GMDD

Interactive
comment

Interactive comment on "A Bayesian Framework Based on Gaussian Mixture Model and Radial Basis Function Fisher Discriminant Analysis for Flood Spatial Prediction (BayGmmKda V1.1)" *by* Dieu Tien Bui and Nhat-Duc Hoang

Dieu Tien Bui and Nhat-Duc Hoang

hoangnhatduc@dtu.edu.vn

Authors' reply to Reviewer 1's comments

Journal Name: Geosci. Model Dev Manuscript: Bayesian Framework Based on Gaussian Mixture Model and Radial Basis Function Fisher Discriminant Analysis for Flood Spatial Prediction (BayGmmKda V1.1)

Reply to Reviewer's Comments

Reviewer's comment: The authors use a data-driven tool to establish a relationship

between 10 "flood influencing factors" and the flood itself in a district in Vietnam. While there is a merit in experimenting with statistical tools for trying understand flood occurrence, the current version of the manuscript lacks to demonstrate such a value in using data-driven algorithms for flood prediction. particularly, I have concern with the description and evaluation of the tool. Unless a substantial improvement is made both in the methodology framing and presentation of the manuscript, I would not decremented for publication on GMD.

Response to reviewer's comment: We thank the Reviewer 1 for giving times and expertise to constructively comment on our manuscript. To address your concerns, we have carefully revised and made a substantial improvement in the description and evaluation of the tool. In addition, the methodology framing and presentation of the manuscript have been carefully checked and improved. We believe that the manuscript is a meaningful contribution to the literature because this is the first time the BayGmmKda tool is proposed for flood study with very promising results.

1) There is no clear objective of the work. And it is not clear how the "tool" can be used for flood mapping or prediction (?). A more focused and tailored description of the tool would be helpful to understand and potentially use for the readers of GMD.

Response to reviewer's comment: The objective of the work is to construct a probabilistic model, named as BayGmmKda, for spatial modeling and prediction of flood in Central Vietnam. This region has been critically damaged by floods in recent years due to climate changes and poor land planning. Thus, this model can be very useful since it helps to accurately and reliably construct a flood susceptibility map for this region. Another objective is to employ advances machine learning algorithms including the Gaussian mixture model with the expectation maximization as well as unsupervised training methods and the Radial Basis Function Fisher Discriminant Analysis. The superiority of the proposed model is demonstated via comparisons with previously used machine learning approaches including metaheuritic-trained Adaptive neuro fuzzy inference system, Support Vector Machine, and Bayesian classifier.

In this study, prediction of flood zones relies on an assumption that future flood events are governed by the very similar conditions of flooded zones in the past (Tehrany et al. 2015; Tien Bui et al. 2016). Thus, past records of flood occurrences, coupled with conditioning factors of the areas, are employed as data instances that help to establish the probabilistic model. We formulate the flood assessment problem as a supervised learning task. Therefore, the data samples collected in the past are employed to train the proposed BayGmmKda. With the model structure identified through the training phase, the model can then be used to make assessment on the flood susceptibility for all studied region. The probabilistic model is coded in Matlab enivronment as an easy-to-use toolbox to assist decision makers in flood prediction.

The application of the tool as well as it practical usefulness are demonstrated in the section 5.2 and 5.3 of the manuscript. In these two sections, the model's outstanding accuracy is clearly shown and the flood susceptibility map of the studied region constructed by the tool is demonstrated. Thus, we believe the tool can also be a promising alternative for similar tasks in other studied regions. Based on the reviewer's suggestion, we will address the reviewer's concern by adding more focused and clear decription of the BayGmmKda tool in the revised version of the manuscript for the sake of GMD's readers.

2) The abstract is too short and lacks details of what they attempt to do.

Response to reviewer's comment: Thanks for your comment. We will extend the abstract to describe the study with more details.

3) There is no definition of flood/no flood. In fact it is not clear at all as to what flood is made in the paper. I think flood extent maps should be used for the evaluation, instead of just the selected points. As it is currently used, then streamflow should be used for the evaluation.

Response to reviewer's comment: We thank the reviewer for the comment and would like to explain to you as follows:

Flood points are flood locations that occurred in the study areas, and have been determined based on documentary sources of the Tuong Duong district and interpretation of Landsat 8 Operational Land Imagery. Using DEM, these flood areas were converted to flood points. In addition, flood locations were collected during field works using handhold GPS. A total of 76 flood locations that occurred during the last five years were prepared.

Non-flood points were randomly generated from non-flood areas within the study area based on DEM, i.e. ridges (we has used DEM to generate topographical shades i.e. flat, Ridge, Saddle Ravine, Convex hillside, Saddle hillside, Slope hillside, Concave hillside, Inflection hillside).

