

University of Nebraska - Lincoln

## DigitalCommons@University of Nebraska - Lincoln

---

Agronomy & Horticulture -- Faculty Publications

Agronomy and Horticulture Department

---

5-15-2020

### Assessing approaches for stratifying producer fields based on biophysical attributes for regional yield-gap analysis

Spyridon Mourtzinis

*Agstat Consulting, Athens, Greece, agstat001@gmail.com*

Patricio Grassini

*University of Nebraska - Lincoln, pgrassini2@unl.edu*

Juan I. Rattalino Edreira

*University of Nebraska-Lincoln, rattalino@unl.edu*

José F. Andrade

*University of Nebraska-Lincoln, jandrade@unl.edu*

Peter M. Kyveryga

*Iowa Soybean Association, kyveryga@iastate.edu*

*See next page for additional authors*

Follow this and additional works at: <https://digitalcommons.unl.edu/agronomyfacpub>



Part of the [Agricultural Science Commons](#), [Agriculture Commons](#), [Agronomy and Crop Sciences Commons](#), [Botany Commons](#), [Horticulture Commons](#), [Other Plant Sciences Commons](#), and the [Plant Biology Commons](#)

---

Mourtzinis, Spyridon; Grassini, Patricio; Rattalino Edreira, Juan I.; Andrade, José F.; Kyveryga, Peter M.; and Conley, Shawn P., "Assessing approaches for stratifying producer fields based on biophysical attributes for regional yield-gap analysis" (2020). *Agronomy & Horticulture -- Faculty Publications*. 1355.  
<https://digitalcommons.unl.edu/agronomyfacpub/1355>

This Article is brought to you for free and open access by the Agronomy and Horticulture Department at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Agronomy & Horticulture -- Faculty Publications by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

---

**Authors**

Spyridon Mourtzinis, Patricio Grassini, Juan I. Rattalino Edreira, José F. Andrade, Peter M. Kyveryga, and Shawn P. Conley

# Assessing approaches for stratifying producer fields based on biophysical attributes for regional yield-gap analysis

Spyridon Mourtzinis,<sup>1</sup> Patricio Grassini,<sup>2</sup>  
Juan I. Rattalino Edreira,<sup>2</sup> José F. Andrade,<sup>2</sup>  
Peter M. Kyveryga,<sup>3</sup> and Shawn P. Conley.<sup>4</sup>

<sup>1</sup> Agstat Consulting, Athens, Greece

<sup>2</sup> Department of Agronomy and Horticulture, University of Nebraska-Lincoln,  
Lincoln, NE 68583-0915, USA

<sup>3</sup> Iowa Soybean Association, Ankeny, IA 50023, USA

<sup>4</sup> Department of Agronomy, University of Wisconsin-Madison, Madison,  
WI 53706, USA

Corresponding author – S. Mourtzinis , email [agstat001@gmail.com](mailto:agstat001@gmail.com)

## Abstract

Large databases containing producer field-level yield and management records can be used to identify causes of yield gaps. A relevant question is how to account for the diverse biophysical background (*i.e.*, climate and soil) across fields and years, which can confound the effect of a given management practice on yield. Here we evaluated two approaches to group producer fields based on biophysical attributes: (i) a technology extrapolation domain spatial framework (‘TEDs’) that delineates regions with similar (long-term average) annual weather and soil water storage capacity and (ii) clusters based on field-specific soil properties and weather during each crop phase in each year. As a case study, we used yield and management

---

Published in *Field Crops Research* 254 (2020) 107825

doi:10.1016/j.fcr.2020.107825

Copyright © 2020 Elsevier B.V. Used by permission.

Submitted 12 December 2019; revised 20 April 2020; accepted 20 April 2020; published 15 May 2020.

data collected from 3462 rainfed fields sown with soybean across the North Central US (NC-US) during four growing seasons (2014–2017). Following the TED approach, fields were grouped into 18 TEDs based on the TED that corresponded to the geographic location of each field. In the cluster approach, fields were grouped into clusters based on similarity of in-season weather and soil. To evaluate how the number of clusters would affect the results, fields were grouped separately into 5, 10, 18, and 30 clusters. The two stratification approaches (TEDs and clusters) were compared on their ability to explain the observed yield variation and yield response to key management factors (sowing date and foliar fungicide and/or insecticide). Lack of stratification of producer fields based on their biophysical background ignored management by environment ( $M \times E$ ) interactions, leading to spurious relationships and results that are not relevant at local level. In the case of the cluster approach, a fine stratification (18 and 30 clusters) explained a larger portion of the yield variance compared with a coarse stratification (5 and 10 clusters). However, for our case study in the NC-US region, we did not find strong evidence that the data-rich clustering approach outperformed the TEDs on the ability to explain yield variation and identify  $M \times E$  interactions. Only the stratification into 30 clusters exhibited a small improved ability at explaining yield variation compared with the TEDs. However, the use of the clustering approach had important trade-offs, including large amount of data requirements and difficulties to scale results to different regions and over time. The choice of the stratification method should be based on objectives, data availability, and expected variation in yield due to erratic weather across regions and years.

**Keywords:** Technology extrapolation domain, Cluster, Producer data, Soybean, Yield gap

## 1. Introduction

Analysis of large producer databases, including yield, management, and site-specific weather and soil data from multiple field-years, is becoming increasingly common in agriculture (*e.g.*, Sadras et al., 2002; Lobell et al., 2005; Rattalino Edreira et al., 2017; Mourtzinis et al., 2018; Di Mauro et al., 2018). When data from a large number of fields are available, analysis of these databases can help identify causes for yield gaps. However, the analysis can potentially be confounded as a result of variation in weather and soil across fields and years (hereafter referred to as the ‘biophysical background’). For example, let us consider a hypothetical management practice that has a positive and negative impact on yield in wet and dry environments, respectively, and a producer database collected from a region where frequency of dry and wet environments is similar (50:50). Analysis of the pooled data may lead to the conclusion that the practice does not have a

consistent impact on yield. A corollary of this simple example is that stratification of fields based on biophysical background is needed to capture M×E interactions and to identify management practices with greatest impact on yield for a given climate and soil context.

Stratification may not be needed for analysis of producer data collected from relatively small geographic areas with small field-to-field and year-to-year variation in weather and soil (*e.g.*, Grassini et al., 2011; Silva et al., 2017). In other cases, a single variable is a strong predictor of the type of environment in each field-year; hence, a simple stratification based on that variable may be sufficient. An example is the boundary-function analysis based on the relationship between water-limited yields and water supply (French and Schultz, 1984). However, at issue is how to stratify producer fields for large agricultural areas that include diversity in climate and soil leading to variation in yield potential. Some previous studies did not account for variation in climate and soil among producer fields, using administrative units or loosely defined production areas as criteria to group fields (*e.g.*, Villamil et al., 2012; Carr et al., 2016; Hurley and Mitchell, 2017). Other studies have used agro-ecological zones that are too coarse for yield-gap analysis at local level (*e.g.*, Tanaka et al., 2017; An et al., 2018). Some other studies have considered field-specific biophysical and management data to explain yield variation across producer fields, without an explicit categorization of the fields into discrete climate-soil domains (*e.g.*, Dardanelli et al., 2006; Di Mauro et al., 2018). One limitation of this approach is the need for site-specific weather and soil data, which are usually not available, leading researchers to retrieve these data from distant weather stations and coarse-scale soil maps. Another problem is that, without stratification, yield variation across fields reflects variation in yield potential, yield gaps, or both, making difficult to identify the causes for yield gaps for specific regions. For example, analysis of yield variation over a large geographic region, without any stratification, may reveal spurious relationships due to strong correlations between weather and management practices (*e.g.*, early sowing date and late cultivar maturity associated with environments with long frost-free season and high yield potential).

