Lehrstuhl für Pflanzenernährung

Yield stability, yield development, and breeding progress in conventional and organic agriculture

Samuel Knapp

Vollständiger Abdruck der von der TUM School of Life Sciences der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Agrarwissenschaften (Dr. agr.)

genehmigten Dissertation.

Vorsitzender:	Prof. Dr. Hans-Rudolf Fries
Prüfer der Dissertation:	1. Prof. Dr. Urs Schmidhalter
	2. Prof. Dr. Kurt-Jürgen Hülsbergen

Die Dissertation wurde am 13.07.2020 bei der Technischen Universität München eingereicht und durch die TUM School of Life Sciences am 01.02.2021 angenommen.

Table of Contents

List of Figures

List of Tables

Summary

In sight of the predicted increase in human population, the widespread negative environmental impacts of agricultural production, and more likely extreme weather events, agricultural production needs to increase its production, while at the same time producing more sustainably and more stable. Among many approaches to change production towards greater sustainably, organic production aims to reduce its negative impacts on the environment by not using mineral fertilizers and synthetic pesticides. While the environmental effects of organic farming have been investigated frequently, the effects on yield stability and long-term yield development, as an indicator of sustainability, have been hardly studied. Due to the limited potential of increasing inputs under organic production, improved varieties might be an important way to increase yields. However, while the contributions of plant breeding to overall yield increases have been shown in many studies, the contribution of plant breeding under organic farming has not been investigated yet. Furthermore, it remains an open question, if separate breeding for organic farming will be more beneficial than relying on varieties bred for conventional growing conditions.

Four datasets were used to investigate the different research questions: (1) a global meta-analysis including 165 paired long-term yield observations from several crops; (2) yield data of several crops from a long-term trial in Switzerland, where different conventional and organic management practices with regular and half fertilization were compared for 40 years; (3) yield and quality data from conventionally and organically managed variety recommendation trials of winter wheat from Germany from 2001 to 2017; and (4) conventional and organic on-farm yields of winter wheat from Germany. Yield stability was assessed by the variance and the coefficient variation, correcting for the yield level, across years in the meta-analysis and in the long-term trial. Yield development was assessed by regression on the year of observation in the long-term trials, variety recommendation trials, and on-farm data. Breeding progress was assessed through a mixed model approach to separate genetic and non-genetic and regression on year of variety release within single trials in the variety recommendation trials. Similarity of variety performance between systems was assessed by analysis of variance components and correlation of variety means in the variety recommendation trials.

Results on the yield stability were similar in both the meta-analysis and the long-term trial. Absolute stability, as measured by the variance across years, was similar between both management systems. Relative stability, as measured by the coefficient of variation and thus correcting for the difference in yield level, was more stable under conventional management, which was due to the higher yields under conventional management. While differences in yield development in the long-term trial were minor between the systems, conventional yields increased in the variety recommendation trials and on-farm. In contrast, organic yields stagnated in both latter datasets. In the variety recommendation trials, estimates for breeding progress were significantly positive for both systems. However, the estimates for organic management are possibly over-estimated due to variety ageing because of breaking of disease-resistances, which in-turn could imply that effective breeding progress could tend towards zero under organic management. Organically bred varieties showed an overall higher quality for the traits protein content, sedimentation value and baking volume, but lower yields and similar grain N uptake compared with conventionally bred varieties. For yield and quality traits, analysis of variance components revealed variety x system interaction to be low and variety means were highly correlated between systems.

Overall, results suggest that the lower yield level through reduced inputs in organic farming leads to reduced relative stability, stagnating yields and not benefitting from the increased yield potential of modern varieties. In both the meta-analysis and long-term trial, the yield differences were generally related to the difference in nitrogen (N) input, while in the long-term trial a particularly strong relation to the share of N applied in mineral form (also within organic fertilizers) was observed in winter wheat and potatoes. However, because in the long-term trial, application rates across treatments were correlated to the intensity of plant protection, a clear separation of the contribution of both effects was not possible. Based on the yield level of the untreated conventional intensity in the variety recommendation trials, the yield difference between conventional and organic management in winter wheat in Germany might be for one third due to plant protection and for two-thirds to the difference in fertilization. This indicates that increasing N availability through, e.g., better N management could contribute to increased yields and thus increased relative stability and benefitting from modern varieties. Although plant breeding might have not contributed to yield increases under organic management, it showed to be successful in counteracting the breaking of disease-resistances and thus to maintain the yield level, which is an important part of plant breeding and might have been under-estimated so far. Even though direct selection under organic management did not show to be more beneficial and variety means were strongly correlated between systems, intensity-specific traits like disease resistances or weed suppression might be important for each management intensity.

Zusammenfassung

Angesichts der vorhergesagten Zunahme der Weltbevölkerung, der negativen Umweltauswirkungen der landwirtschaftlichen Produktion und der häufigeren extremen Wetterereignisse muss die landwirtschaftliche Produktion gesteigert und gleichzeitig nachhaltiger und stabiler produziert werden. Unter vielen Ansätzen, um die Produktion nachhaltiger zu gestalten, verfolgt der ökologische Landbau das Ziel, die negativen Auswirkungen auf die Umwelt zu verringern, indem auf den Einsatz von Mineraldüngern und synthetischen Pestiziden verzichtet wird. Während die Umweltauswirkungen des ökologischen Landbaus häufig untersucht wurden, sind die Auswirkungen auf die Ertragsstabilität und die langfristige Ertragsentwicklung als Indikator für die Nachhaltigkeit kaum untersucht worden. Aufgrund der begrenzten Möglichkeiten erhöhter Inputs im ökologischen Landbau könnten verbesserte Sorten einen wichtigen Beitrag zur Ertragssteigerung darstellen. Während jedoch der Beitrag der Pflanzenzüchtung zur allgemeinen Ertragssteigerung in vielen Studien gezeigt wurde, ist der Beitrag der Pflanzenzüchtung im ökologischen Landbau noch nicht untersucht worden. Darüber hinaus ist es weiterhin eine offene Frage, ob eine spezielle Züchtung für den ökologischen Landbau vorteilhafter ist als auf Sorten zurück zu greifen, die für konventionelle Anbaubedingungen gezüchtet wurden.

Zur Untersuchung der verschiedenen Forschungsfragen wurden vier Datensätze verwendet: (1) eine globale Meta-Analyse mit 165 gepaarten Langzeit-Ertragsbeobachtungen mehrerer Kulturen, (2) Ertragsdaten mehrerer Kulturen aus einem Langzeitversuch in der Schweiz, bei dem verschiedene konventionelle und ökologische Bewirtschaftungsmethoden mit normaler und halber Düngung über 40 Jahre verglichen wurden, (3) Ertrags- und Qualitätsdaten aus konventionell und ökologisch bewirtschafteten Sortenempfehlungsversuchen mit Winterweizen aus Deutschland von 2001 bis 2017, und (4) konventionelle und ökologische Praxis-Erträge von Winterweizen aus Deutschland. Die Ertragsstabilität wurde mit der Varianz und dem Variationskoeffizienten, der für das Ertragsniveau korrigiert, über Jahre in der Meta-Analyse und im Langzeitversuch gemessen. Die Ertragsentwicklung wurde durch Regression auf das Beobachtungsjahr im Langzeitversuch sowie in den Sortenempfehlungsversuchen und den Praxiserträgen bestimmt. Der Züchtungsfortschritt wurde durch ein gemischtes Modell zur Trennung von genetischem und nicht-genetischem Trend und durch Regression auf das Jahr der Sortenzulassung innerhalb einzelner Versuche in den Sortenempfehlungsversuchen untersucht. Das Verhältnis der Sortenleistung zwischen den Systemen wurde durch eine Analyse der Varianzkomponenten und durch Korrelation der Sortenmittelwerte in den Sortenempfehlungsversuchen untersucht.

Die Ergebnisse zur Ertragsstabilität waren sowohl in der Meta-Analyse als auch im Langzeitversuch ähnlich. Die absolute Stabilität, gemessen durch die Varianz über die Jahre, war bei beiden Anbausystemen ähnlich. Die relative Stabilität, gemessen durch den Variationskoeffizienten und damit für die Unterschiede im Ertragsniveau korrigiert, war bei konventioneller Bewirtschaftung höher, was auf die höheren Erträge bei konventioneller Bewirtschaftung zurückzuführen war. Während die Unterschiede in der Ertragsentwicklung im Langzeitversuch zwischen den Systemen geringfügig waren, stiegen die konventionellen Erträge in den Sortenempfehlungsversuchen und in den Praxiserträgen. Im Gegensatz dazu stagnierten die ökologischen Erträge in den beiden letztgenannten Datensätzen. In den Sortenempfehlungsversuchen waren die Schätzer für den Züchtungsfortschritt für beide Systeme signifikant positiv. Es zeigte sich jedoch, dass der Schätzer für die ökologische Bewirtschaftung möglicherweise überschätzt wurde, da bei älteren Sorten Krankheitsresistenzen weniger wirksam sind. Dies könnte wiederum bedeuten, dass der effektive Züchtungsfortschritt unter ökologischer Bewirtschaftung gegen Null gehen könnte. Ökologisch gezüchtete Sorten zeigten insgesamt eine höhere Qualität bei den Merkmalen Proteingehalt, Sedimentationswert und Backvolumen, aber geringere Erträge und eine ähnliche Korn-N-Aufnahme im Vergleich zu konventionell gezüchteten Sorten. Bei den Ertrags- und Qualitätsmerkmalen zeigte die Analyse der Varianzkomponenten, dass die Interaktion zwischen Sorte und System gering war und die Sortenmittelwerte zwischen den Systemen hoch korreliert waren.

Insgesamt deuten die Ergebnisse darauf hin, dass das niedrigere Ertragsniveau durch den geringeren Input im ökologischen Landbau zu einer geringeren relativen Stabilität führt, die Erträge stagnieren und nicht vom erhöhten Ertragspotenzial moderner Sorten profitiert werden kann. Sowohl in der Meta-Analyse als auch im Langzeitversuch hingen die Ertragsunterschiede im Allgemeinen mit Stickstoff (N) Input zusammen, während im Langzeitversuch eine besonders starke Beziehung zum Anteil des in mineralischer Form ausgebrachten N (auch in organischen Düngern) bei Winterweizen und Kartoffeln beobachtet wurde. Da jedoch im Langzeitversuch die Aufwandmengen über die Behandlungen hinweg mit der Intensität des Pflanzenschutzes korreliert waren, war eine klare Trennung des Beitrags beider Effekte nicht möglich. Basierend auf dem Ertragsniveau der unbehandelten konventionellen Stufe in den Sortenempfehlungsversuchen könnte der Ertragsunterschied zwischen konventioneller und organischer Bewirtschaftung zu einem Drittel auf den Pflanzenschutz und zu zwei Dritteln auf den Unterschied in der Düngung zurückzuführen sein. Dies deutet darauf hin, dass eine Erhöhung der N-Verfügbarkeit, z.B. durch ein besseres N-Management, zu höheren Erträgen und damit zu einer besseren relativen Stabilität beitragen könnte und besser von modernen Sorten profitiert werden könnte. Obwohl die

Pflanzenzüchtung möglicherweise nicht zu Ertragssteigerungen unter biologischem Management beigetragen hat, zeigte sich, dass sie erfolgreich dem Brechen von Krankheitsresistenzen entgegenwirken und so das Ertragsniveau aufrechterhalten konnte, welches ein wichtiger Bestandteil der Pflanzenzüchtung ist und bisher möglicherweise unterschätzt wurde. Auch wenn sich die direkte Selektion unter biologischer Bewirtschaftung nicht als vorteilhafter erwies und die Sortenmittel zwischen den Systemen stark korreliert waren, könnten bewirtschaftungsspezifische Merkmale wie Krankheitsresistenzen oder Unkrautunterdrückung für die unterschiedlichen Bewirtschaftungsintensitäten wichtig sein.

1 General introduction

Environmental impacts, growth in world population and climate change

Intensive conventional agriculture has more than tripled yield in the last century (Godfray et al., 2010; Foley et al., 2011; Tilman et al., 2011). However the use of pesticides and mineral fertilizers in conventional agriculture often has a negative impact on the environment like soil degradation, loss of biodiversity, increases in greenhouse gas emission, pollution and eutrophication of water (McLaughlin and Mineau, 1995; Johnson et al., 2007; Moss, 2008; Foley et al., 2011; Tilman et al., 2011; Godfray and Garnett, 2014).

As climate change will lead to greater fluctuations, more extreme weather events, and changing and less predictable climate, agricultural production has to be more resilient against such fluctuations to guarantee future regional and global food security (Howden et al., 2007; Schmidhuber and Tubiello, 2007).

Continuing population and consumption growth will mean that the global demand for food will increase for at least another 40 years (Godfray et al., 2010; Foley et al., 2011; Tilman et al., 2011). If the further increase in production shall not be based on additional land usage and thus reduction of natural ecosystems, production per area needs to be increased, or changes towards less animal-based diets are needed (Phalan et al., 2011; Springmann et al., 2016).

The challenge is thus, to maintain or better increase productivity, in a more stable way, and with less environmental impact, i.e. more sustainable (Tilman et al., 2011). This requires a multifaceted global strategy at all scales, from farm to global level, including factors such as reducing food production limits, reducing temporal yield variability, reducing food waste and changing diets (Godfray and Garnett, 2014).

Organic agriculture

Among several approaches to reduce negative environmental impacts, organic farming has been established to achieve this goal through the non-usage of mineral fertilizers and synthetic pesticides combined with improved crop rotations, increased biodiversity on several levels and soil improvement to deal with diseases and maintain productivity (Reganold and Wachter, 2016). Several studies have shown the reduced negative environmental impacts of organic compared to intense conventional systems (Reganold and Wachter, 2016; Smith et al., 2019).

Due the non-usage of mineral fertilizers and synthetic pesticides, in contrast to conventional high-input farming, organic farming conditions can thus be generally characterized by decreased levels of available plant nutrients (mainly nitrogen), and less abilities to control pests and diseases. Yields of organic farming are thus often reduced in comparison to conventional farming. When analyzing the yield gap between organic and conventional farming, several studies have found similar estimates of around 20% for the overall yield gap, but all of them noted that the variation in yield gap between crops and regions is substantial (de Ponti et al., 2012; Seufert et al., 2012; Ponisio et al., 2015).

Importance of temporal yield stability

Many factors can cause yield of crop species to vary across years, including differences in precipitation, temperature, pest outbreaks, weed pressure, soil fertility, soil structure and agricultural management (Seufert and Ramankutty, 2017). Due to the predicted increase in extreme weather events and more unpredictability of climate, stability of cropping systems will become more important as a mean to buffer against this greater and more unpredictable variability.

The concept of yield stability was originally developed in plant breeding (Lin et al., 1986; Becker and Léon, 1988). In recent years it has also received increased interest from ecologists, especially in relation to the stability of ecosystem functioning (Tilman et al., 2006; Hautier et al., 2015), and in comparing the temporal yield stability of different management systems (Becker and Léon, 1988; Smith et al., 2007; Reckling et al., 2018).

Plant breeding and organic agriculture

Due the non-usage of mineral fertilizers and synthetic pesticides, in contrast to conventional farming, organic farming conditions can thus be generally characterized by decreased levels of available plant nutrients (mainly nitrogen), and less abilities to control pests and diseases. As growing conditions differ substantially from conventional farming, several authors have argued for the necessity of direct breeding and separate testing of varieties for organic farming, particularly with regard to nitrogen use efficiency and improved resistances (Murphy et al., 2007; Lammerts van Bueren et al., 2011). In response, public initiatives, scientific projects, and plant breeding companies have initiated special breeding programs and testing of available varieties under organic conditions. However, as several authors have found that the performance of varieties is well correlated between both systems (Przystalski et al., 2008; Hildermann et al., 2009), this issue remains a matter of debate both in the scientific community and among breeders and farmers (Reynolds and Braun, 2019; Voss-Fels et al., 2019).

Objectives and datasets

The objectives of this thesis were thus to compare the yield stability, yield development and breeding progress between conventional and organic farming.

Four different datasets were used to assess these objectives:

- Data from a previous meta-analysis comparing the yield levels between conventional and organic farming, containing yield observations from several crops and of global origin (Ponisio et al., 2015)
- Data from the DOK long-term trial in Switzerland, in which four different conventional and organic systems at regular and half fertilization were compared for 40 years. These data included yield observations from several crops and the amounts of applied nutrients.
- Data from variety recommendation trials of winter wheat in Germany from 2001-2017, including observations on yield and quality traits.
- Data from on-farm yields of winter wheat in Germany

Structure of the thesis

As the datasets and corresponding methods were rather different, the thesis is structured into three main chapters, where each chapter addresses one or two of the objectives on a separate dataset. A specific introduction addressing the main objectives will be given in each chapter.

In chapter A, the yield stability of conventional and organic farming will be compared employing a meta-analysis approach and using the data from Ponisio et al. (2015).

In chapter B, the yield stability and yield development of conventional and organic farming will be compared in the DOK long-term trial. Additionally, the mean yields will be compared and related to nutrient inputs.

In chapter C, the overall yield development of winter wheat on-farm and in the variety recommendation trials will be compared between conventional and organic agriculture. Furthermore, the breeding progress and variety performance will be compared between both systems, and it will be assessed if organically bred varieties perform better than conventionally bred varieties under organic conditions.

2 Chapter A: A global meta-analysis of yield stability in organic and conservation agriculture

2.1 Abstract

One of the primary challenges of our time is to enhance global food production and security. Most assessments in agricultural systems focus on plant yield. Yet, these analyses neglect temporal yield stability, or the variability and reliability of production across years. Here we perform a meta-analysis to assess temporal yield stability of three major cropping systems: organic agriculture and conservation agriculture (no-tillage) versus conventional agriculture, comparing 193 studies based on 2896 comparisons. Organic agriculture has, per unit yield, a significantly lower temporal stability (-15%) compared to conventional agriculture. Thus, although organic farming promotes biodiversity and is generally more environmentally friendly, future efforts should focus on reducing its yield variability. Our analysis further indicates that the use of green manure and enhanced fertilisation can reduce the yield stability gap between organic and conventional agriculture. The temporal stability (-3%) of no-tillage does not differ significantly from those of conventional tillage indicating that a transition to no-tillage does not affect yield stability.

2.2 Introduction

Continuing population and consumption growth will mean that the global demand for food will increase for at least another 40 years (Godfray et al., 2010; Foley et al., 2011; Tilman et al., 2011). It is, thus, a key challenge to enhance food security. This requires a multifaceted global strategy at all scales, from farm to global level, including factors, such as reducing food production limits, reducing temporal yield variability, reducing food waste and changing diets (Godfray and Garnett, 2014). Moreover, stable food production will be a greater challenge under a changing and less predictable climate (Schmidhuber and Tubiello, 2007).

In addition to the challenges of enhancing food security, there is a growing recognition that agriculture must produce more sustainably. Intensive conventional agriculture has more than tripled yield in the last century (Godfray et al., 2010; Foley et al., 2011; Tilman et al., 2011). However, the use of pesticides and mineral fertilisers in conventional agriculture often has a negative impact on the environment through decreasing biodiversity, pollution and eutrophication of water, and degrading soil quality (Foley et al., 2011; Tilman et al., 2011; Godfray and Garnett, 2014). Thus, there is the challenge to simultaneously enhance global food security and to reduce the environmental impact of agriculture.

Organic farming and conservation agriculture are two alternatives to conventional agriculture and are often promoted as more environmentally friendly practices (Mäder et al., 2002; Hobbs et al., 2008; Tuck et al., 2014). Organic agriculture is defined as having no synthetic inputs (no synthetic pesticides and no mineral fertilisers) (Reganold and Wachter, 2016; Seufert and Ramankutty, 2017), and a range of studies show that organic farming enhances biodiversity and has reduced environmental impact (Mäder et al., 2002; Tuck et al., 2014; Bender et al., 2016). Conservation agriculture represents a set of three crop management principles: (A) direct planting of crops with minimum soil disturbance (that is, reduced or no-tillage), (B) permanent soil cover by crop residues or cover crops, and (C) crop rotation (Pittelkow et al., 2015). Several studies indicate that conservation agriculture has a positive effect on soil quality and a range of soil biota (Hobbs et al., 2008; Köhl et al., 2014; Briones and Schmidt, 2017).

So far, studies comparing organic or conservation agriculture with conventional agriculture have tested whether organic agriculture or conservation agriculture differ in yield, biodiversity or environmental services compared to conventional agriculture. However, an important issue that is relevant for the discussion on food security is that of yield stability (i.e. the variability of yield across years). So far, it has not been tested whether yield stability in organic and conservation agriculture differs from that in conventional agriculture.

The concept of yield stability was originally developed in plant breeding (Lin et al., 1986; Becker and Léon, 1988), but in recent years it has also received increased interest from ecologists, especially in relation to the stability of ecosystem functioning (Tilman et al., 2006; Hautier et al., 2015). Yield stability can be measured in various ways (Lin et al., 1986). One way to measure temporal yield variability is the standard deviation of yield across years. We refer to this as the absolute stability. However, this measure does not account for the differences in yield. Hence, various investigators have calculated the coefficient of variation, which divides the variability across years (expressed as standard deviation) by the mean yield over the same period (Tilman et al., 2006; Hautier et al., 2015; Raseduzzaman and Jensen, 2017). In order to distinguish from absolute stability, we refer to this as relative stability. Different from absolute stability, relative stability is scaled per unit yield produced. This means that both the variability across years and the mean yield level influence relative yield stability (e.g., a treatment with reduced yield but equal absolute stability (standard deviation) has a reduced relative yield stability (greater coefficient of variation) because the amount of variation per unit yield is higher. Many factors can cause yield of crop species to vary across years, including differences in precipitation, temperature, pest outbreaks, weed pressure, soil fertility, soil structure and agricultural management (Seufert and Ramankutty, 2017).

Two recent meta-analyses compared the yield of conventional agriculture with organic agriculture and conservation agriculture. A study by Ponisio et al. (2015), building upon Seufert et al. (2012) and de Ponti et al. (2012), compared 1071 paired yield observations of 115 studies and showed that organically managed fields have on average 19.2% less yield compared to conventionally managed fields. It was further observed that the yield gap between organic and conventional agriculture depends on crop species, and it was lower when both systems used crop rotations or received similar amounts of fertiliser. Another recent meta-analysis by Pittelkow et al. (2015) compared no-tillage, the original and central concept of conservation agriculture, with conventional tillage and observed that no-tillage on average reduced yields by 5.7% compared to conventional tillage. The effects were variable, depended on crop species (Pittelkow et al., 2015) and under certain conditions no-tillage produced equivalent or even greater yields than conventional tillage.

We applied a meta-analysis procedure using the datasets by Ponisio et al. (2015) and Pittelkow et al. (2015) and compared temporal yield stability of (A) organic vs. conventional agriculture and, (B) no-tillage vs. conventional tillage. We used 191 studies (39 studies from Ponisio et al. (2015) and 154 studies from Pittelkow et al. (2015)) resulting in 532 multiple year observations that were based on 2896 comparisons. We demonstrate that relative yield stability of organic agriculture, assessed per unit yield produced, is significantly lower compared to conventional agriculture. Moreover, absolute stability (i.e. the temporal variation in plant yield without correcting for yield level) did not differ between organic and conventional agriculture. Our analysis further indicates that enhanced fertilisation and the application of green manure can help to reduce the yield stability gap with conventional agriculture and reduce relative yield stability in organic agriculture. We further show that no-tillage and conventionally tilled systems have similar yield stability, especially in dry climates and on fields with residue retention and crop rotation.

2.3 Methods

Data generation

We used two datasets: (1) a dataset on organic farming by Ponisio et al. (2015) comparing the yields of organic and conventional farming and (2) a dataset on no-tillage by Pittelkow et al. (2015) comparing the yields of no-tillage and conventional tillage. Both datasets were generated for meta-analysis studies, comprising data from published experiments, and were published as supplemental material. Only field experiments containing side-by-side yield comparisons were included in the database to ensure comparability of the cropping system treatments. Because the focus of this study was on temporal yield stability, i.e. annual variability across years, single year comparisons from the original datasets were combined in order to create observations that were based on several years for each crop investigated (i.e. multiple year observations (MYO)). We focused on studies with a minimum of 4 years of observation for the same crop, thus excluding short-term studies.

Dataset on organic agriculture

The original dataset from Ponisio et al. (2015) was modified in order to calculate temporal yield stability across years. In order to do this we performed the following steps: First, we corrected a number of minor errors in the original dataset (Supplementary Table 3). Second, we removed all comparisons where the years of observations were not the same for organic and conventional farming. Third, in order to calculate the standard deviation across years, multiple year observations (MYO) had to be compiled: Comparisons from the same experiment that originated from single years were combined into MYOs (for examples see Supplementary Table 4, Supplementary Figure 7), and comparisons where the collected error term was the variance across years were used as they were. Fourth, MYOs that contained more than one observation from the same year were removed. Fifth, comparisons based on units which could not be transformed to tonnes ha^{-1} (i.e. lb $plant^{-1}$, boxes ha^{-1}, kg $plant^{-1}$, bales ha^{-1}, trays ha^{-1}, bales ha^{-1}, kg (square centimeter of limb cross-sectional $area)^{-1}$, ka ha^{-1}, g, kg $tree^{-1}$, kg Fw $plant^{-1}$) were removed (this affected a total of 19 comparisons). Sixth, in order to have a robust estimate of the temporal yield stability, we required a minimum of 4 years of observation for each MYO and thus all comparisons based on <4 observations were removed. Finally, when investigating the standardized residuals one extreme outlier was detected and was removed to achieve normal distribution of residuals (Supplementary Figure 8).

After these steps, the final dataset on organic agriculture contained 165 multiple year observations from 39 studies that were based on 443 comparisons from the original dataset.

<u>Dataset on no-tillage</u>

The original dataset from Pittelkow et al. (2015) was processed in the same way as mentioned above for the dataset of Ponisio et al. (2015) with the following additions: Comparisons containing a zero (i.e. no yield data available) for either no-tillage or conventional tillage were removed. As we used the duration value as year of observation, all comparisons containing NA in the “study duration” column were removed. Similarly to the approach for the dataset by Ponisio et al. (2015), MYOs were created by combining comparisons of different years from the same experiments. However, when creating the MYOs, we observed that for some comparisons the number of rows per MYO was greater than the duration length (see column subtreatments in Supplementary Table 5). After we compared these observations with the original publications, it was clear that these MYOs were derived from different subtreatments. Single observations used for creating MYOs were thus either collected in subsequent or alternating order (see column order in Supplementary Table 5). In order to separate subtreatments within MYOs the number of rows needed to be in agreement with the duration of the study, and all MYOs that did not fulfill these criteria were removed. The remaining MYOs containing subtreatments were then further split into separate MYOs (column MYO in Supplementary Table 5). Similarly to the dataset on organic agriculture, standardized residuals were investigated and two extreme outliers were removed (Supplementary Figure 9).

In the end, the final dataset on conservation tillage contained 367 multiple year observations from 154 studies that were based on 2453 comparisons from the original dataset.

Statistical analysis

After the creation of the multiple year observations, for each MYO the mean yield (X), standard deviation (SD) and number of years of observation (N) was available for the experimental (e) and the control (c) treatment. In the dataset on organic farming, the organic treatment was used as the experimental treatment, and the conventional treatment was used as the control treatment. In the dataset on no-tillage, the no-tillage treatment was used as the experimental treatment, and the conventional tillage treatment was used as the control treatment.

In order to determine the overall difference in mean yield we used the log response ratio (expressed as mean yield ratio) as effect size, which is the natural log of the ratio of the mean yield of both cropping systems (Gurevitch and Hedges, 2001). The log-transformation has the property to produce normally distributed data (Hedges et al., 1999). Following Nakagawa et al (2015) we used the two following measures to asses temporal stability: (1) the “absolute stability ratio”, which is based on the standard deviation of both treatments as an indicator for variability, and (2) the “relative stability ratio”, which is based on the coefficient of variation (CV: standard deviation across years divided by the mean across those years) of both treatments as indicator for variability. Therefore, in the latter measure the variability is standardized per unit yield (i.e. the variability relative to the yield level).

For each of the three measures, the ratio was calculated by dividing the respective response of the experimental treatment (organic farming or no-tillage) by the respective response of the control treatment (conventional farming or tillage, respectively). A ratio greater than one indicates greater yield or greater variability (i.e. reduced stability), respectively, for the experimental treatment. The equations for the respective responses were:

$$\ln(\text{mean yield ratio}) = \ln\left(\frac{X_\mathrm{e}}{X_\mathrm{c}}\right), \quad (1)$$

$$\ln(\text{absolute stability ratio}) = \ln\left(\frac{\mathrm{SD_e}}{\mathrm{SD_c}}\right) + \frac{1}{2(N_\mathrm{e}-1)} - \frac{1}{2(N_\mathrm{c}-1)},$$

which simplifies with $N_\mathrm{e} = N_\mathrm{c}$ to

$$\ln(\text{absolute stability ratio}) = \ln\left(\frac{\mathrm{SD_e}}{\mathrm{SD_c}}\right), \quad (2)$$

$$\ln(\text{relative stability ratio}) = \ln\left(\frac{\mathrm{CV_e}}{\mathrm{CV_c}}\right) + \frac{1}{2(N_\mathrm{e}-1)} - \frac{1}{2(N_\mathrm{c}-1)},$$

which again simplifies to

$$\ln(\text{relative stability ratio}) = \ln\left(\frac{\mathrm{CV_e}}{\mathrm{CV_c}}\right), \quad (3)$$

with $\mathrm{CV_e} = \left(\frac{\mathrm{SD_e}}{X_\mathrm{e}}\right)$ and $\mathrm{CV_c} = \left(\frac{\mathrm{SD_c}}{X_\mathrm{c}}\right)$.

In order to account for the sampling uncertainty in each observation we used the sampling variances as proposed in Nakagawa et al. (2015). Through the inclusion of the sampling variance, observations with better sampling quality (lower sampling variance) receive a greater weight in the analysis. Following Nakagawa et al. (2015) the equations for the sampling variances for three different response ratios were as follows:

$$\mathrm{var}(\ln(\text{mean yield ratio})) = \frac{\mathrm{SD_e^2}}{N_\mathrm{e}X_\mathrm{e}^2} + \frac{\mathrm{SD_c^2}}{N_\mathrm{c}X_\mathrm{c}^2}, \quad (4)$$

$$\mathrm{var}(\ln(\text{absolute stability ratio})) = \frac{1}{2(N_\mathrm{e}-1)} + \frac{1}{2(N_\mathrm{c}-1)}, \quad (5)$$

$$\mathrm{var}(\ln(\text{relative stability ratio})) = \frac{\mathrm{SD_e^2}}{N_\mathrm{e}X_\mathrm{e}^2} + \frac{1}{2(N_\mathrm{e}-1)} + \frac{\mathrm{SD_c^2}}{N_\mathrm{c}X_\mathrm{c}^2} + \frac{1}{2(N_\mathrm{c}-1)}. \quad (6)$$

Note that we modified the equation for the sampling variance of the relative stability ratio, because for normally distributed data the variance and the mean are not correlated. Calculations were performed as implemented in the *metaphor* package.

As some observations shared common control or experimental treatments, we employed a variance-covariance (VC) matrix to correct for correlations among observations following Lajeunesse (2011). When multiple treatments share a common control or common experimental treatment, the assumption of independence is violated. Thus, the effects should be aggregated by using an appropriate variance-covariance matrix. When all observations are independent, this variance-covariance matrix only holds the variance on the diagonal. Following Lajeunesse (2011), for two experimental treatments A and B, which have both been compared to the same control treatment C, the variance-covariance matrix holds the variance of the comparisons of A to C (resp. B to C) on the diagonal and the variance of the log of the mean of the control treatment ($\mathrm{var}\left(\ln(X_\mathrm{c})\right) = \frac{\mathrm{SD}_\mathrm{c}^2}{N_\mathrm{c} X_\mathrm{c}^2}$) on the off-diagonal:

$$\mathrm{VC}(\ln\,(\text{mean yield ratio})) = \begin{bmatrix} \frac{\mathrm{SD}_\mathrm{c}^2}{N_\mathrm{c}X_\mathrm{c}^2} + \frac{\mathrm{SD}_\mathrm{A}^2}{N_\mathrm{A}X_\mathrm{A}^2} & \frac{\mathrm{SD}_\mathrm{c}^2}{\mathrm{N}_\mathrm{c}\mathrm{X}_\mathrm{c}^2} \\ \frac{\mathrm{SD}_\mathrm{c}^2}{N_\mathrm{c}X_\mathrm{c}^2} & \frac{\mathrm{SD}_\mathrm{c}^2}{N_\mathrm{c}X_\mathrm{c}^2} + \frac{\mathrm{SD}_\mathrm{B}^2}{N_\mathrm{B}X_\mathrm{B}^2} \end{bmatrix} \tag{7}$$

For the responses absolute stability ratio and relative stability ratio, the respective sampling variance-covariance matrices are then:

$$\mathrm{VC}(\ln\,(\text{absolute stability ratio})) = \begin{bmatrix} \frac{1}{2(N_\mathrm{c}-1)} + \frac{1}{2(N_\mathrm{A}-1)} & \frac{1}{2(N_\mathrm{c}-1)} \\ \frac{1}{2(N_\mathrm{c}-1)} & \frac{1}{2(N_\mathrm{c}-1)} + \frac{1}{2(N_\mathrm{B}-1)} \end{bmatrix} \tag{8}$$

$$\mathrm{VC}(\ln\,(\text{relative stability ratio})) = \begin{bmatrix} \frac{\mathrm{SD}_\mathrm{c}^2}{N_\mathrm{c}X_\mathrm{c}^2} + \frac{1}{2(N_\mathrm{c}-1)} + \frac{\mathrm{SD}_\mathrm{A}^2}{N_\mathrm{A}X_\mathrm{A}^2} + \frac{1}{2(N_\mathrm{A}-1)} & \frac{\mathrm{SD}_\mathrm{c}^2}{N_\mathrm{c}X_\mathrm{c}^2} + \frac{1}{2(N_\mathrm{c}-1)} \\ \frac{\mathrm{SD}_\mathrm{c}^2}{N_\mathrm{c}X_\mathrm{c}^2} + \frac{1}{2(N_\mathrm{c}-1)} & \frac{\mathrm{SD}_\mathrm{c}^2}{N_\mathrm{c}X_\mathrm{c}^2} + \frac{1}{2(N_\mathrm{c}-1)} + \frac{\mathrm{SD}_\mathrm{B}^2}{N_\mathrm{B}X_\mathrm{B}^2} + \frac{1}{2(N_\mathrm{B}-1)} \end{bmatrix} \tag{9}$$

For the generation of the variance-covariance matrix we used a modified version of the *covariance_commonControl()* function from the *metagear* package (Lajeunesse, 2016). When testing the effect of moderators (see below), the structure of common control or experimental treatments changed because observations within studies were derived from different levels of the moderator variable. Therefore, a new variance-covariance matrix was created for each moderator.

We employed a mixed model approach using the *rma.mv()* function from the *metafor* package in R (Viechtbauer, 2010) with REML estimation. To account for variation between studies, a random effect for study was included, and the respective sampling variances (as described above) were included. To estimate the overall effect, a mixed model containing only a fixed intercept and the random study effect was run.

Both datasets contained additional explanatory variables (e.g. crop species or information on management practices such as fertilisation level or the use of green manure). These explanatory variables (moderators) were tested with a separate model for each variable, in which the variable was included as a categorical, fixed effect variable. In order to get the average estimates of the factor levels of the moderator variables, a model was fitted without the intercept. For both, the overall effect and average estimates of the factor levels, 95% confidence intervals, as provided by the rma.mv() function for the coefficients, were used to test the significant difference from 1. All calculations were done with the R statistical package (R Core Team, 2018).

2.4 Results

2.4.1 Yield stability of organic and conventional agriculture

We used the dataset of Ponisio et al. (2015) to compare temporal yield stability of organic and conventional agriculture. Our analysis demonstrates that the relative yield stability (i.e. yield stability per unit yield produced) in conventionally managed fields was, averaged across all crops, 15% [2% to 30%] higher compared to organically managed fields, and this difference was significant (Figure 1). A closer look at the data further confirmed this, and out of the 165 multiple year comparisons (observations) in the dataset, 79% (131 observations) had higher relative stability in conventionally managed fields (Figure 2). We observed no significant difference in absolute stability between organic and conventional agriculture (Figure 1) demonstrating that the overall temporal variability in yield, independent of yield level, was similar between organic and conventional agriculture.

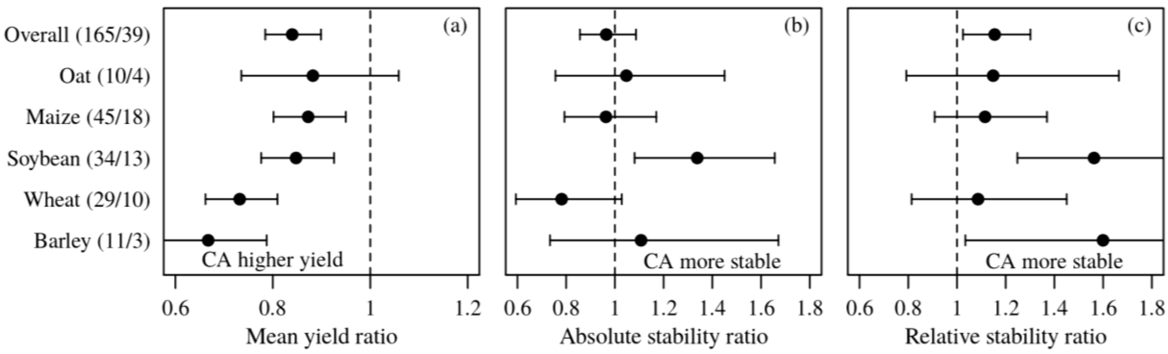


Figure 1: Yield and yield stability comparing organic and conventional agriculture. Mean yield ratio (a), absolute stability ratio (b), and relative stability ratio (c) for organic (OA) versus conventional (CA) agriculture for all crops (Overall) and for crops, for which at least 10 observations were available. Numbers in parentheses denote the number of observations and studies. A ratio of 1 means that there is no difference between organic and conventional managed systems while values <1 indicate higher yield for conventional agriculture. For both stability measures a ratio >1 indicate greater absolute and relative stability for conventional agriculture. Values are mean effect sizes with 95% confidential intervals. Mean yield or stability were deemed significantly different between organic and conventional agriculture if the 95% confidential intervals of the ratios did not overlap one.

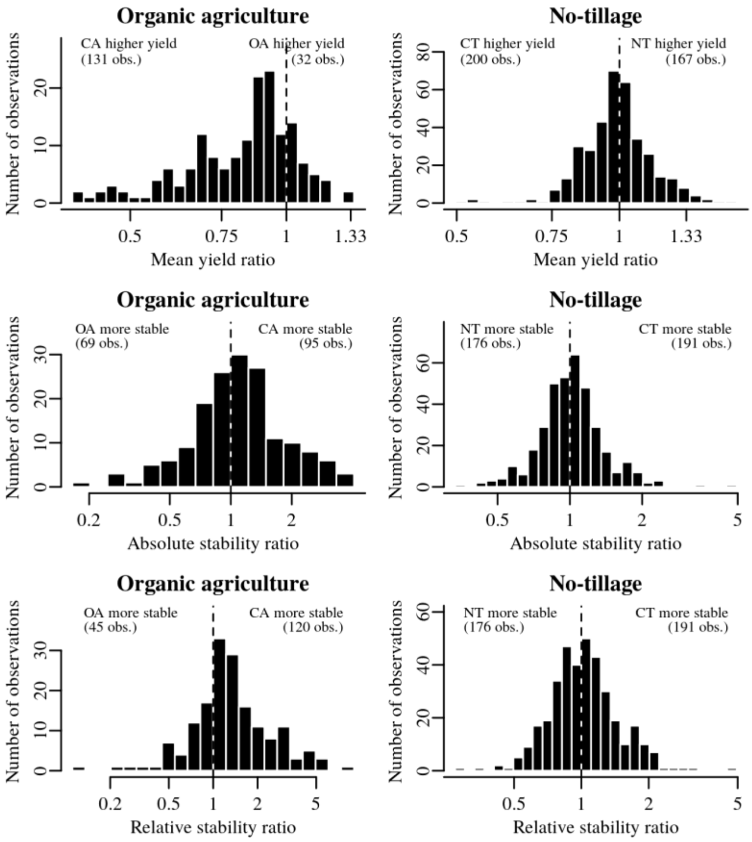


Figure 2: Histograms of the yield and yield stability ratios. Mean yield ratios (top row), absolute stability ratios (middle row), and relative stability ratios (bottom row), for the dataset comparing organic (OA) and conventional agriculture (CA) (left column), and the dataset comparing no-tillage (NT) and conventional tillage (CT) (right column), respectively. The ratios on the x-axis are on the ln scale.

