International Journal of Biometeorology
https://doi.org/10.1007/s00484-023-02544-x

ORIGINAL PAPER

Impacts of meteorological variables and machine learning algorithms on rice yield prediction in Korea

Subin Ha[1] · Yong-Tak Kim[1] · Eun-Soon Im[1,2] · Jina Hur[3] · Sera Jo[3] · Yong-Seok Kim[3] · Kyo-Moon Shim[3]

Received: 16 November 2022 / Revised: 20 July 2023 / Accepted: 15 August 2023

Abstract

As crop productivity is greatly influenced by weather conditions, many attempts have been made to estimate crop yields using meteorological data and have achieved great progress with the development of machine learning. However, most yield prediction models are developed based on observational data, and the utilization of climate model output in yield prediction has been addressed in very few studies. In this study, we estimate rice yields in South Korea using the meteorological variables provided by ERA5 reanalysis data (ERA-O) and its dynamically downscaled data (ERA-DS). After ERA-O and ERA-DS are validated against observations (OBS), two different machine learning models, Support Vector Machine (SVM) and Long Short-Term Memory (LSTM), are trained with different combinations of eight meteorological variables (mean temperature, maximum temperature, minimum temperature, precipitation, diurnal temperature range, solar irradiance, mean wind speed, and relative humidity) obtained from OBS, ERA-O, and ERA-DS at weekly and monthly timescales from May to September. Regardless of the model type and the source of the input data, training a model with weekly datasets leads to better yield estimates compared to monthly datasets. LSTM generally outperforms SVM, especially when the model is trained with ERA-DS data at a weekly timescale. The best yield estimates are produced by the LSTM model trained with all eight variables at a weekly timescale. Altogether this study shows the significance of high spatial and temporal resolution of input meteorological data in yield prediction, which can also serve to substantiate the added value of dynamical downscaling.

Keywords Rice yield prediction · Machine learning model · Dynamical downscaling

Introduction

Weather and climate conditions are the key factors in agriculture as they greatly affect the growth and development of agricultural products. The yield variability of crops is reported to be highly susceptible to heat and water stress, particularly during their late vegetative and early reproductive phases (Teasdale and Cavigelli 2017). Given the great influence of meteorological factors on crop yields, numerous attempts have been made to investigate how weather and climate variability affects the yields of different commercial crops worldwide such as rice, wheat, and maize (Alexandrov and Hoogenboom 2000, 2001; Auffhammer et al. 2012; Oguntunde et al. 2018; Ray et al. 2015; Rodríguez-Puebla et al. 2007; Tao et al. 2014). The inextricable connection between meteorological variables and crop yields underlies the statistical approach for yield estimation. With readily available and easily accessible weather data, crop yield prediction models are developed based on empirically-driven regression equations between the yield and meteorological variables. This weather-based statistical model has a comparative advantage in terms of simplicity and data requirements over a bio-physical model, which usually incorporates extensive amounts of input data including information about cultivars, irrigation, management, and soil conditions in addition to weather data (Lobell and Asseng 2017). The abundant variety of datasets enables

Subin Ha and Yong-Tak Kim contributed equally to this work.

✉ Eun-Soon Im
ceim@ust.hk

1 Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong SAR, China

2 Division of Environment and Sustainability, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong SAR, China

3 National Institute of Agricultural Sciences, Rural Development Administration, Wanju-gun, Jeollabuk-do, Korea

Published online: 05 September 2023

the bio-physical model to describe dynamical processes of crop growth and yield formation, so if all these datasets are readily provided, bio-physical crop modeling could be very useful in predicting yields with wide applicability from a local or regional level to a national level (Morell et al. 2016; van Wart et al. 2013). However, not all parts of the world possess all these datasets required by bio-physical models, and uncertainty in many parameters makes the calibration and evaluation of data difficult (Lobell and Burke 2010). Weather-based statistical models, on the other hand, require meteorological data only as input data, so they gain advantages over bio-physical models in terms of accessibility and scalability (Joshi et al. 2021; Mathieu and Aires 2016). Moreover, they can be easily employed to predict future crop yields (Lobell and Burke 2010).

A variety of machine learning and deep learning algorithms have been actively utilized for developing yield prediction models for various crops. As one of the simplest forms of machine learning algorithms, multiple linear regression has been widely used to predict yields of numerous crops including potato, maize, soybean, hulless barley, arabica coffee, and rice in different parts of the world (Abrougui et al. 2019; Das et al. 2018; Joshi et al. 2021; Kittichotsatsawat et al. 2022; Matsumura et al. 2015; Zaefizadeh et al. 2011). As the temporal and spatial resolution of meteorological information has improved, recent studies tend to adopt more advanced machine learning techniques such as an artificial neural network (Crane-Droesch 2018; Ji et al. 2007; Kaul et al. 2005), support vector machine (SVM; Gandhi et al. 2016; Jaikla et al. 2008; Joshi et al. 2021; Su et al. 2017), generalized additive model (Chen et al. 2019; Joshi et al. 2021; Onwuchekwa-Henry et al. 2022), and long short-term memory (LSTM; Sun et al. 2019; Tian et al. 2021). The wide variety of algorithms allows performance comparison between them, providing valuable information about which machine learning technique is more suitable for the development of crop yield prediction. For example, Joshi et al. (2021) show the superiority of SVM in corn and maize estimation in the US central Corn Belt, whereas Cao et al. (2021) demonstrate that LSTM shows a better performance in predicting rice yield in China compared to random forest.

In this study, we adopt SVM and LSTM algorithms to predict rice yields in Korea based on fine-scale meteorological information. Rice is the most important staple food in South Korea, providing more than 60% of the starch-based calories to the population and accounting for more than half of the total farmland usage (Jeong et al. 2021). In accordance with the high demand and supply of rice in the nation, the estimation of rice yields has been actively studied in Korea using various methods (Hong et al. 2012; Jeong et al. 2018; Na et al. 2012, 2013; Yun 2003). South Korea is reported to be one of the nations where the rice yield is highly dependent on weather conditions. Annually up to half of the rice yield variations in Korea are explained by weather variability, especially that of temperature (Ray et al. 2015). Other weather factors like duration of sunshine, diurnal temperature range (DTR), and precipitation have also substantial influences on rice productivity, together and separately (Chen et al. 2016; Prabhjyot-Kaur et al. 2021). Especially, DTR is demonstrated to play a significant role in the growth and development of crops (Hu and Buyanovsky 2003; Lobell 2007; Tack et al. 2015; Verón et al. 2015). Besides, although more than 80% of paddy fields in Korea are irrigated (MAFRA and KRC 2022), precipitation is still important for rice production in Korea because spring drought due to low precipitation may lower the water level in the reservoir and thus delay rice planting, or fields may be flooded by prolonged rain season or heavy rain in summer. In this regard, this study investigates the effect of input meteorological variables on rice yield prediction by comparing the yield estimates obtained from different combinations of input data.

The input meteorological datasets for statistical crop yield prediction models have been mostly sourced from in-situ observational or satellite data. However, observational data are limited in terms of spatial coverage and sometimes omitted depending on the condition of weather instruments, for example, in case of a temporary failure, damage, or disruption. In contrast, climate models enhance the consistency and coverage of agrometeorological information by seamlessly providing weather information even for areas with limited or no observation stations, and more importantly, enable future yield prediction by providing long-lead-time forecasts. A few studies have employed global reanalysis data in crop yield or phenology phase prediction (Cho et al. 2021; Oses et al. 2020), but due to the coarse resolution of a Global Climate Model (GCM), the direct application of GCM data into a small domain like South Korea is not highly desirable. Therefore, we conduct dynamical downscaling using a Regional Climate Model (RCM) and use the downscaled data to build yield prediction models in order to investigate if the climate model data can replace observations and if the benefit of higher-resolution simulations is recognized not only in the representation of climatology but also in rice yield estimation. In addition to the spatial resolution, we also evaluate if increasing the temporal resolution of the climate data helps improve yield estimates by dividing the datasets into two different timescales, weekly and monthly. The overall goal of this study is to develop rice yield prediction models for South Korea based on SVM and LSTM algorithms fed by observational data as well as climate model output data. A comparison of yield prediction derived from various combinations of meteorological input data with different temporal and spatial resolutions will provide valuable insight into the dependence of rice yield on meteorological variables, demonstrating the added value of dynamical downscaling from an application-wise perspective.

Data and methods

Study area and yield data

For model training and testing, this study utilizes rice yields during the last 40 years (1982–2021) in South Korea, which are obtained from the KOrean Statistical Information Service (KOSIS, https://kosis.kr/). Out of a total of 17 provincial-level administrative divisions, we use the yield data of 11 divisions after excluding those whose share in the national total yield has remained less than 2% for the past ten years or whose yield data is not available for the whole 40-year study period. The list of the divisions included in the analysis and those excluded from the analysis is available in Table S1 in Electronic Supplementary Material (ESM). The annual rice yields are linearly detrended for each administrative division in order to remove non-climate effects such as advances in agricultural technology. In our study, the linear trend in the 40-year annual rice yields is removed using the "signal.detrend" function (type = "linear") from SciPy.

