Can crop simulation models be used to predict local to regional maize yields and total production in the U.S. Corn Belt?

Francisco J. Morell¹, Haishun S. Yang¹, Kenneth G. Cassman¹, Justin Van Wart¹, Roger W. Elmore¹, Mark Licht², Jeffrey A. Coulter³, Ignacio A. Ciampitti⁴, Cameron M. Pittelkow⁵, Sylvie M. Brouder⁶, Peter Thomison⁷, Joe Lauer⁸, Christopher Graham⁹, Raymond Massey¹⁰, Patricio Grassini^{1*}

¹ Department of Agronomy and Horticulture, University of Nebraska-Lincoln, P.O. Box 830915, Lincoln, NE 68583-0915

² Department of Agronomy, Iowa State University, Ames, IA 50011-1010

³ Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN 55108-6026

⁴ Department of Agronomy, Kansas State University, Manhattan, KS 66506

⁵ Department of Crop Sciences, University of Illinois, Urbana, IL 61801

⁶ Department of Agronomy, Purdue University, West Lafayette, IN 47907-2054

⁷ Department of Horticulture and Crop Science, The Ohio State University, Columbus, OH 43211

⁸ Department of Agronomy, University of Wisconsin, Madison, WI 53706

⁹ Plant Science Department, South Dakota State University, Rapid City, SD 57702

¹⁰ Department of Agricultural and Applied Economics, University of Missouri, Columbia, MO 65211

* Corresponding author. Tel.: +1 402-472-5554; fax: +1 402-472-6904. E-mail address:
pgrassini2@unl.edu (P. Grassini).

Abstract

Crop simulation models are used at the field scale to estimate crop yield potential, optimize current management, and benchmark input-use efficiency. At issue is the ability of crop models to predict local and regional actual yield and total production without need of site-year specific calibration of internal parameters associated with fundamental physiological processes. In this study, a well-validated maize simulation model was used to estimate yield potential for 45 locations across the U.S. Corn Belt, including both irrigated and rainfed environments, during four years (2011-2014) that encompassed diverse weather conditions. Simulations were based on measured weather data, dominant soil properties, and key management practices at each location (including sowing date, hybrid maturity, and plant density). The same set of internal model parameters were used across all site-years. Simulated yields were upscaled from locations to larger spatial domains (county, agricultural district, state, and region), following a bottom-up approach based on a climate zone scheme and distribution of maize harvested area. Simulated yields were compared against actual yields reported at each spatial level, both in absolute terms as well as deviations from long-term averages. Similar comparisons were performed for total maize production, estimated as the product of simulated yields and official statistics on maize harvested area in each year. At county-level, the relationship between simulated and actual yield was better described by a curvilinear model, with decreasing agreement at higher yields (>12 Mg

ha⁻¹). Comparison of actual and simulated yield anomalies, as estimated from the yearly yield deviations from the long-term actual and simulated average yield, indicated a linear relationship at county-level. In both cases (absolute yields and yield anomalies comparisons), the agreement increased with increasing spatial aggregation (from county to region). An approach based on long-term actual and simulated yields and year-specific simulated yield allowed estimation of actual yield with a high degree of accuracy at county level ($RMSE \leq 18\%$), even in years with highly favorable weather or severe drought. Estimates of total production, which are of greatest interest to buyers and sellers in the market, were also in close agreement with actual production ($RMSE \leq 22\%$). The approach proposed here to estimate yield and production can complement other approaches that rely on surveys, field crop cuttings, and empirical statistical methods and serve as basis for in-season yield and production forecasts.

Keywords: crop simulation model, upscaling, yield anomaly, yield potential, regional production

1. Introduction

Given the scenario of increasing global demand for feed, fuel, and food, potential impact of climate change, and limited room for further crop area expansion, it is crucial to develop robust and generic tools to forecast local, regional, and national crop production in order to inform policy, farm management, and marketing decisions at all levels of the supply chain (FAO, 2000; Hansen, 2005; Jayne and Rashid, 2010; Basso et al., 2013). There are several examples of yield forecast systems currently in use. The yield forecasting program of the U.S. Department of Agriculture (USDA) provides late-season and end-of-season estimates of maize yields, harvested area, and total production. Forecasts are released the first day of each month, beginning Aug 1st and ending Nov 1st (USDA-NASS, 2012). This approach is based on farmer-reported surveys for the estimation of sown area, and assessments of yield through objective field measurement surveys and field crop cuttings performed by trained personnel. At each forecast time during the growing season, the end-of-season predictions assume normal weather conditions for the remainder of the growing season. Similar yield predictions based on field measurements and/or surveys are performed in national programs of other countries: Field Crop Reporting Series in Canada (<http://www.statcan.gc.ca/eng/survey/agriculture/3401>), Crop Estimates Committee (CEC) in South Africa (CEC, 2008, <http://www.daff.gov.za/daffweb3/Home/Crop-Estimates/Methodology>), and General Crop Estimation Survey (GCEP) in India (Rathi, 2012). While reasonably accurate, this approach requires substantial resources in terms of logistics, labor, and time. Additionally, this method is intrinsically empirical and likely

116 to be uncertain in years when weather deviates substantially from long-term averages.
117 Other operational systems to forecast actual yield have been developed based on
118 empirical relationships between agro-climatic indexes and regional yields (e.g., Kestle et
119 al., 1982; Thompson, 1986; Potgieter et al., 2005; Chipanshi et al., 2015). While this
120 approach may be accurate at reproducing actual yield for specific geographic areas, its
121 applicability in other regions with different agro-climatic conditions is limited.

122

123 Mechanistic crop simulation models can capture major genotype x environment x
124 management (G x E x M) interactions that govern crop development, growth, and yield
125 (Carberry et al., 2009; van Ittersum et al., 2013). Hence, in principle, crop simulation
126 models could potentially represent a cost-effective way to estimate actual yields, or at a
127 minimum, to help fine tune yield estimates derived from empirical approaches, allowing
128 for yield estimates at finer spatial and temporal resolution (Gleason, 1982; Whisler et
129 al., 1986; Boogaard et al., 2013). However, simulation models are limited by several
130 factors related to their capacity to predict actual crop yields. First, crop models can only
131 account for a few of the many factors governing crop yields (e.g., solar radiation,
132 temperature, water supply) compared to the large list of yield limiting and reducing
133 factors typically present in farmer fields (Sinclair and Seligman, 1996). Second, lack of
134 high-quality weather, soil, and management data required to run these models is a
135 major constraint, especially in developing countries where these data are not accessible
136 or do not exist (Grassini et al., 2015). Due to the lack of measured data, many studies
137 have relied on coarse-gridded weather databases and unrealistically broad assumptions
138 on cropping system management, which typically results in biased yield estimates (Van

Wart et al., 2013a, b, 2015). Third, it is challenging to define the inference domain for a crop model simulation based on weather, soil, and crop management data for a specific location and how to upscale this ‘point’ estimate to larger spatial levels (van Wart et al., 2013a; van Bussel et al., 2015). Finally, simulation models should be able to reproduce yields across a wide range of environmental and management conditions without need of site-year specific calibration of internal parameters related to leaf area expansion, carbon assimilation and partitioning, and resource efficiency, though such calibration is common in the literature (e.g., Hodges et al., 1987; O’Neal et al., 2002; Rosenzweig et al., 2014). Examples of use of crop simulation models for estimating yields for specific site-years can be found in the literature (e.g., Bannayan and Crout, 1999; Baigorria et al., 2008; Quiring and Legates, 2008). At issue is, however, the ability of these models to capture the influence of G x E x M interactions on crop yields when they are used across a large number sites and years that cover a wide range of weather, soil, and management conditions.

