ORIGINAL PAPER

High-resolution wind speed forecast system coupling numerical weather prediction and machine learning for agricultural studies — a case study from South Korea

Ju-Young Shin[1] · Byunghoon Min[1] · Kyu Rang Kim[1]

Abstract

Forecasting wind speed near the surface with high-spatial resolution is beneficial in agricultural management. There is a discrepancy between the wind speed information required for agricultural management and that produced by weather agencies. To improve crop yield and increase farmers' incomes, wind speed prediction systems must be developed that are customized for agricultural needs. The current study developed a high-resolution wind speed forecast system for agricultural purposes in South Korea. The system produces a wind speed forecast at 3 m aboveground with 100-m spatial resolution across South Korea. Logarithmic wind profile, power law, random forests, support vector regression, and extreme learning machine were tested as candidate methods for the downscaling wind speed data. The wind speed forecast system developed in this study provides good performance, particularly in inland areas. The machine learning–based methods give the better performance than traditional methods for downscaling wind speed data. Overall, the random forests are considered the best downscaling method in this study. Root mean square error and mean absolute error of wind speed prediction for 48 h using random forests are approximately 0.8 m/s and 0.5 m/s, respectively.

Keywords Wind speed at 3 m aboveground · Post-processing · Random forests · Machine learning · Downscaling

Introduction

Wind speed affects many agricultural activities and crop characteristics, including growth and development (Retuerto and Woodward 2004; Gardiner et al. 2016), and the high value of wind speed can cause fruit falls (McAneney et al. 1984; Gravina et al. 2011), lodging of cereals (Sterling et al. 2003; Feng et al. 2019), and physical damage to crops (Cleugh et al. 1998; Van Gardingen and Grace 1991; Retta et al. 2000). Providing farmers with future wind speed information will increase profits by enabling them to take mitigating actions against adverse impacts of high-speed winds and will improve crop cultivation efficiency (Gardiner et al. 2016). Thus, forecasting wind speed near the land surface is beneficial to agricultural management.

✉ Ju-Young Shin
jyshin83@korea.kr

1 High Impact Weather Research Department, National Institute of Meteorological Sciences, 33 Seohobuk-ro, Seogwipo-si, Jeju-do 63568, South Korea

Agricultural production predicted by agricultural simulation model (ASM) using weather forecasts from climate models and numerical weather prediction systems varies with the spatial resolution of weather forecast data (Mearns et al. 1999; Kim et al. 2015). For example, field scale is often preferred for spatial resolution of climate models and numerical weather prediction systems for agricultural use (Hansen and Indeje 2004; Takle et al. 2014; Shin et al. 2020b). For the simulation and forecast of agricultural production, weather prediction data with high spatial resolution should be developed.

To observe and predict wind speed directly affecting agricultural environments, the measuring height of wind speed should be similar to the height of crops. For example, wind speed data at 3 m aboveground are desired for agriculture (Cermak 1979). The standard protocol for measuring wind speed, recommended by WMO (2018), is at a height of wind speed of 10 m aboveground; therefore, climate models and numerical weather prediction system have been designed to simulate and predict wind speed at 10 m aboveground (WS10M). However, due to the difference between anemometer heights and crops, the direct use of wind speed

predictions by climate models and numerical weather prediction systems sometimes may be inappropriate in agricultural simulations and modeling for specific purposes such as scheduling pesticide spraying, pollen disposal, lodging prediction, and estimating canopy-top evapotranspiration. Therefore, a methodology is required to downscale wind speed from 10 m aboveground to a lower height aboveground. Wind profiling methods have been developed to predict wind speed at different heights aboveground using WS10M observations, including LOGarithmic wind profile (LOG) and POWer law (POW) methods (Monin and Obukhov 1954; Peterson and Hennessey 1978), and performances and limitations of these methods have been explored in studies (Lubitz 2009; Optis et al. 2016). Recently, machine learning (ML) algorithms have been tested for wind profiling methods (Mohandes et al. 2016; Bodini and Optis 2020; Vassallo et al. 2020); however, most of these studies focused on predicting WS10M; therefore, the utility of these methods for agricultural simulation and modeling is unclear. Moreover, despite its applicability to various fields, climate models and numerical weather prediction systems predict WS10M with a coarse spatial resolution, which causes a discrepancy between the information required by agricultural management and that produced by the weather service. To improve crop yield and increase the income of farmers, wind speed prediction systems that are customized for specific agricultural purposes, need to be developed.

The current study aimed to develop a wind speed forecast system for agricultural purposes where the wind speed data has a 100-m spatial resolution at 3 m aboveground.

Korea Meteorological Administration Post Processing (KMAPP) was employed for obtaining WS10M prediction data having 100 m × 100 m horizontal resolution. For agricultural purposes, wind speeds at 3 m aboveground (WS3M) were measured across South Korea. Furthermore, the WS10M prediction from the KMAPP was downscaled to WS3M using ML algorithms and traditional methods, such as LOG and POW that are constructed by relationship between WS10M predictions and WS3M observations; these downscaled products were compared. Geographical and geological information, as well as other meteorological variables, was tested before using them as input features for the ML methods. The backward elimination method was used to select relevant input features for each ML algorithm for downscaling WS3M.

High-resolution wind speed forecast system for agriculture

In this study, the wind speed prediction system was developed using weather forecast data from numerical weather prediction system and post-processing this data, followed by downscaling wind speed data from 10 to 3 m aboveground. The KMAPP data were used as the weather forecast data. For downscaling methods, several methods including traditional and ML methods are employed. A schematic diagram of the high-resolution wind speed forecast system for agriculture is presented in Fig. 1.
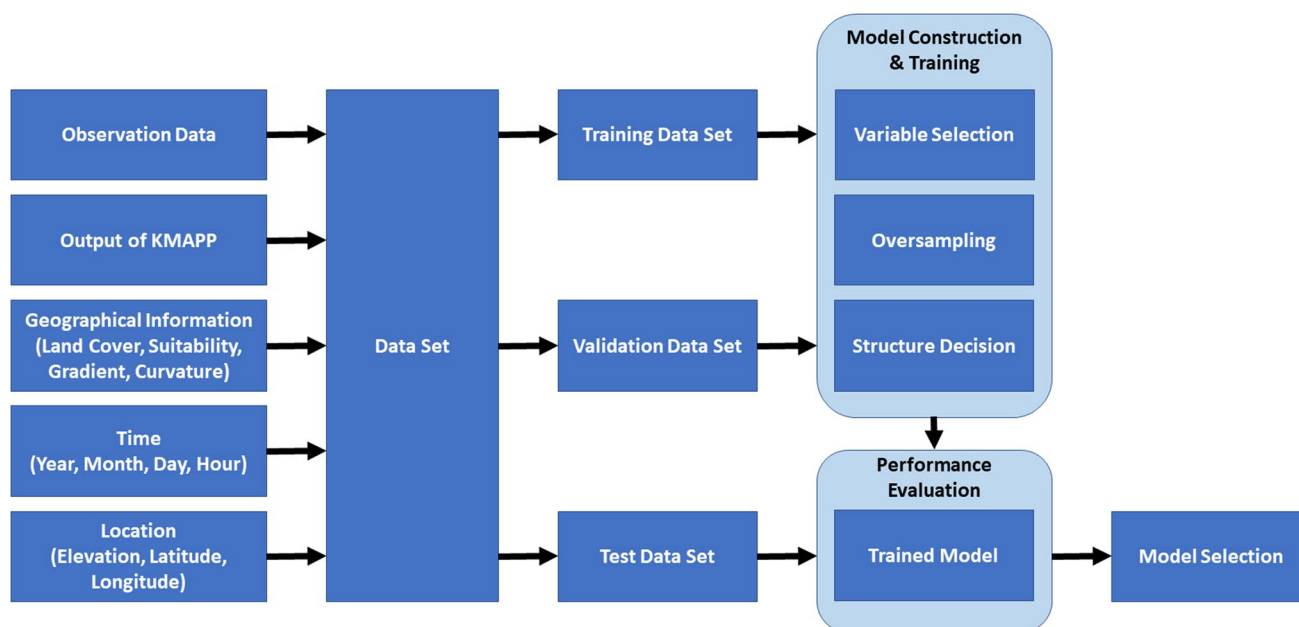

Fig. 1 Schematic diagram of the high-resolution wind speed forecast system for agriculture in the current study

Weather forecast data from numerical weather prediction system

The Local Data Assimilation and Prediction System (LDAPS) was configured for weather prediction over the Korean peninsula and surrounding waters (KMA 2011). The output of LDAPS has a spatial resolution of 1.5 km (H602 × V781) and consists of 70 vertical levels up to 40 km (Cho et al. 2020; Kim et al. 2020). Since the spatial resolution is too coarse to be used in other fields, the National Institute of Meteorological Sciences developed KMAPP for producing high-resolution weather forecast data. This model was developed based on the United Kingdom Post Processing method which was proposed by UK met Office. KMAPP produces high-resolution (100 m × 100 m) weather forecast data using the outputs of LDAPS by spatial downscaling, which is essentially the spatial interpolation of coarse outputs (Yun et al. 2021). Orographic adjustment is adopted for locations with complex terrain. For downscaling wind speed data, additional adjustments are made, based on roughness length and height (Howard and Clark 2007). The reduction rate of air temperature based on height is accounted for while downscaling air temperature (Sheridan et al. 2010, 2018).

