been more than 53 billion Euros. This region is of particular interest as Piedmont has been identified as being within a landslide 'hotspot' (Jaedicke et al., 2014). For this reason it is an area that may particularly benefit from the development of a model capable of predicting the spatio-temporal location of future landslides.



(A) (B)

Figure 2: (A) Location of Piedmont study area within Italy (B) Location of landslides associated with rainfall events and location of rainfall raster centre points.

2.3 Landslide Inventory data

The landslide inventory used in this study is SiFRAP (Sistema Informativo Frane in Piemonte-Landslide information system in Piedmont) is a dataset containing 30439 landslides dating from the early 20th century to 2006, mapped at a scale of 1:10000 (Lanteri & Colombo, 2013). This is an update of the IFFI (Inventario dei fenomeni franosi in Italia- Inventory of Landslide in Italy) project (Amanti et al., 2001). A comprehensive description of the classification taxonomy is available from SiFRAP (2009). Most of the landslides have not been dated, meaning it is impossible to associate them with rainfall conditions. There are however 3636 dated landslides in the SiFRAP dataset which are associated with caused by eight rainfall events (the majority of these landslides, however, were caused by a highly significant rainfall event on 05/11/1994). The locations of these landslides are shown in Figure 2B.

2.4 Rainfall data

The majority of landslides in Italy are triggered by rainfall (Aleotti, 2004). To apply RF models to the prediction of rainfall thresholds, spatial measurements of rainfall and dated instances of landslides are required. The RF model relates landslide occurrence with rainfall intensity, duration, antecedent conditions and the propensity for landslide occurrence (susceptibility). The daily rainfall data on a 25km grid was taken from the European Climate Assessment & Dataset (Hofstra et al., 2009). A total of 64 grid cells can be used to characterise the Piedmont region. The centre points of each grid square are shown in Figure 2B.

Input data	Description
Rainfall data	Rainfall data on a 25km grid was taken from the European Climate Assessment & Dataset. For each grid cell that covered the Piedmont region (64 in total), had daily rainfall recorded derived from measured data, spanning from 01/01/1950 to 31/12/2014. Empirical thresholds typically based on rainfall intensity and duration metrics, however, in this instance; data covering Piedmont was limited to daily temporal resolution. For this reason, both the rainfall on the day that the landslide occurred and the previous day's rainfall are considered (as well as antecedent rainfall, average annual rainfall and susceptibility). While hourly data is typically used to predict empirical thresholds, the advantage of using daily data is that the data are freely available at the European scale, meaning if models are successfully developed for Piedmont they can potentially be developed for any other region in Europe.
Average annual rainfall	Average annual rainfall (AAR) taken at the 64 cell location is calculated based on daily rainfall values between 01/01/1950 to 31/12/2014. AAR is commonly used to differentiate between areas that typically experience high rainfall and those that do not (Giannecchini et al., 2012).
Antecedent rainfall	Previous rainfall and subsequent soil moisture conditions are an important factor which will influence the rainfall thresholds required to trigger landslides (Wieczorek & Glade, 2005). The precise effects are difficult to quantify and will depend on physical soil characteristics, in particular soil permeability. In soils with high permeability, antecedent rainfall is not considered a significant predictor of landside occurrence, however, in soils with low permeability, antecedent rainfall can have a marked effect on pore-water pressure and hence the amount of rainfall required to trigger a landslide (Aleotti, 2004). Generally, the intensity/duration of rainfall required to trigger shallow landslides decreases as antecedent rainfall increases. The number of days over which antecedent rainfall is considered relevant ranges from three (Dahal, & Hasegawa, 2008) to over 50 (Giannecchini et al., 2012). After testing, the optimal length of antecedent rainfall considered in this study was 10 days, which reflects the findings of Crozier (1999) and Glade at al. (2000).
Landslide susceptibility	Landslide susceptibility maps (LSM) show the relative probability of landslide occurrence across space without considering temporal probability. This can be used to identify the likely areas that a landslide will occur given a rainfall event large enough to trigger mass movement. A landslide susceptibility map for Piedmont was developed in Infrarisk Deliverable 5.2.

Table 1: Input data description

2.5 Training and validation datasets

This study will present landslide threshold predictions and spatio-temporal predictions of landslide occurrence at a 100 m² resolution grid cell format. Although the rainfall data are on a 25 km² grid, the susceptibility map developed in Infrarisk Deliverable 5.2 is at 100 m² resolution. A finer resolution may be more useful for decision makers when planning mitigation strategies and issuing landslide warnings. As the purpose of the model is to identify thresholds values for the initiation of landslides, a seed cell sampling approach was used (Yilmaz, 2010). Each of the 3636 landslides which were known to be associated with a specific rainfall event were sampled once to create a training dataset. The location of the sample was taken from the highest elevation point that intersected each landslide footprint after the landslide locations had been overlaid on a 100 m² resolution digital elevation model (DEM). This method of landslide sampling is used to best represent the pre-failure conditions. Landslide susceptibility, rainfall on the day that the landslide occurred, rainfall on the day before the landslide occurred, antecedent rainfall (previous 10 days) and average annual rainfall is then samples for each of the 3636 points, creating a training dataset.

