Field Crops Research 302 (2023) 109072

Contents lists available at ScienceDirect

Field Crops Research

journal homepage: www.elsevier.com/locate/fcr

Organic cropping systems maintain yields but have lower yield levels and yield stability than conventional systems – Results from the DOK trial in Switzerland

Samuel Knapp[a,c], Lucie Gunst[a], Paul Mäder[b], Shiva Ghiasi[a], Jochen Mayer[a,*]

[a] Agroscope, Dept Agroecology and Environment, Reckenholzstrasse 191, 8046 Zurich, Switzerland
[b] Research Institute of Organic Agriculture FiBL, Soil Science Department, Ackerstrasse 113, 8070 Frick, Switzerland
[c] Technical University of Munich, Chair of Plant Nutrition, Emil-Ramann-Straße 2, 85354 Freising, Germany

ARTICLE INFO

Keywords:
Yield trend
Long-term experiment
Nitrogen input
Crop protection

ABSTRACT

Sufficient and stable crop yields are the basis for feeding a growing world population. Limited cropland, climate change, degradation of soil quality and loss of biodiversity coupled with excessive use of non-renewable resources require new solutions for future cropping systems beyond existing management practices. Here we analyzed mean yields, temporal yield trends and the stability of organic and conventional cropping systems from the currently longest-lasting cropping system comparison, the DOK long-term systems comparison trial (DOK) comparing biodynamic, bioorganic and conventional cropping systems, over a period of 40 years. We used yield data of winter wheat, potatoes, grass-clover, maize and soybean in a seven-year rotation, where bioorganic and biodynamic farming systems were compared with conventional mixed and sole mineral fertilized systems. System treatments have been established at a reduced half and a regular fertilization level, which corresponds to standard Swiss farming practices. Yields were significantly lower in organic systems in non-legumes between 13% and 34%, depending on the investigated crop, whereas in legumes, no yield reduction was observed in soybean and only 10% was observed in grass-clover. Half the amount of fertilizer reduced yields by around 10% in all systems and crops. Applied mineral N determined yields mainly in winter wheat and potatoes. Temporal yield trends did not differ between organic and conventional systems, nor between half and regular fertilization over all crops. However, in winter wheat, both conventional and biodynamic management with regular fertilization showed a stronger temporal increase in yield, while yield of grass-clover under biodynamic management with half-fertilization decreased. Increased yield differences between systems in single years were due to poor performance of organic systems rather than better performance of conventional systems. Absolute stability (measured by the variance) did not differ, but conventional systems were more stable than organic ones in relative stability, measured by the coefficient of variation, expressing the stability in relation to the yield level. We found no difference in both absolute and relative stabilities between half and regular fertilization. Long-term organic management results in lower yields than conventional management, but not in a decrease of yields over time. The similarity in both stability measures between half and regular fertilization suggests that the variation in relative stability between organic and conventional management might be more related to plant protection than to fertilization intensity.

1. Introduction

As climate change will result in greater fluctuations and more extreme weather events in the future, agricultural production aims to be more resilient against such fluctuations to guarantee future regional and global food security (Howden et al., 2007). Thus, the challenge is to maintain or better increase productivity, with higher stability and with less negative environmental impact, i.e. achieving sustainability (Tilman et al., 2011).

Organic farming has been developed with the aim to reduce the negative impacts of intensified agriculture on the environment by avoiding the application of mineral fertilizers and synthetic pesticides.

---

* Corresponding author.
E-mail address: jochen.mayer@agroscope.admin.ch (J. Mayer).

https://doi.org/10.1016/j.fcr.2023.109072
Received 17 April 2023; Received in revised form 22 July 2023; Accepted 25 July 2023

However, because of excluding mineral fertilizers and synthetic pesticides, crop yields in organic farming systems are often lower than in conventional farming. When analyzing the yield gap between organic and conventional farming systems, several studies have found similar estimates of around 20% for the overall yield gap, while all of them have noted that the variation in yield gap is substantial between crops and regions (de Ponti et al., 2012; Seufert et al., 2012; Ponisio et al., 2015).

A temporal change in productivity can be evaluated by analyzing the yield development using regression analysis over time. The maintenance of productivity over time can be used as an indicator for the long-term sustainability of a certain management practice (Hejcman et al., 2012). As the majority of comparisons of organic and conventional farming systems are limited to short-term observations, we could not find any study investigating the long-term trends of productivity.

Originating in the field of plant breeding, stability analysis has turned increasing attention toward 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 variance and relative stability for the coefficient of variation. A particular focus of stability analysis has been on comparing conventional and organic management practices, and several studies have found that conventional management tends to have a higher relative yield 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 a greater amount of N fertilization can thus increase relative stability in organic farming.

The main reason for varying stability between management systems is that they react differently to yearly growing conditions, e.g. water availability or pest and disease pressure. Thus, the ratio between the yields of the management systems 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 point of analysis of temporal stability by correlating the yield ratio of each year to the yields of the respective treatments in comparison.

Long-term trials with consistent treatments over time offer a valuable source for determining long-term effects of different management systems on productivity, soil and the environment.

The DOK long-term trial was established in 1978 with the objective of comparing different farming systems including bio-Dynamic, bio-Organic, and “Konventionell” (DOK). However, it was not designed as a pure 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 units per area, while the fertilization intensity of the conventional system 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 45 years under practice, it is the world longest-lasting experiment comparing organic and conventional management, providing a unique dataset for analyzing long-term effects (Mayer and Mäder, 2016).

The objectives of this 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 during the course of the DOK long-term trial. In addition, we analyzed how mean yields, yield trends and stability were related to the amount of applied nutrients.

2. Material and methods

2.1. Description of the trial and measured 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 annual temperature is 10.9 °C (climate norm 1991 – 2020, see also yearly values in Fig. S11). 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 organic management, a biodynamic (BIODYN) system and a bioorganic (BIOORG) system were established. Both systems represent Swiss mixed farming systems with livestock, characterized by fertilization through farmyard manure and slurry, and mechanical weed control. Within conventional management, a mixed farming system combining manure and mineral fertilization (CONFYM) and a system with only mineral fertilization (CONMIN) were established. Both conventional systems received chemical plant protection, i.e. herbicides, insecticides and fungicides. The pesticides were continuously adapted to the currently recommended preparations. The organic treatments received also crop protection agents, mainly fungicides, insecticides and molluscicides that are approved for use in organic farming, like copper and bacillus thuringiensis preparations (Table 1, Table 2). The CONMIN treatment was introduced in the second rotation period from 1985; while in the first period, from 1978 to 1984, it was an unfertilized control treatment with the same plant protection scheme as CONFYM2 (Table 1).