There are very few regional yield forecast systems based on crop simulation models. An example is the MARS Crop Yield Forecasting System developed by the European Commission (MCYFC; <https://ec.europa.eu/jrc/en/research-topic/crop-yield-forecasting>). MCYFC makes use of copious amounts of weather, soil, and management data available at a very high spatial resolution (e.g., Boogaard *et al.*, 2002, de Wit *et al.*, 2010, Bussay *et al.*, 2015). This approach requires an extremely large number of simulations, which, in turn, makes quality control and validation of the results very difficult. Likewise, this approach is not applicable in regions where measured weather

and soil data at high level of spatial resolution do not exist (Grassini *et al.*, 2015). An alternative approach is to simulate yields for a relatively small number of sites, strategically selected based on their location within the crop harvested area and major agro-climatic zones, and upscale the simulated results to region and country level using a robust spatial framework (e.g., van Wart et al., 2013, van Bussel et al., 2015, 2016).

To our knowledge, no previous study has attempted to evaluate the use of a mechanistic crop simulation model for estimating actual yields at local to regional levels, using the same set of internal parameters across all site-years, and based on measured weather, soil, and management data retrieved from a relatively small number of sites strategically selected to portray crop area distribution and variation in agro-climatic conditions. The main goal of this study is to test the hypothesis that a well-validated model that simulates maize yield potential is able to reproduce actual grain yield and total grain production, without need of site-year specific calibration, with this hypothesis being specific to cropping systems where (i) farmers have access to information, technologies, and required inputs to utilize recommended management practices, (ii) high-quality weather, soil, and management data are available, and (iii) a robust upscaling method is available. As a proof of concept, a well-validated maize model was tested at four spatial scales (county, agricultural district, state, and Corn Belt region) relative to the ability to predict actual maize yields and total production. The main factors affecting the prediction power of actual yields were investigated to identify avenues for potential improvement.

2. Materials and methods

2.1. Study area, time period, and data inputs

The U.S. accounts for 36% of global maize grain production, and plays a major role in the global market, accounting for ca. 38% of global maize exports (FAOSTAT, 2008-2012), with the U.S. Corn Belt accounting for ca. 90% of total U.S. maize production (USDA-NASS, 2005-2014). During the last 35 years (1980-2014), U.S. maize production has increased by 85% due to expansion of cropland area (15%) and yield gain (60%) (Fig. 1). Agronomic and genetic factors, and their interactions, have supported a steady rise in U.S. maize yields of $110 \text{ kg ha}^{-1} \text{ y}^{-1}$ during the last five decades (Grassini *et al.*, 2014). However maize demand is accelerating and current domestic consumption for ethanol as a biofuel (39% of production) is equal to maize used as feed (ERS-USDA, 2009-2013) (Fig. 1).

Fig. 1. (a) Total U.S. maize grain production and uses during the last 35 years (1980-2014). (b) Average U.S. maize yield and harvested area during the same time period. Data from USDA-NASS and USDA-ERS.

201 The study area comprises the Corn Belt region of the U.S. and a description of climate,
202 soil, and management practices is provided elsewhere (Connor et al., 2011; Grassini et
203 al., 2014). The western edge of the Corn Belt includes the eastern Great Plains states
204 of North Dakota, South Dakota, Nebraska, and Kansas, where irrigated agriculture
205 accounts for ca. 30% of the total maize harvested area. Actual yield in the Corn Belt
206 region represents ca. 70-80% of yield potential (Grassini et al., 2011; Van Wart et al.,
207 2013a, www.yieldgap.org). Weather and soil data required for crop simulation models
208 are readily available and general management data (sowing date, hybrid maturity, and
209 plant density) can be retrieved from official statistics, seed companies, extension
210 educators, and Natural Resource Conservation Services.

211

212 Weather stations used for this study are part of MESONET state-operated networks
213 (<http://mrcc.isws.illinois.edu/gismaps/mesonets.htm>; Supplementary Table S1). High-
214 quality, long-term (≥ 15 most recent years) daily weather data are available. This time
215 period is sufficient to portray inter-annual weather variability in the favorable rainfed
216 production domain known as the U.S. Corn Belt while avoiding the confounding effect of
217 long-term climate change trends (van Wart et al., 2013a, Grassini et al., 2015). These
218 networks were explicitly designed for agricultural applications; hence, weather stations
219 are located in proximity to major crop production areas (in contrast to other networks in
220 which stations are located within or proximal to cities and airports). Measured daily
221 weather data at these stations contains all variables needed for crop yield simulations,
222 including solar radiation, maximum and minimum temperature (T_{max} and T_{min}),
223 relative humidity, precipitation, and wind speed. Missing precipitation and precipitation

224 data were filled with data from adjacent National Weather Station-Cooperative Observer
225 Network (NWS-COOP) stations while missing radiation data were filled with data from
226 the NASA-POWER database. When data gaps were smaller (<3 days) these were filled
227 by interpolating values between dates, except for precipitation gaps, which were always
228 filled with data from nearby NWS-COOP stations.

229

230 The U.S. Corn Belt was subdivided into climate zones (CZ), each representing a unique
231 combination of annual growing degree-days (GDD; $T_{base} = 0^{\circ}\text{C}$), aridity index, and
232 temperature seasonality (van Wart et al., 2013c) (Fig. 2). Following van Bussel et al.
233 (2015), a buffer zone area of 100 km radius, centered on each weather station, was
234 delimited and clipped by the CZ borders. Subsequently, weather stations used for this
235 study, hereafter called reference weather stations (RWS), were selected based on crop-
236 specific harvested area within each buffer zone until reaching ca. 50% maize-harvested
237 area coverage while ensuring that all CZs with >5% harvested area have at least one
238 RWS located within their borders. Previous work by Van Wart et al. (2013a) and van
239 Bussel et al. (2015) showed that achieving 50% coverage of the production area was
240 sufficient to obtain a robust estimate of yield potential at a national scale. Crop
241 harvested area maps (around year 2005), used for RWS selection for irrigated and
242 rainfed maize, were retrieved from the Spatial Production Allocation Model (SPAM; You
243 et al., 2014, <http://mapspam.info/>). RWS selection was performed separately for
244 irrigated and rainfed maize, with respective 16 and 41 RWS selected (Fig. 2,
245 Supplementary Table S2). Of these, 12 RWS were selected for both rainfed and
246 irrigated crops, which corresponded to areas where both rainfed and irrigated maize

247 production are important. Hence, a total of 45 different RWS across the U.S. Corn Belt,
248 located in 17 different CZ, were selected for the present study (Fig. 2, Supplementary
249 Table S2). The 45 RWS accounted for 59% of US maize-harvested area, while the 17
250 CZs where these RWS were located accounted for 88% of US maize cropped area.
251

Fig. 2. (A) Location of reference weather stations (RWS) selected for this study. For each RWS, yield was simulated for rainfed maize (yellow circles), irrigated maize (blue squares), or both (red triangles). County boundaries (red lines), agricultural districts (solid thin lines), and state boundaries (thick solid lines) are shown. State names are shown and RWS names were identified with a two-letter code (see Supplementary Table S2). Maize harvested area is indicated in green shading. B) Climate zones within the area of study and RWS, black dots. Each color corresponds to a different climate zone, each defined by a unique combination of annual growing degree-days, aridity index, and temperature seasonality (Van Wart et al., 2013c).

To avoid the confounding effect of technological changes while accounting for weather variation, our study focused on the most recent four years (2011-2014). Yield detrending would have been needed if our analysis had included earlier years (e.g., Ye *et al.*, 2015). Relative to mean U.S. maize yield of 9.6 Mg ha⁻¹ during the last decade (2004-2014), the 4 years evaluated in this study included a wide range of inter-annual weather variation typical of the U.S. Corn Belt: yield in 2012 was 20% below the national mean due to drought, 2011 and 2013 had yield near the mean ($\pm 5\%$), and 2014 was 12% above the mean yield (USDA-NASS, <http://quickstats.nass.usda.gov/>). The large number of RWS, years and water regime combinations allowed comparison of simulated and actual yields across a wide range of environments (*i.e.*, RWS-years) and management practices (water regimes, sowing dates, hybrid maturities, and plant densities).

