Natural Hazards
and Earth System
Sciences

Discussions

EGU

Interactive comment on "Bag-of-words-based anomaly-detection principal component analysis and stochastic optimization for debris flow detection and evacuation planning" *by* Chia-Chun Kuo et al.

Anonymous Referee #1

Received and published: 15 November 2017

The ms presents an interesting model to detect debris flows by a "bag-of-words" based procedure and to improve the evacuation planning by stochastic optimization technique. The model is applied to two villages located in Taiwan.

The topic is adequate for NHESS and the proposed model is rather novel, but unfortunately, it cannot be accepted in the present form, since there are many major and minor critiques that should be corrected before acceptance (or re-submission).

Major critiques:

1) The explanation of the model is not very clear and really difficult to follow for a reader, who is not expert in machine-learning and stochastic analysis. I propose two major improvements: a) the structure of the entire model and its different parts must be clearly described at the beginning of the Method-Section. You should use Fig. 2, which is now in the Result-Section, for this general description. b) You may use less equations and more explanations. The description is now very complex: you incorporate 25 equations and many (too much) parameters.

2) The Result-section must significantly improve, because there is a lack of information or description. In the present form (now only two pages!), two tables are the main results. The four graphs in Fig 3-6 are not clear and only commented in 6 lines without adding details, interpretation etc. The lack of explanations and details is a major problem throughout this section. Some examples: a) Time series for rainfall and soil water content are analysed without basic information on the location of these sensors, scan rate, data transmission etc. b) No values and details on the thresholds are given. In L185, the authors introduce a rainfall method and warning without additional explanations or at least references. c) Different parameters were examined (L211-212), but no values (or thresholds) of this sensibility analysis is discussed in detail. I propose to show some graphs on this topic.

3) The authors mention a separation into training and test data (L183), but after reading the Result-section, it's not clear, what was training and what was test/validation of the model. Please, clarify this aspect.

4) The part of the evacuation planning is explained in the Method-section, but not even mentioned in the Results. Please comment this.

5) Introduction-section must be improved: - L34-51 is a long description of the methodology and must principally be placed in the Methods-Section. - the description of the goals is missing

6) Add information on the monitoring stations (already criticised in 2a). Location and

detailed description of all sensors are necessary to understand the results. Only some doubts that I have: how the exact time of debris flow was determined? The detection of the events is recorded near the villages, not in the debris flow initiation zone (normally there is a delay between triggering and arrival in the village).

Minor critiques: I) English should be corrected and improved. II) citation must be corrected (L25: order; L109: format etc.) III) The Section "Results and discussion" lacks of a substantial discussion. IV) L175-181: improve the description of the debris flow episodes and add references. "Debris flows of 450000m3 were triggered" is not very illustrative. V) L199: explain, how you define a "correct prediction" (prediction within the same day, the same floating 24h-window etc.?) VI) "remaining training data" is not clear VII) L225-230: as explained above, additional information and discussion is needed (soil water content sensor is at which depth? How many sensors do you use? What is the "average water content" and how it was determined? etc.) VIII) Table 1 and 2: indicate the correct predictions by bold event times IX) Figures must be improved: - more figures are necessary. - Fig 3-6 are unsubstantial in the present form and must be improved. Merge Fig3-4 and Fig5-6. Add arrows to indicate debris flow triggering, warning levels etc.