Field-scale experiments reveal persistent yield gaps in low-input and organic cropping systems

Alexandra N. Kravchenko[a,1], **Sieglinde S. Snapp**[a], **and G. Philip Robertson**[a,b]

[a]Department of Plant, Soil and Microbial Sciences, Michigan State University, East Lansing, MI 48824; and [b]W. K. Kellogg Biological Station, Michigan State University, Hickory Corners, MI 49060

Edited by David Tilman, University of Minnesota, St. Paul, MN, and approved December 7, 2016 (received for review July 26, 2016)

Knowledge of production-system performance is largely based on observations at the experimental plot scale. Although yield gaps between plot-scale and field-scale research are widely acknowledged, their extent and persistence have not been experimentally examined in a systematic manner. At a site in southwest Michigan, we conducted a 6-y experiment to test the accuracy with which plot-scale crop-yield results can inform field-scale conclusions. We compared conventional versus alternative, that is, reduced-input and biologically based–organic, management practices for a corn–soybean–wheat rotation in a randomized complete block-design experiment, using 27 commercial-size agricultural fields. Nearby plot-scale experiments (0.02-ha to 1.0-ha plots) provided a comparison of plot versus field performance. We found that plot-scale yields well matched field-scale yields for conventional management but not for alternative systems. For all three crops, at the plot scale, reduced-input and conventional managements produced similar yields; at the field scale, reduced-input yields were lower than conventional. For soybeans at the plot scale, biological and conventional managements produced similar yields; at the field scale, biological yielded less than conventional. For corn, biological management produced lower yields than conventional in both plot- and field-scale experiments. Wheat yields appeared to be less affected by the experimental scale than corn and soybean. Conventional management was more resilient to field-scale challenges than alternative practices, which were more dependent on timely management interventions; in particular, mechanical weed control. Results underscore the need for much wider adoption of field-scale experimentation when assessing new technologies and production-system performance, especially as related to closing yield gaps in organic farming and in low-resourced systems typical of much of the developing world.

field experiments | scaling | corn soybean wheat rotation | weed control | organic agriculture

In most regions of the world, row-crop farming is primarily conducted in fields of 25–100+ ha, but knowledge of processes and management recommendations are typically based on research in experimental plots of 0.005–0.01 ha. This scale mismatch has long raised questions about inferences that are made from plot-scale experiments to entire fields or farms (1). Although the limitations of plot-scale research are well known for processes that involve mobile taxa, such as insects and livestock, for field crop management the perils of such limitations are commonly overlooked. Calls to close yield gaps, that is, differences between yields currently achieved in farmer fields and those that can be potentially attained, form the basis for addressing future global food needs (2–4). However, because assessments of potentially attainable yields are often informed by inputs from plot-scale research (5), understanding the role of experimental scale is needed to identify the causes for yield gaps and the ways to close them. Experimental-scale–related yield gaps for organic grain crops can be especially substantial. Recent meta-analyses of plot-scale studies suggest organic yield penalties of 20–25% on average (6, 7), although possibly as low as 8% (8). Farmer surveys, on the other hand, report organic grain yield penalties of 27–34% (9). Why the disparity?

First, all crops grown at field scales are subjected to significant soil and topographic diversity, which leads to well-recognized spatial variability in plant growth and crop yield (e.g., 10, 11). This variability is typically minimized in plot-scale studies by blocking, contiguous designs, and judicious plot layouts (12, 13), as well as by locating experiments on more favorable soils (1). Thus, the yields from plot-scale experiments can exceed those from field-scale studies due to inherent locational bias.

Second, and specific to low-input management, field-scale edaphic variability may be more difficult to mitigate by biological than chemical means. Biologically based management systems depend on a wide variety of diverse practices to replenish soil fertility, manage moisture, and control pests. Especially challenging are plant-based fertility systems, that is, those reliant on cover crop and rotational diversity with mixtures of functional plant types, such as legume and grass combinations (14, 15). In these systems, spatial and temporal variations can affect cover crop growth and biomass production (16) as well as the soil fertility benefits that result (17, 18). Even in manure-based fertility systems, biologically based management may be less resistant to biotic stresses, such as weed competition. Low-input management typically relies on mechanical weed control, and key to its success is optimal soil moisture. Either inadequate or excess soil moisture can cause management delays that reduce the efficacy of mechanical control, and thus organic systems are more vulnerable than conventional systems to weather and edaphic variability for effective weed control. Moreover, because organic systems do not have the option of applying "rescue" chemical treatments if pest pressure builds unexpectedly, they

Significance

Meeting future food needs requires a substantial increase in the yields obtained from existing cropland. Prior global analyses have suggested that these gains could come from closing yield gaps—differences between yields from small-plot research versus those in farmer fields. However, closing this gap requires knowledge of causal factors not yet identified experimentally. Results here suggest that yield gaps can be closed using farming practices that use conventional synthetic chemicals, but practices that rely more on biological management—as is the case throughout much of the developing world and in organic agriculture—require renewed attention to field-scale resource demands and place greater emphasis on the importance of field-scale experimental research.

Author contributions: G.P.R. designed research; A.N.K., S.S.S., and G.P.R. performed research; A.N.K. analyzed data; and A.N.K., S.S.S., and G.P.R. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1] To whom correspondence should be addressed. Email: kravche1@msu.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1612311114/-/DCSupplemental.

