



## Assessing causes of yield gaps in agricultural areas with diversity in climate and soils



Juan I. Rattalino Edreira<sup>a,\*</sup>, Spyridon Mourtzinis<sup>b</sup>, Shawn P. Conley<sup>b</sup>, Adam C. Roth<sup>b</sup>, Ignacio A. Ciampitti<sup>c</sup>, Mark A. Licht<sup>d</sup>, Hans Kandel<sup>e</sup>, Peter M. Kyveryga<sup>f</sup>, Laura E. Lindsey<sup>g</sup>, Daren S. Mueller<sup>h</sup>, Seth L. Naeve<sup>i</sup>, Emerson Nafziger<sup>j</sup>, James E. Specht<sup>a</sup>, Jordan Stanley<sup>e</sup>, Michael J. Staton<sup>k</sup>, Patricio Grassini<sup>a</sup>

<sup>a</sup> Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Lincoln, NE 68583-0915, USA

<sup>b</sup> Department of Agronomy, University of Wisconsin-Madison, Madison, WI 53706, USA

<sup>c</sup> Department of Agronomy, Kansas State University, Manhattan, KS 66506, USA

<sup>d</sup> Department of Agronomy, Iowa State University, Ames, IA 50011-1010, USA

<sup>e</sup> Department of Plant Sciences, North Dakota State University, Fargo, ND 58108-6050, USA

<sup>f</sup> Iowa Soybean Association, Ankeny, IA 50023, USA

<sup>g</sup> Department of Horticulture and Crop Science, The Ohio State University, Columbus, OH 43210, USA

<sup>h</sup> Department of Plant Pathology and Microbiology, Iowa State University, Ames, IA 50011, USA

<sup>i</sup> Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN 55108, USA

<sup>j</sup> Department of Crop Sciences, University of Illinois, Urbana, IL 61801, USA

<sup>k</sup> Michigan State University Extension, Allegan, MI 49010, USA

### ARTICLE INFO

#### Keywords:

Soybean  
Yield potential  
Yield gap  
Survey  
Spatial framework  
Interaction

### ABSTRACT

Identification of causes of gaps between yield potential and producer yields has been restricted to small geographic areas. In the present study, we developed a novel approach for identifying causes of yield gaps over large agricultural areas with diversity in climate and soils. This approach was applied to quantify and explain yield gaps in rainfed and irrigated soybean in the North-Central USA (NC USA) region, which accounts for about one third of soybean global production. Survey data on yield and management were collected from 3568 producer fields over two crop seasons and grouped into 10 technology extrapolation domains (TEDs) according to their soil, climate, and water regime. Yield potential was estimated using a combination of crop modeling and boundary functions for water productivity and compared against highest producer yields derived from the yield distribution in each TED-year. Yield gaps were calculated as the difference between yield potential and average producer yield. Explanatory factors for yield gaps were investigated by identifying management practices that were concordantly associated with high- and low-yield fields. Management  $\times$  TED interactions were then evaluated to elucidate the underlying causes of yield gaps. The chosen spatial TED framework accounted for about half of the regional variation in producer yield within the NC USA region. Across the 10 TEDs, soybean average yield potential ranged from 3.3 to 5.3 Mg ha<sup>-1</sup> for rainfed fields and from 5.3 to 5.6 Mg ha<sup>-1</sup> for irrigated fields. Highest producer yields in each TED were similar ( $\pm 12\%$ ) to the estimated yield potential. Yield gap, calculated as percentage of yield potential, was larger in rainfed (range: 15–28%) than in irrigated (range: 11–16%) soybean. Upscaled to the NC USA region, yield potential was 4.8 Mg ha<sup>-1</sup> (rainfed) and 5.7 Mg ha<sup>-1</sup> (irrigated), with a respective yield gap of 22 and 13% of yield potential. Sowing date, tillage, and in-season foliar fungicide and/or insecticide were identified as explanatory causes for yield variation in half or more of the 10 TEDs. However, the degree to which these three factors influenced producer yield varied across TEDs. Analysis of in-season weather helped interpret management  $\times$  TED interactions. For example, yield increase due to advances in sowing date was greater in TEDs with less water limitation during the pod-setting

**Abbreviations:** ETo, grass-reference evapotranspiration; ETC, crop evapotranspiration; HY, high-yield fields; I, irrigated; LY, low-yield fields; MG, cultivar maturity group; M  $\times$  E, management  $\times$  environment interaction; NC USA, North-Central United States of America; PAWHC, plant-available water holding capacity in the rootable soil depth; P95, yield potential derived from the 95th percentile of the field yield data distribution; R, rainfed; SD<sub>P5</sub>, sowing date derived from the 5th percentiles of the sowing date data distribution (i.e., earliest sowing dates); TED, technology extrapolation domain; TED  $\times$  M, TED  $\times$  management interaction; Yg, yield gap; Yp, yield potential; Yw, water-limited yield potential

\* Corresponding author.

E-mail address: [rattalino@unl.edu](mailto:rattalino@unl.edu) (J.I. Rattalino Edreira).

<http://dx.doi.org/10.1016/j.agrformet.2017.07.010>

Received 25 April 2017; Received in revised form 7 July 2017; Accepted 17 July 2017

Available online 08 August 2017

0168-1923/ © 2017 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

phase. The present study highlights the strength of combining producer survey data with a spatial framework to measure yield gaps, identify management factors explaining these gaps, and understand the biophysical drivers influencing yield responses to crop management.

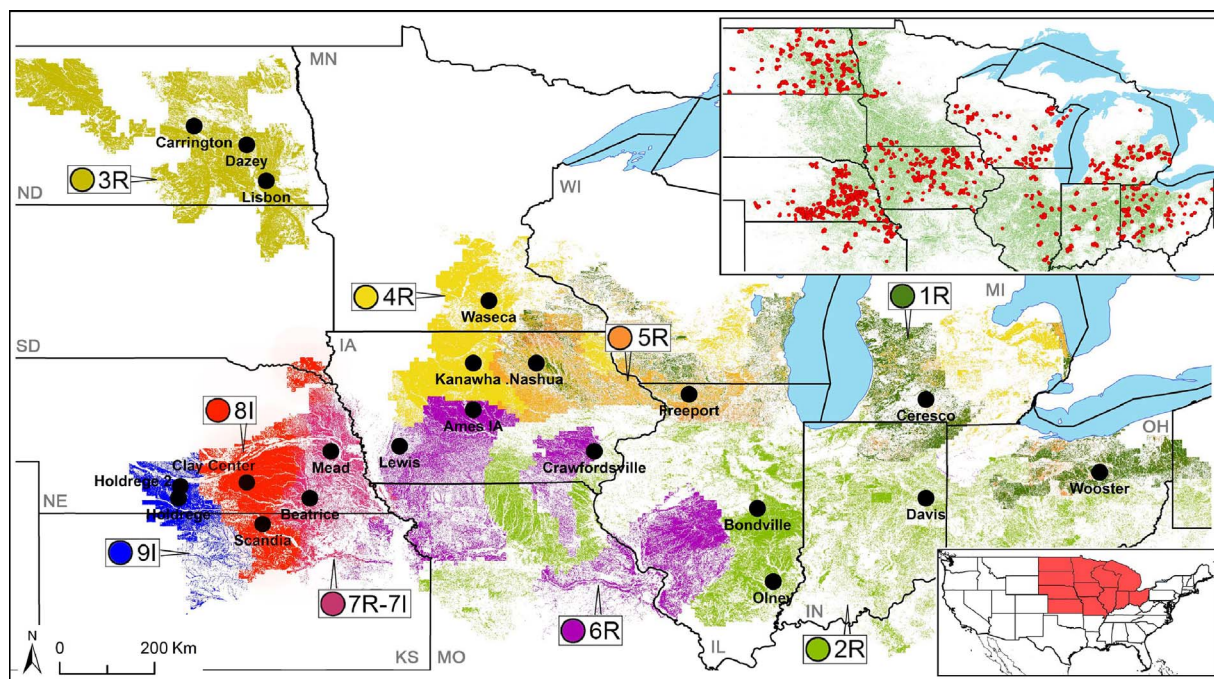
### 1. Introduction

Yield potential ( $Y_p$ ) is the yield of a crop cultivar when grown in an environment to which it is adapted, with non-limiting water and nutrient supplies, and with pests, weeds, and diseases effectively controlled (Evans, 1993; Evans and Fisher, 1999; van Ittersum and Rabbinge, 1997). In these optimal conditions, crop growth is determined by solar radiation, temperature, atmospheric  $CO_2$  concentration, and management practices which influence crop cycle duration and light interception, such as sowing date, cultivar maturity, and plant density. In rainfed systems where water supply from stored soil water at sowing and in-season rainfall is not enough to meet crop water requirement, water-limited yield potential ( $Y_w$ ) is determined by water supply amount and its distribution during the growing season, and by soil properties influencing the crop water balance, such as rootable soil depth, available-water holding capacity, and terrain slope (van Ittersum et al., 2013). Crop simulation models, boundary functions defining maximum yield for a given level of resource availability, and measured yields in highest-yielding farmer’s fields have been used to estimate  $Y_p$  and  $Y_w$  (Sadras et al., 2015; van Ittersum et al., 2013). The difference between  $Y_p$  (or  $Y_w$  in rainfed conditions) and producer average yield is termed the yield gap ( $Y_g$ ). Closing the  $Y_g$  via a fine-tuning of current management practices provides an opportunity to increase crop production on existing cropland (Cassman et al., 2003; van Ittersum et al., 2013).

