

Nat. Hazards Earth Syst. Sci. Discuss., referee comment RC1  
<https://doi.org/10.5194/nhess-2022-118-RC1>, 2022  
© Author(s) 2022. This work is distributed under  
the Creative Commons Attribution 4.0 License.

## **Comment on nhess-2022-118**

Anonymous Referee #1

---

Referee comment on "A Comparative Analysis of Machine Learning Algorithms for Snowfall Prediction Models in South Korea" by Moon-Soo Song et al., Nat. Hazards Earth Syst. Sci. Discuss., <https://doi.org/10.5194/nhess-2022-118-RC1>, 2022

---

This manuscript compared the skill of four machine learning algorithms, including multiple linear regression (MLR), support vector regression (SVR), random forest regressor (RFR), and eXtreme gradient boosting (XGB) for snowfall estimation in South Korea. Meteorological data (minimum temperature, maximum temperature, precipitation, and relative humidity) from 1991–2020 during the winter season (October to April) collected from the automated synoptic observing system, and geographic data (latitude, longitude, and altitude) were used as the input variables and the measured snow depth was used as the output variable for machine learning model training. The results indicate the RFR performs the best among the four machine learning algorithms with an  $R^2$  of 0.64.

The work is interesting, however, the main drawback of this work is that it is too basic and simple. A great deal of similar works have been carried out in previous studies, and some of them have been summarized by the authors (Line 55-97). In the introduction, the authors only mentioned such previous works, but did not point out the problem which remains to be solved in the current work (i.e., the motivation of this study). In other words, if the paper is only a simple imitation of previous studies, it is not innovative.

Other comments:

Line 41-45: add references.

Line 81: where is the reference of "Liang et al. (2015)"?

Section 2.1: only meteorological data were used in the study. Due to the limited spatial coverage of the stations, why the authors did not consider other large-scale data such as remote sensing data or model (reanalysis) based data?

Fig. 1: this figure lacks longitude and latitude information. Moreover, its quality can be improved, e.g., you can use the legend information to represent the stations but do not need to list all the station names.

Line 129: where is the reference of "Ainiyah et al., 2016"?

Line 130: where is the reference of "Mallick et al., 2021"?

Line 140: should be seven inputs and one output?

Fig. 2: isn't the average temperature excluded due to the high collinearity issue?

Line 153: it is better to add a section entitled "2.2 Machine Learning Methods" before "2.2 MLR". Moreover, there are numerous machine learning methods, why did you select the four methods?

Line 198: delete "Tianqi".

Line 215: MSE and RMSE play the same role in the evaluation. You can only preserve RMSE.

Line 255-256: add unit for MAE, MSE, and RMSE.

Table 4: add unit for MAE, MSE, and RMSE.

Fig. 5: add unit for snowfall.