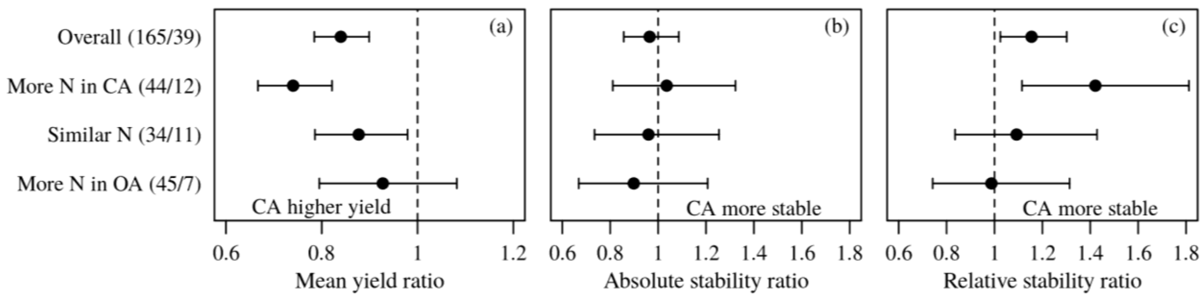


Figure 3: Effect of nitrogen input on yield and yield stability comparing organic and conventional agriculture. Mean yield ratio (a), absolute stability ratio (b), and relative stability ratio (c) for organic (OA) versus conventional (CA) agriculture for different levels of nitrogen input. Numbers in parentheses denote the number of observations and studies. A ratio of 1 means that there is no difference between organic and conventional managed systems while values <1 indicate higher yield for conventional agriculture. For both stability measures a ratio >1 indicate greater absolute and relative stability for conventional agriculture. Values are mean effect sizes with 95% confidential intervals. Mean yield or stability were deemed significantly different between organic and conventional agriculture if the 95% confidential intervals of the ratios did not overlap one.

We observed a significantly increased relative yield stability under conventional management for two (soybean and barley) out of five crop species for which enough data (>10 comparisons) were available (Figure 1). Interestingly, a significant difference in absolute stability (i.e. not corrected for yield level) was only observed for soybean. The absolute stability of soybean was higher in conventionally managed fields compared to organically managed fields. Results for many other crop species were highly variable (Supplementary Table 1) and should be interpreted carefully because few data (often only one or two comparisons) were available.

We evaluated the effects of other factors on yield stability, and our analysis indicated that the increased relative and absolute yield stability of conventional management was related to differences in N-fertilisation (Figure 3). If organically and conventionally managed fields received similar amounts of nitrogen fertiliser, relative yield stability did not vary significantly between both management systems; although it was still lower (9%) in organically managed systems. However, if organically managed fields received less nitrogen, the relative yield stability was much lower (42% [-11% to -81%]) compared to conventionally managed fields. This indicates that the increased relative stability of conventionally agriculture is, in part, due to higher fertilisation levels and related to the higher yield. Still, even with equal amounts of nitrogen fertilisation, organic agriculture had a significantly lower yield (12% [-2% to -21%]); although this difference was less than for the overall dataset where it was 16% [-10% to -22%] (Figure 3). Interestingly, our analysis also indicates that the level of P fertilisation influenced, in a similar way to N, differences in yield and yield stability between organic and conventional agriculture (Supplementary Figure 1). Our analysis further indicated that the addition of green manure had a positive impact on yield and the relative yield stability of organic agriculture (Supplementary Figure 2).

2.4.2 Yield stability of conservation and conventional agriculture

We used the data set of Pittelkow et al. (2015) to compare temporal yield stability of conservation agriculture (focusing on no-tillage) and conventional agriculture. Our analysis indicated that both absolute and relative yield stability did not differ between no-tilled and conventionally tilled fields for the overall data-set and for crop species with at least 10 observations (Figure 4, see Supplementary Table 2 for all species contained in the dataset).

We then tested whether the application of crop rotation and residue management, two of the main conservation agriculture principles, influenced yield stability. The application of crop rotation and residue management in no-tillage had, compared to conventional tillage, no effect on absolute and relative stability (Figure 5). However, without crop rotation and residue management, no-tillage had a 23% [-1% to -50%] reduced relative stability compared to conventional tillage. This result has to be interpreted carefully, as the group where none of the principles of conservation agriculture was followed, was only based on 15 observations (11 studies).

We further tested whether effects of no-tillage and conventional tillage on yield stability depended on climate conditions, comparing dry and humid climate. There was no difference in absolute stability between dry and humid climate and also no difference in relative stability in dry conditions. In contrast, in humid climates, conventionally tilled fields had higher relative yield stability (Supplementary Figure 3).

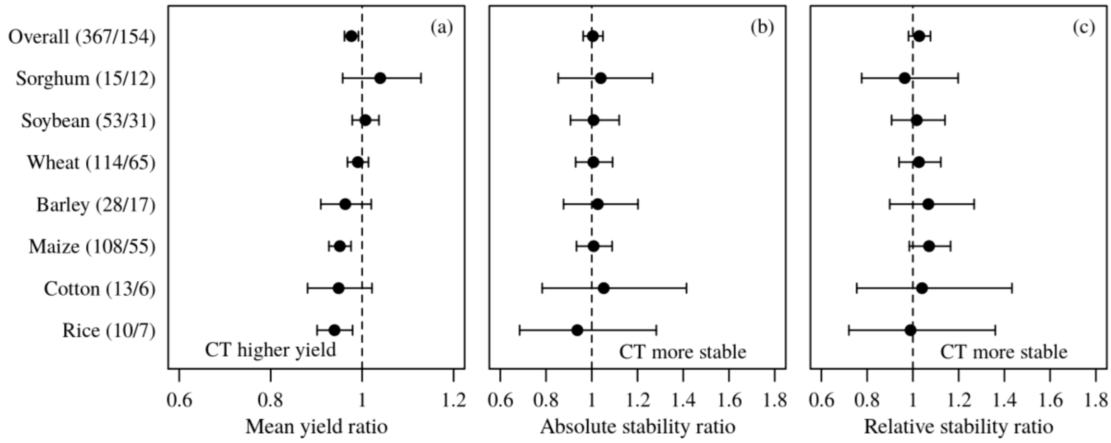


Figure 4: Yield and yield stability comparing no tillage and conventional tillage. Mean yield ratio (a), absolute stability ratio (b), and relative stability ratio (c) of no-tillage (NT) versus conventional tillage (CT) for all crops (Overall) and for crops, for which at least 10 observations were available. Numbers in parentheses denote the number of observations and studies. A ratio of 1 means that there is no difference between no-tillage and conventional tillage while a value <1 indicates higher yield for conventional tillage. For both stability measures ratios >1 indicate greater stability for conventional tillage. Values are mean effect sizes with 95% confidential intervals. Mean yield or stability were deemed significantly different between no-tillage and conventional tillage if the 95% confidential intervals of the ratios did not overlap one.

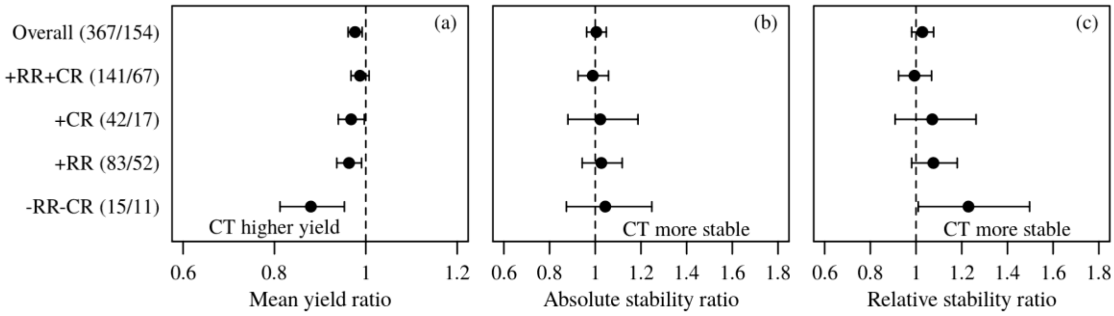


Figure 5: Effect of crop rotation and residue retention on yield and yield stability comparing no-tillage and conventional tillage. Mean yield ratio (a), absolute stability ratio (b), and relative stability ratio (c) of no-tillage (NT) versus conventional tillage (CT) for subcategories of observations regarding residue retention (RR) and crop rotation (CR): +RR+CR (residue retention and crop rotation), +RR (only residue retention), +CR (only crop rotation), or –RR–CR (without residue retention or crop rotation). Numbers in parentheses denote the number of observations and studies A ratio of 1 means that there is no difference between no-tillage and conventional tillage while values <1 indicate higher yield for conventional tillage. For both stability measures values >1 indicate greater stability for conventional tillage. Values are mean effect sizes with 95% confidential intervals. Mean yield or stability were deemed significantly different between no-tillage and conventional tillage if the 95% confidential intervals of the ratios did not overlap one.

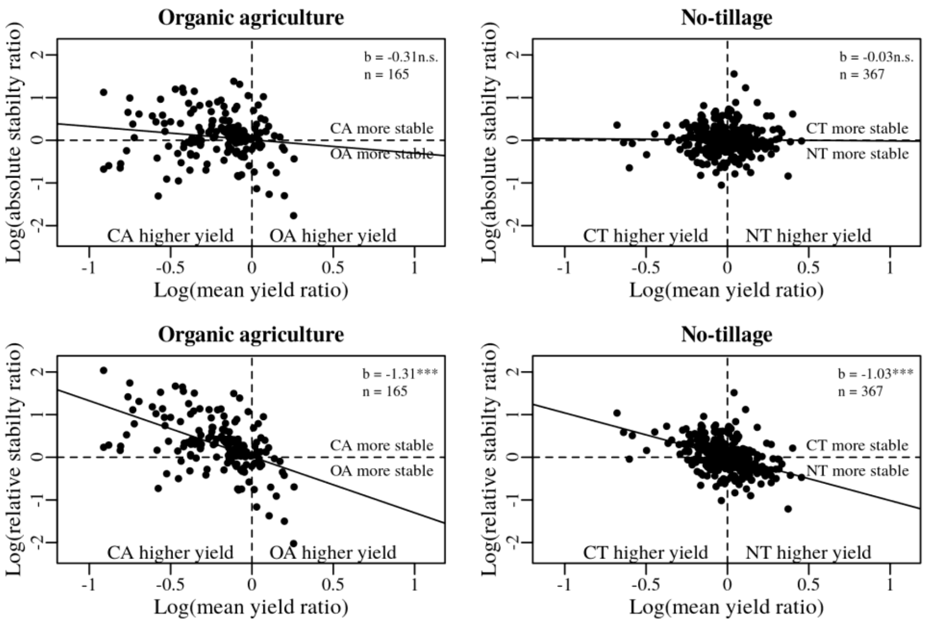


Figure 6: Relationship between the mean yield ratio and the absolute stability ratios. Relationship of the mean yield ratio to the absolute stability ratio (top row), relationship of the mean yield ratio tor the relative stability ratio (bottom row) for the dataset comparing organic (OA) and conventional agriculture (CA) (left column) and the dataset comparing no-tillage (NT) and conventional tillage (CT) (right column), respectively. Each dot represents one multiple year observation (MYO) and ratios are on the natural log scale. The regression line was fitted on log-transformed values, i.e. $\log(y) = a + b * \log(x)$, where y was the respective stability ratio and x was the mean yield ratio. *** denote significance at $P<0.001$ for a t-test with H0: $b=0$, and n.s. denotes non-significant ($P>0.05$).

2.5 Discussion

Our work adds a new perspective to earlier meta-analyses (Pittelkow et al., 2015; Ponisio et al., 2015) and reveals the effects of different cropping systems on the variability and reliability of food production across years (e.g. temporal yield stability). Our analysis demonstrated that conventional agriculture has, on average, and per unit food produced, a higher relative yield stability compared to organic agriculture. Yield stability depended on crop species and nutrient management. Notably, the absolute stability of crop yield was the same in organic and conventional agriculture. However, relative stability, which is the temporal variation per unit yield produced, was significantly higher under organic agriculture due to reduced yields in organic agriculture. Thus, per unit food produced, there is higher temporal variation in yield in organic agriculture.

Enhanced fertilisation and the application of green manure were identified as tools to reduce the yield stability gap of organic agriculture with conventional agriculture (Figure 3; Supplementary Figure 1 and Supplementary Figure 2). The observation that fertilisation enhances yield stability is in agreement with Deguines et al. (2014) observing that relative yield stability increased with increasing land use intensity. Further experiments need to test whether enhanced fertilisation can reduce the yield gap and enhance yield stability under organic farming. Recommendations for enhanced fertilisation would rely on the assumption that sufficient organic fertilisers are available (e.g. see Muller et al. (2017) , but see Connor (2018)). Moreover, it is important to note that increased fertilisation may raise additional environmental concerns, including the loss of nutrients through leaching and subsequently, enhanced levels of nitrate in drinking water or enhanced production of the greenhouse gas N_2O (Galloway et al., 2003). The positive effect of green manure on yield stability is in agreement with a recent study that showed that green manure (e.g. cover crops) are especially suitable to enhance yields in less intensive cropping systems such as organic agriculture (Wittwer et al., 2017).

The reasons for reduced relative yield stability under organic farming can be manifold and include, beside fertilisation level, enhanced disease pressure (and fewer opportunities to rapidly control pests with pesticides). Also, the timing of fertilisation influences plant yield, and appropriate timing is more difficult with organic fertilisers because nutrient release is delayed compared to readily available mineral fertilisers in conventional agriculture. Moreover, past and current breeding programs have largely focused on high-yielding varieties adapted to work well with conventional inputs (Ponisio et al., 2015) and there has been little selection for traits being important in organic agriculture (e.g. increased disease resistance, enhanced cooperation with plant symbionts, better weed suppressing abilities or higher resistance and competitive ability against weeds).

Compared to conventional agriculture, organic agriculture generally has a positive effect on a range of environmental factors, including above and belowground biodiversity (Birkhofer et al., 2008; Verbruggen et al., 2010; Tuck et al., 2014; Lichtenberg et al., 2017), soil carbon stocks (Gattinger et al., 2012) and soil quality (Seufert and Ramankutty, 2017). Moreover, organic farming can reduce soil erosion (Reganold et al., 1987) and has a reduced global warming potential (Prechsl et al., 2017). However, higher productivity and increased relative stability in conventional agriculture are strengths compared to organic agriculture. Thus, in order to benefit from the strengths of organic farming (e.g. reduced environmental impact and enhanced biodiversity) a multi-faceted strategy is necessary to improve its yield and relative yield stability. Such a strategy should focus on enhanced plant nutrition (see above), breeding, weed and disease control, and consider the use of state of the art technologies including precision farming, remote sensing (e.g. through drones or satellites) to detect disease or nutrient deficiency, and robotics (e.g. for weed control) (Niggli et al., 2016). Moreover, measures such as the inclusion of cover crops (see above) or active stimulation of soil life through soil ecological engineering are especially promising for lower intensity systems such as organic agriculture, and this can further help to reduce the yield gap and the yield stability gap between organic and conventional systems (Bender et al., 2016; Wittwer et al., 2017). Further studies also need to assess how environmental stresses, such as drought or the negative effects of climate change, influence yield stability in organic and conventional production systems. Finally, when comparing organic and conventional agriculture, it is important to provide an “output and input footprint” and assess the overall impact of organic and conventional farming practices including yield, yield stability, energy use, pesticide use, fertiliser use, and overall environmental performance.

Absolute and relative yield stability on average did not vary between no-tillage and conventional tillage indicating that a transition to no-tillage generally does not affect yield stability. Interestingly however, yield and yield stability were affected by climate, and no-tillage systems in humid climate had a reduced yield and yield stability compared to dry climate. These differences are probably due to better soil water retention and slightly higher yields of no-tilled soils in dry climate versus the negative effects of delayed soil warming, nutrient mineralization and reduced soil aeration in no-tilled, wet and heavy soils (Hobbs et al., 2008; Martínez et al., 2016). Note that selection and breeding of crops varieties for conservation agriculture is not yet widespread (Newton et al., 2012). Hence, further breeding efforts may enhance yield and yield stability in conservation agriculture.

In our analysis, we employed two different stability measures: absolute stability (measured by the standard deviation in yield across the investigated years) and relative stability, which corrects for yield (measured by the coefficient of variation). While there was a significant difference for relative stability between organic and conventional agriculture, there was no significant difference for absolute stability (Figure 1). This was also indicated by the negative relationship between the mean yield ratio and relative stability ratio (meaning that relative yield stability increased with increasing yield) (Figure 6). Hence, the reduced relative stability in organic agriculture is most likely related to reduced mean yield. The absence of a correlation between the absolute stability ratio and the mean yield ratio in the dataset suggests that absolute stability is less affected by yield level. A similar negative relationship between the coefficient of variation and mean yield has been shown previously by Döring et al. (2015). They associated this with Taylor's power law (Taylor, 1961), which predicts that the natural logarithm of the variance is proportional to the natural logarithm of the mean. This can lead to a spurious negative relationship of the coefficient of variation and the mean. We therefore investigated the relationship between both stability measures and mean yield, and found that in both data-sets absolute stability is not related to the mean yield and relative stability is inversely related to the mean yield (see Supplementary Note 1, Supplementary Figure 4, and Supplementary Figure 5). The coefficient of variation has been used extensively to quantify stability (Lin et al., 1986; Smith and Gross, 2006; Tilman et al., 2006; Schrama et al., 2018), but its relationship to the mean yield has rarely been investigated (Döring et al., 2015). In light of this, we stress the importance– also for future studies – of distinguishing between relative and absolute stability and, in particular, comparing the relationship to the mean when interpreting results.

The estimated yield gap between organic and conventional agriculture in this study (16%) was slightly smaller than the 19% estimated by Ponisio et al. (2015). This is because we only used 41% of the observations (and 34% of the studies). In our analysis, we only included comparisons with a minimum of 4 years of observation per crop (see Methods) explaining this lower number. This approach was necessary in order to be able to calculate the year-to-year temporal variation, which is necessary for a robust assessment of yield stability. Similarly, Pittelkow et al. (2015) demonstrated that, on average, no-tillage reduced yield by 5.7% compared to conventionally tilled fields, while we only observed a difference of 2% [-1% to -4%] (Figure 4) using 45% of the observations (and 25% of the studies) used in the original –analysis. The advantage of our approach is that short-term studies are removed. This reduces the effect of extreme outlier years and generally provides a more robust analysis of differences between these cropping systems. Moreover, this approach also reduces potential transition effects of previous management (e.g.

plant yield levels of organic fields that had previously been managed conventionally might be higher because such fields generally still contain enhanced nutrient levels).

For our meta-analysis, we used a different model approach compared to Pittelkow et al. (2015) and this may further explain some of the observed differences with that study. Pittelkow et al. (2015) applied a weighted mean calculation with bootstrapping, which does not account for the nested structure in the dataset, and leads to non-independence of observations. We corrected for the nested structure of observations derived from the same study by adding a random study effect and by combining observations of several years into multiple year observations. Note, that the datasets used for this study are still based on relatively short-term experiments, i.e. observations with a duration of 4 or 5 years represent 60% of all observations in the dataset for no-tillage and 39% for organic agriculture (Supplementary Figure 6), pointing to the need for long term experiments.

It is important to mention that our meta-analysis uses data from diverse systems, geographic areas, and crop species. For instance, the reduced relative yield stability of organically managed fields provides an average response. Studies that aim to enhance yield stability or reduce the yield gap for organic agriculture should evaluate those experiments and conditions where yield or yield stability are higher (or not lower) under organic agriculture (Schrama et al., 2018) and investigate the causes (e.g. soil type, field management, land use intensity, crop varieties, etc.). Similarly, it is important to investigate under which conditions no-tillage has the most beneficial effects on yield and yield stability.

Our analysis is based on field-scale measurements, and it did not assess yield stability at the farm scale (with a range of crops planted in different fields) or at a regional, national or global scale. To enhance the overall farm level yield stability, farmers could cultivate different crops in different fields (e.g. this reduces the impact of poorly performing crop species at one particular field). Another important strategy to achieve increased yield stability is to grow mixtures of crop species or mixtures of genotypes to exploit positive interaction effects and thus reduce the risk of crop failure (Brooker et al., 2015; Litrico and Violle, 2015). Further modelling and work at different scales (e.g. farm, regional, national and global) is necessary to understand how farmers and policy makers can enhance the stability of the food supply. For instance, farm specialisation and the growing of a few crops may lead to increased regional synchrony, increasing the risk of regional crop failures because of climate or pest/disease outbreaks (examples are e.g. wheat yield losses in Australia (2006) or Russia (2010)). Beside temporal yield stability, there are other measures to evaluate management systems, such as, the resilience of different farming practices to disturbance or climate change or the ability of a particular system to produce enough food or income.

Overall, this work provides further information about the performance of organic and conservation agriculture. The assessment of yield stability and the resilience of cropping systems to environmental variability should receive increased attention because reliable agricultural production is a key issue in light of the growing world population and enhanced demands for food. Moreover, climate change and the predicted increase of extreme weather events will provide additional challenges for stable food production.

Acknowledgments

We would like to thank Fabian Scheipl for mathematical support and Raphael Wittwer, Kyle Hartman and Thomas Döring, for constructive and helpful discussion and comments on the manuscript. Funding was provided by Agroscope and the Swiss National Science Foundation (grant number 166079).

3 Chapter B: Yield, yield development and stability of conventional and organic farming in the DOK long-term experiment

3.1 Abstract

Farming needs to produce sufficient amounts of food in an efficient and sustainable manner and to maintain yield and soil fertility. Here, we compare mean yields, temporal trends, and stability of different organic and conventional farming systems at two fertilization intensities using the currently longest-lasting organic-conventional cropping system comparison, the DOK experiment.

We used yield data of winter wheat, potatoes, grass-clover, maize, and soybean cropped in a seven-year rotation on a silty loam, where bioorganic and biodynamic farming practices have been compared with mixed and sole mineral fertilization conventional practices at regular and half fertilization levels over a period of 35 years from cropping cycle two to six. Mean yields were related to the amount of applied nutrients.

Yields were significantly reduced in the organic systems by between 10% and 34%, dependent on the investigated crop, concomitant to overall 35%, 36%, and 26% lower fertilizer inputs in organic systems for total N, P, and K, respectively, while there was no yield reduction in soybean. This yield reduction was similar under regular and half fertilization and was mainly due to lower nutrient input and the omission of chemical pesticides. Half fertilization led to a yield reduction of around 10% in both conventional and organic systems. In winter wheat and potatoes, mean yields of the treatments were strongly determined by average applied mineral N. Temporal yield trends were not different between organic and conventional systems, and between half and regular fertilization. However, in winter wheat, conventional and biodynamic management at regular fertilization, showed a stronger increase in yield, but in grass-clover, half-fertilized treatments and biodynamic management showed a stronger decrease. There was no difference in absolute stability (measured by the variance) between organic and conventional management. However, conventional management was more stable in relative stability, measured by the coefficient of variation, expressing the stability in relation to the yield level. The difference in relative stability was therefore due to the difference in mean yield. We found no difference in absolute and relative stability between half and regular fertilization. A further analysis revealed that an increased yield

difference between organic and conventional management in single years is mainly due to lower yields of the organic treatments.

The strong effects of applied mineral N on yields in winter wheat and potatoes indicates that there could be a potential to increase yields in organic farming through better N management without increasing the overall applied amount of fertilizer, thus using available N more efficiently. Yield development being overall similar between the systems indicates that organic management did not result in declining yields when compared to conventional management. The similarity in relative stability between fertilization levels suggests that the difference in relative stability between organic and conventional management might be more related to plant protection than to fertilization intensity.

3.2 Introduction

Agricultural production has led to a multitude of negative impacts on the environment like soil degradation, loss of biodiversity, increases in greenhouse gas emission or eutrophication of water (McLaughlin and Mineau, 1995; Johnson et al., 2007; Moss, 2008). With ongoing growth of the world population, agriculture needs to increase its production and reduce its negative impacts (Godfray et al., 2010). Furthermore, as climate change will lead to greater fluctuations and more extreme weather events, agricultural production has to be more resilient against such fluctuations to guarantee future regional and global food security (Howden et al., 2007). The challenge is thus, to maintain or better increase productivity, in a more stable way, and with less environmental impact, i.e. more sustainable (Tilman et al., 2011).

Organic farming has been established with the aim to reduce the negative impacts of agriculture on the environment through the avoidance of mineral fertilizers and synthetic pesticides. Several studies have shown the reduced negative environmental impacts of organic compared to intense conventional systems (Reganold and Wachter, 2016; Smith et al., 2019). However, due to the omission of mineral fertilizers and synthetic pesticides, yields of organic farming are often reduced. When analyzing the yield gap between organic and conventional farming, several studies have found similar estimates of around 20% for the overall yield gap, but all of them noted that the variation in yield gap between crops and regions is substantial (de Ponti et al., 2012; Seufert et al., 2012; Ponisio et al., 2015).

Besides the immediate effects of different management practices on the productivity, there can also be long-term effects on the productivity and the environmental impacts. These long-term effects can be two-sided: While, e.g. too little fertilization could lead to a decrease in soil fertility and the avoidance of pesticides to a build-up in pests over time, over-fertilization and the over-usage of pesticides could result in negative impacts on the environment and human health. In turn, a decrease in soil fertility and build-up of pests could lead to a decrease in productivity. A change in productivity can be assessed by analyzing the yield development through regression on time. Furthermore, the maintenance of productivity over time can be indicative of the sustainability of a certain management practice (Hejcman et al., 2012). As the majority of comparisons of organic and conventional are limited to short-term observations, we could not find any study investigating the long-term trend of productivity.

Originating in plant breeding, stability analysis has gained increased attention in comparing the temporal yield stability of different management systems (Becker and Léon, 1988; Smith et al., 2007; Reckling et al., 2018). The most common measures to compare the stability management systems are the variance (or standard deviation) and the coefficient of variation, which corrects for the difference in yield level between the systems. Knapp and van der Heijden (2018) have introduced the terms absolute stability for the further and relative stability for the latter. A particular focus of stability analysis has been on comparing conventional and organic management practices and several studies have found that conventional management is more stable in relative stability than organic management (Smith et al., 2007, 2019; Knapp and van der Heijden, 2018). The difference in relative stability has been attributed to the difference in mean yield and Knapp and van der Heijden (2018) have argued that increasing N fertilization can thus increase relative stability in organic farming.

The main reason for the different stability of treatments or management systems is that they react differently to yearly growing conditions, including, e.g. water availability and pest pressure. Thus, the ratio of the yields will also vary between years. Based on the variation in yield ratio between years, we propose to use the temporal variation of the yield ratio as an additional analysis of temporal stability by correlating the yield ratio of each year to the yields of the respective treatments to be compared.

Long-term trials with consistent treatments over time offer a valuable source for investigating long-term effects of different management systems on the productivity and impacts on soil and environment. The DOK long-term trial has been established in 1978 in order to compare the farming systems bio-Dynamic, bio-Organic, and “Konventionell” (DOK). However, it was not designed as a static experiment with an orthogonal set of treatments, but rather to dynamically reflect current agricultural practices as conducted in Switzerland (Krause et al., 2020). In this regard, the fertilization intensity of both organic systems is based on the number of livestock per area, while the fertilization intensity is determined by Swiss official regulations. Besides the different farming systems, two levels of fertilization (regular and half fertilization) were established within each system, allowing for an additional assessment of the effect of fertilization intensity. With now 42 years under use, it is the longest-lasting experiment comparing organic and conventional management and provides a unique dataset for the investigation of long-term effects (Mayer and Mäder, 2016).

The objectives of the study were to investigate the effects of long-term organic vs. conventional management and different fertilization intensities on the mean yield, temporal yield development, and temporal yield stability from the 2nd to the 6th seven-year crop rotation. In addition, we investigated whether mean yields were related to the amount of applied nutrients.

3.3 Material and Methods

Description of the trial and assessed parameters

The DOK long-term systems comparison trial is located in Therwil, Switzerland (47° 30.158'N, 7° 32.347'E), 308 m above sea level. Average yearly precipitation is 840 mm and the mean temperature is 10.5°C (climate norm 1991 – 2010). The soil type is a haplic luvisol on deep deposits of alluvial loess. It contains 12% sand, 72% silt, and 16% clay. Eight different treatments corresponding to different management systems and fertilization intensities were compared (see Table 1 for a detailed description). Within the organic system group, a biodynamic (BIODYN) system and a bioorganic (BIOORG) system were included. Both systems represent Swiss mixed farming systems with livestock, characterized by fertilization through farmyard manure and slurry, and mechanical weed control. Within the conventional system group, a mixed farming system combining manure and mineral fertilization (CONFYM) and a system with only mineral fertilization (CONMIN) were included. Both conventional systems received chemical plant protection. The CONMIN treatment was introduced in the second rotation cycle from 1985; in the first cycle, from 1978 to 1984, it was an unfertilized control treatment with the same plant protection scheme as CONFYM2. In general, all systems aimed to represent common Swiss agricultural management systems with 1.4 livestock units (LU) per hectare (1.2 LU in the first two rotation cycles) at regular fertilization. In particular, the CONFYM system represented conventional management according to Swiss integrated production (IP) standards. In the conventional systems (CONFYM, CONMIN) the amount of mineral fertilizers were applied according to Swiss fertilization guidelines (Richner and Sinaj, 2017). In CONFYM, this meant on top of nutrients already applied via 1.4 LU/ha. In addition to treatments with regular fertilization system BIODYN, BIOORG, and CONFYM with half fertilization level were included. Average applied nutrients per treatment are given in Table 2, and average nutrient contents of the applied organic fertilizers can be found in Supplementary Table 7. Lastly, a NOFERT treatment, not receiving any fertilization but with the same plant protection as BIODYN served as control. Soil tillage, sowing and harvesting was conducted in the same way in all treatments. For more detailed information, see Mäder et al. (2002) and Mayer et al. (2015).

The experimental design of the DOK trial can be described as split-strip-plot design with four replicate blocks (see layout in Supplementary Figure 10). The main-plot factor corresponds to three shifted crop rotations, where the crop rotation was shifted by one, respectively four, years. Although rotations were nested within replicates, the different rotations will be called fields here to be in accordance with common long-term experiment terminology. Within main-plots, the horizontal factor is the system, with CONMIN and NOFERT being combined here, and the vertical factor is the fertilization level.

One crop rotation cycle lasted seven years and was the same for all treatments. To mimic common management practices in Switzerland, the crops and crop rotations as well as the management were adapted to current practice over the course of the experiment, while changes occurred always after the completion of one crop rotation cycle (Supplementary Table 6). In the last three cycles the crops remained constant – although changing in order – and in the last cycle the crop rotation was maize, soybean, winter wheat, potato, winter wheat, and two years of grass-clover (see Supplementary Table 6 for all crop rotation cycles). The varieties of the different crops also changed over the duration of the experiment to be in parallel with common practice and to deal with possible breakdowns of resistances. The plot size was 100 m² (20 m x 5 m, Supplementary Figure 10).

Crop yield was determined by harvesting a central strip with a width of 1.5 m and 10 m length, omitting the outer area to avoid any border effects. Yields will be reported here as dry matter weight. Samples of all applied farmyard manure and slurry were analyzed for total N (TotN), mineral N (MinN), phosphorus (P), potassium (K), calcium (Ca), magnesium (Mg), and organic matter (OM) content. Contents of mineral products were used as specified in the product description.

Table 1: Management of DOK treatments

System group		Organic				Conventional			Unfertilized
System		**Biodynamic**		**Bioorganic**		**Mixed**		**Mineral**	
Fertilization level		**Half**	**Regular**	**Half**	**Regular**	**Half**	**Regular**	**Regular**	**None**
Abbreviation		**BIODYN1**	**BIODYN2**	**BIOORG1**	**BIOORG2**	**CONFYM1**	**CONFYM2**	**CONMIN2**	**NOFERT**
Average of yearly applied nutrients (kg/ha)	**Total N**	47	94	47	95	85	169	119	0
	Mineral N	13	26	15	30	56	112	119	0
	P	12	24	12	24	19	37	38	0
	K	89	177	91	183	123	246	242	0
Farmyard manure & slurry		Manure compost, slurry		Rotted manure, slurry		Manure, slurry		None	None
Mineral fertilization		none		Rock powder, patentkali (potassium sulfate with magnesium)		Additional NPK (as recommended)		Only NPK (as recommended)	None
Plant protection	**Weed control**	Mechanical				Mechanical and chemical			Mechanical
	Disease control	None		Copper for potatoes		Chemical, according to thresholds			None
	Pest control	Plant extracts and antagonists				Chemical, according to thresholds			Plant extracts
Additional treatments		Biodynamic preparations		---		Growth regulator			Biodynamic preparations

[1]The CONMIN2treatment was established in 1985 with the start of the second crop rotation cycle. In the first crop rotation cycle from 1978 to 1984 it was an unfertilized control treatment with the same plant protection scheme as CONFYM2.

Table 2: Applied nutrients through fertilizers(kg/ha/year) over the duration of the experiment excluding the first crop rotation cycle and averaged by year. Values in parentheses indicate the amounts, which were applied through manure, slurry, and mineral and other fertilizers combined, respectively. Atmospheric N deposition was around 20 kg N/ha/year (Seitler et al., 2016), and not included in the N inputs.

Treatment	Total nitrogen (TotN)	Mineral nitrogen (MinN)	Phos-phorus (P)	Potassium (K)	Calcium (Ca)	Magne-sium (Mg)	Organic matter (OM)
NOFERT	0 (0,0,0)	0 (0,0,0)	0 (0,0,0)	0 (0,0,0)	0 (0,0,0)	0 (0,0,0)	0 (0,0,0)
BIODYN1	48 (26,21,0)	13 (0,13,0)	12 (9,3,0)	89 (29,61,0)	80 (64,13,4)	14 (9,4,1)	957 (686,271,0)
BIOORG1	48 (25,22,1)	15 (2,13,0)	12 (9,3,1)	92 (35,46,14)	52 (30,10,15)	11 (6,3,3)	1016 (784,224,10)
CONFYM1	86 (26,34,28)	57 (5,26,28)	19 (7,4,9)	124 (26,60,42)	87 (18,13,58)	16 (5,4,7)	1157 (906,294,0)
BIODYN2	95 (52,43,0)	26 (1,25,0)	24 (18,6,0)	179 (58,121,0)	160 (128,26,8)	28 (18,9,2)	1911 (1368,543,0)
BIOORG2	96 (51,44,1)	30 (4,27,0)	24 (17,6,2)	184 (70,92,27)	104 (60,20,31)	22 (12,5,5)	2032 (1568,448,19)
CONFYM2	171 (52,69,55)	113 (9,52,55)	37 (13,7,18)	248 (52,121,83)	144 (36,25,86)	32 (11,7,15)	2314 (1812,587,0)
CONMIN2	121 (0,0,121)	121 (0,0,121)	38 (0,0,38)	246 (0,0,246)	168 (0,0,168)	31 (0,0,31)	0 (0,0,0)

Statistical analysis

In order to remove carry over effects from previous management before start of the experiment until 1977 and because CONMIN2 was started in the second rotation cycle, we removed the first rotation cycle from the dataset. For grass-clover, there were three years of continuous grass-clover in the third rotation cycle, while in all other cycles there were only two years. As the yield of the third year was considerably lower than the first two years, yield observations from the third year were completely removed. Due to the shifted crop rotation between fields, in some years the same crop was grown on two fields in parallel. Thus, we used the combination of field x year as the level of environment in the analysis. Although other crops were grown during the experiment, we only investigated the crops winter wheat, potatoes, grass-clover, maize, and soybean, to have a sufficient number of field x year combinations to produce reliable estimates. To check the quality of the data, we conducted a linear model with the factors *block* and *treatment* within each field x year combination. We assessed normal distribution of the residuals through Kolomogorov-Smirnov (KS) test, the CV of a trial as the square root of the residual error variance divided by the overall mean, and the significance of the treatment effect through ANOVA. KS test was never significant, CV was below 9%, and treatment effect was significant (F-Test, $P<0.05$) for all field x year combinations. Thus, no further observations were removed.

Because in long-term experiments plots are resampled every year, observations could thus be correlated between years (Richter and Kroschewski, 2006). Thus, we compared different models that take into account of this correlation structure (see Supplementary Method 1) and tested, if residuals from an RCBD model (*block* and *treatment* effect) within each year were correlated across years. We found that estimates and statistics did only marginally differ between models taking account of correlation structures and an analysis on simple field x year means (Supplementary Table 10). Residuals were not correlated and correlations did not decrease over time (Supplementary Figure 12), which has also been found by Richter and Kroschewski (2006). We therefore did not consider it necessary to correct for possible autocorrelation in the analysis. As some models did not converge, when analyzing data on the plot level, and because there were no missing data in the dataset, we conducted all analyses on treatment by field x year combination means, i.e. means over replicates.

All statistical analyses were conducted in R (R Core Team, 2014) and mixed models were fit with the R-packages *lme4* (Bates et al., 2015) and *sommer* (Covarrubias-Pazaran, 2016). We estimated treatment means and temporal trends with the following model:

$$\mathrm{YIELD} \sim \mathrm{T} + \mathrm{Yn} + \mathrm{T{:}Yn} + \underline{\mathrm{F{:}Y}} + \mathit{F{:}Y{:}T} \qquad (1),$$

where T stands for treatment, F for field, Yn for year used as numeric, Y for year used as factor, underlines indicate the random effects, italics the residuals, and a colon (:) an interaction. We did not include a field main effect, due to the shifted rotations across fields, the field effect is confounded with the year effect. The combined effect of year and field was necessary, because in some years the same crop occurred on two fields due to appearing twice in the crop rotation (winter wheat) or due to double cropping (grass-clover). Least square means, estimated trends per treatment, and linear contrasts were computed with the R-package *emmeans* (Lenth, 2018). For the calculation of linear contrasts, the NOFERT treatment was excluded, because it was not part in any contrast. A letter display indicating significance of pairwise differences (α=0.05) for means and trends, based on the suggested algorithm by Piepho (2004), was produced with the R-package *multcomp* (Hothorn et al., 2008). As preceding crops have changed during the course of the experiment, yield trends were analyzed in winter wheat only after potatoes (14 field x year combinations and during the whole experiment) and in potatoes after grass-clover (9 field x year combinations in cycles 2 to 4), to avoid any effects from changes in preceding crops on the estimated yield trends.

To investigate any relation between the average amount of applied nutrients over all crops (as shown in Table 2) on the mean yield per treatment, we compared a linear ($y = a + bx$) and a square root regression model ($y = a + bx + cx^{0.5}$), which has been found to fit well for fertilization-yield relationships (Bélanger et al., 2000). The model returning the greater adjusted R^2 was chosen. In this analysis, we excluded the NOFERT treatment to avoid an overestimate of the fit statistic, as NOFERT is very distant to the other treatments, and because we were rather interested in the differences between the fertilized treatments occurring in practice.

Although absolute stability can equivalently be assessed by standard deviation or variance, we used the variance here, as we aimed to calculate standard errors (SE) and significances of pairwise comparisons (Ahn and Fessler, 2003). As a probable yield trend could lead to an increased estimate of the variance, we estimated the variance with the following mixed model:

YIELD ~ T + Yn + T: Yn + *VS(T,F:Y)* (2),

which resembles model (1), except that the field-year effect (F:Y) and the residuals (F:Y:T), which represent the treatment by field-year interaction, are replaced by *VS(T,F:Y),* which symbolizes a diagonal variance matrix with the diagonal elements being the stability variances of the treatments (Piepho, 1999). Standard errors were used as reported by the R-package *sommer* (Covarrubias-Pazaran, 2016). Significances of pairwise comparisons of variances were calculated by an F-Test with n-2 degrees of freedom (df), where n is the number of field x year combinations, and 2 df were subtracted because in the calculation of the variance a mean and a slope was estimated. Pairwise comparisons were turned into a letter display using the same approach as for means and trends.

Relative stability was assessed with the coefficient of variation (CV), which was calculated by dividing the estimated variances from model (2) by the estimated means from model (1). SEs were calculated as $\left(\frac{CV^2}{2n} * (1 + 2 * CV^2)\right)^{0.5}$, where again *n* was the number field x year combinations (Rao et al., 1966). Pairwise comparisons were calculated using the asymptotic test by Feltz and Miller (1996), and subsequently turned into a letter display as for means.

Linear contrasts of variances and CV were computed by fitting two models where treatment codings were modified before model fitting. For the first model (null model), treatments to be compared formed one group and treatments not in the comparison formed the second group; i.e. two variances were estimated. For the second model (contrast model), treatments to be compared were coded into two groups following the respective contrast, and treatments not in the comparison then formed the third group, i.e. three variances were estimated. Model (2) was fit for both and twice the difference in log likelihood of both models was tested with a χ^2-test with one degree of freedom. Ratios were calculated from the estimated variances in the contrast model and 95% confidence intervals (CI) around the ratio were constructed based on an F-distribution with the total number of field x year combinations of each of the treatment groups of the respective contrast as degrees of freedom. For the CV, observed mean yields were first divided by each treatment's overall mean yield, and the square root of the ratios and CI is reported. Similarly, to the linear contrasts on means and trends, the NOFERT was omitted for this linear contrast.

To check any relation between the variance and the mean as stated by the Taylor-Power-Law (Döring et al., 2015), we regressed the natural log of the variance on the log of the mean. However, we did not find this predicted relationship in any of the investigated crops (Supplementary Figure 19).

As an additional analysis of stability, we investigated if an increased yield ratio between the organic and conventional systems in certain years is due a lower performance of the organic systems or a better performance of the conventional systems. We used only the observations from *regular* fertilization and regressed the ratio of the mean of both organic systems to the mean of both conventional systems on the mean of the organic systems and on the mean of the conventional systems.

3.4 Results

3.4.1 Mean yield

The conventional systems CONFYM and CONMIN showed significantly higher yields than the organic systems BIOORG and BIODYN for all crops except soybeans, and the yield difference was consistent at both fertilization levels (Figure 7 and Figure 8). However, the yield difference between organic and conventional systems varied substantially between crops. Under regular fertilization, the highest yield difference was observed for potatoes with the organic systems reaching 66% [95% confidence interval: 63%-69%] of the conventional yield under regular fertilization, followed by wheat (79% [77%-81%]), maize (87% [84%-91%]), and grass-clover (90% [88%-92]). The conventional treatment with half fertilization (CONFYM1) had significantly higher (potatoes, wheat) or equivalent (maize, grass clover, soybean) yields than regularly fertilized organic treatments (BIODYN2 and BIOORG2).

