The Use of Random Forest in Rock-glacier Automatic Detection based on Satellite Imagery

****
University of British Columbia
Vancouver BC V6T 1Z4
****

Abstract

Rock-glacier is an important geomorphological landform in high mountain area. Satellite remote sensing imagery is often used to detect rock-glacier. In this paper, the use random forest in automatic classification of rock-glacier is explored. Five predictor variables derived from remote sensing imagery, together with the truth label (presence/absence of rock-glacier), are used to train the random forest. The testing result exhibits impressive accuracy (>90%). A number of forest parameters are cross-validated. In addition, a novel and intuitive procedure (drop-one test) is proposed to test the relative importance of each predictor variable.

1 Introduction

1.1 Rock-glacier: a geomorphological landform - the label

Rock-glacier (Figure 1) is a fascinating geomorphological landform in high mountain environments (Barsch, 1996). Intuitively, rock-glaciers can be thought of as rocks that have ice and permafrost inside and that have special curved ridge and furrow surface outside. More precisely, rock-glaciers are associated with the presence of ground ice and mountain permafrost belts, and possess a high geo-dynamic and geo-ecologic information value (Harris & Murton, 2005). Their unique surface pattern is mainly resulted from ice core deformation.

Figure 1 Rock-glacier

In the Andes of Santiago de Chile, rock-glaciers occupy c. 10% of the total land surface between 3500 - 4200 m a.s.l. An estimated water equivalent of 0.3 km^3 per 1000 km^2 of mountain area is stored within them (A. Brenning, 2005). The water stored is of great importance to the water supply for the surrounding populous area (A. Brenning, 2005; A. Brenning, 2008). However, not only the stability of high mountain environments is endangered by the predicted and observed global warming (Barsch, 1996), rock-glaciers are also threatened by major human activities, especially large mining projects intending to exploit copper and gold reserves in this area (A. Brenning, 2008).

It is therefore important to explore more accurate and innovative means to monitor rock-glaciers. And particularly, it is interesting to explore what environmental variables (i.e. the predictors) have statistically significant correlation with the presence/absence of rock-glacier (i.e. the label) in certain area. In this paper, the predictive power of a number of environmental variables (e.g. elevation, temperature) will be explored through the use of machine learning algorithm random forest in classifying and predicting rock-glacier.

1.2. Environmental variables of interest – the predictors

There have been a number of studies exploring the geomorphologic, topologic, climatic and environment conditions or controls that determine the limits, continuity and status of rock-glaciers and high mountain permafrost occurrence (Apaloo, Brenning, & Bodin, 2011; Bodin et al., 2009; Bodin, Rojas, & Brenning, 2010; A. Brenning & Azocar, 2010; A. Brenning, 2005; Johnson, Thackray, & Van Kirk, 2007; Smith & Riseborough, 2002). Among these studies, regional and zonal climatic conditions, especially some thermal conditions, such as air temperature, land surface temperature and solar radiation, are found to have close relationship with occurrence and status of regional rock-glacier and high mountain permafrost.

Our study has the aim to explore a number of thermal variables in delineating rock-glaciers in a study area that is rock-glacier abundant. These thermal variables include surface albedo, daytime land surface temperature (LST), nighttime LST and thermal inertia. Because rock-glaciers mostly reside in high mountain area (generally > 3000 m a.s.l), it is imaginably difficult for researchers to conduct on-site observations. As a result, remote sensing technology has been widely used in monitoring of rock-glaciers in high mountain areas. For the purpose of our study, the environmental thermal variables are also derived from satellite remote sensing imagery using digital image processing algorithms, which will be introduced later.

1.3. Machine learning with random forest – the classifier

The machine learning algorithm chosen for this study is random forest. Random forest is a form of "ensemble learning" - methods that generate many classifiers (i.e. decision trees) and aggregate their results (Liaw & Wiener, 2002). The decision trees will be classification trees in our study, as the output will be binary variable (presence/absence of rock-glacier). Each decision tree consists of a number of binary splitting nodes that splits the input dataset into two branches. The leaf nodes will be used for marking the binary classes. Information gain is used as the measure in selecting besting splitting criterion.

