Global Within-Season Yield Anomaly Prediction for Major Crops Derived Using Seasonal Forecasts of Large-Scale Climate Indices and Regional Temperature and Precipitation

Toshichika Iizumi, a Yuhei Takaya, b,c Wonsik Kim, a Toshiyuki NakaeGawa, b and Shuhei Maeda d

a Institute for Agro-Environmental Sciences, National Agriculture and Food Research Organization, Tsukuba, Japan
b Meteorological Research Institute, Japan Meteorological Agency, Tsukuba, Japan
c Global Environment and Marine Department, Japan Meteorological Agency, Tokyo, Japan
d Aerological Observatory, Japan Meteorological Agency, Tsukuba, Japan

(Manuscript received 15 June 2020, in final form 9 December 2020)

ABSTRACT: Weather and climate variability associated with major climate modes is a main driver of interannual yield variability of commodity crops in global cropland areas. A global crop forecasting service that is currently in the test operation phase is based on temperature and precipitation forecasts, while recent literature suggests that crop forecasting services may benefit from the use of climate index forecasts. However, no consistent comparison is available on prediction skill between yield models relying on forecasts from temperature and precipitation and from climate indices. Here, we present a global assessment of 26-yr (1983–2008) within-season yield anomaly hindcasts for maize, rice, wheat, and soybean derived using different types of statistical yield models. One type of model utilizes temperature and precipitation for individual cropping areas (the TP model type) to represent the current service, whereas the other type relies on large-scale climate indices (the CI model). For the TP models, three specifications with different model complexities are compared. The results show that the CI model is characterized by a small reduction in the skillful area from the reanalysis model to the hindcast model and shows the largest skillful areas for rice and soybean. In the TP models, the skill of the simple model is comparable to that of the more complex models. Our findings suggest that the use of climate index forecasts for global crop forecasting services in addition to temperature and precipitation forecasts likely increases the total number of crops and countries where skillful yield anomaly prediction is feasible.

KEYWORDS: Hindcasts; Seasonal forecasting; Agriculture; Climate services; Crop growth

1. Introduction

Oceanic and atmospheric variability and associated variations in regional temperature and moisture conditions during the growing season lead to departures of crop yield from the long-term trend (Sivakumar and Hansen 2007; Rosenzweig and Hillel 2008; Heino et al. 2018; Anderson et al. 2019). Among others, El Niño–Southern Oscillation (ENSO) is the dominant mode of interannual climate variability affecting global crop production, with positive or negative yield impacts depending on the life cycle of ENSO, crop type, location of the cropping areas, timing of the growing season, and agronomic management (Cane et al. 1994; Stone et al. 1996; Iizumi et al. 2014a; Anderson et al. 2017, 2019; Heino et al. 2018, 2020). The Indian Ocean dipole (IOD), North Atlantic Oscillation (NAO), Eastern Atlantic pattern (EA), Eastern Atlantic/western Russia pattern (EAWR), and, to a lesser extent, El Niño Modoki (EM), tropical Atlantic variability (TAV), Madden–Julian oscillation (MJO), and Pacific–North American pattern (PNA) are also known to have production impacts for particular cropping regions of the world (Porter and Semenov 2005; Rosenzweig and Hillel 2008; Yuan and Yamagata 2015; Ceglar et al. 2017; Heino et al. 2018, 2020; Zambrano et al. 2018; Nobre et al. 2019; Anderson et al. 2019, 2020; Najafi et al. 2020).

Currently, major weather centers provide operational seasonal climate forecasts on a routine basis [although not an exhaustive list, these include Graham et al. (2011) and Kim et al. (2016, 2021)]. State-of-the-art atmosphere–ocean coupled general circulation models (AOGCMs) used for operational seasonal climate forecasting present meaningful prediction skill not only in the tropics but also in the extratropics due to remote influences of tropical oceanic states through teleconnections (Stan et al. 2017; Smith et al. 2012). However, due to the inherent limitation of predictability and misrepresentation of teleconnections in climate models, the skill in predicting extratropical regional atmospheric variables (e.g., temperature and precipitation) or circulation patterns is generally lower than the skill in predicting climate indices tailored for representing major climate modes, such as ENSO and other tropical oceanic variability (Jin et al. 2008; Barnston et al. 2012, 2019; Scaife et al. 2019; Shi et al. 2012; Struzzo et al. 2019; Takaya et al. 2018; Wang et al. 2009).

It has been well documented that the prediction skill of seasonal temperature and moisture forecasts largely varies by geographic region, season and climatic variable (Wilks and Godfrey 2002; NakaeGawa et al. 2003; Wang et al. 2009). This varying prediction skill sometimes makes agricultural application of temperature and moisture forecasts for specific regions challenging (Semenov and Doblas-Reyes 2007; Hayashi et al. 2018). Consequently, possible agricultural applications of
climate index forecasts, in addition to temperature and moisture forecasts, have been explored (Yuan and Yamagata 2015; Ceglar et al. 2018; Nobre et al. 2019). For instance, international organizations, such as the Food and Agricultural Organization (FAO) of the United Nations, have utilized ENSO forecasts to help food-insecure countries and food security experts take early actions against possible production shocks (FAO 2016). Another example of such application is the mapping of global yield impacts associated with an anticipated ENSO event found in a report released in July 2014 by the Ministry of Agriculture, Forestry and Fisheries of Japan (Iizumi and Kim 2019).