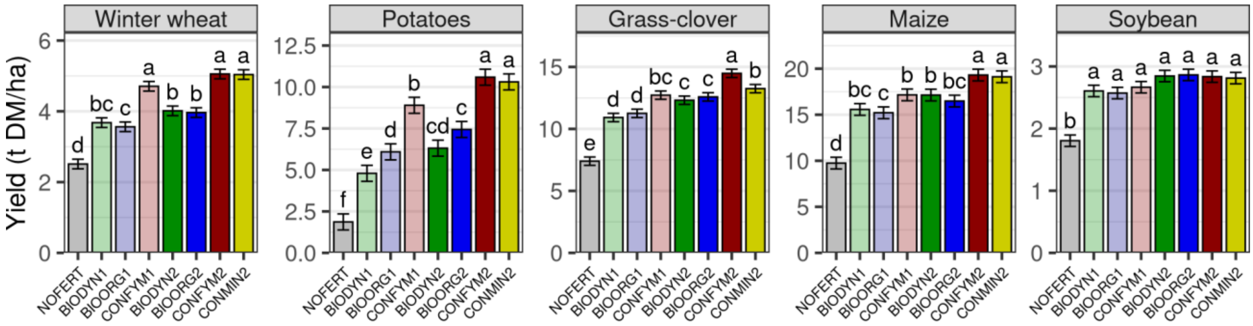


Figure 7: Treatment means for all treatments and investigated crops. Error bars represent the standard error of the mean. Treatments that do not carry the same letters are significantly different at P<0.05.

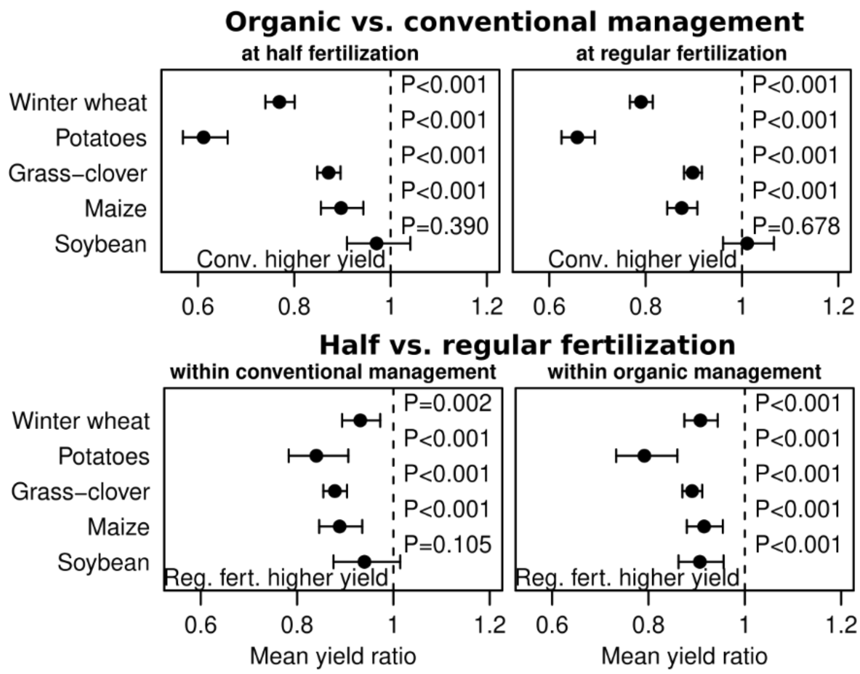


Figure 8: The effect of organic vs. conventional management (top) and of *half* vs. *regular* fertilization (bottom) on the mean yield. Effects of organic vs. conventional were compared under *half* (left; BIODYN1, BIOORG1 vs. CONFYM1) and *regular* (right; BIODYN2, BIOORG2 vs. CONFYM2, CONMIN2) fertilization, and the effect of fertilization within conventional (left; CONFYM1 vs. CONFYM2) and organic management (right; BIODYN1, BIOORG2 vs. BIODYN2, BIOORG2). The ratio was calculated as organic divided by conventional yields, and as half divided by regular fertilization. A ratio smaller than one indicates that organic yields were lower than conventional yields, and that yields at half fertilization were smaller than yields at regular fertilization, respectively. Error bars are 95% CI. Effects are significant at P=0.05, if CIs do not bracket one.

The treatments with half fertilization showed a significant reduction in yield across all crops compared with their corresponding treatments with regular fertilization (Figure 7). Interestingly, this yield reduction was of equal magnitude within organic and within conventional management (Figure 8). However, the reduction in yield through half fertilization was different between crops. While for all crops except potatoes, yields under reduced fertilization were around 90% of the yields under regular fertilization under both conventional and organic management; in potatoes, the yield under reduced fertilization was 84% [78%-91%] of yield under regular fertilization in the CONFYM system and 79% [73% - 86%] in the organic systems. While the unfertilized treatment (NOFERT) achieved around 50% of the yield of the regularly fertilized conventional treatments in winter wheat, grass-clover and maize, in potatoes it was only 18% and in soybean 64%.

Within the organic and conventional systems, there were no significant yield differences between BIOORG vs. BIODYN and CONFYM vs. CONMIN in all crops and at both fertilization levels except in potatoes and grass-clover (Figure 7). In potatoes, the BIODYN treatments showed lower yields than the BIOORG treatments at both fertilization intensities (BIODYN: 79% [70%-90%] of BIOORG at half fertilization and 85% [77%-95%] reduction at regular fertilization). In grass-clover, yields of CONMIN2 were 91% [89%-94%] of CONFYM2.

To test if the observed yields were related to the average amounts of applied nutrients, we compared a linear and a square root function excluding NOFERT. For mineral N, a square root function showed a greater adjusted R^2 than a linear function for all crops except for soybean (Table 3). For all other nutrients, a linear function did fit better for all crops except soybeans. The average amount of applied mineral N showed a considerably higher relation to mean yields in winter wheat and potatoes than other nutrients (Table 3 and Figure 9). Interestingly, in grass-clover total nitrogen showed a better relationship among than mineral N and for this relationship a square root function did fit better than a linear function. In maize, all nutrients except Ca, were highly correlated. Soybean showed a very different pattern than all other crops. While the amount of applied mineral N was not related to yield, P, K, and Mg showed the strongest relation. However, for the interpretation of the observed correlations, it has to be noted that due the design of the treatments applied nutrients were highly correlated between treatments (Table 2 and Supplementary Table 8).

Table 3: Regression of mean yield of the treatments on the average amounts of applied nutrients over all crops and over the whole course of the experiment. Values are adjusted R^2. A linear (l) and a square root (r) regression function excluding the NOFERT treatment was compared and the better fit was chosen by adjusted R^2 (indicated in parentheses). For abbreviations of nutrients, see Table 2.

Crop	TotN	MinN	P	K	Ca	Mg
Winter wheat	0.65 (l)	0.96 (r)	0.65 (l)	0.52 (l)	0.30 (l)	0.39 (l)
Potatoes	0.68 (l)	0.95 (r)	0.69 (l)	0.57 (l)	0.18 (l)	0.34 (l)
Grass-clover	0.95 (r)	0.80 (r)	0.78 (l)	0.74 (l)	0.35 (l)	0.58 (l)
Maize	0.84 (l)	0.89 (r)	0.90 (l)	0.81 (l)	0.63 (l)	0.75 (l)
Soybean	0.69 (r)	0.08 (l)	0.82 (r)	0.90 (r)	0.66 (r)	0.84 (r)

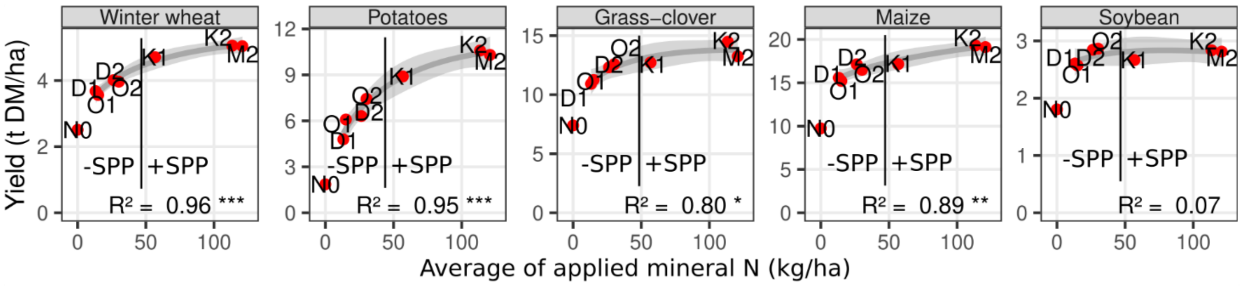


Figure 9: Regression of the mean yield of the treatments on the average amounts of applied mineral N through a square root function. R^2 indicates adusted R^2. NOFERT (N0) was excluded for the regression fit to avoid an overestimation of the fit statistic (R^2), but included in the plot for comparison. O: BIODYN, D: BIODYN, K: CONFYM, M: CONMIN; 2: regular, 1: half fertilisation. –SPP and +SPP indicates the treatments without and with, respectively, synthetic plant protection.

3.4.2 Yield development

The overall trends were positive for all treatments in winter wheat, potatoes, and soybean, but negative in grass-clover and maize (Figure 10). As the preceding crops of winter wheat and potatoes changed during the experiment, we separately analyzed wheat that followed potatoes and potatoes that followed grass-clover. In winter wheat after potatoes, the observed trends were overall slightly lower but correlated to the estimated trends using all preceding crops (r=0.94**, omitting NOFERT). In potatoes after grass-clover, estimated trends were negative for all treatments and the order of treatments was different compared to potatoes using all preceding crops (r= -0.37 ns, omitting NOFERT). Only CONFYM2 in winter wheat showed a significant increase, and NOFERT and BIODYN1 in grass-clover a significant decrease. However, NOFERT showed the strongest decline or the least increase in yield in all crops except in soybean. The treatment with the strongest increase in yield was CONFYM2 in winter wheat and potatoes, and CONMIN2 in grass-clover and maize. While in all crops except soybean the organic treatments ranked in between, they showed a greater yield increase than the conventional treatments in soybean. A similarity in yield change between crops was only observed between potatoes and grass-clover (Spearman rank r=0.68 ns, and r<0.36 for all other pairs), and the trends in soybean were negatively correlated to all other crops (Supplementary Table 9).

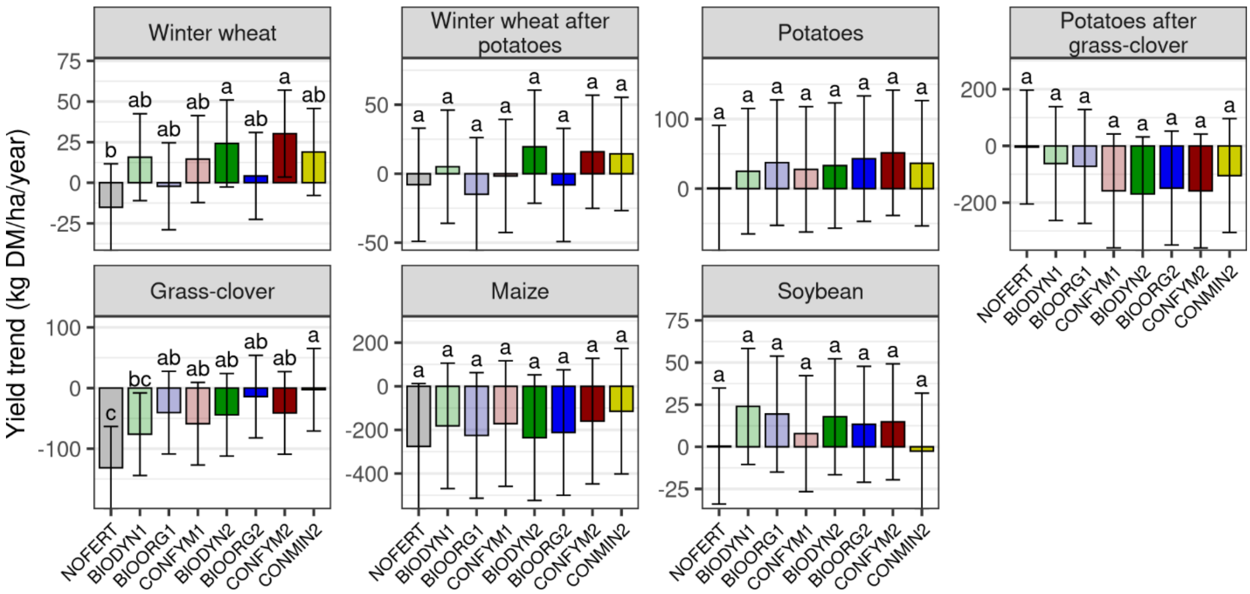


Figure 10: Estimated yield trends per crop and treatment. As preceding crops have changed during the course of the experiment, the yield trends were additionally analyzed in winter wheat only after potatoes and in potatoes only after grass-clover. Values greater than zero denote yield increase, and smaller than zero yield decrease. Error bars indicate the 95% confidence interval of the estimated trend and trends are significantly different from zero at P<0.05 if error bars do not overlap zero. Treatments that do not carry the same letters are significantly different at P<0.05.

To test, if there were any differences in yield trend between treatment groups, we calculated linear contrasts (Table 4). In the comparison between organic and conventional management, the most significant difference was observed in winter wheat with organic treatments showing a smaller increase (P=0.07). When comparing the effect of half vs. full fertilization, in winter wheat after all preceding crops regular fertilization led to a stronger increase (P=0.1); when using only winter wheat after potatoes, this difference was more significant (P=0.05) and half fertilization showed a decrease in yield. In grass-clover, half fertilization showed a significantly stronger decrease in yield. Interestingly, we found significant differences between the BIODYN and BIOORG treatments: In winter wheat, the yields of the BIODYN treatments increased significantly more than the BIOORG treatments (P=0.01), while this was reversed in grass-clover, but also significant (P=0.03).

Table 4: Linear contrasts of yield trends for organic vs. conventional (BIODYN1, BIOORG1, BIODYN2, BIOORG2 vs. CONFYM1, CONFYM2), half vs. regular fertilization (BIODYN1, BIOORG1, CONFYM1 vs. BIODYN2, BIOORG2, CONFYM2), and BIODYN vs. BIOORG (BIODYN1, BIODYN2 vs. BIOORG1, BIOORG2). Δb denotes the difference in slope between the compared groups, SE the standard error, and P the significance of Δb being different from zero. Bold P-values indicate P<0.1, * and ** indicate P<0.05 and P<0.01, respectively. For comparison, the mean trends of the compared groups are shown (b_1 and b_2).

	Contrast (group 1 - group 2)				**Group 1**	**Group 2**
Crop	**Δb = b_1-b_2 (kg/ha/year)**	**SE**	**P**		**b_1 (kg/ha/year)**	**b_2**
Organic vs. conventional					**Organic**	**Conventional**
Winter wheat	-11.9	6.5	**0.07**		10.5	22.4
Winter wheat after potatoes	-6.7	6.9	0.33		0.4	7.1
Potatoes	-4.9	21.7	0.82		34.7	39.6
Potatoes after grass-clover	45.6	62.6	0.47		-113.1	-158.7
Grass-clover	6.2	12.6	0.63		-43.7	-49.9
Maize	-48.1	62.4	0.44		-213.8	-165.6
Soybean	7.3	11.6	0.53		18.6	11.3
Half vs. regular fertilization					**Half fertilization**	**Regular fertilization**
Winter wheat	-10.1	6.2	**0.10**		9.4	19.5
Winter wheat after potatoes	-12.9	6.5	**0.05**		-3.8	9.1
Potatoes	-12.5	20.4	0.54		30.1	42.6
Potatoes after grass-clover	61.2	59.0	0.31		-97.7	-158.9
Grass-clover	-25.4	11.9	**0.03**	*	-58.5	-33.0
Maize	9.9	58.9	0.87		-192.8	-202.7
Soybean	1.7	10.9	0.87		17.1	15.3
BIODYN vs. BIOORG					**BIODYN**	**BIOORG**
Winter wheat	18.9	7.5	**0.01**	*	19.9	1.0
Winter wheat after potatoes	23.8	8.0	**<0.01**	**	12.3	-11.5
Potatoes	-11.1	25.0	0.66		29.1	40.2
Potatoes after grass-clover	-5.2	72.3	0.94		-115.7	-110.5
Grass-clover	-32.8	14.6	**0.03**	*	-60.1	-27.3
Maize	10.1	72.1	0.89		-208.7	-218.9
Soybean	4.5	13.3	0.74		20.9	16.4

As applied rates of fertilizer were not constant during the course of the experiment (particularly in the conventional treatments in winter wheat, see Supplementary Figure 11), we also regressed the estimated yield trends on the change of mineral N per year. However, there were never any significant relationships (Supplementary Figure 13).

3.4.3 Yield stability

To identify the effects of conventional vs. organic management and of the fertilization levels, we calculated linear contrasts of the stability measures (Figure 11, see Supplementary Figure 18 for the estimates per treatment). As for the stability measures smaller values indicate less variation and thus better stability, we will use the term “more stable” for lower values. In the comparison of the stability of conventional vs. organic farming, we found no significant differences in absolute stability, measured by the variance, for all investigated crops. However, relative stability, measured by the CV, was significantly more stable in conventional management in all crops except soybean. In winter wheat and grass-clover, the CV of organic management was around 34% higher, and in potatoes and maize about 65% higher, than of conventional management, indicating better stability for conventional management.

When comparing the two fertilization levels, half fertilization revealed to be more stable in absolute stability in all crops except soybean, although never significantly. Interestingly, relative stability was very similar between half and regular fertilization for all crops.

As relative stability is measured by the CV, and can thus be influenced by the mean yield, we assessed the relation between relative stability and the mean yield through correlation analysis (Figure 12). There was a negative relationship between both measures in all crops. In winter wheat, potatoes, and grass-clover, we furthermore observed a grouping with all organic treatments showing a lower mean yield and lower relative stability (indicated by a greater value, as lower numbers express enhanced stability), and conventional treatments showing a higher yield and better relative stability.

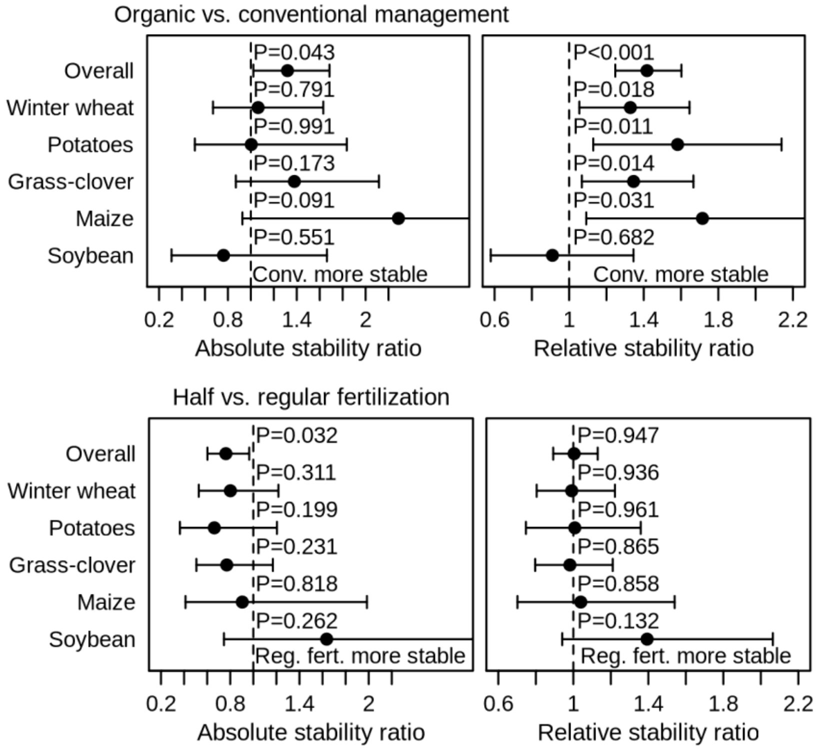


Figure 11: Contrasts of absolute (left) and relative (right) stability comparing organic vs. conventional management (top; BIODYN1, BIOORG1, BIODYN2, BIOORG2 vs. CONFYM1, CONFYM2) and half vs. regular fertilization (bottom; BIODYN1, BIOORG1, CONFYM1 vs. BIODYN2, BIOORG2, CONFYM2). Ratios were calculated as the respective stability measure of organic to conventional management and half to regular fertilization. A ratio of 1 indicates that that the stability is the same between groups, and >1 indicates that conventional management, respectively regular fertilization is more stable. Error bars are 95% CI of the ratio and error bars not bracketing 1 indicated that the ratio is significantly different from 1 at P<0.05.

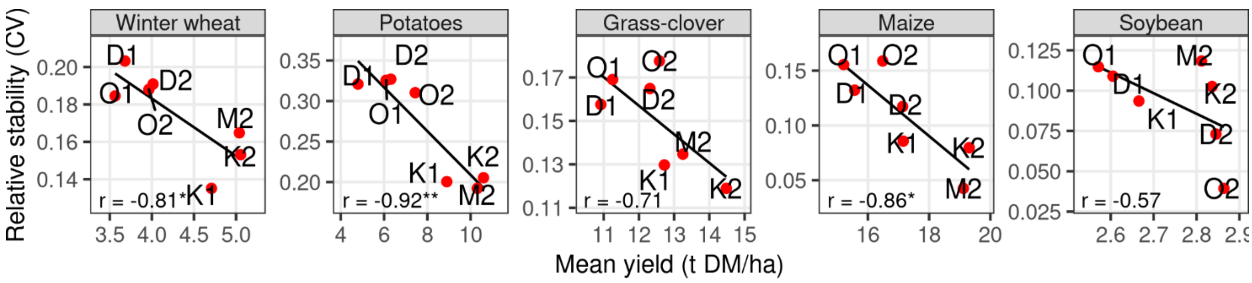


Figure 12: Relation of relative stability (CV) and mean yield per treatment. Lower numbers of relative stability express better stability. r is the Pearson correlation coefficient, * and ** indicate significance at P<0.05, and P<0.01, respectively. NOFERT has been omitted, as it is distant to other treatments and could thus lead to overestimation of the correlation.

As we found that the yield ratio between organic and conventional management at regular fertilization varied substantially between years (see variation on x-axes values in Figure 13), we tested whether an increased yield difference between organic and conventional management was due to better performance of the conventional treatments or worse performance of the organic treatment through correlation analysis (Figure 13). For all investigated crops except soybean, the yield ratio was significantly correlated to the mean yield of the organic treatments (P<0.01 for potatoes and P<0.001 for winter wheat, grass-clover, and maize), but not to the mean yield of the conventional treatments. Thus, in these crops the yield difference was due to lower yields of the organic treatments. In contrast, in soybean, the yield ratio was significantly correlated to the mean yield of the conventional treatments (P<0.001), but not to the mean yield of the organic treatments (P>0.05).

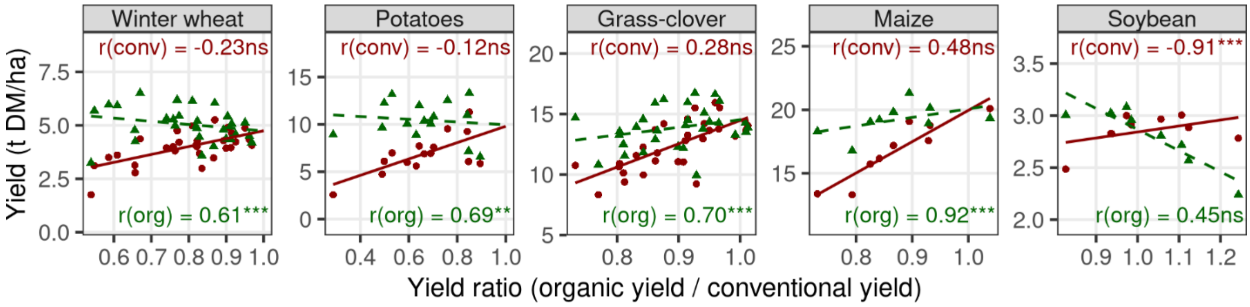


Figure 13: Relation of yield ratio between organic and conventional management to the mean yield (organic: BIOORG2 and BIODYN2, conventional: CONFYM2 and CONMIN2) of organic treatments (green dots and solid line) and conventional treatments (red triangles and dashed line), respectively. One dot represents one year x field combination and r indicates the Pearson correlation coefficient. ** and *** indicate significance at P<0.01, and P<0.001, respectively

3.5 Discussion

3.5.1 Mean yield

We found that organic yields were reduced between 10% and 34% compared with conventional management dependent on the investigated crop at regular fertilization, while we found no yield difference in soybean (Figure 8). This reduction was similar at half fertilization. Half fertilization resulted in a yield reduction of around 10%. This difference also varied between crops but was consistent in organic and conventional management. When we compare these yield differences between organic and conventional systems, we should always keep in mind the discerning factors fertilizer amount and form, and pesticide input.

The average yield difference over all investigated crops, and across both fertilizer, levels is around 15%. This difference is similar to estimates from previous meta-analyses (e.g. 19% in Ponisio et al. (2015) and 16% in Knapp and van der Heijden (2018)). While we found a yield difference of 21% in winter wheat and 34% in potatoes under regular fertilization in the DOK experiment, in a comparison of winter wheat yield from variety trials in Switzerland, Herrera et al. (2020) found that organic yields were 38% and 33% lower than conventional yields with high- inputs and low-inputs, respectively. However, these differences are probably slightly over-estimated as conventional trials were conducted on experimental stations and organic trials on farmers' fields. Rudmann and Willer (2005) using on-farm yields in Switzerland reported a yield difference of 30% for winter wheat and 37% for potatoes. In Germany, it was found that the on-farm yield difference in cereals and potatoes is around 50% (BLE, 2018). Analyzing the variation in yield difference between organic and conventional farming, De Ponti et al. (2012) argued that the variation in yield difference might be largely due to the intensification level of the conventional farming system. Regarding the smaller difference in winter wheat in Switzerland, it has to be noted that conventional wheat cropping in Switzerland focuses on the production of high baking quality wheat under integrated production with reduced inputs of fertilizer and often without fungicides. This could result in lower yields and thus explain the smaller difference between organic and conventional winter wheat yields in Switzerland compared to Germany.

The observation that the yield difference between organic and conventional management differs substantially between crops was also made in previous studies (Seufert et al., 2012; Ponisio et al., 2015). The greatest yield difference was observed in potatoes. This might have been due to a stronger dependence on in-season fertilization, which is supported by the greatest yield reduction in the unfertilized treatment and the strongest reduction in half vs regular fertilization. Furthermore, differences in plant protection might have been another reason for this difference, as control of *Phytophtora infestans* (see also below), *Alternaria solania,* and Colorado potato beetle (*Leptinotarsa decemlineata*) is challenging in organic management. In a further investigation, we found, that the yearly yield differences were partly explained by the day of planting (r=0.58*, Supplementary Figure 17). An underlying reason could be that at earlier planting plants might have matured more before infestation with *Phytophtora infestastans*. The smallest yield difference was observed in soybean and this difference was not significant. Although soybeans received minor amounts of fertilization (no applications in the organic treatments) and weeds were controlled chemically, this did not result in significant yield difference.

Regressing the mean yields on applied nutrients revealed that yields of winter wheat and potatoes were best related to the average amounts of applied mineral N (Table 3). It is important to note that the design of the treatments was rather to reflect common agricultural practices than to identify limiting nutrients. Therefore, applied nutrients were highly correlated between treatments (Supplementary Table 8), which partly hinders the determination of yield limiting nutrients and explains the many significant relations observed. Nonetheless, mineral N was considerably better related to yields in winter wheat and potatoes than other nutrients. This strong relation could suggest that the amount of applied mineral N is the primary reason for the observed yields in these species. Applied nitrogen as a main source for the yield difference between organic and conventional farming was also observed in recent meta-analyses (Seufert et al., 2012; Ponisio et al., 2015). However, as the organic treatments and the unfertilized treatment received less fertilizer and less disease control than the conventional treatments (see Figure 9), fertilization intensity is strongly confounded with disease control intensity in this experiment. It is thus difficult with the given set of treatments to disentangle the effects of fertilization and plant protection. Bilsborow et al. (2013) analysed the effects of organic vs. conventional fertilization management and crop protection in winter wheat in an orthogonal design. They found both effects to be of equal magnitude, which would indicate that the yield difference between organic and conventional farming in winter wheat would be equally due to the difference in fertilization and in disease control. Furthermore, Berry et al. (2010) pointed out that the effects of disease control interact with N fertilization, resulting in stronger effects of disease control under higher N fertilization on the one hand, and in stronger effects of N fertilization under disease control on the other hand. This, in turn, would support that the yield differences between the organic and conventional treatments could be due to the interaction of disease control and fertilization.

Interestingly, the reduction in yields through half fertilization was the same in relative measures under organic and conventional management in investigated crops and for each crop. However, as yields were higher in the conventional treatments, the absolute decrease in yield through half fertilization is therefore greater under conventional treatment. This would be in accordance with the stronger effect of fertilization in combination with higher crop protection intensity as noted by Berry et al. (2010).

The strong dependence of winter wheat and potato yields to applied mineral N forms suggests a potential to increase yields in organic farming through increased application of mineral N forms and a better synchronisation of crop N demand and supply. This could be achieved through e.g. separation of NH_4-N from liquid manures and a more targeted application or direct application of N rich legume residues (Berry et al., 2002). However, it has to be noted that organic N, not immediately available to the crop, is important for maintaining soil fertility (Mäder et al., 2002; Gutser et al., 2005).

Differently to the strong determination of yields through mineral N in winter wheat and potatoes, yields of grass-clover were more related to total N, but also to P and K, and maize yields to all nutrients, but less to Ca (Table 3). In both crops, yield differences between organic and conventional management were smaller than in wheat and potatoes (Figure 8). These observations suggest, that yield differences between organic and conventional management might be smaller in crops that are less dependent on in-season applied N due to biological N fixation, which is also supported by the absent yield difference in soybean. In grass-clover yield difference might then be determined by other nutrients, like P and K, which have been shown to be important for the establishment of clover in grass-clover mixtures (Andrew, 1960; Fortune et al., 2004).

Although BIODYN and BIOORG systems differ in the form of fertilizer (manure compost vs slightly rotted manure) and additional biodynamic preparations in BIODYN, we could not detect any yield differences except in potatoes. The increased yield of BIOORG in potatoes is probably due to an increased amount of mineral N and due to the application of copper against *Phytophtora infestans* in BIOORG. The latter is supported by the observation that in five years where no copper was applied there was no significant yield difference (P=0.93), while in the remaining ten years the yield of the BIODYN treatment was significantly reduced by 20% at regular fertilization (P=0.001, see Supplementary Figure 16). The absent effects of compost vs. manure on yields, and the amounts of applied total and mineral N, might have been due a large share being applied as slurry.

3.5.2 Yield development

To evaluate the long-term effects of the treatments on yield, we regressed yields on the year. Overall, the estimated yield trends did hardly differ significantly from zero (Figure 10). The absence of significant effects might partly be due to the great variations in yield across years (see also Supplementary Figure14), which are common in yield observations, and lead to rather high standard errors of the estimates, which in turn lowers the ability to find significant increases or decreases in yield.

In the analysis of differences between groups of treatments in winter wheat, we found three significant contrasts: (1) conventional management (CONFYM and CONMIN) showed a more positive trend than organic management, (2) regular fertilization a more positive trend than half fertilization, and (3) BIODYN treatments a more positive trend than BIOORG treatments (Table 4). While the first two might be related to overall fertilization intensity, the latter seems more difficult to explain. In a more detailed analysis, we found that in 4 years in the early stage of the experiment BIOORG2 had significantly ($P<0.05$) higher yields than BIODYN2, while in the later stage this was reversed in 3 years, with an intermediate stage without any significant differences (data not shown). The major differences in applied nutrients between BIOORG and BIODYN were considerably higher rates of applied Ca in BIODYN (104 kg/ha/year in BIOORG vs 160 kg/ha/year in BIODYN, Table 2) and that farmyard manure was applied through rotted manure in BIOORG and through composted manure in BIODYN. The higher Ca rate is mainly due to the origin of the compost from a farm on a more calcareous soil, which contained more than double the amount of Ca, and thus represents an external input of Ca. Mäder et al. (2002) reported that the soil of the BIODYN treatments had a substantially higher Ca content, a slightly higher pH value and organic carbon content, and a higher aggregate stability than the BIOORG treatments. Pocknee & Sumner (1997) found that Ca and Mg bound in organic form can have similar effects on soil pH as mineral lime. Furthermore, Ca can have positive effects on soil structure (Bronick and Lal, 2005). Although it can only be speculative, the more positive trend in BIODYN than in BIOORG might have been due the higher Ca input and/or composted manure vs. rotted manure application. While positive interactions of organic carbon content and pH-values in soil are known, underlying mechanisms and their interactions are still not well understood (Paradelo et al., 2015).

When analyzing only potatoes proceeding grass-clover, the yields were decreasing considerably with the regularly fertilized treatments showing the strongest decrease and NOFERT the smallest decrease (Figure 10). Although we have no data on the occurrence, we speculate that this overall decrease might have been due to wireworm (*Agriotes* spp.) infestation, which is a particularly increased when potatoes are grown after grass-clover (Parker and Howard, 2001). However, it seems difficult to explain the yield decreases of the different treatments, as it was neither explained though the contrasts of organic vs. conventional management nor through half vs. regular fertilization (Table 4). The overall positive yield trend, when analyzing potatoes after all preceding crops, was most likely due to the effect of soybean as preceding crop in cycle 5, which resulted in higher yields in all crops (Supplementary Figure 15), and thus representing a recovery from depressed yields due to the wireworm infestation. It is interesting to note that the yields of the NOFERT treatment neither did decrease after grass-clover like all other treatments, nor responded to the change in the preceding crop (Supplementary Figure 15).

The yield of grass-clover was decreasing in all treatments (Figure 10). As grass-clover was always grown after cereals during the whole experiment, any effect on the trend estimates from changing preceding crops can be excluded. However, in the contrast analysis we found half fertilization resulted in a stronger decrease than regular fertilization and that BIODYN yields decreased more than BIOORG yields (Table 4). We speculate that this overall decrease in yield might have been due to limitation of sulphur (S) application, as clover has been shown to react strongly to S application, which alters the plant composition and thus reduces N fixation through clover, resulting in a reduction of yield (Walker and Adams, 1958; Tallec et al., 2008). The slightly stronger decrease in yield of CONFYM2 than of CONMIN2, might have been due to CONMIN2 receiving consistently. The stronger yield decrease in BIODYN than in BIOORG could be due to BIORG receiving occasional applications of potassium sulfate in cycles 3-5 with around 40 kg S/ha per application. S fertilization has not been deemed necessary up to the 1990s as S was deposited in sufficient amounts due to high S emissions from burning of coal and fossil fuels. However, following restrictions on S emissions and technical inventions, S emissions were significantly reduced after 1989, which in turn led to a decrease of deposition rates (Stern, 2005). It is by now widely accepted that current S depositions are insufficient to maintain crop yields, and S fertilization has to be applied (Webb et al., 2016).

3.5.3 Yield stability

In the comparison of organic vs. conventional management, we found no difference in absolute stability (as measured by the variance) but conventional management was more stable in relative stability (as measured by the coefficient of variation), which sets absolute stability in relation to the yield level (Figure 8). The finding of similar absolute stability but different relative stability is in agreement with the meta-analysis of Knapp & van der Heijden (2018) and of Smith et al. (2019). Both studies argue that increased N fertilization can help to increase relative stability of organic management through increased yields. However, we found no significant difference in relative stability between half and regular fertilization. While the overall difference in relative stability between organic and conventional management can be attributed to difference in mean yield (Figure 12), the difference in mean yield due to higher fertilization did not result in a difference in relative stability. As in the discussion on mean yield, it is difficult to disentangle the effects of plant nutrition and plant protection on stability as intensity of plant nutrition is concomitant with intensity of plant protection in the design of the treatments. However, the observation that increased fertilization resulting in a significant increase in mean yield but not in increased relative stability, suggests that relative stability is also determined by plant protection. Particularly, in potatoes relative stability is equal within all organic and within all conventional treatments, which furthermore supports that there is no effect of the fertilization intensities on relative stability (Figure 12). In turn, this suggests that relative stability in potatoes is strongly determined by plant protection.

We found a considerable variation in the yield difference between conventional and organic management across years (Figure 13). While in some years, organic treatments had almost similar yields as the conventional treatments, in other years, organic treatments yielded only half of the conventional treatments. We therefore conducted a regression of the yield ratio on treatments' yields as an in-depth analysis of stability. We found a strong correlation of the yield ratio to the yields of the organic treatments in all crops except soybean. This indicates that the greater yield difference in certain years is due to the lower performance of the organic treatments. Our proposed concept of investigating the yield ratio across years can also be related to stability analysis. As the variation in yield ratio is not correlated to the yield of the conventional treatments, the conventional treatments show lower variation and thus increased stability than the organic treatments.

3.6 Conclusion

Yields were significantly reduced in the organic systems between 10% and 34%, dependent on the investigated crop, while there was no yield reduction in soybean. In winter wheat and potatoes, the mean yields of all fertilized treatments were strongly related to the amount of applied N in mineral form, pointing towards the importance of plant available N for the yield difference between conventional and organic management. We found only marginal differences in yield development between conventional and organic management, suggesting that organic management did not lead to a yield decline compared to conventional management. In accordance to previous studies, we found that absolute stability (measured by the variance) was similar between both systems, but relative stability (measured by the CV) was more stable in conventional management. As relative stability corrects for the difference in mean yield, the better stability of conventional management was due to its higher yields. However, we found that half fertilization did not result in reduced relative stability, as would be expected due to the decreased yield. This suggests that the difference in relative stability between conventional and organic might be partly also due to the difference in the control of weeds, pests and diseases. Furthermore, we found the increased yield differences in certain years between conventional and organic management was more due to lower yields of the organic management than due to higher yields of the conventional management.

Acknowledgments

We acknowledge the financial and personnel support by Agroscope, the Research Institute of Organic Agriculture FiBL and the Federal Office for Agriculture FOAG during 40 years of DOK management and research.

4 Chapter C: Comparing the breeding progress in winter wheat under conventional and organic management in Germany

4.1 Abstract

Challenged by a predicted increase in world population and growing environmental concerns regarding negative environmental impacts of agricultural production farming needs to increase its production and produce more sustainably. Organic farming has been established with the aim to reduce environmental impacts but has been shown to be limited in its production potential. Plant breeding plays a key role in increasing the productivity and decreasing environmental effects. With regard to plant breeding for organic farming, it is still unclear if organic farming has benefitted from plant breeding, and if direct selection and separate testing under organic conditions is necessary and allows for yield improvement in organic farming. With its comparably long history of organic farming, activities of organic breeding companies and official variety-testing networks under organic management, Germany can be used as a valuable case study to address these questions.

We used publicly available on-farm yield data and data from German variety recommendation trials of winter wheat from the years 2001 to 2017, including around 1900 conventional trials and 500 organic trials. Breeding progress was assessed employing a mixed model approach to separate genetic and non-genetic trends, and by calculating the genetic-trend within each single trial. The performance between management intensities was assessed by correlating variety means.

The overall yield difference was around 50% both on-farm and in the trials. While conventional yields increased, organic on-farm yields are stagnating in the last 30 years. Genetic trends were significant for both management systems. However, we found that these estimates might be over-estimated due to variety ageing because of breaking disease-resistances. The effective genetic trend in organic management might thus tend towards zero, which would explain the absent yield trend. Varieties bred under organic conditions showed lower yields but higher baking quality. Analysis of variance components revealed variety x system interaction to be low and variety performance for yield and quality traits was highly correlated between conventional and organic management.

Although plant breeding was successful to maintain the yield level under organic management through counteracting the breaking of disease-resistance, it might not have contributed to yield improvement under organic management. This might be due to limiting growing conditions, particularly nitrogen availability, and wheat cropping might be already very efficiently scavenging available nutrients. While organically bred varieties allow for the production of sufficient baking quality under limiting conditions, the increased quality comes at the cost of reduced yield. As the higher quality was not constrained by less favorable growing conditions and because of strong correlations between management systems, the higher quality might have equally been achieved by selection under conventional management. However, due to the focus on higher quality and a limited number of varieties, the available dataset seems partly inappropriate to address the question of direct vs. indirect selection. While strong correlations of variety means for yield and quality traits might suggest that separate testing is not necessary, we argue that rather trait-specific experiments should be conducted to yield better information on management-specific traits.

4.2 Introduction

Due to a growing population and increased consumption, food demand is expected to increase for at least another 40 years (Godfray et al., 2010; Tilman et al., 2011). If the further increase in production shall not be based on additional land usage and thus reduction of natural ecosystems, production per area needs to be increased, or changes towards less animal-based diets are needed (Phalan et al., 2011; Springmann et al., 2016).

Past increases in production have been driven by a highly successful improvement in management, mainly through increased external inputs like synthetic fertilizers, pesticides and herbicides, and due to cultivar improvement through plant breeding (Shiferaw et al., 2013). However, the indiscriminate use of external inputs has been recently seen in a more and more critical light for its negative impacts on the environment like, e.g. increases in greenhouse gas emission, eutrophication of water, and loss of biodiversity (McLaughlin and Mineau, 1995; Johnson et al., 2007; Moss, 2008). In addition, in many environments, global climate change is expected to lead to more frequent occurrence of drought and heat stress, to more frequently changing occurrence of pests and diseases and to less reliable weather conditions in general (Schmidhuber and Tubiello, 2007). The challenge of the future will thus be to further increase production, to produce more sustainably, and to cope better with unreliable environmental conditions and with biotic and abiotic stress (Foley et al., 2011).