Seeking for an alternative to the previous approaches, two recent studies (Rattalino Edreira et al., 2017; Mourtzinis et al., 2018) assessed causes for soybean yield gaps in the North Central US (NC-US) region

using technology extrapolation domains (TEDs) as basis to stratify fields based on climate and soil similarity (<http://www.yieldgap.org>; Rattalino Edreira et al., 2018). The TED framework aims to reach a compromise between (i) accounting for biophysical factors with greatest influence on crop yield and its stability, and (ii) allowing a reasonable level of aggregation and geographic specificity to be relevant at informing research and extension programs. Briefly, the TED framework delineates regions based on (i) annual total growing degree days ( $T^b = \sum C$ ), which correlates with the length of crop growing season, (ii) aridity index (calculated as the ratio between mean annual total precipitation and mean annual total evapotranspiration), which defines the degree of water limitation in rainfed cropping systems, (iii) annual temperature seasonality, which differentiates between temperate and tropical climates, 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. Details on TEDs development and data sources are available elsewhere (Rattalino Edreira et al., 2018). The TED framework is static, that is, it does not account for year-to-year fluctuation in weather which, in turn, can lead to different responses to management practices. Another limitation is that TEDs are based on weather variables computed on an annual rather than crop-season basis. Annual means may mask the influence of weather during specific crop stages that are important for yield determination. A question is whether a more granular approach that accounts for year-to-year variation in weather and crop stages can improve the ability to identify causes for yield gap at regional level.

As a starting point to evaluate the ability of different approaches to group producer fields for regional yield-gap analysis, we compare here two stratification approaches: one based on a framework delineating ‘static’ climate-soil domains (TEDs), and another based on a data-rich clustering method using means of meteorological factors computed separately for each crop phase and field-specific soil properties (hereafter referred to as ‘clustering’). Our hypothesis is that a more granular approach should lead to improved capacity to identify causes for yield gaps. For this comparison, we used a large database including 3462 fields planted with soybean during four crop growing seasons (2014–2017) across the NC-US region - a region that accounts

for ca. 30% of global soybean production (FAOSTAT, 2016). We discussed advantages and disadvantages of the methods used for stratifying fields and implications for yield gap analysis.

## 2. Methods

### 2.1. Producer database and associated weather and soil information

Soybean yield and management data were collected from rainfed fields sown with soybean across 10 states located in the NC-US region. The database included four crop growing seasons (2014–2017) and has been extensively described in Rattalino Edreira et al. (2017) and Mourtzinis et al. (2018). Producers reported data on field location, average yield (adjusted to 13% grain moisture content), and management practices (*e.g.*, sowing dates, application of foliar fungicide and/or insecticide, cultivar, *etc.*). Producers also reported incidence of other field adversities (*e.g.*, pests, diseases, weeds, iron deficiency chlorosis, hail, waterlogging, and frost). Because the goal was to identify management practices that can be adopted by producers to increase their yields, fields subjected to unmanageable field adversities (*e.g.*, hail, frost, flooding) leading to substantial yield losses were excluded from the database. However, farmers tended to report damage by wind, frost, or other adversities, even when damage was mild or *nil*, or when only affected a small portion of the field. Hence, we only excluded those fields reporting any of the aforementioned adversities when they fall below the 25th percentile of the yield distribution within each TEDyear (4% of all fields contained in the database). After quality control, 3462 fields were included in the study. When aggregated at the county year level, yields were compared well with those reported by USDANASS for the same counties and years (Rattalino Edreira et al., 2017).

Daily weather data for each field were retrieved from the DAYMET dataset (Thornton et al., 2017). DAYMET daily weather data is reasonably accurate when means or totals are computed over extended periods (*e.g.*, monthly or over a crop phase) (Mourtzinis et al., 2017). Weather variables retrieved for each field included incident solar radiation, daylength, daily minimum (Tmin) and maximum (Tmax)

temperature, and precipitation. For each day, Penman-Monteith grass-based reference evapotranspiration (ET<sub>o</sub>) was calculated following Allen et al. (1998). Similarly, daytime vapor pressure deficit (VPD) was estimated as 2/3 of the difference between the saturated vapor pressure at T<sub>max</sub> and the saturated vapor pressure at T<sub>min</sub> following Keenan et al. (2004).

Soil pH and PAWHC were retrieved for each field from the SSURGO database. For each field, pH (0–150 cm) was calculated as a weighted sum of the pH values reported for the topsoil (0–30 cm) and subsoil (30–150 cm). We used the weighted sum of both depths due to their strong correlation (Pearson's  $r=0.67$ ,  $p < 0.05$ ). Mean pH in a given field was derived from the pixel pH distribution within each field (ca. 9 pixels per field). Using the mode, instead of the mean, would have resulted in a negligible change in the calculated field-level pH (< 1%). Slope and terrain differences across fields can influence the crop water balance and final seed yield. Hence, we also calculated the topographic wetness index (TWI) for each field which has been used to characterize the potential for surface run-off and run-on in landscapes (Moore et al., 1993). High values are associated with flat terrain whereas smaller values are associated with more uneven fields (e.g., fields with slopes). Thus, fields with high TWI are more likely to receive runoff water from adjacent areas and *vice versa*. TWI is usually correlated with other soil attributes, including soil organic matter, soil texture, and phosphorous content; hence, higher TWI values are generally associated with more productive soils. TWI was calculated using the *rsaga.wetness.index* package in R (R development Core team, 2016) using the 30-m resolution National Elevation Dataset (USDA:NRCS: Geospatial Data Gateway; <https://datagateway.nrcs.usda.gov/>). For a given field, mean TWI was derived from the pixel TWI distribution within each field.

## *2.2. Estimation of means of meteorological variables for each crop phase*

The most commonly used staging system for soybean was developed by Fehr and Caviness (1977). In this system, vegetative development begins at emergence (VE) when cotyledons appear above the soil surface. A given vegetative stage (V<sub>n</sub>) is reached when  $n$  number



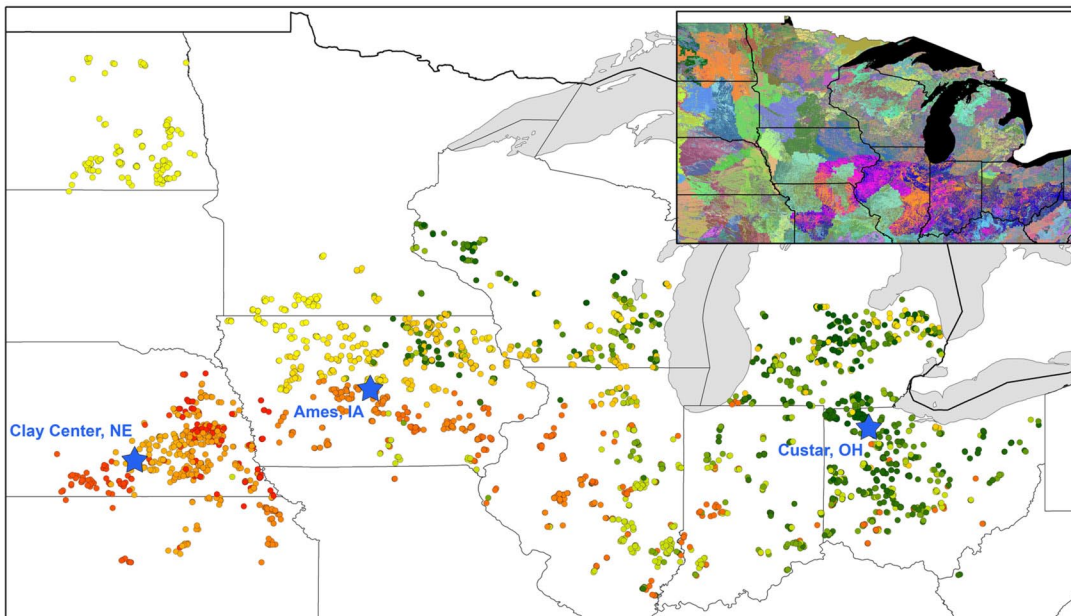
of nodes appear on the main stem with fully developed leaves beginning with the unifoliolate nodes. Number  $n$  can be any number, beginning with 1 for V1, which corresponds to the first-node stage. The beginning of flowering (R1) marks the initiation of the reproductive phase and is easily recognized by the first open flower at any node on the main stem. The beginning of pod-setting (R3) occurs when a pod of at least 5mm length is observable at one of the four uppermost nodes on the main stem. The beginning of seed filling (R5) is marked by the presence of seeds of at least 3mm in any of the four uppermost nodes. Physiological maturity (R7) indicates the end of the reproductive phase and is marked by the presence of one mature pod of brown color at any node on the main stem.