These data will be used to train the RF models, however, as RF is essentially a black-box modelling procedure (meaning it is difficult to interpret how variables interact within the model and hence how predictions are made), it is important to validate results. As there are limited training data, 10-fold cross validation using the caret package in R will be used to assess the accuracy of the models. Results will be further validated using landslide from a further two rainfall events. 218 landslides associated with a rainfall event on 24/08/1987 and 34 landslides associated with a rainfall event on 14/10/2000 will be used to independently validate threshold predictions. While it would be preferable to validate the model with more rainfall data, the lack of dated landslides in the inventory prohibits this in this instance. If more data became available (future landslides recorded) then the model can be further validated. The known rainfall conditions leading to these landslides will be used to predict threshold values. Areas in exceedance of the threshold can be compared to the known location of landslides. Overlap between threshold exceedance and landslide location (234/252 of the landslides occurred in areas that exceeded the threshold values) suggests the models are performing well.

2.6 Scenario testing

A useful application of the RF rainfall threshold model which has been developed is to test how certain rainfall scenarios, such as major rainfall events may lead to future landslides. This is essential a landslide occurrence 'stress test' based on rainfall return periods. A return period is an estimate of the likelihood of an event, usually based on historic data over a long time period. The return period is the inverse of the probability that an event will be exceeded in any one year. In this case it is used to estimate the magnitude of a rainfall event. For example, a 10 year rainfall event has a 1/10 or 0.1 probability of being exceeded in a given year. The use of return periods are frequently applied in risk-assessment methods as structures and mitigation strategies are designed to withstand an evet of a given return period.

For each of the 64 rainfall points covering the study area, rainfall return periods ranging from five to 200 years were calculated. First daily rainfall data between 01/01/1950 to 31/12/2014 was converted into a time series using the R package "hydroTSM" (Zambrano-Bigiarini, 2014). Then the rainfall data was fit to a Gumbel univariate extreme value distribution function using the "exTremes" R package (Gilleland & Katz, 2011). This was used to estimate the magnitude of the various return period rainfall events over a 2h-hour period.

3.0 RESULTS

3.1 Rainfall thresholds

Predicting when landslides will occur is challenging, as when a landslide occurs on a slope for the first time, the slope is in 'peak strength' conditions, whereas, a subsequent landslide occurs somewhere between peak and residual strength. The two states may have little relation with each other, as once a landslide has occurred, the geomorphological conditions at a given slope may have changed considerably (Guzzetti et al., 1999). In terms of thresholds, this means that two identical slopes (given that most of the conditioning factor data is considered static for the purpose of this analysis) could have very different rainfall thresholds if one of the slopes has previously experienced a landslide. This data is not available as the undated landslides in the inventory make it impossible to specify whether overlapping landslides occurred before or after a dated event.

Due to a lack of data, the RF model to predict rainfall thresholds was validated using 10-fold cross validation. This gave an R^2 value of 0.94 suggesting that the model is highly accurate. RF minimises the chances of overfitting by only drawing on a subset of both the training data to build each tree and only testing a subset of predictor variables at each split. To further ensure the model has not been overfit, independent validation is used. Validation using independent data is limited to two rainfall events, relating to 252 landslides due to a lack of data, however, Figure 3 and Figure 4 act as proof of concept for two independent landslide-triggering rainfall events. Moreover, in another Italian case study, four rainfall events were deemed sufficient to develop empirical rainfall thresholds without validation (Aleotti, 2004). Figure 3 shows the predicted rainfall threshold and observed rainfall triggering event for 24/08/1987. The difference between the two shows the level of threshold exceedance. Landslides observed on 24/08/1987 were located in an area of high exceedance. While there are no landslides shown in the areas of 'high exceedance' (e.g. the South East corner of the region) this does not mean that they did not occur, rather it could be that they were not recorded in the inventory. As the quality and completeness of the landslide inventory used is difficult to quantify, independent validation will always have limitations.