KMAPP data consist of surface and level data with an hourly temporal resolution and a lead time of 48 h. Outputs of KMAPP are operationally produced at 00, 06, 12, and 18 UTC within a day. The u (east direction) and v (north direction) components of WS10M, air temperature, relative humidity, downward shortwave flux, visibility, and mean sea level are produced for surface data in the KMAPP. The level data consists of 29 levels within approximately 3 km that is considered the mixing height. For the level data, u and v components, air temperature, and air pressure are produced. In the current study, six variables, including u and v components of WS10M, air temperature, relative humidity, mean sea level pressure, and downward shortwave flux in the surface data of KMAPP, were used as input features for downscaling methods. The information of KMAPP data is summarized in Table S1 in Supplementary Information (SI).

Downscaling methods

The WS3M for agriculture was predicted by downscaling the WS10M from climate models and numerical weather prediction system. ML algorithms were employed as the primary algorithm for downscaling. However, for comparison, LOG and POW methods were also used as the traditional downscaling method. Random Forests (RF), support vector regression (SVR), and extreme learning machine (ELM) are selected for the candidate of ML algorithm.

Logarithmic wind profile method

Based on the LOG method, the mean wind speed at a specific height aboveground (z m aboveground) can be calculated by Eq. (1) (Blackadar and Tennekes 1968; Tennekes 1973; Kent et al. 2018):

$$\overline{U}(z) = \frac{u_*}{\kappa} \ln\left(\frac{z - z_d}{z_0}\right) \tag{1}$$

where $\overline{U}(z)$, u_*, κ, z_d, and z_0 are mean wind speed at z m aboveground, roughness velocity, von Karman's constant, zero-plane displacement, and roughness length, respectively. The value of κ is 0.4, obtained from a wind tunnel experiment (Garratt 1994). WS3M can be predicted by Eq. (2), based on the prediction of WS10M.

$$\overline{U}(3) = \overline{U}(10) \frac{\ln\left(\frac{3 - z_d}{z_0}\right)}{\ln\left(\frac{10 - z_d}{z_0}\right)} \tag{2}$$

Power law method

The POW method could describe the relationship between wind speeds at two different heights aboveground. This is defined in Eq. (3) (Emeis 2014; Kent et al. 2018):

$$\overline{U}(z_1) = \overline{U}(z_2)\left(\frac{z_1 - z_d}{z_2 - z_d}\right)^{\alpha} \tag{3}$$

where z_1, z_2, and α are wind speed at first height aboveground, wind speed at second height aboveground, and wind shear exponent, respectively. The wind shear exponent ranges from 0 to 1. In this study, Eq. (4) was employed for predicting WS3M based on WS10M using the POW method:

$$\overline{U}(3) = \overline{U}(10)\left(\frac{3 - z_d}{10 - z_d}\right)^{\alpha} \tag{4}$$

Random forests

RF has been widely used as an alternative for classification and regression problems (Shin et al. 2020a, 2021; Hansen and Indeje 2004; Pal 2005; Smith 2010; Cho et al. 2020; Watt-Meyer et al. 2021). Breiman (2001) proposed the RF algorithm, which uses many decision trees constructed by the bagging method; thus, the RF can be considered as an ensemble of decision trees. Each decision tree is grown using a selected subset that is randomly resampled with randomly selected features from the original datasets. The RF consists of randomness and ensemble learning. The

randomness comes from random resampling of the entire dataset and the selection of features with which every classification and regression tree is built. The ensemble learning method in RF means that all individual decision trees in a collection of decision trees (ensemble) contribute to a final prediction. The classification and regression tree, without pruning, is used to construct a single decision tree. The final predicted label is the most frequent among the predicted labels of all individual trees. The "ranger" library in R was used to construct the RF model (Wright and Ziegler 2017).

Extreme learning machine

Extreme learning machine is a feed-forward network consisting of a single network with randomly generated weights and bias between input and hidden layers (Huang et al. 2006). The weights and biases of conventional neural network are iteratively optimized while they are tuned in a single iteration because of the randomized weights and biases. The ELM can be expressed in Eq. (5):

$$Y = H\beta \tag{5}$$

where Y, H and β are labels, the output vector of the hidden layer, and weight matrix between hidden layer to output layer, respectively. H is nonlinear feature mapping. H is defined by Eq. (6):

$$H = f_a(XW + B) \tag{6}$$

where $f_a(\bullet)$, X, W, and B are activation function, input feature, weight matrix between input layer to hidden layer, and bias, respectively. The sigmoid function ($f_a(x) = \frac{1}{1+\exp(-x)}$) was employed as the activation function in the ELM in this study.

Support vector regression

SVR is applied as a regressor to many regression problems in various fields (Yang et al. 2006; Yao et al. 2017; Abbas et al. 2020; Liu et al. 2021; Zhang et al. 2021). The SVR algorithm is a modified version of support vector machine (SVM). The SVR was developed for solving classification problems based on mathematics, unlike other ML algorithms, such as ELM and RF. SVM was developed for building classifiers that maximize the margin, that is, the distance between any two groups. The distance between two groups is determined by the distance between support vectors, that is, the nearest vector to another group. Unlike the SVM, the support vectors indicate the two most distant vectors in the datasets for the SVR. Additionally, the regression line that minimizes the margin represents the regressor in the SVR. To figure out the hyperplane minimizing the margin, the

optimization problem (Eqs. (7)–(9)) with its constraints has to be solved:

$$\min \tau(W, \xi, \rho) = \frac{1}{2}\|W\|^2 - \nu\rho + \frac{1}{2}\sum_{i=1}^{n}\xi_i \tag{7}$$

$$\text{subject to } y_i\left(\Theta(X_i)W + b\right) \leq \rho - \xi_i, i = 1, \ldots, n \tag{8}$$

$$\xi_i \geq 0, \rho \geq 0 \tag{9}$$

where, ν ξ_i, and $\Theta(\bullet)$ are regularization constant that ranges from 0 to 1, slack variable for ith data point, and kernel function. The radial basis function is used for kernel function in the SVR used in this study, and the "e1071" library in the R was employed for conducting SVR (Meyer et al. 2021).

Data

Meteorological data

In this study, meteorological data from the Automated Agricultural Observing System (AAOS) were used. AAOS stations are located in agricultural areas, such as farms and orchards, and measure 14 meteorological variables related to the agricultural environment, e.g., air temperature, wind speed, and relative humidity. In this study, hourly observed WS3M data from AAOS were employed for the target data. Hourly wind speed depicts 10-min mean wind speed data observed at 00 min in each hour. The wind speed data can be downloaded from Nongeupnalssi 365 (weather.rda. go.kr). The recording period was from July 2019 to June 2020 (12 months). To obtain reliable data, the meteorological observation data were inspected by a quality check (QC) procedure based on KMA guideline of QC for automatic weather station, using the relocation of the measuring instrument, location of the measuring instrument, proportion of missing data, and proportion of zero values. The stations where the displacement of instrument is higher than 150 m were extracted; in addition, when the instruments were installed in improper location such as near road and high building, the data in these stations were not used. If proportions of missing data and null wind speed were higher than 10%, measures in these stations were not employed. Based on the QC results, 104 stations were of sufficient quality to be used in the modeling. The selected 104 stations are presented in Fig. 2.

Terrain-based spatial data

Surface wind speed varies with surface characteristics such as topography and land cover (Wu et al. 2016). Use of these surface characteristics for input features of machine learning algorithm led to improvements in near-surface wind speed