The most common approach for assessing the magnitude and causes of  $Y_g$  in localized areas involves conducting controlled research trials in which researchers experimentally evaluate various input levels or management practices to identify whether a particular input or practice improve yield, and if the degree of yield improvement justifies input

costs (Lollato and Edwards, 2015; Salvagiotti et al., 2008; Yang et al., 2004). However, assessing the causes of  $Y_g$  over large geographic regions has been an elusive goal for three main reasons. First, it is difficult and costly to run field experiments to evaluate each potential factor that might limit producer yields. Second, it is problematic to extrapolate results from these localized experiments to far-flung producer fields, especially if there is lack of an appropriate description of the biophysical environment (e.g., climate, soil) where these experiments are conducted. Finally, even with a large number of site-year experiments, management  $\times$  environment ( $M \times E$ ) interactions are difficult to interpret without a rational understanding of what the word “environment” means beyond “site” and “year”. Consequently, most studies addressing the causes of  $Y_g$  through on-farm trials have been confined to small geographic areas where field-to-field variation in weather is small (e.g., Kravchenko et al., 2017; Subedi and Ma, 2009; Villamil et al., 2012). Without an objective way to contextualize and extrapolate their findings, it remains uncertain how these local studies can help support more effective research prioritization and impact assessment of technology adoption on crop production and natural resources at local and regional scales.

The present study addresses the aforementioned limitations by proposing a novel, cost-effective approach that combines producer survey data with a robust spatial framework to identify causes of  $Y_g$  across large geographic areas. We argue that having a database containing yield and management data from producer fields across multiple regions and years, properly contextualized relative to the biophysical environment, can be considered equivalent to running hundreds of field experiments to capture both major management effects and  $M \times E$  interactions. Such analysis of large-scale producer data can provide a focus as to what treatments are the most promising to



**Fig. 1.** Map of the North-Central USA (NC USA) region showing nine technology extrapolation domains (TEDs) and meteorological stations (solid circles) selected for the present study. A coding system (from TED 1 to 9) is used to identify each TED (shown with a unique color) and its associated water regime (I: irrigated, R: rainfed). There were actually 10 TED-water regimes (denominated “TEDs” for simplicity) because rainfed and irrigated fields co-existed in TED 7 (7R and 7I, respectively). Top inset: soybean harvested area in year 2015 (green area; USDA-NASS, 2016b) and location of the 3216 surveyed soybean fields (red dots). Bottom inset: location of NC USA region of 12 states within the conterminous USA. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

evaluate in more cost-effective agronomic field trial evaluations. And while there have been examples of local studies addressing the causes of Yg using producer survey data collected from relatively small regions (e.g., Grassini et al., 2011, 2015b; Silva et al., 2016), these studies do not provide an objective way to extrapolate results and measure impact over large geographic areas.

We developed here a novel approach that combines producer-reported data and a spatial framework to identify explanatory causes of Yg over large geographic regions with diversity of climate, soils, and water regimes (rainfed and irrigated). We focused on soybean in the North-Central USA (NC USA) region, which accounts for ca. 30% of global soybean production (2010–2014 period; FAOSTAT, 2016), as a study case to provide a proof of concept on the proposed approach. Specific objectives were to evaluate the proposed approach for its ability to: (i) benchmark producer soybean yields in relation to yield potential of their fields, (ii) identify key management practices explaining Yg, and (iii) elucidate the drivers for some of the observed M × E interactions.

## 2. Material and methods

### 2.1. Study region and database

United States is the world largest soybean producer, accounting for 34% of global soybean production during the 2010–2014 time interval (FAOSTAT, 2016). About 81% of USA soybean is produced on 25.7 Mha located in the NC USA region, which includes the Corn Belt and parts of the US Great Plains (2010–2014; USDA-NASS, 2016a) (Fig. 1, bottom inset). Soybean in the NC USA region is commonly grown in rotation with maize. Average (2010–2014) soybean yield in the NC USA region was 3 Mg ha<sup>-1</sup>, yet previous studies have shown that some producers in favorable environments can attain yields around 6 Mg ha<sup>-1</sup> (Grassini

et al., 2015b; Villamil et al., 2012).

Data on soybean yield and management practices were collected over two crop seasons (2014 and 2015) from fields sown with soybean in 10 states in the NC USA region: Illinois (IL), Indiana (IN), Iowa (IA), Kansas (KS), Michigan (MI), Minnesota (MN), Ohio (OH), Nebraska (NE), North Dakota (ND), and Wisconsin (WI) (Fig. 1). Soybean producers provided data via returned surveys distributed by local crop consultants, extension educators, soybean grower boards, and Natural Resources Districts (Fig. 2). Briefly, producers were asked to report the range of average field yield across the fields sown with soybean in each year and water regime and to provide data for a number of fields that portray well that yield range. Requested data also included field location, average field yield (at 13% seed moisture content), crop management (e.g., sowing date, seeding rate, row spacing, cultivar, and tillage method), applied inputs (e.g., irrigation, nutrient fertilizer, lime, manure, and pesticides), and incidence of biotic and abiotic adversities (e.g., insect pests, diseases, weeds, hail, waterlogging, and frost). Most surveyed fields were rainfed (82% of total fields), except for those in NE where rainfed (34% of NE collected fields) and irrigated (66%) production co-exist within the same geographic area. Maize was the predominant prior crop (88% of total fields), except for a few fields where soybean was grown after wheat (5%) or soybean (4%).

### 2.2. Data quality assessment

Survey data were inputted into a digital database and screened to remove erroneous or incomplete data entries. We were interested in yield variation as related with management factors; hence, a few fields with extremely low yield due to incidence of unmanageable production site adversities (hail, waterlogging, wind, and frost) were excluded from the analyses. The procedure to exclude these fields consisted on three steps: (i) grouping fields within regions with similar soil and climate

	EXAMPLE:	2014 Soybean	2014 Soybean	2015 Soybean	2015 Soybean
Specify field location by Section: Township: Range. →	NE ¼ 25 : 20N : 26W				
Please sketch-in the boundaries of your field location within the Section →					
OR GPS coordinates of field centroid: OR County & field location relative to Rd Intersection:	41.678, -100.257 Saunders Co, SW of Rd 11 & N				
Dryland? OR Pivot, Gravity? Indicate field size (acres)	Pivot (130 ac)	Pivot (137ac)	Gravity (30ac)	Pivot (84ac)	Dryland (111ac)
Does this field have drainage? (no, old clay tile, new systematic tile, surface drainage, other)	No	No	No	No	No
Total Inches of Irrigation Applied to crop?	5 inches	3.5 in.	4.5 in	3.5 in	0
SOYBEAN YIELD (bushels/acre) for this FIELD:	70	80	70	85	39
Lowest   Highest Yield (bu/ac) of your soy fields that year *Use Irrigated fields yield range if this crop was Irrigated: *Use Dryland fields yield range if this crop was Dryland:	Low: 62 High: 80	Low: 61 High: 90	Low: 55 High: 86	Low: 61 High: 90	Low: 13 High: 66
Planting Date in this FIELD (Month/Day/Year):	5/15/2014	4/28/2014	5/2/2014	4/29/2014	5/14/2014
Variety Name (Brand & Number):	Pioneer P93M11	Channel 3402 RR2	Channel 3402 RR2	Channel 2402 RR2	Channel
Seeding Rate (seeds/ac):	125,000	140,000	140,000	140,000	140,000
Row spacing (inches):	30	30	30	15	15
Seed Treated (Yes/No)? What Brand Name Product(s)?	Yes (Cruiser-Max)	yes Accelaron	yes Accelaron	yes Accelaron	yes Accelaron
Prior Crop in this FIELD? Residue harvested or grazed?	Corn - Grazed	Corn - Grazed	Corn - No	Corn - Grazed	Corn - No
Tillage after prior crop? No-Till (NT); Ridge (RT); Strip (ST); Disk (D); Chisel (C); Vertical (V) – Indicate timing (month-year)	ST (March-2014)	NT	D (April 2014)	NT	NT
Any (non-starter) fertilizer after prior crop? Specify rate (pounds NUTRIENT/ac) and timing (month-year)	P <sub>2</sub> O <sub>5</sub> : 70 K <sub>2</sub> O: 30 Other: S (11 lbs) Time: March-2014	P <sub>2</sub> O <sub>5</sub> : K <sub>2</sub> O: Other: None Time:	P <sub>2</sub> O <sub>5</sub> : K <sub>2</sub> O: Other: None Time:	P <sub>2</sub> O <sub>5</sub> : K <sub>2</sub> O: Other: None Time:	P <sub>2</sub> O <sub>5</sub> : K <sub>2</sub> O: Other: None Time:
Any STARTER fertilizer (Yes/No)? If Yes, specify nutrients	Yes (N, P, Zn)	No	No	No	No
Any Lime (L) or Manure (M)? If yes, specify timing (mm-yy)	M (Nov-2013)	No	No	No	No
PRE- or POST-emergence herbicide program or BOTH?	Both	Both	Both	Both	Both
Any in-season foliar fungicide (F) / insecticide (I)?	F and I	No	No	No	No
Soy Cyst Nematodes (Yes/No/I don't know)?	No	No	No	No	No
Iron Deficiency Chlorosis (Yes/No)?	No	No	No	No	No
Any significant yield loss due to Insects, Diseases, Weeds, Frost, Hail, Flood, Lodging? Specify problem	Frost (Sept-2014)	None	None	None	yes Hail (July 2014)

Fig. 2. Example of an actual survey form filled out by a Nebraska soybean producer, providing information for three irrigated fields and one rainfed field sown with soybean in 2014 and 2015. This survey was used to collect information from producer fields across 10 states in the North-Central USA region. Note that producer name is not shown and field location was hatched in order to keep personal information confidential.