The most important characteristic of random forest is its randomness in tree construction process (Pal, 2005). Each decision tree is constructed using a bootstrap (sampling with replacement) sample of the dataset. In addition, unlink in standard trees where each node is split using the best split among all variables, in a random forest, each node is split using the best among a subset of predictors randomly chosen at that node (Liaw & Wiener, 2002).

2. Methods

2.1. Study area

The Andes of Central Chile (33–35°S) is a high mountain area that presents a strong southward trend of climatic conditions and relief (A. Brenning, 2005). Our study area is the Punta Negra valley in the Laguna Negra catchment, in the Western Principal Cordillera near the Andes of Santiago de Chile (33°35' S, 70°5' W, Figure 0). The valley reaches from 2900 m a.s.l to 4100 m a.s.l and is oriented approximately SW – NE, and is bound by SE – and NW – facing ridges that rise to a maximum of 4500 m a.s.l. The Punta Negra valley is composed of a predominantly glacial upper part above ~ 3700 m a.s.l and a lower part with several active and inactive rock-glaciers. In Figure 1, labels with Ax are active rock-glacier abundant areas, and labels with Xx are inactive rock-glacier abundant areas.

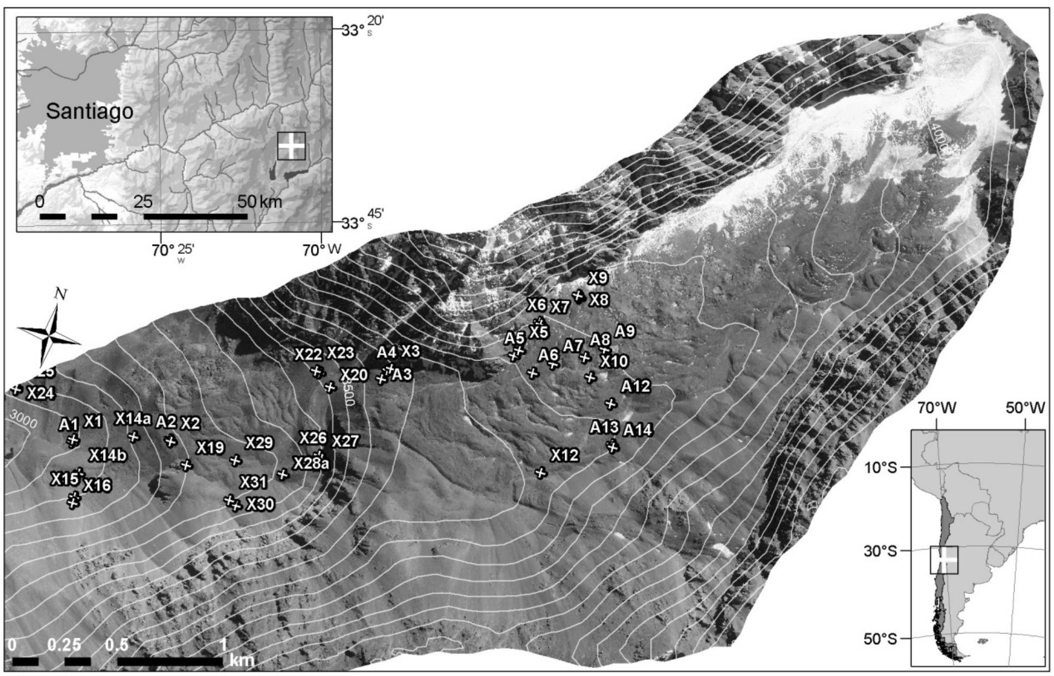

Figure 2 Study Area

2.2. Acquisition of predictor variables

The thermal variables are derived from two ASTER satellite remote sensing images. Of the two remote sensing images, one is a daytime image and the other one is a nighttime image. The two images were taken within 36 hours. ASTER, Advanced Space-borne Thermal Emission and Reflection Radiometer, is an imaging instrument equipped on satellite Terra, which is part of NASA's Earth Observing System. ASTER was launched with Terra in 1999. It is used to obtain detailed remote sensing imageries of land surface temperature, reflectance and elevation (NASA, 2007).