Fig. 2 Location of the used weather stations. Note that the red and green colored stations are selected for performance evaluation at individual stations while all stations are employed for evaluating overall performances of the developed model
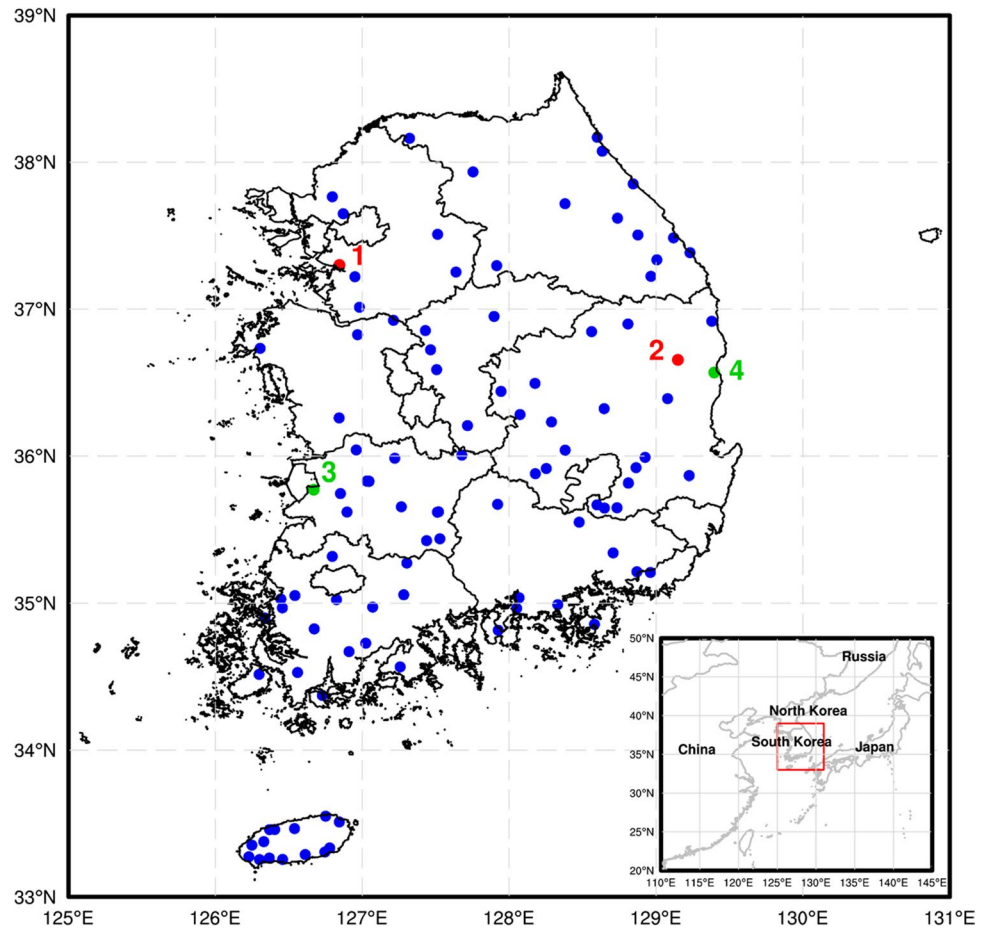
prediction (Jung and Schindler 2020). In this study, land cover and agricultural suitability maps were used as input features for the ML methods as they are related to the agricultural environment. Land cover maps represent the land cover type at the specific location, based on predefined categories, and have three levels: top, medium, and low, based on the hierarchy of land cover types. These types in the map provide the most detailed information at the low level. Medium-level map, with a scale of 1:25,000 that horizontal resolution is approximately less than 10 m, was used in this study as it has sufficient detail to represent land cover characteristics in agricultural areas. This data was obtained from the environmental geographic information service (egis.me.go.kr). The agricultural suitability map, with a scale of 1:25,000 scale, includes information related to cultivation, e.g., soil properties, yield potential, and degree of yield constraint, to help decision-making for increasing agricultural production and yield. The agricultural suitability map used in this study can be downloaded from the Korean soil information system (soil.rda.go.kr). The land cover and agricultural suitability maps are transformed to grid data that matches the KMAPP grid, using the nearest neighbor resampling method.

Terrain is critical in determining wind speed, because air flow is largely associated with terrain, such as mountains, valleys, and flat planes (Cao and Tamura 2006; Schmidli and Rotunno 2012). Elevation data at a specific location may have a limited capacity to represent topological characteristics; hence, the slope and curvature were used as input features representing topological characteristics in this study. The slope and curvature were calculated from the digital elevation map used in KMAPP. These calculated slope and curvature were then employed as input features of the ML algorithm. The slope indicates a first-order differential elevation, with respect to location, and has been widely used in geomorphometrics (Evans 1972; Ghandehari et al. 2019). The curvature is defined as a second-order differential elevation, with respect to location (Ghandehari et al. 2019).

Application

Traditional methods estimate WS3M using WS10M predictions from KMAPP. To compute WS3M using traditional methods, some parameters, such as z_0, α, and z_d should be defined. In this study, z_0 and α were given based on the land

cover type of the location (Aghbalou et al. 2018; Chavan et al. 2017). The values of z_0 and α used in this study are listed in Tables S2 and S3 in SI. The value of z_d was set to 2/3 of the canopy height, assuming that the canopy height is 1 m (De Bruin and Verhoef 1997).

For WS3M prediction using ML algorithms, 801,769 data points were collected. Sixty (481,051), 20 (160,359), and 20% (160,359) of the all data points used were randomly resampled for training, validation, and test datasets, respectively, without replacement. Thus, the number of data points for the validation and test datasets was 160,359 and 160,359, respectively. The observed WS3M is unbalanced data in that the values of data are not uniformly distributed. The use of unbalance data leads to overfitting of model for the most frequent values. Thus, this unbalance data can worsen performances for less frequent values. To avoid overfitting from the unbalanced data, oversampling was performed for the training dataset. In oversampling, the number of data point in each bin that has 1 m/s interval becomes the same by resampling data points in the training dataset. Hence, the number of data points for the training data set was 16,113,150. To optimize input feature selection, backward elimination was carried out (Xu and Zhang 2001). Forty-eight features were considered as the initial input feature in the backward elimination method, including six meteorological variables predicted by the KMAPP, 30 features from the agricultural suitability map, one feature from the land cover map, two features from the slope, two features from the curvatures, four features indicating time, and three features indicating geological location. The input features of all ML methods were selected using the backward elimination method, based on root mean square error (RMSE) for the validation dataset. According to the backward elimination method, the ML methods employed a total of 18 input features: six from KMAPP (u and v components of WS3M, air temperature, relative humidity, mean sea level pressure, and downward shortwave flux), three from the agricultural suitability maps (drainage level, land form, and soil suborder), one from the land cover map, two from the slopes (of north and east direction), two from the curvatures (of north and east direction), one feature from time (month), and three features from geological information (elevation, latitude, and longitude).

The hyperparameters of the ML methods were optimized based on the RMSE value of the trained models for the validation dataset. To determine the number of trees in RF, 300–700 trees were tested. When there were 500 trees, the RMSE of the trained RF was the lowest; hence, 500 trees were used in the RF. For the SVR, 0.01–1 were tested for gamma, and 1–10 were tested for the cost using validation dataset. Based on the cross validation, the smallest RMSE was obtained when the gamma and cost were equal to 0.05 and 1, respectively. The hyperparameters of ELM used in this study were the number of nodes and the tuning

parameter. For ELM, 1000–4000 nodes were tested and 0.01–10 were tested for the tuning parameter, and 2000 and 0.1 were used as the number of nodes and tuning parameter, respectively, based on the cross validation.

Results

Overall performance of models

All evaluation measures were calculated using the test dataset. The proposed system predicts WS3M at 4.6 million points across South Korea. WS3M prediction at 104 points that match the AAOS weather stations were used for a performance evaluation. Correlation, RMSE, mean absolute error (MAE), and mean bias error (MBE) of WS3M predictions from all the models are presented in Fig. 3. Based on correlation, RMSE, and MAE, the ML-based methods performed better than LOG and POW methods. Among the ML algorithms, RF was the best method, while ELM performed the worst. The RMSE value of the prediction for RF was < 0.8 m/s, while that for ELM was > 1.1 m/s. The RMSE values for traditional wind profiling methods, LOG and POW, were > 1.4 m/s. Similarly, RF had the best performance for all lead time with MBE < 0.2 m/s. The MBE of ELM and LOG was > 0.6 m/s. Among the methods, POW had the largest MBE value > 0.9 m/s with a long lead time. Prediction performances of ML-based methods gradually decreased as lead time increased, whereas the prediction performances of traditional methods drastically reduced after a 2-h lead time. The mean, median, and standard deviation of observed WS3M for all employed stations are presented in Figure S1. The mean value of mean and median wind speed is 1.27 m/s and 0.98 m/s, respectively. Thus, the approximate relative absolute error ($= \frac{0.5}{1.27} \times 100$) based on mean value of MAE (0.5 m/s) for WS3M prediction by RF is 39.4%.

Figure 4 represents the overall performances of the tested models for all stations evaluated using density plots comparing WS3M and the KMAPP prediction for WS10M. The predictions of KMAPP are higher than the observed wind speed due to the difference in measuring height resulting in a correlation value of 0.557. Though the LOG and POW provided lower wind speed than the KMAPP, large positive biases remain in their predictions. The correlations for POW and LOG were 0.536 and 0.460, respectively. The predictions of RF and SVR were found to have good agreement with the observations, as their scatter points are located around the diagonal line. Unlike RF and SVR, the ELM led to an overestimation. The RF method showed the highest correlation with a value of 0.827. However, the RF algorithm overestimated wind speed lower than 1.5 m/s and