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 an additional input to the nutrients already applied via 1.4 LU/ha. In addition to treatments with regular fertilization systems, BIODYN, BIOORG and CONFYM with half fertilization levels were included. The organic fertilization was performed with system specific manure types. All systems except CONMIN received a liquid manure fraction (slurry) and a solid manure fraction in addition: CONFYM received stacked farmyard manure, BIOORG rotted farmyard manure and BIODYN composted farmyard manure prepared with biodynamic preparations. Average applied nutrients per treatment are given in Table 1 (details table S1), and average nutrient contents of the applied organic fertilizers can be found in Table S3. Lastly, a NOFERT treatment, without any fertilization but with the same plant protection as BIODYN, served as control. Soil tillage, sowing and harvesting were conducted similarly in all treatments. For more detailed information, see Mäder et al. (2002) and Mayer et al. (2015).

The experimental design can be described as split-strip-plot with four replicate blocks (see Fig. S1). The main-plot factor corresponds to three shifted crop rotations, where the crop rotation was shifted by one and four years. The different rotations will be called fields here, in accordance with common long-term experiment terminology. Within main-plots, the horizontal factor is the system, with combined CONMIN and NOFERT, and the vertical factor is the fertilization level.

One crop rotation cycle lasted seven years and was the same for all treatments. To imitate common management practices in Switzerland, the crops and crop rotations along with their management, were adapted to current practice over the course of the experiment, while changes always occurred after the completion of one crop rotation cycle (Table S2). In the last three cycles the crops remained constant – although changing in order – and in the last cycle the crop rotation was silage maize, soybean, winter wheat, potato, winter wheat, and two

Table 1
Management of the DOK trial. Details of fertilization shows Table S1.

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**[a]	**NOFERT**
Average of yearly applied nutrients (kg/ha)	**Total N**[b]	48	95	48	96	86	171	121	0
	Mineral N	13	26	15	30	57	113	121	0
	P	12	24	12	24	19	37	38	0
	K	89	179	92	184	123	248	246	0
Organic matter		957	1911	1016	2032	1157	2314	0	0
Farmyard manure & slurry		Manure compost, slurry		Rotted manure, slurry		Manure, slurry		None	None
Mineral fertilization		none		Rock powder, 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		—		Plant growth regulators[c]			Biodynamic preparations

[a] The CONMIN2 treatment 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.

[b] Atmospheric N deposition was around 20 kg N/ha/year (Seitler et al., 2016), and not included in the N inputs. Data are available from 1990. Amounts were constant over time. Inputs from symbiotic N_2 fixation are not considered.

[c] Plant growth regulators are used to regulate straw height and strength of cereals by targeting the plant hormonal system.

Table 2
Applied amounts of pesticide active agents (average per year) over six crop rotation periods.

	BIODYN	BIOORG	CONFYM	CONMIN
	kg/ha/year			
Fungicide		0.31	1.25	1.28
Herbicide			1.32	1.05
Insecticide	0.06	0.05	0.15	0.09
Molluscicide	0.05	0.05	0.09	0.10
Seed dressing			0.01	0.01
Growth regulator			0.13	0.11
Pesticides total	0.11	0.41	2.95	2.64

years of grass-clover (see Table S2 for all crop rotation cycles). The varieties of the different crops also changed over the duration of the experiment to parallel common practice and to deal with possible breakdowns of resistances (Table S8). Varieties were always the same for all treatments (Table S8). The plot size was 100 m^2 (20 m x 5 m, Fig. S1).

Crop yield was determined by harvesting a central strip with a width of 1.5 m and 10 m length. Cereal straw was generally removed from the field. Crops residues from soybean, potatoes, cabbage and beetroots were left in the field. 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, according to Swiss analytical reference methods (Agroscope, 2023). The ammonium-N fraction of farmyard manure and slurry was determined by alkalinization and steam distillation (Agroscope, 2023). The nitrate-N fraction in solid farmyard manure was determined by $CaCl_2$ extraction and further ICP-OES measurement (Agroscope, 2023). The respective nutrient contents of the applied mineral products were taken from the product descriptions.

2.2. *Statistical analysis*

2.2.1. *Data pre-treatment and quality check*

To remove carry over effects from the management until 1977 before the start of the experiment and because CONMIN2 was introduced 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 in the third rotation cycle was between 30% and 50% lower than the first year (compared to a reduction of 9–18% in the second year compared to the first year), 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 in the analysis. Although other crops were grown during the experiment, we only investigated the crops winter wheat, potatoes, grass-clover, maize, and soybean, in order 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 analysis with the factors *replicate block* and *treatment* for each field x year combination. We assessed normal distribution of the residuals using the Kolomogorov-Smirnov (KS) test, the coefficient of variation (CV) as the square root of the residual error variance divided by the overall mean, and the significance of the treatment effect by ANOVA. The KS test was never significant, the 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.

2.2.2. *Testing correlation structure*

Since in long-term experiments plots are sampled every year, observations could thus be correlated between years (Richter and Kroschewski, 2006). To test if such correlation structures are present, we first compared different models that take this correlation structure into account and tested if residuals from a randomized complete block (RCB) model (*block* and *treatment* effect) within each year were correlated across years (see Supplementary Method 1 for the detailed description of this analysis and also the results in the supplemental material). This analysis showed that estimates, standard errors and pairwise differences only marginally differed between models. Taking account of correlation structures and an analysis on simple field x year means not correcting for such a correlation structure (Table S6). Furthermore, residuals of the different years were not correlated, and the correlations of residuals did not decrease over time (Fig. S3), which has also been found by Richter and Kroschewski (2006). Based on these results, 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 replicate blocks. In cases where models on the plot level converged, we compared statistics to models on the mean level, and found no differences in statistical outcomes (standard errors and pairwise differences).

2.2.3. *Mean yield and yield development*

All statistical analyses were conducted in R (R Core Team, 2014) and mixed models were fitted using the R-packages *lme4* (Bates et al., 2015) and *sommer* (Covarrubias-Pazaran, 2016). We estimated treatment means and the yield development for the eight different treatments with the following model for each crop separately:

$$\text{YIELD} \sim \text{T} + \text{Yn} + \text{T:Yn} + \underline{\text{F:Y}} + \text{F:Y:T} \quad (1)$$

where YIELD is the mean yield over replicates per treatment and per field x year combination (see before) and per crop, 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. The model was run for each crop separately. We did not include a field main effect, as the field effect is confounded with the year effect due to the shifted rotations across fields. The combined effect of year and field was necessary, as 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). Estimated marginal means, coefficients for temporal trends, 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 of any contrast. A letter display indicating significance of multiple pairwise differences ($\alpha = 0.05$) 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 2–4), to avoid any effects from changes in preceding crops on the estimated yield trends.