Data on hybrid maturity (commercial relative maturity) and plant density, for the specific county where each RWS was located, were provided by local crop consultants, extension educators, and seed company agronomists in each state (Supplementary Table 2). The average sowing date was calculated from the average (2011-2014) calendar date at which 50% of the maize area was sown in each county based on data

provided by the USDA Risk Management Agency (RMA). Using year-specific versus average sowing dates had little impact on simulated yields and did not improve model's ability to reproduce actual yields, which was probably due to the relatively small inter-annual variation in the timing of the maize sowing across years (Supplementary Fig. S1). Management data (sowing date, hybrid maturity, and plant density) were specified separately for rainfed and irrigated maize at RWS locations where both water regimes co-existed. This study encompassed a wide range of sowing dates (April 10 to May 16), hybrid maturities (93 to 120 days), and planting density (3.7 to 8.5 plants m⁻²) across RWSs. The dominant soil type within each RWS buffer zone area was obtained from the SSURGO database (Soil Survey Staff, USDA-NRCS) and confirmed by local experts.

2.2 Crop yield simulation and upscaling method

Yield potential is defined as the maximum yield of an adapted crop cultivar as determined by solar radiation, temperature, carbon dioxide, and genetic traits that govern the length of growing period (Evans, 1993). Water-limited yield potential is determined by these previous factors and also by the amount of precipitation and its distribution during the crop growth period, in addition to field and soil properties that affect soil water storage such as slope, plant-available soil water holding capacity, and depth of the root zone (Lobell et al., 2009; van Ittersum and Rabbinge, 1997; van Ittersum et al., 2013). In the present study, maize yield potential and water-limited yield

potential were simulated for irrigated and rainfed crops, respectively, but, hereafter, both are referred to as simulated yields.

Hybrid-Maize is a crop simulation model used to simulate maize growth and development on a daily time step for both rainfed and irrigated crops (Yang et al., 2004, 2006). The model has been calibrated with data from well-managed rainfed and irrigated crops grown without nutrient limitations and kept free of biotic stresses (Yang et al., 2004). The Hybrid-Maize model does not account for the effect of inadequate nutrient supply, incidence of biotic stresses (weeds, insect pests, and pathogens), lack of stand uniformity, waterlogging, or hail. Previous studies have documented the Hybrid-Maize model's ability to reproduce maize yields obtained in well-managed experiments in which maize was grown without nutrient limitations and kept free of biotic stresses, with the model being robust at portraying key G x E x M interactions across a wide range of yields, from 0.5 to 18 Mg ha⁻¹ (Grassini et al., 2009; Liu et al., 2015). Perhaps more importantly, the model does not require site-specific parameter calibration for simulating yield potential, that is, all model parameters governing photosynthesis, respiration, leaf area expansion, light interception, biomass partitioning, and grain filling are generic (Yang et al., 2004).

Simulations utilized weather, soil, and management data obtained for each RWS location (see Section 2.1). For each RWS-year case, soil water content at sowing was determined dynamically by initializing the model run at harvest of the prior crop (*i.e.*, about 6 months before sowing), assuming 50% of available soil water content at that

time, and actual weather data from that prior harvest to sowing. The choice of 50% soil water content at harvest of prior crop was supported by data reported by a previous simulation study conducted for rainfed maize in the US Corn Belt (Grassini *et al.*, 2009). Perhaps more importantly, our sensitivity analysis indicated little change in simulated yields due to variation in available soil water content at harvest of prior crop: simulated yields based on 25% and 75% of soil water content were, across all RWS-year cases, within ± 0.7 t/ha ($\pm 6\%$) of simulated yields based on 50% soil water content. Soil surface residue coverage at harvest of the prior crop was set at 50% to reflect the large amount of non-grain aboveground biomass left by previous crop and the large proportion of maize production area under no-till or conservation tillage practices (Horowitz *et al.*, 2010). Rooting depth was set at 150 cm based on observed soil water extraction patterns for maize (Payero *et al.*, 2006). Likewise, surface runoff was assumed negligible to reflect producer preference for avoiding highly sloping land for maize production, and wide adoption of conservation tillage and no-till practices, which greatly reduce runoff.

Yield was simulated at all RWS for each year (2011, 2012, 2013, and 2014), and these location-specific simulated yields were upscaled to larger spatial domains following van Bussel *et al.* (2015) with slight modifications to estimate simulated yield for all U.S. administrative units for which actual average yield and total production data are available (county, agricultural district, and state; USDA-NASS, <http://quickstats.nass.usda.gov/>). The yield at the level of the Corn Belt region was estimated as a weighted average of yields in the Corn Belt states, accounting for the

different cropped area under each water regime. Total production was calculated as the total sum of production in the Corn Belt states. Some RWS were excluded from the analysis because no disaggregated data for rainfed and irrigated production was available, although irrigated production is locally important. These sites included Clarkton (MO), Hancock (WI), and all RWS in SD. In this study, simulated yield for a given RWS was considered to be representative of the county in which the RWS was located, and thus contrasted to county yields as reported by USDA-NASS. Simulated yield was upscaled from RWS to CZ based on the relative contribution of harvested area within a RWS buffer zone to total maize harvested area of each CZ as follows:

$$Y_{SCZ} = \frac{\sum_{i=1}^q Y_{SRWS_i} \times Area_{RWS_{buffer\ zone\ i}}}{\sum_{i=1}^q Area_{RWS_{buffer\ zone\ i}}} \quad \text{Equation 1}$$

where Y_s is the mean simulated yield for a given CZ, q is the number of RWS within the CZ, and $Area_{RWS_{buffer\ zone\ i}}$ is the maize harvested area in RWS buffer zone i as estimated from SPAM2005 (You et al., 2014). Upscaling was performed separately for irrigated and rainfed maize, resulting in 11 irrigated and 15 rainfed CZ-level simulated yield values. Simulated yield at agricultural district, state or Corn Belt region levels was estimated based on the relative contribution of each CZ to total maize harvested area within each administrative unit as follows:

$$Y_{S_{agricultural\ district,\ state,\ Corn\ Belt\ region}} = \frac{\sum_{i=1}^s Y_{SCZ_i} \times Area_{CZ_i}}{\sum_{i=1}^s Area_{CZ_i}} \quad \text{Equation 2}$$

where s is the number of CZ within the administrative unit (agricultural district, state, or Corn Belt region) and $Area_{CZ_i}$ is the maize harvested area of the administrative unit that

falls within each climate zone *i*. Following this procedure, maize yield was simulated for each year for each RWS, and then upscaled to agricultural district, state, and Corn Belt region. The upscaling method performed reasonably well as indicated by the close agreement between actual yields upscaled from county to agricultural district, state, and U.S. Corn Belt and respective USDA-NASS yields, with *ca.* 90% of upscaled yields at agricultural district and state levels within $\pm 15\%$ of corresponding USDA-NASS yields (Supplementary Fig. S2). Finally, potential maize grain production was estimated by multiplying the simulated yield at each administrative unit by the maize harvested area reported by USDA-NASS for the same administrative unit and year, separately for irrigated and rainfed maize. The objective was to expand the evaluation of simulation models, not only on their ability for simulating regional yield but also total grain production.