The variables used as input predictors are:

2.2.1. Surface albedo

Surface albedo can be defined as the fraction of incident solar energy reflected by the surface. It indicates the ability of a given surface to absorb energy, which consequently influences its potential to release heat (Peña, 2009). From a macro-perspective, earth surface albedo is an important parameter affecting the global climate (Liang, Strahler, & Walthall, 1999). From a micro-perspective, local surface albedo is governing regional LST and influencing ground thermal regime (Peña, 2009). In terms of glaciology and geocryology studies, albedo of surface rocks was found to relate to depth to ice-cemented permafrost (Bockheim & Hall, 2002). In this study, surface albedo was retrieved from the reflection bands (Band 1 to Bands 9) of the ASTER dataset. For its calculation, a Lambertian surface (i.e. isotropic reflector) was assumed, and conversion formula from narrowband albedo to broadband shortwave albedo was applied according to Liang (2000) (Liang, 2001; Peña, 2009):

$$a_{short} = 0.484 \times a_1 + 0.335 \times a_3 - 0.324 \times a_5 + 0.551 \times a_6 + 0.305 \times a_8 - 0.367 \times a_9 - 0.0015 \quad (1)$$

where a_{short} = shortwave broadband albedo, and a_1, … , a_9 = reflectance of the respective band number.

2.2.2. Daytime and nighttime land surface temperature

Land surface temperature (LST) is the radiant temperature of the land surface layer (Weng & Quattrochi, 2006), and is one of the key parameters in the land-surface processes combining the results of all surface atmosphere interactions and energy fluxes between the atmosphere and the ground. Previous studies have also shown some direct impacts of LST on rock-glaciers: For example, Kääb (2007) noted that variations in surface temperature could indeed affect rock-glacier creep (Kääb, Frauenfelder, & Roer, 2007). In this study, we derived both daytime and nighttime LST of the study area from the pair of ASTER images.

2.2.3. Thermal inertia

Thermal inertia is a volume property that measures the thermal response or resistance power of a material to the changes in its temperature (Nasipuri et al., 2006). Thermal inertia of a material is expressed as:

$$P = (K\rho C)^{1/2} \quad (2)$$

where K is the thermal conductivity, r is its density, and C is the specific heat. Its SI unit is $J/m^{-2} \cdot K^{-1} \cdot s^{-1/2}$. Thermal inertial is an important parameter controlling the thermal regime of a surface, especially affecting its LST. In this study, the algorithm developed by Chen et al. (2008) was used for deriving thermal inertia (Chen et al., 2008) from ASTER images.

2.2.4. Ground elevation

Ground elevation is acquired from the digital elevation model (DEM), which is derived from ASTER imagery.

2.3. Acquisition of output label – presence/absence of rock-glacier

One recently acquired IKONOS remote sensing imagery (spatial resolution = 1 m) was used to manually map the rock-glacier presence/absence within our study area. The resulting imagery is a raster dataset with rock-glacier and non-rock-glacier pixels. The imagery was later converted to ASCII grid format, and can be used as a binary (categorical) variable that has all the pixels as its observations. For each pixel, value = 1 represents presence of rock-glacier, and value = 0 represents absence of rock-glacier. This is the output label for both training and testing purposes.

2.4. Training: construction of random forest

A total of 1798 data points (predictors-label pairs) are available. For training purpose, we randomly selected 2/3 of the points (1198 points). The rest of the data points (600 points) are left for testing purpose.