2.2.4. *Relation to applied nutrients*

To investigate any relation between the average amount of applied nutrients over all crops (as shown in Table 1) and 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 more interested in the differences between the fertilized treatments occurring in practice.

2.2.5. *Yield stability*

Although absolute stability can equivalently be assessed by standard deviation or variance (as variance is the square root of the standard deviation), 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 temporal trend could lead to an increased estimate of the variance, we estimated the variance to measure absolute stability with the following mixed model:

$$\text{YIELD} \sim \text{T} + \text{Yn} + \text{T: Yn} + \text{VS(T,F:Y)} \quad (2)$$

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 symbolize a diagonal variance matrix with the diagonal elements being the absolute 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 were 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 is the standard deviation divided by the mean. Here, the CV to measure relative stability was calculated by dividing the square root of the estimated variances from model (2) by the estimated means from model (1). SEs were calculated as $\left(\frac{CV^2}{2n} * \left(1 + 2 * CV^2\right)\right)^{0.5}$, where 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.

To compare effects of treatment groups, e.g. organic vs. conventional management, 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 the null and the contrast model, 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 (Fig. S9).

As an additional analysis of stability, we investigated whether an increased yield ratio between the organic and conventional systems in certain years is due to a poorer performance of the organic systems or a better performance of the conventional systems. Using only the observations from regular fertilization, we first calculated the mean yield of both organic systems (BIODYN2 and BIOORG2) and the mean yield of both conventional systems (CONFYM2 and CONMIN2) for each year and each crop. To test the relation of the yield ratio both organic and conventional management on the performance of either system, we regressed the yield ratios of the different years (conventional divided by organic) on the organic mean yields and the conventional mean yields over years, and for each crop separately.

3. Results

3.1. *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 (Fig. 1 and Fig. 2). 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% CI: 63%– 69%] of the conventional yield under regular fertilization,

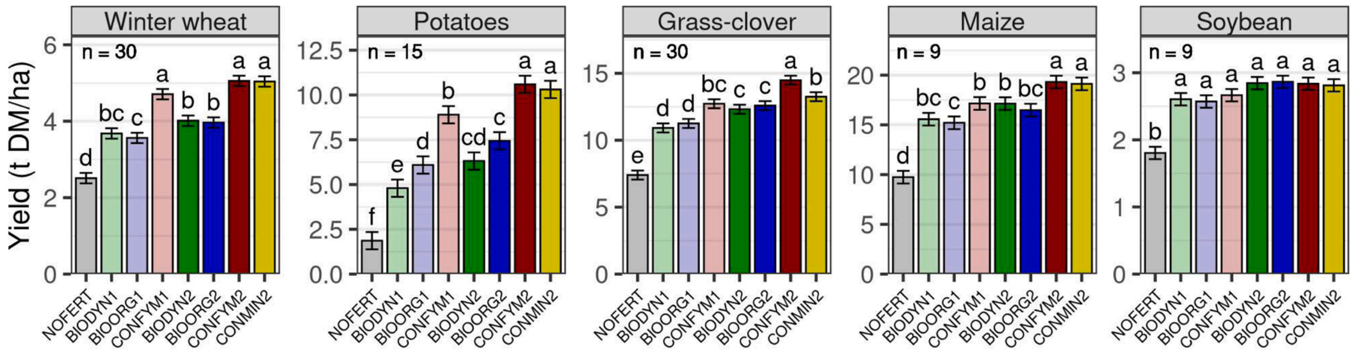

Fig. 1. Treatment means for all treatments and investigated crops. Error bars represent the standard error of the mean; n indicates the number of field x year combinations used for the calculation of the mean yields. Treatments that do not carry the same letters are significantly different at $P < 0.05$.

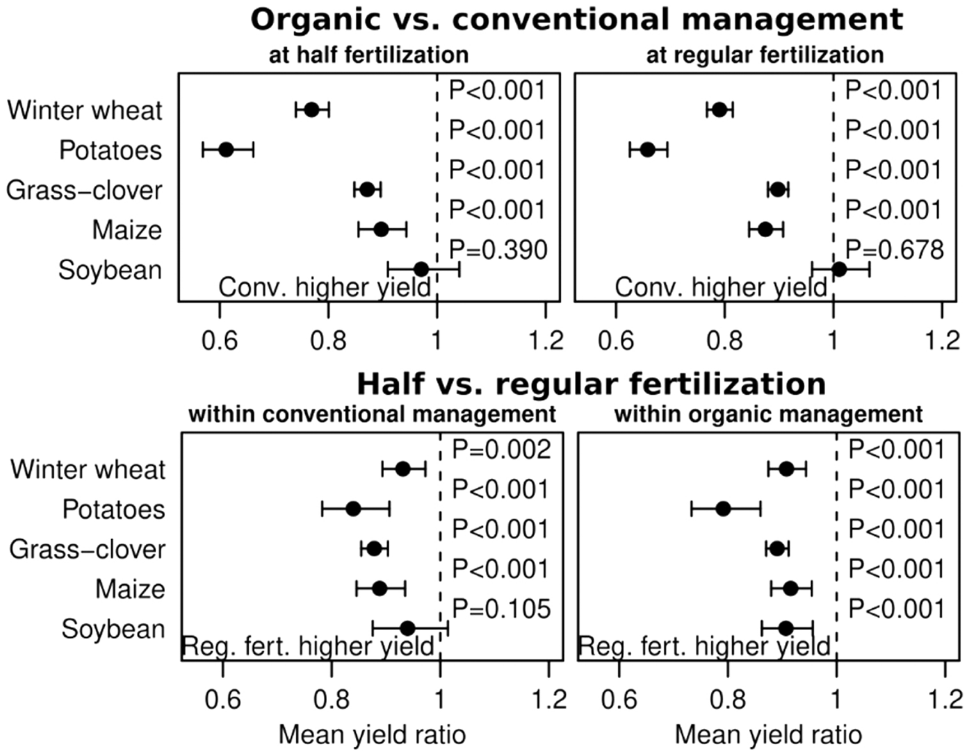

Fig. 2. The effect of organic vs. conventional management (top) and of *half* vs. *regular* fertilization (bottom) on the mean yield. Effects of organic vs. conventional were compared under *half* (left; BIODYN1, BIOORG1 vs. CONFYM1) and *regular* (right; BIODYN2, BIOORG2 vs. CONFYM2, CONMIN2) fertilization, 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 of the mean yield ratio do not overlap 1.

