**Authors response to comments by referee #1**

Manuscript reference number: nhess-2018-85

Manuscript title: Evaluation of predictive models for post-fire debris flows occurrence in the western United States.

We would like to thank the reviewer for his/her constructive comments. We have revised the manuscript according to these comments and below we provide our detailed response to each of their comments. Comments from the reviewer are in black font and our response in blue font.

Note: Revised manuscript is attached as supplementary material.

**Reviewer #1**

GENERAL COMMENTS

**1.1** The contribution “Evaluation of predictive models for post-fire debris flows occurrence in the western United States” by Efthymios I. Nikolopoulos and co-Authors is good and potentially publishable. The Authors in this study we investigate the potential to improve the efficiency of current predictive models with machine-learning approaches. The analysis is based on a database on post-fire debris flows published by United States Geological Survey. In general, the topic well address scientific questions within the scope of NHESS.

Response

We would like to thank the reviewer for his/her comment.

**1.2** The theoretical background is well-argued. Review of literature seems complete. The description of study area is not sufficiently complete. The description of methodology and successive parts of paper are well organized but can be improved. Results, and discussion sections are short compared to amount of work done. They should be widely increased. The readability of the whole paper is sufficient with a quite good English. Overall, the work presents some carelessness and incompleteness. It can be published on NHESS journal only after a major revision.

Response

We have revised the manuscript according to reviewer’s general comments (expanding results and discussion section) and specific comments (see response to specific comments below).

SPECIFIC COMMENTS

**1.3** I have some specific comments that should be addressed before the manuscript can be accepted for publication.

Response

We have addressed in detail all specific comments (see response below)

**1.4** Section 1, “Introduction” (Page 2-3).

At Page 3, Lines 23 - 25 the Authors should add references about the rainfall accumulation-duration thresholds, and the two model used.

Response

We have added references for the rainfall accumulation-duration thresholds as follows

“…*include i) rainfall accumulation-duration thresholds (Guzzetti et al., 2007, Cannon et al., 2011; Rossi et al. 2017; Melillo et al. 2018),…”.*

The two models (RF-ED and RF-all) were developed in this work, therefore at that point of the manuscript only the random forest algorithm is referenced.

**1.5** Section 2, “Study area and data” (Page 4, Lines 1-30).

This paragraph is confuse and not complete. The Fig.1 does not show well the study area, the four different regions, the location of the DF, and the area of fire-affected catchments. In addition, the Authors should better explain the differenced normalized burn ratio (dNBR) used in your analysis (Page 4, Lines 13-14). For a better clarity and understanding of the text, Authors should specify in detail how the values of the average erodibility index (KF-Factor) for the four different regions were chosen (Page 4, Lines 16-21).

Response

We have revised Figure 1, please see our response to comment 1.19 below.

We have revised text to include more info on dNBR as follows

*“Information on burn severity was based on the differenced normalized burn ratio (dNBR) (Key and Benson, 2006), calculated from near infrared and shortwave infrared observations, which is frequently used for classification of burn severity (Miller and Thode, 2007; Keeley, 2009). Severity classification from dNBR was validated from field observations provided by local burned area emergency response teams.”*

For soil erodibility index we have updated text and references. Specifically, we provide reference to STATSGO database that was used to derive information on Kf factor and revised the text as follows

*“Finally, since in burned areas changes in recovery vegetation increase erosion, the average erodibility index (Kf factor) derived from the STATSGO database (Schwartz and Alexander, 1995) is reported in the database as well. Kf factor provides evidence of erodibility of soil, taking into account the fine-earth fraction (<2mm).).”*

**1.6** Section 3, “Seasonality and characteristics of rainfall events” (Page 5-7).

At the page 5, Lines 16-24 the Authors refer to statistics about the seasonality and characteristics of rainfall events reported in Table 2. The total number of the events is normalized in terms of percentage, this well describe the distribution of rainfall event analyzed for the different regions. However, this does not allow a comparison by number of events for each region. This comparison could be useful to guarantee the correctness of the considerations set out in Section 3 (Page 6, lines 4-28) and in Section 4.1.3 (Page 9, lines 6-8). I suggest to integrate in the Table 2 the absolute and relative frequency of the DF and noDf events.