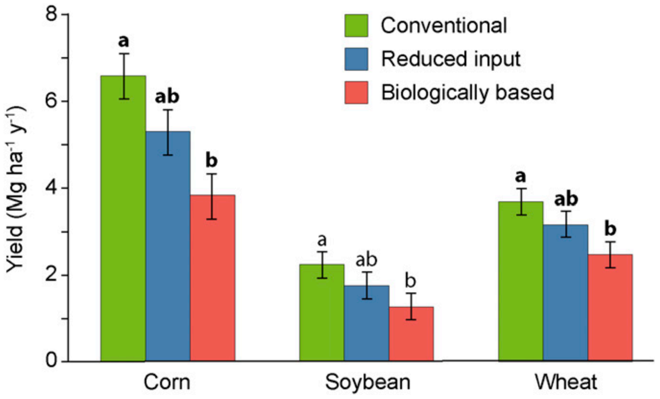

Fig. 1. Field-scale average yields of corn, soybean, and wheat in Conv, RI, and Bio management practices for 2007–2012 ($n = 3$). Error bars represent SEs. Different letters within each crop denote statistically significant differences among the management practices at $P < 0.05$ (boldface type) and $P < 0.1$ levels.

can be especially vulnerable to situations that require extraordinary efforts. An inability to provide proper weed control was regarded as the main reason for much greater yield reductions in organic management in wet compared with dry or normal precipitation springs in plot-scale studies by Posner et al. (19) and Cavigelli et al. (20). At the field scale the influence of spatial and temporal variability on biologically based operations is likely to be even more noticeable than in plot-scale studies and magnified still further at the scale of an entire farm. Because it is not possible to simultaneously operate on all fields, farmer decisions often involve tradeoffs in sequencing, with the likelihood of poorer weed control under suboptimal soil moisture levels. Weed pressure is indeed the factor cited most often by farmers as their greatest management challenge (9). In contrast, in experimental plots, mechanical weed control can generally be conducted at optimal times.

Dependence on experiments with small, relatively homogeneous plots can limit progress in understanding the contribution of spatially variable soil, topographic, and hydrological conditions to outcomes of different agronomic management practices and thus inadvertently add to scale-related incongruities in our understanding of yield gaps. However, at present there is no empirical evidence for how scale can influence the performance of different agricultural management systems and comparisons among them.

Here we compared a replicated experiment conducted for 6 y (2007–2012) at field (6–36 ha) scale with two plot-scale studies at large-plot (1 ha) and small-plot (0.02 ha) scales, referred to as field-scale experiment, main cropping system experiment (MCSE), and living field laboratory (LFL) experiment, respectively. The null hypothesis of the study is that yields of conventional (Conv), reduced-input (RI), and biologically based–organic (Bio) management systems are not affected by the experiment's spatial scale. The three studied row crops are corn (*Zea mays* L.), wheat (*Triticum aestivum* L.), and soybean (*Glycine max* L.). Our objectives are to evaluate the effect of plot versus field scales on the yield performance of these row crops under different management systems and to assess the influence of seasonal rainfall and management factors on differences in performance.

Results

Environment. Growing season precipitation varied greatly over the 6 y of this study, with a high of 540 mm in 2010 and a low of 207 mm in 2012, a regional drought year (Fig. S1). During this period there was a year with an extremely dry summer (2012), two relatively dry years (2007 and 2009), and three normal-wet years (2008, 2010, and 2011). Growing season rainfall distribution varied and was generally favorable for crop growth in the normal years of 2008, 2010, and 2011; generally poor with dry spells in 2007 and 2012; and intermediate in 2009.

The field-scale fields varied substantially in soil and topographical characteristics (Table S1). The within-field topographical gradient ranged from ∼2 m to almost 10 m in different fields with terrain slopes ranging from <1–9%. Within individual fields, soil sand content varied from 2 to >95%, whereas clay content ranged from 3–5% to 20–30%. Bulk density in the soil plow layer (0–20 cm) also varied substantially, from 1.0 to 1.9 g cm^{-3}. Although in the majority of the studied fields, soil organic C in the plowed layer was in the 0.5–1.5% range, two fields in Conv and RI managements had isolated depression areas with soil C as high as 2.5–4.5%.

Field-Scale Yields.

Total yields. Across all 6 y (2007–2012) of the field-scale experiment the yields of all three crops (corn, soybean, and wheat) were the highest in Conv and the lowest in Bio, and RI yields were intermediate, not significantly different from either Conv or Bio (Fig. 1). During the studied period, corn yields varied by a factor of 5, from >10 Mg ha^{-1} in 2010 to ∼2 Mg ha^{-1} in the drought year of 2012 (Table S2). Soybean yields varied by a factor of 3, from ∼1 Mg ha^{-1} to >3 Mg ha^{-1}, whereas wheat yields remained remarkably consistent from year to year in all three studied treatments.

During the studied period (162 crop × management × replicate years) there were 13 instances of crop failure at the field scale (yields too low to harvest and recorded as zero): one event in Conv, six in RI, and six in Bio. Eleven of the failed crops were soybean, with only two instances of a failed corn crop, and none for wheat. Failures appeared to be related to weed infestations, particularly in wet springs, and, in some years, to intense deer browsing. There were no crop failures in either MCSE or LFL sites during this period.

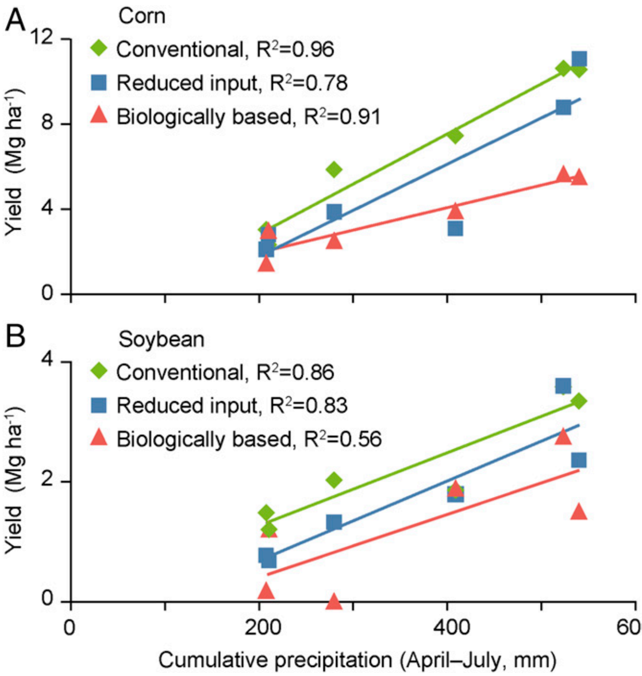

Fig. 2. Relationship between (*A*) average corn and (*B*) soybean yields and cumulative April–July precipitation in the field-scale experiment. All regression models are statistically significant at $P < 0.05$.