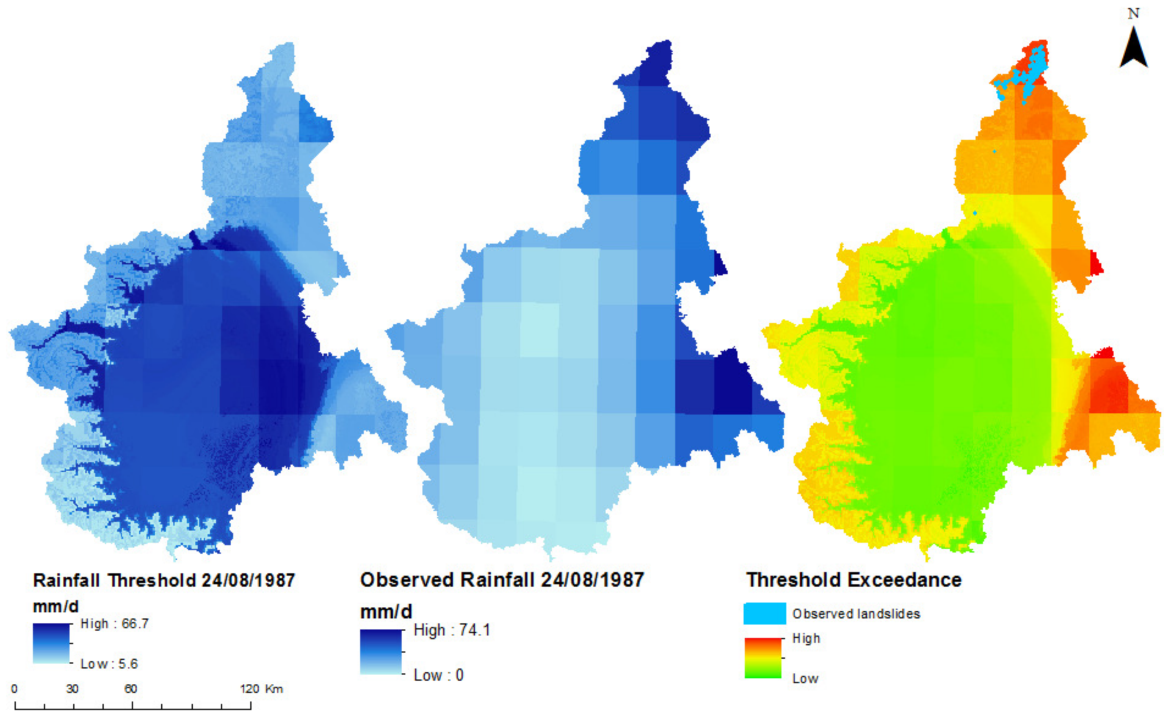


Figure 3: Rainfall thresholds, observed rainfall and levels of exceedance for 24/08/1987

Figure 4 shows the predicted rainfall threshold and observed rainfall triggering event for 14/10/2000. The difference between the two shows the level of threshold exceedance. Landslides observed on 14/10/2000 were located in an area of medium exceedance. The issue for validation using this method is the lack of dated landslide data as well as an incompleteness of landslide inventory data. For example, in Figure 3 there is a second area of high exceedance in the East of the stud area. It is possible that despite this, no landslides occurred and the prediction in that area is wrong. It is also possible that landslides did occur and appear in the inventory but are not dated. It is also possible that landslides did occur but were not reported/recorded. On the strength of the dated landslides within the inventory, is it difficult to dismiss the predictions as incorrect, it is only possible to confirm that the accuracy of the thresholds using observed landslides (of which there are relatively few in the region).

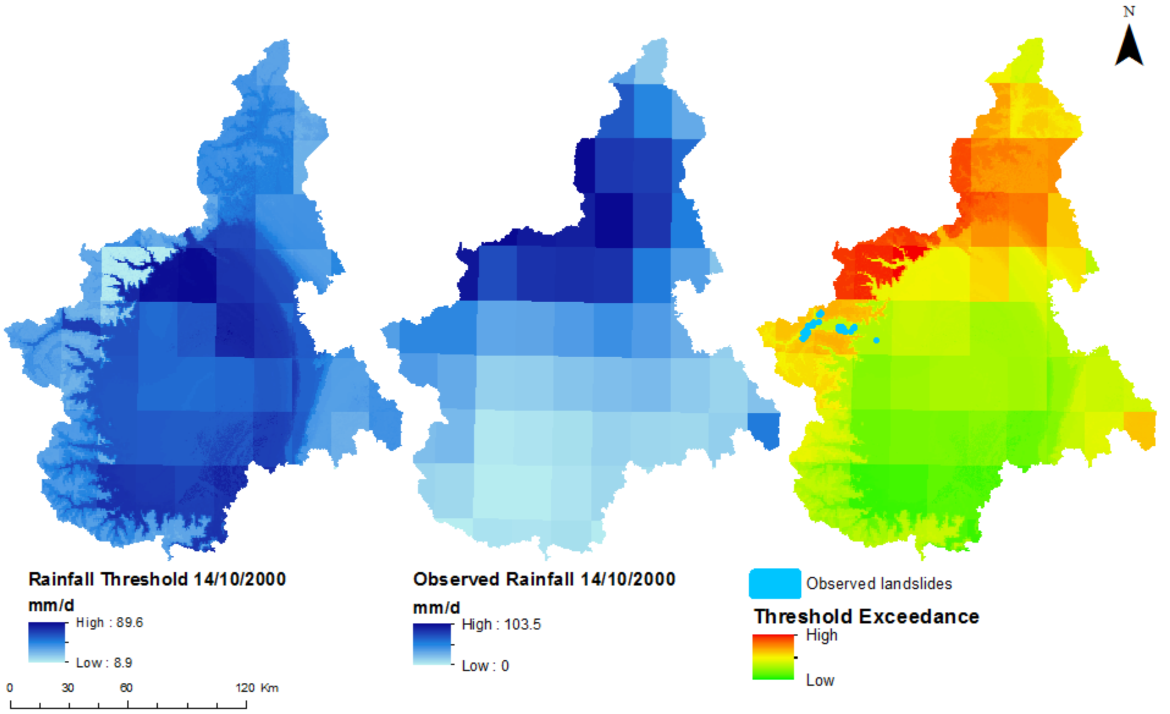


Figure 4: Rainfall thresholds, observed rainfall and levels of exceedance for 14/10/2000

Despite the limitations on validation, the results show that a RF data mining model is a promising technique for rainfall threshold prediction for landslide triggering. Potentially, this could be developed into an early warning system to identify areas which may be affected by landslides on a day-to-day basis. This would require the integration with a forecasting system to compare rainfall predictions with threshold values.

