Improvement of Spatial Interpolation Accuracy of Daily Maximum Air Temperature Using Stacking Ensemble Technique

Dongjin Cho (1), Cheolhee Yoo (1), Jungho Im (1), Yeonsu Lee (1)

[1] Ulsan National Institute of Science and Technology., 50 UNIST-gil, Beomseo-eup, Ulju-gun, Ulsan 44919, Korea

Email: djcho@unist.ac.kr; yoclhe@unist.ac.kr; ersgis@unist.ac.kr; leeysu0423@unist.ac.kr

Abstract: Air temperature is an important indicator of social problems damaging human society such as heat island and heatwave. Monitoring air temperature with high spatio-temporal resolutions is necessary to mitigate those problems. Air temperature is measured by weather stations with a high temporal resolution, but the number of weather stations is not enough to monitor the complex cities. So, numerical models and satellite monitoring are used to monitor the distribution of air temperature. However, numerical models have generally coarse spatio-temporal resolution due to the requirement of lots of computation, and satellite monitoring is usually vulnerable by cloud contamination. Therefore, it is hard to monitor seamless air temperature distribution through numerical models and satellite monitoring. One of the methods to monitor the seamlessly spatial distribution of air temperature is spatial interpolation with observation data provided by weather stations, but it does not reflect heterogeneity within complex cities. Since satellite-based geographical variables such as Digital Elevation Model (DEM), Slope, Aspect and impervious area help to reflect the complex cities, spatial interpolation accuracy is expected to improve when spatial interpolated air temperature and they are merged. Therefore, the objective of this study is to develop and evaluate the performance of machine learning models for improvement of spatial interpolation of maximum air temperature in Seoul, South Korea using a fusion of spatial interpolated air temperature and satellite-based geographical variables. Estimated temperature and height using kriging and 9 satellite-based auxiliary variables including DEM, slope, aspect, latitude, longitude, global man-made impervious surface (GIMS) and human built-up and settlement extent (HBASE) were used as the input variables and three machine learning methods including random forest (RF), support vector machine (SVR) and artificial neural network (ANN) were applied in this study. Finally, outputs of those three machine learning models with prior input variables were input for stacking models with SVR. Leave-one-station-out cross-validation (LOSOCV) was conducted to evaluate the models. The Kriging model had an R^2 of 0.94 and an RMSE of 0.87 °C (Fig. 1a), whereas three single machine learnings resulted in R^2 ranging from 0.87 to 0.94, RMSEs from 0.81 °C to 1.26 °C (Fig. 1b-d). The Stacking model showed the best performance with R^2 of 0.96 and an RMSE of 0.66 °C (Fig. 1e). Besides, the Stacking model

showed produced better robustness with high accuracy than any single machine learning model, and also showed a better simulation of temperature spatial distribution in Seoul.

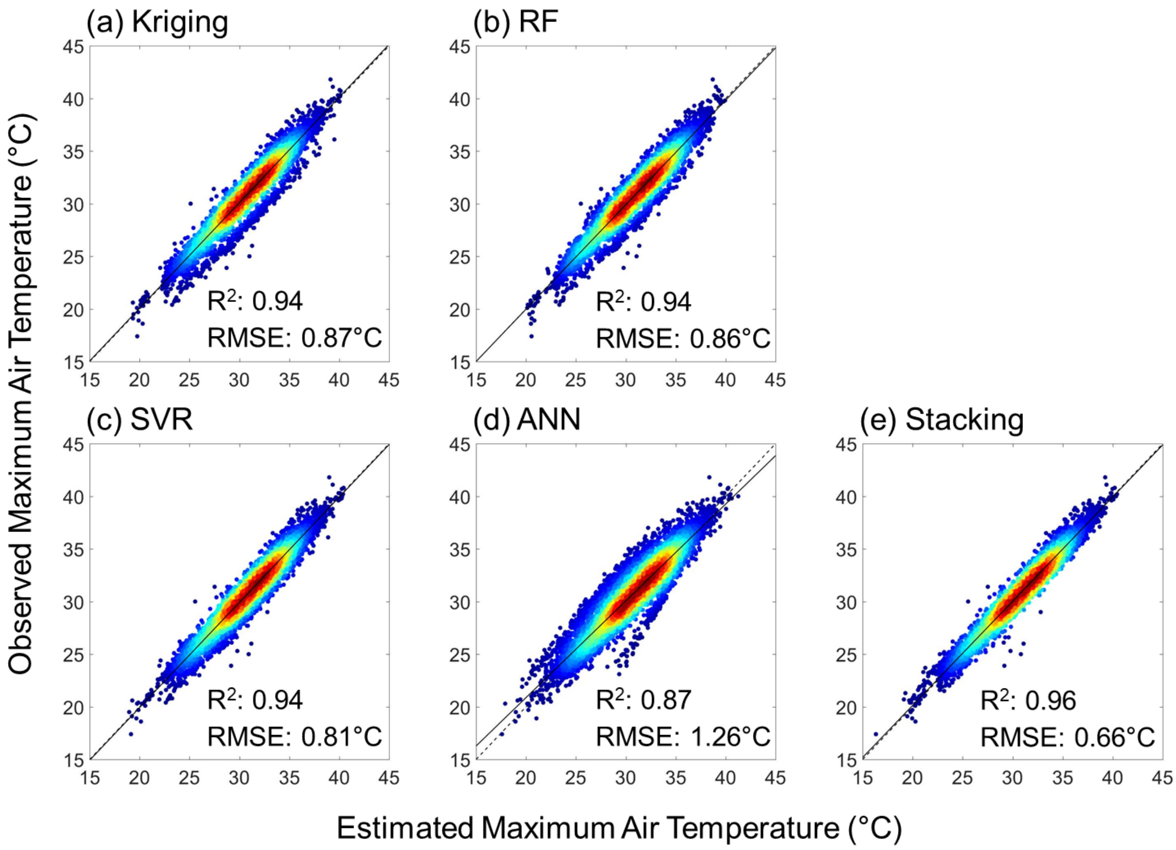

Fig. 1. Scatter plots between estimated and observed maximum air temperatures from LOSOCV results based on (a) Kriging and machine learning models using (b) RF, (c) SVR, (d) ANN, and (e) Stacking

Keywords: Maximum air temperature, Kriging, Satellite-based auxiliary variables, Stacking Ensemble