Meteorological data

In this study, we utilize meteorological data from three different sources: observations, reanalysis, and dynamically downscaled reanalysis. Firstly, daily observational data are obtained from 56 in-situ stations across the nation during the 40-year study period of 1982–2021. The locations of the stations are plotted in Fig. 1(a). The obtained weather data include daily mean temperature (Tmean), maximum temperature (Tmax), minimum temperature (Tmin), precipitation, relative humidity (RH), solar irradiance (SI), and mean wind speed (WSmean). One additional variable, DTR, is calculated as the difference between Tmax and Tmin. Next, the data of these variables are obtained from ECMWF ERA5 Reanalysis hourly data at the $0.25° \times 0.25°$ resolution (Hersbach et al. 2020). Finally, to generate higher-resolution meteorological data, we conduct dynamical downscaling of ERA5 over South Korea using Weather Research and Forecasting. The topography of the domain is shown in Fig. 1(b). With the 5-km resolution, the complicated topographical features of the Korean Peninsula are well described, such as a long, narrow mountain range along the eastern coast with steep elevation gradients, low-lying basins, and complicated coastlines. The physical parameterizations of the modeling system follow the optimal setting that Qiu et al. (2020) found through various sensitivity experiments. The initial and boundary conditions, ERA5, are obtained at the resolution of $0.25° \times 0.25°$ at 6-h intervals. The observations, the original ERA5 data, and the 5-km downscaled ERA5 data are referred to as OBS, ERA-O, and ERA-DS, respectively. As OBS is obtained from 56 in-situ observational stations, meteorological data of 56

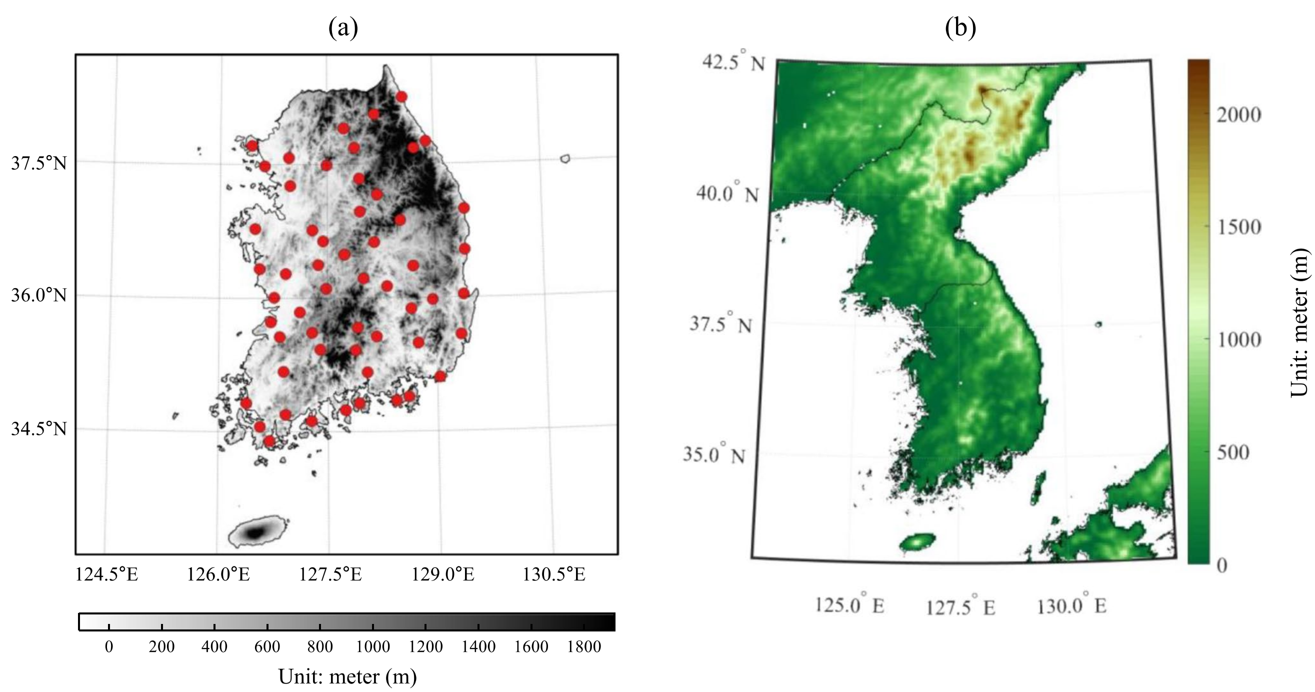

Fig. 1 (**a**) Locations of 56 in-situ observational stations and (**b**) WRF model domain for dynamical downscaling

grid points, each of which is located nearest to the corresponding in-situ station, are extracted from ERA-O and ERA-DS.

The meteorological data from OBS, ERA-O, and ERA-DS are prepared as input datasets at two different timescales, weekly and monthly, in order to evaluate the impact of the temporal resolution of the input data on model performance. In Korea, transplanting is conducted in May, and harvesting starts in late September in the northern region and finishes by late October in the southern region (Rural Development Administration, personal communication). The data are therefore obtained to cover from May to September each year, the period where rice productivity is largely affected by atmospheric conditions. The weekly (monthly) dataset consists of weekly (monthly) averages of Tmean, Tmax, Tmin, DTR, RH, and WSmean, and weekly (monthly) accumulated precipitation and SI. Since the yield data is available at a division level, the weekly and monthly weather input datasets are prepared for each administrative division based on the Thiessen polygon method so that the meteorological data and the yield data are properly matched for each division. The input data are normalized before being utilized to train and test the models. Each variable is measured in a different unit and varies significantly in magnitude, but data normalization rescales all the data features to the same standard range and therefore eliminates data inconsistencies arising from different units, magnitudes, and ranges of input variables. In our study, the z-score normalization method is adopted, which is relatively robust against outliers and minimizes their influence on the data.

Meanwhile, DTR itself may be misleading when used alone, because DTR is simply the difference between Tmax and Tmin within one day, so days with totally different Tmax and Tmin can have the same DTR value. This issue can be addressed by pairing DTR with Tmean, which is in the range between Tmax and Tmin. To compare the impacts of input variables, four datasets are composed of different combinations of meteorological variables. Dataset 1 is composed of Tmean and precipitation, the two most common variables used in yield prediction. Dataset 2 consists of Tmean, precipitation, and DTR, which is designed with the purpose of evaluating the impact of the inclusion of DTR on the performance of the model compared to the one developed with Tmean and precipitation only. RH, SI, and WSmean are then added to form a richer dataset (Dataset 3), advanced from the traditional dataset of temperature and rainfall. Lastly, the full dataset with all eight variables (Dataset 4) is prepared to examine if Tmax and Tmin play further roles in yield forecasting, aside from their previous role as the components of DTR.

Model development

In this study, rice yield prediction models are developed using two well-known machine learning algorithms, SVM and LSTM, in order to estimate the annual yields of the most recent decade (2012–2021) for each administrative division. The Bayesian optimization algorithm is applied to optimize the hyperparameters of the SVM and LSTM models. (The selection result of the parameters for each model is available in Table S2 in ESM.) Also, both models adopt the Leave-One-Out Cross-Validation (LOOCV) method. The LOOCV approach is often used to assess the reliability and robustness of a crop yield prediction model (Kogan et al. 2013; Kouadio et al. 2014, 2021; Liu et al. 2020; Mkhabela et al. 2011). For the SVM model, the yields of 39 years (e.g., 1982–2014, 2016–2021) are used to train the model, and the predicted yield for the remaining one year (e.g., 2015) is used to test the performance of the developed model through comparison to the actual yield of that year (See Figure S1). In the case of the LSTM model, since it is a time-series model, yield data beyond a target year cannot be included in a training dataset. As a result, the training dataset used in the LSTM model development contains yield data from 1982 to the year right before the target year. For example, the yield data from 1982 to 2011 are used to estimate the yield in 2012, the data from 1982 to 2012 to estimate the yield in 2013, and so on. For both SVM and LSTM, the process is repeated 10 times, from 2012 to 2021, resulting in 10-year predicted yields for each administrative division. The LSTM model takes into consideration not only sequential but also geographical features of data in each division (Clauss et al. 2018; Garg et al. 2013; Oyoshi et al. 2016), which differentiates itself from the SVM model. The conceptual diagram of the LOOCV method applied for SVM and LSTM is presented in Figure S1 in ESM.