(see Section 2.3), (ii) selecting fields within the 25th percentile of yield data distribution within each region-year, and (iii) excluding fields affected by any of the aforementioned adversities reported by producers. Because producers tended to overestimate the impact of adversities on average field yield, even when a very small portion of the field was affected, the aforementioned protocol helped distinguish fields with substantial yield losses due to the reported adversity from other fields where yield loss was negligible. After quality control, the database contained data from a total of 3216 fields sown with soybean in 2014 and 2015 (92% of total surveyed fields). A full detailed description of the database is available at: <http://cropwatch.unl.edu/2016-soybean-survey>.

To assess quality of the producer self-reported data, database yields were compared against estimated county-level yield data independently collected by USDA-NASS (<http://quickstats.nass.usda.gov/>). Annual average irrigated and rainfed soybean yields reported by USDA-NASS were retrieved for the 2014 and 2015 crop seasons for the counties that overlap with locations of surveyed fields (Fig. 3). Agreement between yield data sources was evaluated by calculating root mean square error (RMSE) and absolute mean error (ME) as follows:

$$RMSE = \sqrt{\frac{\sum (Y_{PR} - Y_{NASS})^2}{n}} \tag{1}$$

$$ME = \frac{\sum (Y_{PR} - Y_{NASS})}{n} \tag{2}$$

where  $Y_{PR}$  and  $Y_{NASS}$  are the producer-reported yield average and the USDA-NASS county yield average, and  $n$  is the number of pairs of  $Y_{PR}$  and  $Y_{NASS}$ . RMSE was also calculated as percentage (RMSE%) of the mean producer-reported yield. Linear regression analysis was performed to assess any deviation in the regression of producer data yields versus USDA-NASS yields. Confidence intervals and  $t$ -tests were used to detect statistically significant departures of the slope and intercept estimates from null hypothesized values of unity and zero, respectively. Also, paired  $t$ -tests were conducted to detect significant differences between producer data yield and USDA-NASS estimated yields.

The analysis indicated that when averaged over all site-years, the county means for producer-reported yield ( $3.4 \text{ Mg ha}^{-1}$ ) were slightly higher (9%,  $p < 0.01$ ) than the mean of the USDA-NASS yields ( $3.1 \text{ Mg ha}^{-1}$ ). However, the high coefficient of determination ( $r^2 = 0.79$ ) and a slope value undistinguishable from one ( $p = 0.26$ ) between the producer-reported and USDA-NASS yields indicated that the 3216 field-year database was reliably representative of the wide range of soybean yields in the NC USA region, ranging from  $1.5$  to  $5.2 \text{ Mg ha}^{-1}$  across counties, years, and water regimes.

### 2.3. Categorization of fields based on their biophysical context

A challenge is how to cluster producer fields in order to identify management factors that consistently lead to higher yields for a given climate-soil combination. In the present study, surveyed fields were grouped based upon their climate and soil using the spatial framework developed for the central and eastern USA by the Global Yield Gap Atlas (<http://www.yieldgap.org>; van Wart et al., 2013). This framework delineates regions [hereafter called technology extrapolation domains (TEDs)] based on four biophysical attributes that govern crop yield and its inter-annual variability: (i) annual total growing degree-days, which, in large part, determines the length of crop growing season (10 classes), (ii) aridity index, which largely defines the degree of water limitation in rainfed cropping systems (10 classes), (iii) annual temperature seasonality, which differentiates between temperate and tropical climates (3 classes), and (iv) plant-available water holding capacity in the rootable soil depth (PAWHC), which determines the ability of the soil to supply water to support crop growth during rain-free periods (10 classes; 50-mm class interval). Each TED corresponds to a specific

combination of growing-degree days, aridity index, temperature seasonality, and plant-available water holding capacity. Detailed description of TEDs is available at: <http://www.yieldgap.org/web/guest/cz-ted>

We selected TEDs that best portrayed the diversity of climate, soils, and water regimes in the NC USA region (Fig. 1). Six TEDs included only rainfed soybean fields (1R, 2R, 3R, 4R, 5R, and 6R) while two TEDs included only irrigated soybean fields (8I and 9I). One TED included both irrigated and rainfed soybean fields (7I and 7R). Because the impact of management factors on yield is influenced by water supply (e.g., Grassini et al., 2015b; Heatherly, 1988), we separated water regimes (WR; rainfed and irrigated) within the same TED. Hence, a total of 10 TED-WR combinations were eventually used in this study, which are referred hereafter as ‘TEDs’ for simplicity (total of 10 TEDs). Selected TEDs included 38% of the surveyed fields (1343 fields, 38% of the total) and accounted for 25 and 45% of USA rainfed and irrigated soybean area, respectively. Each individual TED contained  $\geq 98$  (rainfed) and  $\geq 59$  (irrigated) surveyed fields (including both years), with an average of 137 fields per TED (Table S1). *Ex-ante* power analysis indicated that the number of fields within each TED was sufficient to detect relatively small yield differences (ca.  $200 \text{ kg ha}^{-1}$ ) attributable to management factors. The lower threshold used in irrigated (59) versus rainfed (98) is justified by the smaller field-to-field yield variation in irrigated fields within the same TED. To assess the degree to which TEDs were able to discriminate amongst biophysical environments and their consistency over years, two-way analysis of variance (ANOVA) was conducted to examine the partitioning of sum of squares amongst year, TED, and TED  $\times$  year sources of variation relative to yield and management practices. The residual variation was taken as a measure of the field-to-field variability within TED.

### 2.4. Estimation of soybean yield potential and yield gap

Annual yield potential ( $Y_p$ ) and water-limited yield potential ( $Y_w$ ) were estimated using measured daily weather data (including solar radiation, rainfall, and maximum and minimum air temperature) collected at 2–3 meteorological stations located within each TED, preferably in proximity to the areas with highest density of surveyed fields (Fig. 1). Previous assessments on the variation of  $Y_p$  and  $Y_w$  within TEDs indicates that the number of weather stations used in the present study was sufficient for a robust estimation of both parameters (Hochman et al., 2016; van Wart et al., 2013). Likewise, our analysis

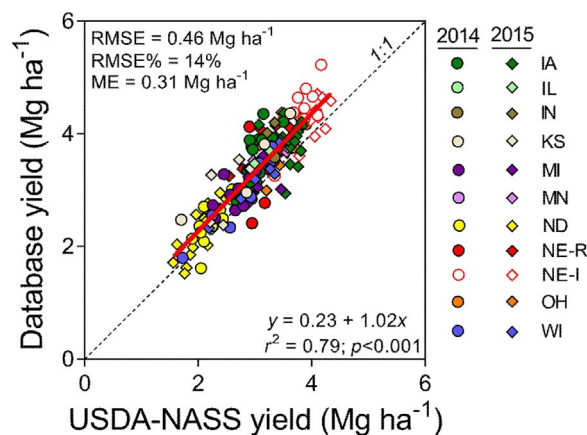


Fig. 3. Comparison between producer-reported yield and USDA-NASS yields in ten NC USA states. Each datapoint corresponds to the average yield for a given county-year combination (For Nebraska, R: rainfed; I: irrigated). The 1:1 line (dashed black line), fitted linear regression (red solid line), root mean square error (RMSE), RMSE as percentage of mean database yield (RMSE%), and absolute mean error (ME) are also shown. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

indicated that there was a very little variation in simulated yield ( $Y_p$  or  $Y_w$ ) among weather stations located within the same TED, even in large TEDs such as 2R and 6R (coefficient of variation = 6% and 7%, respectively). Hence, our estimates of yield potential based on 2–3 weather stations per TED can be considered robust. The stations are managed by MESONET state-operated networks (<http://mrcc.isws.illinois.edu/gismaps/mesonets.htm>). Details on weather data sources and quality control can be found in Morell et al. (2016) and Mourtzinis et al. (2017).  $Y_p$  and  $Y_w$  were calculated for each of the 2–3 locations within each TED, and then averaged to calculate a single  $Y_w$  and  $Y_p$  for that given TED, separately for each year (2014 and 2015).

For irrigated fields,  $Y_p$  was calculated using a well-validated soybean simulation model (SoySim; Setiyono et al., 2010) based on daily weather data and reported sowing date, variety maturity group<sup>1</sup> (MG), and seeding rate (Table S1). Because early sowing date is critical to achieve high soybean yields in the USA Corn Belt region (Bastidas et al., 2008; De Bruin and Pedersen, 2008; Egli and Cornelius, 2009),  $Y_p$  was simulated using an early sowing date, which was calculated from the 5th percentile of the producer sowing date distribution for each TED-year (Table S1). Average reported seeding rate and MG for each TED were used for the simulations because (i) producer seeding rates largely exceeded seeding rate needed to maximize yield (De Bruin and Pedersen, 2009; Grassini et al., 2015b and references cited therein), and (ii) there was a very narrow range of MG within TEDs (typically less than one unit).