In this study, we used SoySim model (Setiyono et al., 2007, 2010) to simulate dates of key developmental stages in soybean for each field year case, including VE, V4, R1, R3, R5, and R7. Model performance at reproducing the observed vegetative and reproductive phenology in producer fields in the US-NC region has been satisfactorily evaluated in previous studies, with an overall RMSE for the comparison between observed and predicted crop stages of 3.6 d (Torrión et al., 2011). Simulations for each of the 3462 field-years was based on daily DAYMET weather and producer-reported sowing date (range: early March to late June) and cultivar maturity group (MG) (range: 0.09 and 4.8). Based on the simulated dates of VE, V4, R1, R3, R5, and R7 for each field-year, average T<sub>min</sub>, T<sub>max</sub>, solar radiation, day length, and VPD were calculated for five crop phases: VE-V4, V4-R1, R1-R3, R3-R5, and R5-R7. Additionally, we calculated a simple water balance to determine the degree of water limitation during each phase. Water balance was calculated as the difference between total precipitation and the non-water limiting crop ET derived from SoySim. The latter provides a proxy of the crop water demand and it was modelled based on the leaf area index, canopy characteristics, and E<sub>T0</sub>. Water balance was calculated for all phases but for VE-V4. Except for extreme years with severe drought, which was not the case in any of the years included in the present study, soils in the NC-US region are fully recharged at sowing time and water is not limiting in early vegetative stages (Morell et al., 2016; Rattalino Edreira et al., 2017).

### 2.3. Field stratification

Selecting which approaches to compare and exactly how to implement them have, to a certain extent, a subjective component (*e.g.*, selection of soil-weather variables, number of clusters, *etc.*). In our study, we used, first, a less-data intensive approach based on a spatial framework that defines geographic regions with similar climate and weather (*i.e.*, TEDs), and, second, a data-intensive approach that uses all available data and tools (*i.e.*, clustering). The selected methods portray well the range of approaches that have been used in the literature.

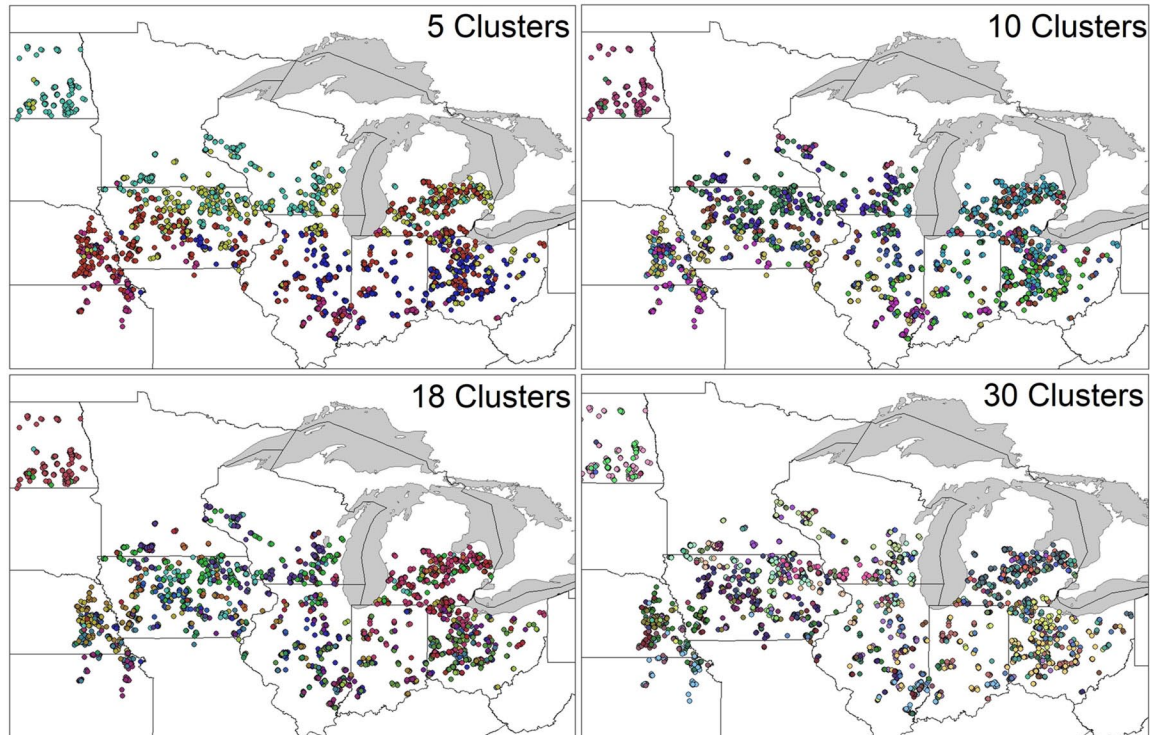
Following the first approach (*i.e.*, TEDs), each field was assigned to a unique TED based on its geographic coordinates (Fig. 1). Fields were grouped into a total of 18 TEDs. Within a given TED, all fields would have, in principle, fairly similar (within a range) annual growing degree day, aridity index, temperature seasonality, and PAWHC. Again, TEDs are based on annual-based long-term weather; hence, they do



**Fig. 1.** Stratification of 3462 farmer fields in 18 technology extrapolation domains (TEDs). Fields shown with same color belong to the same TED, that is, a combination of long-term average climate and water storage capacity. Inset shows the TED spatial framework. Note that each field was assigned to a unique TED based on its geographic coordinates. Stars indicate the location of meteorological stations used to describe seasonal weather patterns shown in Fig. 3.

not account for weather variation across years. Similarly, they do not account for other soil and terrain properties (besides PAWHC) that can affect crop yield and stability such as pH and TWI. On average, there were 192 fields per TED, with >95 fields in all TEDs. The number of fields per TED was reasonable to detect yield differences due to management practices as reported elsewhere (Mourtzinis et al., 2018). To understand the degree of yield variation accounted for by TEDs *versus* year and TED×year interaction, we calculated the percentage of sum of squares (%SS) attributed to each term, after excluding the error. We also evaluated TEDs on their ability to group fields based on their yield potential. Following Rattalino Edreira et al. (2017), the yield potential was estimated from the 95th percentile of the yield distribution in each TED-year; an average (4-y) yield potential was computed for each TED. Effect of TED on yield potential was evaluated using analysis of variance (ANOVA).

In the second approach (*i.e.*, clustering), crop-phase specific weather and soil properties were standardized to mean=0 and standard deviation= 1 and clusters were created using PROC FASTCLUS in SAS 9.4 (SAS Institute Inc., 1999). As mentioned previously, weather variables (water balance, solar radiation, Tmax, Tmin, VPD, and photoperiod) were calculated separately for five crop phases (VE-V4, V4-R1, R1-R3, R3-R5, and R5-R7) and soil properties included PAWHC, TWI, and pH. Briefly, the iterative algorithm minimizes the sum of squared distances from the cluster means, using Euclidean distances computed from numeric variables. This clustering method is called a *k*-means model, since the cluster centers are the means of the observations assigned to each cluster. In each iteration, the least-squares criterion is reduced until convergence is achieved. We used adaptive training (DRIFT option) in which the cluster seed is updated each time an observation is added. We specified LEAST=2 to minimize the root mean square difference between the data and the corresponding cluster means. The LEAST option increases the maximum number of iterations and the chances that the optimization process will converge. The desired number of clusters for each scenario was achieved by using the MAXCLUSTER and DELETE options keeping the rest aforementioned parameters constant. The number of clusters was pre-defined and, to a large extent, subjective. In principle, one would expect greater similarity in climate and soil among fields as the number of



**Fig. 2.** Stratification of 3462 farmer fields in 5, 10, 18, and 30 clusters. Fields shown with same color belong to the same cluster. Fields were clustered based on similarity of site-year weather during the crop season and soil properties.

