
Title: Predicting wildfire burned area in South Central US using integrated machine learning techniques

This paper proposed a machine learning method to predict gridded monthly wildfire burned area during 2002-2015 over the South Central United States and identify the relative importance of the predicted factors on the burned area for both the winter-spring and summer fire seasons. The method is able to alleviate the problem of unevenly-distributed burned area data due to the grid-level resolution. The result is interesting and constructive to some extent.

The authors said the machine learning method can achieve the R2 value of ~ 0.4 . However, this result is hard to say it as a high accuracy. The authors can consider to compare more machine learning algorithms, such as AdaBoost, XGBoost. The low accuracy will also affect the reliability of the importance of predicted factors.

Therefore, I would recommend a major revision.

- P1L25: have been seen increasing?
- P2L51: You said the machine learning methods were used to estimate total burned area aggregated over a large-scale domain in past studies. In this study, you focus on the grid-level resolution. Could you describe what the resolution of past studies is? Do these works have the issue of the unbalanced distributed burned area?
- P3L83: The small fire is less than 10 ha, and the large fire is greater than 100 ha. However, the small fire is defined as less than 25 ha, and the large fire is greater than 150 ha in P2L58. Could you explain why they are different?
- I don't think Figure 2 looks nice and it should be re-organized better. For example, the arrow between step 1 and step 2 confuses that there may be an input-output relationship. In fact, is it correct that they are independent processes?
- P3L90: The description of the four steps is not very clear. This paragraph should be rewritten. Is it correct the quantiles are the x-axis of frequency histogram? Why do you choose these quantiles? Will the pre-defined parameters induce uncertainties?
- P3L90: Although the authors claim that the four steps method will alleviate the problem of uneven-distributed dataset, the multi-steps will introduce some risks. For example, if the second step wrongly classifies the burned area as the non-burned area, the bias will be amplified because it won't enter into step 3.
- P4L126: Please explain the assumption or give the reference that grids with larger burned area will have more right-shifted burned area distribution than the

distributions of the training set.

- P6L171: Please add references.
- P8L223: Please add the F1-score performance criteria because you mentioned it in L235.
- P8L235: Could you please plot the AUC curve so that it could help to analyze the TP rate and FP rate. You can also analyze the F1-score performance by Precision and Recall. This will help to understand whether the classifier is underprediction or overprediction.
- P8L235: The performance accuracy of the classifier and the regressor in Table 2 is not very high. Typically, the F1-score of a good classifier can achieve over 0.8 and the RMSE of a good regressor is lower than 0.2. Could you compare your results with some other machine learning methods, such as Adaboost, XGBoost.
- What does “630” mean in Table 2?
- P8L247: Please add some references for past studies.
- P8L251: You compare your results with Chen et al. (2016) and Liu and Wimberly (2015). I wonder whether they are comparable if they are under different factors, different periods and different regions.
- Please explain the meaning of the blue line in Figure 3.
- P9L281: Although the importance of Random Forest help to identify some key factors, they depend on the accuracy of the machine learning method. If the accuracy is not very high, it will reduce the reliability of the information. On the other hand, the importance can't provide how the change trend of factors affect the prediction.
- P9L299: I can't find the reference “Westerling and Bryant (2008)” in the reference list.
- There are several obvious typos in the manuscript, and the English language is poor. I think the authors should be asked to have the manuscript proofread by a native English speaker before the article can be considered for publication in a scientific journal.