Plant breeding is likely to play a central role in this challenge, as improved cultivars may use available resources more efficiently, reduce the need for pest control through better resistances, and react more stable against unreliable environmental conditions (Bailey-Serres et al., 2019). In cereals, yields have increased dramatically in the 20th century, whereas the increases were due to improvement in management, plant breeding and, importantly, the interplay of both (reviewed in Bingham (1981) and Feil (1992)). Mackay (2011) estimated that while between 1948 and 1982 yield improvement in winter cereals in the UK was about equally due to improvement in management and due to plant breeding, since then at least 88% of the yield improvement was due to plant breeding. However, breeding of cereals has focused and been successful mostly in high-input cropping systems (Foley et al., 2011; Tilman et al., 2011). It thus remains an open question, if plant breeding has been equally successfully in different cropping systems and environmental conditions (Evenson and Gollin, 2003). While lacking breeding progress could also be due to few breeding activities for minor crops due to a limited seed market, or due to the fact that farmers are not able to afford commercial varieties, certain cropping systems or environments might not have benefitted from improved varieties due to genotype x environment (G x E) interaction (Ceccarelli, 1996). In general, the term "G x E interaction" describes a situation, where the ratios of performance within a set of genotypes differs among environments, while a reversal of the performance is called crossover G x E interaction (Annicchiarico, 2000). Note that here environment refers to both environment in the actual sense but also to the cropping system.

With regard to breeding progress, presence of G x E interaction might result in reduced or unsuccessful breeding progress in one although being successful in another environment or cropping system. By comparing a set of cultivars under different rates of applied N, several studies have found that the observed breeding progress is lower or even non-significant the less N is applied (Canevara et al., 1994; Ortiz-Monasterio et al., 1997; Brancourt-Hulmel et al., 2003; Rueda-Ayala et al., 2018).

In effect, strong G x E interaction needs specific breeding in the target environment, which is termed direct selection, as opposed to indirect selection, where breeding is assumed to be equally successful in several environments with no to little G x E interaction (Ceccarelli, 1996).

The necessity of direct vs. indirect selection is a matter of long debate in plant breeding (well summarized in Ceccarelli (1996)). In a recent study, Voss-Fels et al. (2019) found that the performance of winter wheat varieties in Germany is highly correlated between optimal management and management with reduced N fertilization and without fungicide application.

They concluded that plant breeding has been successful for both high and reduced input management. Another example is breeding for drought conditions. While the selection under non-drought conditions has been shown to be also effective under drought conditions (Richards, 1996), Cecarelli (1996) could show that for barley in Syria, direct selection under low yielding conditions was more effective for target low-yielding sites. However, he also noted that direct selection is only more effective below a certain yield level. Further evidence for the advantage of direct selection for different species and under limiting conditions are given in Atlin and Frey (1989).

Among the many approaches to reduce negative environmental impacts, organic farming has been established to achieve this goal through avoidance of mineral fertilizers and synthetic pesticides combined with improved crop rotations, increased biodiversity on several levels and soil improvement to deal with diseases and maintain productivity (Reganold and Wachter, 2016).

Due to the non-usage of mineral fertilizers and synthetic pesticides, in contrast to conventional high-input farming, organic farming conditions can thus be generally characterized by decreased levels of available plant nutrients (mainly nitrogen), and less abilities to control pests and diseases. As growing conditions differ substantially from conventional farming, the aforementioned issues of G x E interaction and direct vs indirect selection also apply to breeding for organic farming. Likewise, several authors have argued for the necessity of direct breeding for organic farming, particularly with regard to nitrogen use efficiency and improved resistances (Murphy et al., 2007; Lammerts van Bueren et al., 2011). This is supported by Hildermann et al. (2009) and Jones et al. (2010), who, both growing a set of winter wheat varieties under organic and conventional management in Switzerland and the UK, respectively, found a higher breeding progress under conventional than under organic management. Similarly, Baresel (2006) examined 70 varieties and breeding lines in Germany and found considerably lower genetic gains under organic than under conventionally managed environments. In contrast, though not having investigated organic growing conditions, Voss-Fels et al. (2019) concluded that conventionally bred varieties are also well adapted to organic farming. Comparing the performance of barley, wheat and triticale under organic and conventional management in several European countries, Przystalski et al. (2008) have found high genetic but only intermediate rank correlations between both systems.

Furthermore, it has to be noted that within the organic farming movement the importance of breeding has gained additional attention due the skepticism about hybrid breeding and genetic engineering and concerns about intellectual property on genetic resources (Ammann, 2008).

With the foundation of the bio-dynamic and biological organic movement around 100 years ago, Germany has a fairly long tradition of organic farming (Kirchmann et al., 2016). In 2018, 9% of total German agricultural land, and 5% of the cereals area, was farmed organically (Schlatter et al., 2020). Since around 25 years, there are three breeding companies (two in Germany and one in Switzerland), that select winter wheat varieties under and for organic farming. Furthermore, a trial network has been established by the federal states for about the same time range, in which conventionally and organically bred varieties are evaluated for usage in organic farming. Data from these variety recommendation trials can be seen as a valuable case study to investigate breeding for organic farming.

Breeding progress can be assessed using historic data from variety trials or national agricultural statistics (e.g. Mackay et al. (2011) or Piepho et al. (2014)) or in trials where a historic set of varieties is grown (e.g. Ahlemayer and Friedt (2012) or Voss-Fels et al. (2019)). Using historic data allows through mixed model approaches the separation of the genetic trend, i.e. breeding progress, from the non-genetic trend, which can be due to improvement in management, e.g. cultivation technology, fertilization, or pesticides. However, because diseases constantly evolve and develop new virulence, older varieties will be more susceptible due to the breakdown of disease-resistances (Summers and Brown, 2013). This decrease in yield to broken resistances is also referred to as variety aging. Regarding the analysis of breeding progress, the lower yield of older and more susceptible varieties can thus lead to an over-estimation of breeding progress. While such over-estimation can occur, the approach based on historic data is particularly prone because older varieties in the dataset have been assessed longer and are thus more likely to have aged than younger ones (Piepho et al., 2014). One way to analyze the ageing effect is to assess the change in yield difference between treatments with and without fungicide applications, assuming that in the treatment with fungicide application diseases are fully controlled (Mackay et al., 2011). The increase in yield difference between both treatments can be used as a measure for the ageing effect.

Using on-farm data and data from conventional and organic variety recommendation trials of winter wheat in Germany as a case study, the objectives of this study were thus (1) to compare overall yield developments and the contribution of plant breeding, (2) to assess the effect of selection under organic management, (3) to compare variety performance between conventional and organic farming. To assess the over-estimation of the breeding progress due to variety ageing we compared the analysis based on historic data to the analysis within single trials.

4.3 Material and Methods

Description of datasets

Variety recommendation trials

Yield data of winter wheat were extracted from publicly available reports of the conventional and organic networks for variety recommendation in Germany (Landessortenversuche) for the years 2001 to 2017. In contrast to the trials for the assessment of value of cultivation and use (VCU), recommendation trials are conducted by the federal states (Bundesländer) independently and used for recommendation of varieties to farmers. However, a set of check varieties is included in all trials across Germany. Varieties in the trials are chosen by the responsible authorities of each state if considered of interest for regional farming practices and growing conditions. Therefore, the set of varieties can vary between the federal states. In the organic trials, varieties considered of special value for organic farming are included. Due to the focus and challenge of producing wheat with sufficient baking quality, besides organically bred varieties, varieties from Austria and Switzerland are included as well, as they often have higher baking quality.

Regular management intensity followed best local agronomic practice. The reduced management intensity received the same herbicide treatment and the same amount of fertilizer as the regular intensity, but no growth regulator and fungicide applications. However, in some federal states, nitrogen (N) application was reduced by around 30-50 kg N/ha up to the year 2004. In the states Lower Saxony, North Rhine-Westphalia, and Schleswig-Holstein, three different management intensities were performed, with an additional intermediate intensity, which differs from the regular intensity by a reduced fungicide application, which has not been specified in detail. Because only the mean yield of this intermediate and the regular intensity has been reported, we used these yield data as regular intensity. Furthermore, no yield data from the reduced management intensity was published from these states and could thus not be included in this analysis. In the organic trial network, trials were managed according to local practice and often included in farmers’ fields.

Additional to yield data, we also extracted data on protein content, sedimentation value, and baking volume from the organic trial network in Bavaria (also included in the German wide dataset) and received the respective data from the Bavarian State Institute for Agriculture, Freising, Germany. Grain N uptake was calculated for each observation by calculating grain N content by dividing grain protein content by 5.7, and subsequently multiplying grain N content by grain yield. Data from organic trials covered the years 2001 to 2019 and data from conventional trials were only available for the regular intensity and for the years 2006 to 2019.

Data on the year of release and quality class of the varieties were collected from the national variety catalogues of Germany, Austria, and Switzerland and from the European database on plant varieties. For varieties that were not classified for their quality class in Germany, we either used breeders' information, where available, or transformed the classification into the German system.

The German quality classification system categorizes varieties based on a set of quality traits into four different quality classes: E (elite wheat), A (quality wheat), B (bread wheat), and C (other wheat) (BSA, 2019). As the focus in organic production is on the production of wheat for baking usage, mainly E and A varieties were tested in the organic trials (Supplementary Figure 21), and we thus included only data from these two quality classes in the analysis.

Varieties were further classified by their breeding origin. Varieties where the full breeding cycle took place under organic management were classified as bred organically (Mikó et al., 2014). Such varieties originated from three different breeding companies: GZPK (Switzerland), LBS Dottenfelderhof (Germany), and Cultivari Darzau GmbH (Germany). These varieties belonged all except one to quality class E and were only assessed in the organic network.

Only varieties were included for which the year of release and information on quality class was available. After the subsetting, there were 1905 conventional trials and 477 organic trials from whole Germany, resulting on average in 112 trials per year for the conventional network and 28 trials per year for the organic network.

On-farm yields

On-farm yields for conventional management in Germany were obtained from FAOSTAT (2019). Although these data are on wheat (including spring wheat) and not specifically from conventional farming they are assumed to represent conventional winter wheat yields, because wheat is predominantly grown as winter wheat in Germany and the proportion of wheat grown organically was 2% in 2017 and lower in the time span before (BMEL, 2017). On-farm yields for organic farming were obtained from the farm network of the German Federal Ministry of Food and Agriculture (BMEL, 2012, 2019, 2005). However, this farm network primarily serves for economic assessments, and yields were mostly estimated by famers. Furthermore, the number of organic farms in the network was only 200 to 400 farms, and in some but not all reports, yields were interpolated. The organic yields were thus only used for visual representation.

Field trials

To compare the performance of varieties of different breeding origin under organic and conventional management, a field trial near Freising in the south-east of Germany, was performed. Seven varieties of varying breeding origin and quality class were grown following standard local practices (180 kg N in the conventional trials and no applied fertilizer in the organic trials) in a randomized complete block design with three replicates in the years 2015/16 and 2016/17. The conventional trials were conducted on the research station Dürnast, and organic trials on the research station Viehhausen, which is located around 5 km from the former. Soils in both locations consisted mainly of homogeneous Cambisols, while soil in Dürnast was loamy clay and in Viehhausen sandy clay. Average yearly precipitation is 800 mm and average temperature is 7.5°C. Protein content was determined by near-infrared spectroscopy.

Statistical analysis

Yield and quality data from the variety recommendation trials were reported as means over replicates, thus one observation value per combination of year, location, variety, and intensity.

Analysis of breeding progress through mixed model approach

To separate genetic and non-genetic trend we employed a mixed model approach as described by Piepho et al. (2014). For this approach, we removed data from varieties that were tested less than three years, in order to avoid bias from varieties that were only tested in one or two years, as their mean will be strongly influenced by the year effect in which they were tested (Mackay et al., 2011).

Models were described using a short-hand notation in a similar fashion as suggested by Piepho et al. (2003). The basic model for the analysis follows (Laidig et al., 2008):

VAR + YEAR + LOC + VAR:YEAR+ VAR:LOC + LOC:YEAR,   (1)

where VAR symbolizes variety, YEAR the year of observation and LOC the location, and a colon (:) indicates interaction. The residual term is then VAR:LOC:YEAR and comprises both variety × location × year interaction as well as the error of a mean, which cannot be separated due to data being available as variety x location x year means within each management intensity (Piepho et al., 2014). In the following uppercase letters will denote variables used as factors and italic letters will denote variables used as numeric variables. Underlines indicate variables used as random, whereas variables without underlines indicate fixed effects.

In order to separate genetic progress (breeding progress) from non-genetic progress (agronomic and climate-related), VAR as *rel* + VAR and YEAR is modelled as *year* + YEAR and, where *rel* is the year of release and the estimated coefficient is the genetic trend, and the coefficient for *year* is the non-genetic trend (Piepho et al., 2014). VAR and YEAR are random deviations for the respective regressions. The model is thus:

$$rel + \underline{\text{VAR}} + year + \underline{\text{YEAR}} + \underline{\text{LOC}} + \underline{\text{VAR:YEAR}} + \underline{\text{VAR:LOC}} + \underline{\text{LOC:YEAR}}. \tag{2}$$

Variety means were estimated by taking VAR as fixed and all other terms as random:

$$\text{VAR} + \underline{\text{YEAR}} + \underline{\text{LOC}} + \underline{\text{VAR:YEAR}} + \underline{\text{VAR:LOC}} + \underline{\text{LOC:YEAR}}. \tag{3}$$

To assess the overall trend, varieties were considered as nested within years, thus dropping VAR and VAR:LOC (Laidig et al., 2014). Year means for visual representation were thus modelled by taking YEAR as fixed:

$$\text{YEAR} + \underline{\text{LOC}} + \underline{\text{VAR:YEAR}} + \underline{\text{LOC:YEAR}}, \tag{4}$$

and overall trend was modelled as

$$year + \underline{\text{YEAR}} + \underline{\text{LOC}} + \underline{\text{VAR:YEAR}} + \underline{\text{LOC:YEAR}}, \tag{5}$$

where the coefficient for *year* is the overall trend and $\underline{\text{YEAR}}$ the random deviation.

To assess the ageing effect, we used the difference between regular and reduced intensity of the conventional trials, where both intensities were available as response and modelled the increase in yield difference between both intensities dependent on the age of the variety. The ageing effect was thus modelled on the G:Y effect of the basic model:

$$\underline{\text{VAR}} + \underline{\text{YEAR}} + \underline{\text{LOC}} + age + \underline{\text{VAR:YEAR}} + \underline{\text{VAR:LOC}} + \underline{\text{LOC:YEAR}}, \tag{6}$$

with *age = year – release*. Yield differences for each year of age were accordingly estimated with AGE as categorical variable for visual representation:

$$\underline{\text{VAR}} + \underline{\text{YEAR}} + \underline{\text{LOC}} + \text{AGE} + \underline{\text{VAR:YEAR}} + \underline{\text{VAR:LOC}} + \underline{\text{LOC:YEAR}}, \tag{7}$$

Regression coefficients were tested to be different from zero using a z-ratio test, as the number of degrees of freedom to be used for a t-test is debated (Gałecki and Burzykowski, 2013). Furthermore, the number of observation was high, and Piepho et al. (2014) found that using a Kenward-Roger adjustment (Kenward and Roger, 1997) had little effect in a similar analysis. The z-ratio was calculated as the ratio of the estimated slope divided by its standard error. A P-value was subsequently calculated assuming a standard normal distribution of z. As the years of release of the varieties were not equally distributed between quality classes (Supplementary Figure 21), which could result in a bias of the estimated breeding progress due to the differences in yield level, we conducted all analyses separately per quality class.

Analysis of breeding progress through single trial approach

As a comparative analysis to the mixed model approach, we performed analysis of breeding progress in single trials, which is supposed to be less prone to over-estimation of the breeding progress (see Introduction). Though not particularly designed for this purpose, each variety recommendation trial could be seen as trial comparing a set of historic varieties. We therefore regressed the yield on the year on the year of release within each available trial and per quality class. As this analysis is less prone to bias from varieties that have been assessed less frequent, no subsetting of data as in the mixed model approach was performed. However, we only used trial x class combination, where a minimum of five varieties were available, to get reliable estimates. To assess the ageing effect, within each trial x class combination, the yield difference between regular and reduced intensity was regressed on the year of release. The dependence of the observed ageing effect on the disease pressure in a given trial was evaluated using the mean difference between regular and reduced intensity over all varieties (class A and E) in the trial. The mean difference was used as a surrogate measure for disease pressure, assuming that at lower disease pressure the yield difference is smaller due to less infection in the reduced intensity.

Variance components

To assess the variety interaction with the management intensity, we estimated variance components on combined data including conventional – regular and organic management. As this analysis was also conducted for quality traits, we used only data from the trials in Bavaria, where quality data were available. To achieve a sufficient number of varieties that were tested in both intensities this analysis included both quality classes, E and A. As intensity can be visualized as a grouping of locations, LOC effect in basic model was replaced by INT and LOC, where INT represents the intensity level and LOC the variation, which is not captured by INT. The model was thus:

$$\underline{VAR} + \underline{YEAR} + \underline{INT} + \underline{LOC} + \underline{VAR{:}YEAR} + \underline{VAR{:}INT} + \underline{VAR{:}LOC} + \underline{INT{:}YEAR} + \underline{LOC{:}YEAR}. \quad (8)$$

To estimate variance components all effects were taken as random. Because the motivation of this analysis was to investigate the size of the variety x intensity interaction in relation to the remaining effects including the variety term, we calculated the percentages for all variety effects and the residual error, as this contains the variety x trial interaction.

All statistical analyses were conducted in *R, version 4* (R Core Team, 2014) and mixed models were fit with the R package *ASReml-R, version 4.1* (Gilmour et al., 1995).

4.4 Results

4.4.1 Comparison of trial and on-farm yields

To assess the validity of the data we compared the yield from the national variety trials to German on-farm yields (Figure 14). Yield variation in the years 2001-2017, for which trial data were available, was similar in the conventional management between trial and on-farm data (r = 0.94). However, the mean yield in the conventional-regular management in the trials was 9.3 t/ha while the mean of conventional on-farm yields was 7.6 t/ha in the respective period, thus trial yields being 18% smaller than on-farm yields. The mean yield in the organic trials was 4.6 t/ha and 3.6 t/ha on-farm, resulting in a relative difference of 23%, slightly greater than in conventional management. The observed yield trend in the period 2001-2017 was similar in both conventional trial intensities and on-farm, though only significant in the conventional – reduced intensity in the trials. Interestingly, the relative yield difference between conventional on-farm yields and trial yields did not change during the investigated period (P = 0.85). Contrasting to the conventional on-farm yields having increased steadily until to around 2000 but still increasing afterwards, organic on-farm yields seemed to remain constant from the beginning of the dataset (1980) to 2017. In the investigated period, organic yields amounted to 49% of the conventional yields on-farm, while in the trials organic yields were 47% of the conventional yields.

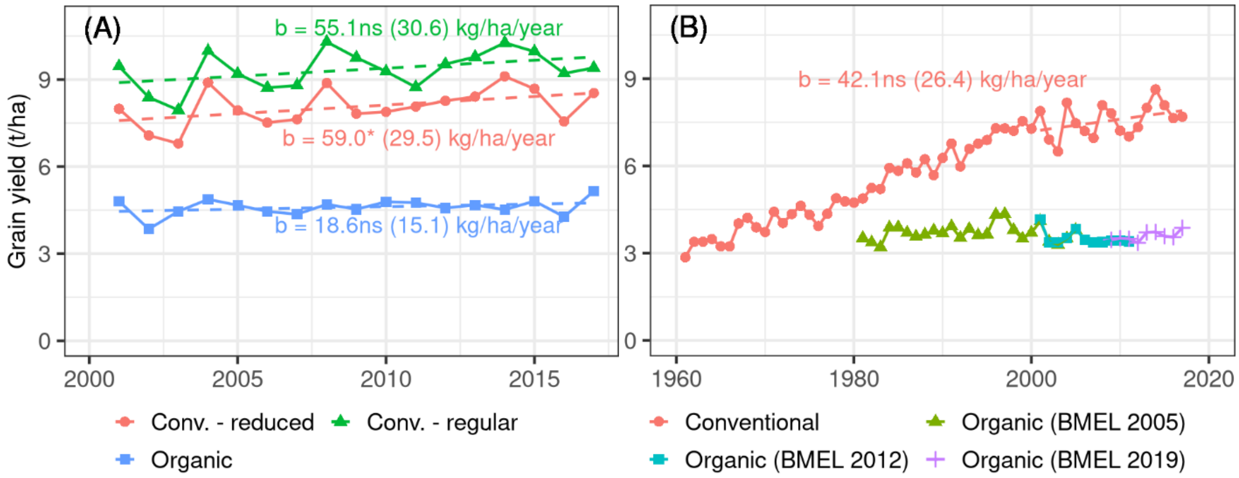


Figure 14: Overall yield development in the variety testing trials and on-farm. A: development in variety testing trials of Germany for quality class A for both conventional intensities and quality class E for organic management to represent on-farm quality distribution, B: on-farm yields in Germany. Note the different time scale in both figures. Yearly means for the trials were calculated with model (4). Dashed lines are linear regressions in the years 2001 to 2017. b is the estimated overall yield trend, using model (5) for the trials, with SE in parentheses. ns, and * indicate non-significant and significance at P<0.05, respectively. No regression analysis was performed on the organic on-farm yields, because available datasets differed in their reference bases, and were primarily recorded for economic assessments.

4.4.2 Breeding progress

As we were interested in the contribution of genetic and non-genetic factors underlying the observed yield trends (Figure 14), we performed a mixed model approach to separate genetic and non-genetic trend using trial data from Germany (Table 5). While the non-genetic trend was never significantly different from zero, in the conventional network the genetic trend was significantly positive for both quality classes and in both intensities. In the conventional – reduced intensity, the genetic trend was around 20 to 40 kg/ha/year greater than in the conventional – regular intensity. In contrast, the overall trend was similar under regular and reduced intensity. Within the organic trials, the genetic trend was significantly positive in quality class A, and the estimated genetic trend was of similar magnitude as in the conventional – regular intensity. In quality class E in the organic trials, the genetic trend was slightly positive, but not significantly different from zero. However, when we performed separate analyses within different breeding origins, we found that when including only conventionally bred varieties, the breeding progress was significantly positive, though of smaller magnitude than in quality class A. Within the organically bred varieties, the estimated genetic trend was slightly negative and not significantly different from zero.

Table 5: Genetic, non-genetic, and overall trend estimated by the mixed model approach. Genetic and non-genetic trend were estimated with model (2), and overall trend with model (5). SE is the standard error of b and P the significance of b being different from zero. b and SE are given in kg/ha/year. Within variety of quality class E, separate analyses were performed for conventionally and organically bred varieties. Data from trials in Germany, 2001-2017.

Manage-ment	**Quality class (breeding origin)**	**Number of**			**Genetic trend**		**Non-genetic trend**		**Overall trend**	
		obser-vations	**vari-eties**	**trials**	**b (SE)**	**P**	**b (SE)**	**P**	**b (SE)**	**P**
Conv. – regular	E	4688	30	1321	33.1 (9.3)	<0.001	21.9 (28)	0.435	70.1 (27.4)	0.011
	A	19983	95	1904	33.9 (4.3)	<0.001	17.9 (30.4)	0.555	55.1 (30.6)	0.072
Conv. - reduced	E	3831	29	929	71.9 (11.6)	<0.001	-26.7 (29.8)	0.370	66.4 (28)	0.018
	A	13320	91	1015	55.1 (6)	<0.001	2.2 (29.9)	0.942	59 (29.5)	0.045
Organic	E	4452	64	477	5.4 (5.6)	0.331	10.8 (15.5)	0.487	18.6 (15.1)	0.218
	A	2491	36	473	37.6 (8.6)	<0.001	-0.5 (18)	0.977	32.4 (16.3)	0.046
	E (conv.)	3202	45	477	18.1 (5.3)	<0.001	9 (16.7)	0.592	28.9 (16.3)	0.077
	E (org.)	1250	19	360	-2.0 (9.1)	0.825	34 (17.7)	0.055	31.4 (16.8)	0.061

When estimating the genetic trend within each single trial and subsequently taking the mean over all trials, estimates were smaller for all intensities and quality classes than in the mixed model approach (Table 5). However, this difference varied between management intensities and class from 6 to 33 kg/ha/year, with a mean of 19 kg/ha/year. In all three management intensities, the genetic trend was higher in quality class A than in quality class E. Interestingly, the difference between regular and reduced intensity within conventional management, was around 34 kg/ha/year, which was in a similar range as the respective difference in the mixed model approach.

As the estimated genetic trends could be over-estimated due to variety ageing through breaking of disease-resistances, we evaluated the ageing effect on the yield difference between regular and reduced intensity within conventional management. The overall yield difference between regular and reduced management was slightly increasing in both quality classes (Figure 15). However, the increase might be influenced by the great difference in 2016 with heavy stripe rust infection.

Table 6: Mean genetic trend (b) as estimated by the single trial approach. Within each trial and quality class yields were regressed on the year of release to estimate the genetic trend (b), using only trial x class combinations with minimum 5 varieties. Due to the observed bias from organically bred varieties (Table 5), only conventionally bred varieties were included, as the focus was on the comparison with the mixed model approach. Average SE of b denotes the mean of the SEs estimated or each trial. Data from trials in Germany, 2001-2017.

Manage-ment	**Quality class (breeding origin)**	**Number of trials**	**Mean number of varieties per trial**	**Mean b (kg/ha/year)**	**Average SE of b (kg/ha/year)**
Conv. – regular	E	467	6.4	3.4	69.2
	A	1704	11.9	18.0	62.5
Conv. - reduced	E	421	6.5	38.8	71.8
	A	1015	13.8	45.7	64.2
Organic	A	277	7.3	18.3	50.1
	E (conv.)	432	7.6	12.3	30.3

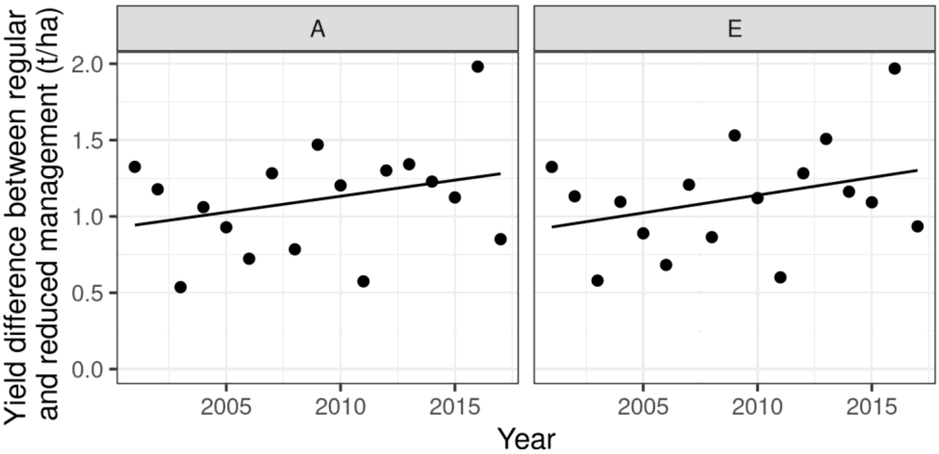


Figure 15: Change of the yield difference between regular and reduced management in the variety recommendation trials. Yield differences were calculated using yearly means for each intensity as calculated with model (4) including only trials for which both intensities were available. Regression slopes for the absolute difference were for quality class A: b=21.0 (SE 17.5) kg/ha/year and for quality class E: b=23.2 (SE 17.2) kg/ha/year. Data from trials in Germany, 2001-2017.

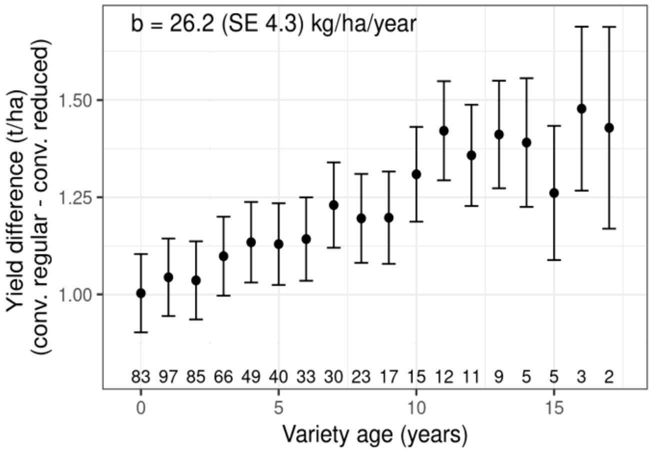


Figure 16: Ageing effect as the increase of the yield difference between regular and reduced intensity with variety age. Means for each year of age were estimated with model (7), and the regression coefficient was estimated with model (6). Error bars are standard errors and numbers indicate the number of varieties for each year of age. Quality class A and E were pooled for this analysis to achieve sufficient numbers of observations. Data from trials in Germany, 2001-2017.

Using the mixed model approach with yield difference as response, the ageing effect was estimated to be at 26 kg/ha/year (Figure 16), indicating that with each year of age the yield difference between regular and reduced intensity increases by 26 kg/ha due to the breaking of resistances. When conducting the same analysis within each single trial, the ageing effect was 33 kg/ha/year for quality class E and 26 kg/ha/year for quality class A, thus being in a very similar range as in the mixed model analysis. As this quantification of the ageing effect, assumes that no ageing occurs under regular management due to full control of diseases, we used the yield differences over all varieties in a trial as a surrogate measure for disease infection. In agreement with the expectation, Figure 17 shows that the observed genetic trend under reduced intensity and the ageing effect within a trial is dependent on the disease pressure. However, also the estimated genetic trend in quality class A under regular intensity seems to be dependent on the disease pressure, thus questioning the assumption of full control of diseases in the regular management. It is interesting to note that relations showed to be mostly linear, although we used a loess smoother, not assuming linearity.

To assess if the estimated genetic trend within a single trial is related to the yield level, we correlated the genetic trend with the mean yield of the given trial (Figure 18). In both quality classes in organic management and in quality class A in the conventional – reduced intensity, the correlation was significantly positive. Within conventional – regular intensity, in quality class A, the estimated genetic trend seemed to be not related to the yield level, while in quality class E there was even a slight negative correlation.

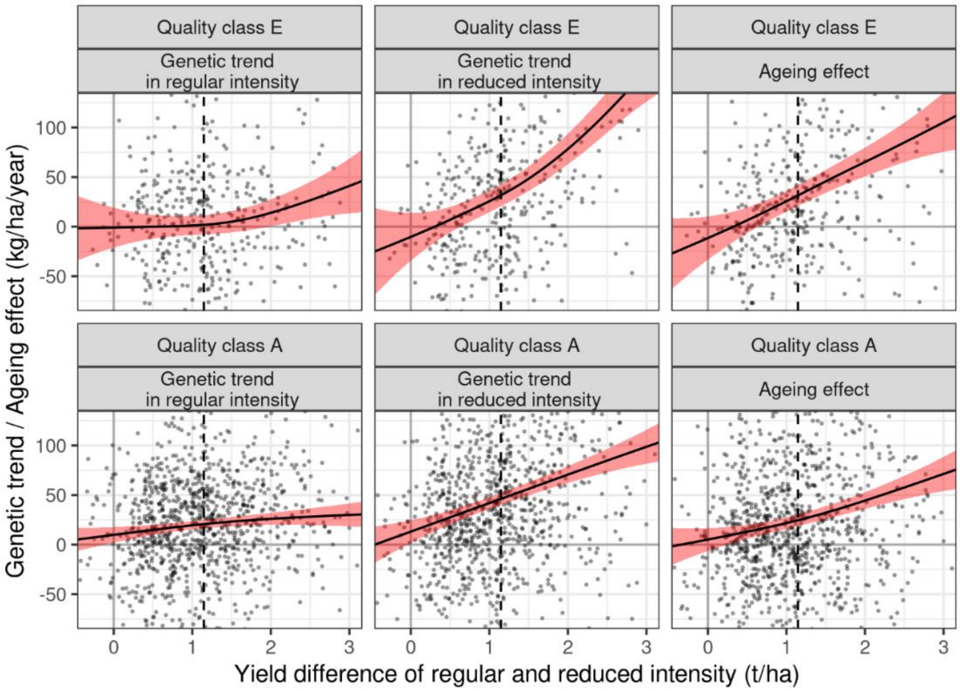


Figure 17: Estimated genetic trend and ageing effect within single trials as influenced by the overall yield difference as a measure for disease pressure. Each dot represents one trial; black line is a loess smoothing function, not assuming linearity; red stripes are the 95% confidence interval bands of prediction, and the vertical dashed line is the mean yield difference over all trials. Data from trials in Germany, 2001-2017.

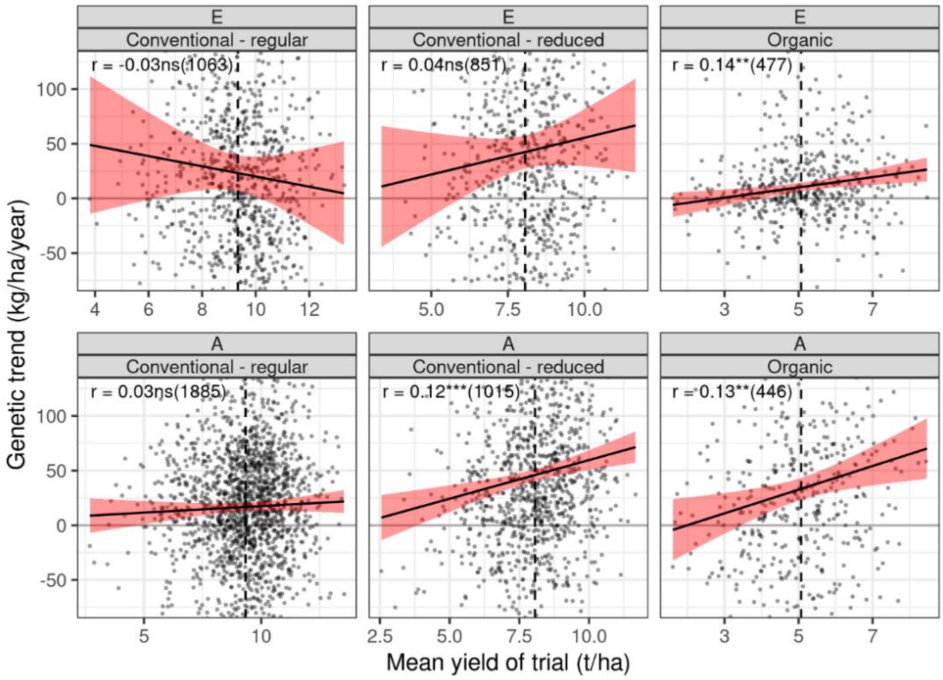


Figure 18: Estimated genetic trend within single trials as influenced by the mean yield of a trial. Each dot represents one trial; black line is a loess smoothing function, not assuming linearity; red stripes are the 95% confidence interval bands, and the vertical dashed line is the mean yield over all trials. Data from trials in Germany, 2001-2017.

4.4.3 Effect of breeding origin

To investigate the effect of breeding origin, i.e. breeding under conventional vs. organic conditions, we plotted the mean yield per variety against their year of release (Figure 19). It became evident that the absent genetic trend in Table 5, when including all varieties of quality class E, in the organic trials was due to the organically bred varieties showing an overall lower yield having been released in the later stage. Thus, including organically bred varieties in the analysis lead to a downwards bias in the regression slope.

In the comparison of the mean values for the different breeding origins, as assessed in the organic trials in Bavaria, organically bred varieties had overall a significantly lower grain yield but higher protein content than conventionally bred varieties (Figure 20). However, the grain N uptake was not significantly different between both groups. Furthermore, baking volume and sedimentation value were slightly higher in organically bred varieties, but this difference was not significant.

To test if the organically bred varieties react differently than conventionally bred varieties in environments differing in their yield potential, we regressed the average performance per breeding origin and quality class against the average grain N uptake per trial, as a measure for yield potential of a trial (Figure 21). Testing the difference in slope through ANCOVA revealed no significant differences ($P>0.05$) between both breeding origins in quality class E and quality class A, indicating that the organically bred varieties reacted similarly and that the higher quality of organically bred varieties is not constrained to lower yielding conditions, under which they might have been selected.

As organically bred varieties were not included in any of the conventional trials, we included data from a field trial where a set of varieties of different breeding origin and quality class were evaluated for grain yield and protein content under conventional and organic management. Grain yield, protein content, and grain N uptake were highly correlated between management systems (Figure 22). Both organically bred varieties showed the highest protein content in both management systems. Interestingly, one organically bred showed the highest value for grain N uptake under conventional management, while the other organically bred variety was among the varieties with the lowest grain N uptake.

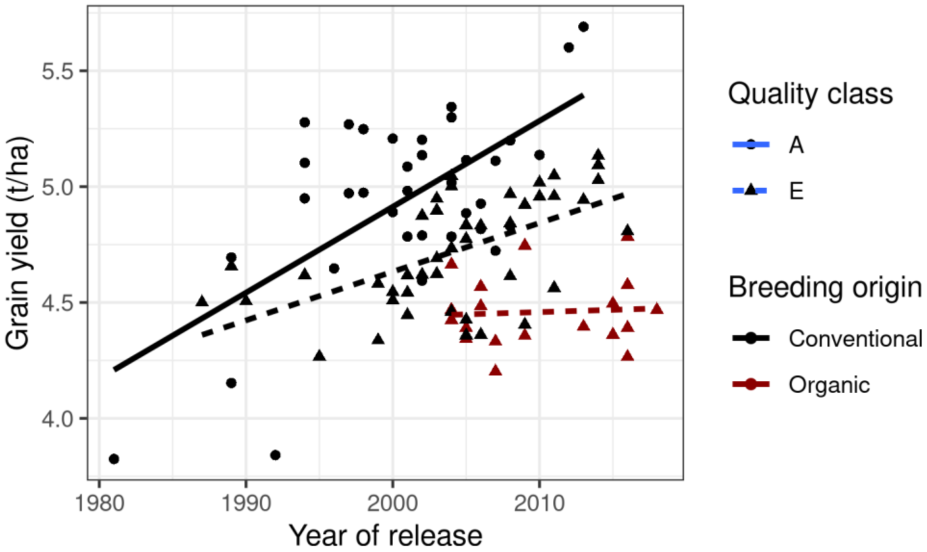


Figure 19: Change in grain yield per quality class and breeding origin in the organic trials from Germany. Each dot represents one variety. Slopes of the regression lines differ slightly from genetic trends in Table 5 due to the different modelling approaches. Data from trials in Germany, 2001-2017.

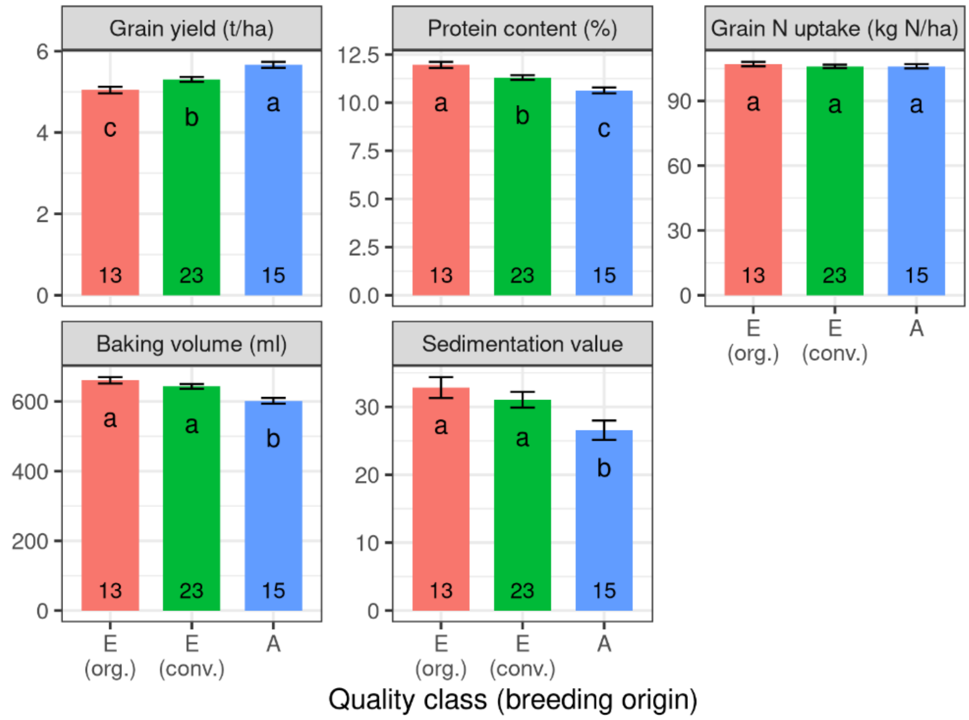


Figure 20: Mean values per quality class and for conventionally and organically bred varieties of quality class E , as assessed in trials in Bavaria under organic management, 2001 - 2019. Letters indicate significant differences at P<0.05, and numbers are the number of varieties.