Results

Relationship between rice yields and meteorological variables

The relative importance of each meteorological variable to the rice yield variation can be estimated by the correlation coefficient (r) between the variable and the yield. Figure 2 depicts the temporal evolution patterns of the correlation between each meteorological variable (obtained from OBS) and the yield from May to September (MJJAS) at weekly and monthly timescales, respectively. It is evident that weather conditions in July and August are most correlated with rice yields, whether positively or negatively, as depicted by the vivid orange or blue colors for many variables during those two months (Fig. 2(a)). From a weekly perspective

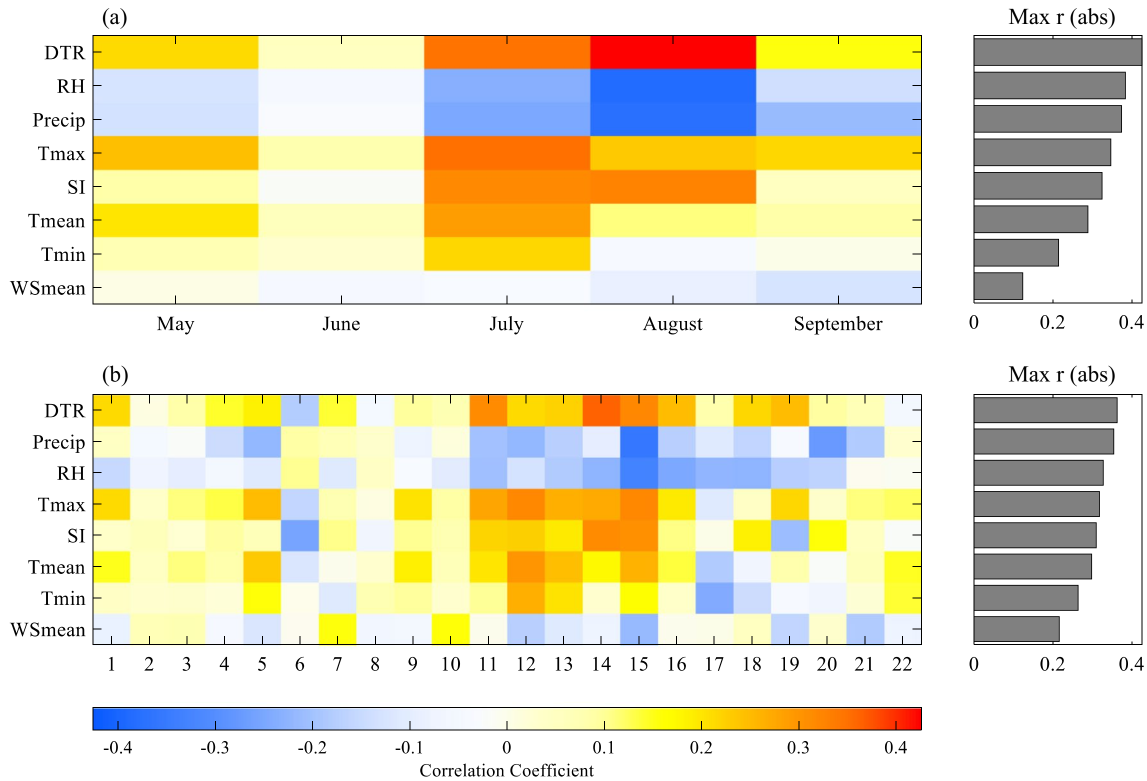

Fig. 2 Temporal evolution patterns of the correlation coefficient (r) between meteorological parameters and the rice yield at the (**a**) monthly and (**b**) weekly levels. Positive correlations are colored in red and negative correlations in blue. The variables are ranked by their maximum correlation coefficient (max r), which is plotted in bars on the right

(Fig. 2(b)), stronger correlations between the meteorological conditions and the yields are observed from Week 11 to Week 15 (mid-July to mid-August), namely during the reproductive stage. At both temporal levels, DTR plays the most significant role in the variation of rice yields, showing the maximum r value among the eight variables. It consistently exhibits a positive correlation with the yield during MJJAS at a monthly level, and a particularly strong positive correlation is observed in August with the peak in week 14 (July 31st–August 6th). The prominent pattern of DTR is in accordance with Chen et al. (2016), which states that DTR contributes to rice yield variability more than precipitation, RH, and Tmean. The next significant variables are RH and precipitation. With a small difference in their maximum r values, these two variables have very similar temporal patterns in their correlation with rice productivity, which may be explained by the close link between rainfall and RH. They are negatively correlated with the yield throughout MJJAS at a monthly timescale with the largest degree in August, which is the same pattern as DTR but in the opposite direction. At a weekly level, the negative correlations are consistently observed from week 11 to week 20, from mid-July to mid-September, which corresponds to the rainy season. The next significant variables are Tmax and SI. The positive response of the yield to Tmax gets intense during July at the monthly level, or from week 11 to week 15 at the weekly level which corresponds to the period from mid-July to mid-August, when extremely hot weather events are frequent. SI also generally affects the rice yield positively, especially in July and August which are characterized by high radiation as the pinnacle of summer. The benefit of increased solar radiation on rice production is reported by Zhang et al. (2010). Temporal changes of Tmean and Tmin have a relatively small impact on the variations of rice yield. They show a positive correlation with the yields across the board, particularly strongly in July. It is notable that Tmean, which is the weather data most commonly used to develop a rice yield prediction model, is indeed not very influential in the yield variation. This is in agreement with the relative significance of meteorological variables to rice productivity in order of DTR, precipitation, RH, and Tmean, as stated in Chen et al. (2016). Also, it is worthy of notice that Tmin itself does not form a strong correlation with the yield, but its difference from Tmax has a great influence on the yield dynamics. WSmean is revealed to be most weakly correlated with yield variances. Temporal changes of WSmean barely elicit a weak negative response from the rice yield with the maximum r value of 0.2 at the weekly level and 0.1 at the monthly level. Though the degree is small, the negative correlation between WSmean and the yield is consistently shown from June to September. The negative impacts of wind speed on crop plants have been addressed in many previous studies. Firstly, strong winds may abrade plants, mainly by allowing their leaves or bodies to rub against each other, which can consequently damage their cuticle and accelerate water loss (Gardiner et al. 2016; Van Gardingen and Grace 1991). Also, as a herbaceous plant, rice is particularly vulnerable to stem or root lodging caused by mighty winds (Kashiwagi and Ishimaru 2004; Rajkumara 2008).

Validation of meteorological input data

With regards to the aforementioned variables, we evaluate the capability of ERA-O and ERA-DS for reproducing the historical climatology in comparison with OBS so as to assess whether they are adequate as input data to replace OBS for prediction models. The spatial distributions of the 40-year average MJJAS Tmax, Tmean, and Tmin of OBS, ERA-DS, and ERA-O are displayed in Fig. 3. Due to the limited number of in-situ observation stations and their uneven distribution over the country, the interpolated OBS spatial