2.3. Comparison of simulated versus actual yield and production

Simulated maize yield potential and potential production were compared against actual yield and production data on a harvested area basis as reported by USDA-NASS (<http://quickstats.nass.usda.gov/>). Comparisons were performed at four spatial scales: county (roughly 2,000 km²), agricultural district (roughly 15,000 km²), state (roughly 150,000 km²), and Corn Belt (roughly 1,600,000 km²). Forecasted yields are subsequently aggregated to the U.S. agricultural reporting district, state, and national levels. Irrigated and rainfed maize were distinguished for the analysis. Both simulated and actual yields and production were reported at the standard 155 g kg⁻¹ grain

moisture content. Simulated yields were not adjusted by harvest losses because these are, with few exceptions, very small and its exact magnitude varies unpredictably across fields and years. The number of paired actual-simulated data at the county level was slightly less than at agricultural district level due to missing data for a small number of county-years. Root mean square error (RMSE), both in absolute and relative terms, were calculated to analyze agreement between simulated and actual yield and production:

$$RMSE = \sqrt{\frac{\sum_{i=1, j=1}^n (Y_{s_{ij}} - Y_{a_{ij}})^2}{n}} \quad \text{Equation 3}$$

$$RMSE (\% \text{ of mean}) = \frac{\sqrt{\sum_{i=1, j=1}^n (Y_{s_{ij}} - Y_{a_{ij}})^2 / n}}{\sum_{i=1, j=1}^n Y_{a_{ij}} / n} \times 100 \quad \text{Equation 4}$$

where $Y_{s_{ij}}$ is the potential yield or production at i^{th} location (county, agricultural district, state, or Corn Belt region) in year j^{th} , $Y_{a_{ij}}$ is the corresponding actual yield for same location-year, and n is the number of pairs of observations. Regression analyses were performed to analyze biases in the agreement between actual and simulated yields. Quadratic terms of fitted regression models were only kept when their significance was statistically significant ($P < 0.01$). Factors contributing to large differences between simulated and actual yields were assessed. Additional simulations were performed to explore the yield differences relative to variation in precipitation quantity and distribution across short distances (< 40 km). For this analysis, we used precipitation data from four nearest-neighbor NWS-COOP stations to the targeted RWS. We chose the NWS-COOP network for this analysis due the higher density of weather stations with precipitation data relative to the MESONET network.

416

417 Simulated and actual yields were also compared in terms of annual yield anomaly, that
418 is, yield deviation relative to the long-term average. For this analysis, the long-term
419 mean yield was calculated for the period 2004-2014, for both simulated and actual
420 yields, to avoid the effect of the technological trend and climate change. Simulated
421 (SYA) and actual (AYA) yield anomalies, expressed as a percentage of the long-term
422 simulated and actual mean yield, respectively, were calculated as follows:

423

$$424 \quad SYA_{ij} = \frac{(Y_{sij} - LT_{si})}{LT_{si}} \times 100 \quad \text{Equation 5}$$

425

$$426 \quad AYA_{ij} = \frac{(Y_{a_{ij}} - LT_{ai})}{LT_{ai}} \times 100 \quad \text{Equation 6}$$

427

428 where Y_{sij} and $Y_{a_{ij}}$ are the simulated and actual yield at location i and year j , and LT_{si}
429 and LT_{ai} are the long-term (2004-2014) mean simulated and actual yield at location i .
430 Knowing the sign of the yield anomaly, regardless of its magnitude, can be useful for
431 many agricultural applications, including early yield forecasting, marketing, and grain
432 trading. Hence, anomalies of simulated yield were compared against anomalies of
433 actual yield to assess model ability to reproduce years. For the purpose of this study,
434 plus or minus 10% was considered as near to the long-term mean yield, while $\leq 10\%$
435 and $\geq 10\%$ were considered as below and above average, respectively.

Finally, we tested a method to improve the actual yield predictions at each spatial scale based on adjusting the location-year specific simulated yield by the ratio between long-term mean actual and simulated yield at a given location as follows:

$$PY_{ij} = \frac{LTA_i}{LTS_i} \times Y_{sij} \quad \text{Equation 7}$$

where PY_{ij} is the predicted yield at location i in year j , LTA_i and LTS_i is the long-term (2004-2014) mean actual and simulated yield at location i , respectively, and Y_{sij} is the simulated yield at location i and year j . The predicted yields were compared against actual yields reported by USDA-NASS at the four spatial scales. A similar comparison was performed for total production. The objective was to determine the degree to which agreement between simulated and actual yields improves with the abovementioned method based on the long-term actual and simulated yields and the year-specific simulated yield.

3. Results and discussion

3.1. Weather variability during 2011-2014

The study period (2011-2014) portrayed the inter-annual and spatial weather variation across the U.S. Corn Belt as shown in cumulative precipitation and GDD at three RWS representative of the west-to-east gradient in precipitation: Custar OH (east), Ames IA (central), and Clay Center NE (west) (Fig. 3). Drought and high temperatures in 2012 led to a reduction in rainfed maize yield due to widespread crop water stress and shorter grain-filling duration. In contrast, cooler temperatures and above-average

precipitation in 2014 resulted in above average maize yields while 2011 and 2013 were closer to a more typical year. These contrasting weather conditions resulted in a wide range of actual yields across sites, years, and water regimes, from 1.3 to 14.2 Mg ha⁻¹. The lowest yield corresponded to rainfed maize in Nebraska in 2012 (state mean yield of 3.7 Mg ha⁻¹) while the highest yields were observed in favorable rainfed environments in Illinois in 2014 (state mean yield of 12.6 Mg ha⁻¹) and irrigated maize in Nebraska (state mean ranging from 11.3 to 12.6 Mg ha⁻¹ across years).

465

Fig. 3. Cumulative precipitation and growing-degree days (GDD, T_{base} and $T_{\text{max}} = 10^{\circ}\text{C}$ and 30°C , respectively) from May 1st until Sept 30th, which roughly coincides with maize emergence and physiological maturity, at three representative locations across the west-to-east precipitation gradient in the U.S. Corn Belt in 2011 (green), 2012 (red), 2013 (black), and 2014 (blue). Dashed line indicates the long-term historical average based on ≥ 25 year of measured daily weather data.

472

3.2. Comparison of simulated yields against actual yields and total production

Simulated maize yield at the county-scale was typically greater than actual yield (Fig. 4), as would be expected in commercial-scale fields where it is not feasible or cost effective to ensure adequate supply of all required nutrients in time and space and have perfect control of biotic stresses (Cassman et al., 2003). Simulated and actual maize yield at the county level exhibited an overall RMSE of 3.3 Mg ha⁻¹, which represented 34% of the mean actual yield across site-years. The fitted quadratic model ($P < 0.001$) indicated a trend in the relationship between actual and simulated yield. Paired observations were randomly scattered around the 1-to-1 line at low and intermediate yield levels, but simulated yields were consistently above actual yields at yields >12 Mg ha⁻¹. Indeed, the curvilinear relationship between actual and simulated yield at the county level suggests that a trade-off exists between the management needed to capitalize on the high yield potential of favorable years and the management that provides the highest return over years with an acceptable level of risk. Current management of rainfed crops is conservative to avoid risk of inadequate rainfall and may not provide optimal management that can take full advantage of years with above-normal precipitation. Interestingly, within the upper range of simulated yields from 10 to 18 Mg ha⁻¹ at the county level, actual yield of irrigated maize was closer to simulated yield potential than for rainfed maize (87 versus 81% of yield potential for irrigated and rainfed, respectively).

494

495 **Fig. 4.** Comparison of simulated and actual maize yields at four spatial levels: county
 496 (A), agricultural district (B), state (C), and U.S. Corn Belt region (D). Each symbol
 497 represents a different state. Rainfed and irrigated yields are indicated with solid and
 498 empty symbols, respectively, in 2011 (orange), 2012 (yellow), 2013 (green), and 2014
 499 (gray). The 1:1 line (solid red line) and $\pm 15\%$ deviation (dotted black lines) are shown.
 500 Fitted regression models and their parameters are shown (solid black line). Root mean
 501 square error (RMSE), expressed in both absolute and relative (% of actual mean yield)
 502 terms, is also shown.