Models were trained in R using the RandomForest package (Liaw & Wiener, 2002). The computer used to train the models had an Intel(R) Xeon(R) CPU E5520 @ 2.27GHz, 2261 Mhz, 4 Core(s), 4 Logical Processor(s) and 12.0 GB installed physical memory (RAM). Training the model on a data frame with 335544 rows and 20 columns took 792.56 seconds. Making predictions for 2538135 records took 105.77 seconds. This shows that this modelling approach may be scaled to larger domains

3.2 Scenario testing

The daily and monthly rainfall time series for a single point are plotted in Figure 5. The results of fitting these data to a Gumbel distribution are shown in Figure 6. This procedure was repeated for all 64 rainfall points covering the Piedmont study area.

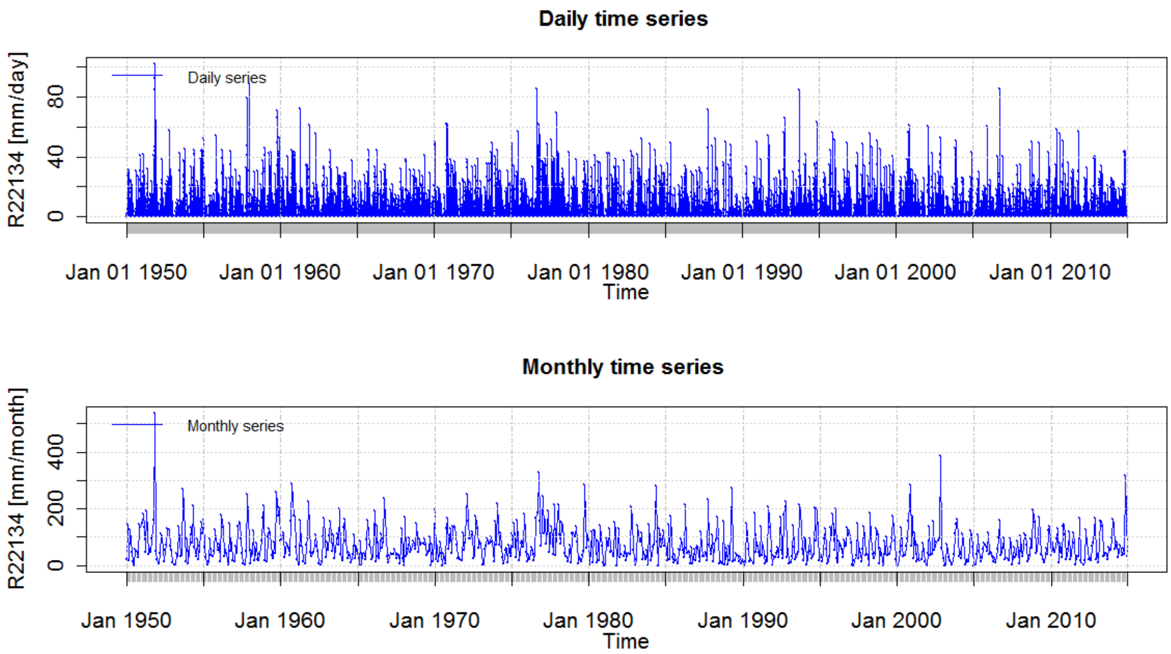


Figure 5: Daily and monthly rainfall time series measurements for a single rainfall point covering the Piedmont region

Fitting the Gumbel distribution to the rainfall time series data of each point means it is possible to calculate the magnitude of rainfall associated with any return period.