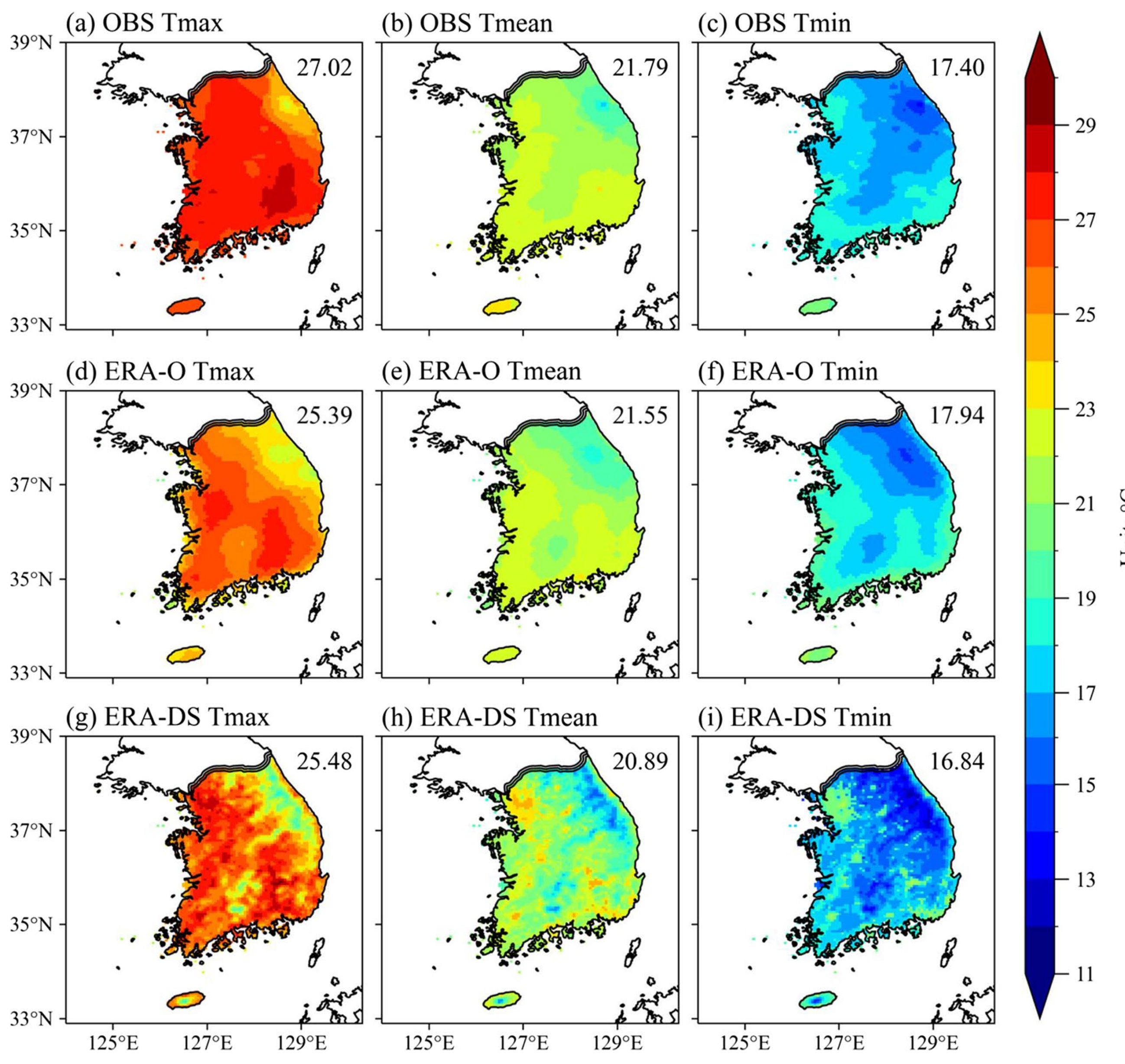

Fig. 3 **a**–**i** Spatial distributions of 40-year (1982–2021) MJJAS average Tmax (left column), Tmean (middle column), and Tmin (right column) from OBS (top row), ERA-O (middle row), and ERA-DS (bottom row). For fair comparisons between OBS (obtained from 56 in-situ stations), ERA-O, and ERA-DS, OBS and ERA-O are interpolated into the 5 km grid of the ERA-DS domain. The value at the top right of the individual figures indicates the spatial average over South Korea

maps may not be the perfect description of the temperature climatology of South Korea, but the general topographical impact on temperature is well described. The temperature is relatively low in the high-altitude mountainous areas along the northeast coast with the coldest temperature at the peak. Meanwhile, the western and southern areas have relatively high temperatures, and the highest temperature is observed in the southeastern basin. ERA-O is able to simulate the observed temperature pattern of low temperature at high altitudes and high temperature at low altitudes, but due to its coarse resolution, the temperature gradient is rather gradual over the mountains, and region-specific temperature patterns are not fully reproduced. Due to the underestimation of elevation in the GCM, ERA-O retains a positive bias for Tmin, whereas a substantial amount of negative bias is exhibited for Tmax. Benefitting from the higher spatial resolution, ERA-DS is able to more realistically describe the region-distinct temperature features, such as the low temperature over the mountainous areas and the high temperature in the southeastern basin. The ERA-DS temperature distributions may seem different from those of OBS, but it is partly attributed to the limited number and the uneven distribution of observational stations. The resolution of ERA-DS is even high enough to simulate the sharp temperature gradient along the northeastern coastlines induced by the steep elevational gradient between high-latitude mountains and low-latitude coasts over a relatively short distance, which is not even clearly illustrated by OBS due to the scarcity of observations over this area.

Another analysis of temperature is conducted to prove that the benefit of dynamical downscaling stands out in reproducing extremes. The 40-year average of annual MJJAS top 5% Tmax and lowest 5% Tmin (8 hottest and coldest days during MJJAS in each year, respectively) values of the 56 stations simulated by ERA-O and ERA-DS are compared with OBS (see Figure S2 in ESM). The underestimation of ERA-O for Tmax is clearly seen with almost all ERA-O-simulated Tmax values located below the 1-to-1 line and almost half of them even located below the 2 °C negative bias boundary. Meanwhile, most of the ERA-DS-simulated Tmax values are much closer to the observed values, with biases less than 2 °C, showing that dynamical downscaling greatly improves the predominant underestimating tendency of ERA-O. In the case of Tmin, the added value of dynamical downscaling is not as distinct as Tmax, yet ERA-DS-simulated Tmin values are generally closer to the observed values. The superiority of ERA-DS in reproducing temperature extremes, especially Tmax, gives great benefits to reproducing DTR, which is most highly correlated with yield variation.

Figure 4 depicts the interannual variations of MJJAS average precipitation, RH, SI, and WSmean anomalies calculated from OBS, ERA-O, and ERA-DS during the 40-year study period. Both ERA-O and ERA-DS well reproduce the yearly variability of precipitation, successfully capturing the wet years (e.g., 1998, 2003, 2020) and dry years (e.g., 1994, 2015). The biases are small in most years, but some relatively large gaps between OBS and ERA-O, or between OBS and ERA-DS are randomly observed in a few years, like the overestimation of ERA-DS in 1998 and 2001, and underestimation of ERA-O in 2010 and 2011. In terms of the simulation of the yearly RH variations, ERA-DS outperforms ERA-O. Although both entail biases in a number of years, the biases of ERA-DS are usually smaller in magnitude and less frequently observed compared to ERA-O (e.g., 1983–1987, 2001, 2011–2012, 2019). This may be attributed to the better temperature simulation achieved by dynamical downscaling as temperature, along with precipitation, is one of the factors that are closely connected with humidity. Regarding SI, although the two model simulations sometimes cannot grasp sharp increases or decreases (e.g., 1994–1995, and 2015–2017) in radiation, they still successfully simulate the years with greater radiation (e.g., 1994, 2004) and the years with less radiation (e.g., 1998, 2003) compared to their respective previous year, proving that they do possess a fair level of ability to simulate the radiation climatology of Korea. Similarly, in the year-to-year variations of the simulated WSmean, despite some occasional large biases, the sign of the yearly WSmean variability is well captured by the climate models in most years. For precipitation, SI, and WSmean, the benefit of dynamical downscaling is not clearly demonstrated. Since the interannual variations of an RCM-downscaled simulation come from its initial and boundary conditions (GCM data), there is a limit to the degree of improvement that can be achieved through dynamical downscaling.

Yield prediction using machine learning techniques

Our analysis of the performance of the prediction models begins with examining the results of the models developed with OBS datasets, which serve as reference results. The comparisons of the predicted and actual yields are presented in the form of scatterplots in Fig. 5. At a glance, it is obvious that models developed with weekly datasets (Fig. 5(e–h)) give better yield estimates than those with monthly datasets (Fig. 5(a–d)). Regardless of the model type and dataset, the yield dots in Fig. 5(a–d) are far more widely scattered than those in Fig. 5(e–h), indicating that many yield values predicted by the monthly models differ a lot from the actual yield values. From the viewpoint of the resemblance between the linear regression line and the 1-to-1 line, the LSTM models perform better than the SVM models with Dataset 1 and 2, but there is little difference between the two with Dataset 3 and 4 at a monthly timescale. When weekly meteorological data are used for model development, on the other hand, the LSTM models always outperform the SVM