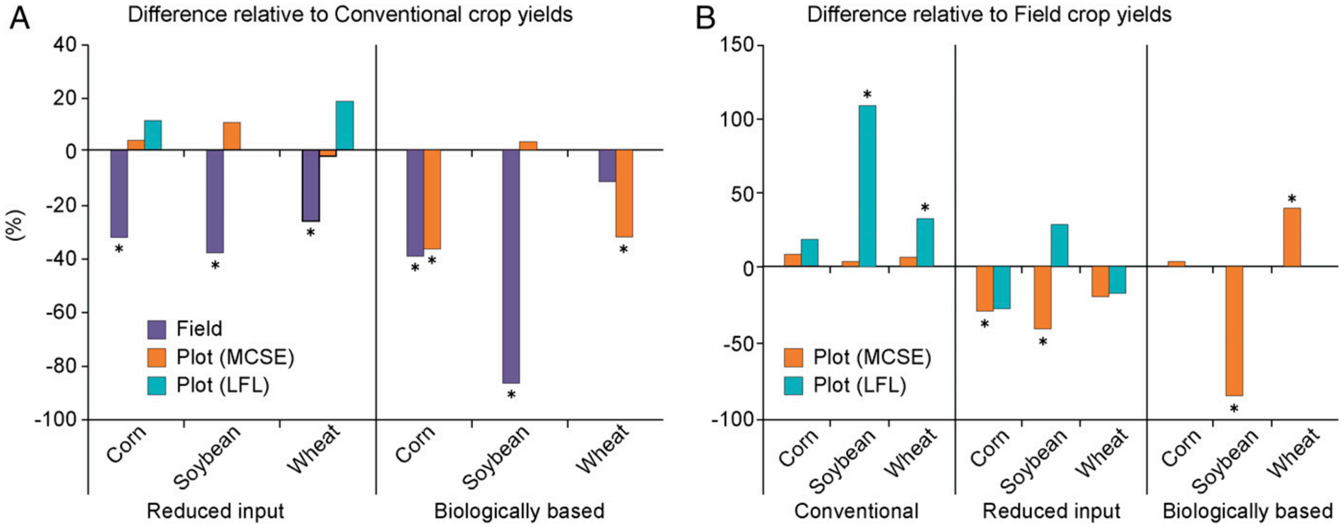

Fig. 3. Relative differences in crop yields (*A*) between conventional vs. reduced-input and biologically based management practices within each experiment and (*B*) between field-scale and plot-scale (MSCE and LFL) experiments for each management practice. The differences are calculated for corn (2008 and 2011), soybean (2009 and 2012), and wheat (2007 and 2010) as percentage of decrease in yield compared with the yield of the conventional management (shown in *A*) and as percentage of difference in yield compared with plot-scale practices for the field ($n = 3$) and plot-scale [MCSE ($n = 6$) and LFL ($n = 4$)] experiments (shown in *B*). Asterisks in *A* mark the differences that were significantly less than zero at $P < 0.05$, and asterisks in *B* mark differences significantly different from zero at $P < 0.1$.

Relationship between precipitation and crop yields. One of the leading factors influencing corn and soybean yields in the field-scale experiment was precipitation in the first part of the growing season from April to July (Fig. 2). In all three management practices there were significant positive correlations between yields and April–July rainfall. The relationship was particularly strong for corn, with R^2 values ranging from 0.78 to 0.96, and somewhat weaker for soybean, with R^2 values of 0.56–0.86. The lowest soybean and corn yields were observed in the three dry years (2007, 2009, and 2012), and the highest yields were observed in the 2 y with the largest amounts of April–July precipitation (2010 and 2011).

The corn yield under Conv management was almost twice that in the Bio management when precipitation levels exceeded 500 mm (Fig. 2). At the same time, corn yield differences among the three managements were minor for the years with low April–July precipitation.

Effect of Experimental Scale. Comparisons in yield performance of RI and Bio managements versus Conv in field-scale, MCSE, and LFL experiments are shown in Fig. 3*A*, and comparisons between the field-scale and plot-scale experiments for each management practice are shown in Fig. 3*B*. Because the MCSE does not have all three crop rotation phases present every year (Materials and Methods), to obtain the most consistent comparison among the experiments we used the field-scale and LFL data of the same crop that was present in that year at the MCSE site: corn yields from 2008 and 2011, soybean yields from 2009 and 2012, and wheat yields from 2007 and 2010.

Conclusions regarding the performance of our management practices obtained in plot-scale experiments did not well match the field-scale results (Fig. 3*A*). The discrepancy was most pronounced for the RI management. In both plot-scale experiments, MCSE and LFL, the yields from RI and Conv managements were very similar to each other. Occasionally, in MCSE and LFL the RI yields were even numerically higher than those of Conv. However, in the field-scale experiment the RI yields were significantly lower than those of Conv for all three crops.

Comparisons between Bio and Conv managements for corn and wheat performance produced consistent results in the field-scale and in the plot-scale experiments (Fig. 3*A*). The Bio yields were lower than those of Conv in both field-scale and MCSE experiments, either numerically (wheat in the field-scale experiment) or statistically (corn in both experiments and wheat in MSCE). However, there was a discrepancy in conclusions for soybean. In the field-scale experiment, the soybean yields were much lower in Bio than in the Conv practice, whereas Bio and Conv yields were very similar in the plot-scale MCSE experiment.

In Conv practice the field-scale yields of all three crops were either very similar to plot-scale yields or even exceeded them; for example, in soybean and wheat the field-scale yields were significantly higher than those from LFL (Fig. 3*B*). However, for RI management there was a tendency for lower yields in field-scale than in plot-scale experiments. In Bio practice the results were inconclusive, with field-scale and MCSE yields similar for corn, field-scale yields lower than MCSE for soybean, and field-scale yields higher than MCSE for wheat. Overall, the differences between the experiments were less pronounced for wheat than for corn and soybean yields.