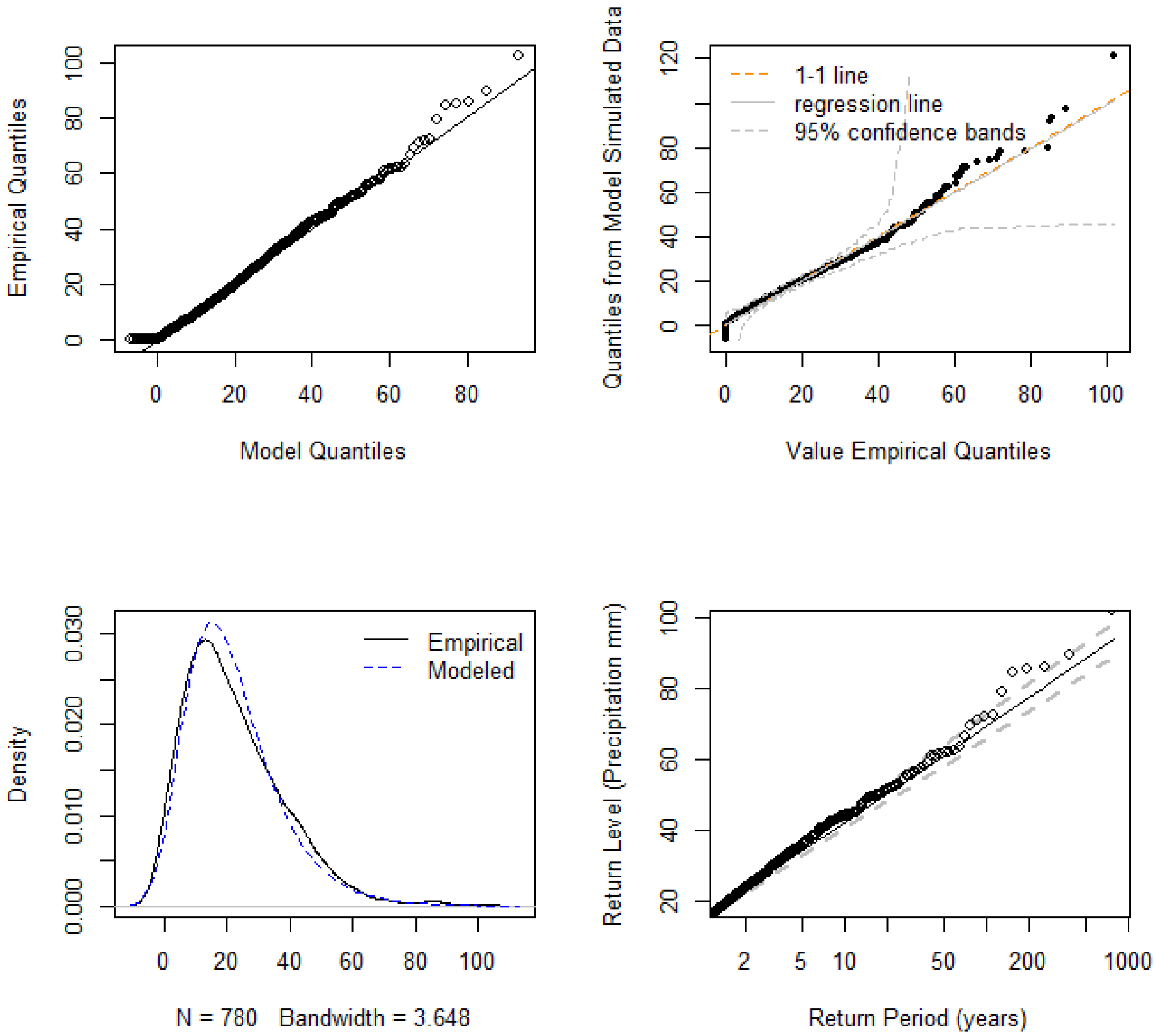


Figure 6: Fitting a Gumbel distribution to the rainfall time series data

For this study, the 5, 10, 50, 100 and 200 year rainfall return periods are mapped in Figure 7. These show daily rainfall ranging from 26 mm for the lowest 5 year return period value in Piedmont, up to 119 mm the highest 200 year return period value.

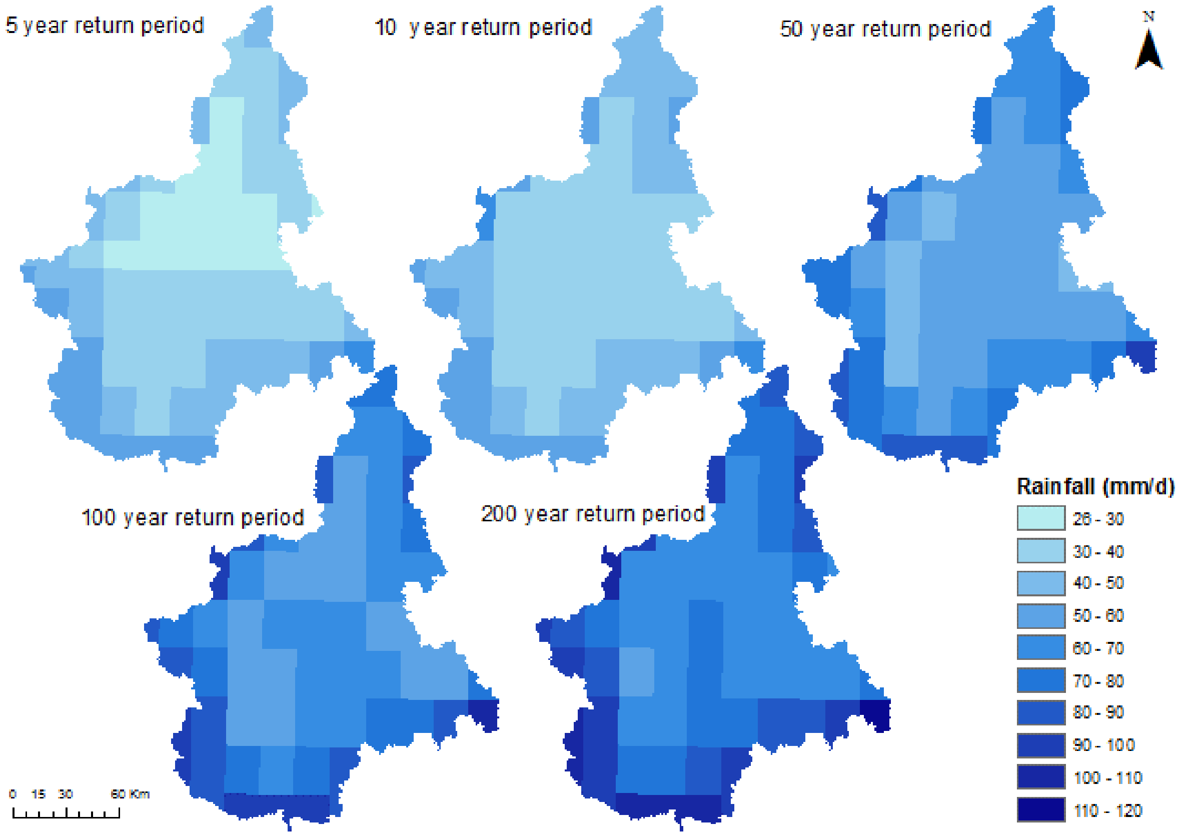


Figure 7: Daily rainfall values at return periods ranging from 5 to 200 years.