Fig. 3 Performance evaluation of five models for all stations using the test data set based on four evaluation measures: **a** correlation, **b** RMSE, **c** MAE, and **d** MBE
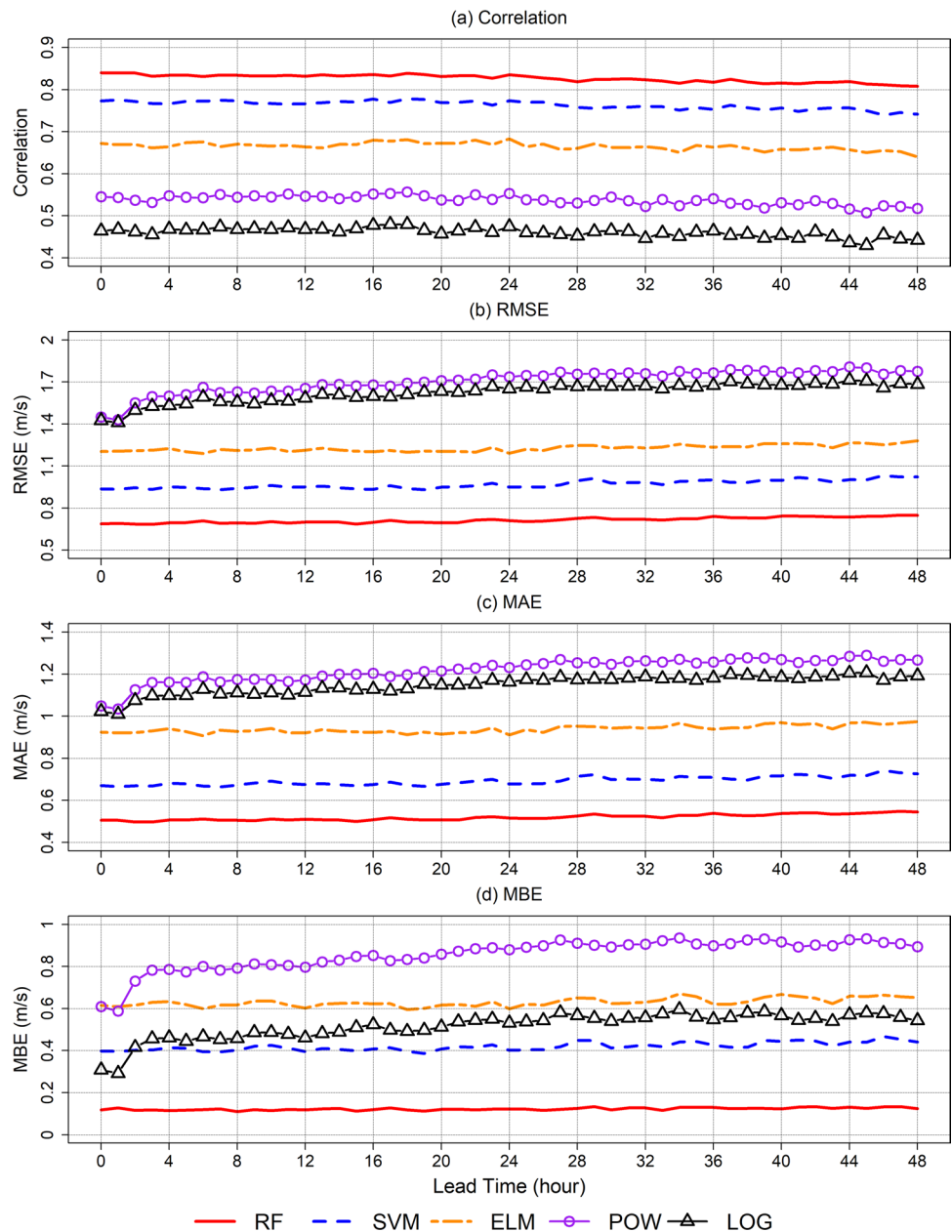
underestimated above 2 m/s. The SVR had the second-highest correlation with a value of 0.763. Overall, the SVR overestimates wind speed. Based on the evaluation measures, RF was determined as the best method; hence, the prediction performance of RF for various lead times was described using the density plot. These density plots for predictions at nowcasting (0 h), 8, 16, 24, 32, and 40 h are presented in Fig. 5. The correlations decrease as lead time increases, and the distributions of scatter points for different lead times were found to be similar. The highest correlation (0.84) was found for predictions at current time.

Spatial distribution of evaluation measures

To further investigate the prediction performance of the RF algorithm, correlation, RMSE, MAE, and MBE for each lead time and station are presented in Fig. 6. The medians of correlations were approximately 0.8, which decrease as the lead time increases. Spatial variation of evaluation measures (depicted as ranges of boxes) ranged from 0.7 to 0.85 and was consistent for all lead times. The results indicate that the spatial variation of prediction performances for all models is consistent for all lead times. Overall tendencies of RMSE, MAE, and MBE are similar to the correlation. For MAE, the
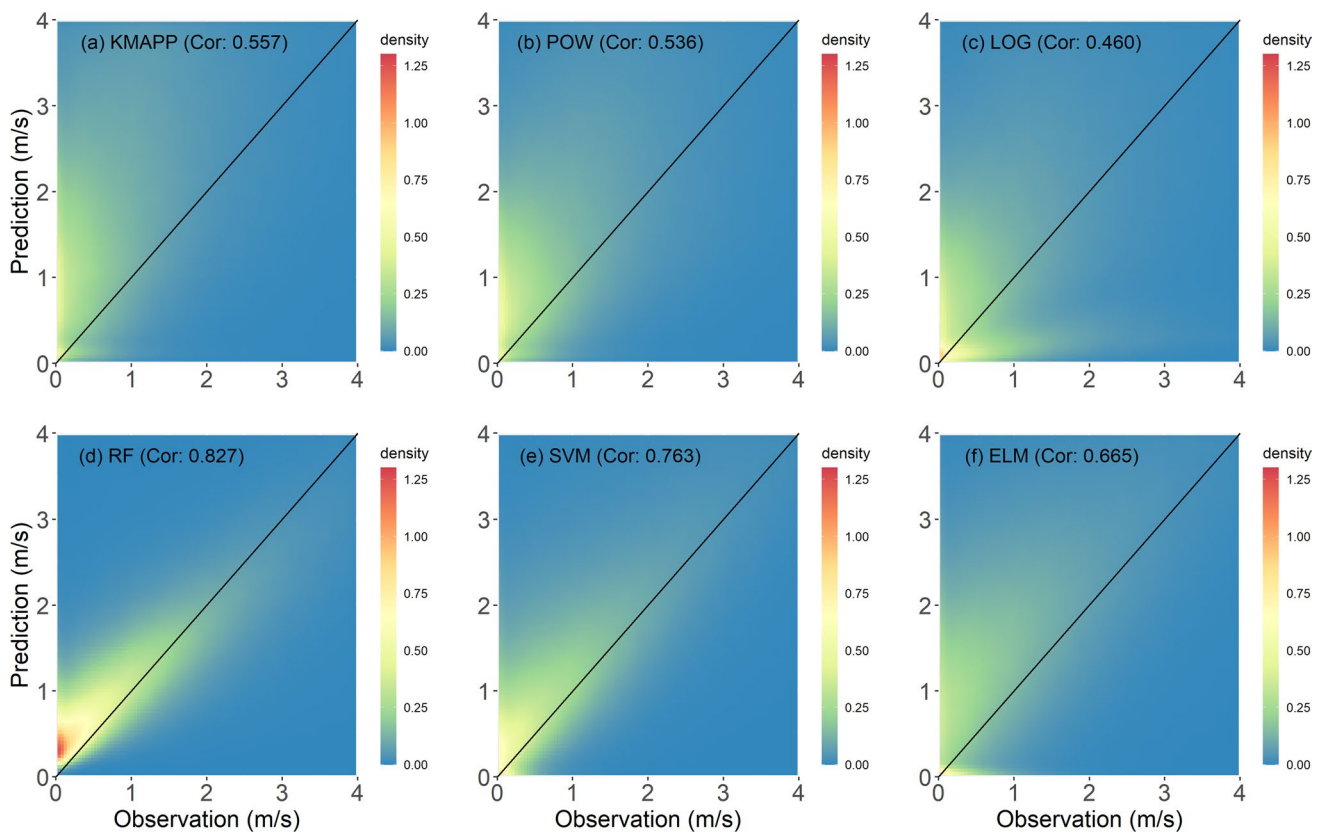
Fig. 4 Density plots of five models as well as KMAPP for all stations using the test data set: **a** KMAPP, **b** POW, **c** LOG, **d** RF, **e** SVR, and **f** ELM

medians are approximately 0.5 m/s, and its spatial variation ranged from 0.3 to 0.7 m/s. Medians of RMSE were between 0.6 and 0.7 m/s, and the its spatial variation ranged from 0.5 to 0.9 m/s. For MBE, the medians were approximately 0.1 m/s, and its spatial variation was between 0 and 0.2 m/s.

The spatial distribution of RMSE from POW, LOG, RF, SVR, and ELM is presented in Fig. 7, to investigate the spatial characteristics of prediction performances for all models. The spatial distribution of RMSE for KMAPP is also illustrated for comparison. For all models, RMSEs for stations in coastal region are larger than in inland regions. The spatial distribution of RMSEs for KMAPP was also similar. The RMSEs for stations in inland regions are > 1.8 m/s, while the RMSEs for stations in coastal regions are > 3.6 m/s. Traditional methods, such as POW and LOG, lead to smaller RMSE than KMAPP, but RMSEs for some coastal stations are > 3.6 m/s, even though the wind speed is downscaled at 3 m aboveground. The ML-based models provide smaller RMSEs than the traditional methods; the RF gives the smallest RMSEs of all the models (< 1.8 m/s for all stations). To investigate the variation of performance depending on lead time, the RMSEs for RF in six lead-times are presented in Fig. 8. For six lead times, the RMSEs for stations in coastal regions were larger than in inland regions. RMSEs increase

as the lead time becomes longer. The increments of RMSEs for stations in coastal region are larger than in inland regions.