Response

We have modified/updated Table 2 per reviewer’s suggestion. The revised Table 2 now reports both the actual number of events and monthly distribution of both DF and noDF events.

**1.7** Section 4.1.2 “Logistic regression” (Page 7-8).

The LR model works with either continuous or categorical independent variables, or a combination of the two types, regardless whether they present a normal distribution or not (Costanzo et al., 2014). The underlying mathematical relationship between the dependent dichotomous variable (presence/absence of a landslide in the mapping unit and the *n* independent variables (*X*1, ...,*X*n). Authors adopt the logistic regression (LR) model using a set of explanatory variables (*X*1, *X*2, and *X*3) that have in common the cumulated rainfall (max 15 min rainfall accumulation). My question is: the explanatory variables used for the Authors are really independent? For a better clarity and understanding, Authors should specify in detail this aspect.

Response

Based on our interpretation of the reviewer’s comment, we assume that he/she wonders whether inclusion of the max 15min rainfall accumulation in all explanatory variables affects the dependence between explanatory variables.

As it is shown in the original version of the manuscript, the three variables *X1,X2,X3* are defined as the product of max 15min rainfall accumulation with the 1)average normalized dNBR, 2) proportion of upslope area burned with gradients >23deg,and 3) soil KF-Factor. Therefore, the correlation (i.e. dependence) between those 3 variables (dNBR, upslope burned % and KF-factor) is not affected by the multiplication with the rainfall accumulation (i.e. correlation between x1 and x2 is the same as correlation between a\*x1 and a\*x2). Essentially, in the formulation of explanatory variables (proposed by Staley et al. 2016, 2017) the 15min rainfall accumulation is used to weigh the three variables mentioned above.

We clarify this in the revised version by stating that

“*Based on this formulation, information on the maximum 15 min rainfall accumulation is used to weigh the other three parameters (upslope burned area, average dNBR and KF-factor) considered.”*

**1.8** Section 4.1.3 “Random forest” (Page 8-9).

Authors stated that “*The first model (RF-ED) was developed using the variables of rainfall accumulation and duration. It is the model that we consider as the one with minimum data requirements, given that only two rainfall variables are used for the prediction.*” (Page 9, Lines 10-12), but in the Page 9, (Lines 2-5) the Authors indicate they have used also an extra categorical variable (named “Region class”). The authors should better explain the total number of variables used and verify if it is in agreement with Table 3.

Response

We would like to thank the reviewer for pointing this inconsistency. We have revised the text appropriately according to the following

*“The first model (RF-ED) was developed using the variables of rainfall accumulation, duration and region class. It is the model that we consider as the one with minimum data requirements, given that only two rainfall variables and a region classification is used for the prediction.”*

**1.9** Section 4.3 “Identification of thresholds” (Page 10, Lines 11-25).

This section is very short. Authors should better argue and comment the obtained results. I suggest to include in the paper a new Figure that show the distribution of the (D,E) pairs with the rainfall thresholds and their uncertainties. In addition the Authors should better explain how they reconstructed the rainfall conditions responsible of the DF and noDF. In particular, if have been used the subjective or objective methods (Vessia et al., 2014; Rossi et al. 2017; Melillo et al. 2015, 2018).

Response

We would like to note that section 4.3. belongs to “Methodology” section and as such, it is not the appropriate section to present results. In section 4.3. we describe the methodology for selecting the thresholds in E,D space (for ED model) or probability space (for LR and RF). To improve our description for the ED threshold we have updated Figure 4 to demonstrate an example of selected ED threshold that maximizes threat score.

Rainfall characteristics were calculated using the approach described in Kean et al. (2011). We have added the following in the manuscript to clarify this aspect.