To see how these return values compare to the threshold values, it is necessary to give the model further rainfall values (e.g. previous day and antecedent rainfall values). The values used were the mean values from the 3636 previously recorded landslides which can be thought of as representative conditions for the onset of mass movement events. Under these conditions, the rainfall return periods were used as the rainfall event which is compared to the threshold value. The results are shown in Figure 8. As expected, as the return period increases, the percentage of the study areas where the threshold has been surpassed increases. By the 100 year event, the entire region has exceeded the rainfall threshold and is therefore at risk of landslide occurrence. To demonstrate how the RF model can be used for stress testing, the antecedent and previous day's rainfall is kept constant and the stress test focuses on the rainfall event on the day of landslide occurrence. This approach shows how the RF methodology can be applied to stress testing. There are a vast number of scenarios which could be tested. This method can be used to test how the rainfall thresholds are affected by different return period antecedent rainfall, different previous day rainfall and any combination of the predictor variables. If this method became the basis of an early warning system, recorded data and predicted data could be used to predict future thresholds.

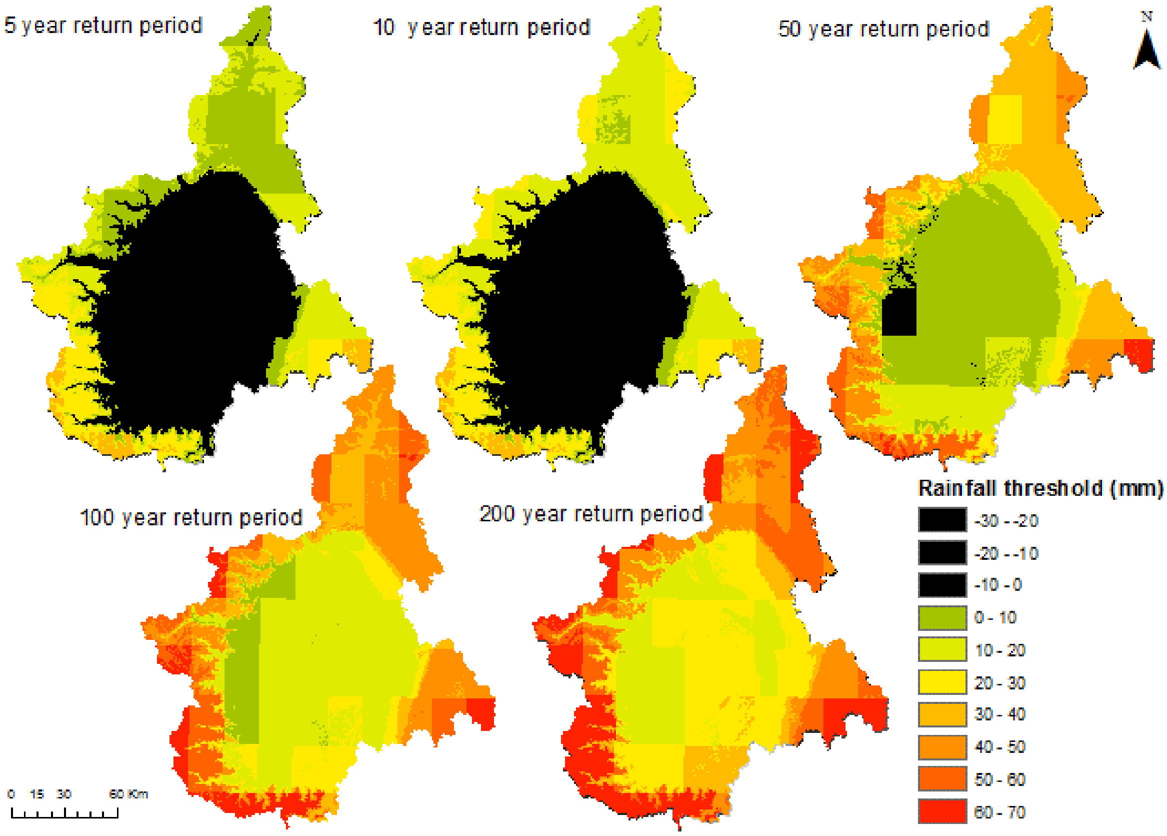


Figure 8: Rainfall threshold exceedance based on 5 to 200 year return period rainfall events.

4.0 DISCUSSION AND CONCLUSION

Results have demonstrated that RF modelling has substantial potential for the development of empirical rainfall threshold predictions and hence the implementation of early warning systems. The cross validation results show that the model is highly accurate and the independent validation shows strong correlation with recorded events (234/252 of the landslides occurred in areas that exceeded the threshold values).