With the randomly selected 1198 training points, Breiman's (1999) classic algorithm was used to construct the random forest. The greedy philosophy was applied picking the best split at each node. The splits chosen are the "best" at each step, which maximizes information gain, i.e. the difference between pre-split entropy and post-split expected entropy. The general steps of forest construction can be described as follows:

A. For each tree in the forest (for b = 1 to n_{trees}):
   a. Draw a bootstrap sample Z^* of size 1198 points
   b. Grow a decision tree based on the sample drawn from step a, by recursively repeating these steps for each node until the minimum node size $n_{min\text{-}node\text{-}size}$ or maximum depth $n_{max\text{-}depth}$ of tree is reached:
      a) Select $n_{dimentions}$ variables at random from the 5 predictor variables
      b) Using information gain calculation, pick the best split (variable-threshold pair, i.e. the certain value in certain variable that best splits)
      c) Split the node into two daughter nodes

B. Output the ensemble of trees $\{T_b\}^{ntrees}{}_1$. The majority votes will be taken for classification purpose.

In order to construct the random forest, a number of user-defined parameters have to be decided. Different combinations of these parameters have been experimented and cross-validated (in step 2.5) to find the optimal choice. These parameters and their range are as follows (the range is chosen by taking into consideration the computing resource and time available):

---

n_{trees}: the number of trees – [8, 9, 10, 11, 12]

$n_{min\text{-}node\text{-}size}$: the minimum node size – [1, 2, 3, 4]

$n_{max\text{-}depth}$: the maximum tree depth – [5, 6, 7, 8, 9]

$n_{dimentions}$: the number of dimensions used in each best-split finding – [2, 3, 4, 5]

The complete python code (adapted from instructor's) can be downloaded from [this link].

2.5. Testing

Testing was performed using the remaining 600 data points. The predictor variables of each testing point are run through the predictive function of the random forest constructed in the training process. We took the majority votes of the decision trees in determining the predicted class label of each point. The predicted labels were then compared with the truth labels of the testing points for classification accuracy evaluation and analysis.

2.6. Exploring relative importance of each predictor variable – drop-one test

The procedures described in 2.4 and 2.5 focused on examining all predictor variables as a whole in constructing random forest and predicting presence or absence of rock-glacier. However, the relative importance of each predictor variable is not clearly revealed (the best-splitting condition cannot accomplish this since the splitting criterion is a dimension-threshold pair rather than purely dimension). In order to fix this deficiency, a procedure named drop-one test is proposed and performed.

In this test, five new random forests are constructed. Each random forest is constructed with only 4 predictor variables. In other words, we intentionally drop one specific predictor variable in each one of the five forests. The predictive accuracy of the five new drop-one forests are calculated and compared with their corresponding complete forests' (i.e. predictive accuracy acquired from step 2.5). The forest that has the largest accuracy decrease indicates that the variable it drops has the most importance, and vice versa.

In terms of the user-defined forest parameters (e.g. n_{trees}, $n_{min\text{-}node\text{-}size}$), the same combination that produces the best overall accuracy in step 2.4 and 2.5 is used for configuring all five forests.

3. Results

3.1. Predictive accuracy

As is introduced before, a number of combinations of random forest user-defined parameters were experimented. Because there are 5 different tree numbers, 5 different max tree depths, 4 different min leaf nodes number and 4 different dimension numbers (see section 2.4), these result in 400 (=5*5*4*4) different combinations of parameters. In Table 1 below, only the top 20 combinations that have the best accuracy (in terms of the mean of training accuracy and testing accuracy) are shown. The complete set of 400 testing results can be found from this link.

Table 1 Top 20 Results with Best Accuracy

n_{trees}	$n_{max\text{-}depth}$	$n_{min\text{-}node\text{-}size}$	$n_{dimentions}$	Train Accuracy	Test Accuracy	Mean
12	9	1	4	0.9958	0.9850	0.9904
9	9	2	5	0.9950	0.9850	0.9900
10	8	1	5	0.9958	0.9833	0.9896
11	8	1	4	0.9942	0.9850	0.9896
12	7	1	5	0.9942	0.9850	0.9896
8	9	2	5	0.9950	0.9833	0.9892
10	9	1	5	0.9950	0.9833	0.9892
12	8	1	4	0.9950	0.9833	0.9892
12	9	1	5	0.9950	0.9833	0.9892
10	8	1	4	0.9925	0.9850	0.9887
11	9	1	5	0.9950	0.9817	0.9883