Discussion

In this study, traditional and ML methods were employed for downscaling wind speed data and predicting WS3M. Overall, the ML methods lead to a better prediction performance than traditional methods. The poor performances of traditional methods for downscaling may result from the use of inaccurate values for z_0, α, and z_d. The values of these parameters can be accurately obtained by considering complex conditions, such as type, density, and height of vegetation. Some studies have reported that these parameters are associated with the type and density of vegetation (Shaw and Pereira 1982); therefore, information on these parameters is needed to successfully implement traditional methods for downscaling wind speed data in agricultural areas. The crop growth stage depends on various conditions, such as weather, vegetation type, soil properties, farmers, and days after sowing. The crop growth stage can represent the height and density of vegetation; thus, these parameters are strongly linked to the crop growth stage. However, collecting all the
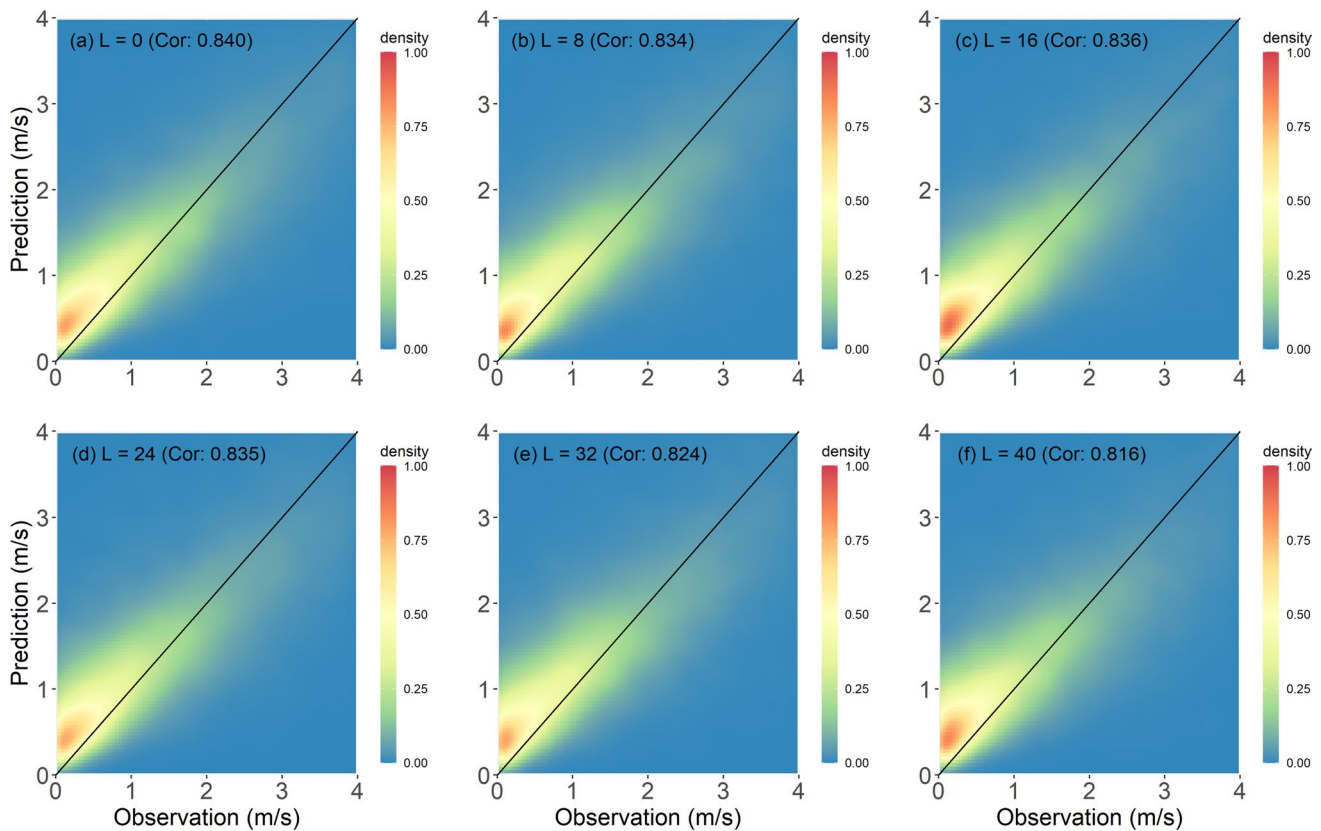
Fig. 5 Density plots of prediction from the RF model for different lead times using the test data set: **a** lead time (*L*)=0 h, **b** *L*=8 h, **c** *L*=16 h, **d** *L*=24 h, **e** *L*=32 h, and **f** *L*=40 h

information related to changes in the crop growth stage for farm areas nationwide is virtually impossible. Thus, in traditional methods, the inaccurate estimations of these parameters are an inevitable limitation. Overcoming some of the limitations of traditional methods, the ML methods provide a better prediction performance for downscaling wind speed in agricultural areas. ML algorithms can incorporate detailed information of surface, meteorology and meteorological prediction, and time. Although these parameters cannot directly represent vegetation properties in the area of interest, they can indirectly represent vegetation, and thus, improving the performance of the ML method in predicting WS3M.

Among the tested ML methods, RF performed the best at downscaling the wind speed data. The ELM algorithm performed the worst, but its performance was better than traditional methods. Agricultural suitability map and land cover map (categorical variables) were used as input features, e.g., mountain, flat plane, and slope near mountain. The ELM method may have poor performance because of the activation functions, e.g., sigmoid and hyperbolic tangent functions, which are normally used for continuous numeric variables. The categorical variable such as the farm for land cover cannot be directly used in the ELM, and this variable has to be digitized like one or two for use in the

ELM. The ELM inherently accounts for the magnitude of digits for input feature as strength of signal due to the activation function. However, because the magnitude of digits for the digitized categorical variable is meaningless, the ELM may not fully use information from the categorical variable.

Performance of wind speed prediction in inland areas is better than that in coastal areas for all tested downscaling methods. The performance difference between two areas results from wind speed prediction skill of KMAPP. The KMAPP provides better prediction skill for WS10M in inland areas than that in coastal regions (Yun et al. 2021). Because the WS3M prediction comes from the WS10M prediction by KMAPP, the prediction performance of WS3M is strongly associated to WS10M prediction. The result supports that the difference comes from the difference between prediction skills in inland and coastal areas. Subsequently, the proposed WS3M prediction system provides better performance in inland region as compared to the performance in coastal areas.

For some agricultural purposes, wind speed prediction at lower height than 10 m is required. WS3M data is valuable information in modeling canopy-top evapotranspiration (Allen et al. 1998). Additionally, when unmanned aerial vehicles (UAVs) are used to spraying pesticide, the flying

Fig. 6 Boxplots of the employed evaluation measures for all lead time in all stations: **a** correlation, **b** RMSE, **c** MAE, and **d** MBE
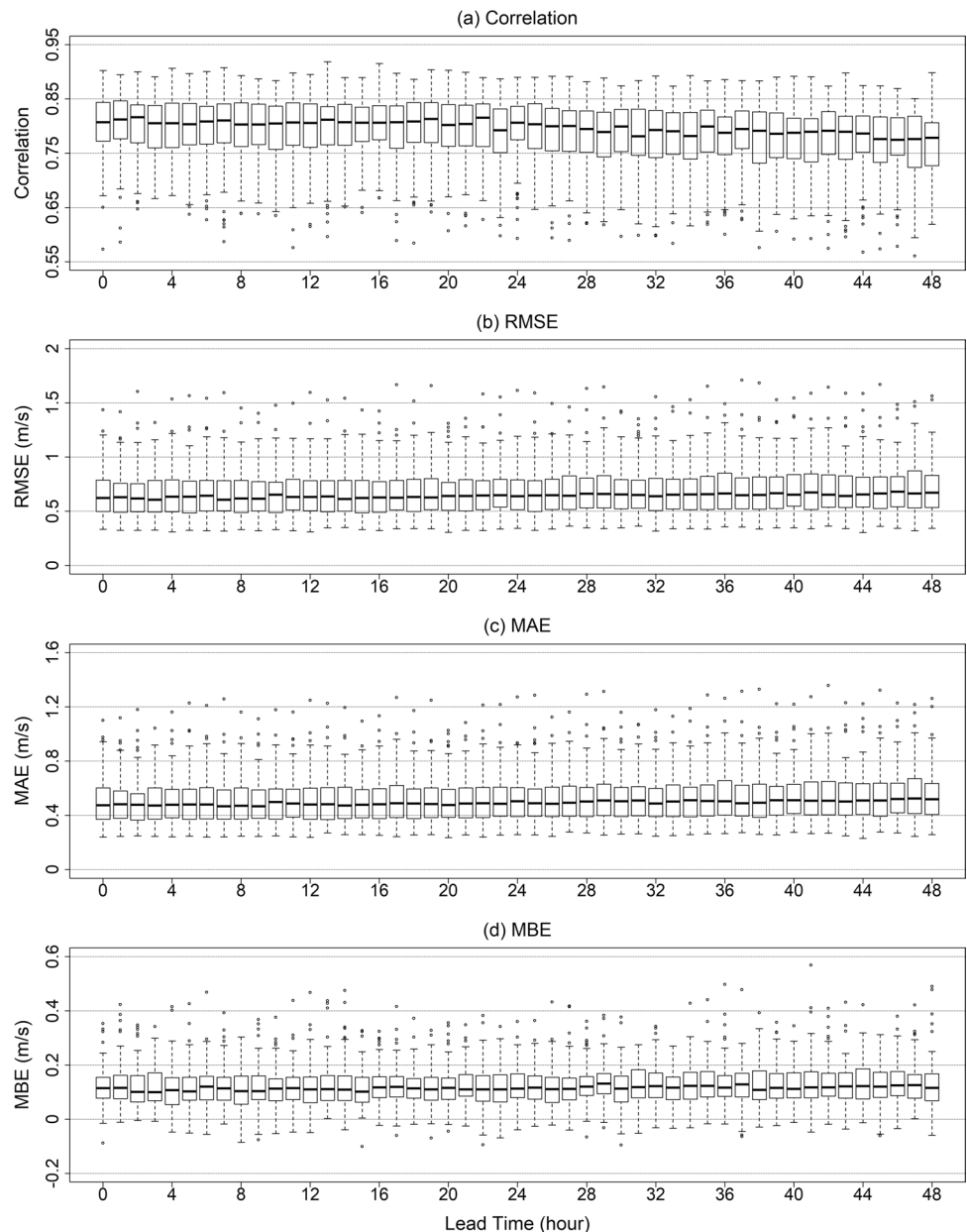
height should be lower than 4 m in South Korea due to the government regulation on avoiding long drift of pesticide (NAAS 2018). Though there is no standard flying height for UAVs, the flying height is less than 5 m aboveground (Martin et al. 2019). Thus, WS3M can be used to make decision to spray pesticides onto farm using the UAV in South Korea. For modeling serial crop lodging, wind speed at 2 m aboveground often has been adopted in some studies (Wen et al. 2019). In addition, the information of wind speed at lower height than 10 m can be used to model pollen disposal, which can be used in yield modeling (Tackenberg 2003). The height of wind speed data should be selected based on type of plants. Hence, there is no consensus for use of WS3M in agricultural modeling and simulation. For