*“According to the description of the dataset provided in Staley et al. (2016), the rainfall characteristics (peak intensities, accumulation etc) were calculated using a backwards differencing approach (Kean et al., 2011).”*

Uncertainty is a very important aspect and we would like to thank that reviewer for highlighting this. Although we do not explicitly present the uncertainty in the model parameters (e.g. for ED models) we present the variability in the model performance as a result of the sampling uncertainty during the random sampling validation. To better describe this and discuss in more detail the importance of uncertainty, we have revised the text in section 5.1 as follows:

*“Additionally, an important note is that overall the variability of the performance of all models, for a given sample size, is considerable and this essentially highlights the effect of sampling uncertainty; an aspect that requires careful consideration for the development and application of such predictive models. “*

and conclusions section as follows:

*“Uncertainty is a very important element to consider when developing and evaluating predictive models of this nature. Two important sources of uncertainty pertain to estimation of input variables (e.g. rainfall, burn severity) and sampling. In this work, we implicitly demonstrated the impact of sampling uncertainty on model’s prediction skill through the random sampling exercise but we did not account for uncertainty in input parameters. The impact of input parameter uncertainty will be a topic of future research.”*

**1.10** Page 1, Line 21: After Page 1, Line 20 it seems that there are two parts in the text that are not well connected. I suggest to rewrite better the start of this part.

Response

Page 1 refers to the abstract of the manuscript. We believe that rewriting the abstract is not required.

**1.11** Page 1, Line 24: I suggest to introduce here the acronyms of the four regions (AZ, CA,CO, and NM), because are here that are cited for the first time.

Response

We agree with the reviewer and we revised accordingly.

**1.12** Page 7, Line 21: I suggest to replace “both debris and no debris flow events” with “both DF and noDF events”

Response

We revised according to reviewer’s suggestion.

**1.13** Page 8, Line 1: The acronym “PFDF” does not exist. I suggest to introduce PFDF in the previous part (Page7, Line 7)

Response

We revised according to reviewer’s suggestion.

**1.14** Page 8, Line 12, 14, 15: I suggest to replace “*X1*, *X2*, *X3*” with “*X*1, *X*2, *X*3”.

Response

We revised according to reviewer’s suggestion.

**1.15** Page 10, Line 18, 22: I suggest to replace “NO DF” with “noDF”.

Response

We revised according to reviewer’s suggestion.

**1.16** Page 11, Line 7: I suggest to replace “NO DF” with “noDF”.

Response

We revised according to reviewer’s suggestion.

**1.17** Page 19, Table 1: I suggest to replace “(DF=1,NO DF=0)” with “(DF=1, noDF=0)”

Response

We revised according to reviewer’s suggestion.

**1.18** Page 22, Table 1: I suggest to replace “(NO Debris Flow)” with “no Debris Flow” and “(No Debris Flow)” with “no Debris Flow”.

Response

We revised according to reviewer’s suggestion.

**1.19** Page 26, Figure 1: Please show better in the Figure the study area, the four different regions, the location of the DF, and the area of fire-affected catchments. In addition, insert horizontal scale and North indicator symbol.

Response

We have revised Figure 1 following reviewer’s suggestions. Specifically, the figure is further zoomed over the regions of interest and the boundaries of the different regions are now outlined to improve visibility. Horizontal scale and North symbol was added. However, it is not possible to show the area of fire affected catchments. The area of fire affected catchments is in the majority less than 1km2 and is not possible to show it on the same figure for all catchments involved. Please note that the purpose of Figure 1 is to demonstrate the geographic extent of the database used and also provide an overview of the location of the events in the database.

**1.20** Page 27, Figure 2: In the blue (AZ region) and black (CO region) boxplots the median is not well visible. Please change the colours using other with major contrast. In addition I suggest to replace in the caption of the Figure “Figure 2 Boxplot for storm duration, storm accumulation and peak 15-min storm accumulation” with “Figure 2 Boxplot for (a) storm duration, (b) storm accumulation and (c) peak 15-min storm accumulation”.

Response

We have revised caption and figure according to reviewer’s suggestion. We changed the color of the horizontal line corresponding to median values to white, to make it visible for the blue and black boxplot.

**1.21** Page 28, Figure 3: I suggest to replace the caption with “(a) Total rainfall accumulation vs duration and (b) peak 15 minute rainfall accumulation vs duration for Arizona (AZ), California (CA), Colorado (CO) and New Mexico (NM). Colored dots and x symbols correspond to DF and noDF occurrence respectively”.

Response

We revised according to reviewer’s suggestion.

**1.22** Page 30, Figure 5: I suggest to use a different type of the line (solid line) to represent the full range of variation. The dashed-line produce confusing.

Response

We revised according to reviewer’s suggestion.