Yield anomaly predictions relying on climate index forecasts can be accurate, as already demonstrated in part in a few earlier regional studies (Yuan and Yamagata 2015; Ceglar et al. 2018; Nobre et al. 2019). However, a global overview of the skill of climate index-based yield anomaly predictions for major commodity crops has not yet been presented. No consistent comparison is available between statistical yield models relying on temperature and precipitation and those relying on climate indices. Yield models applicable to global crop forecasting depend on temperature and moisture forecasts (e.g., Iizumi et al. 2013, 2018b). This is the case for the National Agriculture and Food Research Organization and Asia–Pacific Economic Cooperation Climate Center (NARO-APCC) Joint Crop Forecasting Service, which is in the test operation phase from June 2019 to March 2021 (Iizumi and Kim 2019; Iizumi 2020). The underlying yield model of the service is based on temperature and precipitation forecasts from the APCC multimodel ensemble (Min et al. 2014; Iizumi et al. 2018b).

In this article, we present an evaluation and intercomparison of 26-yr (1983–2008) within-season yield anomaly hindcasts for maize, rice, wheat and soybean. The hindcasts are generated three months before the harvesting of the crop of interest and referred to as the within-season prediction. The findings derived from this study are expected to provide a unique input to the knowledge base, which will ultimately improve global and regional crop forecasting services.

2. Materials and methods

The data flow used in this study is shown in Fig. 1. The yield data used as the predictand and climate data used as the

FIG. 1. Data flow used in this study to estimate the statistical yield models and subsequent analyses.
predictors are described in section 2a and section 2b, respectively. The model structures are described in section 2c; with details for the model type that utilizes temperature and precipitation as predictors [the TP model type; section 2c(1)] and the climate indices–based model [the CI model; section 2c(2)]. Using the cross-validated models and regression coefficient values, we conducted a crop hindcast experiment (section 2d) and assessed the skill score (section 2e) as well as the key limiting factor of forecast skill (section 2f) and dominant climate modes for yields (section 2g).

a. Crop yields

The 0.5°-resolution 30-yr annual yields for the four crops from 1981 to 2010 were obtained from the global dataset of historical yields version 1.2 (Iizumi et al. 2014c, 2018a). This dataset is a hybrid of satellite-based net primary production and FAO country-level yield statistics. Although the gridcell yields are model estimates, earlier analyses, including comparisons with subnational yield statistics, indicate that the dataset is a valuable source of information on global yields for analyzing climate–yield relationships (Iizumi et al. 2014c, 2018a; Müller et al. 2017; Schauburger et al. 2017; Iizumi and Sakai 2020). In this study, we treated the data as a representation of the actual yields.

Using the yield data, the yield anomaly is computed as follows:

$$\Delta Y_{t-g,s} = \frac{Y_{t-g,s} - Y_{t-2+2+g,s}}{Y_{t-2+2+g,s}} \times 100,$$

where the subscript $t$ is the harvesting year, $g$ is the grid cell, $s$ is the cropping season of the crop, $\Delta Y$ is the yield anomaly (% of normal yield), and $Y$ is the yield (t ha$^{-1}$). We assume that the 5-yr moving average for the period from $t - 2$ to $t + 2$ ($Y_{t-2+2+2}$) represents the normal yield for year $t$. Two cropping seasons are available for maize, rice, and wheat (winter and spring seasons for wheat and major and second seasons for the remaining two crops). Soybean has a single cropping season. These distinctions in cropping season follow the global crop calendars used in this study (Sacks et al. 2010).

In many crops and countries, yields have an increasing trend, predominantly due to improvements in agronomic technology and management. Therefore, the annual time series of the yield anomaly calculated relative to the long-term average would have many negative values in earlier years of the study period and many positive values in later years; this is not conducive to analyzing the relationships between climate conditions and yield. As departures of yield from the trend line (i.e., yield anomaly) occur mainly due to variations in growing season climate conditions, the removal of the trend line is a necessary preprocessing step. However, yield trend patterns are often nonlinear (Grassini et al. 2013; Lu et al. 2017). Among the nonlinear detrending methods, the moving average is simple and often utilized (e.g., Iizumi et al. 2014a; Lu et al. 2017) alongside a Gaussian filter (Anderson et al. 2017, 2019, 2020) and the locally weighted scatterplot smoother (LOESS) (Ceglar et al. 2017). The use of a longer time window in the moving average leads to a more robust estimate of the yield trend, but it decreases the number of effective yield anomaly samples. We used a 5-yr interval to keep as many effective samples as possible while removing trends. Importantly, climate–yield relationship analysis results, including the estimated yield impacts of relatively long-lasting climate modes, such as ENSO, are sensitive to the length of the time window in the moving average in a quantitative manner but less sensitive in a qualitative manner (Iizumi et al. 2014a). In this study, the crop forecast skill in distinguishing between yield loss and yield gain is of primary interest (see section 2e for details). This justifies the adoption of the 5-yr moving average method for this study.

b. Temperature, precipitation, and climate indices

Daily mean 2-m air temperature and precipitation calculated from the Japanese 55-year Reanalysis (JRA-55; Kobayashi et al. 2015; Harada et al. 2016) 3-hourly data were utilized to represent the actual climate conditions. Reanalysis data are not observations but outputs of a climate model constrained by using a data assimilation technique. Thus, reanalysis data are often treated as actual climate conditions in climate–yield analysis (Challinor et al. 2005; Iizumi et al. 2014b; Toreti et al. 2019). Reanalysis daily data spanning from 1958 to the present were interpolated onto a 0.5° regular grid from the original grid size of 0.5625° using the inverse distance weighted averaging method to be consistent with the yield data.