instances, the American Society of Agricultural and Biological Engineers recommended wind speed data at from 2 to 3 m (ASBAE 2006), the American Association of State Climatologists recommended WS3M (AASC 1985), and WMO recommended 2 m for height of anemometer (Gommes et al. 2010). In downscaling wind speed at low height that is lower than 3 m, use of WS3M may lead to more accurate data than use of WS10M. However, it needs to explore what is an appropriate height for wind speed measure in agriculture.

The downscaling methods proposed in this study have certain limitations. First, the downscaling methods provide poor performances for downscaling WS3M near the coast. The RMSE values of stations in inland areas are < 1 m/s, and it varies depending on whether lead time is small. However,
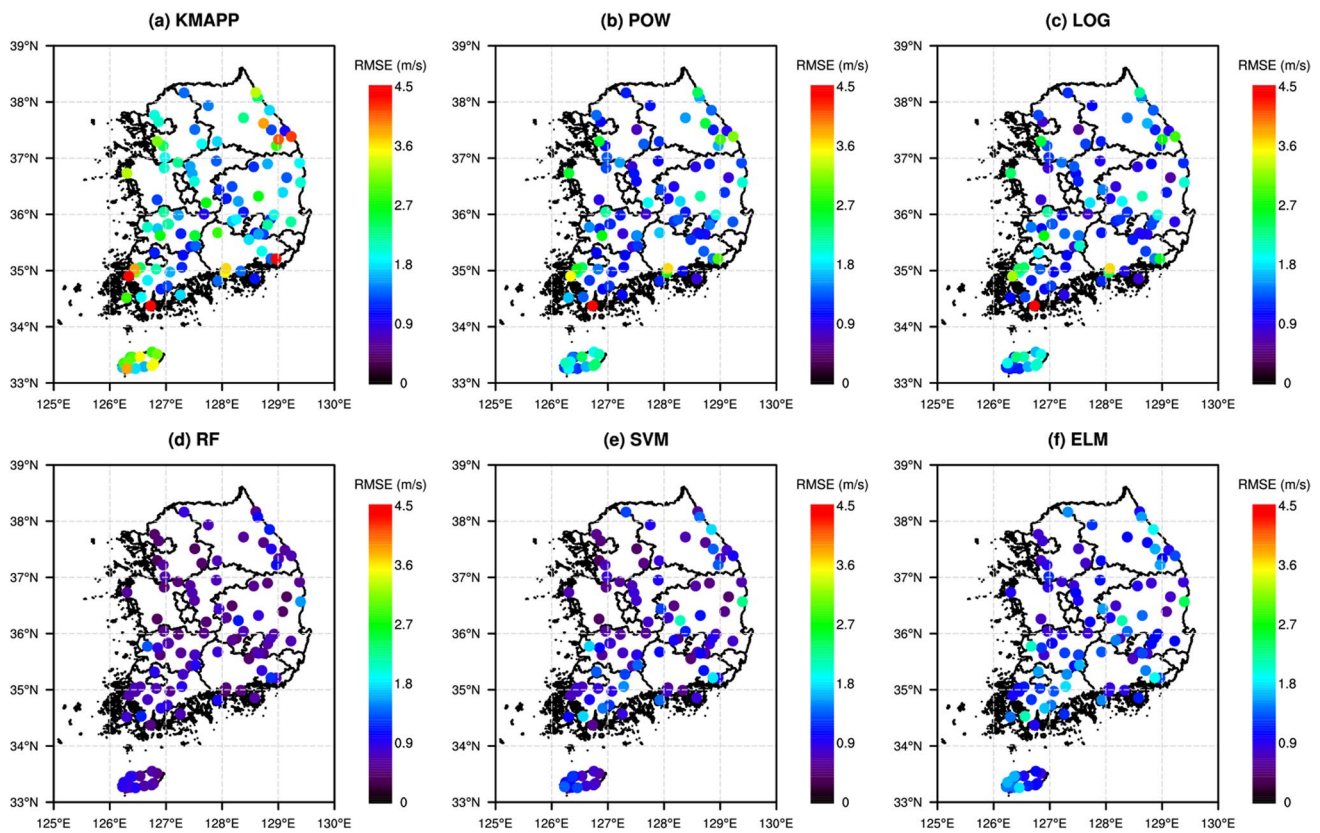
Fig. 7 Spatial distributions of RMSE for six models in the used stations: **a** KMAPP, **b** POW, **c** LOG, **d** RF, **e** SVR, and **f** ELM

the RMSEs of stations near the coast are larger than in inland areas, and the variation of RMSE depending on lead time is large. Hence, further research is required to improve the performance of the downscaling method for wind speed near the coast. Second, the proposed downscaling system provides an inaccurate WS3M prediction for wind speed greater than 4.5 m/s. The RF methods led to an underestimation of high value of wind speed. Hence, the developed models have a limited capacity for warning the risk or damage from high winds. High values of wind speed near surface are strongly associated to the turbulent kinematic energy under mixing height (Seibert et al. 2000; Brasseur 2001). In this study, because near-surface wind speed predictions from numerical weather prediction model were used for predicting WS3M, turbulent kinematic energy was not considered. This limitation may lead to poor performance for predicting high wind speed. Thus, to improve predictability of high wind speed, the prediction data for high altitudes will be considered in wind speed prediction. Third, the proposed ML-based models can only predict WS3M. The ML-based models were trained using WS3M; therefore, these models cannot predict wind speed at different heights. Although the traditional methods can predict wind speed at various heights, a large amount of information is required to implement

these methods accurately. Thus, these limitations should be explored in future research. Also, based on results of backward elimination method, slope and curvature representing characteristics of terrain were selected for input variables. The characteristics of terrain would be associated to wind speed mechanism at low altitude. In this study, quantifying how terrain is relevant in WS3M prediction was not carried out for consistency. Investigating impact of terrain on wind speed at low altitude is a good research question to improve our understanding on wind speed mechanism, particularly in agriculture areas. Hence, a relationship between terrain and wind speed, particularly in agriculture areas, should be investigated in future research.

Conclusions

The current study developed a high-resolution wind speed prediction system for agriculture purposes. The developed system consists of two parts: WS10M prediction and downscaling WS10M to WS3M. For downscaling method, traditional and ML-based methods were employed. The performance of WS3M prediction from the developed system was evaluated using the observed WS3M across South Korea.
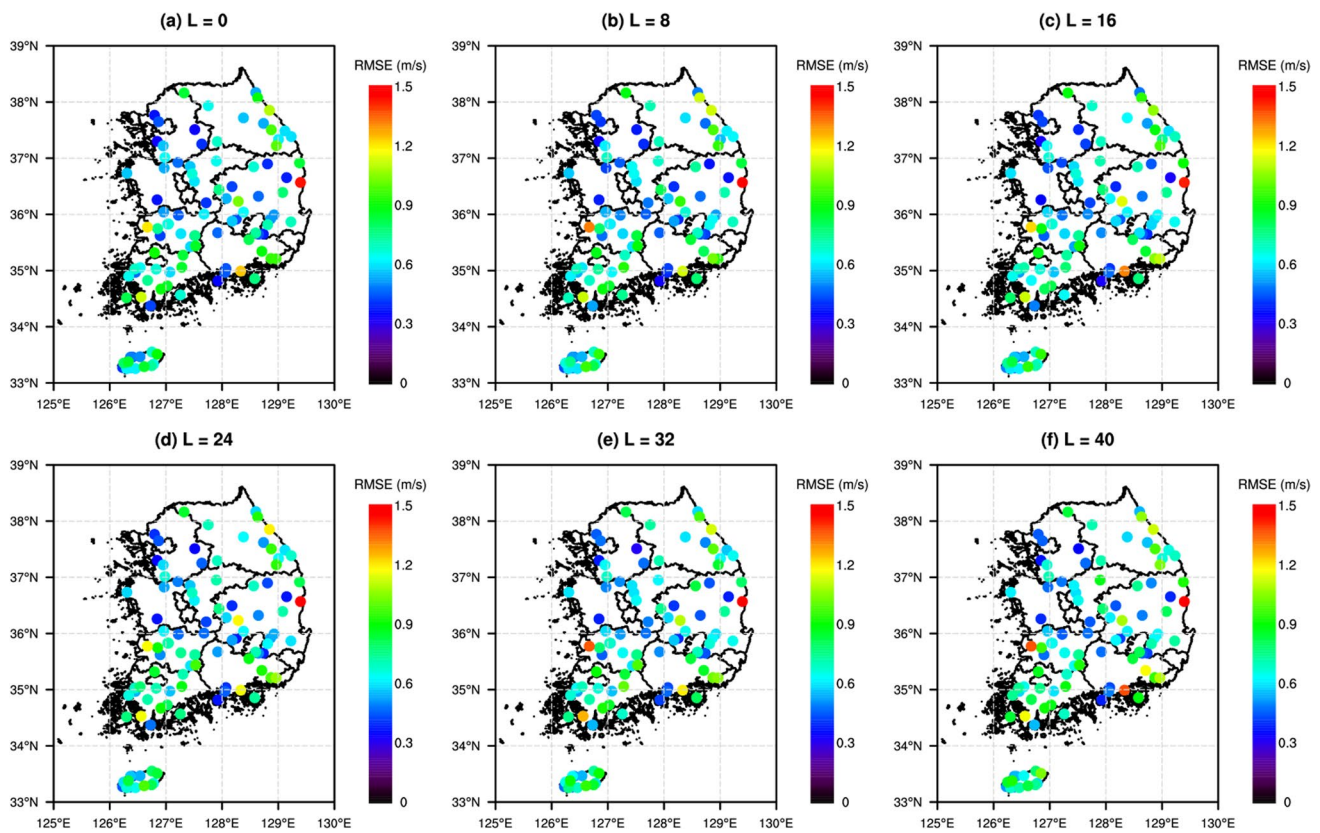
Fig. 8 Spatial distributions of RMSE from RF for different lead time (L) in the used stations