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).

The treatments with half fertilization showed a significant reduction in yield across all crops compared with their corresponding treatments with regular fertilization (Fig. 1). Interestingly, this yield reduction was of equal magnitude within organic and within conventional management (Fig. 2). 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%.

No significant yield differences were observed, within the organic (BIOORG vs. BIODYN) and conventional (CONFYM vs. CONMIN) systems, in all crops and at both fertilization levels except in potatoes and grass-clover (Fig. 1). In potatoes, BIODYN showed lower yields than the BIOORG at both fertilization levels (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 whether 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 for soybean. The average amount of applied mineral N showed a considerably higher relation to mean yields in winter wheat and potatoes with R^2 of 0.96*** and 0.95***, respectively than other nutrients (Table 3 and Fig. 3). Interestingly, in grass-clover total nitrogen showed a better relationship than mineral N (R^2 of 0.95 versus 0.80) and a square root function did fit better than a linear

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. Displayed values are adjusted R^2. A linear (l) and a square root (r) regression function excluding the NOFERT treatment (n = 7) was compared, and the better fit was chosen by adjusted R^2 (indicated in parentheses). For abbreviations of nutrients, see Table 1.

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)

function. In maize, all nutrients except Ca showed a relation to the mean yield. 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 relationship. However, for the interpretation of the observed correlations, it has to be noted that due to the design of the experiment, applied nutrients were highly correlated between each other (Tables 1 and S4).

3.2. *Yield development*

The overall trends were positive for all treatments in winter wheat, potatoes, and soybean, but negative in grass-clover and maize (Fig. 4). As the preceding crops of winter wheat and potatoes changed during the experiment, we additionally 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. 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

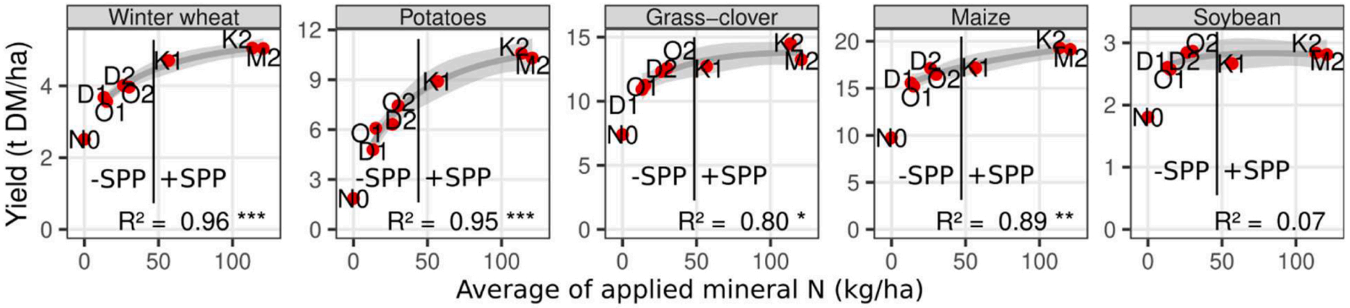

Fig. 3. Regression of the mean yield of the treatments on the average amounts of applied mineral N through a square root function. R^2 indicates adjusted R^2. NOFERT (N0) was excluded for the regression fit to avoid an overestimation of the fit statistic (R^2), but included in the plot for comparison. The grey area around the regression line represents the 95% prediction CI of the regression. 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.

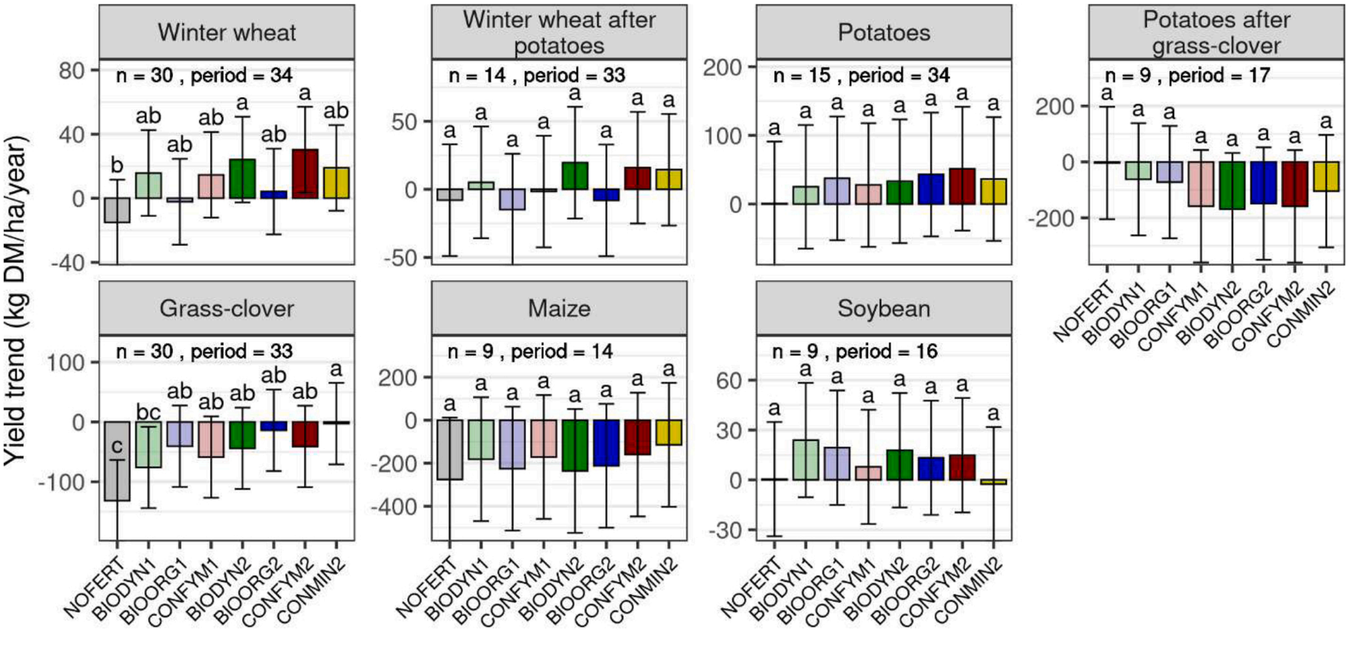

Fig. 4. Estimated yield trends per crop and treatment. As preceding crops have changed during the experiment, the yield trends were additionally analyzed in winter wheat only after potatoes and in potatoes only after grass-clover. Values greater than zero denote yield increase, and smaller than zero yield decrease. n indicates the number of field x year combinations, and period shows the number of years between first and last year of observations. 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 $P < 0.05$.

except soybean the organic treatments ranked between NOFERT and the conventional treatments, they showed a greater yield increase than the conventional treatments in soybean. A similar yield change was 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 (Table S5).