As expected, the variability in the field-scale fields was substantially higher than that in the MCSE and LFL plots (Table 1). In all studied years the SEs from the field-scale experiment were

Table 1. SEs for the crop yield means obtained by analyzing data from each crop of each of the three experiments separately in each year

		Experiment*		
Year	Crop	Field scale	MCSE	LFL
2007	Wheat	0.44	0.03	0.15
2008	Corn	0.96	0.41	0.16
2009	Soybean	0.55	0.10	0.10
2010	Wheat	0.35	0.22	0.29
2011	Corn	1.70	0.43	0.36
2012	Soybean	0.36	0.12	0.02
Average		**0.73**	**0.22**	**0.18**

*To ensure that the SEs of the MCSE and LFL experiments are comparable to those of the field-scale study, only three randomly selected replications were used from MCSE and LFL sites data. Bold values represent averages across all years.

substantially greater—occasionally as much as 5–10 times greater—than those from the MCSE and LFL sites.

Discussion

Results demonstrate that for conventionally managed corn, soybean, and wheat, plot-scale experiments can accurately represent the yield performance of commercial fields. However, for corn and soybean under reduced-input and biologically based management there was a tendency for higher plot-scale than field-scale yields. This resulted in overall greater yield penalties due to reduced-input and organic managements in the field-scale experiment compared with the plot-scale observations. In the field-scale experiment the magnitudes of the differences between yields of conventional (Conv) and alternative (RI and Bio) managements were affected by weather variations and were particularly large in the years with optimal precipitation.

Effect of Environmental Factors on Management Performance in the Field-Scale Experiment. Greater management differences in corn and soybean performances in years with more spring precipitation (Fig. 2) are consistent with other studies that have reported more pronounced differences between biologically based–organic and conventional practices in years with sufficient precipitation (21) and in irrigated systems (7). In some instances, improvement in soil hydraulic properties under organic management, including soil water holding capacity, has been proposed as a buffering mechanism leading to better plant performance in organically managed soils (21). However, in our study the modest difference in corn yields among Conv, RI, and Bio management in dry years was apparently not an outcome of soil improvement, but of a much greater drop in performance of Conv management crops compared with Bio during dry years. Corn yield under Conv management dropped from 10 Mg ha^{-1} y^{-1} in optimal precipitation years to as low as 2 Mg ha^{-1} y^{-1} in the drought summer of 2012. In Bio, the maximum corn yield was only 4 Mg ha^{-1} y^{-1}, whereas in 2012 it was also ~2 Mg ha^{-1} y^{-1}, that is, similar to that obtained in Conv management.

One possible explanation is that for Conv crops, which are adequately supplied with chemical fertilizer and pest control inputs, the availability of water is likely the main yield-limiting factor. Early season precipitation is known to be positively associated with corn and soybean grain yields and to influence yield response to management (22–24). Unlike Conv crops, RI and especially Bio crops experience nutrient deficiencies and weed and pest pressures even in years with optimal precipitation. Nitrogen deficiencies might be particularly expressed when rainfall is sufficient to support high corn yield and are likely responsible for the most contrasting results in corn yield response to management and their relationships with precipitation.

Scale Effects: Comparisons of Field Versus Plot Scales. Yield penalties for using alternative practices, such as Bio or RI, appear to be greater in the field-scale than in the plot-scale experiments (Fig. 3*A*). In the plot-scale MCSE the yields in Bio were 19% lower than Conv, the difference consistent in size with the published plot-scale literature results (6–8, 25). For example, an average yield loss across six corn, soybean, and/or wheat North American experiments (19, 24–27), where organic management relied on the use of cover crops and not on animal manure inputs, was 19% [from a review by Seufert et al. (7)]. Posner et al. (19) reported that for the majority of times the organic yields were about 10–15% lower than the conventional, whereas 20–30% yield losses in organic management occurred only in one third of the studied site-years. However, Bio yields from the field-scale experiment of this study were >40% lower than Conv (Fig. 1). Specifically, in all studied years after the establishment year of 2007, corn yields in the field-scale experiment Bio treatment were >45% lower than in Conv management, and wheat Bio yields were >30% lower than in Conv management (Table S2). Bio's soybean crop failed in 2009, whereas soybean yield reductions were 55%, 27%, and 88% in 2010, 2011, and 2012, respectively.

Likewise, the yield penalty due to using RI management was higher in the field-scale experiment than in the plot-scale MCSE and LFL (Fig. 3*A*). For 2008–2012, in contrast to >20% losses observed for RI management in the field-scale experiment, RI yields were numerically higher than Conv by 8% in the MCSE and by 13% in the LFL experiments.

We explored two plausible reasons for the observed differences between the field-scale and the plot-scale results. First, the field-scale study and the plot-scale MCSE and LFL experiments are of different durations. The MCSE experiment was established in 1989, and the LFL experiment was established in 1993. Over the years, Bio management practices in both the MCSE (26, 27) and the LFL (28, 29) have resulted in higher levels of soil organic matter compared with Conv management. Thus, the more modest yield losses under Bio management in the MCSE and LFL could have resulted from improvements in soil organic matter. However, this does not appear to be the case: MCSE yields from a 5-y period at the beginning of the study, that is, 1993–1997 (Fig. S2), also show an ~19% reduction in Bio yields compared with Conv management, similar to that in the 2008–2012 period of the present study. Likewise, the difference between RI and Conv yields in 1993–1997 was very similar to that observed in 2008–2012, with RI crop yields 4% higher than Conv yields. Taken together, then, gains in soil organic matter under Bio and RI managements at the MCSE do not seem to explain the much better yield performance at the plot-scale study compared with the field-scale study.

Second, RI and, especially, Bio managements heavily rely on mechanical weed control. Timely and effective administration of mechanical weed control is routinely reported as a challenge even in plot-scale experiments. For example, Cavigelli et al. (20) showed that weed competition accounted for about one quarter of yield losses in an organic management system. Posner et al. (19) found that nearly all instances of substantially poorer organic yield performance in their >20 site-year data set were associated with lack of sufficient weed control in the organic system. Likewise, 20–30% lower yields were associated with poor weed control in a summary of five experiments compiled by Posner et al. (19), whereas no notable yield losses were observed when weed control was adequate. Similar results have been shown in a 16-ha study in Australia, where yields in biologically based management systems were consistently suppressed by weed infestations (30).