We also used seasonal temperature and precipitation hindcasts from 1981 to 2014 derived from the Japan Meteorological Agency/Meteorological Research Institute-Coupled Prediction System version 2 (JMA/MRI-CPS2; Takaya et al. 2018). A 5-member ensemble of daily mean temperature and precipitation hindcasts for a 217-day interval after the initialized date is available, with an initialization frequency of twice per month. Hindcasts with the initialization date that is the closest to the beginning of the key season of a crop of interest [see section 2c(1)(i) for details] were used. Although the AOGCM used in the JMA/MRI-CPS2 system has a grid size of 1°, climate model outputs are aggregated to a 2.5° resolution (Takaya et al. 2018). The hindcasts were reinterpolated to the 0.5° resolution in the same manner as the reanalysis data. No elevation correction was conducted for either the reanalysis or hindcast data, as this procedure does not alter temperature and precipitation anomalies.

Monthly data for multiple climate indices, including the DMI, EA500, EMI, NAO, NINO34, and PNA500 (see Table 1 for these abbreviations), were calculated from the ensemble mean of the JMA/MRI-CPS2 hindcasts for this study. Extratropical teleconnection indices (EA500, NAO, and PNA500) were calculated based on a rotated empirical orthogonal function (REOF) analysis using a monthly 500-hPa height hindcast. There are many more indices than considered here. However, we limited our analysis to these six indices because their yield impacts are likely evident from the literature, in particular the impacts from the IOD, EA, NAO and ENSO and, to a lesser extent, PNA (Porter and Semenov 2005; Rosemberg and Hillel 2008; Iizumi et al. 2014a; Yuan and Yamagata 2015; Ceglar et al. 2017; Heino et al. 2018, 2020;
DMI Indian Ocean dipole mode index—This index is defined by an anomalous sea surface temperature (SST) gradient between the western (10°S–10°N, 50°–70°E) and southeastern equatorial Indian Ocean (10°S–0°N, 90°–110°E) (Saji and Yamagata 2003)

EA500 Eastern Atlantic pattern of the 500-hPa height (Wallace and Gutzler 1981)

EMI El Niño Modoki index—This index is computed by combining area-averaged SST anomalies over three regions: A (10°S–10°N, 165°E–140°W), B (15°S–5°N, 110°–70°W), and C (10°S–20°N, 125°–145°E) (Ashok et al. 2007)

NAO North Atlantic Oscillation (Barnston and Livezey 1987)

NINO34 El Niño SST index for region 3.4 in the eastern central tropical Pacific (5°N–5°S, 170°–120°W)

PNA500 Pacific–North American pattern of the 500-hPa height (Wallace and Gutzler 1981)

Zambrano et al. 2018; Anderson et al. 2019, 2020; Nobre et al. 2019; Najafi et al. 2020). The reanalysis-based climate indices were also used and treated as actual climate conditions in this study.

c. Statistical yield models

Two different model types were considered here. One type utilizes temperature and precipitation as predictors (the TP model type). This type has three specifications with different model complexities, as described later. The other type uses climate indices as predictors and has a single specification (the CI model). One specification of the TP model type is close to the model used in the global crop forecasting service. However, the yield anomaly used in the service is differently defined from the model used in the global crop forecasting service. However, the consideration of these terms increases the number of predictors. Therefore, we first detrended the yield data and then associated the yield anomaly with climate variables. Such a two-step approach can be seen in Ceglar et al. (2017).

According to Iizumi et al. (2018b), we adopted the leave-three-out cross-validation technique to avoid the possible influences of adjacent years in the temporarily correlated time series. This method indicates that the samples for years \( t - 1, t, \) and \( t + 1 \) were removed when the model for year \( t \) was estimated. The model estimation and variable selection were conducted for each year, crop, cropping season of the crop and grid cell using the ordinary least squares approach and the Akaike information criterion (AIC) available in R version 3.5.0 [“lm” and “stepAIC” functions; R Core Team (2018)].

For maize, rice and wheat, the modeled yield anomalies from multiple cropping seasons of the crop are combined by

\[
Y_{t,g} = \sum_{s=1}^{2} \frac{w_{g,s} Y_{t,s}}{\sum_{s=1}^{2} w_{g,s}},
\]

where \( w \) is the production volume of the crop in tons yielded from growing season \( s \). The average country-wide production volumes in the 1990s obtained from a report of the U.S. Department of Agriculture (USDA 1994) were used throughout the study period. Production volumes from different cropping seasons have changed with time, as shown in Anderson et al. (2017) for early- and short-season maize in Brazil, but historical data are rarely accessible at the global scale.

While climate conditions for the presowing season and the vegetative growth stage (emergence to flowering) are also important, those for the reproductive growth stage (flowering to harvesting) are more directly influential on yield formation than those for these earlier periods. However, it is difficult to precisely determine the crop- and location-specific time period in which the reproductive growth stage occurs for global croplands, because the crop duration from sowing to harvesting largely varies by crop and location and spans from 90 to 180 days (Brouwer and Heibloem 1986). Therefore, we simply assumed the 90-day interval just before harvesting as the key season for associating climate conditions with yield based on previous work (Iizumi et al. 2013, 2014a, 2018b).
(ii) Intermediate specification

We further considered two more specifications of the TP model type. For both, the normalization of the response variable and predictors, leave-three-out cross validation, and the adoption of parsimonious models via variable selection were commonly practiced.