For rainfed fields, a boundary function relating soybean seed yield and seasonal water supply reported by Grassini et al. (2015b) was used to determine  $Y_w$ . Boundary functions have been widely used for yield-gap analysis (Passioura and Angus, 2010; Sadras and Angus, 2006). The boundary function had a slope (attainable water productivity) of  $9.9 \text{ kg mm}^{-1} \text{ ha}^{-1}$  and x-intercept (seasonal soil evaporation) of 73 mm. Seasonal water supply was calculated as the sum of available soil water at sowing in the upper 1.5 m soil depth and in-season precipitation from sowing to physiological maturity (soybean stage R7; Fehr et al., 1971). Available soil water at sowing in the upper 1.5 m soil depth was determined dynamically using the Hybrid Maize model (Yang et al., 2017; Yang et al., 2004) by initializing the model run at harvest of the prior maize crop (i.e., about 6 months before soybean sowing), assuming 50% of available soil water content at that time, and measured weather data from that prior harvest to soybean sowing. The choice of 50% soil water content at harvest of prior maize crop was supported by data reported by a previous simulation study conducted in the USA Corn Belt (Grassini et al., 2009). In-season precipitation was calculated for the time interval between sowing date and the calendar date of R7 stage as simulated using SoySim model.

$Y_w$  and  $Y_p$  were used as benchmarks for calculating  $Y_g$  for rainfed (TEDs 1R, 2R, 3R, 4R, 5R, 6R, and 7R) and for irrigated TEDs (7I, 8I, and 9I). The  $Y_g$  was calculated as the difference between  $Y_p$  (or  $Y_w$ ) and average producer yield and expressed as percentage of  $Y_p$  (irrigated) or  $Y_w$  (rainfed). For rainfed TED-year cases in which  $Y_w \geq Y_p$ , rainfed crops were assumed not to be limited by water supply; in those cases, simulated  $Y_p$  was taken as an estimate of yield potential and used to calculate the  $Y_g$  for rainfed crops. Finally,  $Y_w$  (or  $Y_p$ ) and average producer yield were upscaled to the entire NC USA region based on the values calculated for each TED (Table S1), weighted by the relative contribution of each TED to the regional soybean harvested area (USDA-NASS, 2010–2014). The upscaling was performed separately for irrigated and rainfed TEDs.

Lobell et al. (2009) and van Ittersum et al. (2013) have shown that, in high-input cropping systems without severe water limitations, highest producer yields for a given year and region can be taken as

rough estimates of  $Y_p$  (or  $Y_w$ ). To evaluate the robustness of the approach used in the present study for calculating  $Y_g$ , we compared our estimates of  $Y_w$  (or  $Y_p$ ) derived from crop modeling and boundary functions against independent estimates of yield potential derived from the 95th percentile of the field yield distribution (P95) for each TED and year. Agreement in yield potential calculated using the two independent approaches was assessed using RMSE, ME, and RMSE%.

Weather data and simulated crop stages were used to compute means of meteorological factors (incident solar radiation, and maximum and minimum temperature) for four different crop phases: early vegetative phase, late vegetative phase, pod-setting, and seed-filling. Pod-setting was defined as the period between beginning of pod-setting (R3 stage, Fehr and Caviness, 1977) and beginning of seed-filling (R5 stage). Seed-filling was defined as the time interval between R5 and physiological maturity (R7 stage). The period between sowing and R3 was divided into two equal parts, with the mid-point corresponding roughly to the first flower (R1 stage). For the indeterminate cultivars grown in the NC USA, the vegetative period overlaps with the R1 to R2 reproductive period of flowering. An apparent water balance was also calculated for each phase as the difference between total rainfall and simulated non-water limiting crop evapotranspiration (ETc). A negative and positive water balance values indicate an apparent water deficit and surplus, respectively. Patterns for each meteorological factor and the water balance across the different crop phases were shown for four TEDs that portrayed well the spatial variation in weather across the soybean-producing region in the NC USA region (Fig. 4). Magnitude of water deficit increased following an E-W gradient, while solar radiation followed the opposite trend. In contrast, there was a N-S temperature gradient, with southern TEDs exhibiting warmer temperatures. PAWHC ranged from 200 to 300 mm across fields located in the selected TEDs, except for TED 1R where it ranged from 100 to 150 mm (Table S1).

## 2.5. Identification of causes of yield gaps

As a first approach to identify factors explaining  $Y_g$ , high-yield (HY) and low-yield (LY) field classes were identified based on their respective presence in the upper and lower terciles of the field yield distribution within each TED. Differences in each management practice and applied input between the HY and LY fields were then evaluated for significance using *t*-tests. Association between field classes and categorical variables (e.g., artificial field drainage, seed treatment, and lime) was evaluated using Chi-square ( $\chi^2$ ) tests. For some management practices involving more than two distinguishable techniques, fields were grouped in two categories to facilitate the analysis. Following Grassini et al. (2015a,b), fields were categorized as either no-till or tilled, with the latter including chisel, disk, strip-till, ridge-till, vertical, field cultivator, and moldboard plow. Likewise, because row spacing distribution exhibited a strong bimodal shape, field were grouped into the two most common row spacing classes: narrow (38 cm) and wide (76 cm). Some practices have already been widely adopted by producers in some of the TEDs; hence, it was not possible to make comparisons when one of the alternatives for a given practice predominated, resulting in too few fields for a balanced comparison (e.g., herbicide application, seed treatment). Finally, to avoid confounding effects, fields treated with fungicide only, insecticide only, or both fungicide and insecticide were pooled for the analysis because in-season canopy fungicide and insecticide applications were commonly applied together (51% total treated fields).

Variables identified as statistically significant on their influence on seed yield, as revealed from comparison between HY versus LY fields, were further investigated. Quantile regression was used to derive a boundary function for the relationship between producer yield and sowing date delay based on the 90th percentile using the *quantreg* package in R (R Development Core Team, 2016). For categorical variables (e.g., tillage, artificial drainage, pesticide application), average yields calculated for contrasting management categories were

<sup>1</sup> Soybean varieties are divided into groups according to their relative times of maturity. Maturity groups are usually designated using triple zero, double zero, zero and Roman numerals from I to X for very short- and long-season varieties, respectively.

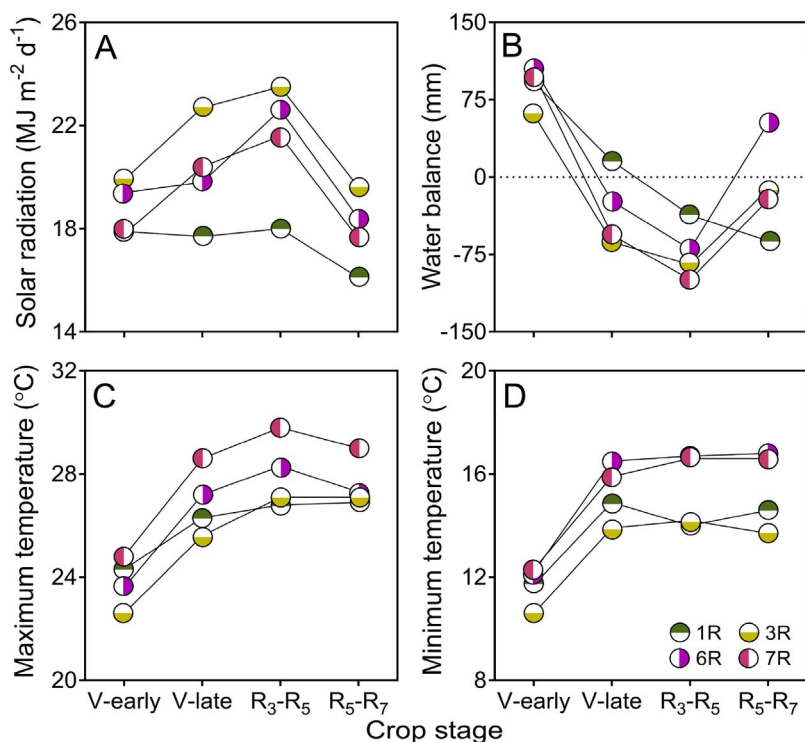


Fig. 4. Average solar radiation (A), water balance (B), and average maximum (C) and minimum air temperature (D) in four technology extrapolation domains during four crop phases: early vegetative phase (V-early), late vegetative phase and R1–R2 flowering (V-late), pod-setting (R3–R5), and seed-filling (R5–R7). Water balance was calculated as the difference between total rainfall and non-water limiting crop evapotranspiration. The four TEDs region were selected to portray the variation in weather over the North-Central USA region. Each data point corresponds to the average value for a given crop phase calculated based on 2–3 meteorological stations located within each TED and two crop seasons (2014 and 2015).

compared (e.g., no-till versus tilled fields) using paired *t*-tests. ANOVA was performed to evaluate the statistical significance of the yield impact of each management (M) practice main effect and its interaction with TED (M × TED) and year (M × Y). Finally, Pearson’s correlation analysis, based on yield responses to different management factors (dependent variable) and meteorological factors calculated for each four crop phases in each TED (independent variables), was used to investigate the biophysical basis for some of the observed M × TED interactions.

### 3. Results

#### 3.1. Sources of variation in regional yields and management practices

There was a large variation in average annual yield across TEDs, ranging from 2.6 to 4.9 Mg ha<sup>-1</sup> (Table 1). TEDs accounted for 96% of the treatment sum of squares (*i.e.*, excluding the error) and of the remaining sums of squares, the TED × Y interaction explained at least three times more than the contribution of year. These findings were consistent with observed differences in seasonal weather patterns among TEDs (Fig. 4), and similarities in weather (and yield) between the two crop seasons within each TED (Table S2). Overall, these findings indicated that the TED framework was robust at capturing the influence of key biophysical factors on crop yield *per se*, and was 31 × more explanatory than the TED × Y interaction. This analysis indicates that the TED framework can be used to delineate climate-soil domains that predictably account for seed yield potential. This finding also can be extended to some key agronomic practices for which the ten TEDs account for 50–99% of the variation in the producer choices of tillage, MG, sowing date, foliar pesticide, and row spacing.