In the commercially managed fields of the field-scale experiment, achieving appropriate mechanical weed control appears to have been a greater challenge than at the plot scale. The organic management of corn and soybean relied on the use of early summer cultivation and rotary hoeing to control weeds. The intensity of these weed control operations differed in the field-scale and MCSE experiments. During the studied period the number of weed control operations in the MCSE experiment ranged from 4 to 7 per year (Table S3). In the Bio fields of the field-scale experiment the number of weeding operations was never greater than five, and sometimes as few as two. Soil finishing and preparation of the MCSE was also conducted more thoroughly than could be achieved in the field-scale fields. For example, during the studied period, double-cultivation (cultivation conducted in opposite directions across the plots) was conducted in the MCSE plots twice, whereas only one field of the field-scale experiment was double-cultivated in only one of the studied years (MCSE and field-scale experiment agronomic logs at lter.kbs.msu.edu/datatables). Our results suggest that the experience and resources needed to manage weeds with biologically based–organic management are substantial. These findings are also consistent with the role that timely field operations in general have been hypothesized to play in explaining the yield gap between plot-scale and field-scale crop performance (31).

Implications

There are at least three important implications of these findings. The first is that scaling from plot-level experiments to farm fields can be straightforward for conventional, chemically based management systems, but is significantly more challenging for systems that are more biologically based. Moreover, extrapolation from plot to field scale can be expected to be a problem for any management practice that involves challenges in performing crop and soil management operations, for example, sowing or weed control operations requiring optimal timing, regardless of whether the system in question is conventional or organic. Thus, researchers must be extremely careful when extrapolating plot-level results to farm fields. Likewise, farmers must appropriately resource their adaptations of plot-tested management systems. For example, if plot-scale success is tied in part to frequent mechanical weed control applied in a timely manner, then equivalent field-scale success will likely be possible only when sufficient labor and equipment is available to duplicate these efforts at the farm scale. Although this seems an obvious conclusion, it appears to be overlooked in practice. Extrapolations and appropriate resourcing will be especially challenging in locations and futures with more variable climate.

Second is the need for field-scale research, especially in low-input systems, largely absent today. High costs, along with land and labor requirements, are main reasons that large-scale experiments are so scarce. However, lack of realization by the scientific community of how substantial the differences in plot-scale versus field-scale findings can be, especially for low-input systems with high management requirements, appears to be yet another reason for the shortage of replicated large-scale studies. Our findings provide data to raise awareness that (*i*) despite its costs, field-scale research is crucial for designing resource efficient systems that better match field resource heterogeneities and that (*ii*) farmer recommendations and policies based on poorly scaling plot-level research can cost the public more than investing in large-scale research.

Third, findings suggest that addressing yield gaps among poorly resourced farmers may be especially intractable without additional resources. Seufert et al. (7) found organic performance in developing countries to be even more challenged than in developed countries (43% yield penalties for developing vs. 20% for developed countries). Our findings suggest that this difference will be still greater at the farm scale, emphasizing the special need in developing countries to create technologies that are less time-sensitive and make efficient use of labor. This will be especially important where reduced-input farming is pursued out of necessity rather than choice, for example, in sub-Saharan Africa (32).

Research is sorely needed to design and extend the incorporation of adaptive management to low-input systems at field scales. Mechanisms are needed to better buffer these systems from edaphic and climatic variability and to better respond to fast-acting disturbances, such as weed outbreaks and short-term drought.

Conclusions

Corn, soybean, and wheat yields from plot-scale experiments well represented the yields obtained from commercial-size fields for conventional chemical-input intensive crop management. However, findings from the plot-scale experiments were not always consistent with field-scale outcomes for corn and soybean in reduced-input and biologically based management systems. Challenges associated with timely weed control appear to be a main cause for the discrepancy in our field-scale versus plot-scale comparisons. Differences in performance were particularly large during years with sufficient precipitation, whereas during drought years the differences in corn and soybean yields between alternative and conventional management practices were minimal. Findings emphasize a critical need for field-scale studies to complement widely relied upon plot-scale experimentation. Such studies would be important for any management practice with extra labor, equipment, and/or timing requirements, but are especially important for low-input and biologically based–organic cropping systems.

Materials and Methods

Experimental Sites. The three experiments that represented three spatial scales in this study were located in close proximity to one another on soils of the same series, were randomized replicated experiments, and implemented identical conventional and alternative management systems.

The experiments were conducted at the W. K. Kellogg Biological Station (KBS) Long-Term Ecological Research (LTER) site located in southwest Michigan (42° 24′ N, 85° 24′ W, 288-m elevation). Soils at KBS are well-drained loams developed on glacial outwash; dominant soil series are comingled Kalamazoo (fine-loamy, mixed, mesic Typic Hapludalfs) and Oshtemo (coarse-loamy, mixed, mesic Typic Hapludalfs). KBS receives 1,005 mm of annual precipitation and has mean annual temperature of 10.1 °C. Yields of conventionally managed corn, soybean, and wheat at KBS are similar to both county and average US yields for rainfed crops (33). Additional soil, climatic, and conventional yield details appear in Robertson and Hamilton (33).

The field-scale experiment included 27 cropped fields and was initiated in 2006, with grain yield data collection starting in 2007. Fields ranged in size from 6 to 36 ha, with two thirds of the fields >20 ha. Fields were managed as corn/soybean/winter wheat rotations using conventional, low-input, and biologically based–organic management practices. Each phase of the rotation was represented in each of the studied years for a total of 3 crop phases (corn, soybean, or wheat) × 3 management practices × 3 replicate blocks = 27 fields. Field assignments were blocked by field size such that each year there was one field <20 ha and two fields >20 ha in each crop of each management practice.

Conventional management (Conv) followed Michigan-recommended field crop production practices. Reduced- input management (RI) followed conventional practices with three exceptions: winter cover crops were integrated into the rotation, nitrogen fertilizer was reduced by 70% in corn and by 50% in wheat, and herbicide applications were two thirds reduced by banding herbicides within the rows and using mechanical cultivation to control between-row weeds. The cover crops consisted of cereal rye (*Secale cereale* L.), planted after harvest of corn and plowed under before soybean planting in May, and red clover (*Trifolium pratense* L.), frost-seeded into the winter wheat crop in February and incorporated into the soil with tillage before corn planting 15 mo later. Biologically based–organic management (Bio) followed organic management-recommended practices and included the same cover crops as those used in RI, no nitrogen fertilizer, and only mechanical cultivation to control weeds. No treatments received manure or compost. Chisel plowing (a conservation tillage practice common in the region) was implemented in spring in all three managements. A detailed description of management practices is provided in Robertson and Hamilton (33).