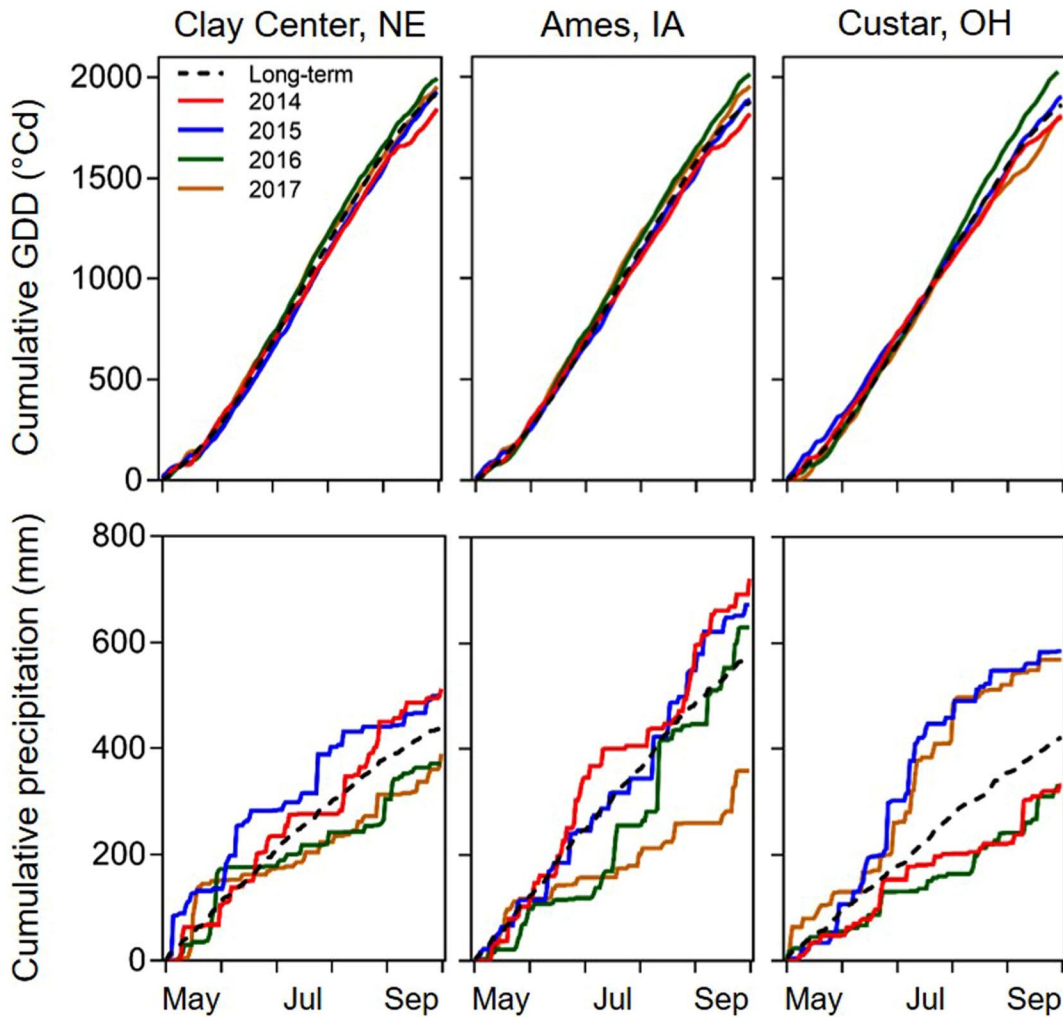
clusters increases, which should help at identifying the specific management practices that explain the greatest portion of the field-to-field yield variation. At the same time, higher number of clusters means fewer fields per cluster, which can result in insufficient power to detect statistically significant differences in yield between management practices. Given this uncertainty, we included a stratification with 18 clusters, which was comparable to the number of TEDs ( $n=18$ ), and then some variants ( $n=5, 10, 30$ ) to evaluate how the number of clusters can influence the results. We did not include stratification with larger number of clusters as that would have led to an insufficient number of fields in many clusters to conduct a statistical analysis with adequate power. While we recognize that the analysis does not allow to establish a precise ‘optimal’ number of clusters, it does explore the plausible range of options. To summarize, we evaluated the clustering approach for four different number of clusters (5, 10, 18, and 30), hereafter referred to as C5, C10, C18, and C30, respectively (Fig. 2).

As a first step, frequency of each cluster within each TED was calculated to determine if one type of cluster prevailed over the others; the analysis was performed separately for C5, C10, C18, and C30. Subsequently, analysis of variance (ANOVA) using PROC MIXED in SAS 9.4 was used to evaluate each of the stratification methods (TED and the four cluster sizes) on their ability to explain the observed variation in yield across field-year observations. To do this, we computed the %SS (relative to the overall SS) accounted for each of the five stratifications. We focused the analysis on understanding the effect of sowing date and use of foliar fungicide and/or insecticide (FI) on soybean yield because previous studies using the same database have determined that those were the most important management factors explaining soybean yield gaps in the US NC region (Rattalino Edreira et al., 2017; Mourtzinis et al., 2018). Hence, we computed the %SS (after excluding the error) associated with each of these management practices (and their interactions) for each stratification method. Of particular importance was the analysis of management×TED (or cluster) interactions between stratification methods as this would give an indication of the ability of the stratification methods to capture M×E interactions. We note that the objective of this work was not to re-validate the impact of sowing date and FI on soybean yield but rather to evaluate how different stratification methods can capture the effect of these management factors on soybean yield across producer fields with diverse biophysical background.

### **3. Results**

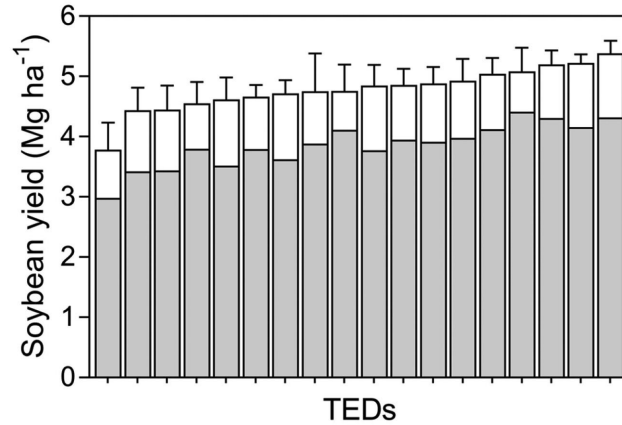
#### *3.1. Weather variation across sites and years*

The four years included in the database portrayed well the expected year-to-year and spatial variation in seasonal precipitation and temperature in the region (Fig. 3). For example, total rainfall ranged from 380 to 700mm at Ames, IA. The database also included years with relatively warm and cool weather as indicated by the contrasting seasonal growing-degree days accumulated in 2014 *versus* 2016. To summarize, the database can be considered representative of the temporal variation in weather across soybean fields in the NC-US region.



**Fig. 3.** Cumulative precipitation and growing-degree days (GDD, Tbase and Topt 8 °C and 31 °C, respectively) from May 1st until September 30th, which roughly coincides with soybean emergence and physiological maturity, at three representative locations across the west-to-east precipitation gradient in the U.S. Corn Belt in 2014 (red), 2015 (blue), 2016 (green) and 2017 (orange). Dashed line indicates the long-term (1988-2018) average. The three locations are shown in Fig. 1.

Variation in weather across regions and years led to variation in producer yield (Fig. 4). The effect of TED, year, and TED×year interactions were statistically significant ( $p < 0.001$ ), but TEDs explained the largest portion of the modelled yield variation (67% of sum of SS after excluding the error); the rest of the variation was explained by year and TED×year interaction (16 and 17% of SS after excluding



**Fig. 4.** Yield potential and average producer yield across technology extrapolation domain (TED). Values are 4-y means, computed from the average and 95th percentile derived from the field yield distribution in each TED-year combination. Bars show the yield potential while the solid portion of the bar represents the average producer yield. Vertical lines indicate the standard deviation for yield potential. TEDs are sorted from lowest to highest yield potential.

error). Similarly, the effect of TED on yield potential was statistically significant ( $p < 0.001$ ), with yield potential ranging from 3.8 to 5.4  $\text{Mg ha}^{-1}$  across TEDs (Fig. 4).