To test any differences in yield trends between treatment groups, i.e. between organic and conventional management and between half and full fertilization, we calculated linear contrasts (Table 4). Between organic and conventional management, the most significant difference was observed in winter wheat with organic treatments showing a smaller increase ($P = 0.07$). Winter wheat after all preceding crops increased stronger at full vs. half fertilization ($P = 0.1$). This effect was pronounced when analyzing only winter wheat after potatoes ($P = 0.05$) where yield at half fertilization resulted in a decrease. 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 ($P = 0.03$).

As applied rates of fertilizer were not constant during the experiment (particularly in the conventional treatments in winter wheat, see Fig. S2), we also regressed the estimated yield trends on the change of mineral N per year. However, there were never any significant relationships (Fig. S4).

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 slopes between the compared groups, SE the standard error, and P the significance level 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).

Crop	Contrast (group 1 - group 2)				Group 1	Group 2
	$\Delta b = b_1 - b_2$	SE	P		b_1	b_2
	(kg/ha/year)				(kg/ha/year)	
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**	**Regular**
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

3.3. *Yield stability*

To identify the effects of conventional vs. organic management and of the fertilization levels, we calculated linear contrasts of the stability measures (Fig. 5, see Fig. S8 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, as estimated by the variance, for all investigated crops. However, relative stability, as estimated 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 that of conventional management, indicating a better stability for conventional management.

When comparing the two fertilization levels, half fertilization showed 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 (Fig. 6). 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.

As we have found that the yield ratio between organic and conventional management at regular fertilization varied substantially between years (see variation on x-axes values in Fig. 7), 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 (Fig. 7). 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 increased yield difference (as indicated by a smaller yield ratio) was always 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 ($P > 0.05$).

4. Discussion

We investigated mean yield, temporal yield development, and temporal yield stability of long-term organic vs. conventional management and different fertilization intensities. Mean yields reflected the overall observations of organic vs. conventional yield gaps with 10–30% lower organic yields and, as expected, relative yield stability over time was lower in organic management. However, surprisingly yield development showed only small differences between managements systems but was different among crop types. It increased significantly for winter wheat, potatoes and soybeans, but decreased for grass-clover and maize. The comparison of full vs. half fertilization intensity revealed a strong yield dependence of nitrogen fertilization whereas crop protection determined mainly yield stability.

4.1. *Mean yield*

The organic-conventional yield gap has been intensely studied in the

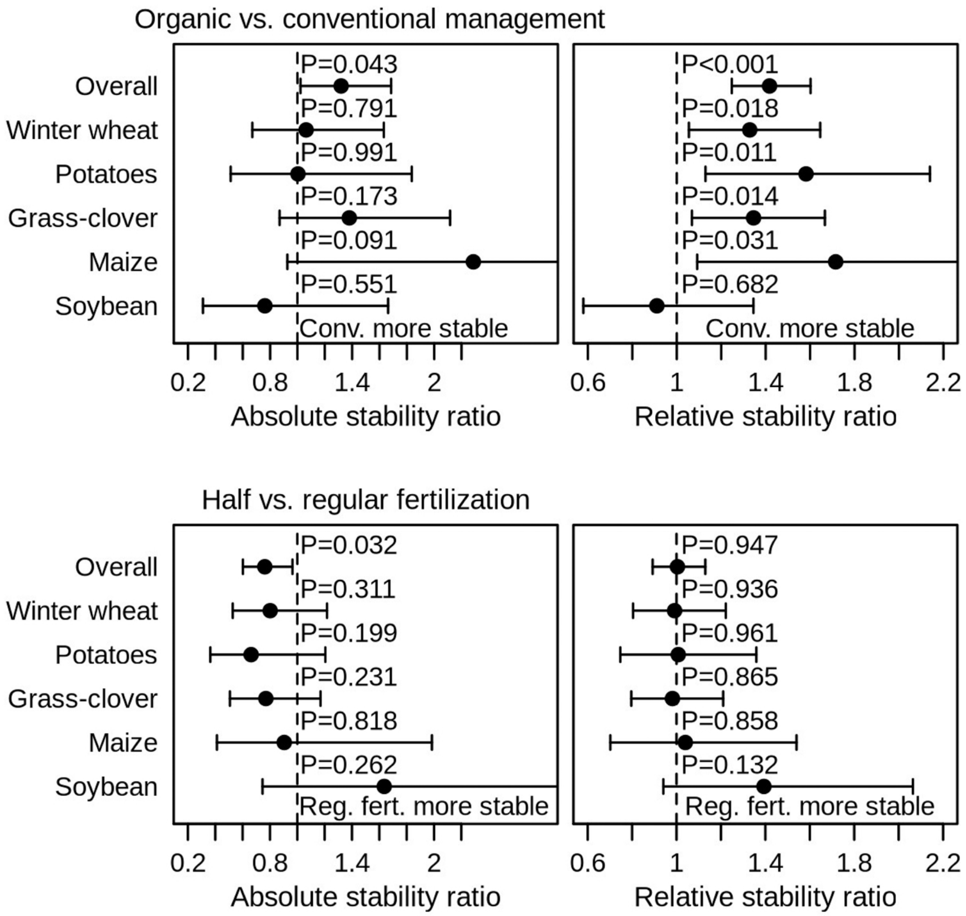

Fig. 5. Contrasts of absolute (left) and relative (right) stability comparing organic vs. conventional management (top; BIODYN1, BIOORG1, BIODYN2, BIOORG2 vs. CONFYM1, CONFYM2) and half vs. regular fertilization (bottom; BIODYN1, BIOORG1, CONFYM1 vs. BIODYN2, BIOORG2, CONFYM2). Ratios were calculated as the respective stability measure of organic to conventional management and half to regular fertilization. A ratio of 1 indicates that the stability is the same between groups, and > 1 indicates that conventional management, regular fertilization is more stable. Error bars are 95% CI of the ratio and error bars not overlapping 1 indicate that the ratio is significantly different from 1 at P < 0.05.