Crop varieties grown in Conv and RI were those recommended for conventional production (corn variety Dekalb DKC52-59; soybean variety Pioneer 92Y30 RR; soft red winter wheat variety Pioneer 25R47). In Bio management we used varieties recommended for organic production (corn variety Blue River Hybrid 22A10; soybean Blue River 19AR1; soft red wheat variety Pioneer 25R47 untreated seed).

We compared the results from the field-scale experiment with results from two adjacent plot-scale experiments: the LTER main cropping system experiment (MCSE), established in 1989, and the living field laboratory (LFL), experiment established in 1993. The management practices, main crops, and cover crops of the MCSE were identical to those used in the field-scale experiment, with the exception that only one phase of the rotation, that is, one crop species, was present each year. In the MCSE each management practice is applied to six replicated 1-ha experimental plots. Detailed descriptions of the MCSE experiment and yield and yield variability comparisons among management practices have been reported by Kravchenko et al. (13), Smith et al. (34), and Robertson et al. (35).

The LFL experiment included Conv and RI treatments that were identical to those used in the MCSE and field-scale experiments. The LFL experimental design is a split-split plot with four randomized complete blocks: management practice is the main plot factor, phase of the rotation is the split-plot factor, and cover crop is the split-split plot factor. Every crop phase was

AGRICULTURAL SCIENCES

SUSTAINABILITY SCIENCE

present every year, as in the field-scale experiment. Individual plots were 9.1 m × 20 m (0.018 ha), which accommodated 12 rows, spaced 0.76 m apart, for corn and soybean and rows 0.19 m apart for wheat.

Soil analytical data for the MCSE and LFL and crop yield data for each of the five studied years (2007–2012) from each studied field-scale field are available at lter.kbs.msu.edu/datatables.

Sampling. In 2010 we collected 65 deep (0–90 cm) soil cores from 10 of the field-scale fields, with 4, 3, and 3 fields in Conv, RI, and Bio managements, respectively. Fields were selected to represent the diversity of soil and topographical characteristics of the studied area. Soil texture, bulk density, and soil organic C were measured in each core within 0–20, 20–35, 35–50, 50–70, and 70–90 cm depth increments. Soils in the MCSE and LFL experiments were systematically soil-sampled on a schedule that included decadal deep profile sampling for the same soil properties.

Crop grain yields were collected via a combine harvester. Crop yields were measured by calibrated truck scale, and grain moisture was monitored using a grain moisture tester (DICKEY-John Corporation, Auburn, IL).

Statistical Analysis. Data analyses were conducted using the PROC MIXED procedure of Statistical Analysis Software (SAS) (36). To compare management practices in the field-scale experiment across all years the statistical model included management practices, crop species, and their interaction as fixed factors; and years and interactions between years, crops, and management practices as random factors. In addition, the blocking factor that grouped fields by size, the effect of individual fields nested within management practices and within the size blocks, and the interaction between crop species and fields were included as random factors in the model. Because there were marked differences in variability of the three studied crops, we conducted heterogeneous variance analysis using the Repeated/group = crop statement in the PROC MIXED procedure (37).

Interactions among the fixed factors were examined using a slicing approach. Mean comparisons were conducted when the respective effects, either main, interaction, or slicing, were found to be statistically significant. The differences are reported either at 0.05 or at 0.1 levels of statistical significance. In all conducted analyses the assumptions of normality of the residuals and homogeneity of variances were checked, and either data transformations or analyses with heterogeneous variances were used as needed.

To address the influence of experiment size on performance of the studied management practices the data of the field-scale, MCSE, and LFL experiments were combined. For the field-scale and LFL experiments in these comparisons we used only data from the crop that was grown in that year in the MCSE experiment, where, as mentioned earlier, only one phase of the rotation was present every year. The statistical model for the combined analysis included experiments, crop species, management practices, and interactions among them as fixed factors, and blocks nested within each respective experiment as a random factor. Unequal variances among crops and experiments were handled by using the Repeated/group = crop × experiment statement in PROC MIXED.

The SEs for treatment means can be viewed as a measure of how accurately the means of the studied management practices can be estimated, given the study design, the number of replications, and the variability of different random sources present in the experiment. To assess the accuracy of the studied experiments, we used their SEs as a response variable in the data analysis that was conducted separately in each year of each experiment. For SE comparisons among the management practices the model consisted only of the fixed effect of management practice and the random effect of blocks. The studied experiments had different numbers of replications, that is, three in the field scale, six in the MCSE, and four in the LFL experiments. Thus, to ensure unbiased comparisons among them in terms of their SE values, the data from only three randomly selected replications of MCSE and LFL studies were used in this analysis.

ACKNOWLEDGMENTS. We are indebted to B. D. Knezek for early discussions that led to the initiation of this study; to J. A. Bronson for agronomic management of the field-scale experiment; to J. T. Simmons, M. A. Halvorson, G. Parker, T. Martin, and others for agronomic management of the MCSE and LFL experiments; and to S. M. Swinton, S. K. Hamilton, K. L. Gross, S. H. Gage, D. A. Landis, and T. M. Schmidt for helpful discussions during the course of the experiment and manuscript preparation. We are grateful to R. R. Harwood for initiating the LFL experiment. Funding has been provided by the National Science Foundation's Long-Term Ecological Research Program (DEB 1027253) and by Michigan State University AgBioResearch.