### 3.2. Field stratification into TEDs and clusters and associated drivers

Number of field-years per cluster decreased with number of clusters (Table 1). On average, there were 692, 346, 192, and 115 fields per cluster for the stratification based on 5, 10, 18, and 30 clusters, respectively. Except for the stratification based on 30 clusters, all TEDs

**Table 1** Minimum, mean, and maximum number of fields (n) and field-to-field coefficient of variation (CV) for soybean yield within technology extrapolation domains (TEDs) and clusters (C). Separate values are shown for fields grouped into different number of clusters (C5, C10, C18, and C30).

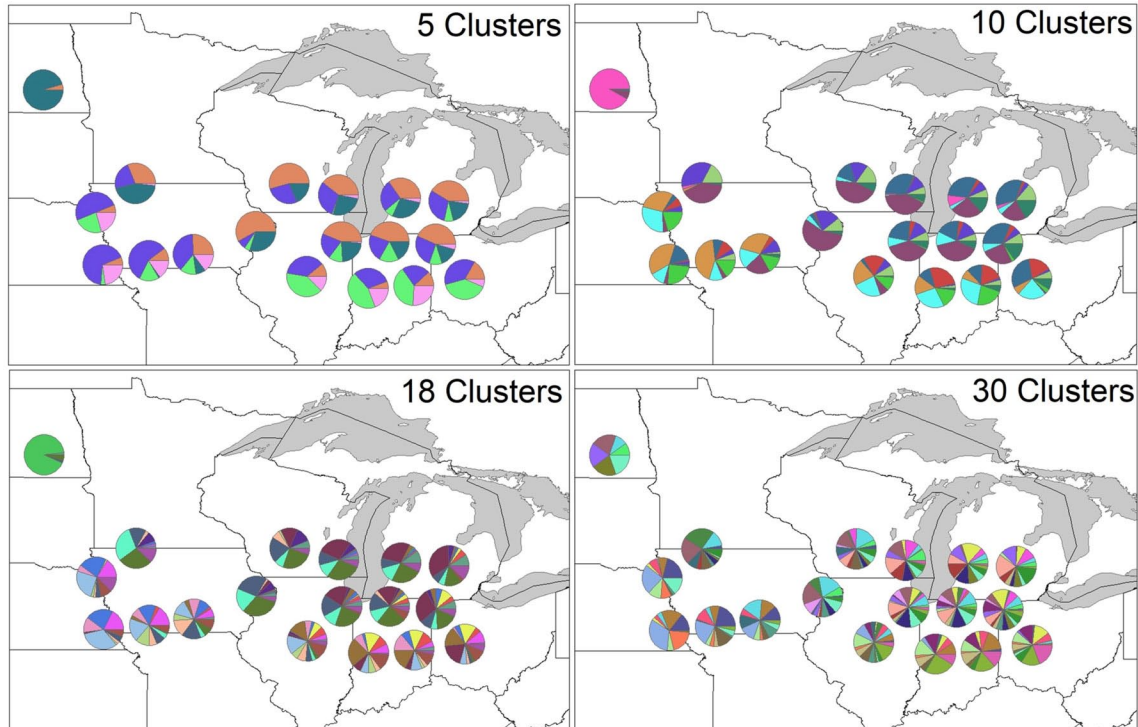
	C5		C10		C18		C30		TEDs	
	n	CV	n	CV	n	CV	n	CV	n	CV
Minimum	290	16	185	13	73	13	41	12	96	11
Mean	692	19	346	18	192	17	115	17	192	17
Maximum	1031	22	772	22	454	23	266	22	354	23

and clusters had a minimum number of fields near or larger than 100 fields which, as mentioned previously, would allow to capture expected yield differences as a result of management practices (Mourtzinis et al., 2018). Coefficient of variation (CV) for yield was  $\leq 23\%$  across all clusters and TEDs, without any detectable difference between stratification methods.

The two approaches to stratify producer fields lead to different grouping of the fields. In the case of the TEDs, the geographic distribution of the different groups of fields portrayed gradients in temperature (north-south) and precipitation (west-east) across the region and, to a lesser extent, variation in PAWHC (Fig. 1). For example, fields located in the western fringe of the region were grouped in different TEDs compared with fields in the eastern and central portions of the region as a result of the west-east gradient in seasonal precipitation. In contrast, the clustering approach grouped fields according to field-year in-season weather and a number of soil properties (Fig. 2). As a result, the same field could be grouped in one cluster in a given year and in another cluster in the following year. For example, despite differences in long-term average precipitation, ca. 270 fields located in Nebraska were grouped together with fields in Iowa and Illinois because in-season precipitation (as well as other parameters used for clustering) was not too different among these fields in specific years. Similarly, analysis of frequency of clusters within each TED showed that, except for the case of the TED in the northwestern portion of the region (North Dakota) and the analysis based on a small number of clusters (C5), none of the clusters clearly prevailed within a given TED (Fig. 5).

Reported  $r^2$  values in Table 2 indicate the explained variability in a given variable when clusters were used as explanatory variables (PROC FASTCLUS, SAS 9.4). Crop phase-specific weather parameters were strong clustering predictors (Table 2). Most informative variables for cluster delineation were Tmax followed by VPD, daylength, and Tmin (Table 2). In all cases, the predictive power increased with higher number of clusters. Across cluster number  $\times$  weather parameter combinations, differences in  $r^2$  were relatively large among crop phases for water balance and Tmin, especially in C5 and C10. In contrast, differences among crop phases were small for the other cluster number  $\times$  weather parameter combinations. The  $r^2$  tended to be higher in reproductive *versus* vegetative phases for Tmax, Tmin, and





**Fig. 5.** Pie charts showing frequency of clusters within each technology extrapolation domain (TED). Separate panels are shown for stratification based on different number of clusters (5, 10, 18, and 30 clusters). Each pie chart is centered on the centroid that corresponds to each TED.

solar radiation, but this was not the case for the other weather parameters. Indeed, water balance early in the season was an important factor for field clustering. In contrast to weather parameters, soil variables tended to be relatively poor clustering predictors, especially when the number of clusters was small (*i.e.*, C5). Among soil parameters, PAWHC was the most important clustering predictor, especially when the number of clusters was high (*i.e.*, C30).

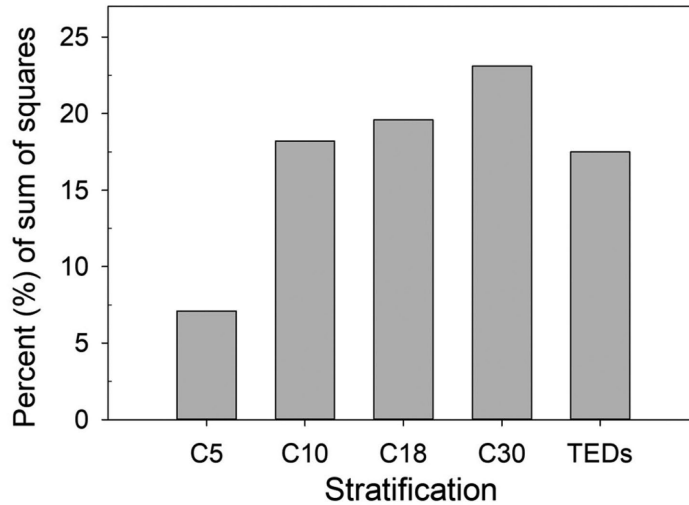
### *3.3. Portion of yield variation accounted by each stratification method and effect of management practices*

Quantifying the fraction of overall yield variation accounted for by each stratification method provided an objective way to evaluate their performance at grouping fields based on their biophysical background (Fig. 6). Across stratification methods, the overall modelled

**Table 2** Coefficient of determination ( $r^2$ ) for soil and phase-specific weather variables for the fields stratified into different number of clusters: 5, 10, 18, and 30 (C5, C10, C18, and C30, respectively). Variables with high  $r^2$  were strong cluster predictors.