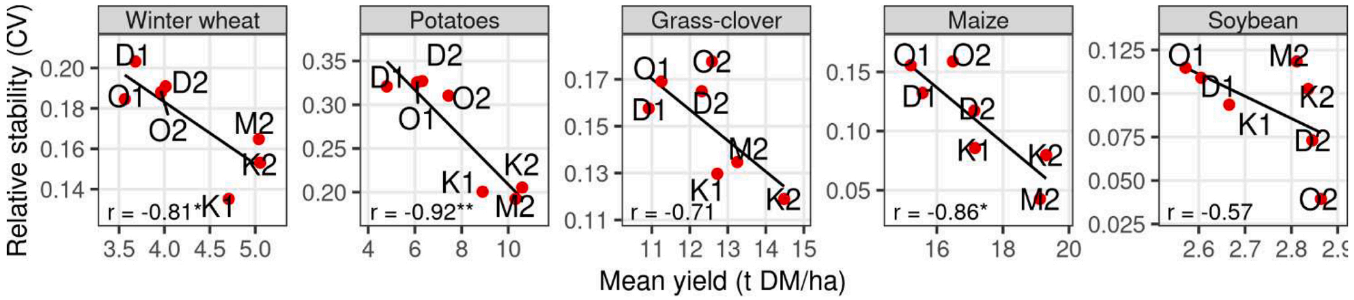

Fig. 6. Relation of relative stability (CV) and mean yield per treatment. Lower numbers of relative stability express better stability. r is the Pearson correlation coefficient, * and * * indicate significance at P < 0.05, and P < 0.01, respectively. NOFERT has been omitted, as it is distant to other treatments and could thus lead to overestimation of the correlation.

last decade (de Ponti et al., 2012; Seufert et al., 2012; Ponisio et al., 2015). In these studies, organic management reached average yields between 75% and 81% of conventional depending on region and crop type. The DOK results represent very well the observations made in the meta-analyses with 85% of conventional yields across investigated crops and both fertilization levels. In line with the meta-analyses, the highest yield gap was observed in non-legumes, mainly in potatoes with 66% and winter wheat with 79% of conventional yields at regular fertilization (Fig. 2). Organic winter wheat yields in variety testing trials in Switzerland were 62% and 77% of conventional management with high and low inputs, respectively (Herrera et al., 2020). The Swiss on-farm yield gap is also within this range, with 70% for winter wheat and 63% for potatoes (Rudmann and Willer, 2005). However, in high yielding regions like Germany, the on-farm yield difference in cereals and potatoes was found to be around 50% (BLE, 2018; Hülsbergen et al., 2022). The differences in the yield gap are due to management restrictions in organic farming and the intensification level of conventional systems (de Ponti et al., 2012; Döring and Neuhoff, 2021). Compared to other European countries, Swiss conventional cropping systems, in particular the wide-spread integrated production systems, are restricted in fertilizer amounts and pesticide use (Richner and Sinaj, 2017), which results in a relatively smaller yield gap compared to e.g. Germany. However, as organic farming relies on biological nitrogen fixation and fodder production for animals in the case of manure fertilization, organic yields of non-legume crops will be limited by N availability and the area for legume cropping and fodder production should also be taken into account when calculating the organic to conventional yield gap (Connor, 2018; Döring and Neuhoff, 2021). On the

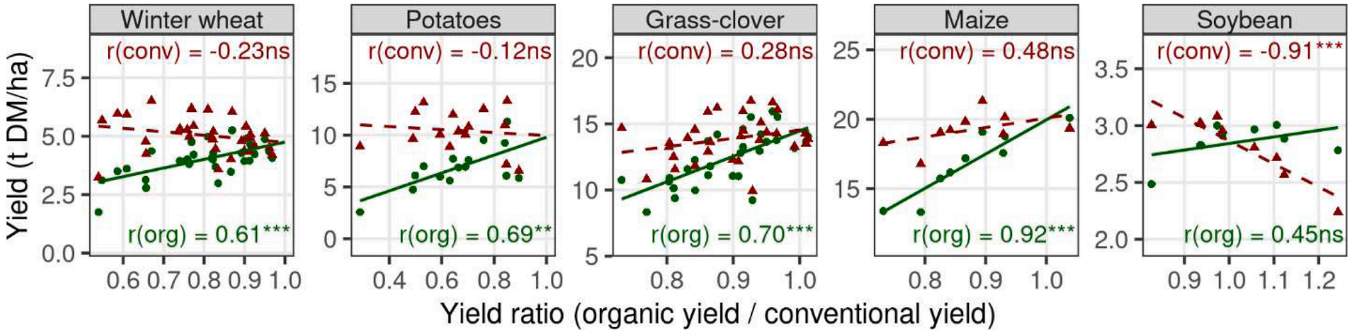

Fig. 7. Relation of yield ratio between organic and conventional management to the mean yield (organic: BIOORG2 and BIODYN2, conventional: CONFYM2 and CONMIN2) of organic treatments (green dots and solid line) and conventional treatments (red triangles and dashed line), respectively. One dot represents one year x field combination and r indicates the Pearson correlation coefficient. * * and * ** indicate significance at $P < 0.01$, and $P < 0.001$, respectively.

other hand, legumes, which do not depend on external N sources, reached comparable yields in soybean (100% of conventional) and were only slightly reduced in grass-clover (90% of the conventional) in the DOK trial. When comparing overall productivity, it also needs to be noted that the crop rotation in the DOK trial was similar in all systems. In practical farming, crop rotations often differ between organic and conventional systems, e.g. grass-clover is not very common in conventional rotations. However, in Switzerland, grass-clover is very common in conventional mixed-cropping farms to provide a fodder base for dairy cattle.

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, as assessed by R^2 (Table 3, Fig. 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 each other (Table S4). Nonetheless, mineral N was considerably better related to yields in winter wheat and potatoes than other nutrients. This strong relation suggests that the amount of applied mineral N is the primary reason for the observed yields in these species (Finckh et al., 2006; Seufert et al., 2012; Ponisio et al., 2015; Mayer et al., 2015). Available soil nutrients decreased largely for P in all treatments and for K in NOFERT, BIOORG1, BIODYN1 and CONFYM1 (Table S7). However, up to the end of the study period in 2019 yield limitations were observed only for K in NOFERT. Initial soil available P contents were very high in the DOK and did not lead to any detectable limitation during 40 years (Gunst et al., 2013; Hammelehle et al., 2018).