Because the above information is available in our previous paper published in Journal of Hydrology, we have provided a citation for this reference in section 3.1 Flood inventory map and flood conditioning factors of the study area) within the revised manuscript. We copy the text here for your review:

"In this study, the flood inventory map established by Tien Bui et al. (2016) was used to analyse the relationships between flood occurrences and influencing factors"

Regarding your comment "it is not clear at all as to what flood is made in the paper", all the floods in this study are flash flood. This is the main flood type in this study area due to characteristics of the terrain. Moreover, we use flood points because flood extent maps are not available. Thus, we employs flood points provided from the sources of local authority and handhold GPS.

Regarding your comment on the streamflow being used for the flood evaluation, in fact, we have performed a literature review on the use of streamflow for evaluating flood susceptibility. However, we found no relevant or feasible guidances to construct the flood susceptibility model for the studied area based on the available data. Thus, we'd like to consider the possibility of using streamflow for flood evaluations in a future research. This direction will be stated in the conclusion of the revised version.

4) While describing the methodology (section 3), there is no connection made between the statistics and the physical flood characteristics? For example, what are the classes (in the classification of section 3.1) deal with?.

Response to reviewer's comment: We'd to thank reviewer for these comments and we totally agree with the reviewer's opinion at this point. We will provide explanations on the connection made between the statistics and the physical flood characteristics in the beginning of the section 3 of the revised manuscript. We copy the texts in the revised manuscript for your review: "The flood modeling in this study is considered to be a binary classification problem within which 'flood' and 'non-flood' are the two class labels of interest. As a result, the probability of pixels belonging to the flood class, which are derived from the model, will be used as susceptibility indices. These susceptibility indices of the pixels are then used to generate the flood susceptibility map."

5) The paper fails to explain the physical relationship between the "Influencing factors" (Table 1) and the flood processes. And why were those particular factors selected?. How about antecedent soil moisture and other potential factors?

Response to reviewer's comment: We agree with the reviewer on this comment. Based on the reviewer's comment, we have provided texts in the revised manuscript with the pertinent reference to explain the physical relationship between the "Influencing factors" (Table 1) and the flood processes as well as the reason why we choose those particular influencing factors. We copy the texts from the revised manuscript here for your review: "In our previous works of Tien Bui et al. (2016), the physical relationships between influencing factors and flood processes were analyzed. Based on the findings, a total of ten influencing factors were selected in this study, including slope (o), elevation(m), curvature, TWI, SPI, distance to river (m), stream density (km/km2), NDVI, lithology, and rainfall (mm)."

Regarding the comment "How about antecedent soil moisture and other potential factors?", we'd like to explain as follows: In fact, the selection of the conditioning factors

varies from one study area to another based on different characteristics of each place. One variable can have high degree of impact in flooding in a specific area, but it can be without any influence in another regions (Kia et al. 2012). In this study, due to the data availability, we have not employed antecedent soil moisture as a conditioning variable. However, we appreciate the reviewer's suggestions and we think that further studies should be carried out to investigate the influences of antecedent soil moisture and other potential factors for the study regions in Vietnam. This point will be addressed in our conclusion in the revised version.

6) Poor writing throughout. The following is partial list. - L14: to facilitate - L20: cause heavy loss of - L23-24 is that number refers to annual deaths? - L26: the country - L28: 60% of the area in the country is ... a report produced by - L33: It is possible - L47: what does "sceintific manner" means? - L51-52: that is not an accurate description of Dottori etal., because they also provide a water depth. The are based on physical models as well. - L63: can yield - L77: Wha tis that exactly the limitations of the hydrolgical models? And what is the limitations of the proposed method? - L109: by far a heavly affected - L110: located between - L112: Doesn't watershed include mountains and rivers? - L119: have been damaged... must be relocated - L125; reasonable strategy - many more language corrections through out the text! - A more common terms in flood community such as probability of detection and false alarm ratio (rate) can be used - Remove the background color from figure 2 (the region outside of the study region should be white)

Response to reviewer's comment: We'd like to thank the reviewer for your great help. All addressed grammatical and presentation issues will be addressed in the revised version. The whole manuscript has been proofread to improve the writing.

Reference Kia, M. B., Pirasteh, S., Pradhan, B., Mahmud, A. R., Sulaiman, W. N. A., and Moradi, A. (2012). "An artificial neural network model for flood simulation using GIS: Johor River Basin, Malaysia." Environ Earth Sci, 67(1), 251-264. Tehrany, M. S., Pradhan, B., Mansor, S., and Ahmad, N. (2015). "Flood susceptibility assessment

using GIS-based support vector machine model with different kernel types." CATENA, 125, 91-101. Tien Bui, D., Pradhan, B., Nampak, H., Bui, Q.-T., Tran, Q.-A., and Nguyen, Q.-P. (2016). "Hybrid artificial intelligence approach based on neural fuzzy inference model and metaheuristic optimization for flood susceptibilitgy modeling in a high-frequency tropical cyclone area using GIS." J. Hydrol., 540, 317-330.

—————————————————————

Interactive comment on Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-311, 2017.