The developed wind speed prediction system provides a good performance for predicting WS3M, and this can be beneficial for agricultural applications. This performance can provide valuable wind speed information for scheduling pesticide spraying, pollen disposal, lodging prediction, and estimating canopy-top evapotranspiration in South Korea. The ML-based methods are more appropriate to predict wind speed at a fixed height, e.g., 3 m, than the traditional methods. Because the ML-based methods can integrate various variables related to wind speed and the developed model inherently pursue predicting wind speed at a fixed height, they can lead to better performances for predicting wind speed at a fixed height than the traditional methods which use roughness length, zero-plane displacement, and wind shear exponent. RF is considered as the most appropriate algorithm of the tested ML algorithms for downscaling wind speed to the fixed height. The ML method, e.g., ELM, which has the limited capacity to consider categorical variable may need delicate procedure to preprocessing the categorical variables for considering them. Thus, applying an algorithm that is easy to consider categorical variables would lead to successful implementation for wind speed prediction. These results bolster the fact that the ML algorithm, which can successfully consider categorical variables, would be a good option for predicting weather variables related to agriculture.

Author contribution J.S." conceptualization, methodology, validation, formal analysis, investigation, visualization, writing-original draft preparation, and writing-review and editing, B.M.; conceptualization, methodology, software, validation, formal analysis, data acquisition, modeling, analysis, resources, visualization, writing-original draft preparation, K.K.; conceptualization, validation, investigation, supervising, writing-original draft preparation, and writing-review and editing. All the authors have read and agreed to the published version of the manuscript.

Declarations

Competing interests The authors declare no competing interests.

References

AASC (1985) Heights and exposure standards for sensors on automated weather stations. The State Climatologist vol 9. American Association of State Climatologists,

Abbas F, Afzaal H, Farooque AA, Tang S (2020) Crop yield prediction through proximal sensing and machine learning algorithms. Agronomy 10(7):1046. https://doi.org/10.3390/agronomy10071046

Aghbalou N, Charki A, Elazzouzi SR, Reklaoui K (2018) A probabilistic assessment approach for wind turbine-site matching. In J Elec Power & Energy Syst 103:497–510. https://doi.org/10.1016/j.ijepes.2018.06.018

Allen RG, Pereira LS, Raes D, Smith M (1998) Crop evapotranspiration-guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56. Fao, Rome 300(9):D05109

ASBAE (2006) Measurement and reporting practices for automatic agricultural weather stations. vol ASAE S526. American Society of Agricultural and Biological Engineers, MI, USA

Blackadar AK, Tennekes H (1968) Asymptotic similarity in neutral barotropic planetary boundary layers. J Atmos Sci 25(6):1015–1020

Bodini N, Optis M (2020) The importance of round-robin validation when assessing machine-learning-based vertical extrapolation of wind speeds. Wind Energy Sci 5(2):489–501

Brasseur O (2001) Development and application of a physical approach to estimating wind gusts. Mon Weather Rev 129(1):5–25. https://doi.org/10.1175/1520-0493(2001)129%3c0005:daaoap%3e2.0.co;2

Breiman L (2001) Random forests. Mach Learn 45(1):5–32. https://doi.org/10.1023/A:1010933404324

Cao S, Tamura T (2006) Experimental study on roughness effects on turbulent boundary layer flow over a two-dimensional steep hill. J Wind Eng Indl Aerodyn 94(1):1–19. https://doi.org/10.1016/j.jweia.2005.10.001

Cermak JE Applications of wind data and definition of needs. In: Mehta KC (ed) The workshop on wind climate, Asheville, North Carolina, November 12–13 1979. pp 9–24

Chavan DS Gaikwad S Singh A Parashar D Saahil V Sankpal J Karandikar P Impact of vertical wind shear on wind turbine performance. In: 2017 International Conference on Circuit, Power and Computing Technologies (ICCPCT), 2017. IEEE, pp 1–6. https://doi.org/10.1109/ICCPCT.2017.8074395

Cho D, Yoo C, Im J, Cha D-H (2020) Comparative assessment of various machine learning-based bias correction methods for numerical weather prediction model forecasts of extreme air temperatures in urban areas. Earth and Space Sci 7(4):e2019EA000740. https://doi.org/10.1029/2019EA000740

Cleugh HA, Miller JM, Böhm M (1998) Direct mechanical effects of wind on crops. Agrofor Syst 41(1):85–112. https://doi.org/10.1023/A:1006067721039

De Bruin H, Verhoef A (1997) A new method to determine the zero-plane displacement. Bound-Layer Meteorol 82(1):159–164. https://doi.org/10.1023/A:1000233230943

Emeis S (2014) Current issues in wind energy meteorology. Meteorol Appl 21(4):803–819. https://doi.org/10.1002/met.1472

Evans IS (1972) General geomorphometry, derivatives of altitude, and descriptive statistics. In: C. RJ (ed) Spatial analysis in geomorphology. Routledge, London, pp 17–90

Feng S, Kong D, Ding W, Ru Z, Li G, Niu L (2019) A novel wheat lodging resistance evaluation method and device based on the thrust force of the stalks. PLoS ONE 14(11):e0224732. https://doi.org/10.1371/journal.pone.0224732

Gardiner B, Berry P, Moulia B (2016) Wind impacts on plant growth, mechanics and damage. Plant Sci 245:94–118. https://doi.org/10.1016/j.plantsci.2016.01.006

Garratt JR (1994) Review: the atmospheric boundary layer. Earth-Sci Rev 37(1):89–134. https://doi.org/10.1016/0012-8252(94)90026-4

Ghandehari M, Buttenfield BP, Farmer CJ (2019) Comparing the accuracy of estimated terrain elevations across spatial resolution. Int J Remote Sens 40(13):5025–5049. https://doi.org/10.1080/01431161.2019.1577581

Gommes R Challinor A Das H Dawod M Mariani L Tychon B Krüger R Otte U Vega R Trampf W (2010) Guide to agricultural meteorological practices. WMO 134

Gravina A, Cataldo J, Gambetta G, Pardo E, Fornero C, Galiger S, Pienika R (2011) Relation of peel damage in citrus fruit to wind climate in orchard and its control. Sci Hortic 129(1):46–51. https://doi.org/10.1016/j.scienta.2011.03.002

Hansen JW, Indeje M (2004) Linking dynamic seasonal climate forecasts with crop simulation for maize yield prediction in semi-arid Kenya. Agric for Meteorol 125(1):143–157. https://doi.org/10.1016/j.agrformet.2004.02.006

Howard T, Clark P (2007) Correction and downscaling of NWP wind speed forecasts. Meteorol Appl 14(2):105–116. https://doi.org/10.1002/met.12

Huang G-B, Zhu Q-Y, Siew C-K (2006) Extreme learning machine: theory and applications. Neurocomputing 70(1):489–501. https://doi.org/10.1016/j.neucom.2005.12.126

Jung C, Schindler D (2020) Integration of small-scale surface properties in a new high resolution global wind speed model. Energy Convers Manag 210:112733. https://doi.org/10.1016/j.enconman.2020.112733

Kent CW, Grimmond CSB, Gatey D, Barlow JF (2018) Assessing methods to extrapolate the vertical wind-speed profile from surface observations in a city centre during strong winds. J Wind Eng Indl Aerodyn 173:100–111. https://doi.org/10.1016/j.jweia.2017.09.007

Kim J, Sang W, Shin P, Cho H, Seo M, Yoo B, Kim KS (2015) Evaluation of regional climate scenario data for impact assessment of climate change on rice productivity in Korea. J Crop Sci Biotechnol 18(4):257–264. https://doi.org/10.1007/s12892-015-0103-z