The second specification of the TP model type (the $T^2 + T + P^2 + P$ model) is

$$\Delta Y_{t,g,s} = \beta_{0g,s} + \beta_{1g,s} T^2_{t,g,s} + \beta_{2g,s} T_{t,g,s} + \beta_{3g,s} P^2_{t,g,s} + \beta_{4g,s} P_{t,g,s} + \epsilon,$$

which includes squared terms of the temperature and precipitation anomalies of the key season; $\beta_{1-4}$ are the regression coefficients. Although $T^2$ and $P^2$ are considered in addition to $T$ and $P$, all the predictors considered here are derived from monthly climate data. Therefore, this specification is classified as a model with intermediate complexity.

(iii) Complex specification

The third specification of the TP model type (the GDD + EDD + $P$ model) is

$$\Delta Y_{t,g,s} = \gamma_{0g,s} + \gamma_{1g,s} \text{GDD}_{t,g,s} + \gamma_{2g,s} \text{EDD}_{t,g,s} + \gamma_{3g,s} P_{t,g,s} + \epsilon,$$

where GDD and EDD are the key season growing degree days and extreme degree days, respectively; and $\gamma_{0,3}$ are the regression coefficients. This specification is relatively complex because daily climate data were used to compute the GDD and EDD. The following approaches were used to compute these daily temperature indices:

$$\text{GDD} = \sum_{t=d_{i-90}}^{d_{i}} \max\{0, \min(T_i, T_u) - T_0\},$$

$$\text{EDD} = \sum_{t=d_{i-90}}^{d_{i}} \max\{0, (T_i - T_u)\},$$

where $d_{i}$ is the harvesting day, $T_i$ is the daily mean temperature for the $i$th day ($^\circ\text{C}$), and $T_0$ and $T_u$ are the lower and upper temperature limits ($^\circ\text{C}$), respectively. The lower limits used here were 8°C for maize and rice, 0°C for wheat, and 10°C for soybean (Deryng et al. 2011; van Oort et al. 2011), whereas a common value of 30°C was set for the upper limit across all the crops (Lobell et al. 2012; Parkes et al. 2019). Different upper limit values are found in Roberts et al. (2017), but we consider a single value for this study, mainly because of computational constraints. There is another approach for the GDD and EDD calculation that considers the within-day distribution of temperatures by fitting a sinusoidal curve between daily maximum and minimum temperatures (e.g., Parkes et al. 2019): we do not adopt this approach because these temperature variables are not available in many operational seasonal climate forecasts.

2) The CI model

Finally, the model relying on climate indices considered here (the CI model) is

$$\Delta Y_{t,g,s} = \theta_{0g,s} + \theta_{1g,s} \text{DMI}_{t,s} + \theta_{2g,s} \text{EA500}_{t,s} + \theta_{3g,s} \text{EMI}_{t,s} + \theta_{4g,s} \text{NAO}_{t,s} + \theta_{5g,s} \text{NINO34}_{t,s} + \theta_{6g,s} \text{PNA500}_{t,s} + \epsilon,$$

where DMI, EA500, EMI, NAO, NINO34, and PNA500 are the anomalies in the key season average climate indices and $\theta_{0-6}$ are the regression coefficients. The season average values were computed from the monthly data using the number of days within a month overlapping the season as the weights. The computed season average anomalies were normalized and used as the predictors. Unlike any specifications of the TP models, the CI model associates yield anomalies for individual locations with the large-scale climate indices; thus, there is no subscript $g$ in the predictors [Eq. (8)].

It is well known that some climate indices are highly correlated with each other. The strongest Pearson’s correlation ($-0.492; p < 0.001$) is found between the reanalysis NAO and PNA500 monthly data for the 1980–2014 period ($n = 418$). The ordinary least squares method is not appropriate in this condition because of the multicollinearity between the predictors. We adopt a regularized and variable selection method called the elastic net (Zou and Hastie 2005). The elastic net is more useful than earlier regularized regressions, such as the lasso (Tibshirani 1996) and ridge regressions (Hoerl et al. 1975; Lawless and Wang 1976), when the number of predictors is much larger than the number of samples or in any situation with many correlated predictors (Friedman et al. 2010). In our data, the latter situation applies. We used the elastic net function available in R (the “cv.glmnet” function with the mixing parameter $\alpha = 0.5$). Leave-three-out cross validation was also carried out for the CI model. The variable selection was objectively conducted with the elastic net.