Variable	C5	C10	C18	C30
Soil parameters				
pH	<i>nil</i>	<i>nil</i>	0.01	0.02
PAWH	<i>nil</i>	0.21	0.24	0.36
TWI	0.04	0.07	0.12	0.17
Water balance <sup>δ</sup>				
V4-R1	0.29	0.34	0.44	0.52
R1-R3	0.25	0.24	0.37	0.38
R3-R5	0.02	0.18	0.27	0.39
R5-R7	0.18	0.32	0.37	0.43
Tmin				
VE-V4	0.36	0.37	0.57	0.70
V4-R1	0.32	0.55	0.63	0.69
R1-R3	0.50	0.56	0.68	0.73
R3-R5	0.42	0.63	0.67	0.75
R5-R7	0.29	0.59	0.65	0.75
Tmax				
VE-V4	0.45	0.49	0.64	0.76
V4-R1	0.42	0.61	0.67	0.72
R1-R3	0.51	0.58	0.70	0.79
R3-R5	0.55	0.63	0.72	0.79
R5-R7	0.47	0.66	0.69	0.77
Solar radiation				
VE-V4	0.30	0.47	0.57	0.65
V4-R1	0.34	0.44	0.49	0.60
R1-R3	0.29	0.38	0.56	0.58
R3-R5	0.27	0.42	0.52	0.62
R5-R7	0.24	0.54	0.59	0.68
VPD				
VE-V4	0.47	0.55	0.65	0.75
V4-R1	0.49	0.63	0.68	0.73
R1-R3	0.45	0.53	0.64	0.74
R3-R5	0.38	0.44	0.58	0.64
R5-R7	0.42	0.56	0.62	0.67
Daylength				
VE-V4	0.44	0.60	0.69	0.78
V4-R1	0.55	0.66	0.72	0.75
R1-R3	0.55	0.65	0.71	0.75
R3-R5	0.53	0.63	0.69	0.73
R5-R7	0.51	0.61	0.67	0.71

VE: emergence; V4: fourth vegetative stage; R1: beginning of bloom, R3: beginning of pod setting, R5: beginning of seed filling, R7: physiological maturity; VPD: daytime vapor pressure deficit; Tmin: minimum temperature; Tmax: maximum temperature. <sup>δ</sup> Water balance during the VE-V4 phase was not included because water supply is not limiting in early vegetative stages.



**Fig. 6.** Amount of soybean yield variability (expressed as percentage of total sum of squares) explained by each stratification variable: technology extrapolation domains (TEDs) or 5, 10, 18, and 30 clusters (C5, C10, C18, and C30, respectively). Percentage of sum of squares was calculated using fixed effect models including only stratification method as independent variable in the model.

%SS was relatively small, ranging from 7% to 23%. This result was expected considering the influence of management practices on yield and some residual variation in weather and soil within clusters or TEDs that cannot not be fully captured by the stratification methods. In the case of clusters, the analysis revealed that the fraction of modelled variation increased with increasing number of clusters. Modelled variation was small for C5 (%SS=7) and changed little from C10 to C30 (from 18 to 23%). The portion of variation accounted for by TEDs (%SS=18) was slightly smaller (but comparable) to that of C10, C18, and C30 and more than 2x larger compared with C5 (Fig. 6). To summarize, TEDs and clusters (except for C5) accounted for a relatively similar portion of the observed yield variance.

The percentage of sum of squares explained by two key management factors (sowing date and foliar fungicide and/or insecticide) and their interactions with the stratification method are shown in Table 3. In general, the %SS accounted by the single effects of management practices decreased with increasing number of clusters. For example, in the case of C5, about 27 and 52% of the total variance (after excluding the error) was accounted by the effect of sowing date and FI, respectively, while only 4 and 29% (in the same order) of the total

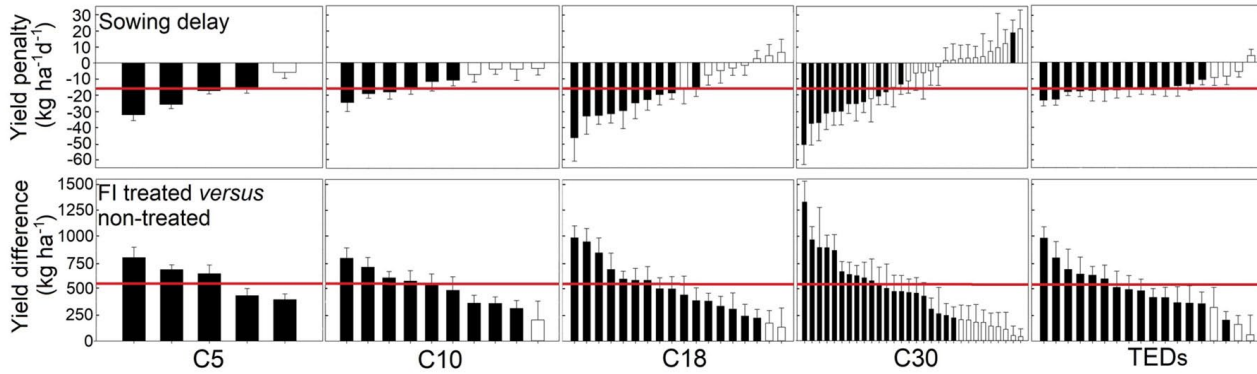
**Table 3** Portion of the modelled variation in soybean yield accounted for by stratification method (SM), sowing date (SD), foliar fungicide and/or insecticide (FI), and their interactions. Stratification methods include technology extrapolation domains (TEDs) and clusters (C). Separate values are shown for fields grouped into different number of clusters (C5, C10, C18, and C30). All sources of variation were statistically significant at  $\alpha=5\%$ .

Source of variation	% of sum of squares <sup>a</sup>				
	C5	C10	C18	C30	TEDs
Stratification method (SM)	5	4	9	13	8
Sowing date (SD)	27	14	12	4	28
Foliar fungicide/insecticide (FI)	52	54	44	29	33
SM×SD	6	6	9	13	8
SM×FI	10	21	27	41	23

<sup>a</sup> Percentage of total sum of squares after excluding the error.

variance was explained by the same two factors in the case of C30 (Table 3). In contrast, the %SS accounted by management×cluster interaction increased with higher number of clusters. These results suggest that a greater number of clusters allowed a better identification of M×E interactions. In the case of the TEDs, its interactions with sowing date and FI accounted for respective 8 and 23% of the total explained variance, which was similar to the %SS accounted by the C18 (9 and 27%, respectively) (Table 3).

Comparison of average yield response to sowing delay and FI on yield across all fields (*i.e.*, without stratification) *versus* range in yield responses across clusters or TEDs highlighted the importance of stratifying fields in order to capture M×E interactions (Fig. 7). However, it was difficult to find strong evidence that one stratification method outperformed the others in detecting the influence of management practices on yield. There were differences in the range of yield penalty (due to delay in sowing) or yield difference (due to FI) among the different stratification methods. For example, only considering cases with statistically significant differences from zero, yield response to sowing delay ranged from ca. -11 to -32 kg ha<sup>-1</sup> d<sup>-1</sup> across clusters in C5 and C10 and from ca. 19 to -50 kg ha<sup>-1</sup> d<sup>-1</sup> across clusters in C18 and C30 (Fig. 7). Number of clusters (as percentage of total clusters) without a statistically significant response to sowing delay increased from 20% (C5) to 56% (C30), and there was only one cluster (in C30) in which there was a statistically significant yield increase as a result of sowing delay (+19 kg ha<sup>-1</sup> d<sup>-1</sup>). As a general trend, the magnitude



**Fig. 7.** Yield penalty due to sowing delay (upper panels) and yield difference between treated and non-treated fields with foliar fungicide and/or insecticide (FI; bottom panels). Each bar corresponds to a specific cluster or technology extrapolation domain (TED;  $n = 18$ ). White bars indicate cases where the effect of sowing date on yield (upper panels) or yield difference between treated and non-treated fields (bottom panels) was not statistically different from zero, respectively ( $P > 0.05$ ). Vertical bars indicate standard error of the mean. Horizontal red lines indicate average yield response to sowing date delay and foliar fungicide and/or insecticide application across fields without considering stratification.

of the yield penalty due to sowing delay became more variable across clusters as the number of clusters increased. Range of yield differences due to FI followed similar patterns. In the case of the TEDs, the range of yield penalty due to sowing delay was similar to that observed for C5 and C10, but smaller compared with C18 and C30, ranging from  $-11$  to  $-23 \text{ kg ha}^{-1} \text{ d}^{-1}$ , with 22% of TEDs exhibiting no statistically significant response to sowing delay. In the case of yield differences due to FI, the range observed across TEDs was similar to that for C18 and C30 (Fig. 7).