There are, however, further issues to address. Validating results is not straightforward. Much like the independent validation in this report, Li et al. (2010) produced predictive grids using a neural network modelling approach and validated results by comparing areas where landslides occurred with grid cells where the probability of threshold exceedance was beyond 0.7. While this method found good agreement between threshold exceedance and the location of landslides, it did not report the locations of 'false positive' results. These are the areas shown to exceed threshold values where no landslides were recorded. To be effective as a warning system, false positive results should be minimised, however, the issue is further complicated by the quality of the landslide inventory used. The IFFI landslide inventory is a comprehensive record of landslide occurrence in Italy between the early 20th century to 2006, but even in this inventory, not every landslide that has occurred will be recorded. This means that when validating results it is not certain that areas where the threshold was exceeded didn't experience a landslide, even if it is not recorded in the inventory. This problem is compounded if the landslide inventory used for validating predictions is less complete. To improve models, more detailed landslide data is required and more independent validation is needed. Moreover, future work is needed to investigate which data mining technique is best applied to model rainfall thresholds.

Although RF is a black box modelling technique, meaning it is very difficult to establish how the predictor variables interact it does have the advantage of being able to rank variables in order of their contribution to the models predictive accuracy (Liaw & Wiener, 2002) (Figure 9). The most critical predictor of rainfall threshold is the previous day's rainfall, with the previous 10 day's rainfall also an important factor. The average annual rainfall (AAR) and prior susceptibility are less influential. This is to be expected, as susceptibility and AAR are static properties in the region whereas antecedent and previous day's rainfall vary spatio-temporally and it is probable that they will be highly correlated to the rainfall threshold on the day of landslide occurrence.

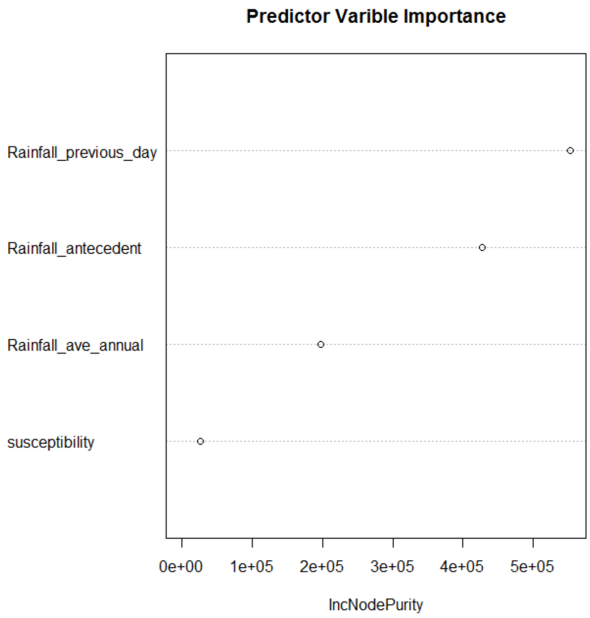


Figure 9: Ranking of predictor variable importance

Scenario testing for various return period rainfall events can be a useful tool to identify the areas most at risk of landslide occurrence. This is also interesting as it can be indicative as to how situations can change in the future. Given current climate change trends, it is likely that the magnitude of current 50 or 100 return period events will change to lower return period events. That is to say the region will experience high magnitude rainfall events more frequently. Potential changes are of interest as in this region, the majority of landsides are triggered by rainfall events in exceedance of the 50-year return period (Alotti, 2004). If this magnitude rainfall occurs more frequently, it is likely that landslide will also occur more frequently.

The results of the threshold predictions and scenario tests are promising for the implementation of a RF modelling approach to inform an early warning system, however, the model does require further refinement. The threshold values are exceeded readily which means widespread warnings would be issued for even the five year return period rainfall event. One option would be to focus on the areas where the exceedance is greatest, however, this approach would require further validation.

Results suggest RF models have a great deal of potential for the analysis of spatio-temporal data relating to natural hazards. Predicting the spatio-temporal relationship between infrastructure behaviour and natural hazards using data mining techniques is limited by data availability. For example, most of the datasets are held by insurance companies and mainly concern the financial cost of damage to infrastructure. Detailed data on the nature of the damage, duration of repair or effect on functionality are typically not available, especially in the quantity that would make a data-mining modelling approach feasible (i.e. a minimum of many hundreds of instances). Where these data are available, RF should be considered a suitable data-mining algorithm for the development of predictive models. The general lack of data is currently a stumbling block preventing the widespread adoption of data-mining modelling techniques of the prediction of the spatio-temporal relationship between hazard and Infrastructure. A further limitation with RF (common with all data mining models) is that they are not readily applied in areas beyond the geographic extent of the training data used in their development. Models would need to be trained again, or at least validated using data from the new area. For Infrarisk this means that it is not possible to apply models developed in Italy to the Croatian case study.

One topic of considerable interest is outlier prediction, which can be used to predict infrastructure failure (e.g. the collapse of a bridge). If enough empirical data were available (e.g. bridge structural parameters, hazard data and data on infrastructure damage/functionality) then, again, RF can be a useful tool to predict failures.

5.0 REFERENCES

Aleotti, P. (2004). A warning system for rainfall-induced shallow failures. Engineering Geology, 73(3), 247-265.