1. Robertson GP, Burger LW, Kling CL, Lowrance R, Mulla DJ (2007) New approaches to environmental management research at landscape and watershed scales. *Managing Agricultural Landscapes for Environmental Quality*, eds Schnepf M, Cox C (Soil and Water Conserv Soc, Ankeny, Iowa), pp 27–50.
2. Foley JA, et al. (2011) Solutions for a cultivated planet. *Nature* 478(7369):337–342.
3. Mueller ND, et al. (2012) Closing yield gaps through nutrient and water management. *Nature* 490(7419):254–257.
4. van Ittersum MK, et al. (2013) Yield gap analysis with local to global relevance—A review. *Field Crops Res* 143:4–17.
5. Lobell DB, Cassman KG, Field CB (2009) Crop yield gaps: Their importance, magnitudes, and causes. *Annu Rev Environ Resour* 34:179–204.
6. de Ponti T, Rijk B, van Ittersum MK (2012) The crop yield gap between organic and conventional agriculture. *Agric Syst* 108:1–9.
7. Seufert V, Ramankutty N, Foley JA (2012) Comparing the yields of organic and conventional agriculture. *Nature* 485(7397):229–232.
8. Badgley C, et al. (2007) Organic agriculture and the global food supply. *Renew Agr Food Syst* 22(2):86–108.
9. McBride WD, Greene C, Foreman L, Ali M (2015) *The Profit Potential of Certified Organic Field Crop Production* (U.S. Dept of Agric, Econ Res Serv, Washington, DC).
10. Kravchenko AN, Robertson GP, Thelen KD, Harwood RR (2005) Management, topographical, and weather effects on spatial variability of crop grain yields. *Agron J* 97(2):514–523.
11. Florin MJ, McBratney AB, Whelan BM (2009) Quantification and comparison of wheat yield variation across space and time. *Eur J Agron* 30(3):212–219.
12. Jaynes DB, Kaspar TC, Colvin TS, James DE (2003) Cluster analysis of spatiotemporal corn yield patterns in an Iowa field. *Agron J* 95(3):574–586.
13. Kravchenko AN, Robertson GP, Hao X, Bullock DG (2006) Management practice effects on surface total carbon: Difference in spatial variability patterns. *Agron J* 98(6):1559–1568.
14. Drinkwater LE, Snapp SS (2007) Nutrients in agroecosystems: Rethinking the management paradigm. *Adv Agron* 92:163–186.
15. Pearson CJ (2007) Regenerative, semi-closed systems: A priority for twenty-first-century agriculture. *Bioscience* 57(5):409–418.
16. Munoz JD, Steibel JP, Snapp S, Kravchenko AN (2014) Cover crop effect on corn growth and yield as influenced by topography. *Agric Ecosyst Environ* 189:229–239.
17. Ladoni M, Basir A, Robertson GP, Kravchenko AN (2016) Scaling-up: Cover crops differentially influence soil carbon in agricultural fields with diverse topography. *Agric Ecosyst Environ* 225:93–103.
18. Ladoni M, Kravchenko AN, Robertson GP (2015) Topography mediates the influence of cover crops on soil nitrate levels in row crop agricultural systems. *PLoS One* 10(11):e0143358.
19. Posner JL, Baldock JO, Hedtcke JL (2008) Organic and conventional production systems in the Wisconsin Integrated Cropping Systems Trials: I. Productivity 1990–2002. *Agron J* 100(2):253–260.
20. Cavigelli M, Teasdale JR, Conklin AE (2008) Long-term agronomic performance of organic and conventional field crops in the mid-Atlantic region. *Agron J* 100(3):785–794.
21. Lotter D, Seidel R, Liebhardt W (2003) The performance of organic and conventional cropping systems in an extreme climate year. *Am J Altern Agr* 18(3):146–154.
22. Oberle SL, Keeney DR (1990) Soil type, precipitation, and fertilizer n effects on corn yieldS. *J Prod Agr* 3(4):522–527.
23. Twine TE, Kucharik CJ (2009) Climate impacts on net primary productivity trends in natural and managed ecosystems of the central and eastern United States. *Agric For Meteorol* 149(12):2143–2161.
24. Wilhelm WW, Wortmann CS (2004) Tillage and rotation interactions for corn and soybean grain yield as affected by precipitation and air temperature. *Agron J* 96(2):425–432.
25. Pacini C, Wossink A, Giesen G, Vazzana C, Huirne R (2003) Evaluation of sustainability of organic, integrated and conventional farming systems: a farm and field-scale analysis. *Agric Ecosyst Environ* 95(1):273–288.
26. Senthilkumar S, Kravchenko AN, Robertson GP (2009) Topography influences management system effects on total soil carbon and nitrogen. *Soil Sci Soc Am J* 73(6):2059–2067.
27. Syswerda SP, Corbin AT, Mokma DL, Kravchenko AN, Robertson GP (2011) Agricultural management and soil carbon storage in surface vs. deep layers. *Soil Sci Soc Am J* 75(1):92–101.
28. Sánchez JE, et al. (2004) Managing soil carbon and nitrogen for productivity and environmental quality. *Agron J* 96(3):769–775.
29. Snapp SS, Gentry LE, Harwood RR (2010) Management intensity - not biodiversity - the driver of ecosystem services in a long-term row crop experiment. *Agric Ecosyst Environ* 138:242–248.
30. Penfold C, Miyan M, Reeves T, Grierson I (1995) Biological farming for sustainable agricultural production. *Aust J Exp Agr* 35(7):849–856.
31. Hillocks RJ (2014) Addressing the yield gap in Sub-Saharan Africa. *Outlook Agric* 43(2):85–90.
32. Snapp SS, Blackie MJ, Gilbert RA, Bezner-Kerr R, Kanyama-Phiri GY (2010) Biodiversity can support a greener revolution in Africa. *Proc Natl Acad Sci USA* 107(48):20840–20845.
33. Robertson GP, Hamilton SK (2015) Long-term ecological research in agricultural landscapes at the Kellogg Biological Station LTER site: Conceptual and experimental framework. *The Ecology of Agricultural Landscapes: Long-Term Research on the Path to Sustainability*, eds Hamilton SK, Doll JE, Robertson GP (Oxford Univ Press, New York, New York), pp 1–32.
34. Smith RG, Menalled FD, Robertson GP (2007) Temporal yield variability under conventional and alternative management systems. *Agron J* 99(6):1629–1634.
35. Robertson GP, et al. (2014) Farming for ecosystem services: An ecological approach to production agriculture. *Bioscience* 64(5):404–415.
36. SAS (2009) *SAS/STAT 9.2 User's Guide* (SAS Inst Inc, Cary, NC), 2nd Ed.
37. Milliken G, Johnson D (2009) Analysis of Messy Data. Volume 1: Designed Experiments (CRC Press, Boca Raton, FL), 2nd Ed.