#### 3.2. Soybean yield potential and yield gap in the NC USA region

The two independent estimates of yield potential (Y<sub>p</sub> or Y<sub>w</sub> versus P95) compared reasonably well, with RMSE of 0.29 Mg ha<sup>-1</sup>, representing 6% of average Y<sub>w</sub> or Y<sub>p</sub> (Fig. 5A). Average difference in yield potential estimated using the two approaches (–0.03 Mg ha<sup>-1</sup>)

was not statistically different from zero for both rainfed and irrigated crops (*t*-test, *p* > 0.60). In all cases, the P95 value derived from the yield distribution was within ± 12% of simulated Y<sub>p</sub> or Y<sub>w</sub>. Similarity in yield potential estimated by the two independent approaches was consistent across the entire range of yields, indicating that our Y<sub>w</sub> (or Y<sub>p</sub>) estimates were robust and can be reliably used as benchmarks for estimating Y<sub>g</sub> for soybean fields across the NC USA region.

Table 1

Analysis of variance for seed yield, tillage, variety maturity group, sowing date, in-season foliar fungicide and/or insecticide, and row spacing reported for producer fields sown with soybean during 2014–2015 years (Y) in ten technology extrapolation domains (TEDs) located to the NC USA region.

Variable (and units)	Source	Degrees of freedom	Sum of squares	% SS <sup>a</sup>	<i>p</i> -value
Seed yield (Mg ha <sup>-1</sup> )	TED	9	631	96%	< 0.01
	Y	1	4	1%	< 0.01
	TED × Y	9	19	3%	< 0.01
	Residual	1353	505		
Tillage (% tilled fields)	TED	9	42	87%	< 0.01
	Y	1	1	3%	0.01
	TED × Y	9	5	10%	< 0.01
	Residual	1338	290		
Maturity group (unitless)	TED	9	772	99%	< 0.01
	Y	1	1	< 1%	0.01
	TED × Y	9	1	< 1%	0.87
	Residual	1228	188		
Sowing date (day of year)	TED	9	22616	50%	< 0.01
	Y	1	129	< 1%	0.32
	TED × Y	9	22943	50%	< 0.01
	Residual	1310	132814		
Foliar fungicide and/or insecticide (% treated fields)	TED	9	23	82%	< 0.01
	Y	1	0	< 1%	0.38
	TED × Y	9	5	17%	< 0.01
	Residual	1353	297		
Row spacing (% wide-row fields)	TED	9	28	90%	< 0.01
	Y	1	0	< 1%	0.5
	TED × Y	9	3	9%	0.2
	Residual	920	204		

<sup>a</sup> %SS: proportion of sum of squares relative to the non-error total sum of squares.

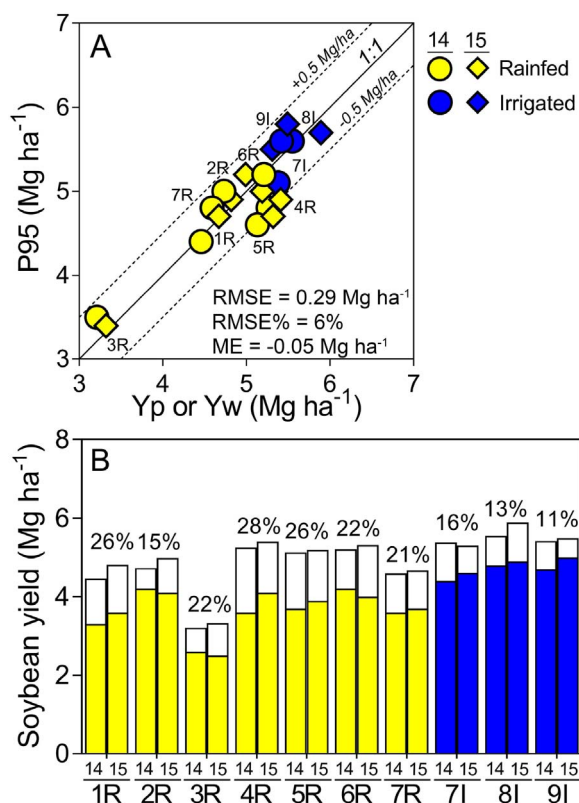


Fig. 5. (A) Comparison between producer yield derived from the 95th percentile of the yield distribution (P95) versus crop model estimates of yield potential for irrigated soybean (Yp, blue symbols) and water-limited yield potential for rainfed soybean (Yw, yellow symbols) across 10 technology extrapolation domains (TEDs) in 2014 (14) and 2015 (15). (B) Yield potential for rainfed (Yw) and irrigated (Yp) soybean in each of the 10 TEDs in 2014 (14) and 2015 (15). Solid and empty portions of the bars represent the average producer yield and yield gap, respectively. Values on top of the bars indicate the (2-year) average Yg, expressed as percentage of Yw (rainfed) or Yp (irrigated). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Average Yw ranged from 3.2 to 5.4 Mg ha<sup>-1</sup>, while Yp varied from 5.4 to 6.1 Mg ha<sup>-1</sup> across TEDs (Fig. 5B). TED 3R exhibited the lowest Yw due to lower seasonal precipitation in relation with other TEDs (Fig. 4). In contrast, Yp was highest in TED 8I due to non-limiting water supply and high incident solar radiation. Upscaled to the entire NC USA region, Yw and Yp averaged 4.8 and 5.7 Mg ha<sup>-1</sup>, respectively. Average producer yield was consistently lower than Yw (or Yp) across all TEDs ( $p < 0.01$ ). Yield gaps, expressed as percentage of Yp (irrigated) or Yw (rainfed), tended to be larger in rainfed (range: 15–28%) than in irrigated crops in relation with size of Yg persisted in TED 7, where rainfed and irrigated fields are located adjacent to each other within the same geographic region (21 versus 16% of Yw and Yp, respectively). At regional level, the rainfed Yg averaged 22% in contrast to the irrigated Yg of 13%.

### 3.3. Underpinning causes for yield variation among fields within TED

Analysis of management practices allowed identification of candidate factors explaining Yg in each TED (Table 2, Table S3). Differences in sowing date, tillage, in-season foliar fungicide and/or insecticide, and MG between HY and LY fields were statistically significant in half or more of the 10 TEDs ( $p < 0.10$ ). Sowing date had the most consistent impact on soybean yield (Fig. 6). HY fields were sown, on average, 7 days earlier than LY fields in both irrigated and rainfed conditions (Table 2). There was a strong sowing date × TED interaction on yield as indicated by the wide range in yield penalty across TEDs,

ranging from  $-1$  to  $-33$  kg ha<sup>-1</sup> d<sup>-1</sup> (Fig. 6). Although differences in variety MG between HY and LY were less than one unit, there was a consistent trend towards shorter-season MGs in the HY field tercile in all TEDs, except for those located in the northern fringe of the NC USA region (3R and 4R).

Similar to sowing date, other management practices also exhibited a significant M × TED interaction (Table 2, Fig. 7). While there was an overall statistically positive impact of foliar fungicide and/or insecticide ( $0.31$  Mg ha<sup>-1</sup>,  $p < 0.01$ ) and artificial drainage ( $0.18$  Mg ha<sup>-1</sup>,  $p = 0.05$ ) on soybean seed yield, the magnitude of these yield differences were not consistent across TEDs, and not even significant in some of them (Table 2, Fig. 7). For example, average yield of fields treated with foliar fungicide and/or insecticide was  $0.75$  Mg ha<sup>-1</sup> higher in relation with untreated fields in TED 7R, but this yield difference was negligible ( $-0.06$  Mg ha<sup>-1</sup>) and not statistically significant in TED 6R. Likewise, artificially drained fields achieved statistically higher yields compared with fields without artificial drainage in only 2 of 6 TEDs. Consistent with these observations, the M × TED term was significant for foliar fungicide and/or insecticide and artificial drainage, explaining a larger portion of the treatment sum of squares in relation to management and M × Y interaction (Fig. 7). We did not find evidence of no-till fields outperforming yield of tilled fields in every TED; indeed, tilled fields yielded significantly more in half of the TEDs ( $0.15$  Mg ha<sup>-1</sup>,  $p = 0.02$ ) (Fig. 7). Still, there may be reasons for producers to adopt no-till despite the observed yield penalty. For example, no-till can help control soil erosion and reduce irrigation water requirements. Indeed, we found that, on average, total irrigation was 65 mm less in no-till versus tilled fields ( $p < 0.01$ ). Hence, differences in irrigation between HY and LY fields observed for 2 of the 3 irrigated TEDs are likely to be the result of lower adoption of no-till in HY fields relative to LY fields (Table 2).

In contrast to the aforementioned variables, there were inconsistent (and generally small) differences between HY and LY fields in relation with row spacing, seeding rate, seed treatment, nutrient (N, P, K) fertilizer application, lime, and manure (Table 2, Table S3). Lack of statistically significant differences between management practices should be interpreted with caution. For example, some practices might influence yield depending upon the level of another management practice [e.g., seed treatment in relation with sowing date (Gaspar and Conley, 2015)]. Likewise, the benefit of other practices may only be realized in crop seasons with unfavorable weather, which was not the case in our study (e.g., narrow row spacing, no-till; Taylor (1980); Wilhelm and Wortmann (2004)). Similarly, yield impact of some practices may be masked by other field variables not accounted here. For example, lack of yield differences between fields that received fertilizer application versus those that did not receive fertilizer might reflect producer tendency to apply fertilizer only in fields where soil nutrient status is inadequate as evaluated using soil nutrient tests. It may also reflect many producers over-fertilizing the previous maize crop, expecting the subsequent soybean crop to benefit from the residual soil fertility. Finally, there are management practices that exhibited a very narrow range (e.g., MG) or inputs that were applied in amounts well above their optimums. For example, on-farm average soybean seeding rate ranged from 36 to 42 plants m<sup>-2</sup> across TEDs. These densities are higher than the required plant density for maximum yields (25–35 plants m<sup>-2</sup>) (Grassini et al., 2015a); hence, our analysis does not fully capture the influence of these management factors on seed yield.