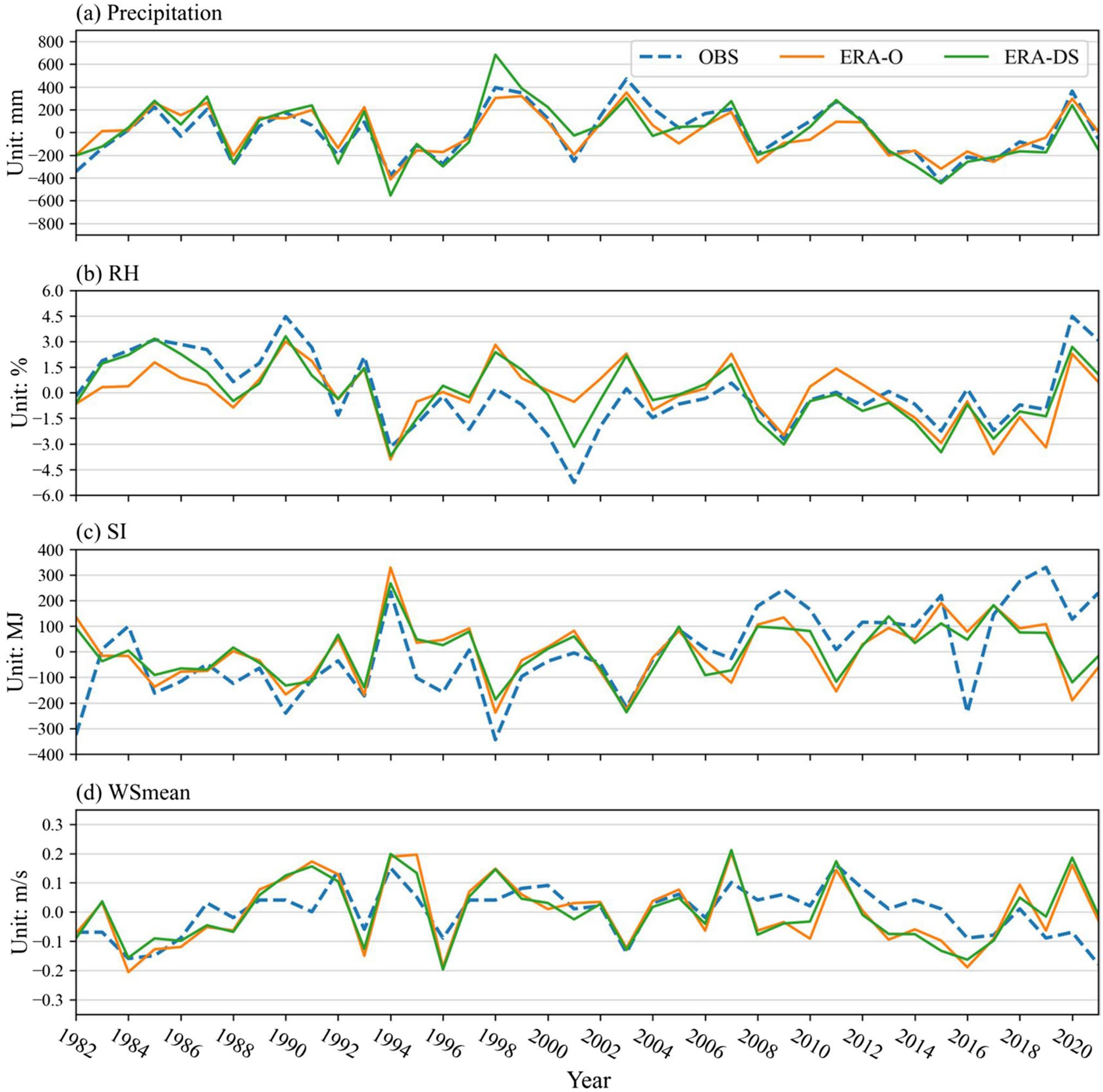

Fig. 4 Interannual variations of MJJAS average (**a**) Precipitation, (**b**) RH, (**c**) SI, and (**d**) WSmean anomalies from 1982 to 2021 (averaged over the 56 stations). OBS, ERA-O, and ERA-DS results are displayed with blue, orange, and green lines, respectively

models, irrespective of the dataset. When the dispersion of the yield points and their regression line are considered together, it is clearly seen that weekly models are superior to monthly models, and the LSTM models tend to perform better than the SVM models.

To quantitatively evaluate the performance of the yield prediction models, correlation coefficient (CORR), root mean square error (RMSE), Index of Agreement (IoA), and Nash–Sutcliffe efficiency coefficient (NSE) values of all 16 OBS models are presented in Table S3 in ESM. In addition, the difference between the weekly and monthly models and the difference between the SVM and LSTM models for each dataset are also calculated and presented in Table S3 in order to assess the impact of the temporal level and the model type on the performance of the model. The CORR, RMSE, IoA, and NSE comparison between the monthly and weekly models proves the benefit of using higher temporal-level weather data for yield prediction. The best RMSE, IoA, and NSE are 31.268 kg/1000m^2, 0.613, and 0.469 for the monthly SVM model group,

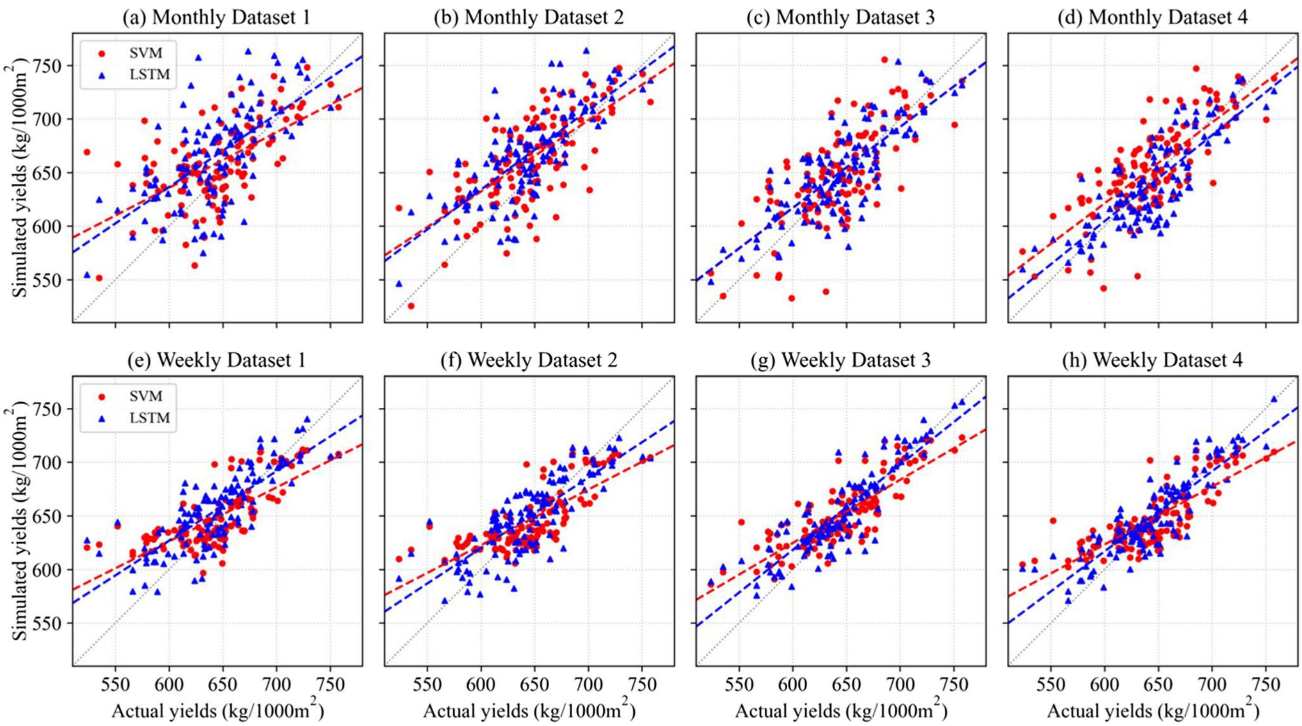

Fig. 5 Scatterplots of the actual versus predicted yields from SVM (red dots) and LSTM (blue triangles) developed with (**a**–**c**) monthly and (**d**–**f**) weekly OBS datasets. The black diagonal line represents the 1-to-1 line, and the red and blue dashed lines represent the linear regression between the actual and predicted yields from SVM and LSTM, respectively

which are poorer than the worst weekly SVM model with RMSE, IoA, and NSE of 28.967 kg/1000m^2, 0.662, and 0.544, respectively. The highest NSE value of 0.469 from the monthly SVM model group is still considered "unsatisfactory", but NSE values of all four weekly SVM models all fall into the "satisfactory" group according to Moriasi et al. (2007). The situation is similar for LSTM models; when Dataset 1 is fed to the LSTM model at a monthly scale, the resulting NSE value is even below zero, which means that the model is unacceptable (Zeybek 2018). Using the same dataset at a weekly level greatly improves the NSE value, from – 0.011 to 0.574, upgrading the model from an unacceptable level to a satisfactory level. Not only NSE, but RMSE is also noticeably reduced when a model is trained with the weekly weather data. Replacing the monthly datasets with the weekly datasets can decrease RMSE by as much as 35% for the LSTM models and by as much as 26% for the SVM models. In addition to the importance of high temporal-resolution meteorological input data, the superiority of LSTM over SVM in yield prediction is also numerically demonstrated. Only when the training dataset consists of monthly Tmean and precipitation, SVM provides slightly better yield estimates in terms of RMSE, IoA, and NSE, and LSTM outdoes SVM in all other cases. The general superiority of LSTM can be attributed to its ability to capture the temporal dependencies in the data and also the complex relationships between meteorological variables and rice yields over a long time.

