Nat. Hazards Earth Syst. Sci. Discuss., 2, C1857–C1863, 2014
www.nat-hazards-earth-syst-sci-discuss.net/2/C1857/2014/

Natural Hazards and Earth System Sciences
Discussions

Interactive comment on "Statistical modeling of rainfall-induced shallow landsliding using static predictors and numerical weather predictions: preliminary results" *by* V. Capecchi et al.

Anonymous Referee #1

Received and published: 20 August 2014

General comments

The study by V. Capecchi et al. examines the performance of landslide susceptibility models that incorporate numerical weather prediction (NWP) data as additional predictors. Generalized linear models (GLM) and random forests (RF) are compared. While this topic is of current scientific interest, there are several methodological shortcomings, and the presentation and discussion of results are partially incomplete and lack depth. Perhaps the most significant methodological limitation is the apparent lack of a performance estimation on an independent test set or by using cross-validation.

Furthermore, the relevance of the findings is limited by the fact that the models used here were trained using landslides that resulted from the rainfall predicted by the NWP, i.e. it is merely a hindcasting exercise rather than a forecasting attempt in which a model would be trained on earlier landslide and NWP data (e.g., trained using data from 2011 but applied to predict landslides in 2013 using NWP for the 2013 event).

Overall, based on my assessment substantial additional analyses would be required as well as rewriting of important parts of the paper, including the Results and Discussion.

The following specific comments and technical corrections focus mainly on methodological aspects. I hope that the authors will find them useful.

Specific comments

P4992L25 and P4994L7 Landslides were mapped based on what kind of image source? Indicate sensor, date, resolution? How complete is the inventory? Provide summary statistics of landslide size. Based on this information and the information on P5002L20 it is unclear if all landslide-affected grid cells (including the landslide deposit) were used as landslide observations for modeling purposes (i.e., value of response variable = 1), or, for example, only their scarps or a point ("initiation point") at the centre of the scarp. The interpretation of the hazard map, model performances and variable importances will vary accordingly. Also, how many non-landslide locations were selected for modeling fitting, and how?

P5000 L9 - 45 seems to be an unusually large number of land cover classes, and an impractical one for the purpose of landslide modeling. In the context of RF, for example, there are more than 10 trillion ways of creating a binary split based on a 45-class predictor variable (see Strobl et al., 2007, in BMC Bioinformatics); in the GLM, 44 coefficients need to be estimated when including this categorical predictor in the model.

P5002 L22-23 NWP data was not downscaled from 3 km x 3 km resolution to 30 m x

30 m resolution. Please provide a justification and discuss (in the Discussion) possible implications due to finer-scale variability, considering that high-intensity rainfall events are often spatially very variable.

P5003 L4 provide reference for application of RF in landslide modeling or geomorphology. However, RF may also overfit to spatial data, as seen in other geomorphological applications.

P5003 L24ff - While RF is often praised based on these general characteristics, in the present context it may be appropriate to think about the characteristics of both RF and GLM that make them adequate choices for landslide modeling. In this study, for example, there don't seem to be missing values that would need to be handled by RF (c). Also, (b) GLM doesn't assume a "formal distribution" of predictors either (this is a common misbelief). Likewise, GLM can handle categorical predictors in addition to quantitative ones (a), perform automatic variable selection (in this study, using the AIC) (c), has "little need to fine-tune parameters" (e). GLM is also less in need of cross-validation (d) since it doesn't tend to overfit less than RF. Variable importance (P5004L4) can also be assessed in the GLM based on model coefficients, which have a fairly straightforward interpretation, while on the other hand the permutation-based approach can also be applied to GLM, although this is not very common. Finally, interaction terms can be incorporated into GLMs (less conveniently than in RF though), and nonlinear extensions such as the generalized additive model (GAM) are known in the landslide literature (e.g., Goetz et al., 2011 in Geomorphology).

Overall, rather than praising random forests and announcing "excellent performances" (L29), a more balanced account of the relative advantages and disadvantages of RF and GLM should therefore be presented.

P5003 L15 Please check if stepwise forward variable selection was performed instead of backward, as stated. Fitting a full GLM with >90 predictor degrees of freedom (see Table 4) as a starting point for backward variable selection seems problematic in the

case of the 18 March 2013 event with only 127 landslide observations.

P5004 L7-23 can be shortened substantially and refocused on the relative utility of RF in these applications. E.g., how well did RF perform in studies that compared performances of various techniques in a landslide or geomorphological context? Was overfitting an issue in any of these studies?

The performance measure (AUC) and the procedure used for its estimation are not mentioned in the Results section. Since there is no specific mention of setting a hold-out data set aside, one might get the impression that the training sample was also used for the estimation of the AUC. The result would be an over-optimistic AUC estimate - probably a highly overoptimistic one in the case of RF.

P5005L3-4 and Table 4: The use of slope aspect as a predictor variable (without any transformation) is problematic since this is a directional variable, i.e. 0 degrees = 360 degrees, and e.g. 359 degrees minus 0 degrees = 1 degree. One way to fix this is to use the cosine and sine of slope aspect instead of aspect itself.