Amanti, M., Bertolini, G., & Ramasco, M. (2001). The Italian landslides inventory–IFFI Project. In Proceedings of III Simposio Panamericano de deslizamientos (pp. 1-2).

Boström, H. (2007). Estimating class probabilities in random forests. In Machine Learning and Applications, 2007. ICMLA 2007. *Sixth International Conference on Machine Learning and Applications* (pp. 211-216). IEEE.

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5-32.

Crozier, M. J. (1999). Prediction of rainfall-triggered landslides: A test of the antecedent water status model. Earth Surface Processes and Landforms,24(9), 825-833.

Dahal, R. K., & Hasegawa, S. (2008). Representative rainfall thresholds for landslides in the Nepal Himalaya. Geomorphology, 100(3), 429-443.

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. Annals of eugenics, 7(2), 179-188.

Giannecchini, R., Galanti, Y., & D'Amato Avanzi, G. (2012). Critical rainfall thresholds for triggering shallow landslides in the Serchio River Valley (Tuscany, Italy). Natural Hazards and Earth System Science, 12(3), 829-842.

Gilleland, E. & Katz, R.W. (2011). "New software to analyze how extremes change over time." Eos, 92(2), pp. 13-14.

Glade, T., Crozier, M., & Smith, P. (2000). Applying probability determination to refine landslide-triggering rainfall thresholds using an empirical "Antecedent Daily Rainfall Model". Pure and Applied Geophysics, 157(6-8), 1059-1079.

Guzzetti, F., Peruccacci, S., Rossi, M., & Stark, C. P. (2007). Rainfall thresholds for the initiation of landslides in central and southern Europe.Meteorology and atmospheric physics, 98(3-4), 239-267.

Hofstra, N., & New, M. (2009). Spatial variability in correlation decay distance and influence on angular-distance weighting interpolation of daily precipitation over Europe. International Journal of Climatology, 29(12), 1872-1880.

Jaedicke, C., Van Den Eeckhaut, M., Nadim, F., Hervás, J., Kalsnes, B., Vangelsten, B. V., Smith, J.T., Tofani, V., Ciurean, R., Winter, M.G., Sverdrup-Thygeson, K., Syre, E. and Smebye, H. (2014). Identification of landslide hazard and risk 'hotspots' in Europe. Bulletin of Engineering Geology and the Environment, 73(2), 325-339.

Liaw, A., & Wiener, M. (2002). Classification and regression by Random Forest. R news, 2(3), 18-22.

Lanteri, L., & Colombo, A. (2013). The integration between satellite data and conventional monitoring system in order to update the Arpa Piemonte landslide inventory. In Margottini, C., Canuti, P., & Sassa, K. (Eds.) *Landslide science and practice* (pp. 135-140). Springer: Berlin Heidelberg.

Latinne, P., Debeir, O., & Decaestecker, C. (2001). Limiting the number of trees in random forests. In Multiple Classifier Systems (pp. 178-187). Springer Berlin Heidelberg.

Li, C., Ma, T., & Zhu, X. (2010). aiNet-and GIS-based regional prediction system for the spatial and temporal probability of rainfall-triggered landslides. Natural hazards, 52(1), 57-78.

Martelloni, G., Segoni, S., Fanti, R., & Catani, F. (2012). Rainfall thresholds for the forecasting of landslide occurrence at regional scale. Landslides, 9(4), 485-495.

Segoni, S., Lagomarsino, D., Fanti, R., Moretti, S., & Casagli, N. (2015). Integration of rainfall thresholds and susceptibility maps in the Emilia Romagna (Italy) regional-scale landslide warning system. Landslides, 12(4), 773-785.

Segoni S, Rossi G, Rosi A, Catani F (2014) Landslides triggered by rainfall: a semiautomated procedure to define consistent intensity-duration thresholds. Comput Geosci 63:123–131

SiFRAP (2009). Guida alla lettura della scheda frane SIFRAP.*Servizio WebGIS Sistema Informativo Frane in Piemonte-http://gisweb. arpa. piemonte. it/arpagis/index. htm*.

Vallet, A., Varron, D., Bertrand, C., & Mudry, J. (2015). Hydrogeological Threshold Using Support Vector Machines and Effective Rainfall Applied to a Deep Seated Unstable Slope (Séchilienne, French Alps). In Engineering Geology for Society and Territory-Volume 2 (pp. 2143-2146). Springer International Publishing.

Varnes, D. J. (1978). Slope movement types and processes. Transportation Research Board Special Report, (176).

Wieczorek, G. F., & Glade, T. (2005). Climatic factors influencing occurrence of debris flows. In Debris-flow hazards and related phenomena (pp. 325-362). Springer Berlin Heidelberg.

Wu, Y. M., Lan, H. X., Gao, X., Li, L. P., & Yang, Z. H. (2015). A simplified physically based coupled rainfall threshold model for triggering landslides. Engineering Geology, 195, 63-69.

Yilmaz, I. (2010). Comparison of landslide susceptibility mapping methodologies for Koyulhisar, Turkey: conditional probability, logistic regression, artificial neural networks, and support vector machine.Environmental Earth Sciences, 61(4), 821-836.

Zambrano-Bigiarini, M. (2014). hydroTSM: Time series management, analysis and interpolation for hydrological modelling. R package version 0.4-2-1. http://CRAN.R-project.org/package=hydroTSM