The comparison of the results of the four different datasets shows that the yield prediction ability of a model largely depends on the combination and diversity of weather variables. It is indisputable that using monthly Tmean and precipitation (Dataset 1) is insufficient for building good yield prediction models since both SVM and LSTM models trained with Dataset 1 at a monthly level are absolutely unqualified in all four quantitative aspects. Adding DTR to Tmean and precipitation (Dataset 2) improves the prediction ability of the models, but the results still fall short of acceptable standards. When trained with Dataset 3, the LSTM model shows an acceptable level of performance with CORR of 0.787, RMSE of 28.18 kg/1000m^2, which is equivalent to 4.1% of the actual 10-year average yield, IoA of 0.658, and NSE of 0.568 which is regarded "satisfactory" according to Moriasi et al. (2007), even at a monthly timescale. However, if weekly meteorological data are utilized, using Dataset 1 already achieves comparable performance, which shows a clear advantage of high temporal resolution of the input weather data. Same as at the monthly level, training the models with the increased variety of meteorological variables at the weekly timescale helps enhance the prediction ability of the models, as shown by the comparison between the results of Datasets 1, 2, and 3. When trained with the complete variable set (Dataset 4), both the SVM and LSTM models generally perform at their best, except for the SVM model at a weekly timescale. Indeed, the SVM model with weekly Dataset 3 slightly outperforms that with weekly Dataset 4. It shows that a dataset with increased diversity of variables does not always lead to better predictions but may give a different result depending on which model algorithm it is fed to. Nonetheless, the LSTM model trained with Dataset 4 shows exceptionally outstanding performance with the highest CORR (0.878), lowest RMSE (21.313 kg/1000m^2, 3.1% of the actual yield), highest IoA (0.764), and highest NSE (0.753) among all models developed with OBS datasets. It is the one and only OBS model which is regarded as "very good" according to the NSE rating criteria of Moriasi et al. (2007), consolidating its position as the best-performing OBS model. At least for the LSTM algorithm, including respective Tmax and Tmin helps improve the forecasting capability of the model, proving their significance as individual variables in input datasets, which should be considered separately from the impact of their difference (DTR).

The analysis of the results of the OBS models underlies the importance of high temporal-level input data, the superiority of LSTM over SVM, and the positive impact of additional climatic information besides temperature and precipitation. In order to verify the adequacy of climate model data to replace OBS for rice yield model development and assess the impact of increasing the spatial resolution of the model data on yield prediction, the yield estimates from the models developed with ERA-O and ERA-DS data are quantitatively analyzed too. The CORR, RMSE, IoA, and NSE values of ERA-O and ERA-DS models trained with the four weekly datasets are presented in Table 1. The features shown in the OBS models are almost identically observed in the models trained with climate model data. The benefit of the utilization of weekly meteorological datasets is demonstrated with the higher CORR, IoA, and NSE, and lower RMSE values. There is one exception when the ERA-O Dataset 1 leads to a slightly higher CORR value of the SVM model at a monthly timescale, but the difference is insignificant (0.007), so it may safely be said the high temporal-resolution input climate data contribute to the development of good yield prediction models. The superiority of LSTM over SVM is also irrefutably demonstrated from a quantitative perspective, and the best yield estimates are obtained when Dataset 4 is utilized. For SVM, in fact, using the ERA-O dataset leads to better yield estimates than using the ERA-DS dataset. However, the best-performing SVM model still falls behind the LSTM models, whether the training dataset is obtained from ERA-O or ERA-DS. Contrary to SVM, LSTM gains a clear advantage from using ERA-DS data, and the benefit of using ERA-DS data instead of ERA-O for the LSTM development is greater than the benefit of using ERA-O data instead of ERA-DS in the SVM model development. In conclusion, the LSTM model developed with ERA-DS weekly Dataset 4 is evaluated as the best; the CORR is almost close to 0.9, the RMSE of 20.054 kg/1000m^2 is less than 3% of the actual yield, the IoA value of 0.816, and the NSE value of 0.781 means that the model is "very good" according to the standard of Moriasi et al. (2007). The outstanding performance of this model is also graphically well-noticed. The scatterplots of the actual and predicted yields from the ERA-O and ERA-DS models with weekly Dataset 4 are presented in Figure S3 in ESM. The numerical evaluation of the SVM weekly best models tells that ERA-O outperforms ERA-DS, but in the scatterplot, the difference between ERA-O and ERA-DS is not very apparent. On the other hand, when the ERA-DS weekly Dataset 4 is the training dataset, the LSTM model shows a great performance which is demonstrated by the distribution of its estimates along the black 1-to-1 line. The added value of dynamically downscaled climate data is clearly acknowledged in statistical yield prediction based on the LSTM algorithm.

Summary and conclusion

As climate change is expected to bring huge changes in meteorological conditions and unprecedented weather events in the future, the importance of crop yield prediction has come to the fore worldwide. A weather-based statistical model is a useful tool for yield prediction, which requires only weather information as input data without the need for other information about soil conditions, genotypes, or management. However, deterministic forecasts are only available in the short term, which cannot meet the requirements of statistical models. This problem may be solved by employing seasonal forecasts obtained from GCMs, subject to the accuracy of the prediction, but another limitation arises from the coarse resolution of the GCMs. Especially, the complicated geographical features of South Korea are not fully reflected in the usual resolution of global forecasts, which leads to the poor quality of the region-specific climatology simulations in the Korean Peninsula. Dynamical downscaling using RCMs can be used to overcome this limitation of GCM simulations by generating high-resolution simulations with a more realistic description of the topography. This study aims to evaluate the applicability of the dynamically downscaled climate model data in the development of weather-based statistical rice yield estimation models for South Korea, prior to the employment of seasonal forecasts.

As the only input, weather data are the key components of the model, so the RCM simulations are first validated by comparison with the observations. The spatial patterns of the 40-year MJJAS average Tmax, Tmean, and Tmin clearly show the added value of dynamical downscaling.

Table 1 CORR, RMSE, IoA, and NSE between the actual yields and the yields predicted by SVM and LSTM with monthly (skipped) and weekly ERA-O and ERA-DS datasets. The differences between the weekly and monthly results, and between the SVM and LSTM results are also presented. Changes for the better are set in italics while changes for the worse in boldface

Data source	Temporal level	Model	SVM				LSTM				Difference (LSTM – SVM)			
		Dataset	CORR	RMSE	IoA	NSE	CORR	RMSE	IoA	NSE	CORR	RMSE	IoA	NSE
ERA-O	Weekly	1	0.711	30.200	0.622	0.504	0.730	29.719	0.653	0.520	*+0.019*	*−0.481*	*+0.031*	*+0.016*
		2	0.760	28.294	0.656	0.565	0.800	26.181	0.691	0.627	*+0.04*	*−2.113*	*+0.035*	*+0.062*
		3	0.801	25.813	0.689	0.638	0.826	24.378	0.725	0.677	*+0.025*	*-1.435*	*+0.036*	*+0.039*
		4	0.792	26.518	0.678	0.618	0.884*	21.021*	0.768*	0.760*	*+0.093*	*-5.498*	*+0.090*	*+0.142*
	Difference (Weekly –Monthly)	1	**−0.007**	*−5.991*	*+0.099*	*+0.216*	*+0.081*	*−9.100*	*+0.125*	*+0.339*				
		2	*+0.04*	*−8.173*	*+0.137*	*+0.288*	*+0.035*	*−3.266*	*+0.041*	*+0.098*				
		3	*+0.056*	*−3.065*	*+0.041*	*+0.091*	*+0.006*	*−4.608*	*+0.067*	*+0.134*				
		4	*+0.164*	*−9.217*	*+0.094*	*+0.312*	*+0.017*	*−1.049*	*+0.035*	*+0.025*				
ERA-DS	Weekly	1	0.700	30.927	0.631	0.480	0.779	28.181	0.676	0.568	*+0.079*	*−2.746*	*+0.045*	*+0.088*
		2	0.706	30.623	0.628	0.490	0.775	28.378	0.672	0.562	*+0.069*	*−2.245*	*+0.044*	*+0.072*
		3	0.765	28.074	0.662	0.572	0.886	20.283	0.803	0.776	*+0.121*	*−7.791*	*+0.141*	*+0.204*
		4	0.750	28.685	0.657	0.553	0.892*	20.054*	0.816*	0.781*	*+0.142*	*-8.630*	*+0.159*	*+0.229*
	Difference (Weekly –Monthly)	1	*+0.131*	*−8.897*	*+0.071*	*+0.342*	*+0.162*	*−10.648*	*+0.142*	*+0.387*				
		2	*+0.106*	*−7.414*	*+0.068*	*+0.276*	*+0.146*	*−12.377*	*+0.157*	*+0.465*				
		3	*+0.088*	*−5.219*	*+0.057*	*+0.174*	*+0.056*	*−7.479*	*+0.133*	*+0.195*				
		4	*+0.084*	*−4.884*	*+0.065*	*+0.165*	*+0.052*	*−5.731*	*+0.126*	*+0.143*				

- Dataset 1: Tmean + Precipitation
- Dataset 2: Tmean + Precipitation + DTR
- Dataset 3: Tmean + Precipitation + DTR + RH + SI + WSmean
- Dataset 4: Tmean + Precipitation + DTR + RH + SI + WSmean + Tmax + Tmin

The best value of each evaluation metric in each data source is indicated with an asterisk (*)

The low resolution of the original ERA5 is not able to simulate the localized temperature climatology, especially in high-latitude areas. The cold bias and the warm bias are observed in the Tmax and Tmin simulations, respectively, which indicates that GCMs cannot properly simulate extreme temperature events. On the contrary, the high resolution of RCMs is well able to capture the spatial details of temperature patterns as well demonstrated in a number of previous studies (Im et al. 2021; Qiu et al. 2020). Especially, the downscaled ERA5 exhibits a clear advantage over the original ERA5 in simulating Tmax and Tmin. The comparison of the 40-year MJJAS average Tmax values of 56 in-situ stations between ERA-O and ERA-DS shows that dynamical downscaling is not only greatly effective in reducing the prevalent cold bias for Tmax in ERA-O simulations but also able to decrease the warm bias for Tmax and Tmin observed in one specific station. Not only temperature, but RH is also far more realistically simulated in the downscaled data, well following the region-specific characteristics such as high humidity in mountains and low humidity in basins. However, the benefit of dynamical downscaling is not observed in all meteorological variables. In the interannual variations of MJJAS average precipitation, SI, and WSmean during the 40 years, ERA-DS does not outperform ERA-O. Yet, dynamical downscaling has great meaning in that it gains ground on simulating temperature, especially DTR which is the most highly correlated with the rice yield variations in Korea.