Mineral fertilizers allow a better in season distribution and thus improve the synchronization between N fertilizer supply and crop N demand. The yield gap between CONFYM1 and BIOORG2 / BIODYN2 clearly demonstrated these limitations of organic systems here: CONFYM1 received less total N amounts (and other nutrients), but more mineral N fertilizers (Table 1) and reached 18% and 29% higher yields in wheat and potatoes, respectively (Fig. 1, Fig. 3). Furthermore, in organic potatoes, the control of *Phytophtora infestans*, *Alternaria solani,* and the Colorado potato beetle (*Leptinotarsa decemlineata*) is challenging and limits yields (Finckh et al., 2006; Möller et al., 2007; Kühne et al., 2008; Runno-Paurson et al., 2014). An evaluation of yield drivers in organic potato management in Northeast England supports these findings and shows the strong impact of N supply on organic potato yields, which contributed more than limited crop protection (Palmer et al., 2013). Although BIODYN and BIOORG systems differ in the form of solid manures applied (manure compost vs slightly rotted manure) and additional biodynamic preparations in BIODYN, we could not detect any yield differences between these two management practices, except in potatoes. The higher yields of potatoes in BIOORG can be explained by the application of copper fungicides to control *Phytophtora infestans*. This is supported by the fact that in five years where no copper was applied to BIOORG no significant yield difference was observed ($P = 0.93$), while in the remaining ten years BIODYN yields were significantly reduced by 20% ($P = 0.001$, see Fig. S7). Bilsborrow et al. (2013) analyzed the yield effects of organic vs. conventional fertilization management and crop protection in winter wheat and found both measures to be of equal magnitude, but disease control also interacts with N fertilization, resulting in stronger effects of N fertilization under disease control (Berry et al., 2010). The yield gap between CONFYM1 and BIODYN2/BIOORG2, with different amounts of N forms and crop protection (Table 1) supports this strong combined effect.

In maize, clover-grass and soybean organic systems achieved yield levels higher than 87% of conventional (Fig. 1), which underlines a smaller dependence on external fertilizer N, and a higher relevance of other nutrients namely phosphorous and potassium. A first potassium limitation in the DOK trial has been previously observed in grass-clover (Oberson et al., 2013; Hammelehle et al., 2018). Surprisingly, half compared to regular fertilization reduced yields by only 10% over all treatments (Fig. 2) and this reduction was similar under organic and conventional management. Mayer et al. (2015) found for winter wheat in the DOK that the doubling of the N application from half to regular fertilization increased wheat yields in average by only 16%. They explained this observation by long-term integrating system effects such as soil nutrient availability, other soil fertility factors and large N inputs by legumes as part of the crop rotation (Oberson et al., 2007, 2013; Hammelehle et al., 2018).

4.2. *Yield development*

To evaluate the long-term effects of the cropping systems on yield, we regressed yields on the year. Overall, the estimated yield trends hardly differed significantly from zero (Fig. 4). The absence of significant effects might partly be due to the great variations in yield across years (see also Fig. S5), which are common in yield observations. It leads to rather high standard errors of the estimates, which in turn lowers the ability to find significant increases or decreases in yield.

Conventional winter wheat yields showed an overall positive yield trend. This was also true when only considering wheat following potatoes, to avoid any effects from changing preceding crops. The mean trend of the regularly fertilized conventional treatments with potatoes as preceding crop was 15 kg/ha/year. As winter wheat varieties were changing during the experiment (6 different varieties from cycle 2), the increase in yield over time might have been due genetic breeding progress. However, the trend is considerably smaller than the estimate from Laidig et al. (2014) of 70 kg/ha/year as observed in German variety trials from 1983 to 2012 or of even 110 kg/ha/year from Rueda-Ayala et al. (2018) as observed from 1953 to 2009 in a long-term trial in Germany. On the other hand, the small estimate and its non-significance is in agreement with the fact that conventional on-farm yields of winter wheat in Switzerland are stagnating since around 1990 (Erdin, 2018; Herrera et al., 2020). The same observation was noticed in potatoes, where Erdin (2018) observed a stagnation of conventional on-farm yields since 1990. Beside genetic improvements a second factor of the observed increase might be an increase in fertilization level during the experiment, particularly in winter wheat in the conventional treatments, although there was no significant relationship between the estimated trends and the increase in applied mineral N (Fig. S4). Mean annual temperature increased by around 1.5 °C during the experiment duration, while precipitation remained constant and global radiation increased slightly during the experiment duration (Fig. S11). While warmer temperature is often predicted to reduce lower wheat yields, increased global radiation and CO_2 levels can lead to increased wheat yields (Asseng et al., 2015; Holzkämper et al., 2014).

The analysis of differences between groups of treatments in winter wheat revealed three significant contrasts: (1) conventional management (CONFYM and CONMIN) showed a more positive trend than organic management, (2) regular fertilization showed a more positive trend than half fertilization, and (3) BIODYN treatments showed a more positive trend than BIOORG treatments (Table 4). While the first two might be related to the overall fertilization intensity, the latter appears more difficult to interpret. 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 1) and rotted manure in BIOORG vs. composted manure in BIODYN (2032 kg/ha/year OM vs 1911 kg/ha/year, respectively). The higher Ca rate is mainly due to the origin of the BIODYN compost from a farm on a more calcareous soil and thus higher external inputs of $CaCO_3$. In combination with similar inputs of stabilized composted OM in BIODYN (94% of BIOORG), soils had a substantially higher Ca content which resulted in increasing trends of soil pH value (Fig. S10) and soil organic carbon content (Leifeld et al., 2009), as well as a higher aggregate stability (Mäder et al., 2002). A combination of these factors might have led to higher mineral soil N contents in spring in BIODYN vs BIOORG and indicate the build-up of a higher N mineralization potential in BIODYN. That might explain a shift from lower to higher wheat yields in BIODYN than in BIOORG over time (data not shown).

When analyzing only potatoes succeeding grass-clover, the yields were decreasing considerably in all regularly fertilized treatments (Fig. 4). While we have no data on the occurrence, this overall decrease might have been due to wireworm (*Agriotes* spp.) infestation, which is particularly increased when potatoes are grown after grass-clover (Parker and Howard, 2001). In addition, the high proportion of cereals and grass-clover (5 out of 7 fields) provide advantageous conditions for wireworm development. The overall positive yield trend, when analyzing potatoes after all preceding crops, was due to the effect of soybean as preceding crop in cycle 5, which resulted in higher yields in all crops (Fig. S6), and thus representing a recovery from depressed yields due to the wireworm infestation.