503

504 Simulated yields were below actual yields in 20% of the county-year cases. These

505 cases corresponded to rainfed crops in the low yield range ($< 8 \text{ Mg ha}^{-1}$); some of these

biases may reflect spatial variation in rainfall that occurs due to the convective nature of summer thunderstorms that provide much of the rain in the US Corn Belt (Hubbard, 1994). That is, actual rainfall at a given RWS may differ from that received at nearby farm due to the variability of these convective storms over short distances, with a greater impact in years with below-average rainfall. Taking the drought year of 2012 as an example, simulated yields at Lewis IA and Custar OH were 1 to 1.2 Mg ha⁻¹ below their corresponding county actual yields. For these two locations, we performed additional simulations to assess the impact of precipitation variation over short spatial scales (<40 km) on simulated yield in a drought year (Supplementary Fig. S3). The analysis indicated important differences in precipitation amount (up to 300 mm) and distribution among nearby weather stations, which can help explain the difference between simulated and actual yield in a severe drought year (e.g., 2012). In contrast, spatial variation in precipitation was not a sensitive factor influencing simulated yield in years with precipitation near or above normal such as in 2013 (Supplementary Fig. S3). Another factor that is not accounted for by the crop simulation model, with a likely positive impact on yields in drought years, is the water supply from perched water tables in the central and eastern regions of the Corn Belt. For example, average depth to water table is roughly 0.5 to 1 m in north-central Illinois (Hollinger, 1995) and this area coincides with the location of three RWS (Peoria, Springfield, and Bondville) where simulated yields were well below actual yields in 2012.

Relative to these county-level estimates, comparison of actual and simulated yields at larger spatial scales (agricultural districts, state, and Corn Belt region) showed three

distinct patterns (Fig. 4). First, simulated yields were greater than actual yields in almost all site-year cases. Second, the relationship between actual and simulated yield became linear and agreement increased as comparisons moved from county (RMSE = 3.3 Mg ha⁻¹, r^2 = 0.59) to agricultural district (RMSE = 3.2 Mg ha⁻¹, r^2 = 0.61), state (RMSE = 2.9 Mg ha⁻¹, r^2 = 0.65), and Corn Belt region (RMSE = 2.1 Mg ha⁻¹, r^2 = 0.79). The same trend was observed for the RMSE, expressed a percentage of mean actual yield, which was 34, 34, 31 and 20% at the county, agricultural district, state and Corn Belt region, respectively. Third, the range of simulated and actual yields across site-years became smaller with increasing spatial aggregation (Fig. 4). This scaling trend suggests that large spatial scales with large maize area spread across a number of different CZ are buffered against low yields in a given year due to unfavorable climatic events, such as drought, because these events will rarely occur across a majority of the production area. Likewise, highest simulated yields are averaged out when moving from small to large spatial scales, making the plateauing effect to disappear too.

Increasing agreement between simulated and actual production, with increasing spatial aggregation, was consistent with the observed trend for yield, with simulated production ca. 15% above actual production across the four spatial scales (Fig. 5). Correlation between simulated and actual production was substantially stronger ($r^2 \geq 0.89$) relative to correlation between simulated and actual yields ($r^2 \geq 0.59$). This difference was partly due to the use of the same harvested area values for the calculation of both actual and simulated production. But there was also an effect of larger yield differences being related to smaller harvested areas and *vice versa*, leading to smaller differences of their

552 production values which, in turn, led to a better correlation. In few words, a yield
553 belonging to a spatial unit (county, agricultural district, state) with a small harvested
554 area resulted in a relative small production value, ending up in the lower left part of the
555 graph and masking the large difference between simulated and actual yields (Fig. 5).
556 Likewise, the impact of large yield differences in some of the most favorable
557 environments (*e.g.*, irrigated maize in Nebraska) on total production was also smoothed
558 by their relatively small harvested area. Furthermore, regions with small yield
559 differences related mostly to larger harvested areas. Altogether, the observed
560 interdependency between yield differences and harvested areas made forecasting of
561 total production more promising than yield forecasting.

562

Fig. 5. Comparison of maize grain production estimates with simulated and actual yields at four spatial levels: county (A), agricultural district (B), state (C), and U.S. Corn Belt region (D). Each symbol represents a different state. Rainfed and irrigated yields are indicated with solid and empty symbols, respectively, in 2011 (orange), 2012 (yellow), 2013 (green), and 2014 (gray). Simulated production was estimated as the product of simulated yields and reported maize harvested area by USDA-NASS within each aggregation level. The 1:1 line (solid red line) and $\pm 15\%$ deviation (dashed black lines) are shown. Fitted regression models and their parameters are shown (solid black line). Root mean square error (RMSE), expressed in both absolute and relative (% of mean actual yield) terms, is also shown.

Large positive and negative yield anomalies corresponded to respective years with very favorable (2014) and unfavorable (2012) weather conditions for crop growth and yield (Fig. 6). A linear regression model provided the best fit of simulated versus actual yield anomalies across all spatial scales. RMSE values for these yield anomalies were consistently smaller than RMSE values calculated for the comparison of absolute yields (Fig. 4). Degree of agreement between simulated and actual anomalies increased with greater spatial aggregation, which was consistent with the observed trends for yield and total production. Analysis of simulated yield anomalies, relative to their ability to distinguish years with actual yield below, near, and above the long-term average, is shown in Supplementary Table S3. Anomalies of simulated yield were in agreement with actual yield anomalies in 60% (county), 69% (agricultural district), 73% (state), and 100% (Corn Belt region) of the cases (Supplementary Table S3). The probability of simulating the opposite anomaly was remarkably small across all spatial scales: four and two cases (out of respective 162 and 176 total cases) at county and agricultural district levels (Supplementary Table S3). Rainfall spatial variation and/or poor quality of weather data in specific site-years might explain the opposite simulated yield anomalies.

Fig. 6. Comparison of anomalies of simulated and actual maize grain yield at four spatial levels: county (A), agricultural district (B), state (C), and U.S. Corn Belt region (D). Each symbol represents a different state. Rainfed and irrigated yields are indicated with solid and empty symbols, respectively, in 2011 (orange), 2012 (yellow), 2013 (green), and 2014 (gray). Yield anomaly was calculated as the deviation of simulated or actual yield in a given year relative to the long-term (2004-2014) simulated or actual yield mean at the same location (and expressed as %). The 1:1 line (solid red line) and $\pm 15\%$ deviation (dashed black lines) are shown. Gray vertical and horizontal lines indicate $x = 0$ and $y = 0$, respectively. Fitted linear regression models and their parameters are shown (solid black line).

Based on the abovementioned analysis, it was apparent that prediction of actual yields could be improved by adjusting the simulated yield in a given year-location adjusted by the ratio between long-term (2004-2014) mean actual and simulated yields at each location (Fig. 7). This method was effective at increasing agreement between simulated and actual yields across all spatial levels, with RMSE (in %) of 18% and 7% at the county and Corn Belt region levels, respectively, without a detectable bias along the range of actual yield (Fig. 7). These RMSE values were substantially below the RMSE estimated for the comparison between simulated yields (without adjustment) and actual yields, which ranged from 34% (county) to 20% (Corn Belt region) (Fig. 4). The yield correction led also to an almost two-fold increase in the accuracy (*i.e.*, RMSE decreased by ca. 50%) of predicted production at the four spatial scales (Fig. 8) relative to the estimated production using the uncorrected yields (Fig. 5). Across all spatial scales, agreement between simulated and actual yield and production values was better for irrigated *versus* rainfed crops, based on both absolute and relative RMSE calculated separately for each water regime (Supplementary Table 4). While intrinsically empirical, the proposed method proved robust to account for site characteristics that were not accounted for by the model or are unknown and can easily be applied without forcing models to reproduce actual yield through unjustified reconfiguration of internal parameters.

Fig. 7. Comparison of maize grain yield predicted based on the location-year specific simulated yield, adjusted by the ratio between long-term mean actual and simulated yield at each location, at four spatial levels: county (A), agricultural district (B), state (C), and U.S. Corn Belt region (D). Each symbol represents a different state. Rainfed and irrigated yields are indicated with solid and empty symbols, respectively, in 2011 (orange), 2012 (yellow), 2013 (green), and 2014 (gray). The 1:1 line (solid red line) and ±15% deviation (dashed black lines) are shown. Fitted linear regression models and their parameters are shown (solid black line).