Beyond the simulation of Korean climatology, the added value of dynamical downscaling is achieved in the prediction of rice yields in South Korea. Two machine learning algorithms, SVM and LSTM, are utilized to develop rice yield prediction models using OBS, ERA-O, and ERA-DS. The good performance of the OBS models proves the effectiveness of statistical yield prediction models and serves as the reference performance. The comparison with the ERA-O and ERA-DS models shows that the yield prediction ability of the model is increased with the temporal and spatial resolution of input climatological datasets. The use of weekly ERA-DS datasets in the prediction models consistently provides the best yield estimates in terms of all CORR, RMSE, IoA, and NSE, which highlights the added value of dynamical downscaling in yield prediction. Particularly, when the LSTM model is trained with weekly Tmean, Tmax, Tmin, precipitation, DTR, RH, SI, and WSmean simulated by ERA-DS, it leads to the best estimates. In this view, in the absence of long-term forecasts that are detailed enough for yield prediction in the present time, dynamical downscaling of the global seasonal forecasts (e.g., CFSv2) is expected to provide valuable meteorological input data. Therefore, future work will be focusing on utilizing the downscaled seasonal forecast data for yield prediction. In this regard, the need for accurate prediction of future crop yields will keep rising to give useful aid to the decision-making of policymakers and farmers (Peng et al. 2018). The comparison and evaluation of prediction models developed in this study based on dynamically downscaled weather data comprised of different weather parameter sets and timescales could ensure the ability of the model before the actual application of the model to predict future rice yields using downscaled forecasts.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00484-023-02544-x.

Acknowledgements This study was carried out with the support of the "Research Program for Agricultural Science & Technology Development (Project No. PJ014882)", National Institute of Agricultural Sciences, Rural Development Administration, Republic of Korea. We would thank Prof. Axel Garcia y Garcia and Dr. Vijaya Raj Joshi who provided us the R code used for developing and evaluating the statistical models presented in Joshi et al. (2020). We would also like to express our gratitude to Prof. Kwon Hyun-Han for discussing the interpretation of the statistical analysis with us.

Data Availability The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

Abrougui K, Gabsi K, Mercatoris B, Khemis C, Amami R, Chehaibi S (2019) Prediction of organic potato yield using tillage systems and soil properties by artificial neural network (ANN) and multiple linear regressions (MLR). Soil Till Res 190:202–208. https://doi.org/10.1016/j.still.2019.01.011

Alexandrov VA, Hoogenboom G (2000) The impact of climate variability and change on crop yield in Bulgaria. Agric For Meteorol 104(4):315–327. https://doi.org/10.1016/S0168-1923(00)00166-0

Alexandrov VA, Hoogenboom G (2001) Climate variation and crop production in Georgia, USA, during the twentieth century. Clim Res 17(1):33–43. https://doi.org/10.3354/cr017033

Auffhammer M, Ramanathan V, Vincent JR (2012) Climate change, the monsoon, and rice yield in India. Clim Chang 111(2):411–424. https://doi.org/10.1007/s10584-011-0208-4

Cao J, Zhang Z, Tao F, Zhang L, Luo Y, Zhang J et al (2021) Integrating multi-source data for rice yield prediction across china using machine learning and deep learning approaches. Agric For Meteorol 297:108275. https://doi.org/10.1016/j.agrformet.2020.108275

Chen H, Wu W, Liu HB (2016) Assessing the relative importance of climate variables to rice yield variation using support vector machines. Theor Appl Climatol 126(1):105–111. https://doi.org/10.1007/s00704-015-1559-y

Chen K, O'Leary RA, Evans FH (2019) A simple and parsimonious generalised additive model for predicting wheat yield in a decision support tool. Agric Syst 173:140–150. https://doi.org/10.1016/j.agsy.2019.02.009

Cho S, Youn Y, Kim S, Jeong Y, Kim G, Kang J et al (2021) A comparative evaluation of multiple meteorological datasets for the rice yield prediction at the county level in South Korea. Korean J Remote Sens 37(2):337–357. https://doi.org/10.7780/kjrs.2021.37.2.12

Clauss K, Ottinger M, Leinenkugel P, Kuenzer C (2018) Estimating rice production in the Mekong Delta, Vietnam, utilizing time series of Sentinel-1 SAR data. Int J Appl Earth Obs Geoinf 73:574–585

Crane-Droesch A (2018) Machine learning methods for crop yield prediction and climate change impact assessment in agriculture. Environ Res Lett 13(11):114003. https://doi.org/10.1088/1748-9326/aae159

Das B, Nair B, Reddy VK, Venkatesh P (2018) Evaluation of multiple linear, neural network and penalised regression models for prediction of rice yield based on weather parameters for west coast of India. Int J Biometeorol 62(10):1809–1822. https://doi.org/10.1007/s00484-018-1583-6

Gandhi N, Armstrong LJ, Petkar O, Tripathy AK (2016) Rice crop yield prediction in India using support vector machines. In 2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE) (pp. 1–5). IEEE. https://doi.org/10.1109/JCSSE.2016.7748856

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

Garg B, Beg MS, Ansari AQ (2013) Fuzzy time series model to forecast rice production. 2013 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), Hyderabad, India, 2013, pp. 1–8. https://doi.org/10.1109/FUZZIEEE.2013.6622509

Hersbach H, Bell B, Berrisford P, Hirahara S, Horányi A, Muñoz-Sabater J et al (2020) The ERA5 global reanalysis. Q J R Meteorol Soc 146(730):1999–2049. https://doi.org/10.1002/qj.3803

Hong SY, Hur J, Ahn JB, Lee JM, Min BK, Lee CK et al (2012) Estimating rice yield using MODIS NDVI and meteorological data in Korea. Korean J Remote Sens 28(5):509–520. https://doi.org/10.7780/kjrs.2012.28.5.4

Hu Q, Buyanovsky G (2003) Climate effects on corn yield in Missouri. J Appl Meteorol Climatol 42(11):1626–1635. https://doi.org/10.1175/1520-0450(2003)042%3c1626:CEOCYI%3e2.0.CO;2

Im ES, Ha S, Qiu L, Hur J, Jo S, Shim KM (2021) An evaluation of temperature-based agricultural indices over Korea from the high-resolution WRF simulation. Front Earth Sci 9:656787. https://doi.org/10.3389/feart.2021.656787

Jaikla R, Auephanwiriyakul S, Jintrawet A (2008) Rice yield prediction using a support vector regression method. In 2008 5th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (Vol. 1, pp. 29–32). IEEE. https://doi.org/10.1109/ECTICON.2008.4600365

Jeong O, Park HS, Baek MK, Kim WJ, Lee GM, Lee CM et al (2021) Review of rice in Korea: current status, future prospects, and comparisons with rice in other countries. J Crop Sci Biotechnol 24(1):1–11. https://doi.org/10.1007/s12892-020-00053-6

Jeong S, Ko J, Yeom JM (2018) Nationwide projection of rice yield using a crop model integrated with geostationary satellite imagery: a case study in South Korea. Remote Sens 10(10):1665. https://doi.org/10.3390/rs10101665

Ji B, Sun Y, Yang S, Wan J (2007) Artificial neural networks for rice yield prediction in mountainous regions. J Agric Sci 145(3):249–261. https://doi.org/10.1017/S0021859606006691

Joshi VR, Kazula MJ, Coulter JA, Naeve SL, Garcia y Garcia A (2021) In-season weather data provide reliable yield estimates of maize and soybean in the US central Corn Belt. Int J Biometeorol 65(4):489–502. https://doi.org/10.1007/s00484-020-02039-z

Kashiwagi T, Ishimaru K (2004) Identification and functional analysis of a locus for improvement of lodging resistance in rice. Plant Physiol 134(2):676–683. https://doi.org/10.1104/pp.103.029355

