Susceptibility Zoning of Dust Source Areas by Data Mining Methods over Khorasan Razavi Province

Mahdi Boroughani[i]: *Assistant Professor of Research Center for Geoscience and Social Studies, Hakim Sabzevari University*
Sima Pourhashemi**:** *Geomorphology PhD in Geographic Faculty of Hakim Sabzevari University*

Article History (Received: 2016.9.16 Accepted: 2020.01.14)

Extended abstract

1- Introduction

Dust storms are natural hazards that effect on weather conditions, human health and ecosystem. Atmospheric processes are directly affected by the absorption and diffusion of radiation by dust, and dust in the cloud acts as a nucleus of congestion. The main dust areas in the world are flat topographically dry areas with erosion-sensitive soil and poor vegetation that is easily eroded by wind. Atmospheric processes are directly affected by the absorption and scattering of radiation by dust, and the dust in the clouds acts as a condensation nucleus. Iran is exposed to numerous local and trans-regional dust systems due to being in the arid and semi-arid belt of the world. Recent observations indicate that this phenomenon is increasing in the country and region under study. Khorasan Razavi province has arid and semi-arid climates with numerous dust storms occurring annually. In some years, more than 50 dust events have been recorded in the province. The large amount of dust in this area indicates large harvesting areas across the province, indicating the importance of research in this area. The purpose of this study was to evaluate the factors affecting the occurrence of dust and to prepare its susceptibility map in Khorasan Razavi province using logistic regression and random forest regression models.

2- Methodology

Khorasan Razavi Province with an area of 143864.118 square kilometers and with a population of 5.999529 is the fourth largest province in the country. The highest point of the province is in the highlands of Binalood, north of Neyshabur, at an altitude of 3211 meters above sea level, and the lowest point of the province is 300 meters above sea level, on the border of Iran and Turkmenistan. Dust indices include BTD_{2931}, D, NDDI and BTD_{3132}, which were calculated for all satellite images of these indices. Then there are four false color combination methods including 1- R: NDDI, G: B4, B: B3, 2- R: D, G: BTD_{2931}, B: NDDI, 3- R: D, G: BTD_{3132}, B: NDDI and 4- R: BTD_{2931}, G: B4, B: B3, were used to determine the best method for identifying dust source. Initially, 65 dust sources were identified in the study area and for the period 2005-2016. Of the identified focal points, 70% (46 dust source) were used for modeling and 30% (19 dust source) were used for evaluation. Then, 7 factors including soil, lithology, slope, vegetation index (NDVI), distance from the river, geomorphic units and land use as an independent and effective variable on dust were prepared. Then using logistic regression and random forest regression models, the weight of each of the factors in R software was calculated, finally, the dust hazard sensitivity maps were prepared.

3- Results

Four different false color combinations were used to detect dust areas in all images. Each of these methods revealed a dusty mass. The second method (R: D, G: BTD_{2931}, B: NDDI) is the best method for detecting dust and determining dust collection zones. Finally, by employing these techniques and creating false color combinations and applying them to existing images and their visual interpretation,

[i] **Corresponding Author: m.boroughani@hsu.ac.ir**

65 dust collection points were identified throughout the area. Based on the results, slope had the greatest impact on creating dust source zones. Geomorphology, lithology, vegetation index, soil and land use are the next influential categories of dust source area. Negative coefficients indicate the inverse relationship of the dependent variable (dust source) with the independent variables; in fact, negative coefficients do not indicate a lack of correlation, but mean a weaker correlation than other factors. The distance factor from the river has the least impact on the formation of these dust source. The OOB results show a predicted error rate of about 35.16%. In other words, the accuracy of the model can be considered 64.84%. The results of logistic regression algorithm showed that 24.5% of Khorasan Razavi province was in low sensitivity area, 28.2% in medium sensitivity area, 29.4% in high sensitivity area and 17.9% in high sensitivity area. It is very sensitive. The results show that the highest dust source points (57.9%) are located in high sensitivity area. The results indicate that about 84.2% of the dust collection points are located in the area of high and high sensitivity. Surveying the area of dust susceptibility map using random forest algorithm showed that the area with the highest sensitivity had the lowest area (16.7) and the areas of high, medium and low were 23.5, 31 and 8.8, respectively. 28 estimated. About 63.2% of source are located in the area of high sensitivity and there is no source for the area of low sensitivity.

4- Discussion & Conclusions

The results of both models show that slope and land use with the significant coefficients of 0.852 and 0.673 had the most effect on the creation of dust source area. The results also indicate that most dust removal sites (57.9%) are located in high sensitivity area. The results indicate that about 84.2% of the dust collection points are located in the area of high and high sensitivity. The evaluation of models using the ROC curve showed that in terms of success rate, logistic regression models and random forest had a precision of 781 and 907 respectively. Each of the two high correlation models between the sensitivity maps and the distribution of dust source areas can be observed. But in general, the random forest model has a higher accuracy than the other logistic regression models.

Key Words: Khorasan Razavi Province, Zoning, Random Forest, Dust source, Logistic Regression.