The approach described in this paper to estimate regional yields can be applied to other cropping systems where current farm yields approach yield potential and high-quality historical weather records are available for long periods (10 to 30 years, depending on weather variability at the given crop system) in order to portray inter-annual weather variation and reliable management, soil, and actual yield data are accessible. If long-term historical weather data are available and current weather can be accessed real time, the approach described here can also be used to increase accuracy of in-season crop yield and production forecasts (e.g., Bannayan and Crout, 1999; Yang et al., 2006; Quiring and Legates, 2008). Crop growth can be simulated up to the date of the forecast and historical weather data can be used to complete the rest of the growing season and simulate as many possible yields as years of weather data are available. By comparing these possible yields against the long-term simulated yield for the same location, it is possible to derive a probabilistic distribution of simulated yield anomalies for the current season. This distribution, together with the long-term mean actual yield for the specific spatial unit, can then be used to derive a range of forecasted actual yields and probabilities for above-, near-, and below-average yield.

Conclusions

To be useful as decision-support tools, crop simulation models must be able to portray the impact of major G x E x M interactions on yield, using a limited number of accessible data inputs, without need of site-year specific calibration of internal parameters associated with fundamental physiological processes. The approach proposed in the present study makes use of publicly available data (or easily accessible as it is the case

662 for hybrid maturity and plant density) to simulate yield anomalies and provide robust
663 estimates of actual yields at different spatial scales, even in years with highly favorable
664 weather or severe drought. Simulated sites are strategically selected for their location
665 within climate zones with the greatest maize harvested area. Crop yield simulations are
666 performed using location-specific observed weather data, dominant soil types, and crop
667 management information, without the requirement of site-year model calibration. Local
668 uncertainties on both actual and simulated yields are offset when results are upscaled
669 to larger spatial levels and, hence, estimation of actual yields improved substantially as
670 the scale of spatial aggregation increased. And while the degree to which these yield
671 estimates can be considered accurate or not ultimately depends upon the specific user
672 and application, the proposed method seems robust at reproducing local to regional
673 yields, across a diverse range of climate, soil, and management, for both rainfed and
674 irrigated crops. Estimates of total production, which are of greatest interest to buyers
675 and sellers in the market, were also in close agreement with actual production. The
676 cost-effective and generic approach proposed here can complement other approaches
677 that rely on surveys, field crop cuttings or empirical relationships between yield and
678 agro-climate indexes to estimate yield and production (Basso et al., 2013 and reference
679 cited therein). It can also be easily coupled with emerging approaches to estimate
680 actual yields based on remote sensing or statistical methods (e.g., Johnson, 2008;
681 Bsaibes et al., 2009, Guindin 2013, Chipanshi et al., 2015; Lobell, 2013, 2015; López-
682 Lozano et al., 2015) and serve as foundation for monitoring local and regional
683 environmental footprint and climate change impact (e.g., Supit *et al.*, 2012).
684

685 **Acknowledgements**

686 We acknowledge DuPont Pioneer® agronomists for providing information to identify
687 dominant hybrid maturity and planting density at each location and water regime. We
688 are also grateful to Rebecca Davis (Risk Management Agency-USDA) for providing
689 county sowing progress data, Maribeth Milner (University of Nebraska-Lincoln, UNL) for
690 helping to identify the mean sowing date in each county and year, and Greg Krueger
691 and Charles Shapiro (UNL), and Jennifer Rees and Keith Glewen (UNL Extension
692 Educators) for providing crop management information for Nebraska. We also thank
693 Dennis Todey (South Dakota State University), Ken Scheeringa (Purdue University),
694 Rick Wayne (University of Wisconsin-Madison), Mark Seeley (University of Minnesota),
695 Jenny Atkins (WARM Program, University of Illinois at Urbana-Champaign), Daryl
696 Herzmann (Iowa State University), and Bill Sorensen and Natalie Umphlett (UNL) for
697 their help to access historical daily weather data.

698

699 **References**

700 Baigorria, G.A., Jones, J.W., O'Brien, J.J., 2008. Potential predictability of crop yield
701 using an ensemble climate forecast by a regional circulation model. *Agric. For. Meteorol.*
702 148, 1353–1361.

703

704 Bannayan, M., Crout, N.M.J., 1999. A stochastic modelling approach for real-time
705 forecasting of winter wheat yield. *Field Crops Res.* 62, 85-95.

706 Basso, B., Cammarano, D., Carfagna, E., 2013. Review of crop yield forecasting
 707 methods and early warning systems. In: Report Presented to First Meeting of the
 708 Scientific Advisory Committee of the Global Strategy to Improve Agricultural and Rural
 709 Statistics. FAO, Headquarters, Rome, Italy, 18–19 July. Available URL:
 710 [http://www.fao.org/fileadmin/templates/ess/documents/meetings_and_workshops/GS_S](http://www.fao.org/fileadmin/templates/ess/documents/meetings_and_workshops/GS_SAC_2013/Improving_methods_for_crops_estimates/Crop_Yield_Forecasting_Methods_and_Early_Warning_Systems_Lit_review.pdf)
 711 [AC_2013/Improving_methods_for_crops_estimates/Crop_Yield_Forecasting_Methods](http://www.fao.org/fileadmin/templates/ess/documents/meetings_and_workshops/GS_SAC_2013/Improving_methods_for_crops_estimates/Crop_Yield_Forecasting_Methods_and_Early_Warning_Systems_Lit_review.pdf)
 712 [and_Early_Warning_Systems_Lit_review.pdf](http://www.fao.org/fileadmin/templates/ess/documents/meetings_and_workshops/GS_SAC_2013/Improving_methods_for_crops_estimates/Crop_Yield_Forecasting_Methods_and_Early_Warning_Systems_Lit_review.pdf)

713 Boogaard, H.L., Eerens, H., Supit, I., van Diepen, C.A., Piccard, I., Kempeneers, P.,
 714 2002. Description of the MARS Crop Yield Forecasting System (MCYFS). Study
 715 contract no. 19226-2002-02-F1FED ISP. Joint Research Centre, Ispra

716 Boogaard, H., Wolf, J., Supit, I., Niemeyer, S., van Ittersum, M., 2013. A regional
 717 implementation of WOFOST for calculating yield gaps of autumn-sown wheat across
 718 the European Union, Field Crops Res. 143, 130-142.

719 Bsaibes, A., Courault, D., Baret, F., Weiss, M., Olioso, A., Jacob, F., Hagolle, O.,
 720 Marloie, O., Bertrand, N., Desfond, V., Kzemipour, F., 2009. Albedo and LAI estimates
 721 from FORMOSAT-2 data for crop monitoring. Remote Sens. Environ. 113, 716–729.

722 Bussay, A, van der Velde, M, Fumagalli, D., Seguini, L., 2015 Improving operational
 723 maize yield forecasting in Hungary. Agric. Syst. 141, 94-106.

724 Carberry, P.S., Hochman, Z., Hunt, J.R., Dalglish, N.P., McCown, R.L., Whish, J.P.M.,
 725 Robertson, M. J., Foale, M.A.P., Poulton, L., van Rees, H., 2009. Re-inventing model-

726 based decision support with Australian dryland farmers. 3. Relevance of APSIM to
727 commercial crops. *Crop and Pasture Sci.* 60, 1044–1056.

728 Cassman, K.G., Dobermann, A., Walters, D.T., Yang, H., 2003. Meeting cereal demand
729 while protecting natural resources and improving environ- mental quality. *Annual Rev.*
730 *Env. Res.* 28: 315–358.