Kaul M, Hill RL, Walthall C (2005) Artificial neural networks for corn and soybean yield prediction. Agric Syst 85(1):1–18. https://doi.org/10.1016/j.agsy.2004.07.009

Kittichotsatsawat Y, Tippayawong N, Tippayawong KY (2022) Prediction of arabica coffee production using artificial neural network and multiple linear regression techniques. Sci Rep 12(1):1–14. https://doi.org/10.1038/s41598-022-18635-5

Kogan F, Kussul NN, Adamenko TI, Skakun SV, Kravchenko AN, Krivobok AA et al (2013) Winter wheat yield forecasting: A comparative analysis of results of regression and biophysical models. J Autom Inf Sci 45(6):68–81. https://doi.org/10.1615/JAutomatInfScien.v45.i6.70

Kouadio L, Byrareddy VM, Sawadogo A, Newlands NK (2021) Probabilistic yield forecasting of robusta coffee at the farm scale using agroclimatic and remote sensing derived indices. Agric for Meteorol 306:108449

Kouadio L, Newlands NK, Davidson A, Zhang Y, Chipanshi A (2014) Assessing the performance of MODIS NDVI and EVI for seasonal crop yield forecasting at the ecodistrict scale. Remote Sens 6(10):10193–10214

Lobell DB (2007) Changes in diurnal temperature range and national cereal yields. Agric for Meteorol 145(3–4):229–238. https://doi.org/10.1016/j.agrformet.2007.05.002

Lobell DB, Asseng S (2017) Comparing estimates of climate change impacts from process-based and statistical crop models. Environ Res Lett 12(1):015001. https://doi.org/10.1088/1748-9326/aa518a

Lobell DB, Burke MB (2010) On the use of statistical models to predict crop yield responses to climate change. Agric For Meteorol 150(11):1443–1452. https://doi.org/10.1016/j.agrformet.2010.07.008

Mathieu J, Aires F (2016) Statistical weather-impact models: an application of neural networks and mixed effects for corn production over the United States. J Appl Meteorol Climatol 55(11):2509–2527. https://doi.org/10.1175/JAMC-D-16-0055.1

Matsumura K, Gaitan CF, Sugimoto K, Cannon AJ, Hsieh WW (2015) Maize yield forecasting by linear regression and artificial neural networks in Jilin. China J Agric Sci 153(3):399–410. https://doi.org/10.1017/S0021859614000392

Ministry of Agriculture, Food and Rural Affairs (MAFRA), Korea Rural Community Corporation (KRC) (2022) Statistical yearbook of land and water development for agriculture, South Korea

Mkhabela MS, Bullock P, Raj S, Wang S, Yang Y (2011) Crop yield forecasting on the Canadian Prairies using MODIS NDVI data. Agric For Meteorol 151(3):385–393

Morell FJ, Yang HS, Cassman KG, Van Wart J, Elmore RW, Licht M et al (2016) Can crop simulation models be used to predict local to regional maize yields and total production in the US Corn Belt? Field Crop Res 192:1–12. https://doi.org/10.1016/j.fcr.2016.04.004

Moriasi DN, Arnold JG, Van Liew MW, Bingner RL, Harmel RD, Veith TL (2007) Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Trans ASABE 50(3):885–900. https://doi.org/10.13031/2013.23153

Na SI, Hong SY, Kim YH, Lee KD, Jang SY (2013) Prediction of rice yield in Korea using paddy rice NPP index-application of MODIS data and CASA model. Korean J Remote Sens 29(5):461–476. https://doi.org/10.7780/kjrs.2013.29.5.2

Na SI, Park JH, Park JK (2012) Development of Korean paddy rice yield prediction model (KRPM) using meteorological element and MODIS NDVI. J Korean Soc Agric Eng 54(3):141–148. https://doi.org/10.5389/KSAE.2012.54.3.141

Oguntunde PG, Lischeid G, Dietrich O (2018) Relationship between rice yield and climate variables in southwest Nigeria using multiple linear regression and support vector machine analysis. Int J Biometeorol 62(3):459–469. https://doi.org/10.1007/s00484-017-1454-6

Onwuchekwa-Henry CB, Ogtrop FV, Roche R, Tan DK (2022) Model for Predicting Rice Yield from Reflectance Index and Weather Variables in Lowland Rice Fields. Agriculture 12(2):130. https://doi.org/10.3390/agriculture12020130

Oses N, Azpiroz I, Marchi S, Guidotti D, Quartulli M, Olaizola GI (2020) Analysis of copernicus' era5 climate reanalysis data as a replacement for weather station temperature measurements in machine learning models for olive phenology phase prediction. Sensors 20(21):6381. https://doi.org/10.3390/s20216381

Oyoshi K, Tomiyama N, Okumura T, Sobue S, Sato J (2016) Mapping rice-planted areas using time-series synthetic aperture radar data for the Asia-RiCE activity. Paddy Water Environ, 14:463–472

Peng B, Guan K, Pan M, Li Y (2018) Benefits of seasonal climate prediction and satellite data for forecasting US maize yield. Geophys Res Lett 45(18):9662–9671. https://doi.org/10.1029/2018GL079291

Prabhjyot-Kaur SSS, Dhillo BS, Sing H (2021) Rice yield variability in Punjab: an overview of five decades. Paddy and Water Environ 19(4):673–681. https://doi.org/10.1007/s10333-021-00866-3

Qiu L, Im ES, Hur J, Shim KM (2020) Added value of very high resolution climate simulations over South Korea using WRF modeling system. Clim Dyn 54(1):173–189. https://doi.org/10.1007/s00382-019-04992-x

Rajkumara S (2008) Lodging in cereals-a review. Agric Rev 29(1):55–60

Ray DK, Gerber JS, MacDonald GK, West PC (2015) Climate variation explains a third of global crop yield variability. Nat Commun 6(1):1–9. https://doi.org/10.1038/ncomms6989

Rodríguez-Puebla C, Ayuso SM, Frias MD, Garcia-Casado LA (2007) Effects of climate variation on winter cereal production in Spain. Clim Res 34(3):223–232. https://doi.org/10.3354/cr00700

Su YX, Xu H, Yan LJ (2017) Support vector machine-based open crop model (SBOCM): Case of rice production in China. Saudi J Biol Sci 24(3):537–547. https://doi.org/10.1016/j.sjbs.2017.01.024

Sun J, Di L, Sun Z, Shen Y, Lai Z (2019) County-level soybean yield prediction using deep CNN-LSTM model. Sensors 19(20):4363. https://doi.org/10.3390/s19204363

Tack J, Barkley A, Nalley LL (2015) Effect of warming temperatures on US wheat yields. Proc Natl Acad Sci 112(22):6931–6936. https://doi.org/10.1073/pnas.1415181112

Tao F, Zhang Z, Xiao D, Zhang S, Rötter RP, Shi W et al (2014) Responses of wheat growth and yield to climate change in different climate zones of China, 1981–2009. Agric for Meteorol 189:91–104. https://doi.org/10.1016/j.agrformet.2014.01.013

Teasdale JR, Cavigelli MA (2017) Meteorological fluctuations define long-term crop yield patterns in conventional and organic production systems. Sci Rep 7(1):1–10. https://doi.org/10.1038/s41598-017-00775-8

Tian H, Wang P, Tansey K, Zhang J, Zhang S, Li H (2021) An LSTM neural network for improving wheat yield estimates by integrating remote sensing data and meteorological data in the Guanzhong Plain. PR China. Agric For Meteorol 310:108629. https://doi.org/10.1016/j.agrformet.2021.108629

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

Van Wart J, Kersebaum KC, Peng S, Milner M, Cassman KG (2013) Estimating crop yield potential at regional to national scales. Field Crop Res 143:34–43. https://doi.org/10.1016/j.fcr.2012.11.018

Verón SR, De Abelleyra D, Lobell DB (2015) Impacts of precipitation and temperature on crop yields in the Pampas. Clim Chang 130(2):235–245. https://doi.org/10.1007/s10584-015-1350-1

Yun JI (2003) Predicting regional rice production in South Korea using spatial data and crop-growth modeling. Agric Syst 77(1):23–38. https://doi.org/10.1016/S0308-521X(02)00084-7

Zaefizadeh M, Jalili A, Khayatnezhad M, Gholamin R, Mokhtari T (2011) Comparison of multiple linear regressions (MLR) and artificial neural network (ANN) in predicting the yield using its components in the hulless barley. Adv Environ Biol 109–114

Zeybek M (2018) Nash-sutcliffe efficiency approach for quality improvement. J Applied Math Comput 2(11):496–503. https://doi.org/10.26855/jamc.2018.11.001

Zhang T, Zhu J, Wassmann R (2010) Responses of rice yields to recent climate change in China: an empirical assessment based on long-term observations at different spatial scales (1981–2005). Agric for Meteorol 150(7–8):1128–1137. https://doi.org/10.1016/j.agrformet.2010.04.013