P5005L18-20, Figure 5, Tables 2 and 3: Insert a scatterplot showing the correlation between model predictions and observations of 24-h rainfall, in addition to scatterplots of modeled and observed 1-h rainfall. To the tables please add summary statistics of the differences between observed and modeled, since this would provide information on model bias and precision. Discuss these results and performance measures in Section 3.1.

P5008 The text refers to predictors as being "classified" as being important. Increased note purity is a quantitative measure, please indicate in the Methods how you classified based on this measure. Is there a rationale for using this particular importance measure rather than decrease in accuracy or AUROC? In fact decrease in AUROC would seem to be the most natural measure of predictive variable importance in this context since AUROC was chosen as the performance measure and has an actual interpretation, while increase in node purity is a very abstract measure that can only be

interpreted in relative terms.

A table with model coefficients from the fitted GLM should be included in the article, and these coefficients should be interpreted, as far as possible, in terms of odds ratios, especially as far as NWP variables are concerned.

Discussion on P5009 - The text on this page is largely a summary of study objectives and procedures without actually discussing the present findings in the context of the literature, in terms of their broader relevance or with regards to their limitations. The text starting in L23 lists a number of papers that also utilize NWP data in a natural hazards context, without discussing the present results or methods in the context of these studies.

P5010 L13ff - Are AUROC values really comparable among studies in different study areas? The problem with any comparison is that study areas that contain larger "easy-to-predict" portions (such as flat valley floors or less steep forelands) will result in higher AUROC values. In other words, the exact, often arbitrary definition of a study area determines the AUROC to a large extent. Overfitting and the performance estimation technique further influence the reported AUROC values.

A discussion of model results (e.g., model coefficients, variable importances, relationships between predictors and response in RF) in geomorphological or hydroclimatological terms is missing. In the context of landslide prediction, what is the relevance of the present results for actual landslide forecasting (see my General Comment above) - i.e., fitting a model to a landslide inventory and NWP data, and applying it to new NWP data that comes in in near-real time in order to forecast future landslides.

Technical Corrections

Throughout the paper, the authors refer to the statistical models as "indirect" models. There seems to be no justification for using this attribute, it should therefore be omitted.

P4988 In the Abstract, briefly mention AUROC (area under the ROC curve) values (L9)

and variable importance of numerical forecast (L15).

P4988 L2 "a . . . modeling" - rephrase L5 "model's forecast" -> "model forecasts" L5 "combine together": omit "together"

P4999 L27 and throughout the paper: "returning period" -> "return period"

P5000 L11-28 Unfortunately it is not clear from this text how the EVI variable was derived from 'raw' (EVI?) data for the years 2000-2013. Phenology is mentioned at some point (L14) but there is no mention of specific measures of phenology. Precisely what EVI was used, e.g. the overall maximum value of these 14 years, or some multiannual average, or the date of seasonal maximum EVI, etc?

P5002 L22 - "to nudge" - reword L26 (and elsewhere) "In bibliography" -> "In the literature" - provide reference L27 (and possibly elsewhere) "forecasting" - This is not forecasting since the models were trained based on the outcomes that are to be predicted. Prediction is a more general term that would apply here.

P5003 L1 "GLM model" - omit "model" since the "M" in GLM stands for "model" (also in L7 and possibly elsewhere). This sentence can be simplified since the models used "are" the GLM and RF (not "based on") L3 "see below for references" - please insert references here instead - "below" could be anywhere in the article L5-6 "No interactions. . ." - I suggest to omit this sentence since it letting different prediction methods interact seems uncommon. L24 needs to be rephrased

P5004 L24-29 can be omitted

P5005 L1 "R translation" -> "R implementation"

P5005 L20-25 Avoid sentences that either just report a method in the Results section (methods should be introduced in the Methods section), or that simply point to a figure or table without providing an actual summary or interpretation of the information that the reader is referred to. (Please also check the rest of the Results for such sentences, e.g. the first and third sentence on P5007 does not say anything substantial. Same

for P5007L14-16. L21 It would appear that scatterplots would provide more direct evidence of the relationship and difference between modeled and observed rainfall amounts compared to contingency tables and Fig. 6. I suggest to omit Fig. 6.

P5006 L7-8 This “shift” should be referred to as a “model bias”.

P5007 L10-13 can be omitted. L20-22 AUC values how estimated? Use more precise terminology, e.g. “on the training set” or “out-of-bag estimate of”, as applicable. L23-L2 on following page: Move to Methods.

P5008 “All these layers are highlighted. . .” (two occurrence): Please omit this information from the text, it should be in the figure caption.

P5008L24 Please rephrase, adding a particular set of predictor variables does not seem to constitute the development of a “statistical framework”.

P5009L4 “...no interactions. . .” Please omit, this seems obvious and not relevant for the Discussion.

P5012L1 Avoid repetition from P5011L23-24.

Table 4 should mention the 45 classes of the land cover variable. What unit is “h” in “h m-1” for curvature? (Certainly not ‘hour’.)

Figure 2 should be integrated into Figure 1.

Figure 3 could be omitted since this is a standard procedure that is not greatly modified by adding NWP data.

Figure 4: improve readability by using different (larger?) symbols for rain gauges.

Interactive comment on Nat. Hazards Earth Syst. Sci. Discuss., 2, 4987, 2014.