Kim D-J, Kang G, Kim D-Y, Kim J-J (2020) Characteristics of LDAPS-predicted surface wind speed and temperature at automated weather stations with different surrounding land cover and topography in Korea. Atmosphere 11(11):1224. https://doi.org/10.3390/atmos11111224

KMA (2011) Numerical data application manual. Korea Meteorological Administration, Seoul, Korea

Liu L-W, Hsieh S-H, Lin S-J, Wang Y-M, Lin W-S (2021) Rice blast (Magnaporthe oryzae) occurrence prediction and the key factor sensitivity analysis by machine learning. Agronomy 11(4):771. https://doi.org/10.3390/agronomy11040771

Lubitz WD (2009) Power law extrapolation of wind measurements for predicting wind energy production. Wind Eng 33(3):259–271. https://doi.org/10.1260/0309-524X.33.3.259

Martin DE, Woldt WE, Latheef MA (2019) Effect of application height and ground speed on spray pattern and droplet spectra from remotely piloted aerial application systems. Drones 3(4):83

McAneney K, Judd M, Trought M (1984) Wind damage to kiwifruit (Actinidia chinensis Planch.) in relation to windbreak performance. New Zealand J Agric Res 27(2):255–263. https://doi.org/10.1080/00288233.1984.10430427

Mearns LO, Mavromatis T, Tsvetsinskaya E, Hays C, Easterling W (1999) Comparative responses of EPIC and CERES crop models to high and low spatial resolution climate change scenarios. J Geophys Res: Atmos 104(D6):6623–6646. https://doi.org/10.1029/1998JD200061

Meyer D Dimitriadou E Hornik K Weingessel A Leisch F (2021) e1071: misc functions of the Department of Statistics, Probability Theory Group (Formerly: E1071). 1.7–6. edn., TU Wien

Mohandes M, Rehman S, Abido M, Badran S (2016) Convertible wind energy based on predicted wind speed at hub-height. Energ Source, Part A 38(1):140–148

Monin AS, Obukhov AM (1954) Basic laws of turbulent mixing in the surface layer of the atmosphere. Contrib Geophys Inst Acad Sci USSR 151(163):e187

NAAS (2018) Safety manual for spraying pesticide using UAVS. Prevention of crop pests and diseases, vol 1, 1st edn. National Institue of Agricultural Sciences, Jeonrabuk-do

Optis M, Monahan A, Bosveld FC (2016) Limitations and breakdown of Monin-Obukhov similarity theory for wind profile extrapolation under stable stratification. Wind Energy 19(6):1053–1072. https://doi.org/10.1002/we.1883

Pal M (2005) Random forest classifier for remote sensing classification. Int J Remote Sens 26(1):217–222. https://doi.org/10.1080/01431160412331269698

Peterson E, Hennessey JP (1978) On the use of power laws for estimates of wind power potential. J Appl Meteorol 17:390–394. https://doi.org/10.1175/1520-0450(1978)017%3c0390:OTUOPL%3e2.0.CO;2

Retta A, Armbrust DV, Hagen LJ, Skidmore EL (2000) Leaf and stem area relationships to masses and their height distributions in native grasses. Agron J 92(2):225–230. https://doi.org/10.2134/agronj2000.922225x

Retuerto R, Woodward F (2004) Effects of windspeed on the growth and biomass allocation of white mustard Sinapis alba L. Oecologia 92:113–123

Schmidli J, Rotunno R (2012) Influence of the valley surroundings on valley wind dynamics. J Atmos Sci 69(2):561–577. https://doi.org/10.1175/JAS-D-11-0129.1

Seibert P, Beyrich F, Gryning S-E, Joffre S, Rasmussen A, Tercier P (2000) Review and intercomparison of operational methods for the determination of the mixing height. Atmos Environ 34(7):1001–1027. https://doi.org/10.1016/S1352-2310(99)00349-0

Shaw RH, Pereira AR (1982) Aerodynamic roughness of a plant canopy: a numerical experiment. Agric Meteorol 26(1):51–65. https://doi.org/10.1016/0002-1571(82)90057-7

Sheridan P, Smith S, Brown A, Vosper S (2010) A simple height-based correction for temperature downscaling in complex terrain. Meteorol Appl 17(3):329–339. https://doi.org/10.1002/met.177

Sheridan P Vosper S Smith S (2018) A physically based algorithm for downscaling temperature in complex terrain. J Appl Meteorol Climatol 57. https://doi.org/10.1175/JAMC-D-17-0140.1

Shin J-Y, Kim B-Y, Park J, Kim KR, Cha JW (2020) Prediction of leaf wetness duration using geostationary satellite observations and machine learning algorithms. Remote Sens 12(18):3076. https://doi.org/10.3390/rs12183076

Shin J-Y, Kim KR, Ha J-C (2020) Seasonal forecasting of daily mean air temperatures using a coupled global climate model and machine learning algorithm for field-scale agricultural management. Agric for Meteorol 281:107858. https://doi.org/10.1016/j.agrformet.2019.107858

Shin J-Y, Park J, Kim KR (2021) Emulators of a physical model for estimating leaf wetness duration. Agronomy 11(2):216

Smith A (2010) Image segmentation scale parameter optimization and land cover classification using the random forest algorithm. J Spat Sci 55(1):69–79. https://doi.org/10.1080/14498596.2010.487851

Sterling M, Baker CJ, Berry PM, Wade A (2003) An experimental investigation of the lodging of wheat. Agric for Meteorol 119(3):149–165. https://doi.org/10.1016/S0168-1923(03)00140-0

Tackenberg O (2003) Modeling long-distance dispersal of plant diaspores by wind. Ecol Monogr 73(2):173–189

Takle ES, Anderson CJ, Andresen J, Angel J, Elmore RW, Gramig BM, Guinan P, Hilberg S, Kluck D, Massey R (2014) Climate forecasts for corn producer decision making. Earth Interact 18(5):1–8. https://doi.org/10.1175/2013EI000541.1

Tennekes H (1973) The logarithmic wind profile. J Atmos Sci 30(2):234–238. https://doi.org/10.1175/1520-0469(1973)030%3c0234:TLWP%3e2.02.CO;2

Van Gardingen P Grace J (1991) Plants and wind. In: Callow JA (ed) Advances in Botanical Research, vol 18. Academic Press, pp 189–253. https://doi.org/10.1016/S0065-2296(08)60023-3

Vassallo D, Krishnamurthy R, Fernando HJ (2020) Decreasing wind speed extrapolation error via domain-specific feature extraction and selection. Wind Energy Sci 5(3):959–975. https://doi.org/10.5194/WES-5-959-2020

Watt-Meyer O, Brenowitz ND, Clark SK, Henn B, Kwa A, McGibbon J, Perkins WA, Bretherton CS (2021) Correcting weather and climate models by machine learning nudged historical simulations. Geophys Res Lett 48(15):e2021GL092555. https://doi.org/10.1029/2021GL092555

Wen W, Gu S, Xiao B, Wang C, Wang J, Ma L, Wang Y, Lu X, Yu Z, Zhang Y, Du J, Guo X (2019) In situ evaluation of stalk lodging resistance for different maize (Zea mays L.) cultivars using a mobile wind machine. Plant Methods 15(1):96. https://doi.org/10.1186/s13007-019-0481-1

WMO (2018) Guide to instruments and methods of observation. World Meteorological Organisation,

Wright MN, Ziegler A (2017) ranger: a fast implementation of random forests for high dimensional data in C++ and R. J Stat Software 77(1):1–17. https://doi.org/10.18637/jss.v077.i01

Wu J, Zha J, Zhao D (2016) Estimating the impact of the changes in land use and cover on the surface wind speed over the East China Plain during the period 1980–2011. Cilm Dynam 46(3):847–863. https://doi.org/10.1007/s00382-015-2616-z

Xu L, Zhang W-J (2001) Comparison of different methods for variable selection. Anal Chim Acta 446(1):475–481. https://doi.org/10.1016/S0003-2670(01)01271-5

Yang F, White MA, Michaelis AR, Ichii K, Hashimoto H, Votava P, Zhu A, Nemani RR (2006) Prediction of continental-scale

evapotranspiration by combining MODIS and AmeriFlux data through support vector machine. IEEE Trans Geosci Remote Sens 44(11):3452–3461. https://doi.org/10.1109/TGRS.2006.876297

Yao Y, Liang S, Li X, Chen J, Liu S, Jia K, Zhang X, Xiao Z, Fisher JB, Mu Q, Pan M, Liu M, Cheng J, Jiang B, Xie X, Grünwald T, Bernhofer C, Roupsard O (2017) Improving global terrestrial evapotranspiration estimation using support vector machine by integrating three process-based algorithms. Agric for Meteorol 242:55–74. https://doi.org/10.1016/j.agrformet.2017.04.011

Yun J, Kim Y-H, Choi H-W (2021) Analyses of the meteorological characteristics over South Korea for wind power applications using KMAPP. Atmosphere 31(1):1–15. https://doi.org/10.14191/Atmos.2021.31.1.001

Zhang T, Huang Y, Reddy KN, Yang P, Zhao X, Zhang J (2021) Using machine learning and hyperspectral images to assess damages to corn plant caused by glyphosate and to evaluate recoverability. Agronomy 11(3):583. https://doi.org/10.3390/agronomy11030583