d. Crop hindcast experiment

To provide within-season yield anomaly hindcasts, all the yield models considered here were first established using the actual yield anomaly as the response variable and the reanalysis temperature and precipitation or reanalysis-based climate indices as the predictors. Then, we input the hindcast ensemble season average anomalies in temperature and precipitation or climate indices to the models to derive the yield anomalies of the crops for the 1983–2008 period. This method is an analog of the perfect-prog approach (Wilks 2006). The modeled yield anomalies were computed for 0.5° grid cells and aggregated into country averages using the harvested area in 2000 (Monfreda et al. 2008) as the weights. We avoided directly associating climate hindcasts with actual yield anomalies, which is close to the model output statistics (MOS) approach (Wilks 2006), because it makes the interpretation of results difficult if climate hindcasts with large prediction errors lead to accurate yield anomalies. The perfect-prog approach used is more intuitive than the MOS approach because the yield anomaly prediction skill increases as climate forecasts become more accurate. Furthermore, the yield models used here predict yield anomalies, and the forecast skill in distinguishing yield loss and yield gain is assessed in this study. This enables us to avoid conducting bias correction of climate forecasts,
because in general, bias correction does not alter the signs of temperature and precipitation anomalies.

e. Skill score

The yield anomaly hindcast presented here is deterministic and serves as a binary classification tool for distinguishing between yield loss ($\Delta Y < 0$) and yield gain ($\Delta Y \geq 0$). The underlying premise is that for a given crop and location, it is reasonable to anticipate the incidence of yield loss for a harvesting year of interest if a negative yield anomaly value is predicted. Similarly, yield gain is a reasonable expectation if a positive yield anomaly value is predicted. Given these characteristics of yield anomaly prediction, the relative operating characteristic (ROC) score (Wilks 2006) is an appropriate measure of prediction skill. ROC scores equal to or greater than 0.6 indicate skill that is better than random chance, whereas those between 0.5 and 0.6 are associated with predictions that are slightly better than random chance (Met Office 2013). In Iizumi et al. (2018b), ROC scores are separately computed for yield loss and yield gain, and their average is used as an indicator of overall prediction skill. However, the ROC scores are almost the same between yield loss and yield gain. Therefore, ROC scores were computed for only yield loss in this study. The significance of the ROC scores was tested using 500 bootstrap replications, with a null hypothesis that the ROC score of the yield model in the reanalysis mode computed between the actual and reanalysis-based yield anomalies for the 26-yr period (ROC$_{\Delta Y\_\text{Reanalysis}}$) (for instance, see Fig. 2 for maize). The minimum ROC score among the climate variables within the yield model after variable selection was used for the latter. Taking the $T + P$ model as an example, a single climate variable, either $T$ or $P$, for which the ROC score computed between the reanalysis and hindcast of the climate variable for the study period (Figs. S1 and S2 in the online supplemental material) was the lowest, was selected (ROC$_i$, $i$ indicates $T$ or $P$). Then, a single factor, either ROC$_{\Delta Y\_\text{Reanalysis}}$ or ROC$_i$, the value of which was lower than the other, was selected as the key skill-limiting factor of the within-season hindcast (ROC$_{\text{low}}$). The skill-limiting factors for the other three models were addressed in the same manner as for the $T + P$ model.

f. Determination of dominant modes of climate variability

The climate index type that most negatively affects yield in terms of amplitude and frequency was determined for each crop and location. First, for each year, location and cropping season of the crop, we selected the index type with the most negative CI model regression coefficient value: $j_{i,g,s} = \max(|\theta_{i,g,s,t}|; i = 1, \ldots, 6)$, where $j$ indicates the most influential index type to yield loss among the six indices, which is indicated by $i$ ($1 = \text{DMI}$, $2 = \text{EA500}$, $3 = \text{EMI}$, $4 = \text{NAO}$, $5 = \text{NINO34}$, and $6 = \text{PNA500}$). Second, we counted the number of cases in which the $i$th index type was the most influential on yield loss using samples within the years and cropping seasons:

$$\#(j_{s,i,g}) = \{\theta_{i,g,s,t} = 1983, \ldots, 2008, s = 1, \ldots, n\}$$

where $j$ indicates the most influential index type to yield loss among the six indices, which is indicated by $i$ ($1 = \text{DMI}$, $2 = \text{EA500}$, $3 = \text{EMI}$, $4 = \text{NAO}$, $5 = \text{NINO34}$, and $6 = \text{PNA500}$).
where \( \#(i) \) is the number of cases for the \( i \)th index type and \( n \) is the number of cropping seasons (\( 1 = \) soybean and \( 2 = \) maize, rice, and wheat). Finally, we specified the single index type that was identified most frequently as the most influential on yield loss:

\[
k_g = \max\{\#(i_1), i = 1, \ldots, 6\},
\]

(11)

where \( k \) indicates the dominant index type for yield loss.

For yield gain, Eq. (9) was modified to

\[
j_{g,s} = \max\{\theta_{i,g,s}; i = 1, \ldots, 6\} \text{ if } \theta_{i,g,s} > 0,
\]

(12)

where \( j \) indicates the most influential index type for yield gain. This modification was applied to select a single index type that most positively affects yield. The procedure thereafter [Eqs. (10) and (11)] was the same between yield gain and yield loss.

Although the individual climate modes can fluctuate separately or simultaneously, the key growing season differs by location, and the impacts on yield are also different across the climate index types. This helps us identify the single index type that is most negatively (or positively) influential on crop yield in a statistical sense. The use of leave-three-out cross validation also contributed to identifying the most influential climate index type more robustly, as it increases the sample size of the regression coefficient values.