8	9	2	4	0.9933	0.9833	0.9883
8	9	3	4	0.9933	0.9833	0.9883
9	9	1	3	0.9933	0.9833	0.9883
10	9	2	5	0.9933	0.9833	0.9883
12	8	1	5	0.9933	0.9833	0.9883
12	9	2	4	0.9933	0.9833	0.9883
10	9	1	4	0.9958	0.9800	0.9879
9	9	2	3	0.9942	0.9817	0.9879

Every one of the 400 results has exhibited a testing accuracy and a mean accuracy greater than 90%. The highest mean is 99.04% while the lowest is 92.91%. The discrepancy between training and testing accuracy is relatively small.

3.2. Predictor variable importance (drop-one test)

The results of performing drop-one test are shown in Table 2. The predictive accuracy decreased when any one of the five predictor variables is dropped. Among them, dropping DEM, daytime LST and thermal inertia have relatively bigger influence on accuracy. This may imply that these variables are relatively more important ones.

Table 2 Drop-one Test Results

Variable Dropped	Train Accuracy	Test Accuracy	Mean Accuracy	Difference from Non-dropped Mean
Surface Albedo	0.9941	0.9865	0.9903	0.0001
Daytime LST	0.9908	0.9800	0.9854	0.0050
Nighttime LST	0.9933	0.9867	0.9900	0.0004
Thermal Inertia	0.9967	0.9783	0.9875	0.0029
DEM	0.9875	0.9482	0.9679	0.0225

4. Discussion and conclusion

Firstly, the impressive predictive accuracy (generally above 90%, some close to 100%) has indicated that random forest might be a useful classification and machine learning technique that can be used to deal with automatic detection of rock-glaciers based on remote sensing imagery. As for future improvement, it will be interesting to testify the same procedures on other rock-glacier study areas. In addition, since remote sensing topics have certain intrinsic similarities, the utility of random forest in other remote sensing topics can also be explored.

Secondly, by experimenting on around 400 different combinations of random forest parameters, there seems to exist an interesting trend: most of the best-performing (in terms of predictive accuracy) have larger maximum tree depth (mostly 8~9), smaller minimum number of leaf nodes (mostly 1~2), and larger number of dimensions for splitting selection (mostly 4~5). It seems forests which are more complex (i.e. higher depth, smaller leaf, and more dimensions) perform relatively better than simpler trees. Although one can argue that complex trees may be more flexible in fitting or even over-fitting data, the consistently impressive testing accuracy can be used to argue against this. In future work, if more powerful computing resource and time is permitted, it will be interesting to explore even more complex forest parameters. In addition, it is tempting to perform more automatic procedures, such as Bayesian Optimization, to choose those parameters.

Thirdly, the drop-one test has revealed that DEM, daytime albedo, and thermal inertia might be the more influential variables in predicting presence/absence of rock-glaciers. This drop-one test is an intuitive procedure. However, the validity of the test should be verified/falsified through more rigorous mathematical proof in future work.

Lastly, though the data points are sampled from the study area randomly, they inevitably still have some spatial correlation between each other. This effect was not taken into consideration when performing this study. It will be important that in the future more effort is made into exploring this issue.

As a conclusion, this paper examined the use of random forest in automatic detection of rock-glaciers. Impressive predictive accuracy is generated. A large number of cross-validations have revealed the effect of different combinations of parameters on random forest. In addition, the proposed drop-one test may be used to explore relative variable importance.

References:

Apaloo, J., Brenning, A., & Bodin, X. (In preparation). *Interactions between snow cover, ground surface temperature and topography (andes of santiago, chile, 33.5°S)* Unpublished manuscript.

Barsch, D. (1996). *Rockglaciers indicators for the present and former geoecology in high mountain environments*. Berlin: Springer.

Bockheim, J. G., & Hall, K. J. (2002). Permafrost, active-layer dynamics and periglacial environments of continental antarctica. *South African Journal of Science, 98*(1-2), 82-90.