### 3.4. Interpretation of M × E interactions

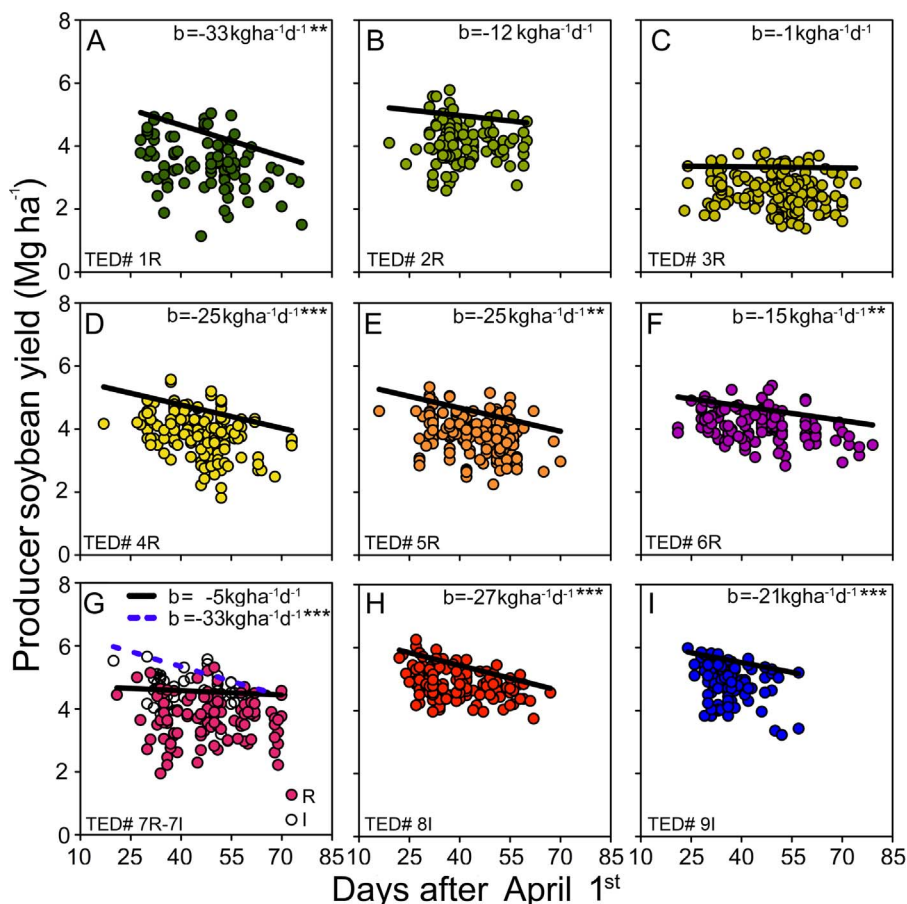
Assessment of the observed TED × M interactions, in relation to weather dynamics during the growing season, revealed a relationship between yield response to sowing date and the degree of water deficit during pod-setting (R3–R5) phase (Fig. 8). Yield penalty (or response) to sowing date was negligible when water balance was  $< -100$  mm, but increased linearly up to nearly  $-40$  mm. Yield response to sowing

**Table 2**

Comparison of producer soybean yield, management practices, and applied inputs between the highest terciles of field yields (HY) and the lowest terciles (LY) in 10 technology extrapolation domains (TEDs) in the NC USA region. Values indicate the mean differences (HY – LY) between the upper and lower yield terciles. Means for each variable in the HY and LY field categories are shown in Table S3.

Variables	Units	NC USA TEDs (see Fig. 1)									
		1R	2R	3R	4R	5R	6R	7R	7I	8I	9I
Seed yield	Mg ha <sup>-1</sup>	1.7 <sup>***</sup>	1.4 <sup>***</sup>	1.2 <sup>***</sup>	1.5 <sup>***</sup>	1.4 <sup>***</sup>	1.2 <sup>***</sup>	1.6 <sup>***</sup>	1.1 <sup>***</sup>	1.1 <sup>***</sup>	1.3 <sup>***</sup>
Field management											
Artificial drainage	% drained fields	12	18	19 <sup>**</sup>	7	-6	11	n.c.	n.c.	n.c.	n.c.
Tillage	% tilled fields	-3	31 <sup>**</sup>	20 <sup>**</sup>	25 <sup>***</sup>	10	3	-1	22	20 <sup>*</sup>	20 <sup>*</sup>
Crop management											
Sowing date	days	-10 <sup>***</sup>	-3	-4 <sup>*</sup>	-8 <sup>***</sup>	-8 <sup>***</sup>	-12 <sup>***</sup>	-6	-9 <sup>**</sup>	-10 <sup>***</sup>	-4 <sup>***</sup>
Row spacing	% wide-row fields	11	-3	-19	20 <sup>*</sup>	3	16	-9	14	-3	14
Seeding rate	seeds m <sup>-2</sup>	-2	-3 <sup>***</sup>	1	-1	0	0	-1	2 <sup>**</sup>	0	3 <sup>***</sup>
Maturity group	unitless	-0.2 <sup>*</sup>	-0.1	0.2 <sup>***</sup>	0.1 <sup>*</sup>	0	-0.3 <sup>**</sup>	-0.2 <sup>**</sup>	-0.6 <sup>***</sup>	-0.3 <sup>***</sup>	-0.1
Applied inputs											
ST fungicide	% ST fields	n.c.	n.c.	11	3	-3	2	2	36 <sup>*</sup>	11	6
ST insecticide	% ST fields	n.c.	n.c.	5	-5	0	23 <sup>***</sup>	15	n.c.	19	-7
ST nematocide	% ST fields	n.c.	n.c.	n.c.	6	10	31 <sup>***</sup>	n.c.	6	n.c.	n.c.
ST growth regulator	% ST fields	n.c.	n.c.	5	4	-24 <sup>*</sup>	13	-15	n.c.	-3	-1
ST inoculant	% ST fields	n.c.	n.c.	4	12	-9	8	-28 <sup>**</sup>	n.c.	n.c.	n.c.
Starter fertilizer	% treated fields	-10	n.c.	-9	3	-4	n.c.	8	n.c.	12	-7
Lime	% treated fields	-11	-13	n.c.	6	4	-3	19 <sup>*</sup>	-11	n.c.	n.c.
Manure	% treated fields	12	15	n.c.	-5	-2	21 <sup>***</sup>	0	5	n.c.	n.c.
Fungicide and/or insecticide	% FT fields	10	13	25 <sup>***</sup>	31 <sup>***</sup>	22 <sup>**</sup>	-13	39 <sup>***</sup>	44 <sup>***</sup>	20 <sup>**</sup>	31 <sup>***</sup>
P fertilizer	kg ha <sup>-1</sup>	10	-2	-4	8	-1	8	4	-1	6	0
K fertilizer	kg ha <sup>-1</sup>	0	-38 <sup>**</sup>	0	11	-1	17	3	n.c.	-4	n.c.
Irrigation amount	mm	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.	7	27 <sup>***</sup>	61 <sup>*</sup>

ST: seed treated. FT: in-season foliar treated. Asterisks indicate significance at  $p < 0.1$ (\*),  $p < 0.05$ (\*\*), and  $p < 0.01$ (\*\*\*). In some cases, comparison between HY and LY fields was not calculated (n.c.) because the presence of a given practice exceeded 95% of the fields, and thus was not suitable for a reasonable comparisons, or because the practice was related with a specific water regime (e.g., irrigation amount).



**Fig. 6.** Producer soybean yield plotted against sowing date in 10 technology extrapolation domains (TED) in the NC USA region, including rainfed (A–G) and irrigated (G–I) production areas. Solid line corresponds to the fitted boundary function using quantile regression (percentile 90th). Separate boundaries were derived for rainfed (empty symbols) and irrigated (solid symbols) soybean fields in TED7. Slope of the fitted boundary function (b) is shown, with asterisks indicating significance at  $p < 0.1$ \*,  $p < 0.05$ \*\*, and  $p < 0.01$ \*\*\* for the null hypothesis of  $b = 0$ .