3. Results

a. The upper limits of the prediction skill

The upper limits of prediction skill in terms of the number of grid cells with significant ROC scores computed between the actual and reanalysis-based yield anomalies vary among the yield models. Taking maize as an example for explanatory purposes, as Figs. 2a–d shows, the number of significant grid cells was highest for the \( T + P \) model (2595) and closely followed by the GDD + EDD + P model (2506). The results for the \( T^2 + T + P^2 + P \) model (1490) and the CI model (1094) were only 57\% and 42\% of the \( T + P \) model result, respectively. These numbers are summarized in the subpanel labeled “Reanalysis” in Fig. 3a.
The number of significant grid cells is a simple indicator of the skillful area extent, that is, the ratio of the harvested area with a significant ROC score to the global harvested area. For maize, the order of the models based on the number of significant grid cells (from high to low) was as follows: $T + P$, GDD + EDD + $P$, $T^2 + T + P^2 + P$, CI. This order was close to the order of the skillful area extent in the reanalysis mode (16.4% for the $T + P$ and GDD + EDD + $P$ models, 9.2% for the $T^2 + T + P^2 + P$ model and 4.5% for the CI model; see the x axis of Fig. 4a). The order of the models was common across maize, wheat, and soybean and across the number of significant grid cells and the skillful area extent (see the Reanalysis subpanel of Figs. 3a, 3b and 3d and the x axis of Figs. 4a, 4b and 4d).

The exception to this pattern was rice. The numbers of significant grid cells for the CI (1396), $T + P$ (1397), and GDD + EDD + $P$ (1390) models were similar, followed by that of the $T^2 + T + P^2 + P$ model (881) (see the Reanalysis subpanel of Fig. 3c). Moreover, the order of the models based on the number of significant grid cells and the skillful area extent was different for rice than for the other crops. The GDD + EDD + $P$ model ranked first (12.6%) and was followed by the $T + P$ (11.4%), CI (7.7%), and $T^2 + T + P^2 + P$ (7.2%) models (see the x axis of Fig. 4c).

b. Within-season prediction skill

The subpanel labeled “Hindcast” in Fig. 3 presents the ROC scores computed between the actual and hindcast yield anomalies. As expected, the average ROC score decreased when the hindcast mode of the models was assessed. Interestingly, for all the crops considered here, the decrease in the average ROC score was much smaller for the CI model than for any specifications of the TP model type (see the subpanels labeled Hindcast minus Reanalysis in Figs. 3a–d). Indeed, for the CI model, the average ROC score decreased...
only slightly, from 0.716–0.730 in the reanalysis mode to 0.682–0.704 in the hindcast mode (see the red boxplots in the Reanalysis and Hindcast subpanels in Figs. 3a–d, respectively). For the TP model type, for instance, the average ROC score of the $T + P$ model dramatically decreased, from 0.721–0.731 in the reanalysis mode to 0.560–0.606 in the hindcast mode (see the gray boxplots in the Reanalysis and Hindcast subpanels in Figs. 3a–d, respectively). An important note here is that although the average ROC score values in the hindcast mode were higher for the CI model than for the TP model type, the number of skillful grid cells in the reanalysis mode for the CI model (for instance, 1129 for wheat) was roughly half of that for the $T + P$ model (2249 for wheat), with rice as a noticeable exception (1396 for the CI model and 1397 for the $T + P$ model) (Figs. 3a–d). These results suggest that the skillful area extent of the CI model in the reanalysis mode is limited but that in areas where the CI model is skillful, it has a better prediction skill than the TP model type.

A decrease in the skillful area extent from the reanalysis mode to the hindcast mode is a common tendency among the model types. This decrease is associated with prediction errors in the climate hindcasts and visually indicated by the downward departures of the symbols from the 1:1 line in Fig. 4. For the CI model, the skillful area extent only slightly decreased, from 4.5%–7.7% in the reanalysis mode (see the x axis in Figs. 4a–d) to 3.7%–7.4% in the hindcast mode (see the y axis in Figs. 4a–d). The decrease in the skillful area extent was substantial for the TP model type. An extreme example was the GDD + EDD + $P$ model for soybean, for which the skillful area extent in the hindcast mode (5.7%) was one-third of that in the reanalysis mode (17.4%) (the blue circle in Fig. 4b). Among the other specifications of the TP model type, the $T + P$ model did a relatively good job in the hindcast mode. For maize, the $T + P$ and $T^2 + T + P^2 + P$ models ranked first (7.9%) and were closely followed by the GDD + EDD + $P$ model (7.8%). The $T + P$ model ranked second for soybean (7.0%) and rice (5.8%) after the CI model. For wheat, the $T + P$ model ranked third (12.1%), but its skillful area extent was almost comparable to that for the top two models (the $T^2 + T + P^2 + P$ (12.3%) and GDD + EDD + $P$ (12.2%) models).

The key skill-limiting factors of the within-season hindcasts were different among the model types. The skill of the CI model was almost always limited by the strength of the climate–yield relationship (Figs. 5m–p). In contrast, both the reliability of the climate forecasts and the strength of the climate–yield relationship emerged as skill-limiting factors for the TP model type (Figs. 5a–l).

c. Country-level prediction skill

The evaluation of prediction skill at the country level further highlighted the uniqueness of the CI model relative to the TP model type. The countries where the country-average yield anomaly hindcast appeared skillful were almost the same among the $T + P$, $T^2 + T + P^2 + P$, and GDD + EDD + $P$ models, with minor differences in the ROC score. Taking soybean as an example, the skillful countries included India, Mexico, Paraguay, South Africa, and the United States for the TP model type (Figs. 6a–c), whereas a quite different set of skillful countries (Bolivia, Canada, India and Vietnam) was obtained from the CI model (Fig. 6d). The tendency for only slight overlap of skillful countries between the two model types was common for all the crops examined here (Fig. 6, Figs. S3–S5). This uniqueness of the CI model is important for increasing the number of skillful countries in global crop forecasting services, as skillful countries can be added from the CI model to those from the TP model type.