Bodin, X., Rojas, F., & Brenning, A. (2010). Status and evolution of the cryosphere in the andes of santiago (chile, 33.5°S.). *Geomorphology, 118*(3-4), 453-464.

Bodin, X., Thibert, E., Fabre, D., Ribolini, A., Schoeneich, P., Francou, B., et al. (2009). Two decades of responses (1986-2006) to climate by the laurichard rock glacier, french alps. *Permafrost and Periglacial Processes, 20*(4), 331-344.

Breiman, L., (1999), Random forests—random features. Technical Report 567, Statistics Department, Universty of California, Berkeley, ftp://ftp.stat.berkeley.edu/pub/users/breiman.

Brenning, A. (2005). Geomorphological, hydrological and climatic significance of rock glaciers in the andes of central chile (33-35 degrees S). *Permafrost and Periglacial Processes, 16*(3), 231-240.

Brenning, A., & Azocar, G. F. (2010). Statistical analysis of topographic and climatic controls and multispectral signatures of rock glaciers in the dry andes, chile (27 degrees-33 degrees S). *Permafrost and Periglacial Processes, 21*(1), 54-66.

Brenning, A. (2005). *Climatic and geomorphological controls of rock glaciers in the andes of central chile: Combining statistical modelling and field mapping.* Unpublished PhD, Humboldt-Universit¨at zu Berlin, Berlin, Germany.

Brenning, A. (2008). *The impact of mining on rock glaciers and glaciers: Example from central chile*

Brenning, A. (2009). Benchmarking classifiers to optimally integrate terrain analysis and multispectral remote sensing in automatic rock glacier detection. *Remote Sensing of Environment, 113*(1), 239-247.

Chen, Z., Li, S., Ren, J., Gong, P., Zhang, M., Wang, L., et al. (2008). In Liang S. (Ed.), *Monitoring and management of agriculture with remote sensing*. DORDRECHT; PO BOX 17, 3300 AA DORDRECHT, NETHERLANDS: SPRINGER.

Harris, C., & Murton, J. B. (2005). *Interactions between glaciers and permafrost: An introduction*

Johnson, B. G., Thackray, G. D., & Van Kirk, R. (2007). The effect of topography, latitude, and lithology on rock glacier distribution in the lemhi range, central idaho, USA. *Geomorphology, 91*(1-2), 38-50.

Kääb, A., Frauenfelder, R., & Roer, I. (2007). On the response of rockglacier creep to surface temperature increase. *Global and Planetary Change, 56*(1-2), 172-187.

Liang, S. (2001). Narrowband to broadband conversions of land surface albedo I algorithms. *Remote Sensing of Environment, 76*(2), 213-238.

Liang, S., Strahler, A. H., & Walthall, C. (1999). Retrieval of land surface albedo from satellite observations: A simulation study. *Journal of Applied Meteorology, 38*(6), 712-725.

Liaw, A,, Wiener, M. (2002). Classification and Regression by randomForest. R News (2/3) 18-22

NASA. (2007). *ASTER homepage.* Retrieved 04/03, 2011, from http://asterweb.jpl.nasa.gov/

Nasipuri, P., Majumdar, T. J., & Mitra, D. S. (2006). Study of high-resolution thermal inertia over western india oil fields using ASTER data. *Acta Astronautica, 58*(5), 270-278.

Pal, M. (2005): Random forest classifier for remote sensing classification,International Journal of Remote Sensing, 26:1, 217-222 Peña, M. A. (2009). Examination of the land surface temperature response for santiago, chile. *Photogrammetric Engineering and Remote Sensing, 75*(10), 1191-1200.

Smith, M. W., & Riseborough, D. W. (2002). Climate and the limits of permafrost: A zonal analysis. *Permafrost and Periglacial Processes, 13*(1), 1-15.

Weng, Q., & Quattrochi, D. A. (2006). Thermal remote sensing of urban areas: An introduction to the special issue. *Remote Sensing of Environment, 104*(2), 119-122.