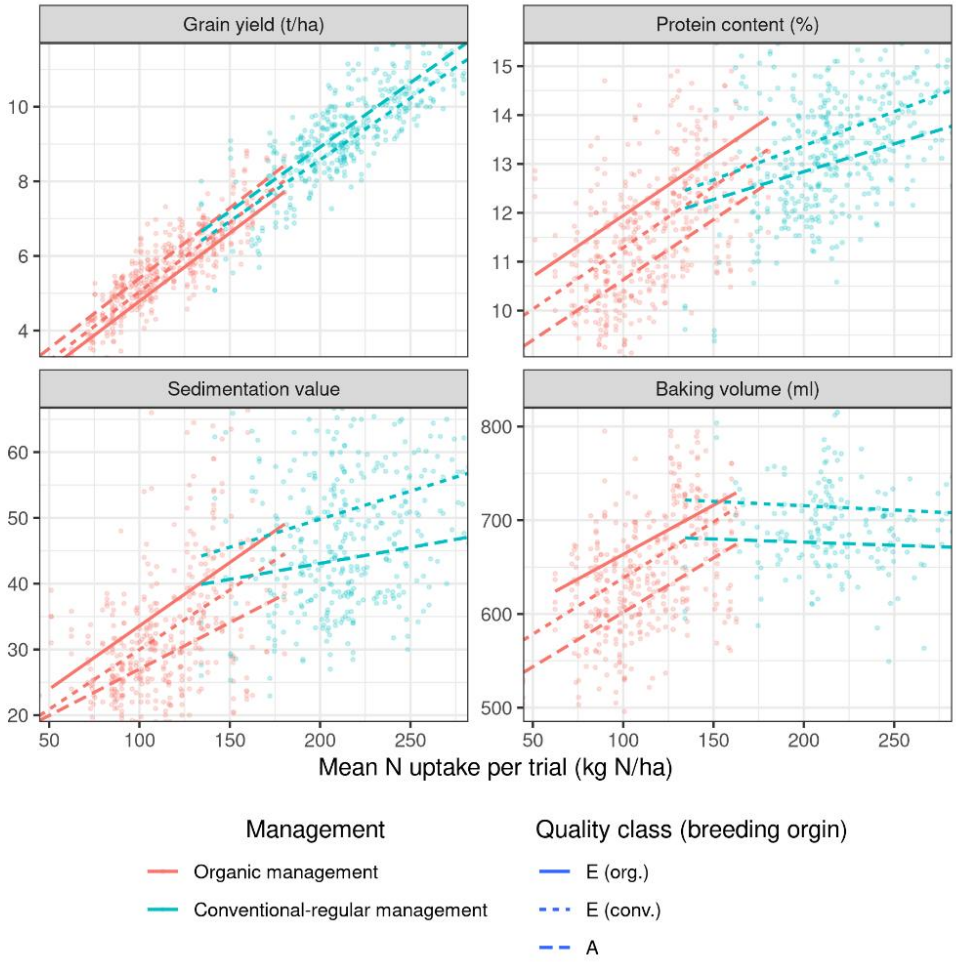


Figure 21: Response of grain yield and quality traits to the yield potential per trial as measured by the mean grain N uptake per trial. Each dot represents the mean value for the respective quality trait per varieties of common quality class plotted against the mean N uptake of all varieties in the given trial.

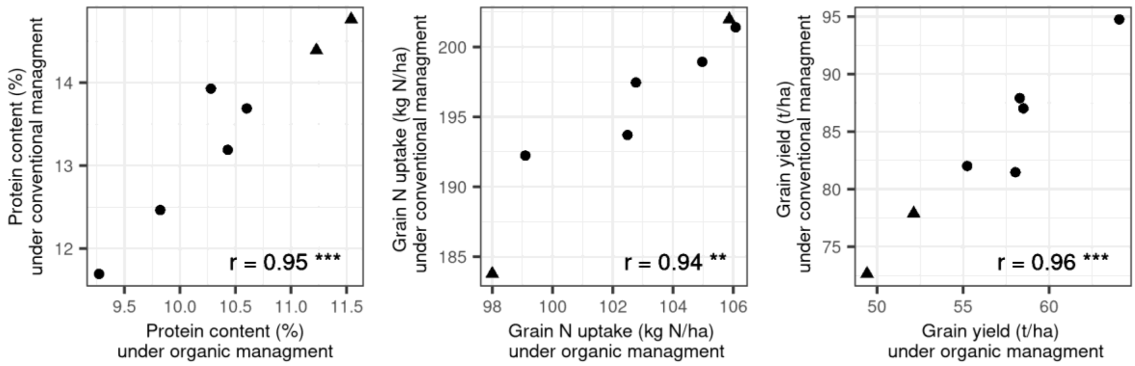


Figure 22: Mean grain yield and protein content under conventional and organic management. Triangles indicate organically bred varieties, dots indicate conventionally bred varieties. r is the Pearson correlation coefficient between the performance under both management systems. Data from a separate field trial (see Material and Methods).

4.4.4 Performance between intensities

To quantify the degree of variety x intensity interaction, which indicates if the performance of varieties differs between conventional and organic management, we assessed variance components. Figure 23 shows the percentage of all variety related terms, which characterize the effect of variety and its interaction. It is thus important to note that the expressed percentages do not represent the contribution to the total variation (see Supplementary Figure 23 for all variance components). The variety main effect and the residual, containing the variety x trial interaction, showed the greatest contribution for all traits. The variance component of the variety x intensity interaction was around 5% for grain yield, grain N uptake, and baking volume. For protein content, it was estimated to be around 1%, and around 20% for sedimentation value. For most traits, the variety x intensity interaction was smaller than the sum of variety x location and variety x year interaction.

As presence of variety x intensity interaction could either indicate a non-linear relation of performance between intensities or crossover interaction, we performed Pearson correlation analysis between variety means estimated within conventional – reduced and within organic management intensity (Table 7). Although the number of varieties, which were assessed in both intensities, was rather small, means between both intensities were highly correlated for all traits. This indicates that varieties showing, e.g., a high grain yield under conventional management also show a high grain yield under organic management.

As a further analysis of the performance between intensities, we calculated Pearson correlations between pairs of single trials (Figure 24). As summary measure over all pairs of trials, we calculated the median. The median correlations were substantially smaller than the correlation of variety means (Table 7). The correlations were lowest for grain N uptake ($r = 0.28$), while they were in a similar range for grain yield, protein content, sedimentation value, and baking volume ($r = 0.59$ to $r = 0.75$). The distribution was rather flat for grain yield and grain uptake, where also a considerable number of negative correlations were observed, while for protein content and sedimentation value most pairs of trials showed very high correlations.

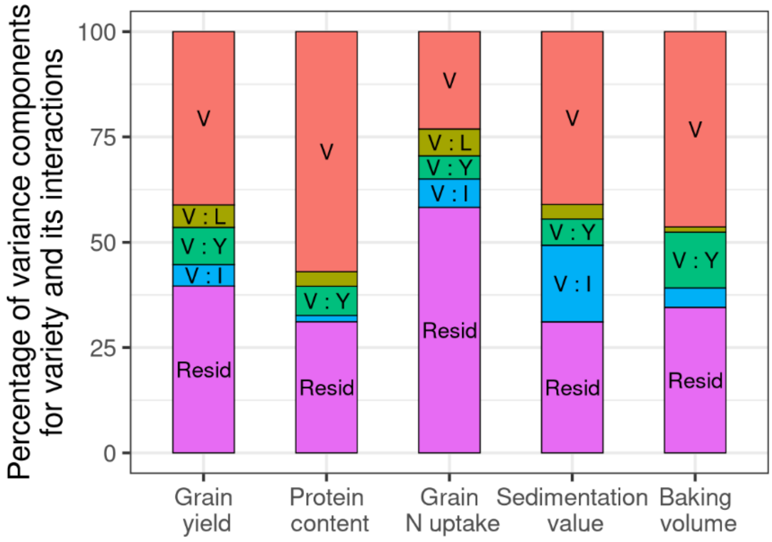


Figure 23: Percentage of variance components for only variety related terms. using model (8), including conventional – regular and organic management and varieties of class E and A. V: variety, L: location, Y: Year, I: intensity, Resid: residual variation containing V:Y:L, V:Y:I and the experimental error. Data from trials in Bavaria, 2001 – 2019.

Table 7: Pearson correlation for grain yield, grain N uptake and quality traits between conventional – regular and organic management. Means were estimated within each intensity using model (3). Data from trials in Bavaria and including varieties of class E and A, 2001 - 2019.

Trait	**Pearson correlation (number of varieties)**
Grain yield	0.86*** (22)
Protein content	0.95*** (20)
Grain N uptake	0.81*** (20)
Sedimentation value	0.91*** (20)
Baking volume	0.89*** (20)

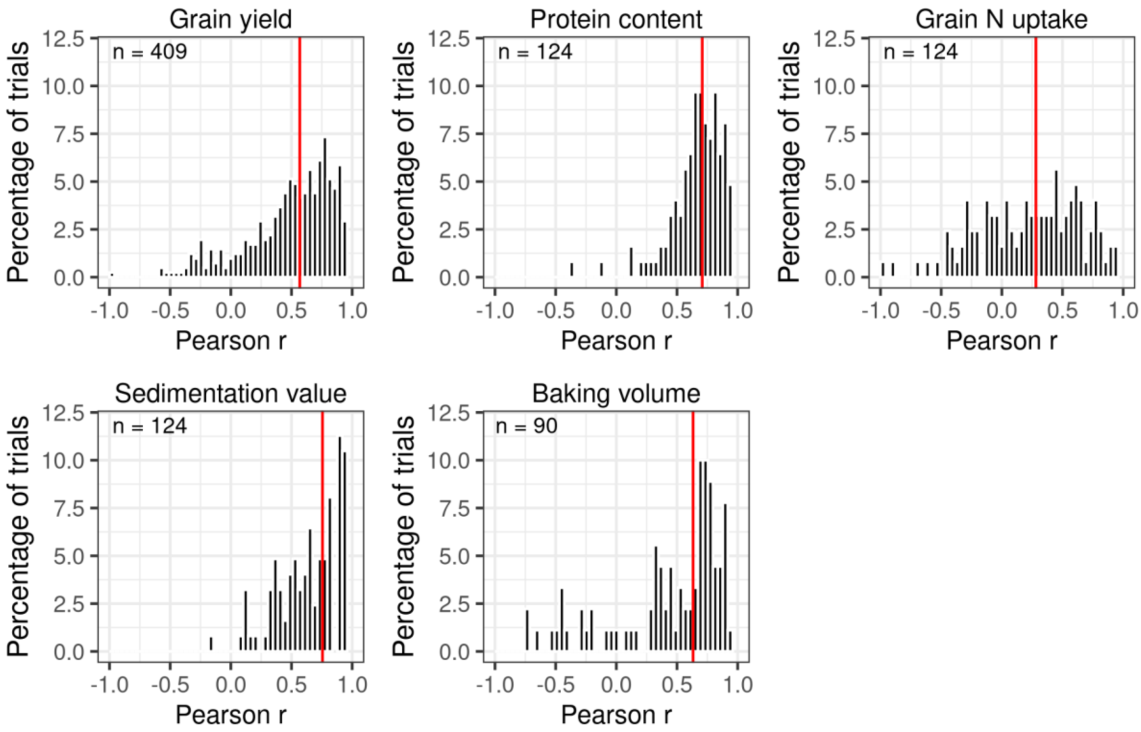


Figure 24: Pearson correlations of variety performance between pairs of conventional and organic trials within the same year , including only pairs, where minimum five varieties were tested in both trial. n indicates the number of pairs of trials, red line is the median. Data from trials in Bavaria, including quality class E and A, 2001 – 2019.

4.5 Discussion

4.5.1 Comparison of trial and on-farm yields

The visual inspection of conventional on-farm yields revealed that yields were steadily increasing up to around 2000, with afterwards slightly levelling off and showing an increased variation (Figure 14). Lin and Huybers (2012) have identified a change point, followed by a yield plateau, for most of the major wheat producing countries roughly around the year 2000. Bönecke et al. (2020), using data from fertilization trials in Germany, found that the levelling-off occurred earlier at sites with lower yield potential. Although this matter was not the focus of our investigation, we find it important to note that the period of our data (2001 – 2017) just starts after the change point around 2000, where additionally the yield variation seemed to increase. Although we identified an overall increase of conventional yields in both trials and on-farm yields in our investigation period of around 50-70 kg/ha/year, this estimate is lower than the increase in on-farm yields from 1960 to 2000, from 3 to 7.5 t/ha, yielding 112.5 kg/ha/year. Similar estimates have been reported from other countries for the time period before the change point (Spink, 2010).

We found a relative yield gap between trial and on-farm yields of 17% for conventional management and 23% for organic management. This yield gap for conventional management did not increase. For organic management, we refrained from this calculation due to the basis of the on-farm data (see Material and Methods). The yield gap might be due edge effects of experimental plots and trial sites being located at sites with higher yield potential (Fischer et al., 2014). Although trial management is supposed to follow standard on-farm management, on-farm management is often constrained by economic reasons, leading to sub-optimal management and thus lower but probably more economic yields, in relation to costs of inputs. If farmers reduce fungicide applications due to economic or environmental reasons, it could thus be questioned, if the regular input level of trials is in general the appropriate control for investigating yield gaps.

For conventional management, Laidig et al. (2014) found a slightly greater yield gap of 24% in 1983, which did minimally decrease to 23% in 2012, in the comparison of VCU trial yields with on-farm yields in Germany. However, they did not specify, which quality classes were included. Thus, most probably the trial yields were based on the full set of quality classes. We only used data from quality class A for the calculation of the mean trial yield as the majority of winter wheat grown across Germany is of quality class A (see also Supplementary Figure 21). However, when using the mean yield of quality class C (9.9 t/ha), the yield gap increases to 23%, which is similar to the yield gap of Laidig et al. (2014), but probably over-estimated, because not all on-farm wheat is quality class C. Because VCU trials are in general similarly as recommendation trials, and often even combined, we do not know of any reasons explaining the difference between their and our different estimate of the yield gap.

Fisher et al. (2014) suggested the analysis of the yield gap between on-farm and trial yields to gain insights into the limitations and progress of on-farm yields. In particular, assuming that trial yields represent the best currently achievable yields (potential yield) with given technology (mainly inputs and varieties), the analysis of the yield gap is meant to identify how well technical improvements are adopted on-farm. They assume a yield gap of 23% as the minimum attainable, assuming prices are reasonable for the farmer. Note that they calculate the yield gap with on-farm yield being the reference, while we use trial yields as the reference. German on-farm yields are therefore within the attainable range, assuming that trial yields represent current potential yield, according to their definition.

In contrast to conventional yields, organic on-farm yields have been constant at around 3.6 t/ha since 1980. However, it needs to be kept in mind, that these data have been mainly collected for economic assessments, and may thus not fully reflect real on-farm yields. The yield gap to the organic trial yields was 23%, and was thus slightly greater, but in a similar range, as observed for the conventional comparison. While the respective yield gap in conventional management might be due to sub-optimal management on-farm (mineral fertilizer and synthetic pesticides), such difference in management might be less the case for the yield gap in the organic comparison, due to the non-usage of such inputs. The direct application of the concept of Fisher et al. (2014) might not be fully valid for organic management, as they assume trial management as making full usage of current available technology and inputs. However, when widening the definition from technology currently available to technology currently allowed within a given management system (which certainly also plays a role in conventional trial management), organic trial yields would represent the potential yield under organic farming restrictions. In this case, organic on-farm yields, being 23% lower, would be within their suggested range of attainable on-farm yields, and thus making the best use of current and allowed technology within the organic farming restrictions.

Properly applying the concept of Fisher et al. (2014) to organic on-farm yields would mean to compare organic on-farm yields to conventional trial yields. In this, case organic on-farm yields are 61% lower than potential yields, suggesting that there is great potential for yield improvement through better usage of available technology.

It is interesting to note, that the organic yield level matches the yield level of conventional yields at around 1965, where average mineral N fertilizer usage per all farmland in Germany was around 50 kg/ha, compared to around 100 kg/ha today (van der Ploeg et al., 2001). Up to 1950, German wheat yields were around 2 t/ha. Acknowledging that a great share of the conventional yield increase was due to mineral N fertilization, and that organic yields are not based on mineral N inputs as it was the case before 1950, organic wheat yields were almost doubled compared to German yields in the times before mineral N usage. This difference might be mostly due to better varieties and better technology for sowing and weed control.

Due to the increase of conventional on-farm yields and the stagnation of organic on-farm yields, the yield difference has dramatically increased. The only study on long-term yield development in organic farming we could find was Kirchmann et al. (2016), who reported a yield decrease in wheat in Sweden from 2003 to 2013, based on official Swedish statistics, though they did no test if the decrease was significant. The increase of on-farm yields in our dataset had a strong effect on the yield difference: While in 1980 organic on-farm yields were 80% of conventional, they were only 49% of conventional yields in the investigated period (2001 – 2017). The same yield difference was observed in the trial yields. In a meta-analysis based on a global dataset, Knapp and van der Heijden (2018) found organic wheat yields to be 73% of conventional yields, while in a similar analysis Ponisio et al. (2015) found this value to be 79%, but for cereals in general.

De Ponti et al. (2012), also using a global meta-analysis, found by regressing the observed yield difference on the conventional yields, that the yield differences were greater when conventional yields were higher. They concluded that greater yield differences were due to more intensive management in conventional farming. This is in accordance with our finding, as the increased yield difference is only due to the increases in the conventional yields. The observed lower yield differences in the meta-analyses might thus be due to less intensive management in conventional management on a global scale.

As organic management differs from conventional management in both fertilization and plant protection, it remains an open question, how much these factors contribute to the yield difference. When taking the yield difference between the reduced and regular management in the trials (ca. 1.3 t/ha) as a measure for the fungicide effect, the fungicide effect accounts for a about a third of the absolute yield difference of organic and regular conventional trial (ca. 3.1 t/ha). Analysing the effects of organic vs. conventional fertilization and crop protection in winter wheat in an orthogonal design, Bilsborow et al. (2013) found both effects on average to be of equal magnitude regarding the absolute yield difference, although effects varied strongly between years. As weeds were controlled chemically in the conventional – reduced management in our dataset, the remaining difference might thus be due weed control. However, it has to be stressed that these effects can strongly interact (Berry et al., 2010) and vary considerably from year to year.

4.5.2 Breeding progress

Using a mixed model approach to separate genetic and non-genetic trend, we found a significant genetic trend in both quality classes and both intensities within conventional management, with the genetic trend being greater under reduced intensity. In organic management, the genetic trend was significant in both quality classes, when organically bred varieties were removed (Table 5). The overall trend, which includes genetic and non-genetic trend, was similar in both conventional intensities. As the non-genetic trend represents the difference between genetic and overall trend, the non-genetic was thus smaller under reduced intensity. Using the ratio between genetic and overall trend, recent studies have estimated the contribution of plant breeding to past yield increases (e.g. Mackay et al. (2011)). Due to great variation of overall yields in the period of our dataset (Figure 14), we found that the estimates for the overall trend were strongly influenced by the chosen period. E.g., when we removed the last three years from the dataset (as this analysis would have been conducted in 2015), the overall trend was estimated to be around 85 kg/ha/year in both conventional intensities and quality classes, while it was only around 25 kg/ha/year when the first three years were removed (Supplementary Table 11). We do thus refrain from making such estimates for the contribution of plant breeding to the overall yield increase. In contrast, estimates for the genetic trend showed to be robust towards this subsetting of data.

In both the mixed model and single trial approaches, estimates for genetic trend where overall around 30 kg/ha/year greater under regular than reduced intensity in conventional management. The same difference was also observed by Laidig et al. (2014) on VCU trial data from Germany. However, as we could not observe any decrease in the overall difference between regular and reduced intensity (Figure 15), the greater genetic trend in reduced intensity does not indicate a breeding progress towards better resistances. This is also supported by the similar overall trend in both intensities (Table 5). Interestingly, the ageing effect, as measured by the increase of the yield difference between regular and reduced management with variety age (Figure 16) was found to be of similar size as the difference in breeding progress between both intensities. This is in agreement with Laidig et al. (2014), who found the same pattern of the ageing effect to be similar to the difference in genetic trend between regular and reduced intensity for several cereal species. Mackay et al. (2011), using a mixed model approach and winter wheat variety trial data from UK, found a genetic trend of 74 kg/ha/year in treated trials and 104 kg/ha/year in untreated trials. While this difference is again 30 kg/ha/year, they estimated the ageing effect to be 80 kg/ha/year. The fact, that we found the ageing effect to be of similar size in the single trial analysis, indicates that this estimate is not biased trough the mixed model approach. However, in the single trial approaches with historic sets of German varieties tested by Voss-Fels et al. (2019) and Ahlemeyer and Friedt (2012), the difference in genetic trend between treated and untreated was considerably lower (17 kg/ha/year in the former and 4 kg/ha/year in the latter, means over quality class E and A). This smaller difference might be due to the historic variety set, including up to 50 years of age. In summary, the generally observed greater genetic trend under reduced intensity suggests, that the observed genetic trend in reduced intensity in both the mixed model approach and single trial analysis is over-estimated due to variety ageing.

An over-estimation of the genetic trend implies an under-estimation of the non-genetic trend, as non-genetic trend is estimated as the increase in overall yield, which is not due genetic trend.

Although we found that the exact estimate of the overall trend is dependent on the period of the dataset it was always significant. The overall increase is most likely due to an increase in N fertilization in three federal states (see Supplementary Figure 22).

We observed that the estimated genetic trends were higher in the mixed model approach than in the single trial approach. Piepho et al. (2014) also noted this issue by comparing their mixed model estimates with other single trial approach studies, and speculate that the trend observed in the single trial approach might reflect “purely genetic trend”, while they termed the trend from the mixed model approach “apparent genetic trend”, which over-estimates genetic trend and under-estimates agronomic trend. Furthermore, they concluded that even the genetic trend estimated for regular intensity is biased due to variety ageing, suggesting that diseases are not fully controlled under regular intensity. This is supported by our finding, that the observed genetic trend under regular intensity for quality class A in the single trial approach, showed to be dependent on the disease pressure at the given trial (Figure 17). These findings on the over-estimation of genetic trend due to the presence of variety ageing in both the mixed model approach and the single trial approach might indicate that previous studies have been prone to over-estimations of the genetic trend and consequently under-estimation of the non-genetic trend. In effect, this implies that the contribution of plant breeding to past yield increases might have been over-estimated (Laidig et al., 2014).

While we found a mean genetic trend in the single trial approach of 4 kg/ha/year and 18 kg/ha/year under regular intensity, for quality class E and A, respectively, these estimates were greater in other studies, where a historic set of varieties was assessed. E.g., Ahlemeyer and Friedt (2012) reported 26 kg/ha/year for both quality classes, and Voss-Fels et al. (2019) reported 20 kg/ha/year and 28 kg/ha/year for class E and A, respectively. Our very low estimation for quality class E might be due to the rather low average number of varieties per trial (only 6.4 varieties per trial). The general observation of greater breeding progress for quality class A than quality class E is probably due to more breeding activity due to the greater growing area and thus seed demand for quality class A, and due to the higher quality requirements for quality class E, limiting the progress in grain yield.

A significant and positive genetic trend for organic management was observed in both the mixed model and single trial approach, when organically bred varieties were removed. However, as in the organic management no chemical plant protection is applied, breaking of diseases resistances occurs. Thus, the observed genetic trend under organic management is prone to over-estimation as well as in the reduced intensity within conventional management. However, due to the intertwined nature of the year of release, the year of observation, and variety age, it is not possible to assess the effect of variety ageing within a given intensity (Piepho et al., 2014). Because the yield difference between regular and reduced management (sometimes also termed fungicide effect) is related to the yield level itself, the absolute rate of variety ageing will most likely be smaller than under conventional management due to the lower yield level in organic management. However, the estimated genetic effect in the single trial approach, which is less prone to the mixed model over-estimation, was 12 kg/ha/year and 18 kg/ha/year for quality class E and A, respectively. If the ageing effect would be proportional to the yield level, the expected ageing effect under organic management, having about 50% of conventional yield and an ageing effect under conventional management with 30 kg/ha/year, would be around 15 kg/ha/year. In this case, the effective genetic trend would be zero for organic agriculture.

While this finding suggests that plant breeding has been successful in at least counteracting the breaking of disease-resistances, as organic trial and on-farm yields have been maintained, the question remains why the “purely genetic trend” has been realised under organic management. Our finding in the single trial analysis that the genetic trend was higher in trials with a higher yield level (Figure 18), could suggest that breeding progress is more efficient at higher yield potential, e.g. more available nitrogen. However, this relation has to be interpreted with care, as the ageing effect could again be higher at higher yield level, and thus the observed higher genetic trend could again be over-estimated. This probable relation would be supported by the strong relation to the yield level under reduced intensity but no relation under regular intensity in quality class A in the same analysis.

The finding that breeding progress is higher under conventional than under organic management was also reported by Hildermann et al. (2009), growing a historic set of varieties under conventional and organic farming. The results of Rueda-Ayala et al. (2018) suggest that applied N strongly influences the observed genetic trend. Although not noted by the authors, it is evident from their Fig. 1 that N application is the strongest factor, in comparison with P, K, and Ca application, determining the observed genetic trend.

The calculation of the N budget could give a further hint for the absent breeding progress under organic management: In the trials from Bavaria, the average grain N uptake was 210 kg N/ha under the conventional – regular management and 110 kg N/ha under organic management. On average 180 kg N/ha were applied in the conventional trials. Soil mineral N in spring was on average 70 kg N/ha in the conventional trials and 80 kg N/ha in the organic trials, most probably due to grass-clover as preceding crop. With additional N from grass-clover of around 20 kg N/ha and around 10 kg N/ha due to long-term organic fertilization, the grain N uptake just equals the input. For conventional management, adding 10 kg/ha from preceding crops (sometimes sugar beet and rapeseed), grain N uptake is 50 kg N/ha smaller than the input. This calculation indicates that organic wheat cropping is already very efficient in comparison with conventional farming and wheat varieties already use most of the available N, which in turn implies that there is little potential for yield increase and thus for breeding progress if not available N is increased.

4.5.3 Effect of breeding origin

In the analysis of the breeding progress, we observed that the genetic trend under organic management in quality class E was only significant after removing organically bred varieties (Table 5). We further found that this effect was due to the organically bred varieties showing a particularly lower yield and thus leading to a downward bias in the regression (Figure 19). However, we observed that besides the lower yield, the organically bred varieties had a significantly higher protein content, higher baking volume, and higher sedimentation value, while grain N uptake was similar (Figure 20). Regressing yield and baking quality traits on the average N uptake per treatment as a measure to characterize the yield potential revealed that the organically bred varieties reacted similar to increased yield potential as the conventionally bred varieties (Figure 21).

Figure 21shows that the assessed quality parameters were not only lower under organic management but also more dependent on the overall yield potential, as indicated by the overall steeper slopes for organic management than for conventional management. This clearly indicates that producing wheat of good baking quality under organic management can be challenging.

The reduced yield of the organically bred varieties is not too surprising due to the genetically determined negative correlation between grain yield and protein content (Oury and Godin, 2007; Sherman et al., 2014). This negative correlation is also indicated by the equal grain N uptake between the different breeding origins.

Baking quality of winter wheat has been shown to be partly determined by available N (Brabant and Levy Häner, 2016). As N is often the most limiting factor for yield under organic production, this might also influence baking quality traits, as there is a general relationship to protein content (Knapp et al., 2017).

The overall higher quality of the organically bred varieties can therefore favourably contribute to the production of wheat with higher quality under organic management. The mean protein content of the organically bred varieties was 0.7 percentage points higher than that of the conventionally bred varieties of quality class E. Interestingly, this is the same difference as between quality class E and A of conventionally bred varieties, under both conventional and organic management. For the sedimentation value and baking volume, this difference was slightly smaller and not significant. With regard to the differences between quality classes, the organically bred varieties could thus be visualized as representing a higher quality class, above quality class E.

The similar response to yield potential (Figure 21) and the observation that the protein content of the organically bred varieties was also higher under conventional management (Figure 22) suggests that such high-quality varieties could have also been selected under conventional management. However, such higher quality might not be beneficial in variety registration trials, as in the German registration system, candidate lines need to fulfil the class-specific threshold values for each quality trait (BSA, 2019). Due to the general negative correlation to grain yield, having higher trait values than the threshold, could thus reduce yield and, in turn, reduce the VCU. Furthermore, as baking quality is to a certain degree sufficiently good under conventional management, there is less demand for wheat of higher quality, and thus breeding varieties of higher quality is not attractive for breeders. However, protein content has been removed from the list of traits for the classification of varieties in Germany in 2019 (BSA, 2019). Due to the negative correlation of protein content to grain yield and because it has been shown that protein content is only partly correlated (Knapp et al., 2017), this will allow for the breeding of varieties with equal baking quality but higher yield.

Some of the organically bred varieties have been registered through a separate registration process for varieties meant to be suited for organic management, which was established in 2012 (BSA, 2019). In this registration process, candidate lines are tested only under organic management, and additionally evaluated for ground cover at the tillering stage and the formation of biomass at shooting, as indirect measure for weed suppression. Therefore, some of the varieties might have had a sufficient VCU rather due to their weed suppression properties than due to their higher quality, as this is not directly rewarded by the German registration scheme.

However, the question of this investigation was actually if direct selection is more beneficial than indirect selection, i.e. if organically bred varieties perform better under organic management than conventionally bred varieties. The general observation was that organically bred varieties show an overall higher quality and lower yield. Thus, the effect of organic breeding can be rather visualized as a shift towards quality on the cost of yield. While this allows production of quality wheat under organic management, the effect might be less due to selection under organic management, than rather due to selection for higher quality. Due to the general trade-off between protein content and yield, the only trait among the investigated ones that remains for comparison is grain N uptake, where both breeding origins were similar. Furthermore, the small variety x intensity interaction (Figure 23) and high correlations of variety means between intensities (Table 7) suggest that due to the similar performance between intensities indirect selection, i.e. selection under conventional management for organic management, is possible.

It has been argued that selection under organic conditions might produce varieties with better weed suppression or better disease-resistances, because both are less controlled and breeders would select more intensively for such traits (Murphy et al., 2007). However, the similarity in grain N uptake does not imply that such selection might have been more beneficial for the overall performance under organic management.

Although our results suggest that direct selection under organic management has not been more beneficial besides the increase in quality, it has to be stressed in order to be fair that the comparison was based on a small set of organically bred varieties from three small breeding companies. The breeding activities of these companies are not comparable to the breeding activities for conventional management. Thus, for a fair and scientific evaluation of the question of direct vs. indirect selection a selection experiment should be conducted where selection is carried with the same intensity under both management systems, and selected lines are subsequently evaluated in the opponent system (Ceccarelli, 1996).

4.5.4 Performance between intensities

In the analysis of variance components, we found that the effect variety x intensity interaction is considerably small and always smaller than the variety main effect (Figure 23). When correlating the variety means of the varieties that were tested under both conventional and organic management, correlations were high for all of the assessed traits (Figure 18).

Our finding is supported by several studies who did not observe a significant genotype x system interaction or found correlated performance when comparing varieties between conventional and organic management (Przystalski et al., 2008; Hildermann et al., 2009), under different rates of nitrogen (Ortiz-Monasterio et al., 1997; Guarda et al., 2004), or between untreated and treated trials (Mackay et al., 2011).

However, calculating correlations between pairs of trials revealed that the performance in the different intensities were not consistently correlated and overall lower than the correlations of variety means (Figure 22). It should be noted that the correlations also capture the variety x location interactions, and thus lower or negative correlations do not necessarily indicate variety x intensity interaction. Variety means represent the overall performance within a given system, and thus the absent interaction and strong correlation indicates organic and conventional management are overall not different in their variety interaction. In contrast, the correlations between trials confirm that there is an overall correlation but also indicate that this overall similarity in performance does not translate into the correlation between single trials. This implies, that while the variety means characterize e.g. the overall yield potential, the variation of yields of different varieties within a given trial is only partly determined by their yield potential. The yield variation within a trial is additionally determined by the variation in specific traits corresponding to the prevailing growing conditions.

The variation for certain traits only shows up under certain management conditions, e.g., variation in a particular disease resistance can only be evaluated, if this disease occurs and is not treated. Likewise, weed suppression can only be evaluated if weeds are present and not treated. This would imply that rather trait-specific experiments should be conducted to evaluate the performance of specific traits and information should be combined for variety registration and recommendation. As the general yield potential is correlated between intensities, yield is probably best evaluated under best growing conditions. In contrast, lodging tolerance would probably be best evaluated under ample N supply to provoke lodging, but untreated with growth regulators.

On-farm variety choice is then determined by the prevailing importance of each trait. This applies equally and in a complex manner to all management intensities. For growing conditions, where diseases, weeds, and lodging can be successfully controlled, only the yield potential might be important. In a conventional production scheme, where chemical weed treatment is avoided for environmental reasons, weed suppression might be more important due to higher N availability, than at certain organic growing conditions where weeds are less problematic due to certain preceding crops, successful mechanical weed control and lower N availability. In summary, regarding variety recommendation for organic farming, trait-specific experiments like artificial and controlled weed or disease infestation could probably be more informative than yield evaluation under organic management. Furthermore, this applies equally for different conventional production schemes and calls for intelligent and specific usage of all available information from and for different management intensities.

4.6 Conclusion

We conclude by addressing the three central questions of this study.

What was the contribution of plant breeding to yield development in conventional and organic management?

The main finding was that variety ageing due to the breaking of disease-resistances could lead to substantial over-estimation of the observed genetic trend. For conventional management, this implies that previous estimates of the contribution of plant breeding to the past yield increases might have been generally over-estimated. However, even when correcting for this over-estimation, the genetic-trend was still significant and yield increases were at least partly due plant breeding. As we observed that the estimated overall trend was strongly influenced by the great yearly variation of the overall yields, we do refrain from making estimates on the contribution of plant breeding to yield increase. In contrast, under organic management the observed genetic trend was in the range of the estimated ageing effect, which is the rate of yield decrease due to breaking of disease-resistances. This implies that plant breeding has been successful to maintain the yield level and thus to counteract the breaking of disease-resistances. The observed genetic trend being in the range of the ageing effect further implies that the effective genetic trend tends towards zero. Combined with the observation of no overall yield increase, this suggests that plant breeding has not contributed any yield increase under organic management. However, we would like to stress, that the non-observed effective breeding progress under organic management does not imply that organic management has generally not benefitted from plant breeding or that older varieties perform equally. Counteracting the breaking of disease-resistances is a major effort in plant breeding and showed to be successful also for organic management. While the non-observed effective genetic trend might be due to little direct selection and particular breeding efforts for organic management, this is probably more due to the limiting growing conditions, particularly N limitation, under organic management. When relating grain N uptake to available N, it could be argued that organic wheat cropping is already highly efficient in comparison with conventional management, which in turn implies little potential of further yield improvement through plant breeding. Therefore, to achieve further yield improvements and thus to make use of the increased yield potential of modern varieties, the amount of available N would need to be increased under organic management.

Has breeding under organic management been more beneficial than breeding under conventional management for organic management?

We found that organically bred varieties showed overall lower yield, better baking quality, as assessed by protein content, sedimentation value and baking quality, but grain N uptake was similar. Because producing wheat with sufficient baking quality is challenging under N limited conditions, varieties with higher quality characteristics can be one way to achieve a good quality. In this regard, organic breeding has thus been successful. However, due to the negative genetic correlation between baking quality and grain yield, this has been achieved through lower yields, and with no improvement in grain N uptake. We found, that there is little to no environmental interaction for quality traits, i.e. that the higher quality of organically bred varieties was independent of the yield potential, and that the higher protein content was achieved similarly under conventional management. This finding suggests that in this case the effect of breeding origin is less related to the issue of direct vs. indirect selection, but more to selection for higher quality characteristics, and that selecting for higher quality characteristics could have been equally successful under conventional management. The similarity in grain N uptake suggests that organic breeding has been equally successful in overall performance, which indicates no advantage of direct vs. indirect selection. However, due to the observed difference in quality and due to the limited number of organically bred varieties combined with overall less breeding activity, our results are limited to the current market situation. To better answer this question, scientific experiments would be needed, where consistent selection within different intensities is applied with equal effort, followed by subsequent cross-evaluation of the selected lines.

Is separate testing under conventional and organic management necessary for variety registration and recommendation?

Our results that yield and quality characteristics are strongly correlated between different management intensities could suggest that evaluation of these traits would be sufficient under one management intensity, preferably the one producing the highest heritability and thus the best information about variety performance. However, as each management intensity has its own constraints, different traits are important for each management intensity. We therefore argue that the general yield potential should probably be assessed under conditions where the performance of varieties is as little as possible influenced by any stress factors, which would be conventional and best treated experiments. Additionally, for recommendation for organic farming or any other management intensity, intensity specific traits should be evaluated in trait-specific experiments.

Acknowledgements

Financial support by the German Ministry for Education and Research (BMBF) through the INSUSFAR project is thankfully acknowledged. We are very thankful to Lorenz Hartl from the Bavarian State Institute for Agriculture, Freising, Germany, for providing the data from the conventional trials in Bavaria.

5 General discussion

Yield stability

The analyses of yield stability in the meta-analysis (chapter A) and in the long-term trial (chapter B) showed very similar results. Absolute stability, as measured by the variance, was similar between conventional and organic management. Relative stability, as measured by the coefficient of variation, was more stable in conventional management. In both analyses, it was found that the difference in relative stability was due to the difference in yield level. In the meta-analysis, it was concluded that increased fertilization or better fertilization management might contribute to improve relative stability. However, in the long-term trial regular and half fertilization showed to be similar in relative stability at both conventional and organic management. This suggests that the difference in relative stability might only partly be due to the lower fertilization level and thus also due to the difference in plant protection. Due to the avoidance of synthetic pesticides, organic management is more prone to yield reduction or loss due to disease and pest infection, which will thus increase the yearly variation of yields (Bilsborrow et al., 2013). This observation is supported by the analysis of the yearly variation in the long-term trial, where the increased yield differences revealed to be due to lower organic yields and not due to higher conventional yields. It is however interesting to note that in both the meta-analysis and the long-term trial, the absolute stability was similar in both systems. The Taylor power law predicts a linear relationship between the natural logarithm of the mean and the natural logarithm of variance (Taylor, 1961), and has to been shown to hold true for crop yields as well (Döring et al., 2015). However, this relationship was not observed in both analyses.

Yield development

The development of yields was investigated with the main question, if long-term organic management can sustain its yield level. In the analysis on the long-term trial (chapter B), a significant difference in yield trend between the conventional and organic treatments was found only for winter wheat, where the conventional treatments showed a higher yield increase than the organic treatments. However, a particular cause for this difference could not be identified, but might be due to the fertilization intensity. With regard to the yield trends being similar for the remaining crops (potatoes, grass-clover, maize, soybean), it could thus be concluded that 40 years of organic management did not result in a long-term yield decline, when compared with conventional management. Similarly, the organic yields of winter wheat in Germany showed no decrease in the variety recommendation trials (chapter C). As the dataset of organic on-farm yields was based on different reference bases across years, no regression analysis was performed. However, visual inspection did not suggest any overall decrease in yields.

The yield increase of winter wheat yields in the long-term trial was considerably below yield increases, found in Germany (chapter C) or other countries (Mackay et al., 2011). In contrast, winter wheat yields of conventional management in Germany showed an increasing trend, while organic yields remained constant, both on-farm and in the recommendation trials. However, it has to be noted that the increase in conventional yields has slowed down since around the year 2000, which has also been found by other authors and for other major wheat producing countries (Spink, 2010; Lin and Huybers, 2012; Bönecke et al., 2020). Thus, while the stagnation under organic management was observed in the long-term trial in Switzerland and in on-farm and trial data from Germany, the yield increase under conventional management, which was observed in Germany, was not present in the long-term trial. Average organic yields of winter wheat were the same in the Swiss long-term trial and German variety trials (4.6 t/ha at 14% moisture, note that in chapter B yields were given as dry matter). In contrast, average conventional yields in the long-term trial were 5.8 t/ha (at 14% moisture) and 9.3 t/ha in the variety trials. In Switzerland, conventional on-farm yields of winter wheat have been found to stagnate since around 1985 (Erdin, 2018; Herrera et al., 2020) at around 6 t/ha. Interestingly, conventional on-farm yields in Germany around 1985 were also around 6 t/ha. Thus, while German on-farm yields have still increased steadily until around the year 2000, Swiss on-farm yields have stagnated at a level, which was at that time similar in Germany. As the Swiss long-term trial was initiated in 1978 and conventional treatments were managed following common Swiss fertilization standard, the stagnation of the trial yields is in agreement with the stagnation of on-farm yields. The lower yields in Switzerland than in Germany might be most probably due to overall lower N fertilization (around 100 kg N/ha in the long-term trial vs. around 180 kg N/ha in the variety trials, which should reflect on-farm applications since both aim to reflect local agronomic practice). Although the yield increase in German on-farm yields of winter wheat was found to slow down after around the year 2000, there was still an increase in both on-farm and variety trials, though smaller, compared with other countries, where yields have fully levelled off (Lin and Huybers, 2012). Data on the amount of applied N fertilizer showed that in some federal states, N rates still increased in the last 20 years. Therefore, the increase in-farm yields might also be due increasing N fertilization However, it was found that due to the high variation in yields, estimates of the yield increase were quite sensible to the chosen period and thus the observation on still increasing yields needs to be treated with care.

Breeding progress under organic agriculture

Significant positive genetic trend was found under both conventional and organic management (chapter C). However, the estimates might be over-estimated due to variety aging because of breaking of disease-resistances, particularly if diseases are not treated. As no synthetic disease control is applied in organic management, the effective genetic trend might tend towards zero. The observation that yields were not declining, despite the constant breaking of disease-resistances, therefore suggests that plant breeding has successfully contributed to the maintenance of yield under organic management. However, increase in yield might be difficult as wheat cropping is already highly efficient with regard to the available N. This might indicate that not much yield increase can be expected from plant breeding if nutrient levels, in particular N, are the limiting production factor.