The difference in skillful countries highlighted the importance of considering teleconnections simulated in climate models into crop forecasting through yield models such as the TP model type. The effects of considering teleconnections were prominent, for instance, for maize in Australia, the United States, Zambia, and Zimbabwe (Figs. S6a–c versus Fig. S6d); soybean in Argentina, Brazil, and South Africa (Figs. S7a–c versus Fig. S7d); rice in India and South China (Figs. S8a–c versus Fig. S8d); and wheat in western Europe and Kazakhstan (Figs. S9a–c versus Fig. S9d).

d. Dominant modes of climate variability

The results corroborated that some climate modes contribute to yield variability in some crops and regions of the world. ENSO was the dominant climate mode causing yield loss for maize in Indonesia, Thailand, Zambia, and Zimbabwe (Fig. 7a); soybean in India (Fig. 7b); rice in Brazil, India, Pakistan, Laos, and Uruguay (Fig. 7c); and wheat in Brazil, Northeast China, and South Africa (Fig. 7d). Other climate modes negatively affecting yields included the EA for maize in France, Italy, Romania, and Spain (Fig. 7a); the EA and PNA for soybean in the United States (Fig. 7b); and the IOD and EM for wheat in Australia, Ethiopia, and the United States (Fig. 7d). The NAO rarely appeared to lead to yield loss (Figs. 7a–d).

The dominant climate modes between yield loss and yield gain showed contrasts. For soybean in the United States, the dominant climate modes for yield loss were the EA and PNA, while ENSO was detected for yield gain (Figs. 7b and 7f). Another example was rice in Brazil, for which the dominant climate modes were ENSO for yield loss and EA and EM for yield gain (Figs. 7c and 7g). The EA was the dominant mode for wheat yield gain in France and Germany, while no climate mode was distinctive for wheat yield loss in these countries (Figs. 7d and 7h).

4. Discussion

a. Implications for crop forecasting services

Our analysis reveals that the reductions in prediction skill and skillful area extent from the reanalysis mode to the hindcast mode are substantially different among the model types. This finding has important implications for improving global crop forecasting services. While the skillful area extent is limited in absolute terms, crop forecasting, which relies on climate index forecasts such as the CI model, can be useful for some crops and countries (Fig. 6 and Figs. S3–S5). Importantly, the countries where the CI model is skillful are quite different from those where the TP models demonstrate meaningful prediction skill. Therefore, the use of climate index forecasts, in addition to temperature and precipitation forecasts, likely increases the
FIG. 5. The relationships between the key skill-limiting factors (ROC$_{low}$) and skill of the within-season yield anomaly hindcasts (ROC$_{\Delta Y\_Hindcast}$) for the four models.
total number of skillful countries and the extent of skillful areas in global crop forecasting services.

In addition, regional crop forecasting services that rely on climate index forecasts are likely useful. More specifically, linking ENSO forecasts with rice forecasting is promising for India. Wheat forecasting for Australia would benefit from IOD and EM forecasts as well as from ENSO forecasts. Utilizing forecasts of atmospheric indices (e.g., EA) for Europe has the potential to further improve existing crop forecasting systems in the region. These findings are in line with those of recent studies (Jha et al. 2016; Yuan and Yamagata 2015; Ceglar et al. 2017; Nobre et al. 2019). The prediction skills for the oceanic indices (IOD, EM, and ENSO) are considerably higher than those for the atmospheric indices (EA, NAO, and PNA) (Fig. S10). Therefore, regional crop forecasting services relying on oceanic indices are particularly worth examining in future research. Differences in forecast skill between winter and summer crops and their association with the forecast skill of climate indices and variables need be explored and characterized more explicitly.

Globally consistent simple climate indices tailored to specific sectors and computed using daily temperature and precipitation data have been developed and promoted by the Expert Team on Sector-Specific Climate Indices (ET-SCI) of the World Meteorological Organization (WMO) to help characterize the climate sensitivity of various sectors, including agriculture (WMO 2019). Our findings suggest the potential of oceanic and atmospheric indices (and temperature and precipitation) often differ by season. Although it is beyond the scope of this study, differences in forecast skill between winter and summer crops and their association with the forecast skill of climate indices and variables need be explored and characterized more explicitly.

b. Consistency of dominant climate modes across studies

The consistency of the dominant climate modes across studies varies by crop and region. A highly consistent tendency is found for wheat in Australia. Heino et al. (2018) report that the IOD is the significant climate mode for the western and southeastern parts of Australia, while ENSO, IOD, and NAO are the significant climate modes for the northeastern part of the country. This tendency is supported by the finding of Yuan and Yamagata (2015) that the impacts on wheat yield in the former area from the IOD are larger than those from ENSO and EM, whereas ENSO plays a dominant role in the latter area. Furthermore, ENSO and the IOD are estimated as the main climatic drivers in the eastern part of Australia, explaining 30% of wheat yield variability (Anderson et al. 2019). Our results that the dominant climate modes are the IOD and EM for wheat yield loss in western and southeastern parts of Australia and ENSO, NAO, and EA for wheat yield gains in the northeastern part of the country (Figs. 7d,h) are consistent with other studies (Yuan and Yamagata 2015; Heino et al. 2018; Anderson et al. 2019). Other examples showing a consistent tendency include the following. The NAO is rarely selected as the significant climate mode for winter wheat in Europe compared to the EA, EAWR, and Scandinavian patterns (Ceglar et al. 2017), which resembles our results (Figs. 7d,h). Jha et al. (2016) found negative impacts of ENSO on rice yield in South India, which is comparable to our result (Fig. 7c).