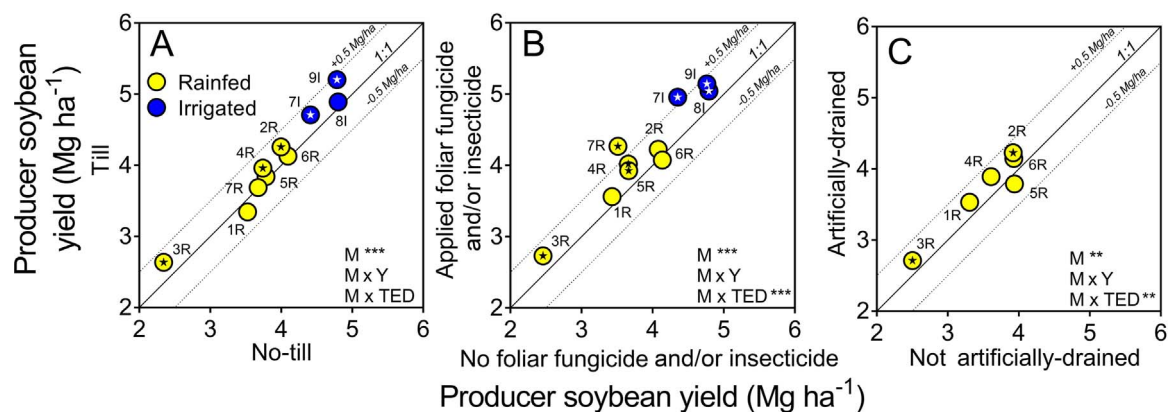


Fig. 7. Comparison of average producer soybean yield between groups of fields with different management across ten technology extrapolation domains (TEDs): (A) tillage (tilled versus no-till), (B) in-season foliar fungicide and/or insecticide (treated versus untreated fields), and (C) artificial drainage (fields with and without artificial drainage system). Star inside symbols indicate statistically significant difference for a given TED (*t*-test;  $p < 0.1$ ). Asterisks indicate significance of the impact on yield with respect to the specified management factor (M), and its interaction with year (M × Y) or with TED (M × TED) as evaluated using F-test at  $p < 0.1$  (\*),  $p < 0.05$  (\*\*), and  $p < 0.01$  (\*\*\*). Data from the two crop seasons were pooled for the analysis because M × Y influence on yield was not statistically significant. TEDs 7R, 7I, 8I, and 9I are not included in (C) because of the low number of fields with artificial drainage.

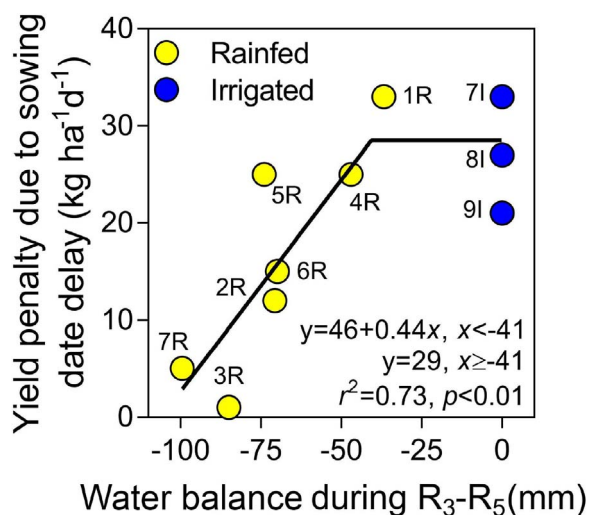


Fig. 8. Soybean yield penalty due to sowing date delay as a function of water balance during the pod-setting ( $R_3$ – $R_5$ ) phase across 10 technology extrapolation domains (TEDs), including rainfed (yellow circles) and irrigated (blue circles) production environments (averaged over 2014–2015). Water balance was estimated as the difference between rainfall and simulated non-water limiting crop evapotranspiration and set to zero for irrigated crops. Parameters of the fitted linear-plateau model (solid line) and coefficient of determination ( $r^2$ ) are shown. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

date remained relatively unchanged at water balance  $> -40$  mm, ranging from 20 to 35  $\text{kg ha}^{-1} \text{d}^{-1}$ . The role of water balance in influencing the yield response to sowing date was evident for TED 7, where irrigated and rainfed crops exhibited a six-fold difference (33 versus 5  $\text{kg ha}^{-1} \text{d}^{-1}$ , respectively) (Fig. 8). In other words, these findings indicated that yield response to sowing date diminished as the degree of water limitation in the pod-setting period of the production environment increases.

It was notable that yield response to sowing date delay exhibited much higher explanatory power with the degree of water deficit during pod-setting phase ( $r^2 = 0.73$ ,  $p < 0.01$ ) relative to the other crop phases (early vegetative phase, late vegetative phase, and seed-filling) or entire crop season ( $r^2 < 0.38$ ,  $p > 0.06$ ). This finding is consistent with the notion of sequential yield determination in field crops and further highlights the importance of a proper description of the biophysical environment in order to decipher the biophysical drivers behind observed M × TED interactions. While the analysis performed

here is a first attempt to interpret some of these interactions in relation with meteorological factors calculated for different crop phases, it is still insufficient. For example, although other management practices also exhibited a strong M × TED interaction in relation with soybean yield (e.g., foliar fungicide and/or insecticide, artificial drainage), there were no clear associations between the variation in yield response across TEDs with any meteorological factor.

#### 4. Discussion

In the present study, soybean in the NC USA region was used as a case study to test a novel approach that combines producer self-reported data, crop modeling, and a spatial framework to quantify Yg and identify explanatory causes. Our study expanded previous Yg analysis performed for relatively small geographic regions to large regions with diversity of climate and soil. With increasing pressure to monitor the productivity and environmental footprint of agricultural systems, efforts have increased to collect evermore producer field data by both private and public sectors (Antle et al., 2015; Thomson et al., 2017). This trend means that there will be opportunities to translate producer field data into useful information for producers, crop consultants, agricultural industry, and regional extension and research programs. We argue here, however, that this will be possible only if producer data are properly contextualized in relation to the climate and soil of each individual field in order to allow valid comparative tests of alternatives in each given management practice in well-defined regional environments, such that any detectable M × E interactions can be better understood and interpreted. The present study provides a first step in this direction, by providing a cost-effective approach to categorize fields using a spatial biophysical framework that accounts for major factors influencing yield, management, and their spatial variation.

While estimates of Yg and Yw (or Yp) that we report here for soybean are consistent with those reported by Grassini et al. (2015b) for a relatively small geographic area in Nebraska (USA), the present study expanded estimation of these parameters to the rest of the NC USA soybean producing region. Our spatial framework allowed upscaling of these parameters from local to regional scales, based on the site-specific yield potential and average producer yield and soybean area within each TED, resulting in an average regional Yg of 22% (rainfed) and 13% (irrigated) of the Yw and Yp, respectively. Our study also confirmed that some farmers in each TED are attaining soybean yields that were close to the yield potential of the production environment, which is consistent with previous reports for high-yield cropping systems (Grassini et al., 2014; Lobell et al., 2009). It also confirms that Yg tends

to be greater in rainfed *versus* irrigated fields, even within the same TED, which is consistent with the lower level of inputs and late sowing dates in rainfed fields reported by Grassini et al. (2015b).

While the size of the regional average Yg was relatively small, this study identified management factors that can be modulated to generate small, but still significant, yield increases in soybean production environments in the NC USA region. For example, there was a consistent effect of sowing date that explained yield variation across fields within the same TED, which is in agreement with Grassini et al. (2015b) study for Nebraska and previous experimental data (Bastidas et al., 2008 and reference cited therein; Rowntree et al., 2013). Sowing date had the most consistent impact on soybean yield, explaining, on average, 28% of total Yg across TEDs (range: 2–56%). The latter values were estimated based on the difference in attainable yield between early and average sowing dates, as derived from the boundary functions shown in Fig. 6, and comparing this yield difference against the Yg in each TED. Tillage methods, fungicide and/or insecticide application, and artificial drainage were other explanatory factors for the Yg. However, identification of 'best' management practices at regional level is complicated by the presence of TED  $\times$  M interactions (*i.e.*, the difference between two alternative methods of a given agronomic practice varies from low to high, depending on the given TED) (Figs. 6 and 7). The present study also made a first attempt to explore the biophysical drivers for some of the observed M  $\times$  TED interactions. For instance, we found that yield response to sowing date across TEDs (range:  $-1$  to  $-33$  kg ha $^{-1}$  d $^{-1}$ ) was strongly related with the degree of water deficit during the pod-setting phase (Fig. 8). Although intrinsically empirical, these relationships between yield response and simple biophysical variables are extremely useful to determine the probability and range of yield response associated with a change or adoption of a given practice in a given region (Calviño et al., 2004; Calviño et al., 2003).

Another contribution of the present study is to provide a solid basis for *ex ante* assessment of the extra crop production, at both local (TED) and regional (NC USA) levels, that would result from complete producer adoption or fine-tuning of a given management practice. For example, the potential extra production derived from earlier soybean sowing can be calculated based on the (i) specific yield response to sowing date in each TED, (ii) the degree to which the current average sowing date differs from the optimal one, and (iii) soybean harvested area in each TED. Hence, a 2-week shift towards early soybean sowing in TED 4R, from current average sowing on May 17 to a hypothetical, but realistic, May 3 sowing, would result in 0.35 Mg ha $^{-1}$  yield increase and 504,000 Mg production increase, leading to a 10% and 0.7% increase in soybean production in TED 4 and NC USA region, respectively. This example illustrates the power of this approach for impact assessment to support policy and investment prioritization and for monitoring the impact of research and extension programs.

## 5. Conclusions

Soybean Yg and its causes were assessed for the NC USA region using a novel approach that combines a spatial framework and producer self-reported data. The framework applied in this study explained the largest portion of the spatial variation in yield and management practices across the NC USA region. Soybean Yg in the NC USA were relatively small, averaging 22% (rainfed) and 13% (irrigated) of the estimated yield potential. Sowing date was the most consistent factor explaining yield variation within the same TED and year, with magnitude of yield response to sowing delay dependent upon degree of water deficit during pod-setting phase. Other practices also explained yield variation (tillage, and in-season foliar fungicide and/or insecticide, and artificial drainage), but the degree to which each of these practices influences yield depended upon TED. The combined use of producer data and a robust spatial framework that captured regional variation in weather and soils represents a cost-effective approach to identify causes of Yg across large geographic regions, which, in turn, can help inform

and strategize research and extension programs at both local and regional levels.