Supporting Information

Kravchenko et al. 10.1073/pnas.1612311114

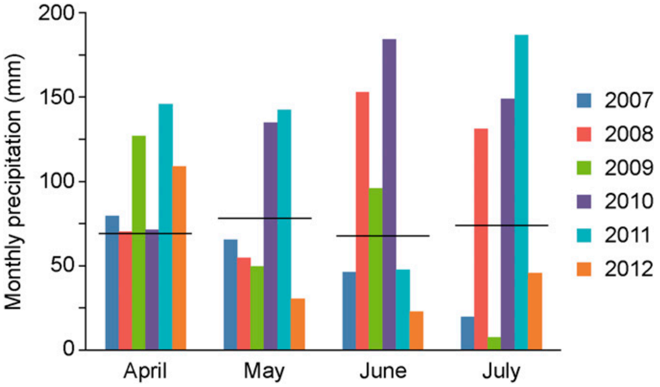

Fig. S1. Monthly precipitation from April–July for 2007–2012. Horizontal lines represent the average monthly precipitation for 1988–2012.

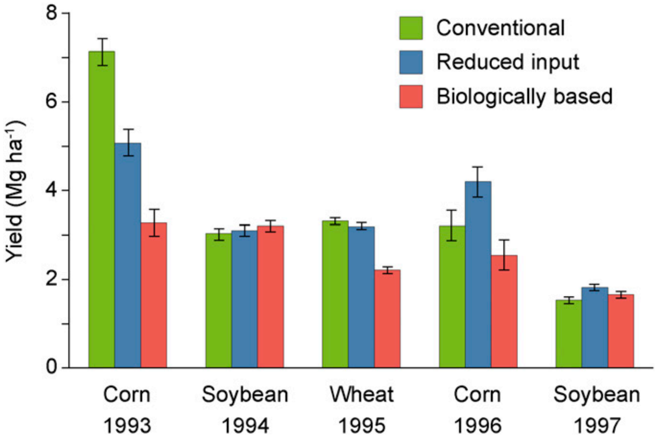

Fig. S2. Average yields from the MCSE experiment during 1993–1997.

Table S1. **Soil properties and topographical characteristics of the field-scale fields (based on 65 soil cores from 10 fields)**

Soil/topographical characteristic	Mean	Coefficient of variation	Minimum	Maximum
Sand content (%)				
0–20cm	57	31	4	90
20–35 cm	57	38	3	91
35–50 cm	64	40	2	95
50–70 cm	73	34	2	96
70–90 cm	77	32	5	98
Clay content (%)				
0–20 cm	11	39	5	24
20–35 cm	13	46	1	26
35–50 cm	13	54	3	36
50–70 cm	10	61	1	35
70–90 cm	8	67	2	24
Bulk density (g cm^{-3})				
0–20 cm	1.52	12	1.02	1.95
20–35 cm	1.73	9	1.29	1.99
35–50 cm	1.78	10	1.37	2.22
50–70 cm	1.80	8	1.38	2.09
70–90 cm	1.83	8	1.42	2.13
Soil organic C (%)				
0–20 cm	1.00	49	0.49	4.07
20–35 cm	0.54	64	0.15	2.06
35–50 cm	0.31	88	0.01	1.04
50–70 cm	0.24	116	0.01	1.06
70–90 cm	0.18	150	0.01	1.11
Range of elevation within the field (m)	5.0	55	1.9	9.8
Terrain slope (%)	2.6	76	0.3	9.0

Table S2. **Average corn, soybean, and wheat yields (Mg ha^{-1} y^{-1}) during individual years for conventional (Conv), reduced input (RI), and biologically based–organic (Bio) management practices in the field-scale experiment**

	Management practice, Mg ha^{-1} y^{-1}		
Year	Conv	RI	Bio
Corn (1.21)*			
2007	2.3	2.8	3.0
2008	$7.5^{\mathbf{a}}$	$3.1^{\mathbf{b}}$	$3.8^{\mathbf{b}}$
2009	5.9	3.9	2.5
2010	$10.5^{\mathbf{a}}$	$11.1^{\mathbf{a}}$	$5.5^{\mathbf{b}}$
2011	10.6^{a}	8.7^{ab}	5.5^{b}
2012	2.9	2.1	1.5
Soybean (0.45)*			
2007	1.2	0.7	1.2
2008	1.9	1.8	1.9
2009	$2.0^{\mathbf{a}}$	$1.3^{\mathbf{b}}$	$0.0^{\mathbf{b}}$
2010	$3.3^{\mathbf{a}}$	$2.4^{\mathbf{ab}}$	$1.5^{\mathbf{b}}$
2011	3.6	3.6	2.8
2012	1.4	0.8	0.4
Wheat (0.44)*			
2007	2.7	2.3	2.8
2008	$3.4^{\mathbf{a}}$	$3.1^{\mathbf{ab}}$	$1.9^{\mathbf{b}}$
2009	$3.9^{\mathbf{a}}$	$3.2^{\mathbf{ab}}$	$2.3^{\mathbf{b}}$
2010	3.5^{a}	2.2^{b}	2.5^{ab}
2011	$4.3^{\mathbf{a}}$	$3.7^{\mathbf{a}}$	$2.3^{\mathbf{b}}$
2012	$4.2^{\mathbf{a}}$	$4.4^{\mathbf{a}}$	$2.8^{\mathbf{b}}$

*SEs for each crop are shown in parentheses. Boldface superscript letters within each year of each crop indicate statistically significant differences among the management practices ($P < 0.05$); nonboldface superscript letters correspond with $P < 0.1$.

Table S3. The number of postplanting mechanical weeding operations in corn and soybean for Bio management at the MCSE and field-scale experiments during 2008–2012

		No. of postplanting weeding operations	
		Experiment	
Year	Rotation phase	MCSE	Field scale
2008	Corn	6	4
2009	Soybean	7	3.3
2011	Corn	4	4.7
2012	Soybean	5	4

The number of weeding operations shown for the field-scale experiment is the average from all of the fields in the particular crop in that year.