Relatively inconsistent tendencies across studies are also found. Anderson et al. (2019) indicate that 23% of maize yield variability in the U.S. Midwest is explained by ENSO, while Heino et al. (2018) report that no significant climate mode is detected for the region. Our determination that there is no dominant climate mode for the crop and region (Figs. 7a,e) is close to the result of Heino et al. (2018). Another example is rice in the Philippines and Indonesia. Koide et al. (2013) and Naylor et al. (2001) indicate the significant impacts of ENSO on rice production in the Philippines and Indonesia, respectively. Heino et al. (2018) suggest that ENSO and the NAO influence the crops in these countries. However, no extensive rice-cropping area where ENSO plays a dominant role is found for these countries in our results (Figs. 7c,g).

The following examples show contrasting tendencies. The positive impacts of the EAWR on maize in South Europe
(France, Italy, Portugal, Spain) are shown in Ceglar et al. (2017), while our analysis detects the EA as the dominant climate mode for maize yield loss in this region (Fig. 7a). Ceglar et al. (2017) also indicate the negative impacts of the NAO on maize in eastern Europe (e.g., Romania). However, the NAO and ENSO influenced maize yield gain in Romania in our results (Fig. 7e). Cane et al. (1994) reported a positive correlation between maize yield in Zimbabwe and the ENSO index (Niño-3), but our result is the opposite—ENSO negatively affects maize yield in the country (Fig. 7a). Explaining these discrepancies is beyond the scope of this study, but an in-depth analysis to reconcile them is needed. Although not an exhaustive list, different sources of yield data, detrending methods, yield model structures, and selection procedures within correlated predictors could result in different climate modes having the greatest influence on yield. This would be especially true for regions where the fingerprint of the climate mode is weak.

c. On the use of statistical yield models for seasonal crop forecasting

Our findings provide a unique view of statistical yield models for seasonal crop forecasting. Recent studies assessing climate change impacts on crop yields often incorporate squared season temperature and precipitation terms or daily temperature index terms (e.g., GDD and EDD) into models to capture nonlinear yield responses to weather and climate extremes (Tack et al. 2015; Roberts et al. 2017; Parkes et al. 2019). The model skill in the reanalysis mode may increase if these terms are considered. However, the incorporation of these terms does not increase the prediction skill or skillful area (Figs. 3 and 4) because the prediction skill of the TP models in many locations is limited by the reliability of climate forecasts rather than by the strength of the climate–yield relationship (Figs. 5a–l versus Figs. 5m–p). Given the current skill levels of temperature and precipitation forecasts, the prediction skill of a simple model, such as the $T_1P_1$ model, is comparable to those of more complex models, such as the $T_2 + P_1$ and GDD + EDD + $P$ models. These findings suggest that although the addition of explanatory variables to yield models, such as temperature and moisture conditions for the presowing period and vegetative growth stage, may be useful in the reanalysis mode, an assessment of the skills in the forecast mode is indispensable.

5. Conclusions

The present study evaluates and compares within-season yield anomaly hindcasts for major crops derived from different types of statistical yield models. The TP model type shows
different skills among the model specifications in explaining yield variability from actual temperature and precipitation conditions. Despite the sizable difference in the model skill in the reanalysis mode, in the forecast mode, the prediction skill of the simple specification is comparable to that of the more complex specifications of the TP model type. This finding underpins the use of simple models in global crop forecasting services.

The CI model is characterized by a small reduction in the skillful area from the reanalysis mode to the forecast mode. Although its skillful area in the reanalysis mode is almost always smaller than that of the TP model type, in the forecast mode, the skillful areas of the CI model are largest for rice and soybean. The CI model leads to skillful areas in different countries from those for the TP model type. This uniqueness of the CI model likely contributes to improving global crop forecasting services.

Acknowledgments. T.I. thanks Misako Hagiwara and Mizuki Kotoku for their support in an early stage of the analysis. T.I. was partially supported by Grants-in-Aid for Scientific Research (18H02317 and 20K06267) from the Japan Society for the Promotion of Science; the Environment Research and Technology Development Fund (JPMEEF20202005) of the Environmental Restoration and Conservation Agency of Japan; the Joint Research Program of the Arid Land Research Center, Tottori University (30F2001); and the Ministry of Education, Culture, Sports, Science and Technology (MEXT) Coordination Funds for Promoting AeroSpace Utilization (JP10000959).

Data availability statement. The yield dataset is available at https://doi.org/10.20783/DIAS.528. The climate hindcast data used here are available for purchase from the Japan Meteorological business support center (http://www.jmbsc.or.jp/en/index-e.html). The JRA-55 reanalysis data are available after user registration at https://jra.kishou.go.jp/JRA-55/index_en.html.

REFERENCES


Unauthenticated | Downloaded 05/30/22 06:13 AM UTC