#### 4. Discussion

Spatial and temporal variation in weather, together with field-to-field variation in soil properties, can confound the effect of management practices on yield. Hence, stratifying fields considering site-specific soil and crop-season weather can potentially increase the ability to identify management practices with greatest influence on yield and  $M \times E$  interactions. Using producer soybean fields in the US NC region

as a case study, we showed that a data-rich approach to group producer fields, such as clustering, resulted in a marginal increased ability to detect management practices influencing crop yield compared with a less-data intensive stratification approach (TEDs) based on annual climate means and soil water storage. Still, the advantage of C30 compared to the TEDs was small as quantified using the percentage of sum of squares for seed yield explained by each stratification method (% SS=18 *versus* 23 for TEDs and C30, respectively).

There are a number of factors explaining the similarity in the analysis based on TEDs *versus* clusters. First, year-to-year variation in yield across TEDs was relatively low as quantified using the inter-annual CV (range: 11–23%). Indeed, we found that year and TED×year interaction explained a much smaller portion of overall yield variance (16 and 17% of modelled SS) compared with TEDs (67% of modelled SS). Second, there is usually a good correlation between seasonal *versus* annual means (or totals) of the meteorological variables for the region evaluated in the study. For example, total precipitation and mean temperature during the crop season were highly correlated with their respective annual values for all the TEDs evaluated in our study ( $r^2=0.57$  and 0.93, respectively;  $p < 0.001$ ). Hence, we consider the findings from our research to be applicable in other production environments that follow similar patterns in terms of yield stability and association between season- and annual-based weather. In contrast, we speculate that the cluster approach may have an advantage in current and future environments with more erratic weather patterns and, hence, higher yield variability due to year and stronger TED×year interactions as it is the case of the US Central Great Plains and Australia (Chapman et al., 2000; Lollato et al., 2017). We also recognize the importance of stratifying fields based on other non-biophysical factors. For example, if there are contrasting differences in the socio-economic background among farmers, it may be prudent to account for them by clustering fields based on other variables such as access to market or farm size (*e.g.*, Jelsma et al., 2017).

Potential advantages and disadvantages of each stratification method are shown in Table 4. For comparison purposes, we also discussed the approach without any stratification. On the one hand, using site-specific soil and crop-phase weather in the cluster approach would allow for a better characterization of the biophysical background associated with each site-year and better assessment of M×E

**Table 4** Advantages and disadvantages associated with different methods for field stratification: no stratification, technology extrapolation domains (TEDs), and clusters.

Attribute	No stratification	TEDs	Clusters
Data requirements	Low	Low	High
Description of biophysical background in each field	Poor	Acceptable	Excellent
Ability to scale out results	Difficult	Easy	Intermediate-difficult
Detection of M×E interactions	Low	Intermediate-high	High

interactions. However, as mentioned previously, this would be relevant in environments with strong variation in weather across years, which was not the case in our study region. On the other hand, the clustering approach requires an enormous amount of ancillary weather and soil data as well as crop modeling to determine crop stages, without a clear benefit in terms of additional explanatory power. Perhaps more importantly, most farmer decisions relative to crop management are made before the season starts (*e.g.*, tillage, fertilizer, sowing), without knowing to which specific cluster the field would be grouped into in that particular year. Hence, extrapolation of result from analysis of cluster to a specific region is difficult. It can still be done if the frequency of clusters in each TED is known (Fig. 5) as long as the most dominant cluster(s) in each TED can be clearly identified, which, again, it is not the case for the US NC region. Conversely, the TEDs approach seems to be a reasonable approach to stratify fields given its low data requirements, ability to detect M×E interactions, and ease to extrapolate results to specific areas based on their average weather and soil conditions. Although not evaluated in the present study, a potential drawback of the TEDs is that only those TEDs with a sufficient number of fields (> 100 in our case) would be considered for the analysis while the others (with smaller number of fields) would be discarded, whereas all fields could potentially be used with the clustering approach. Overall, we argue that further evaluation of stratification methods across regions with contrasting year-to-year weather variation should be performed to understand where the clustering approach would have a benefit that justifies the associated high data requirement. For that cross-region analysis, it would be worthwhile to consider using a more objective and automated approach to determining the ‘optimal’ number of clusters for each region (*e.g.*, Silhouette plots). Investigating whether less data intensive

clustering approaches could also still lead to a justifiable result should be an explicit goal as this is one of the two main disadvantages of the cluster approach.

Stratification of producer fields based on the biophysical background (either using TEDs or clustering) should be taken as a routinely step in any future studies assessing causes for yield gaps to avoid the confounding effect of varying climate and soil conditions, which leads to variation in yield potential across fields (Fig. 4). No stratification of producer data collected across wide geographic regions would lead to spurious relationships and information that has little relevance at local level. For example, a simple analysis of yield variation across the NC-US region, without any type of stratification, would reveal that soybean yield is associated with MG ( $r^2=0.24$ ,  $p<0.001$ ). This is because late MG cultivars are grown in the warmer southern fringe of the region where the length of the frost-free season is longer and, in turn, an earlier sowing is possible; all together this would lead to higher yield. Such analysis would simply confirm the (well-known) fact that late MGs are sown earlier in the southern region (Mourtzinis and Conley (2017) but will have little value to help fine tune sowing date and MG choice for a specific climate-soil domain within the NC-US region. Similarly, lack of (or coarse) stratification would ignore  $M \times E$  interactions, as illustrated for sowing date and FI in Fig. 7, resulting in information that has limited value for farm management and prioritization of research and extension programs. Instead, proper field stratification, as performed in this study, allowed us to determine the effect of management practices on yield, in terms of both sign and magnitude, for clearly defined biophysical environments.

## 5. Conclusions

Lack of stratification of producer fields based on their biophysical background ignores  $M \times E$  interactions, leading to spurious relationships and results that are not relevant at local level. Results from this study showed that a fine stratification into 18 and 30 clusters of similar environmental conditions explained a larger portion of the yield variance compared with a coarse stratification into 5 and 10 clusters. However, for our case study in the NC-US region, we did not



find strong evidence that the data-rich clustering approach outperformed the TEDs on the ability to explain yield variation and identify M×E interactions. Only the stratification into 30 clusters exhibited a small improved ability at explaining yield variation compared with the TEDs. In turn, use of the clustering approach has many trade-offs, including high data requirements and difficulties to scale out results. The choice of the stratification method should be based on objectives, data availability, and expected variation in yield due to erratic weather across years.

**Competing Interest** – None declared.

#### **CRedit authorship contribution statement**

**Spyridon Mourtzinis:** Conceptualization, Methodology, Formal analysis, Writing - original draft. **Patricio Grassini:** Conceptualization, Methodology, Writing - review & editing. **Juan I. Rattalino Edreira:** Formal analysis, Writing - review & editing. **José F. Andrade:** Formal analysis, Writing - review & editing. **Peter M. Kyveryga:** Methodology, Writing - review & editing. **Shawn P. Conley:** Conceptualization, Writing - review & editing.