## Acknowledgements

Authors acknowledge the North-Central Soybean Research Program (NCSRP), Nebraska Soybean Board, and Wisconsin Soybean Marketing Board for their support to this project. We also thank UNL Extension Educators, Nebraska Natural Resource Districts, and Iowa Soybean Association for helping collect the producer data. Finally, we thank Lim Davy, Agustina Diale, Laurie Gerber, Clare Gietzel, Mariano Hernandez, Ngu Kah Hui, Caleb Novak, Juliana de Oliveira Hello, Matt Richmond, and Paige Wacker for inputting and cleaning the survey data.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.agrformet.2017.07.010>.

## References

- Antle, J., Capalbo, S., Houston, L., 2015. Using Big Data to Evaluate Agro-Environmental Policies. Choices AAEA. Available online: <http://choicesmagazine.org/choices-magazine/submitted-articles/using-big-data-to-evaluate-agro-environmental-policies>.
- Bastidas, A.M., et al., 2008. Soybean sowing date: the vegetative, reproductive, and agronomic impacts. *Crop Sci.* 48 (2).
- Calviño, P.A., Andrade, F.H., Sadras, V.O., 2003. Maize yield as affected by water availability, soil depth, and crop management. *Agron. J.* 95 (2), 275–281.
- Calviño, P., Sadras, V.O., Redolatti, M., Canepa, M., 2004. Yield responses to narrow rows as related to interception of radiation and water deficit in sunflower hybrids of varying cycle. *Field Crops Res.* 88, 261–267.
- Cassman, K.G., Dobermann, A., Walters, D.T., Yang, H., 2003. Meeting cereal demand while protecting natural resources and improving environmental quality. *Annu. Rev. Environ. Resour.* 28 (1), 315–358.
- De Bruin, J.L., Pedersen, P., 2008. Soybean seed yield response to planting date and seeding rate in the Upper Midwest. *Agron. J.* 100 (3), 696–703.
- De Bruin, J.L., Pedersen, P., 2009. New and old soybean cultivar responses to plant density and intercepted light. *Crop Sci.* 49 (6), 2225–2232.
- Egli, D.B., Cornelius, P.L., 2009. A regional analysis of the response of soybean yield to planting date. *Agron. J.* 101 (2), 330–335.
- Evans, L.T., Fisher, R.A., 1999. Yield potential: its definition, measurement, and significance. *Crop Sci.* 39 (6), 1544–1551.
- Evans, L.T., 1993. *Crop Evolution, Adaptation, and Yield*. Cambridge University Press, Cambridge, UK.
- FAOSTAT, 2016. *Crops and Livestock Trade Database*. [www.fao.org](http://www.fao.org).
- Fehr, W.R., Caviness, C.E., 1977. Stages of soybean development. Cooperative Extension Service, Agriculture and Home Economics Experiment Station. Iowa State University, Ames, Iowa.
- Fehr, W.R., Caviness, C.E., Burmood, D.T., Pennington, J.S., 1971. Stage of development descriptions for soybeans, Glycine Max (L.) Merrill. *Crop Sci.* 11 (6), 929–931.
- Gaspar, A.P., Conley, S.P., 2015. Responses of canopy reflectance, light interception, and soybean seed yield to replanting suboptimal stands. *Crop Sci.* 55 (1), 377–385.
- Grassini, P., Yang, H., Cassman, K.G., 2009. Limits to maize productivity in Western Corn Belt: a simulation analysis for fully irrigated and rainfed conditions. *Agric. For. Meteorol.* 149, 2154–2165.
- Grassini, P., Thorburn, J., Burr, C., Cassman, K.G., 2011. High-yield irrigated maize in the Western U.S. Corn Belt: I. On-farm yield, yield potential, and impact of agronomic practices. *Field Crops Res.* 120 (1), 142–150.
- Grassini, P., Torrión, J.A., Cassman, K.G., Yang, H.S., Specht, J.E., 2014. Drivers of spatial and temporal variation in soybean yield and irrigation requirements in the western US Corn Belt. *Field Crops Res.* 163, 32–46.
- Grassini, P., Specht, J.E., Tollenaar, M., Ciampitti, I.A., Cassman, K.G., 2015a. High-yield maize–soybean cropping systems in the US Corn Belt. In: Sadras, V.O., Calderini, D.F. (Eds.), *Crop Physiology. Applications for Genetic Improvement and Agronomy*. Academic Press, Oxford, pp. 15–42.
- Grassini, P., et al., 2015b. Soybean yield gaps and water productivity in the western U.S. Corn Belt. *Field Crops Res.* 179, 150–163.
- Heatherly, L.G., 1988. Planting date, row spacing, and irrigation effects on soybean grown on clay soil. *Agron. J.* 80 (2), 227–231.
- Hochman, Z., Gobett, D., Horan, H., Navarro Garcia, J., 2016. Data rich yield gap analysis of wheat in Australia. *Field Crops Res.* 197, 97–106.
- Kravchenko, A.N., Snapp, S.S., Robertson, G.P., 2017. Field-scale experiments reveal persistent yield gaps in low-input and organic cropping systems. *Proc. Natl. Acad. Sci. U. S. A.* 114 (5), 926–931.
- Lobell, D.B., Cassman, K.G., Field, C.B., 2009. Crop yield gaps: their importance, magnitudes, and causes. *Annu. Rev. Environ. Resour.* 179–204.
- Lollato, R.P., Edwards, J.T., 2015. Maximum attainable wheat yield and resource-use efficiency in the Southern Great plains. *Crop Sci.* 55 (6), 2863–2876.

- Morell, F.J., et al., 2016. Can Crop Simulation Models Be Used to Predict Local to Regional Maize Yields and Total Production in the U.S. Corn Belt?.
- Mourtzinis, S., Rattalino Edreira, J.I., Conley, S.P., Grassini, P., 2017. From grid to field: assessing quality of gridded weather data for agricultural applications. *Eur. J. Agron.* 82 (Part A), 163–172.
- Passioura, J.B., Angus, J.F., 2010. Improving productivity of crops in water-limited environments. In: Donald, L.S. (Ed.), *Advances in Agronomy*. Academic Press, pp. 37–75.
- R Development Core Team, 2016. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rowntree, S.C., et al., 2013. Genetic gain  $\times$  management interactions in soybean: I. Planting date. *Crop Sci.* 53 (3), 1128–1138.
- Sadras, V.O., Angus, J.F., 2006. Benchmarking water-use efficiency of rainfed wheat in dry environments. *Aust. J. Agric. Res.* 57 (8), 847–856.
- Sadras, V.O., et al., 2015. *Yield Gap Analysis of Rainfed and Irrigated Crops: Methods and Case Studies*. FAO, Rome, Italy.
- Salvagiotti, F., et al., 2008. Nitrogen uptake, fixation and response to fertilizer N in soybeans: a review. *Field Crops Res.* 108 (1), 1–13.
- Setiyono, T.D., et al., 2010. Simulation of soybean growth and yield in near-optimal growth conditions. *Field Crops Res.* 119 (1), 161–174.
- Silva, J.V., Reidsma, P., Laborte, A.G., van Ittersum, M.K., 2016. Explaining rice yields and yield gaps in Central Luzon, Philippines: an application of stochastic frontier analysis and crop modelling. *Eur. J. Agron.*
- Subedi, K.D., Ma, B.L., 2009. Assessment of some major yield-limiting factors on maize production in a humid temperate environment. *Field Crops Res.* 110 (1), 21–26.
- Taylor, H.M., 1980. Soybean growth and yield as affected by row spacing and by seasonal water supply. *Agron. J.* 72 (3), 543–547.
- Thomson, A.M., et al., 2017. Science in the supply chain: collaboration opportunities for advancing sustainable agriculture in the United States. *Agric. Environ. Lett.* 2 (1).
- USDA-NASS, 2016. *USDA-National Agricultural Statistics Service (NASS), Crop U.S. State and County Databases*. <https://www.nass.usda.gov>.
- USDA-NASS, 2016. *USDA-National Agricultural Statistics Service (NASS), National Cultivated Layer*. <https://www.nass.usda.gov>.
- van Ittersum, M.K., Rabbinge, R., 1997. Concept of production ecology for analysis and quantification of agricultural input-output combination. *Field Crops Res.* 52, 197–208.
- van Ittersum, M.K., et al., 2013. Yield gap analysis with local to global relevance—a review. *Field Crops Res.* 143, 4–17.
- van Wart, J., et al., 2013. Use of agro-climatic zones to upscale simulated crop yield potential. *Field Crops Res.* 143, 44–55.
- Villamil, M.B., Davis, V.M., Nafziger, E.D., 2012. Estimating factor contributions to soybean yield from farm field data. *Agron. J.* 104 (4), 881–887.
- Wilhelm, W.W., Wortmann, C.S., 2004. Tillage and rotation interactions for corn and soybean grain yield as affected by precipitation and air temperature. *Agron. J.* 96 (2), 425–432.
- Yang, H.S., et al., 2004. Hybrid-maize—a maize simulation model that combines two crop modeling approaches. *Field Crops Res.* 87 (2–3), 131–154.
- Yang, H., Grassini, P., Cassman, K.G., Aiken, R.M., Coyne, P.I., 2017. Improvements to the Hybrid-maize model for simulating maize yields in harsh rainfed environments. *Field Crops Res.* 204, 180–190.