It was furthermore found that organically bred varieties show better baking quality, but due to negative correlation of grain yield and protein content lower yields. Judged by the similarity in grain N uptake no advantage over conventionally bred varieties could be identified. As the organically bred varieties reacted similarly as conventionally bred varieties to the overall yield potential and their protein content showed to be very high under conventional as well, it might thus have been also possible to select for such high quality under conventional conditions. This is supported by the overall high correlations of variety means between systems. However, as overall variety means simply reflect overall yield potential, specific traits might be important for specific growing conditions, like weed suppression or disease resistance.

With regard to the small difference between organic and conventional management, plant breeding might thus not contribute to reducing the yield difference. Furthermore, the relation of baking quality to available N, in combination with the demand for good baking quality, presents a further limitation to yield improvement under organic management.

Reduced yields of organic management

Although the difference in mean yield was not among the central research questions of this thesis, the reduced yields of organic management emerged in each chapter. In particular, relative stability was found to be lower in organic management due to the lower yields and the smaller or even absent breeding progress under organic management was most likely also due to yield limitation by limited N availability. The overall yield difference was around 15-20% in both the meta-analysis (chapter A) and the long-term trial (chapter B). However, for winter wheat in Germany, the yield difference war around 50% on-farm and in the variety recommendation trials (chapter

C). The yield differences for wheat were 27% in the meta-analysis and 22% in the long-term trial. Interestingly the yield difference is even smaller in the long-term trial from Switzerland compared to the global estimate from the meta-analysis. This reduced difference could either be due to higher yields than global organic wheat yields or be due to lower than global conventional yields. However, as it was noted above that Swiss conventional wheat yields are considerably low due to a low N application rate, this small difference might rather be due to low conventional wheat yields.

In both the meta-analysis and long-term trial, the yield differences were generally related to the difference in nitrogen (N) input, while in the long-term trial a particularly strong relation to the share of N applied in mineral form (also within organic fertilizers) was observed in winter wheat and potatoes. However, because in the long-term trial, application rates across treatments were correlated with the intensity of plant protection, a clear separation of the contribution of both effects was not possible. Based on the yield level of the untreated conventional intensity in the variety recommendation trials, the yield difference between conventional and organic management in winter wheat in Germany might be due to plant protection for one third and due to the difference in fertilization for two-thirds. However, it has to be noted that fertilization and plant protection do not add up additively but rather interact (Berry et al., 2010). However, although it might be difficult to exactly determine the contribution of fertilization and plant protection, organic yields are strongly limited by available N (Berry et al., 2002).

As mineral fertilization is not allowed in organic farming, it will thus be more important to utilize available N in the best efficient way possible. This will include minimizing losses from the animal to the field (Erickson and Klopfenstein, 2010). Furthermore, additional N inputs from biogas residues, increased nutrient recycling in society, or mineral nitrogen fertilizers from renewable source could be utilized (Röös et al., 2018; Serdjuk et al., 2018). However, Muller et al. (2017) predicted, based on a food model, that adequate N supply might be challenging at 100% conversion to organic production.

Concluding remarks on organic agriculture

The finding that yields trends were not different in the long-term trial between conventional and organic management (chapter B) and organic wheat yields in Germany did not show a decline (chapter C) might suggest that long-term organic management did sustain its yield level, and could therefore be judged as sustainable. However, the lower relative stability was shown to be due to lower yields (chapter A and B) and breeding progress might be difficult under nutrient limiting conditions and it might thus be difficult to increase yields under organic farming (chapter C).

The lower yield level (15-20% on a global scale compared to 50% in wheat in Germany) implies that additional land for agricultural land would be needed for the same demand for food (Muller et al., 2017). Furthermore, the commonly estimated yield ratios of direct comparisons of yields per crop neglect that additional area is needed for legumes for biological nitrogen fixation, often preceding stable crops like cereals (Connor, 2018), indicating that an effective yield ratio taking account of this additional land might be greater. Additional land would in turn lead to reduced natural habitats and thus to reduced biodiversity in natural ecosystems (Guzmán Casado and González de Molina, 2009), counteracting the positive effects of organic farming on biodiversity within the agricultural area (Tuck et al., 2014).

The positive effects of converting land from conventional to organic agricultural production on several indicators like, e.g. biodiversity, reduced nitrate leaching, better soil conditions, or reduced greenhouse gas emissions, have been shown in numerous studies (Kirchmann and Bergström, 2001; Mäder et al., 2002; Bengtsson et al., 2005). This effect is not too surprising, as organic production in a way represents simply an extensification, where inputs with potential negative effects are reduced or omitted, and thus overall negative effects should be reduced as well. However, when calculations are either corrected for the additional need for land or calculated per amount of produced food, these effects are either not present or even reversed (Kirchmann and Bergström, 2001; Mondelaers et al., 2009; Leifeld and Fuhrer, 2010).

However, potential drawbacks and possible negative effects of organic farming should not indicate that high-input conventional farming is advantageous because of its higher productivity. This higher productivity is largely facilitated by mineral fertilizers and synthetic pesticides which can have dramatic effects on the environment (McLaughlin and Mineau, 1995; Johnson et al., 2007; Moss, 2008; Foley et al., 2011; Tilman et al., 2011; Godfray and Garnett, 2014). Besides the importance of changing diets towards less animal-based food in order to reduce overall demand (Godfray and Garnett, 2014), the challenge of the future will thus be to identify combinations of different production factors that allow for sustainable and effective agricultural production. Therefore, future agricultural systems should lie somewhere in between (Reganold and Wachter, 2016), which implies for research to rather identify potential risks of agricultural production systems and develop more effective production methods and technology than fighting trench warfare for the superiority of either one or the other production system, i.e. not like it was done in this thesis.

Publication list and author contributions

Chapter A was published as:

Samuel Knapp and Marcel van der Heijden. 2018. A global meta-analysis of yield stability in organic and conservation agriculture. Nature Communications 9(1): 3632. doi: 10.1038/s41467-018-05956-1.

Marcel van der Heijden (MvdH), Plant-Soil-Interactions, Research Division Agroecology and Environment, Agroscope, Zurich, Switzerland.

SK and MvdH conceived and designed the study. SK performed the analysis. SK and MvdH. wrote the manuscript.

Chapter B has been submitted:

Lucie Gunst (LG), Jochen Mayer (JM), Water Protection and Substance Flows, Research Division Agroecology and Environment, Agroscope, Zurich, Switzerland; Paul Mäder (PM), Soil Science Department, Research Institute of Organic Agriculture FiBL, Frick, Switzerland.

SK performed statistical analysis and wrote the manuscript, LG managed data collection and quality control, JM supervised the statistical analysis, JM and PM conducted the experiment, provided data, and contributed to the writing of the manuscript.

Oral contributions related to this thesis:

Samuel Knapp, J Peter Baresel, Urs Schmidhalter: *Comparing the breeding progress in winter wheat between organic and conventional management in Deutschland*. 70. Jahrestagung der Vereinigung der Pflanzenzüchter und Saatgutkaufleute Österreichs, Raumberg-Gumpenstein, Austria; 11/2019

Samuel Knapp, J Peter Baresel, Urs Schmidhalter: *Zuchtfortschritt in Winterweizen - Unterschiede zwischen konventioneller und ökologischer Bewirtschaftung und zwischen verschiedenen Anbaugebieten in Deutschland*. Jahrestagung der Gesellschaft für Pflanzenbauwissenschaften, Kiel, Deutschland; 09/2018

Samuel Knapp, Marcel van der Heijden: *Comparing the Temporal Stability of Organic and Conventional Cropping Systems - A Meta-Analysis*. 14th ESA Congress, Edinburgh, Scotland; 09/2016

Bibliography

Ahlemeyer, J., and W. Friedt. 2012. Züchterischer Ertragsfortschritt ungebrochen. Praxisnah 3(Juli 2012): 2–4.

Ahn, S., and J.A. Fessler. 2003. Standard errors of mean, variance, and standard deviation estimators. EECS Department, The University of Michigan: 1–2.

Ammann, K. 2008. Integrated farming: why organic farmers should use transgenic crops. New Biotechnology 25(2): 101–107. doi: 10.1016/j.nbt.2008.08.012.

Andrew, C.S. 1960. The effect of phosphorus, potassium, and calcium on the growth, chemical composition, and symptoms of deficiency of white clover in a subtropical environment. Aust. J. Agric. Res. 11(2): 149–161. doi: 10.1071/ar9600149.

Annicchiarico, P. 2000. Genotype x environment interactions: Challenges and opportunities for plant breeding and cultivar recommendations. FAO Plant Production and Protection Papers 174: 126.

Atlin, G.N., and K.J. Frey. 1989. Breeding crop varieties for low-input agriculture. Am J Alt Ag 4(2): 53–58. doi: 10.1017/S0889189300002721.

Bailey-Serres, J., J.E. Parker, E.A. Ainsworth, G.E.D. Oldroyd, and J.I. Schroeder. 2019. Genetic strategies for improving crop yields. Nature 575(7781): 109–118. doi: 10.1038/s41586-019-1679-0.

Baresel, J.P. 2006. Weizenzüchtung für den Ökologischen Landbau. Verlag Dr. Köster, Berlin.

Bates, D., M. Mächler, B. Bolker, and S. Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software 67(1): 1–48. doi: 10.18637/jss.v067.i01.

Becker, H.C., and J. Léon. 1988. Stability Analysis in Plant Breeding. Plant Breeding 101(1): 1–23. doi: 10.1111/j.1439-0523.1988.tb00261.x.

Bélanger, G., J.R. Walsh, J.E. Richards, P.H. Milburn, and N. Ziadi. 2000. Comparison of three statistical models describing potato yield response to nitrogen fertilizer. Agronomy Journal 92(5): 902–908.

Bender, S.F., C. Wagg, and M.G.A. van der Heijden. 2016. An Underground Revolution: Biodiversity and Soil Ecological Engineering for Agricultural Sustainability. Trends in Ecology & Evolution 31(6): 440–452. doi: 10.1016/j.tree.2016.02.016.

Bengtsson, J., J. Ahnström, and A.-C. Weibull. 2005. The effects of organic agriculture on biodiversity and abundance: a meta-analysis. Journal of Applied Ecology 42(2): 261–269. doi: 10.1111/j.1365-2664.2005.01005.x.

Berry, P.M., D.R. Kindred, J.E. Olesen, L.N. Jorgensen, and N.D. Paveley. 2010. Quantifying the effect of interactions between disease control, nitrogen supply and land use change on the greenhouse gas emissions associated with wheat production. Plant Pathology 59(4): 753–763. doi: 10.1111/j.1365-3059.2010.02276.x.

Berry, P.M., R. Sylvester-Bradley, L. Philipps, D.J. Hatch, S.P. Cuttle, et al. 2002. Is the productivity of organic farms restricted by the supply of available nitrogen? Soil Use and Management 18(s1): 248–255. doi: 10.1111/j.1475-2743.2002.tb00266.x.

Bilsborrow, P., J. Cooper, C. Tétard-Jones, D. Średnicka-Tober, M. Barański, et al. 2013. The effect of organic and conventional management on the yield and quality of wheat grown in a long-term field trial. European Journal of Agronomy 51: 71–80. doi: 10.1016/j.eja.2013.06.003.

Bingham, J. 1981. The achievements of conventional plant breeding. Philosophical Transactions of the Royal Society of London. B, Biological Sciences 292(1062): 441–455. doi: 10.1098/rstb.1981.0039.

Birkhofer, K., T.M. Bezemer, J. Bloem, M. Bonkowski, S. Christensen, et al. 2008. Long-term organic farming fosters below and aboveground biota: Implications for soil quality, biological control and productivity. Soil Biology and Biochemistry 40(9): 2297–2308. doi: 10.1016/j.soilbio.2008.05.007.

BLE. 2018. Erträge im biologischen und konventionellen Landbau. www.oekolandbau.de/handel/marktinformationen/der-biomarkt/marktberichte/ertraege-im-biologischen-und-konventionellen-landbau/ (accessed 7 January 2020).

BMEL. 2005. Agrarberichte 1981-2005. Bundesministerium für Ernährung und Landwirtschaft, Bonn, Germany.

BMEL. 2012. Die wirtschaftliche Lage der Landwirtschaftlichen Betriebe 2011/12: Buchführungsergebnisse Landwirtschaft. Bundesministerium für Ernährung und Landwirtschaft, Bonn, Germany.

BMEL. 2017. Statistisches Jahrbuch über Ernährung, Landwirtschaft und Forsten der Bundesrepublik Deutschland. Bundesministerium für Ernährung und Landwirtschaft, Bonn, Germany.

BMEL. 2019. Die wirtschaftliche Lage der Landwirtschaftlichen Betriebe 2018/19: Buchführungsergebnisse Landwirtschaft. Bundesministerium für Ernährung und Landwirtschaft, Bonn, Germany.

Bönecke, E., L. Breitsameter, N. Brüggemann, T.-W. Chen, T. Feike, et al. 2020. Decoupling of impact factors reveals the response of German winter wheat yields to climatic changes. Global Change Biology 26(6): 3601–3626. doi: 10.1111/gcb.15073.

Brabant, C., and L. Levy Häner. 2016. Einfluss der Stickstoffdüngung und ihrer Aufteilung auf die Backqualität von Weizen. Agrarforschung Schweiz 7(2): 88–97.

Brancourt-Hulmel, M., G. Doussinault, C. Lecomte, P. Bérard, B. Le Buanec, et al. 2003. Genetic Improvement of Agronomic Traits of Winter Wheat Cultivars Released in France from 1946 to 1992. Crop Science 43(1): 37. doi: 10.2135/cropsci2003.3700.

Briones, M.J.I., and O. Schmidt. 2017. Conventional tillage decreases the abundance and biomass of earthworms and alters their community structure in a global meta-analysis. Global Change Biology 23(10): 4396–4419. doi: 10.1111/gcb.13744.

Bronick, C.J., and R. Lal. 2005. Soil structure and management: a review. Geoderma 124(1): 3–22. doi: 10.1016/j.geoderma.2004.03.005.

Brooker, R.W., A.E. Bennett, W.-F. Cong, T.J. Daniell, T.S. George, et al. 2015. Improving intercropping: a synthesis of research in agronomy, plant physiology and ecology. New Phytologist 206(1): 107–117. doi: 10.1111/nph.13132.

BSA. 2019. Beschreibende Sortenliste: Getreide, Mais, Öl- und Faserpflanzen, Leguminosen, Rüben, Zwischenfrüchte. Bundessortenamt, Hannover, Germany.

Canevara, M.G., M. Romani, M. Corbellini, M. Perenzin, and B. Borghi. 1994. Evolutionary trends in morphological, physiological, agronomical and qualitative traits of Triticum aestivum L. cultivars bred in Italy since 1900. European Journal of Agronomy 3(3): 175–185. doi: 10.1016/S1161-0301(14)80081-6.

Ceccarelli, S. 1996. Adaptation to low/high input cultivation. Euphytica 92(1): 203–214. doi: 10.1007/BF00022846.

Connor, D.J. 2018. Organic agriculture and food security: A decade of unreason finally implodes. Field Crops Research 225: 128–129. doi: 10.1016/j.fcr.2018.06.008.

Covarrubias-Pazaran, G. 2016. Genome-Assisted Prediction of Quantitative Traits Using the R Package sommer. PLOS ONE 11(6): e0156744. doi: 10.1371/journal.pone.0156744.

Deguines, N., C. Jono, M. Baude, M. Henry, R. Julliard, et al. 2014. Large-scale trade-off between agricultural intensification and crop pollination services. Frontiers in Ecology and the Environment 12(4): 212–217. doi: 10.1890/130054.

Denison, R.F., D.C. Bryant, and T.E. Kearney. 2004. Crop yields over the first nine years of LTRAS, a long-term comparison of field crop systems in a Mediterranean climate. Field Crops Research 86(2–3): 267–277. doi: 10.1016/j.fcr.2003.08.014.

Döring, T.F., S. Knapp, and J.E. Cohen. 2015. Taylor's power law and the stability of crop yields. Field Crops Research 183: 294–302. doi: 10.1016/j.fcr.2015.08.005.

Erdin, D. 2018. Ertragsentwicklung im Schweizer Ackerbau. AGRISTAT - Statistisches Monatsheft (6): 6–11.

Erickson, G., and T. Klopfenstein. 2010. Nutritional and management methods to decrease nitrogen losses from beef feedlots. J Anim Sci 88(suppl_13): E172–E180. doi: 10.2527/jas.2009-2358.

Evenson, R.E., and D. Gollin. 2003. Assessing the Impact of the Green Revolution, 1960 to 2000. Science 300(5620): 758–762. doi: 10.1126/science.1078710.

FAOSTAT. 2019. Crop production data. Food and Agriculture Organization of the United Nations (FAO).

Feil, B. 1992. Breeding Progress in Small Grain Cereals — A Comparison of Old and Modern Cultivars. Plant Breeding 108(1): 1–11. doi: 10.1111/j.1439-0523.1992.tb00093.x.

Feltz, C.J., and G.E. Miller. 1996. An asymptotic test for the equality of coefficients of variation from k populations. Statistics in Medicine 15(6): 647–658. doi: 10.1002/(SICI)1097-0258(19960330)15:6<647::AID-SIM184>3.0.CO;2-P.

Fischer, R.A., D. Byerlee, and G.O. Edmeades. 2014. Crop yields and global food security: will yield increase continue to feed the world? Australian Centre for International Agricultural Research, Canberra.

Foley, J.A., N. Ramankutty, K.A. Brauman, E.S. Cassidy, J.S. Gerber, et al. 2011. Solutions for a cultivated planet. Nature 478(7369): 337–342. doi: 10.1038/nature10452.

Fortune, S., J. Hollies, and E.A. Stockdale. 2004. Effects of different potassium fertilizers suitable for use in organic farming systems on grass/clover yields and nutrient offtakes and interactions with nitrogen supply. Soil Use and Management 20(4): 403–409. doi: 10.1111/j.1475-2743.2004.tb00389.x.

Gałecki, A., and T. Burzykowski. 2013. Linear Mixed-Effects Models Using R: A Step-by-Step Approach. Springer-Verlag, New York.

Galloway, J.N., J.D. Aber, J.W. Erisman, S.P. Seitzinger, R.W. Howarth, et al. 2003. The Nitrogen Cascade. BioScience 53(4): 341–356. doi: 10.1641/0006-3568(2003)053[0341:TNC]2.0.CO;2.

Gattinger, A., A. Muller, M. Haeni, C. Skinner, A. Fliessbach, et al. 2012. Enhanced top soil carbon stocks under organic farming. PNAS 109(44): 18226–18231. doi: 10.1073/pnas.1209429109.

Gilmour, A.R., R. Thompson, and B.R. Cullis. 1995. Average Information REML: An Efficient Algorithm for Variance Parameter Estimation in Linear Mixed Models. Biometrics 51(4): 1440–1450. doi: 10.2307/2533274.

Godfray, H.C.J., J.R. Beddington, I.R. Crute, L. Haddad, D. Lawrence, et al. 2010. Food Security: The Challenge of Feeding 9 Billion People. Science 327(5967): 812–818. doi: 10.1126/science.1185383.

Godfray, H.C.J., and T. Garnett. 2014. Food security and sustainable intensification. Philosophical Transactions of the Royal Society B: Biological Sciences 369(1639): 20120273. doi: 10.1098/rstb.2012.0273.

Guarda, G., S. Padovan, and G. Delogu. 2004. Grain yield, nitrogen-use efficiency and baking quality of old and modern Italian bread-wheat cultivars grown at different nitrogen levels. European Journal of Agronomy 21(2): 181–192. doi: 10.1016/j.eja.2003.08.001.

Gurevitch, J., and L.V. Hedges. 2001. Meta-analysis: combining the results of independent experiments. In: Scheiner, S.M. and Gurevitch, J., editors, Design and analysis of ecological experiments. 2nd ed. Oxford University Press, USA, New York, USA. p. 347–369

Gutser, R., Th. Ebertseder, A. Weber, M. Schraml, and U. Schmidhalter. 2005. Short-term and residual availability of nitrogen after long-term application of organic fertilizers on arable land. Z. Pflanzenernähr. Bodenk. 168(4): 439–446. doi: 10.1002/jpln.200520510.

Guzmán Casado, G.I., and M. González de Molina. 2009. Preindustrial agriculture versus organic agriculture: The land cost of sustainability. Land Use Policy 26(2): 502–510. doi: 10.1016/j.landusepol.2008.07.004.

Hautier, Y., D. Tilman, F. Isbell, E.W. Seabloom, E.T. Borer, et al. 2015. Anthropogenic environmental changes affect ecosystem stability via biodiversity. Science 348(6232): 336–340. doi: 10.1126/science.aaa1788.

Hedges, L.V., J. Gurevitch, and P.S. Curtis. 1999. The Meta-Analysis of Response Ratios in Experimental Ecology. Ecology 80(4): 1150–1156. doi: 10.1890/0012-9658(1999)080[1150:TMAORR]2.0.CO;2.

Hejcman, M., E. Kunzová, and P. Šrek. 2012. Sustainability of winter wheat production over 50 years of crop rotation and N, P and K fertilizer application on illimerized luvisol in the Czech Republic. Field Crops Research 139: 30–38. doi: 10.1016/j.fcr.2012.10.005.

Herrera, J.M., L. Levy Häner, F. Mascher, J. Hiltbrunner, D. Fossati, et al. 2020. Lessons From 20 Years of Studies of Wheat Genotypes in Multiple Environments and Under Contrasting Production Systems. Front. Plant Sci. 10. doi: 10.3389/fpls.2019.01745.

Hildermann, I., A. Thommen, D. Dubois, T. Boller, A. Wiemken, et al. 2009. Yield and baking quality of winter wheat cultivars in different farming systems of the DOK long-term trial. Journal of the Science of Food and Agriculture 89(14): 2477–2491. doi: 10.1002/jsfa.3750.

Hobbs, P.R., K. Sayre, and R. Gupta. 2008. The role of conservation agriculture in sustainable agriculture. Philosophical Transactions of the Royal Society B: Biological Sciences 363(1491): 543–555. doi: 10.1098/rstb.2007.2169.

Hothorn, T., F. Bretz, and P. Westfall. 2008. Simultaneous inference in general parametric models. Biometrical journal 50(3): 346–363.

Howden, S.M., J.-F. Soussana, F.N. Tubiello, N. Chhetri, M. Dunlop, et al. 2007. Adapting agriculture to climate change. PNAS 104(50): 19691–19696. doi: 10.1073/pnas.0701890104.

Johnson, J.M.-F., A.J. Franzluebbers, S.L. Weyers, and D.C. Reicosky. 2007. Agricultural opportunities to mitigate greenhouse gas emissions. Environmental Pollution 150(1): 107–124. doi: 10.1016/j.envpol.2007.06.030.

Jones, H., S. Clarke, Z. Haigh, H. Pearce, and M. Wolfe. 2010. The Effect of the Year of Wheat Variety Release on Productivity and Stability of Performance on Two Organic and Two Non-Organic Farms. The Journal of Agricultural Science 148(03): 303–317. doi: 10.1017/S0021859610000146.

Kenward, M.G., and J.H. Roger. 1997. Small Sample Inference for Fixed Effects from Restricted Maximum Likelihood. Biometrics 53(3): 983. doi: 10.2307/2533558.

Kirchmann, H., and L. Bergström. 2001. Do Organic Farming Practices Reduce Nitrate Leaching? Communications in Soil Science and Plant Analysis 32(7–8): 997–1028. doi: 10.1081/CSS-100104101.

Kirchmann, H., L. Bergström, T. Kätterer, and R. Andersson. 2016. Dreams of Organic Farming - Facts and Myths. Fri Tanke förlag, Stockholm, Sweden.

Knapp, S., C. Brabant, M. Oberforster, H. Grausgruber, and J. Hiltbrunner. 2017. Quality traits in winter wheat: comparison of stability parameters and correlations between traits regarding their stability. Journal of Cereal Science. doi: 10.1016/j.jcs.2017.08.011.

Knapp, S., and M.G.A. van der Heijden. 2018. A global meta-analysis of yield stability in organic and conservation agriculture. Nature Communications 9(1): 3632. doi: 10.1038/s41467-018-05956-1.

Köhl, L., F. Oehl, and M.G.A. van der Heijden. 2014. Agricultural practices indirectly influence plant productivity and ecosystem services through effects on soil biota. Ecological Applications 24(7): 1842–1853. doi: 10.1890/13-1821.1.

Krause, H.-M., A. Fliessbach, J. Mayer, and P. Mäder. 2020. Implementation and management of the DOK long-term system comparison trial. In: Bhullar, G.S. and Riar, A., editors, Long-Term Farming Systems Research. Academic Press. p. 37–51

Krebs, C.J. 1998. Ecological Methodology. Subsequent. Addison Wesley, Menlo Park, California.

Laidig, F., T. Drobek, and U. Meyer. 2008. Genotypic and environmental variability of yield for cultivars from 30 different crops in German official variety trials. Plant Breeding 127(6): 541–547. doi: 10.1111/j.1439-0523.2008.01564.x.

Laidig, F., H.-P. Piepho, T. Drobek, and U. Meyer. 2014. Genetic and non-genetic long-term trends of 12 different crops in German official variety performance trials and on-farm yield trends. Theor Appl Genet 127(12): 2599–2617. doi: 10.1007/s00122-014-2402-z.

Lajeunesse, M.J. 2011. On the meta-analysis of response ratios for studies with correlated and multi-group designs. Ecology 92(11): 2049–2055. doi: 10.1890/11-0423.1.

Lajeunesse, M.J. 2016. Facilitating systematic reviews, data extraction and meta-analysis with the metagear package for r. Methods in Ecology and Evolution 7(3): 323–330. doi: 10.1111/2041-210X.12472.

Lammerts van Bueren, E.T., S.S. Jones, L. Tamm, K.M. Murphy, J.R. Myers, et al. 2011. The need to breed crop varieties suitable for organic farming, using wheat, tomato and broccoli as examples: A review. NJAS - Wageningen Journal of Life Sciences 58(3): 193–205. doi: 10.1016/j.njas.2010.04.001.

Leifeld, J., and J. Fuhrer. 2010. Organic Farming and Soil Carbon Sequestration: What Do We Really Know About the Benefits? AMBIO 39(8): 585–599. doi: 10.1007/s13280-010-0082-8.

Lenth, R. 2018. emmeans: Estimated Marginal Means, aka Least-Squares Means.

Lichtenberg, E.M., C.M. Kennedy, C. Kremen, P. Batáry, F. Berendse, et al. 2017. A global synthesis of the effects of diversified farming systems on arthropod diversity within fields and across agricultural landscapes. Global Change Biology 23(11): 4946–4957. doi: 10.1111/gcb.13714.

Lin, C.-S., M.R. Binns, and L.P. Lefkovitch. 1986. Stability analysis: where do we stand? Crop Science 26(5): 894–900.

Lin, M., and P. Huybers. 2012. Reckoning wheat yield trends. Environ. Res. Lett. 7(2): 024016. doi: 10.1088/1748-9326/7/2/024016.

Litrico, I., and C. Violle. 2015. Diversity in Plant Breeding: A New Conceptual Framework. Trends in Plant Science 20(10): 604–613. doi: 10.1016/j.tplants.2015.07.007.

Mackay, I., A. Horwell, J. Garner, J. White, J. McKee, et al. 2011. Reanalyses of the historical series of UK variety trials to quantify the contributions of genetic and environmental factors to trends and variability in yield over time. Theor Appl Genet 122(1): 225–238. doi: 10.1007/s00122-010-1438-y.

Mäder, P., A. Fliessbach, D. Dubois, L. Gunst, P. Fried, et al. 2002. Soil fertility and biodiversity in organic farming. Science 296(5573): 1694–1697. doi: 10.1126/science.1071148.

Martínez, I., A. Chervet, P. Weisskopf, W.G. Sturny, A. Etana, et al. 2016. Two decades of no-till in the Oberacker long-term field experiment: Part I. Crop yield, soil organic carbon and nutrient distribution in the soil profile. Soil and Tillage Research 163: 141–151. doi: 10.1016/j.still.2016.05.021.

Mayer, J., L. Gunst, P. Mäder, M.-F. Samson, M. Carcea, et al. 2015. Productivity, quality and sustainability of winter wheat under long-term conventional and organic management in Switzerland. European Journal of Agronomy 65: 27–39. doi: 10.1016/j.eja.2015.01.002.

Mayer, J., and P. Mäder. 2016. Langzeitversuche - Eine Analyse der Ertragsentwicklung. In: Freyer, B., editor, Ökologischer Landbau: Grundlagen, Wissensstand und Herausforderungen. UTB, Stuttgart, Germany. p. 421–455

McLaughlin, A., and P. Mineau. 1995. The impact of agricultural practices on biodiversity. Agriculture, Ecosystems & Environment 55(3): 201–212. doi: 10.1016/0167-8809(95)00609-V.

Mikó, P., F. Löschenberger, J. Hiltbrunner, R. Aebi, M. Megyeri, et al. 2014. Comparison of bread wheat varieties with different breeding origin under organic and low input management. Euphytica 199(1): 69–80. doi: 10.1007/s10681-014-1171-8.

Mondelaers, K., J. Aertsens, and G.V. Huylenbroeck. 2009. A meta-analysis of the differences in environmental impacts between organic and conventional farming. British Food Journal 111(10): 1098–1119. doi: 10.1108/00070700910992925.

Moss, B. 2008. Water pollution by agriculture. Philosophical Transactions of the Royal Society B: Biological Sciences 363(1491): 659–666. doi: 10.1098/rstb.2007.2176.

Muller, A., C. Schader, N.E.-H. Scialabba, J. Brüggemann, A. Isensee, et al. 2017. Strategies for feeding the world more sustainably with organic agriculture. Nature Communications 8(1): 1290. doi: 10.1038/s41467-017-01410-w.

Murphy, K.M., K.G. Campbell, S.R. Lyon, and S.S. Jones. 2007. Evidence of varietal adaptation to organic farming systems. Field Crops Research 102(3): 172–177. doi: 10.1016/j.fcr.2007.03.011.

Nakagawa, S., R. Poulin, K. Mengersen, K. Reinhold, L. Engqvist, et al. 2015. Meta-analysis of variation: ecological and evolutionary applications and beyond. Methods in Ecology and Evolution 6(2): 143–152. doi: 10.1111/2041-210X.12309.

Newton, A.C., D.C. Guy, A.G. Bengough, D.C. Gordon, B.M. McKenzie, et al. 2012. Soil tillage effects on the efficacy of cultivars and their mixtures in winter barley. Field Crops Research 128: 91–100. doi: 10.1016/j.fcr.2011.12.004.

Niggli, U., H. Willer, and B. Baker. 2016. A Global Vision and Strategy for Organic Farming Research. TIPI - Technology Innovation Platform of IFOAM, Frick, Switzerland.

Ortiz-Monasterio, J.I., K.D. Sayre, S. Rajaram, and M. McMahon. 1997. Genetic Progress in Wheat Yield and Nitrogen Use Efficiency under Four Nitrogen Rates. Crop Science 37(3): cropsci1997.0011183X003700030033x. doi: 10.2135/cropsci1997.0011183X003700030033x.

Oury, F.-X., and C. Godin. 2007. Yield and grain protein concentration in bread wheat: how to use the negative relationship between the two characters to identify favourable genotypes? Euphytica 157(1–2): 45–57. doi: 10.1007/s10681-007-9395-5.

Paradelo, R., I. Virto, and C. Chenu. 2015. Net effect of liming on soil organic carbon stocks: A review. Agriculture, Ecosystems & Environment 202: 98–107. doi: 10.1016/j.agee.2015.01.005.

Parker, W.E., and J.J. Howard. 2001. The biology and management of wireworms (Agriotes spp.) on potato with particular reference to the U.K. Agricultural and Forest Entomology 3(2): 85–98. doi: 10.1046/j.1461-9563.2001.00094.x.

Phalan, B., A. Balmford, R.E. Green, and J.P.W. Scharlemann. 2011. Minimising the harm to biodiversity of producing more food globally. Food Policy 36: S62–S71. doi: 10.1016/j.foodpol.2010.11.008.

Piepho, H.-P. 1999. Stability Analysis Using the SAS System. Agronomy Journal 91(1): 154. doi: 10.2134/agronj1999.00021962009100010024x.

Piepho, H.-P. 2004. An Algorithm for a Letter-Based Representation of All-Pairwise Comparisons. Journal of Computational and Graphical Statistics 13(2): 456–466. doi: 10.1198/1061860043515.

Piepho, H.P., A. Büchse, and K. Emrich. 2003. A Hitchhiker's Guide to Mixed Models for Randomized Experiments. Journal of Agronomy and Crop Science 189(5): 310–322. doi: 10.1046/j.1439-037X.2003.00049.x.

Piepho, H.-P., F. Laidig, T. Drobek, and U. Meyer. 2014. Dissecting genetic and non-genetic sources of long-term yield trend in German official variety trials. Theor Appl Genet 127(5): 1009–1018. doi: 10.1007/s00122-014-2275-1.

Pittelkow, C.M., X. Liang, B.A. Linquist, K.J. van Groenigen, J. Lee, et al. 2015. Productivity limits and potentials of the principles of conservation agriculture. Nature 517: 365–368.

van der Ploeg, R.R., R.P. Schweigert, and R.J. Bachmann. 2001. Use and Misuse of Nitrogen in Agriculture: The German Story. The Scientific World JOURNAL 1: 737–744. doi: 10.1100/tsw.2001.263.

Pocknee, S., and M.E. Sumner. 1997. Cation and Nitrogen Contents of Organic Matter Determine Its Soil Liming Potential. Soil Science Society of America Journal 61(1): 86–92. doi: 10.2136/sssaj1997.03615995006100010014x.

Ponisio, L.C., L.K. M'Gonigle, K.C. Mace, J. Palomino, P. de Valpine, et al. 2015. Diversification practices reduce organic to conventional yield gap. Proceedings of the Royal Society of London B: Biological Sciences 282(1799): 20141396. doi: 10.1098/rspb.2014.1396.

de Ponti, T., B. Rijk, and M.K. van Ittersum. 2012. The crop yield gap between organic and conventional agriculture. Agricultural Systems 108: 1–9. doi: 10.1016/j.agsy.2011.12.004.

Prechsl, U.E., R. Wittwer, M.G.A. van der Heijden, G. Lüscher, P. Jeanneret, et al. 2017. Assessing the environmental impacts of cropping systems and cover crops: Life cycle assessment of FAST, a long-term arable farming field experiment. Agricultural Systems 157: 39–50. doi: 10.1016/j.agsy.2017.06.011.

Przystalski, M., A. Osman, E. Thiemt, B. Rolland, L. Ericson, et al. 2008. Comparing the performance of cereal varieties in organic and non-organic cropping systems in different European countries. Euphytica 163(3): 417–433. doi: 10.1007/s10681-008-9715-4.

R Core Team. 2014. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

R Core Team. 2018. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Rao, C.R., S.K. Mitra, and A. Matthai, editors. 1966. Formulae and Tables for Statistical Work. Statistical Publishing Company, Calcutta.

Raseduzzaman, Md., and E.S. Jensen. 2017. Does intercropping enhance yield stability in arable crop production? A meta-analysis. European Journal of Agronomy 91: 25–33. doi: 10.1016/j.eja.2017.09.009.

Reckling, M., T.F. Döring, G. Bergkvist, F.L. Stoddard, C.A. Watson, et al. 2018. Grain legume yields are as stable as other spring crops in long-term experiments across northern Europe. Agron. Sustain. Dev. 38(6): 63. doi: 10.1007/s13593-018-0541-3.

Reganold, J.P., L.F. Elliott, and Y.L. Unger. 1987. Long-term effects of organic and conventional farming on soil erosion. Nature 330(6146): 370–372. doi: 10.1038/330370a0.

Reganold, J.P., and J.M. Wachter. 2016. Organic agriculture in the twenty-first century. Nature Plants 2: 15221. doi: 10.1038/nplants.2015.221.

Reynolds, M., and H. Braun. 2019. Benefits to low-input agriculture. Nat. Plants 5(7): 652–653. doi: 10.1038/s41477-019-0462-4.

Richards, R.A. 1996. Defining selection criteria to improve yield under drought. Plant Growth Regulation 20(2): 157–166. doi: 10.1007/BF00024012.

Richner, W., and S. Sinaj. 2017. Grundlagen für die Düngung landwirtschaftlicher Kulturen in der Schweiz (GRUD 2017). Agrarforschung Schweiz 8(6). https://www.agroscope.admin.ch/agroscope/de/home/themen/pflanzenbau/ackerbau/Pflanzenernaehrung/grud.html.

Richter, C., and B. Kroschewski. 2006. Analysis of a Long-term Experiment with Repeated-measurement Models. Journal of Agronomy and Crop Science 192(1): 55–71. doi: 10.1111/j.1439-037X.2006.00167.x.

Röös, E., A. Mie, M. Wivstad, E. Salomon, B. Johansson, et al. 2018. Risks and opportunities of increasing yields in organic farming. A review. Agron. Sustain. Dev. 38(2): 14. doi: 10.1007/s13593-018-0489-3.

Rudmann, C., and H. Willer, editors. 2005. Jahrbuch Biolandbau Schweiz 2005. FiBL, Frick, Switzerland.

Rueda-Ayala, V., H.E. Ahrends, S. Siebert, T. Gaiser, H. Hüging, et al. 2018. Impact of nutrient supply on the expression of genetic improvements of cereals and row crops – A case study using data from a long-term fertilization experiment in Germany. European Journal of Agronomy 96: 34–46. doi: 10.1016/j.eja.2018.03.002.

Schlatter, B., J. Travnicek, J. Lernoud, and H. Willer. 2020. Current Statistics on Organic Agriculture Worldwide: Area, Operators and Market. In: Willer, H., Schlatter, B., Trávnícek, J., Kemper, L., and Lernoud, J., editors, The World of Organic Agriculture - Statistics and Emerging Trends 2020. Research Institute of Organic Agriculture FiBL & IFOAM - Organics International, Frick, Switzerland & Bonn, Germany. p. 337

Schmidhuber, J., and F.N. Tubiello. 2007. Global food security under climate change. PNAS 104(50): 19703–19708. doi: 10.1073/pnas.0701976104.

Schrama, M., J.J. de Haan, M. Kroonen, H. Verstegen, and W.H. Van der Putten. 2018. Crop yield gap and stability in organic and conventional farming systems. Agriculture, Ecosystems & Environment 256: 123–130. doi: 10.1016/j.agee.2017.12.023.

Seitler, E., L. Thöni, and M. Meier. 2016. Atmosphärische Stickstoff-Deposition in der Schweiz 2000 bis 2014. Forschungsstelle für Umweltbeobachtung, Rapperswil, Switzerland.

Serdjuk, M., U. Bodmer, and K.-J. Hülsbergen. 2018. Integration of biogas production into organic arable farming systems: crop yield response and economic effects. Org. Agr. 8(4): 301–314. doi: 10.1007/s13165-017-0196-4.

Seufert, V., and N. Ramankutty. 2017. Many shades of gray—The context-dependent performance of organic agriculture. Science Advances 3(3): e1602638. doi: 10.1126/sciadv.1602638.

Seufert, V., N. Ramankutty, and J.A. Foley. 2012. Comparing the yields of organic and conventional agriculture. Nature 485(7397): 229–232. doi: 10.1038/nature11069.

Sherman, J.D., J.M. Martin, N.K. Blake, S.P. Lanning, and L.E. Talbert. 2014. Genetic Basis of Agronomic Differences between a Modern and a Historical Spring Wheat Cultivar. Crop Science 54(1): 1. doi: 10.2135/cropsci2012.12.0710.

Shiferaw, B., M. Smale, H.-J. Braun, E. Duveiller, M. Reynolds, et al. 2013. Crops that feed the world 10. Past successes and future challenges to the role played by wheat in global food security. Food Sec. 5(3): 291–317. doi: 10.1007/s12571-013-0263-y.

Smith, O.M., A.L. Cohen, C.J. Rieser, A.G. Davis, J.M. Taylor, et al. 2019. Organic Farming Provides Reliable Environmental Benefits but Increases Variability in Crop Yields: A Global Meta-Analysis. Front. Sustain. Food Syst. 3. doi: 10.3389/fsufs.2019.00082.

Smith, R.G., and K.L. Gross. 2006. Weed community and corn yield variability in diverse management systems. Weed Science 54(1): 106–113. doi: 10.1614/WS-05-108R.1.

Smith, R.G., F.D. Menalled, and G.P. Robertson. 2007. Temporal Yield Variability under Conventional and Alternative Management Systems. Agronomy Journal 99(6): 1629–1634. doi: 10.2134/agronj2007.0096.

Spink, J. 2010. Winter wheat yield trends in Europe and at a global scale. p. 111–113

Springmann, M., H.C.J. Godfray, M. Rayner, and P. Scarborough. 2016. Analysis and valuation of the health and climate change cobenefits of dietary change. PNAS 113(15): 4146–4151. doi: 10.1073/pnas.1523119113.

Stern, D.I. 2005. Global sulfur emissions from 1850 to 2000. Chemosphere 58(2): 163–175. doi: 10.1016/j.chemosphere.2004.08.022.

Summers, R.W., and J.K.M. Brown. 2013. Constraints on breeding for disease resistance in commercially competitive wheat cultivars. Plant Pathology 62: 115–121. doi: 10.1111/ppa.12165.

Tallec, T., S. Diquélou, S. Lemauviel, J.B. Cliquet, F. Lesuffleur, et al. 2008. Nitrogen:sulphur ratio alters competition between Trifolium repens and Lolium perenne under cutting: Production and competitive abilities. European Journal of Agronomy 29(2): 94–101. doi: 10.1016/j.eja.2008.04.004.

Taylor, L.R. 1961. Aggregation, Variance and the Mean. Nature 189(4766): 732. doi: 10.1038/189732a0.