**Acknowledgments** – The 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 Adam C. Roth (University of Wisconsin-Madison), Shaun N. Casteel (Purdue University), Ignacio A. Ciampitti (Kansas State University), Hans J. Kandel (North Dakota State University), Mark A. Licht (Iowa State University), Laura E. Lindsey (The Ohio State University), Daren S. Mueller (Iowa State University), Seth L. Naeve (University of Minnesota), Emerson D. Nafziger (University of Illinois), Jordan Stanley (North Dakota State University), Michael J. Staton (Michigan State University Extension), University of Nebraska Extension Educators, Nebraska Natural Resource Districts, and Iowa Soybean Association for helping collect the producer data.

#### **References**

- Allen, R.G., Pereira, L.S., Raes, D., Smith, M., 1998. Guidelines for Computing Crop Water requirements-FAO Irrigation and Drainage Paper 56. Available URL: <http://www.fao.org/3/X0490E/X0490E00.htm>.
- An, N., Wei, W., Qiao, L., Zhang, F., Christie, P., Jiang, R., Dobermann, A., Goulding, W.T.K., Fan, J., Fan, M., 2018. Agronomic and environmental causes of yield and nitrogen use efficiency gaps in Chinese rice farming systems. *Eur. J. Agron.* 93, 40–49.

- Carr, T., Yang, H., Ray, C., 2016. Temporal variations of water productivity in irrigated corn: an analysis of factors influencing yield and water use across central Nebraska. *PLoS One*, e0161944.
- Chapman, S.C., Cooper, M., Butler, D.G., Henzell, R.G., 2000. Genotype by environment interactions affecting grain sorghum. I. Characteristics that confound interpretation of hybrid yield. *Aust. J. Agric. Res.* 51, 197-208.
- Dardanelli, J.L., Balzarini, M., Martinez, M.J., Cuniberti, M., Resnik, S., Ramunda, S.F., Herrero, R., Baigorri, H., 2006. Soybean maturity groups, environments, and their interaction define mega-environments for seed composition in Argentina. *Crop Sci.* 46, 1939-1947.
- Di Mauro, G., Cipriotti, P.A., Gallo, S., Rotundo, J.L., 2018. Environmental and management variables explain soybean yield gap variability in Central Argentina. *Eur. J. Agron.* 99, 186-194.
- FAOSTAT, 2016. Crops and Livestock Trade Database. Available URL: <http://www.fao.org/faostat/>.
- Fehr, W.R., Caviness, C.E., 1977. Stages of Soybean Development. Iowa Coop. Ext. Serv. Spec. Rep. 80. Iowa State Univ., Ames.
- French, R., Schultz, J., 1984. Water use efficiency of wheat in a Mediterranean-type environment. I. The relation between yield, water use and climate. *Aust. J. Agric. Res.* 35, 743-764.
- 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, 142-150.
- Hurley, T., Mitchell, P., 2017. Value of neonicotinoid seed treatments to US soybean farmers. *Pest Manag. Sci.* 73, 102-112.
- Jelsma, I., Schoneveld, G.C., Zoomers, A., van Westen, A.C.M., 2017. Unpacking Indonesia's independent oil palm smallholders: an actor-disaggregated approach to identifying environmental and social performance challenges. *Land Use Policy* 69, 281-297.
- Kemarian, A.R., Stockle, C.O., Huggins, D.R., 2004. Variability of barley radiation-use efficiency. *Crop Sci.* 44, 1662-1672.
- Lobell, D.B., Ortiz-Monasterio, J.I., Asner, G.P., Naylor, R.L., Falcon, W.P., 2005. Combining field surveys, remote sensing, and regression trees to understand yield variations in an irrigated wheat landscape. *Agron. J.* 97, 241-249.
- Lollato, R.P., Edwards, J.T., Ochsner, T.E., 2017. Meteorological limits to winter wheat productivity in the U.S. Southern Great Plains. *Field Crops Res.* 203, 212-226.
- Moore, D.I., Gessler, P.E., Nielsen, G.A., Peterson, G.A., 1993. Soil attribute prediction using terrain analysis. *Soil Sci. Soc. Am. J.* 57, 443-452.
- Mourtzinis, S., Conley, S.P., 2017. Delineating soybean maturity groups across the US. *Agron. J.* 109, 1-7.
- 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, 163-172.

- Mourtzinis, S., Rattalino Edreira, J.I., Grassini, P., Roth, A., Casteel, S.N., Ciampitti, I.A., Kandel, H., Kyveryga, P.M., Licht, M.A., Lindsey, L.E., Muller, D.S., Nafziger, E., Naeve, S., Stanley, J., Staton, M., Conley, S.P., 2018. Sifting and winnowing: analysis of farmer field data for soybean in the US North-Central region. *Field Crops Res.* 221, 130–141.
- Rattalino Edreira, J.I., Mourtzinis, S., Conley, S.P., Roth, A., Ciampitti, I.A., Licht, M., Kandel, H., Kyveryga, P.M., Lindsey, L., Muller, D., Naeve, S., Nafziger, E., Specht, J.E., Stanley, J., Staton, M., Grassini, P., 2017. Assessing causes of yield gaps in agricultural areas with diversity in climate and soils. *Agric. For. Meteorol.* 247, 170–180.
- Rattalino Edreira, J.I., Cassman, G.K., Hochman, Z., van Ittersum, K.M., van Bussel, L., Claessens, L., Grassini, P., 2018. Beyond the plot: technology extrapolation domains for scaling out agronomic science. *Environ. Res. Lett.* 13, 054027.
- Sadras, V., Roget, D., O’Leary, G.O., 2002. On-farm assessment of environmental and management constraints to wheat yield and efficiency in the use of rainfall in management constraints to wheat yield and efficiency in the use of rainfall in the Malle. *Aust. J. Agric. Res.* 53, 587–598.
- SAS Institute Inc, 1999. SAS/STAT® User’s Guide, Version8, Cary, NC: SAS Institute Inc.
- Setiyono, T.D., Weiss, A., Specht, J., Bastidas, A.M., Cassman, K.G., Dobermann, A., 2007. Understanding and modeling the effect of temperature and daylength on soybean phenology under high-yield conditions. *Field Crops Res.* 100, 257–271.
- Setiyono, T.D., Cassman, K.G., Specht, J.E., Dobermann, A., Weiss, A., Yang, H., Conley, S.P., Robinson, A.P., Pedersen, P., De Bruin, J.L., 2010. Simulation of soybean growth and yield in near-optimal growth conditions. *Field Crops Res.* 119, 161–174.
- Silva, J.V., Reidsma, P., Laborte, A.G., van Ittersum, M.K., 2017. Explaining rice yields and yield gaps in Central Luzon, Philippines: an application of stochastic frontier analysis and crop modelling. *Eur. J. Agron.* 82, 223–241.
- Tanaka, A., Johnson, J.M., Senthilkumar, K., Akakpo, C., Segda, Z., Yameogo, P.L., Bassoro, I., Lamare, M.D., Allarangaye, D.M., Gbakatchetche, H., Bayuh, A.B., Jaiteh, F., Bam, K.R., Dogbe, W., Sekou, K., Rabeson, R., Rakotoarisoa, M.N., Kamissoko, N., Mossi, M.I., Bakare, S.O., Mabone, L.F., Gasore, R.E., Baggie, I., Kajiru, J.G., Mghase, J., Ablede, A.K., Nanfumba, D., Saito, K., 2017. On-farm rice yield and its association with biophysical factors in sub-Saharan Africa. *Eur. J. Agron.* 85, 1–11.
- Thornton, P.E., Thornton, M.M., Mayer, B.W., Wilhelmi, N., Wei, Y., Devarakonda, R., Cook, R.B., 2017. Daymet: Daily Surface Weather Data on a 1-km Grid for North America, Version 3. ORNL DAAC, Oak Ridge, Tennessee, USA. <https://doi.org/10.3334/ORNLDAAC/1328>. Available URL:.
- Torrion, J., Setiyono, T.D., Cassman, K., Specht, J.E., 2011. Soybean phenology simulation in the north-central United States. *Agron. J.* 103, 1661–1667.
- Villamil, M.B., Davis, V.M., Nafziger, E.D., 2012. Estimating factor contributions to soybean yield from farm field data. *Agron. J.* 104, 881–887.