731 Chipanshi, A., Zhang, Y., Kouadio, L., Newlands, N., Davidson, A., Hill, H., Warren, R.,
732 Qian, B., Daneshfar, B., Bedard, F., Reichert, G., 2015. Evaluation of the Integrated
733 Canadian Crop Yield Forecaster (ICCYF) model for in-season prediction of crop yield
734 across the Canadian agricultural landscape, *Agric. Forest Meteorolo.* 206, 137-150.

735 Connor, D.J., Loomis, R.S., Cassman, K.G., 2011. *Crop Ecology. Productivity and*
736 *Management in Agricultural Systems.* Cambridge Univ. Press, Cambridge, UK.

737 Crop Estimates Committee (CEC) of South Africa, 2008. Current methodology of official
738 crop forecasting in South Africa. Available at:
739 <http://www.daff.gov.za/daffweb3/Home/Crop-Estimates/Methodology>

740 de Wit, A., Baruth, B., Boogaard, H., van Diepen, K., van Kraalingen, D., Micale, F., te
741 Roller, J., Supit, I., van den Wijngaart, R., 2010. Using ERA-INTERIM for regional crop
742 yield forecasting in Europe. *Climate Research* 44, 41-53.

743 Evans, L.T., 1993. *Crop Evolution, Adaptation and Yield.* Cambridge University Press,
744 Cambridge.

745 Field Crops Reporting Series of Canada. Available at:
746 <http://www.statcan.gc.ca/eng/survey/agriculture/3401>

747 Food and Agriculture Organization of the United Nations, FAO Statistical Databases.
748 Available at <http://faostat.fao.org/>

749 Food and Agriculture Organization of the United Nations, FAO, 2000. Crop forecasting,
750 In: Report of the 18th session of the Asia and Pacific commission on agricultural
751 Statistics, 6-10 November, 2000, Bali, Indonesia. Available at:
752 <http://www.fao.org/docrep/004/ab989e/ab989e05.htm>

753 Gleason, C.P., 1982. Large area yield estimation/forecasting using plant process
754 models. Supervisory Mathematical Statistician Statistical Reporting Service, USDA,
755 Paper No. 82-4569. Winter Meeting American Society of Agricultural Engineers.
756 Chicago, USA. Available at:
757 [http://www.nass.usda.gov/Education_and_Outreach/Reports, Presentations and Conf](http://www.nass.usda.gov/Education_and_Outreach/Reports,_Presentations_and_Conferences/Yield_Reports/Large%20Area%20Yield%20Estimation%20Forecasting%20Using%20Plant%20Process%20Models.pdf)
758 [erences/Yield_Reports/Large%20Area%20Yield%20Estimation%20Forecasting%20Usi](http://www.nass.usda.gov/Education_and_Outreach/Reports,_Presentations_and_Conferences/Yield_Reports/Large%20Area%20Yield%20Estimation%20Forecasting%20Using%20Plant%20Process%20Models.pdf)
759 [ng%20Plant%20Process%20Models.pdf](http://www.nass.usda.gov/Education_and_Outreach/Reports,_Presentations_and_Conferences/Yield_Reports/Large%20Area%20Yield%20Estimation%20Forecasting%20Using%20Plant%20Process%20Models.pdf)

760 Grassini, P., Yang, H., Cassman, K.G., 2009. Limits to maize productivity in Western
761 Corn-Belt: A simulation analysis for fully irrigated and rainfed conditions. Agric. Forest
762 Meteorol. 149, 1254-1265.

763 Grassini, P., Thorburn, J., Burr, C., Cassman, K.G., 2011. High-yield irrigated maize
764 systems in the Western U.S. Corn-Belt. I. On-farm yield, yield-potential, and impact of
765 agronomic practices. Field Crop Res. 120, 142–150.

766 Grassini, P., Specht, J.E., Tollenaar, M., Ciampitti, I., Cassman, K.G., 2014. High-yield
 767 maize–soybean cropping systems in the US Corn Belt. In: Crop Physiology (2nd Edition),
 768 Sadras, V.O., Calderini, D.F. (Eds). Academic Press, San Diego..

769 Grassini, P., van Bussel, L.G.J, Van Wart, J., Wolf, J., Claessens, L., Yang, H.,
 770 Boogaard, H., de Groot, H., van Ittersum, M.K., Cassman, K.G., 2015. How good is
 771 good enough? Data requirements for reliable crop yield simulations and yield-gap
 772 analysis. Field Crops Res. 177, 49-63.

773 Guindin, 2013. Estimating midwestern maize grain yield from crop biophysical
 774 parameters using remote sensing. In: Proceedings of ASA/CSSA/SSSA 2013 Annual
 775 Meetings, Tampa, FL.

776 Hasen, J.W., 2005. Integrating seasonal climate prediction and agricultural models for
 777 insights into agricultural practice. Phil. Trans. R. Soc. B, 360, 2037–2047

778 Hodges, T., Botner, D., Sakamoto, C., Hays Haug, J., 1987. Using the CERES-Maize
 779 model to estimate production for the U.S. Corn Belt. Agric. Forest Meteorol. 40, 293-303

780 Hubbard, K.G., 1994. Spatial variability of daily weather variables in the high plains of
 781 the USA. Agric. For. Meteorol. 68, 29–41.

782 Hollinger, S.E., 1995. Midwestern Climate Center Soils Atlas and Database. Illinois
 783 State Water Survey. Champaign, Circular 179.

784 Horowitz, J., Ebel, R., Ueda, K., 2010. No-till farming is a growing practice. USDA-ERS
 785 Economic Information Bulletin No. 70.

786 Jayne, T.S., Rashid, S., 2010. The value of accurate crop production forecasts. MSU
787 International Development Working Paper 108. Available at:
788 <http://fsg.afre.msu.edu/papers/idwp108.pdf>

789 Johnson, D.M., 2014. An assessment of pre- and within-season remotely sensed
790 variables for forecasting corn and soybean yields in the United States. Rem. Sens.
791 Environ. 141, 116–128.

792 Kestle, R.A., 1982. Evaluation of the Thompson-type yield models for soybeans in Iowa,
793 Illinois, and Indiana. USDA Statistical Reporting Service, Statistical Research Division,
794 Staff Report Number AGES 820715. Available URL:
795 [http://www.nass.usda.gov/Education_and_Outreach/Reports, Presentations and Conf](http://www.nass.usda.gov/Education_and_Outreach/Reports,_Presentations_and_Conferences/Yield_Reports/Evaluation%20of%20the%20Thompson-type%20Yield%20Models%20for%20Soybeans%20in%20Iowa,%20Illinois,%20and%20Indiana.pdf)
796 [erences/Yield_Reports/Evaluation%20of%20the%20Thompson-](http://www.nass.usda.gov/Education_and_Outreach/Reports,_Presentations_and_Conferences/Yield_Reports/Evaluation%20of%20the%20Thompson-type%20Yield%20Models%20for%20Soybeans%20in%20Iowa,%20Illinois,%20and%20Indiana.pdf)
797 [type%20Yield%20Models%20for%20Soybeans%20in%20Iowa,%20Illinois,%20and%20](http://www.nass.usda.gov/Education_and_Outreach/Reports,_Presentations_and_Conferences/Yield_Reports/Evaluation%20of%20the%20Thompson-type%20Yield%20Models%20for%20Soybeans%20in%20Iowa,%20Illinois,%20and%20Indiana.pdf)
798 [Indiana.pdf](http://www.nass.usda.gov/Education_and_Outreach/Reports,_Presentations_and_Conferences/Yield_Reports/Evaluation%20of%20the%20Thompson-type%20Yield%20Models%20for%20Soybeans%20in%20Iowa,%20Illinois,%20and%20Indiana.pdf)

799 MARS Crop Yield Forecasting System (MCYFS). Available URL:
800 <https://ec.europa.eu/jrc/en/research-topic/crop-yield-forecasting>

801 Liu, X., Andresen, J., Yang, H., and Niyogi, D., 2015. Calibration and Validation of the
802 Hybrid-Maize crop model for regional analysis and application over the U.S. Corn Belt.
803 Earth Interactions. 19, 1-16.