Tilman, D., C. Balzer, J. Hill, and B.L. Befort. 2011. Global food demand and the sustainable intensification of agriculture. PNAS 108(50): 20260–20264. doi: 10.1073/pnas.1116437108.

Tilman, D., P.B. Reich, and J.M.H. Knops. 2006. Biodiversity and ecosystem stability in a decade-long grassland experiment. Nature 441(7093): 629–632. doi: 10.1038/nature04742.

Tuck, S.L., C. Winqvist, F. Mota, J. Ahnström, L.A. Turnbull, et al. 2014. Land-use intensity and the effects of organic farming on biodiversity: a hierarchical meta-analysis. Journal of Applied Ecology 51(3): 746–755. doi: 10.1111/1365-2664.12219.

Verbruggen, E., W.F.M. Röling, H.A. Gamper, G.A. Kowalchuk, H.A. Verhoef, et al. 2010. Positive effects of organic farming on below-ground mutualists: large-scale comparison of mycorrhizal fungal communities in agricultural soils. New Phytologist 186(4): 968–979. doi: 10.1111/j.1469-8137.2010.03230.x.

Viechtbauer, W. 2010. Conducting meta-analyses in R with the metafor package. Journal of Statistical Software 36(3): 1–48.

Voss-Fels, K.P., A. Stahl, B. Wittkop, C. Lichthardt, S. Nagler, et al. 2019. Breeding improves wheat productivity under contrasting agrochemical input levels. Nat. Plants 5(7): 706–714. doi: 10.1038/s41477-019-0445-5.

Walker, T.W., and A.F.R. Adams. 1958. Competition for sulphur in a grass-clover association. Plant Soil 9(4): 353–366. doi: 10.1007/BF01343832.

Webb, J., C. Jephcote, A. Fraser, J. Wiltshire, S. Aston, et al. 2016. Do UK crops and grassland require greater inputs of sulphur fertilizer in response to recent and forecast reductions in sulphur emissions and deposition? Soil Use and Management 32(1): 3–16. doi: 10.1111/sum.12250.

Wittwer, R.A., B. Dorn, W. Jossi, and M.G.A. van der Heijden. 2017. Cover crops support ecological intensification of arable cropping systems. Scientific Reports 7: 41911. doi: 10.1038/srep41911.

Supplementary material

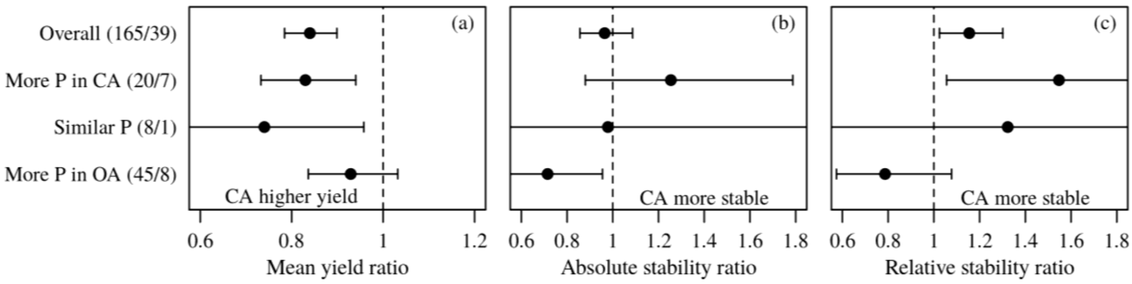
Chapter A

Supplementary Note 1

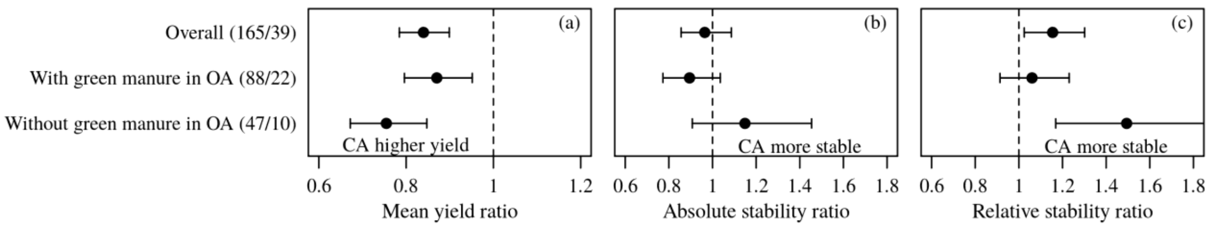
It has previously been observed, that the standard deviation is often correlated to the mean. This relationship, known as Taylor's power law (Taylor, 1961), has been shown to also hold true for crop yields when comparing the natural log of the variance and the natural log of the mean (Döring et al., 2015). The coefficient of variation (CV) is used as one approach to correct for this positive relationship of the standard deviation and the mean. Krebs (1998) noted that the use of CV is only appropriate, when the slope of the log(SD) on log(mean) is equal to 1 (corresponding to a slope of 2 in the more commonly used regression of log(variance) on log(mean)). When looking at observed mean yields and the standard deviation across observations (not ratios), a significant positive correlation can be observed between log(SD) and log (mean), which results in a negative slope of log(CV) on log(mean) in both datasets (Supplementary Figure 4). However, in our analysis we did not compare the variability between observations, but between treatments, as we assessed the ratio between treatments. We therefore regressed log(SD) and log(CV), respectively, on log(mean) of both treatments within each observation. As this regression only used two points (both treatments) from each observation, the estimated slopes showed a wide and non-symmetric distribution (Supplementary Fig. 5) and we therefore used the pseudo-median (estimated by the Hodges-Lehmann estimator) instead of the mean as location parameter. We calculated 95% confidence intervals and tested against the alternative hypothesis of independence of log(SD) from log(mean), i.e. the true location parameter is not equal to 0 for the regression of log(SD) and -1 for the regression of log(CV) using the wilcox.test() function in R. The estimated overall slopes of log(SD) on log(mean) ($b=-0.42$ and $b=-0.09$ for the organic and no-tillage dataset, respectively), were not significantly ($\alpha=0.05$) different from zero ($P=0.06$ and $P=0.65$ for the organic and no-tillage dataset, respectively). This indicates that the SD was independent of the mean, which is also supported by our finding that there is no difference in absolute stability. Furthermore, the slope of log(CV) on log(mean) were, on average, $b=-1.42$ and $b=-1.09$ for the organic and no-tillage dataset, respectively, and not significantly (alpha=0.05) different from -1 ($P=0.06$ and $P=0.65$ for the organic and no-tillage dataset, respectively), indicating that the CV is inversely related to the mean. This is further confirmed by the regression of the ratios of the stability measures on the mean yield ratios (see Figure 6 in the main text). It is important to note, that the significance level for both regressions (log(SD) and log(CV) in the organic dataset is only very slightly above the threshold of alpha=0.05, indicating that there is some negative relation between log(SD) and log(mean). This would mean that conventional agriculture has not only higher mean yield, but also increased absolute stability. However, the actual meta-analysis procedure does not indicate this (Fig. 1 in the main text).

The observed independence (or even negative relationship) of log(SD) and log(mean) does not indicate that Taylor's Power Law occurred. It is therefore important to carefully interpret the difference in variation measured by CV, as the difference is mainly due to the difference in variation measured by SD, which is indicated by the linear negative relationship between log(CV) and log(mean). Furthermore, the positive relationship between SD and mean across observations does not influence our results as we used pairwise comparisons between treatments within observations by using the ratios of the mean yield or the respective stability measures.

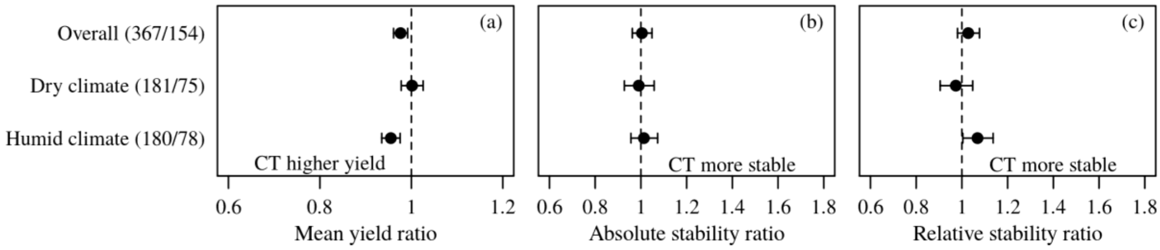
Supplementary Figures



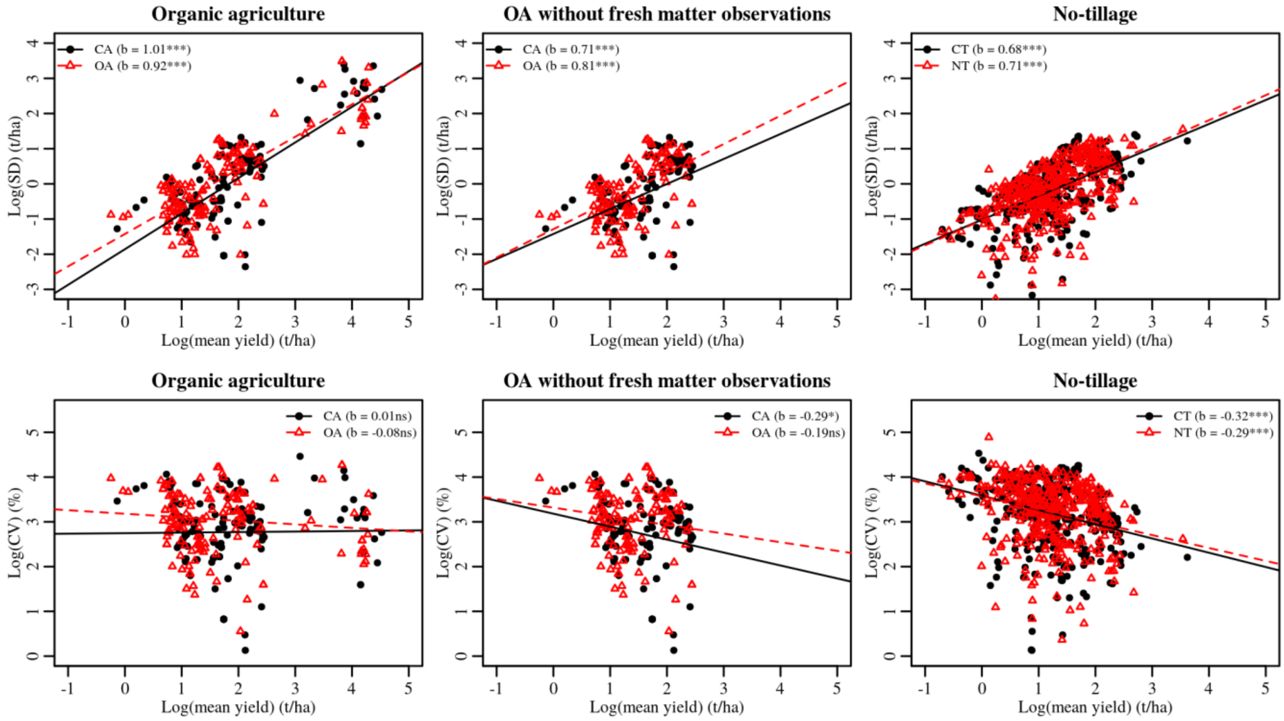
Supplementary Figure 1: Effect of phosphorus input on yield and yield stability comparing organic and conventional agriculture. Mean yield ratio (a), absolute stability ratio (b), and relative stability ratio (c) in organic (OA) versus conventional (CA) agriculture for different levels of phosphorus (P) input in organic and conventional systems. Numbers in parentheses denote the number of observations and studies. A ratio of 1 means that there is no difference between organic and conventional managed systems while values <1 indicate higher yield for conventional agriculture. For both stability measures ratios >1 indicate greater absolute and relative stability for conventional agriculture. Values are mean effect sizes with 95% confidential intervals. Yield or stability were deemed significantly different between organic and conventional agriculture if the 95% confidential intervals of the ratios did not overlap one.



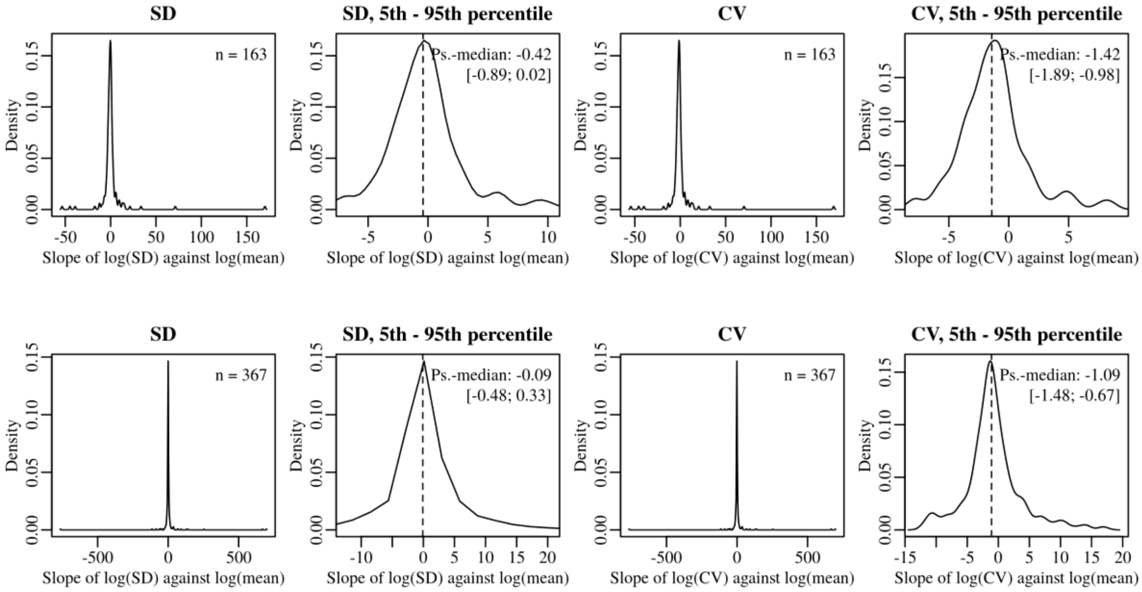
Supplementary Figure 2: Effect of green manure on yield and yield stability comparing organic and conventional agriculture. Mean yield ratio (a), absolute stability ratio (b), and relative stability ratio (c) in organic (OA) versus conventional (CA) agriculture using all observations (top row) or observations with or without the application of green manure in organic agriculture. Numbers in parentheses denote the number of observations and studies. A ratio of 1 means that there is no difference between organic and conventionally managed systems while values <1 indicate higher yield for conventional agriculture. For both stability measures, ratios >1 indicate greater absolute and relative stability for conventional agriculture. Values are mean effect sizes with 95% confidential intervals. Yield or stability were deemed significantly different between organic and conventional agriculture if the 95% confidential intervals of the ratios did not overlap one.



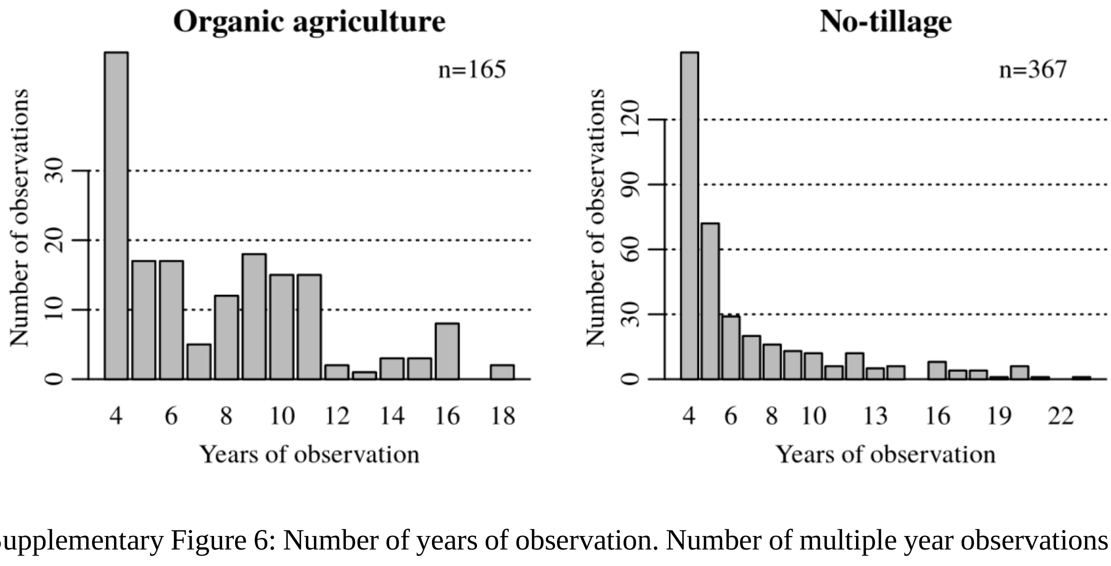
Supplementary Figure 3: Effect of climate on yield and yield stability comparing no-tillage and conventional tillage. Mean yield ratio (a), absolute stability ratio (b), and relative stability ratio (c) of no-tillage (NT) versus conventional tillage (CT) for dry and humid climates. Following Pittelkow et al. (2015) we defined dry climates based on the aridity index with values less than 0.65. Numbers in parentheses denote the number of observations and studies. A ratio of 1 means that there is no difference between no-tillage and conventional tillage while yield values <1 indicate higher yield for conventional tillage. For both stability measures values >1 indicate greater stability for conventional tillage. Values are mean effect sizes with 95% confidential intervals. Yield or stability were deemed significantly different between no-tillage and conventional tillage if the 95% confidential intervals of the ratios did not overlap one.



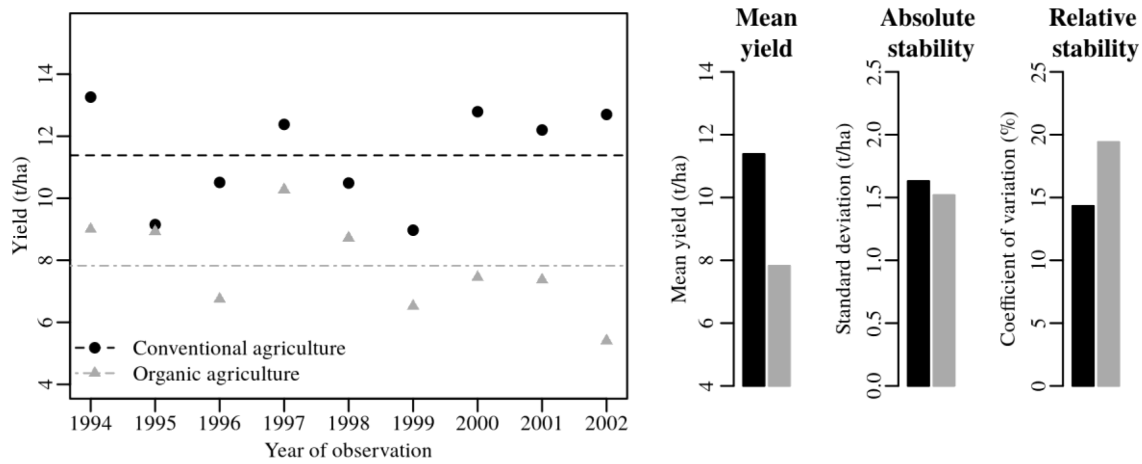
Supplementary Figure 4: Relationship of observed mean yield and measures of variation. Standard deviation (top row) and coefficient of variation (bottom row) for the dataset comparing organic (OA) and conventional agriculture (CA) (left and middle column) and the dataset comparing no-tillage (NT) and conventional tillage (CT) (right column), respectively. Each dot represents one multiple year observation (MYO) and both axes are on the log scale. The regression line was fitted separately for each treatment on log-transformed values, i.e. log(y) = a + b * log(x), where y was the respective stability measure (standard deviation or coefficient of variation) and x was the mean yield ratio. The dataset comparing organic and conventional agriculture (left column) contained several observations with a mean yield greater than 20 t/ha. These are observations for fruits or vegetables, where yield was determined on fresh matter. To account for that, the middle column shows the relationship between the observed mean yield and the observed standard deviation (top row) or the coefficient of variation (bottom row) without observations for fruits and vegetables, where yield was measured on dry matter. * and *** denote significance at P<0.05 and P<0.001, respectively, for a t-test with H0: b=0, and n.s. denotes non-significant (P>0.05).



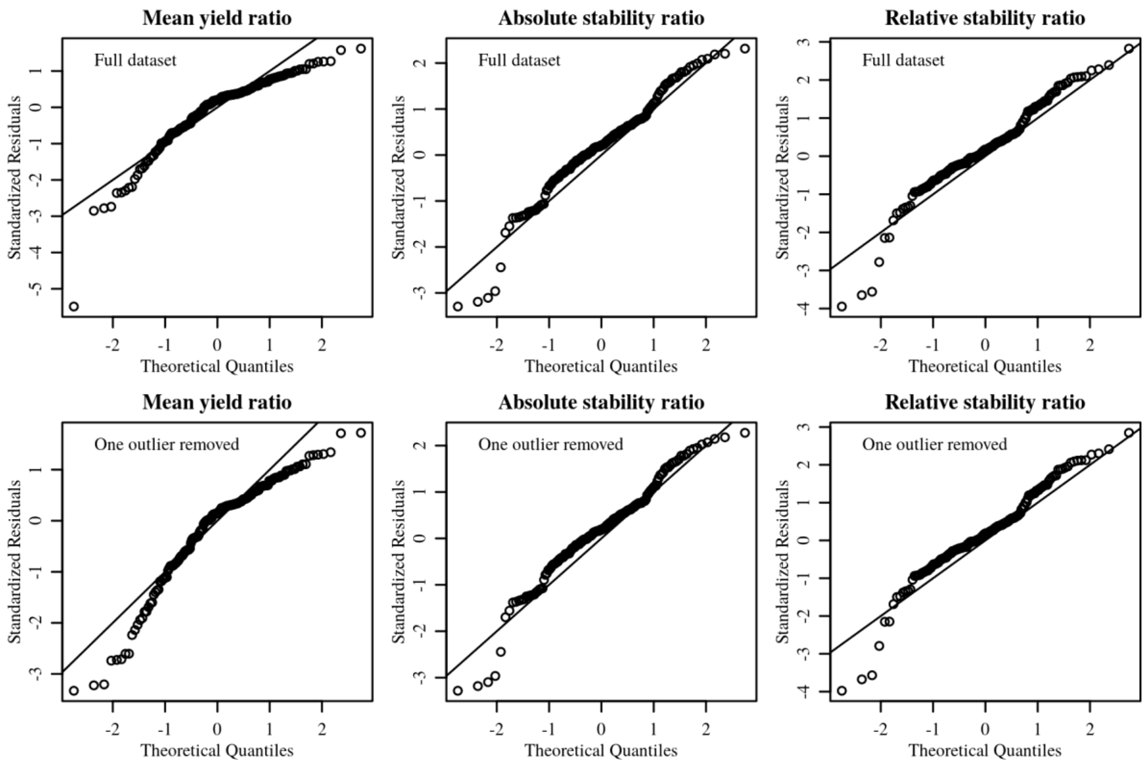
Supplementary Figure 5: Relationship of observed mean yield and measures of variation within each observation. Kernel density plots of the distribution of the slope b from the regressions of log(SD) and log(CV), respectively, on log(mean), within each observation for the dataset on organic agriculture (top row) and no-tillage (bottom row). In the dataset on organic agriculture, two MYOs had exactly the same yield for both treatments, and it was thus not possible to calculate the regression for these two MYOs. In order to better visualize the distribution, only the distribution between the 5th and 95th percentile is shown in the second and fourth column. The dashed line indicates the pseudo-mean (estimated by the Hodges-Lehmann estimator).



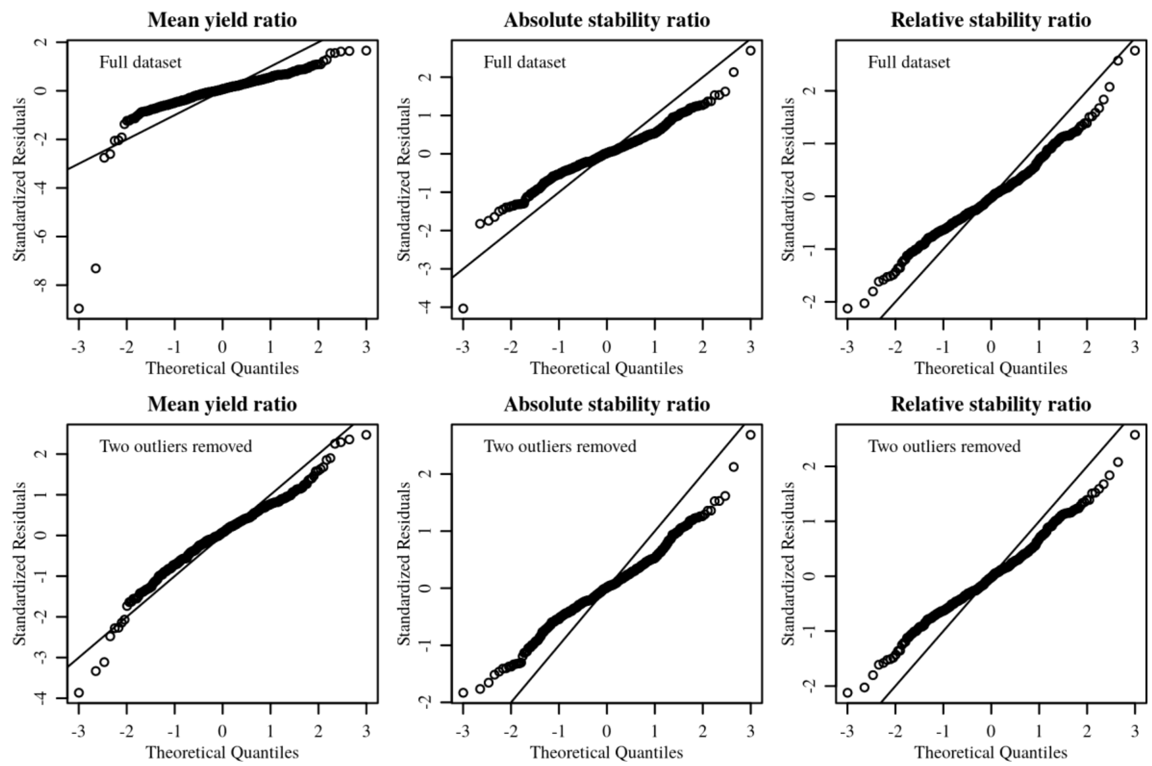
Supplementary Figure 6: Number of years of observation. Number of multiple year observations in relation to the number of years of observation in the dataset on organic agriculture (left) and on no-tillage (right). In the dataset on organic agriculture the observations are spread more equally over the whole range, whereas for the dataset on no-tillage 60% of all observations are based on 4 or 5 years of observation.



Supplementary Figure 7: Example of a study monitoring the effects of organic and conventional agriculture on maize yield from 1994 to 2002 (left panel). The right panel shows the mean yield, the absolute stability and the relative stability (i.e. the coefficient of variation, which measures the temporal variability across years relative to the mean yield level across those years). This study provided one multiple year observation as input for the meta-analysis. Data obtained from Denison et al. (2004).



Supplementary Figure 8: QQ-plots of the standardized residuals from overall model for the dataset on organic farming. The top row shows the distributions including the outlier and the bottom row shows the distribution after one outlier was removed.



Supplementary Figure 9: QQ-plots of the standardized residuals from overall model for the dataset on no-tillage. The top row shows the distributions including the outlier and the bottom row shows the distribution after two outliers were removed.

Supplementary Tables

Supplementary Table 1: Estimates for mean yield, absolute and relative stability for all species contained in the dataset on organic agriculture (Ponisio et al., 2015).

Crop species	Observations/ Studies	Mean yield ratio			Absolute stabilty ratio			Relative stabilty ratio		
		Mean	lower CL	upper CL	Mean	lower CL	upper CL	Mean	lower CL	upper CL
maize	45/18	0.87	0.80	0.95	0.96	0.79	1.17	1.12	0.91	1.37
soybean	34/13	0.85	0.78	0.93	1.34	1.08	1.66	1.56	1.25	1.96
wheat	29/10	0.73	0.66	0.81	0.78	0.59	1.03	1.09	0.81	1.45
oat	10/4	0.88	0.74	1.06	1.05	0.76	1.45	1.15	0.79	1.66
tomato	9/4	1.01	0.86	1.19	0.52	0.34	0.79	0.51	0.33	0.80
barley	11/3	0.67	0.57	0.79	1.11	0.73	1.67	1.60	1.03	2.47
apple	3/2	1.00	0.69	1.46	0.82	0.36	1.87	0.77	0.30	1.97
alfalfa	4/1	1.02	0.85	1.24	0.89	0.56	1.40	0.88	0.53	1.44
beetroot	2/1	1.03	0.74	1.45	1.83	0.57	5.88	1.77	0.54	5.78
carrot	2/1	1.06	0.75	1.51	1.27	0.40	4.08	1.19	0.36	3.90
lettuce	2/1	0.91	0.56	1.46	2.14	0.74	6.17	2.36	0.77	7.28
potato	2/1	1.00	0.65	1.54	0.77	0.24	2.48	0.77	0.23	2.58
rye	2/1	0.75	0.52	1.10	1.13	0.39	3.25	1.49	0.50	4.44
spring wheat	2/1	0.74	0.50	1.09	1.69	0.61	4.68	2.46	0.85	7.17
bean	1/1	0.65	0.27	1.57	0.32	0.11	0.90	0.50	0.13	1.94
chard	1/1	1.02	0.65	1.62	0.66	0.23	1.89	0.67	0.22	2.06
cotton	1/1	1.11	0.60	2.06	1.39	0.38	5.12	1.25	0.30	5.17
elephant foot yam	1/1	1.19	0.73	1.92	1.08	0.33	3.50	0.91	0.26	3.14
flax	1/1	0.56	0.33	0.95	0.66	0.27	1.59	1.18	0.44	3.18
grapes	1/1	0.91	0.54	1.55	0.95	0.35	2.55	1.04	0.35	3.09
pumpkin	1/1	1.18	0.60	2.33	0.92	0.32	2.60	0.80	0.23	2.75
safflower	1/1	0.87	0.53	1.41	0.81	0.34	1.96	0.93	0.35	2.46

Supplementary Table 2: Estimates for mean yield, absolute and relative stability for all species contained in the dataset on no-tillage (Pittelkow et al., 2015).

Crop species	Observations/ Studies	Mean yield ratio			Absolute stabilty ratio			Relative stabilty ratio		
		Mean	lower CL	upper CL	Mean	lower CL	upper CL	Mean	lower CL	upper CL
barley	28/17	0.96	0.91	1.02	1.03	0.88	1.20	1.07	0.90	1.27
cotton	13/6	0.95	0.88	1.02	1.05	0.78	1.41	1.04	0.75	1.43
maize	108/55	0.95	0.93	0.98	1.01	0.93	1.09	1.07	0.98	1.16
rice	10/7	0.94	0.90	0.98	0.94	0.68	1.28	0.99	0.72	1.36
sorghum	15/12	1.04	0.96	1.13	1.04	0.85	1.27	0.96	0.78	1.20
soybean	53/31	1.01	0.98	1.04	1.01	0.91	1.12	1.02	0.91	1.14
wheat	114/65	0.99	0.97	1.01	1.01	0.93	1.09	1.03	0.94	1.12
pea	5/3	1.03	0.85	1.25	1.10	0.68	1.76	1.01	0.60	1.71
bean	4/4	1.04	0.85	1.27	0.86	0.62	1.19	0.75	0.50	1.12
canola	3/3	1.07	0.85	1.35	1.09	0.56	2.09	1.02	0.51	2.05
oat	3/1	1.12	0.87	1.44	0.82	0.51	1.30	0.75	0.44	1.27
chickpea	2/2	0.98	0.72	1.34	1.08	0.66	1.76	1.11	0.62	1.99
sunflower	2/2	1.14	0.79	1.65	0.94	0.45	1.97	0.79	0.33	1.88
lentil	2/1	1.17	0.86	1.59	0.88	0.55	1.39	0.76	0.44	1.32
lupin	2/1	1.07	0.81	1.41	0.97	0.66	1.42	0.90	0.56	1.44
rye	2/1	0.88	0.70	1.11	1.10	0.49	2.45	1.25	0.54	2.87

Supplementary Table 3: Corrections that were applied to the original dataset from Ponisio et al. (2015)

Comparison	Column name	Original value	New Value	Reasoning
43	X_coord	-95.8	95.48	same value as other comparisons from the same study
349	Legume	non-legumes	non-legume	all other entries are are without plural s
349	Rotation	org	more org	to achieve similarity to other entries
476-478	Var ac. Years	yes	no	in year-org and year-con only one year, thus variance across years is not possible
919-920	Var ac. Years	yes	no	in year-org and year-con only one year, thus variance across years is not possible
1037-1039	year-org	2001;2002;2004	2001 to 2003	same format as other observations from same study
1037-1039	year-con	2001;2002;2004	2001 to 2003	same format as other observations from same study
961	year-org	1990 and 1992	1990 to 1991	same format as other observations from same study
961	year-con	1990 and 1992	1990 to 1991	same format as other observations from same study
941-953	Yield unit	Mg/ha	kg/ha	wrong unit, see original publication
220-264	Yield unit	kg/ha	Mg/ha	all maize entries had wrong unit, see original publication
62-65			completely removed	extracted values do not correspond to original publication, also only 3 years in original publication

Supplementary Table 4: Example of the generation of multiple year observation (MYOs) from single-year observations for selected comparisons in the dataset on organic agriculture. The column "study" and "comparison" refer to the original dataset by Ponisio et al. (2015)

Study	Comp-arison	Author (year)	Year of obser-vation	Treatment		Yield		Multiple year observation (MYO)	Mean over years		SD over years		CV over years	
				conv	org	conv	org		conv	org	conv	org	conv	org
3	43	Archer et al. (2007)	2003	conv	org	3.7	2.73							
3	8	Archer et al. (2007)	2004	conv	org	3.73	3.51	A	3.57	2.81	0.31	0.49	11.6	5.7
3	9	Archer et al. (2007)	2005	conv	org	3.75	2.62							
3	10	Archer et al. (2007)	2006	conv	org	3.11	2.38							
27	302	Doltra et al. (2010)	2005	CF+C+F	OF+C+M	5.68	5.19							
27	269	Doltra et al. (2010)	2006	CF+C+F	OF+C+M	5.84	5.14	B	5.47	5.04	0.62	0.27	8.8	18.9
27	267	Doltra et al. (2010)	2007	CF+C+F	OF+C+M	4.54	4.64							
27	271	Doltra et al. (2010)	2008	CF+C+F	OF+C+M	5.81	5.18							
27	270	Doltra et al. (2010)	2005	CF+C+F	OF-C+M	5.68	4.92							
27	282	Doltra et al. (2010)	2006	CF+C+F	OF-C+M	5.84	4.38	C	5.47	3.95	0.62	0.84	8.8	4.7
27	268	Doltra et al. (2010)	2007	CF+C+F	OF-C+M	4.54	3.2							
27	275	Doltra et al. (2010)	2008	CF+C+F	OF-C+M	5.81	3.29							

Supplementary Table 5: Example of splitting multiple year observations (MYO) that contained subtreatments in the dataset on no-tillage from Pittelkow et al. (2015). MYO A did not contain any subtreatments. MYO B and C were listed in subsequent order in the original dataset, and MYO D and E were listed in alternating order.

Study	**duration**	**subtreatments**	**order**	**MYO**
Aase et al. (1997)	1	no		A
Aase et al. (1997)	2	no		A
Aase et al. (1997)	3	no		A
Aase et al. (1997)	4	no		A
Dalal et al. (2013)	6	yes	subsequent	B
Dalal et al. (2013)	7	yes	subsequent	B
Dalal et al. (2013)	8	yes	subsequent	B
Dalal et al. (2013)	9	yes	subsequent	B
Dalal et al. (2013)	10	yes	subsequent	B
Dalal et al. (2013)	6	yes	subsequent	C
Dalal et al. (2013)	7	yes	subsequent	C
Dalal et al. (2013)	8	yes	subsequent	C
Dalal et al. (2013)	9	yes	subsequent	C
Dalal et al. (2013)	10	yes	subsequent	C
Maurya (1986)	1	yes	alternating	D
Maurya (1986)	1	yes	alternating	E
Maurya (1986)	2	yes	alternating	D
Maurya (1986)	2	yes	alternating	E
Maurya (1986)	3	yes	alternating	D
Maurya (1986)	3	yes	alternating	E
Maurya (1986)	4	yes	alternating	D
Maurya (1986)	4	yes	alternating	E

Chapter B

Supplementary Method 1

In a preliminary analysis, we compared an analysis using plot values with different models taking account of the repeated measure design, because the same plots are sampled repeatedly, and plot residuals are thus expected to be correlated. The basic model without any correction for a possible correlation (NC) for the estimation of means and trends was

YIELD ~ B + T + Yn + T: Yn + <u>F:Y</u> + <u>F:Y:T</u> + <u>B:F:Y</u> + *F:Y:T:B* (NC-Means&Trends),

where B stands for block, T for treatment, F for field, Yn for year used as numeric, Y for year used as factor, underlines indicate the random effects, italics the residuals, and a colon (:) an interaction.

The model for the estimation of the treatment variances (absolute stability) was accordingly

YIELD ~ B + T + Yn + T: Yn + <u>US(T,F:Y)</u> + <u>B:F:Y</u> + *F:Y:T:B* (NC-Variances),

where US(T,F:Y) symbolizes an unstructured variance-covariance matrix with the diagonal being the variances of the treatments.

Next, we modelled a compound symmetry structure (CS) through an additional random plot effect. This model assumes a constant correlation between plot residuals, irrespective of the time distance between sampling.

YIELD ~ B + T + Yn + T: Yn + <u>F:Y</u> + <u>F:Y:T</u> + <u>B:F:Y</u> + <u>Plot</u> + *F:Y:T:B* (CS-Means&Trends)

YIELD ~ B + T + Yn + T: Yn + <u>US(T,F:Y)</u> + <u>B:F:Y</u> + <u>Plot</u> + *F:Y:T:B* (CS-Variances)

For this model, we calculated the correlation between plot residuals with $rho = \sigma_p^2/(\sigma_p^2 + \sigma_\varepsilon^2)$, where σ_p^2 is the variance of the random plot effect and σ_ε^2 the residual variance.

Additionally, we allowed for heterogenic residuals per field x year combination (CSH):

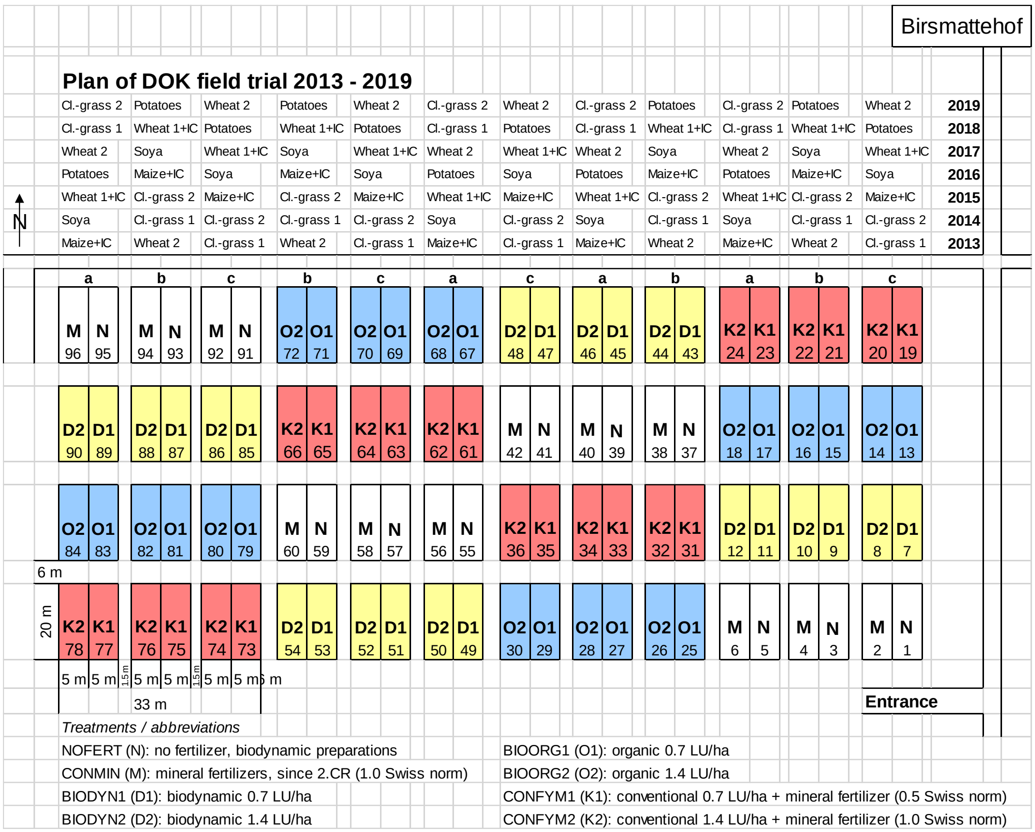
YIELD ~ B + T + Yn + T: Yn + <u>F:Y</u> + <u>F:Y:T</u> + <u>B:F:Y</u> + <u>Plot</u> + *VS(F:Y,T:B)* (CSH-Means&Trends)

YIELD ~ B + T + Yn + T: Yn + <u>US(T,F:Y)</u> + <u>B:F:Y</u> + <u>Plot</u> + *VS(F:Y,T:B)* (CSH-Variances)

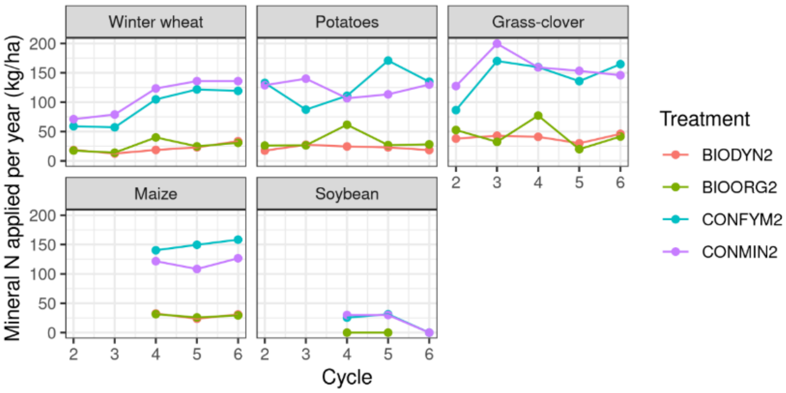
Here, VS(F:Y,T:B) symbolizes a diagonal matrix with the diagonal being the residuals for each field x year combination.