804 Lobell, D.B., Cassman, K.G., Field, C.B., 2009. Crop yield gaps: their importance,
805 magnitude, and causes. Annu. Rev. Environ. Resour. 34, 1–26.

806 Lobell, 2013. The use of satellite data for crop yield gap analyses. *Field Crops Res.*
807 143, 56-64

808 Lobell, D.B., Thau, D., Seifert, C., Engle, E., Little, B., 2015. A scalable satellite-based
809 crop yield mapper. *Remote Sens. Environ.* 164, 324–333.

810 López-Lozano, R., Duveiller, G., Seguini, L., Meroni, M., García-Condado, S., Hooker,
811 J., Leo, O., Baruth, B., 2015 Towards regional grain yield forecasting with 1 km-
812 resolution EO biophysical products: Strengths and limitations at pan-European level.
813 *Agric. Forest Meteorol.* 206, 12-32.

814 O’Neal, M.R., Frankenberger, J.R., Ess, D.R., 2002. Use of CERES-Maize to study
815 effect of spatial precipitation variability on yield. *Agric. Syst.* 73, 205-225.

816 Payero, J.O., Klocke, N.L., Schneekloth, J.P., Davison, D.R., 2006. Comparison of
817 irrigation strategies for surface-irrigated corn in West Central Nebraska. *Irrig. Sci.* 24,
818 257-265.

819 Potgieter, A.B., Hammer, G.L., Doherty, A., de Voil, P., 2005. A simple regional-scale
820 model for forecasting sorghum yield across North-Eastern Australia. *Agric. Forest*
821 *Meteorol.* 132, 143–153.

822 Quiring, S.M., Legates, D.R., 2008. Application of CERES-Maize for within-season
823 prediction of rainfed corn yields in Delaware, USA. *Agric. Forest Meteorol.* 148, 964-
824 975.

825 Rath, R.P., 2012. General crop estimation survey (GCES), chapter 14, in Techniques of
826 Estimation and Forecasting of Crop Production in India, Edited by Sud, U.C., Chandra,
827 H., Aditya, K., ICAR, New Delhi, India

828 Rosenzweig, C., Elliott, J., Deryng, D., Ruane, A.C., Müller, Ch., Arneth, A., Boote, K.J.,
829 Folberth, C., Glotter, M., Khabarov, N., Neumann, K., Piontek, F., Pugh, T.A.M.,
830 Schmid, E., Stehfest, E., Yang, H., Jones, J.W., 2014. Assessing agricultural risks of
831 climate change in the 21st century in a global gridded crop model intercomparison.
832 Proc. Natl. Acad. Sci. U.S.A. 111, 3268–3273.

833 Sinclair, T.R., Seligman, N.G., 1996. Crop modeling: from infancy to maturity. Agron. J.
834 88, 698-704.

835 Soil Survey Staff, Natural Resources Conservation Service, United States Department
836 of Agriculture (NRCS-USDA). Web Soil Survey. Available at:
837 <http://websoilsurvey.nrcs.usda.gov/>

838 Supit, I, van Diepen, C.A., de Wit, A.J.W., Wolf, J., Kabata, P., Baruthc, B., Ludwiga, F.,
839 2008. Assessing climate change effects on European crop yields using the Crop Growth
840 Monitoring System and a weather generator. Agric. For. Meteorol. 164, 96-111.

841 Thompson, L. M., 1986. Climatic Change, Weather Variability, and Com Production.
842 Agron. J. 78, 649-653.

843 USDA–National Agricultural Statistics Service (NASS), 2000-2014. Quick stats 2.0.
844 Available at: <http://quickstats.nass.usda.gov/>

845 USDA-National Agricultural Statistical Service (NASS), 2012. The yield forecasting and
846 estimating program of NASS, by the Statistical Methods Branch, Statistics Division,
847 National Agricultural Statistics Service, U.S. Department of Agriculture, Washington,
848 D.C., April 2012. NASS Staff Report No. SMB 12-01.

849 USDA-ERS- National Agricultural Statistical Service, Economic Research Service.
850 Available at: <http://www.ers.usda.gov>

851 van Bussel, L.G.J., Grassini, P., Van Wart, J., Wolf, J., Claessens, L., Yang, H.,
852 Boogaard, H., de Groot, H., Saito, K., Cassman, K.G., Van Ittersum, M.K., 2015. From
853 field to atlas: Upscaling of location-specific yield gap estimates. *Field Crops Res*, 117,
854 98-108.

855 van Bussel, L.G.J., Ewert, F., Zhao, G., Hoffmann, H., Enders, A., Wallach, D., Asseng,
856 S., Baigorria, G.A., Basso, B., Biernath, C., Cammarano, D., Chryssanthacopoulos, J.,
857 Constantin, J., Elliott, J., Glotter, M, Heinlein, F., Kersebaum, K-C, Klein, C., Nendel,
858 C., Priesack, E, Raynal, H., Romero, C.C., Rötter, R.P., Specka, X., Tao F., 2016.
859 Spatial sampling of weather data for regional crop yield simulations. *Agric. For.*
860 *Meteoro.* 220, 101-115.

861 van Ittersum, M.K., Rabbinge, R., 1997. Concepts in production ecology for analysis
862 and quantification of agricultural input–output combinations. *Field Crops Res.* 52, 197–
863 208.

864 van Ittersum, M.K., Cassman, K.G., Grassini, P., Wolf, J., Titttonell, P., Hochman, Z.,
865 2013. Yield gap analysis with local to global relevance—a review. *Field Crops Res.* 143,
866 4–17.

867 Van Wart, J., Kersebaum, C., Peng, S., Milner, M., Cassman, K.G., 2013a. Estimating
868 crop yield potential at regional to national scales. *Field Crops Res.* 143, 34–43.

869 Van Wart J, Grassini P, Cassman KG, 2013b. Impact of derived global weather data on
870 simulated crop yields. *Global Change Biology* 19:3822-3834.

871 Van Wart, J., Van Bussel, L.G.J., Wolf, J., Licker, R., Grassini, P., Nelson, A.,
872 Boogaard, H., Gerber, J., Mueller, N.D., Claessens, L., Cassman, K.G., Van Ittersum,
873 M.K., 2013c. Reviewing the use of agro-climatic zones to upscale simulated crop yield
874 potential. *Field Crops Res.* 143, 44–45.

875 Van Wart J, Grassini P, Yang HS, Claessens L, Jarvis A, Cassman KG, 2015. Creating
876 long-term weather data from thin air for crop simulation modelling. *Agric. Forest*
877 *Meteoro.* 208, 49-58.

878 Whisler, F.D., Ackokc, B., Baker, D.N., Fye, R.E., Hodges, H.F., Lambert, J.R.,
879 Lemmon, H.E., McKinion, J.M., Reddy, V.R., 1986. Crop simulation models in
880 agronomic systems. *Adv. Agron.* 40 141-208.

881 Yang, H.S., Dobermann, A., Lindquist, J.L., Walters, D.T., Arkebauer, T.J., Cassman,
882 K.G., 2004. Hybrid-maize - A maize simulation model that combines two crop modeling
883 approaches. *Field Crops Res.* 87, 131-154.

884 Yang, H., Dobermann A., Cassman K.G., and Walters D.T., 2006. Features,
885 applications, and limitations of the Hybrid-Maize simulation model. *Agron. J.* 98, 737-
886 748.

887 Ye, T., Nie, J., Wang, J., Shi, P., Wang, Z., 2015. Performance of detrending models of
888 crop yield risk assessment: evaluation on real and hypothetical yield data. *Stoch.*
889 *Environ. Res. Risk Assess.* 29, 109-117.

890 You, L., Wood, S., Wood-Sichra, U., Wu, W., 2014. Generating global crop distribution
891 maps: From census to grid. *Agric. Syst.* 127, 53-60.