We compared these models to the analysis on treatment by field x year combinations means as described in the main text for winter wheat (Supplementary Table 10). The Akaike Information Criterion suggested that the CSH model gave the best fits for both the model for means and trends and for the variance model. In the CS models, rho was estimated to around 0.07, suggesting that there is no correlation between plot residuals. This observation matches with the analysis, correlating plot residuals from models run within each year (Supplementary Figure 12). Because these observations suggest that there is little to no correlation, estimates and statistics did only marginally differ between the analyses on plot level and on field x year means, and some models on the plot values did not converge, we decided to conduct the full analysis on field x year means.

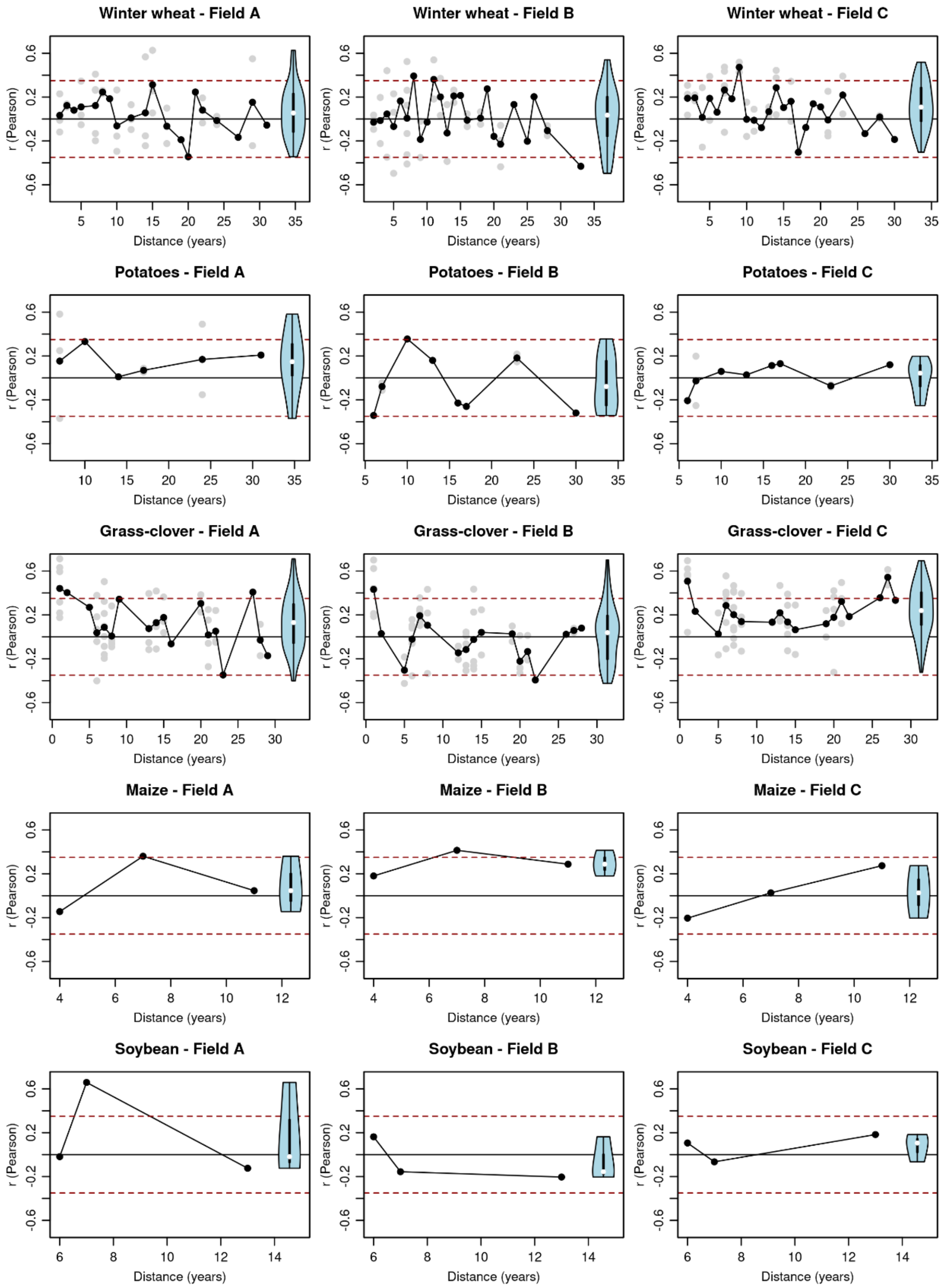
Supplementary Figures



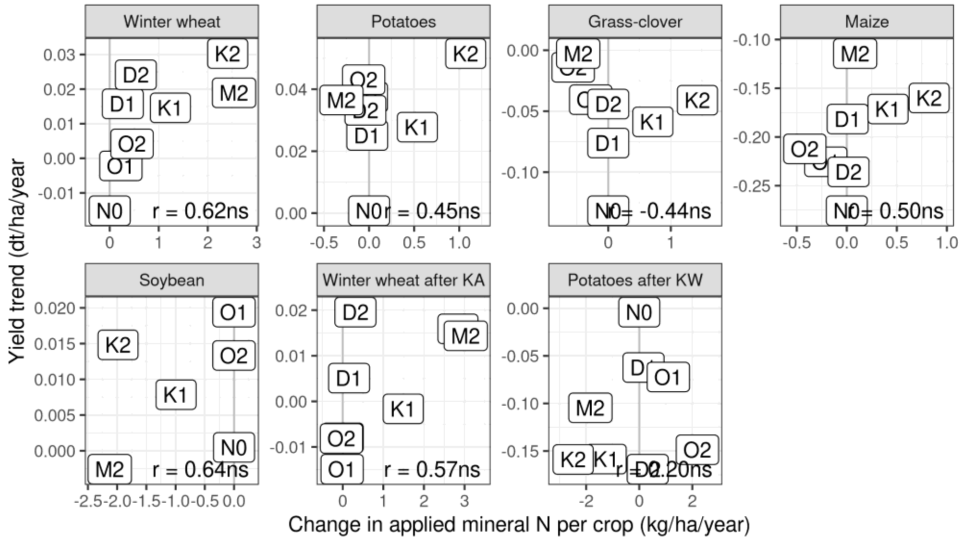
Supplementary Figure 10: Field layout of the DOK trial, with the crop rotation of the 6th crop rotation cycle.



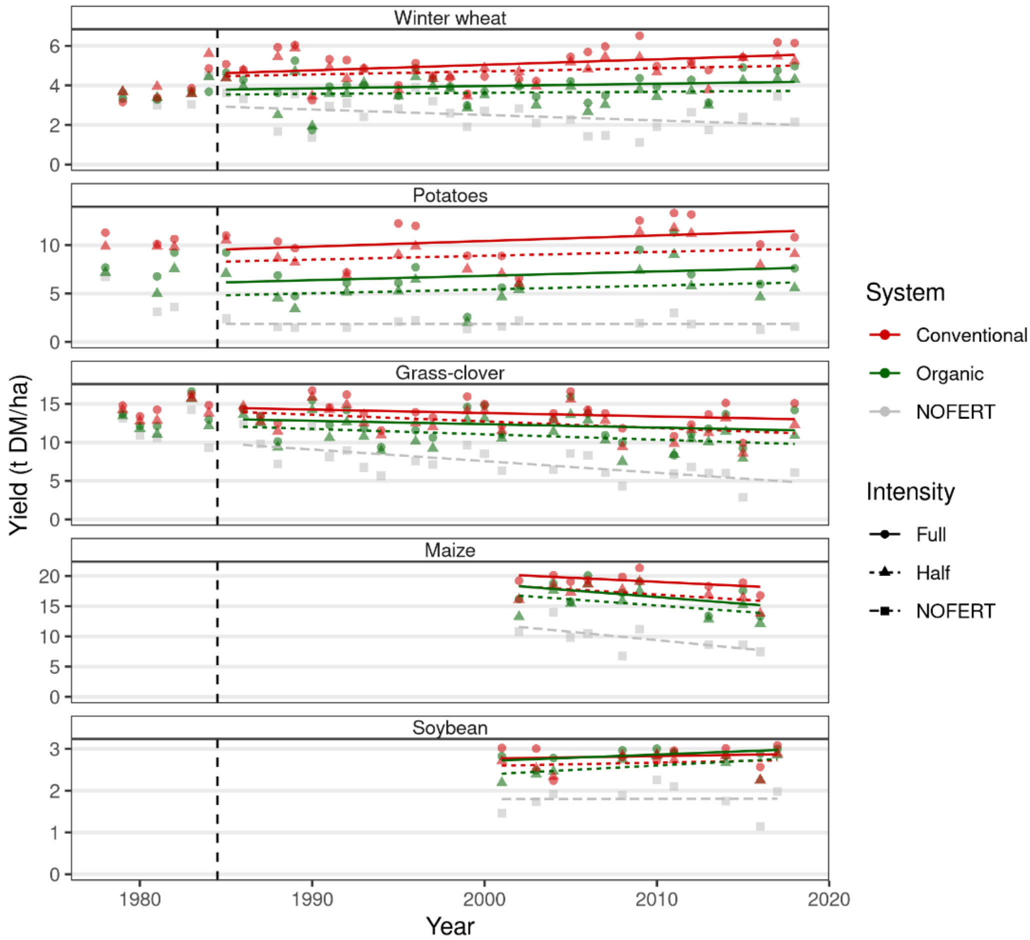
Supplementary Figure 11: Change of applied mineral N per crop over per crop rotation cycles.



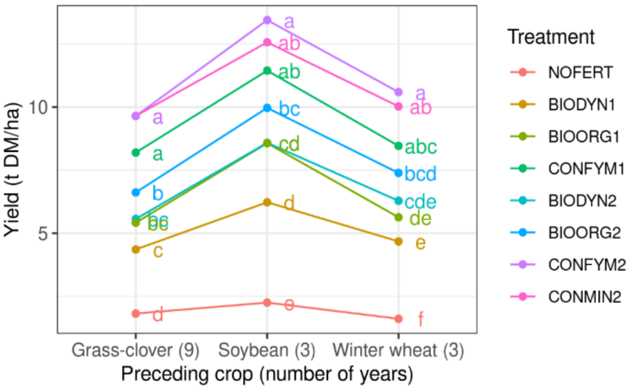
Supplementary Figure 12: Pearson correlation coefficients of the residuals from a model with fixed treatment and block effects conducted within each year. Correlations were calculated between all pairs of years, and means for each year distance plotted against the distance between the years. The red line indicates the threshold level of significance at P=0.05 with 31 of degrees of freedoms, as the correlation is based on 32 plots. Black dots represent the mean of all correlations for a given distance of years, while grey dots indicate the single correlations. The violin plots denote the distribution of all correlations over all pairs of years.



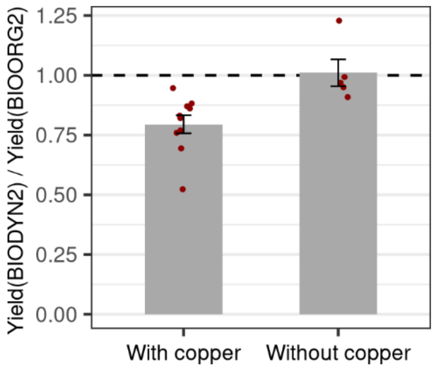
Supplementary Figure 13: Relation of the estimated yield trend and the change in mineral N fertilization. r indicates the Pearson correlation coefficient and ns indicates non-significance (P>0.05). . O: BIODYN, D: BIODYN, K: CONFYM, M: CONMIN; 2: regular, 1: half fertilisation. KA: potatoes, KW: grass-clover.



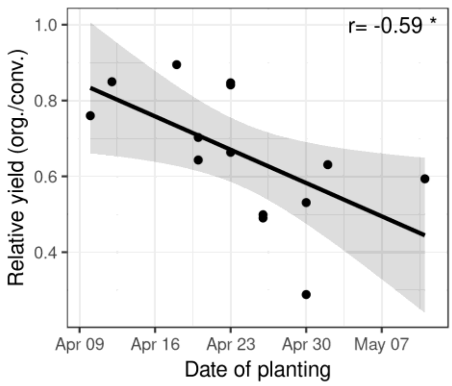
Supplementary Figure14: Observed yields per year. Conventional represents the mean of the CONFYM and CONMIN, and organic the mean of the BIOORG and BIODYN treatments. The vertical dashed lines indicate the start of the second crop rotation cycle. Data from the first crop rotation cycle were omitted from the analysis.



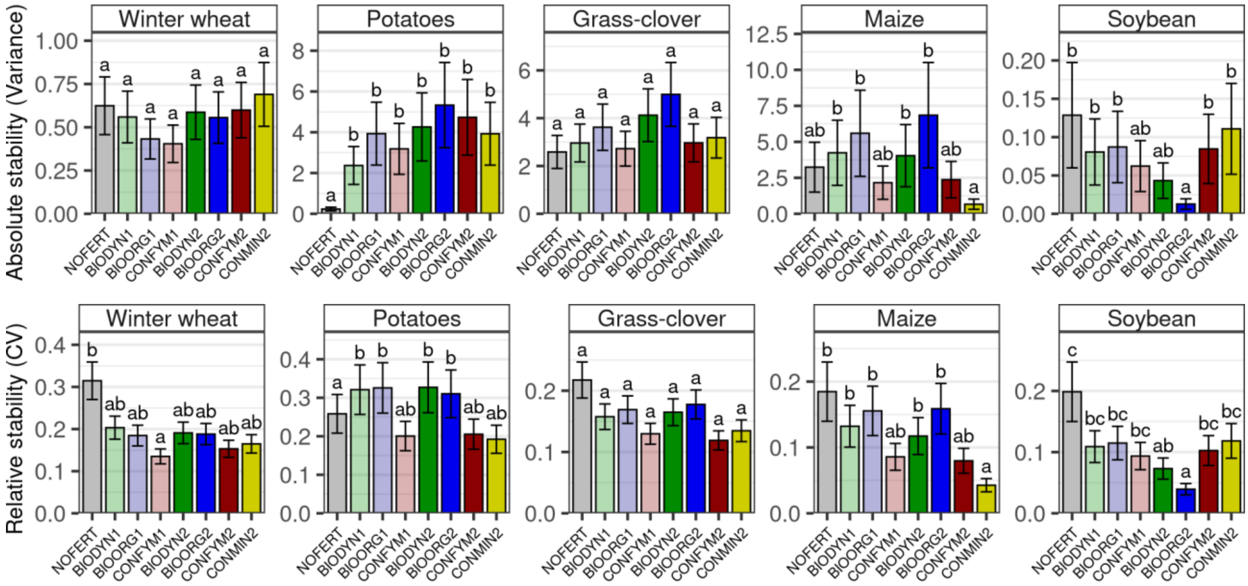
Supplementary Figure 15: Effect of the different preceding crops on the yield of potatoes. Dots are the means over all years with the respective preceding crop.



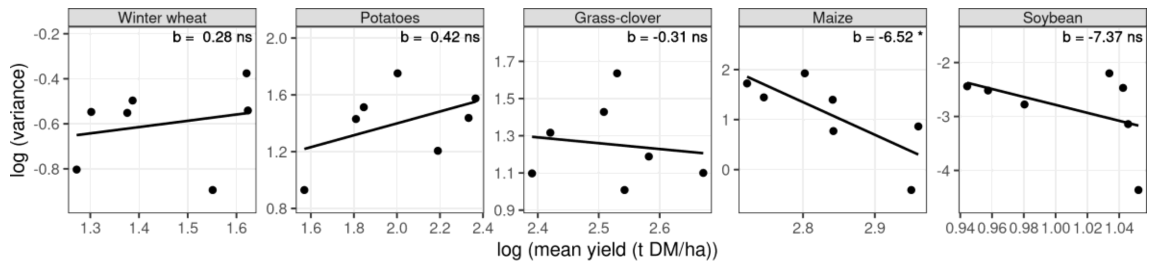
Supplementary Figure 16: Effect of the copper application against *Phytophtora infestans*. The bars represent the average yield ratio of BIODYN2 to BIOORG2 in years where copper was applied in BIOORG2 (left, n=10) and in years where no copper was applied (right, n=5), respectively, where n is the number of years. Red dots indicate the respective ratio in the single field x year combinations and error bars the standard error of the average yield ratio.



Supplementary Figure 17: Relation of the relative yield between organic and conventional treatments at regular fertilization (BIODYN2, BIOORG2 to CONFYM2, CONMIN2) to the day of planting in potatoes. r is the Pearson correlation coefficient and * indicates P<0.05.



Supplementary Figure 18: Absolute (top) and relative (bottom) stability. Error bars are standard errors. Treatments that do not carry the same letters are significantly different at P<0.05.



Supplementary Figure 19: Testing of Taylor-Power Law, which predicts a linear relationship of the natural logarithm of variance and the natural logarithm of the mean yield. One dot represents one treatment and significance indicates if the slope is different from b = 0, assessed by t-test. Treatment NOFERT has been omitted, as its mean yield is much lower than the other treatments and could lead to a spurious correlation through distant clusters. ns indicates P>0.05 and * P<0.05.

Supplementary Tables

Supplementary Table 6: Crop rotations in the different crop rotation cycles. Underlined crops were followed by green-manure.

Cycle 1 1978-1984	Cycle 2 1985-1991	Cycle 3 1992-1998	Cycle 4 1999-2005	Cycle 5 2006-2012	Cycle 6 2013-2019
Potatoes	Potatoes	Potatoes	Potatoes	Maize	Maize
Winter wheat	Winter wheat	Winter wheat	Winter wheat	Winter wheat	Soybean
Cabbage	Beetroot	Beetroot	Soybean	Soybean	Winter wheat
Winter wheat	Winter wheat	Winter wheat	Maize	Potatoes	Potatoes
Winter barley	Winter barley	Grass-clover	Winter wheat	Winter wheat	Winter wheat
Grass-clover	Grass-clover	Grass-clover	Grass-clover	Grass-clover	Grass-clover
Grass-clover	Grass-clover	Grass-clover	Grass-clover	Grass-clover	Grass-clover

Supplementary Table 7: Characterization of the organic fertilizers.

Product	Manure (kg/t)			Slurry (kg/m³)		
System	BIODYN	BIOORG	CONFYM	BIODYN	BIOORG	CONFYM
Total N	5.07	4.69	4.42	1.08	0.96	1.40
Mineral N	0.17	0.46	0.81	0.68	0.59	1.05
P	1.80	1.64	1.13	0.15	0.13	0.16
K	5.00	6.11	4.34	2.82	1.94	2.48
Ca	12.92	5.42	2.96	0.59	0.42	0.51
Mg	1.77	1.12	0.89	0.19	0.12	0.15
Corg	73.43	81.82	85.46	7.29	5.61	7.14
Dry matter	234.50	230.47	188.07	21.88	15.62	20.54
Nmin/N Ratio	0.04	0.10	0.18	0.63	0.62	0.75
C/N Ratio	14.70	18.54	20.11	6.72	5.68	5.06
Samples	90	79	48	202	211	160

Supplementary Table 8: Correlation between applied amounts of the different nutrients applied over all crops. The NOFERT treatment was omitted to avoid bias. *, **, *** indicates P<0.05, P<0.01, P<0.001

	TotN	MinN	P	K	Ca	Mg	OM
TotN		0.85*	0.93**	0.92**	0.74	0.87*	0.37
MinN	0.85*		0.91**	0.84*	0.67	0.75	-0.16
P	0.93**	0.91**		0.99***	0.86*	0.94**	0.08
K	0.92**	0.84*	0.99***		0.87*	0.96***	0.17
Ca	0.74	0.67	0.86*	0.87*		0.97***	0.06
Mg	0.87*	0.75	0.94**	0.96***	0.97***		0.20
OM	0.37	-0.16	0.08	0.17	0.06	0.20	

Supplementary Table 9: Spearman rank correlation between the estimated trends. NOFERT was removed to avoid overfit. None of the correlations were significant at P<0.05.

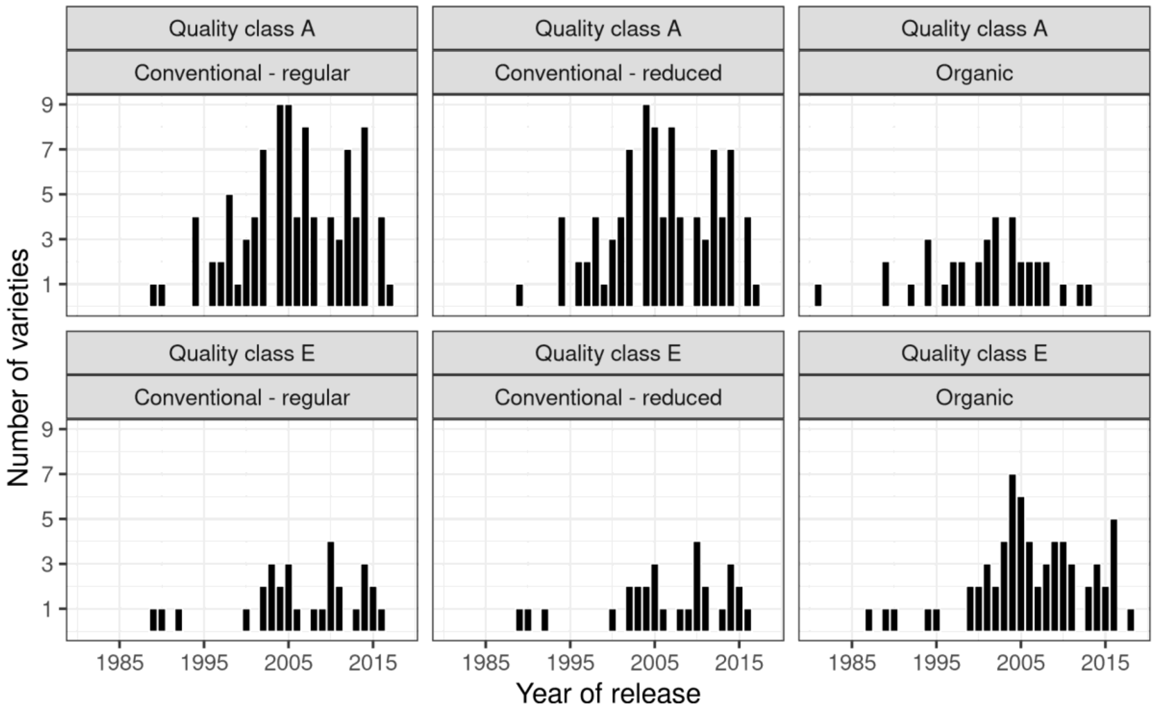
	Winter wheat	Potatoes	Grass-clover	Maize	Soybean
Winter wheat		0.07	-0.14	0.36	-0.11
Potatoes	0.07		0.68	0.11	-0.21
Grass-clover	-0.14	0.68		0.21	-0.54
Maize	0.36	0.11	0.21		-0.61
Soybean	-0.11	-0.21	-0.54	-0.61	

Supplementary Table 10: Estimates, standard errors (SE), pairwise differences (PD), for means, trends, and variances from the analysis on plot values without any correction structure (NC), compound symmetry (CS), and CS with heterogenic residuals per field x year combination (CSH) for winter wheat. P tests if the estimated trend is different from zero. r indicates the Pearson correlation of the respective estimates and SE from the analysis on plot level to the analysis on field x year means, Mean denotes the means of Estimates and SE per model, rho the estimated correlation of residuals between years, and AIC the Akaike information criterion.

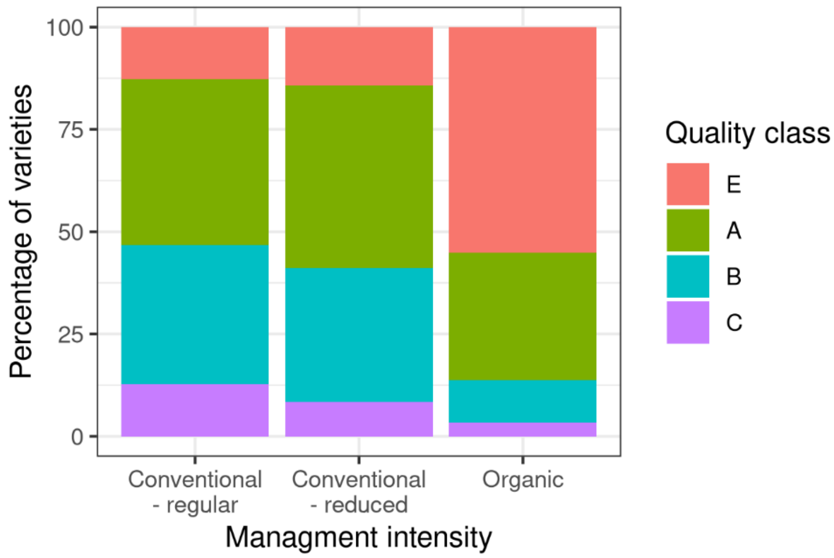
		NC				CS (rho=0.066)				CSH				On field x year means			
	Treatment	**Estimate**	**SE**	**PD**	**P**	**Estimate**	**SE**	**PD**	**P**	**Estimate**	**SE**	**PD**	**P**	**Estimate**	**SE**	**PD**	**P**
Means (t DM/ha)	**AIC (rho)**		-849				-860 (0.071)				-943						
	NOFERT	2.511	0.136	d		2.511	0.138	d		2.514	0.138	d		2.511	0.136	d	
	BIODYN1	3.679	0.136	bc		3.679	0.138	bc		3.672	0.138	bc		3.679	0.136	bc	
	BIOORG1	3.563	0.136	c		3.563	0.138	c		3.558	0.138	c		3.563	0.136	c	
	CONFYM1	4.708	0.136	a		4.708	0.138	a		4.708	0.138	a		4.708	0.136	a	
	BIODYN2	4.014	0.136	b		4.014	0.138	b		4.012	0.138	b		4.014	0.136	b	
	BIOORG2	3.964	0.136	b		3.964	0.138	b		3.962	0.138	b		3.964	0.136	b	
	CONFYM2	5.055	0.136	a		5.055	0.138	a		5.060	0.138	a		5.055	0.136	a	
	CONMIN2	5.039	0.136	a		5.039	0.138	a		5.048	0.138	a		5.039	0.136	a	
	r	1.000	0.250			1.000	-0.482			1.000	0.376						
	Mean	4.067	0.136			4.067	0.138			4.067	0.138			4.067	0.136		
Trends (t DM/ha/year)	**NOFERT**	-0.015	0.013	b	0.258	-0.015	0.013	b	0.256	-0.015	0.013	b	0.261	-0.015	0.013	b	0.262
	BIODYN1	0.016	0.013	ab	0.239	0.016	0.013	ab	0.240	0.016	0.013	ab	0.243	0.016	0.013	ab	0.244
	BIOORG1	-0.002	0.013	ab	0.871	-0.002	0.013	ab	0.869	-0.002	0.013	ab	0.861	-0.002	0.013	ab	0.871
	CONFYM1	0.015	0.013	ab	0.275	0.015	0.013	ab	0.275	0.015	0.013	ab	0.271	0.015	0.013	ab	0.280
	BIODYN2	0.024	0.013	a	0.071	0.024	0.013	a	0.071	0.024	0.013	a	0.076	0.024	0.013	a	0.076
	BIOORG2	0.004	0.013	ab	0.754	0.004	0.013	ab	0.757	0.004	0.013	ab	0.769	0.004	0.013	ab	0.755
	CONFYM2	0.030	0.013	a	0.024	0.030	0.013	a	0.023	0.030	0.013	a	0.023	0.030	0.013	a	0.027
	CONMIN2	0.019	0.013	ab	0.157	0.019	0.013	ab	0.156	0.019	0.013	ab	0.149	0.019	0.013	ab	0.162
	r	1.000	-0.786		1.000	1.000	0.957		1.000	1.000	0.916		1.000				
	Mean	0.011	0.013		0.331	0.011	0.013		0.331	0.011	0.013		0.332	0.011	0.013		0.335
Variances (t^2 DM/ha^2)	**AIC (rho)**		-1074				-1083 (0.067)				-1157 (not converged)						
	NOFERT	0.597	0.167	a		0.595	0.166	a		0.593	0.133	a		0.624	0.167	a	
	BIODYN1	0.532	0.149	a		0.533	0.150	a		0.520	0.118	a		0.559	0.149	a	
	BIOORG1	0.405	0.116	a		0.406	0.116	a		0.395	0.090	a		0.432	0.116	a	
	CONFYM1	0.378	0.108	a		0.363	0.104	a		0.355	0.081	a		0.405	0.108	a	
	BIODYN2	0.560	0.157	a		0.567	0.158	a		0.559	0.127	a		0.587	0.157	a	
	BIOORG2	0.529	0.149	a		0.531	0.149	a		0.521	0.120	a		0.556	0.149	a	
	CONFYM2	0.572	0.160	a		0.572	0.160	a		0.570	0.130	a		0.599	0.160	a	
	CONMIN2	0.663	0.184	a		0.668	0.186	a		0.652	0.147	a		0.690	0.184	a	
	r	1.000	1.000			0.999	0.999			0.998	0.997						
	Mean	0.530	0.149			0.530	0.149			0.521	0.118			0.557	0.149		

Chapter C

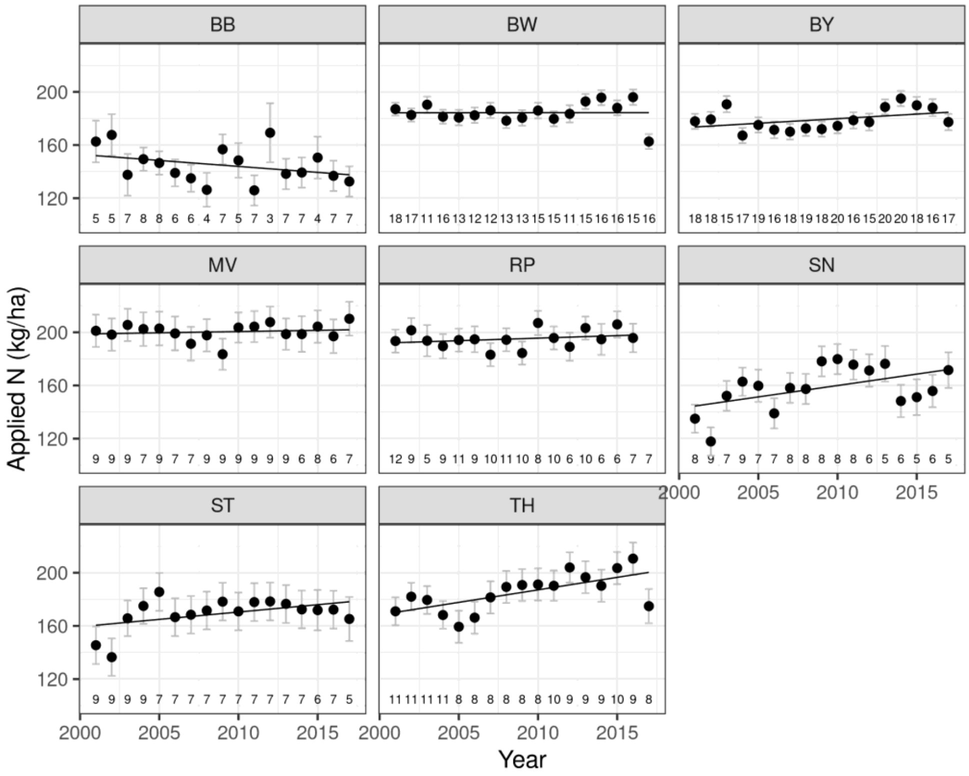
Supplementary Figures



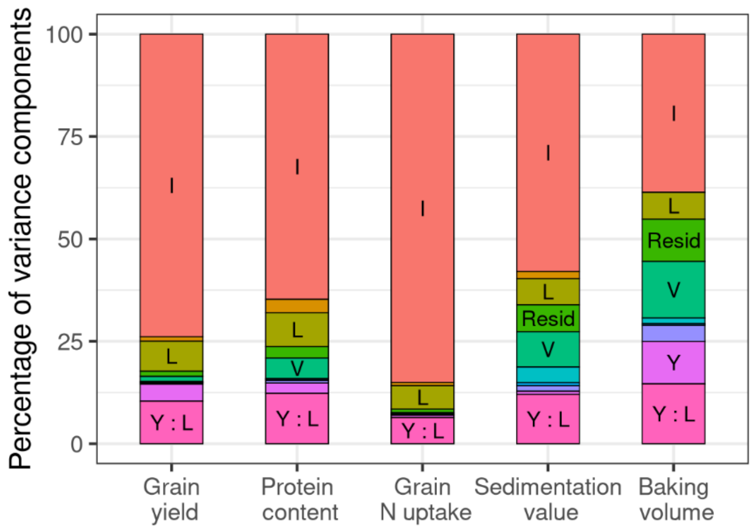
Supplementary Figure 20: Distribution of the year of release of the varieties included in the final dataset per quality class and for the three different management intensities. Data from trials in Germany, 2001 – 2017.



Supplementary Figure 21: Share of the varieties per quality class and per management intensity. Data from trials in Germany, 2001 – 2017.



Supplementary Figure 22: Average applied N per year for the federal states, for which applied N was available. Means were estimated with a mixed model with year as fixed and location as random. Errors bars are standard errors, numbers indicate the number of trials per year. The low value in BW in 2017 was due to one trial receiving only 85 kg N/ha, while the low value in TH in 2017 is due to several trials receiving less N. N applied was not available for NI, NW, and SH.



Supplementary Figure 23: Variance components expressed as percentage of the sum of all effects. Data from Bavaria, including conventional – regular and organic management intensity and varieties of quality class E and A. Only contributions with more than 5% are labelled. I: intensity, L: location, Y: year, Resid: residual. Data from trials in Bavaria, 2001 – 2019.

Supplementary Tables

Supplementary Table 11: Genetic, non-genetic, and overall trend when removing the last three years (2001 to 2014), respectively the first three years (2004 to 2017) from the dataset. Data from trials in Germany.

Time span of subset	Manage-ment	Quality class (breeding origin)	Number of observations	Number of varieties	Number of trials	Genetic trend b (SE)	Genetic trend P	Non-genetic trend b (SE)	Non-genetic trend P	Overall trend b (SE)	Overall trend P
2001 to 2014	Conv. – regular	E	3867	29	1137	29.7 (9.2)	0.001	46.6 (39.2)	0.235	89.1 (38.8)	0.022
		A	17417	90	1591	31.4 (4.6)	<0.001	41 (42.8)	0.338	77.7 (43.1)	0.072
	Conv. - reduced	E	3125	28	792	57.3 (10.4)	0.000	17.7 (35.4)	0.617	88.4 (34)	0.009
		A	11656	86	861	49.2 (5.9)	<0.001	29.9 (39.4)	0.448	81.7 (39.3)	0.038
	Organic	E	3456	59	387	5.9 (5.8)	0.309	8.1 (18.8)	0.666	16.3 (18.3)	0.375
		A	2277	36	387	38.3 (8.9)	<0.001	0.1 (21)	0.998	24 (19.8)	0.225
		E (conv.)	2647	42	387	16.5 (5.8)	0.004	1.5 (19.2)	0.937	18.6 (18.7)	0.321
		E (org.)	809	17	270	3.9 (11.4)	0.730	42.6 (21.8)	0.051	42.6 (20.9)	0.041
2004 to 2017	Conv. – regular	E	4004	30	1042	31.5 (9.9)	0.002	-5 (31.9)	0.876	43.3 (31.6)	0.171
		A	16794	91	1562	37.9 (4.8)	<0.001	-16.3 (36.4)	0.654	22.7 (36.4)	0.533
	Conv. - reduced	E	3293	29	736	71.3 (12.4)	<0.001	-66.5 (37.2)	0.074	34.1 (36.5)	0.350
		A	11178	88	822	57 (6.5)	<0.001	-36.3 (36.5)	0.320	24.3 (36.1)	0.501
	Organic	E	4034	62	395	6.7 (6.3)	0.286	2.4 (17.4)	0.888	9.6 (16.8)	0.568
		A	1877	35	391	37.8 (9.6)	<0.001	-16.6 (21.4)	0.438	18.8 (18.1)	0.299
		E (conv.)	2817	43	395	20 (6.1)	0.001	0.7 (19.5)	0.971	21.7 (18.9)	0.251
		E (org.)	1217	19	346	-1.7 (9.3)	0.857	35 (20.8)	0.092	31.1 (19.8)	0.117

Acknowledgments / Danksagung

An erster Stelle bedanke ich mich ganz herzlich bei Prof. Urs Schmidhalter, dass er mir während meiner Promotionszeit sehr viel Freiheiten ließ, mich aber stets bei allen bei meinen Vorhaben unterstützt hat, und sich immer die Zeit genommen hat, wenn ich mal nicht mehr weiterwusste. Prof. Kurt-Jürgen Hülsbergen möchte ich für die Übernahme des Zweitprüfers und bei Prof. Hans-Rudolf Fries für die Übernahme des Prüfungsvorsitzes danken.

Bei Peter Baresel möchte ich mich dafür bedanken, dass er mich im Rahmen des INSUSFAR Projekts nach Freising geholt hat. Hier geht mein Dank auch an alle beteiligten dieses Projekts.

Bei Marcel van der Heijden möchte ich mich für die Zusammenarbeit an der Meta-Analyse bedanken, und bei Paul Mäder und Jochen Mayer für die Bereitstellung der DOK Daten und ihre Unterstützung beim gesamten Vorhaben. Lucie Gunst danke ich für die immer sehr schnelle und kompetente Auskunft, wenn es Unklarheiten bei den Daten gab. Bei Lorenz Hartl bedanke ich mich für die Bereitstellung der Daten aus den bayerischen Landessortenversuchen.

Sven Böse danke ich, dass er mich als Mentor unterstützt hat, mir sehr interessante Einblicke in die Züchtungswelt und sehr wichtige Ratschläge für meine Auswertungen gegeben hat.

Obwohl aus meiner Feldarbeit in Dürnast wenig in die Dissertation eingeflossen ist, danke ich dem ganzen Dürnast Team für die Unterstützung und den Einsatz bei allen anfallenden Arbeiten. Besonderer Dank geht an Martin Köckeis, mit dem ich viele Stunden im Feld verbracht habe und ohne den die Versuche sicher nicht so gut gelaufen wären. Außerdem möchte ich mich bei Johann Huber und Matthias Parzefall für die Unterstützung im Rahmen ihrer Abschlussarbeiten bedanken.

Dem Sekretariat danke ich für die stets schnelle und unbürokratische Hilfe. Allen Doktorandinnen und Doktoranden am Lehrstuhl danke ich für die gemeinsamen Mittagessen, und dass wir uns immer austauschen konnten über Privates und Fachliches.

Vor allem möchte ich Gilgia danken, dass sie mich in allen Phasen der Doktorarbeit unterstützt

hat und mir immer sehr viel Rückhalt gegeben hat. Vielen Dank für alles!

Curriculum Vitae

Persönliche Daten

Name	Samuel Knapp
Geburtsdatum	6. Dezember 1983
Geburtsort	Esslingen am Neckar, Deutschland
Nationalität	Deutsch

Berufserfahrung

Seit Aug. 2020	**Getreidezüchter (Winterweizen und Hafer)**
	Nordic Seed Germany, Nienstädt
Apr. 2015 - Mai 2019	**Wissenschaftlicher Mitarbeiter - Doktorand**
	TU München, Lehrstuhl für Pflanzenernährung, Freising
Feb. 2017 - Mai 2017	**Forschungsaufenthalt**
	CIMMYT, Gruppe Weizenphysiologie, Ciudad Obregon, Mexico
Feb. 2013 - Mär. 2015	**Wissenschaftlicher Mitarbeiter**
	Agroscope, Institut für Pflanzenproduktion, Zürich, Schweiz
Sep. 2011 - Mai 2012	**Wissenschaftlicher Mitarbeiter - Masterarbeit**
	John Innes Centre, Department of Crop Genetics, Norwich, UK
Jul. 2005 - Aug. 2006	**Ausbildung Landwirt**
	Gut Rheinau, Schweiz
Sep. 2003 - Sep. 2004	**Zivildienst**
	Camphill Village West Coast, Kapstadt, Südafrika

Ausbildung

Okt. 2009 - Okt. 2012	**Universität Hohenheim**
	Master of Science, Crop Science
Aug. 2008 - Mai 2009	**Universität Kopenhagen, Dänemark**
	Erasmus Austauch, Agricultural Sciences
Okt. 2006 - Jul. 2008	**Universität Bonn**
	Vordiplom, Agrarwissenschaften