The yield of grass-clover was decreasing in all treatments over time (Fig. 4). 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 that 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 sulphur containing fertilizers. The stronger yield decrease in BIODYN than in BIOORG could have been due to BIOORG receiving occasional applications of potassium sulfate in cycles 3–5 with around 40 kg S/ha per application. Sulphur fertilization has not been considered necessary up to the 1990 s 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 to less than 1/6 of the emissions in1980 (BAFU, 2020), which in turn led to a strong 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 is recommended (Webb et al., 2016).

4.3. *Yield stability*

In the comparison of organic vs. conventional management, we found no difference in absolute stability (as measured by the variance). However, 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 (Fig. 2). The finding of similar absolute stability but different relative stability is in agreement with the meta-analysis of Knapp and van der Heijden (2018) and Smith et al. (2019). Both studies argue that increased N fertilization can help to improve 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 managements can be attributed to differences in mean yield (Fig. 6), 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 related to 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, in contrast to what has been found by Knapp and van der Heijden (2018) and by Smith et al. (2019). Plant protection and weed control was supported by much larger quantities of pesticides in the conventional systems that received pesticide active agents of 2.8 kg ha^{-1} $year^{-1}$ whereas organic systems received only 0.26 kg ha^{-1} $year^{-1}$ (Table 2). In consequence weed seed banks in the DOK show a higher abundance and species richness in organic systems. Conventional systems selected for nitrophilous species and against herbicide-susceptible species (Rotchés-Ribalta et al., 2017, 2020). Wheat fungal disease infection was elevated in some cases in organic systems. In the study of Gunst et al. (2006) the infestation with *Septoria tritici*, *Septoria nodorum* and *Septoria nivale* was two to three-fold in organic systems, whereas other fungal diseases did not show significant differences among systems. Particularly, in potatoes relative stability is equal within all organic and within all conventional treatments, which furthermore supports the idea that there is little effect of the fertilization intensities on relative stability (Fig. 6). In turn, this suggests that relative stability in potatoes is strongly determined by plant protection. Surprisingly, relative stability in potatoes was similar between BIODYN and BIOORG, although they differed in crop protection, i.e. in the application of copper, which did result in significant yield reduction in BIODYN in certain years (Fig. S7).

We found a considerable variation in the yield differences between conventional and organic managements across years (Fig. 7). While in some years, organic treatments yielded almost similar to conventional, in other years, yields were only half of conventional. We therefore conducted a regression of the yield ratio on treatments' yields as an in-depth analysis of stability. We found a strong correlation between the yield ratios 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.

5. Conclusion

In the world's longest lasting experimental dataset yield reduction in non-legumes of organic systems varied considerably between 13% and 34% compared to conventional systems. However, in legumes, no yield reduction was observed in soybean and only 10% in grass-clover. Winter wheat and potato yields were strongly related to the amount of applied mineral N forms, pointing towards the importance of plant available N for reducing the organic vs. conventional yield gap in non-legumes. Long-term organic management did not lead to any yield decline compared to conventional management. In accordance with previous studies, absolute stability (measured by the variance) was similar between both systems, but relative stability (measured by the CV) was more stable in conventional systems. As relative stability corrects for the difference in mean yield, the better stability of conventional systems was due to their higher yields. However, the fact that half fertilization in both systems did not result in reduced relative stability, suggests that the lower relative stability in organic systems might be to a large extend due to less effective options of weed, pest and disease control. Consequently, significant differences in yield were due to poor performance of organic rather than better performance of conventional systems. We want to stress that the presented results are based on one single experiment. Thus, results should only be generalized with care and further experiments are encouraged in order to gain additional insights in other soil types, crop rotations, climates and regions.

6. Future prospects

The conventional system with half fertilization (CONFYM1) gained higher or similar yields and higher stability over all crops compared to the regular fertilized organic systems but it received fewer absolute amounts of nutrients with fertilizers. However, it received a higher amount of mineral N forms and it was treated with pesticides. Hence, the main drivers to reduce the yield gap and improve yield stability are an improvement of nitrogen availability and synchronization between supply and crop demand. Further improvements in weed control by new robotic technologies and crop protection by cultivars that are more resistant or by crop diversification will be a key measure of future management. New breeding technologies like gene editing which are controversially discussed in organic associations should be critically reviewed. Since the N nutrition in organic cropping is largely limited by the potential of symbiotic N_2 fixation and the land demand for legumes (Döring and Neuhoff, 2021) crop yields in organic cropping systems can be improved if the system boundaries will be redefined in a way without calling into question the basic idea of Organic Agriculture. That can be achieved if the idea of the closed nutrient cycles on farm level will be extended to a regional level with the aim to close cycles in the context of an urban – rural relationships. The use of a limited amount of mineral N forms e.g. from human urine collection or separation from sewage sludge have the potential to close the N gap in Organic Agriculture and to reuse other nutrients. Further, processing of liquid manure is obligatory to reduce ammonia losses and keeps N in the system. Stripping technologies producing "farm ammonium sulphate" or modern N separation technologies can help to reduce the mismatch of crop N demand and supply.

CRediT authorship contribution statement

Samuel Knapp performed statistical analysis and wrote the manuscript. Lucie Gunst managed data collection and quality control. Shiva Ghiasi performed evaluations on disease infestation and did proof reading. Jochen Mayer supervised the analysis and the concept of the paper. Paul Mäder and Jochen Mayer conducted the experiment, provided data, and contributed to the writing of the manuscript.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests Jochen Mayer reports financial support and administrative support were provided by Agroscope.

Data availability

Data will be made available on request.

Acknowledgments

We are very thankful to Urs Schmidhalter for the helpful suggestions on the manuscript and Hans-Peter Piepho and Juliane Hirte for discussions on statistics. We acknowledge the financial and personnel support by Agroscope, the Research Institute of Organic Agriculture FiBL and the Swiss Federal Office for Agriculture FOAG during 40 years of DOK management and research.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.fcr.2023.109072.

References

Agroscope, 2023, Schweizerische Referenzmethoden der Forschungsanstalten Agroscope. ⟨https://www.agroscope.admin.ch/agroscope/de/home/themen/umwelt-ressourcen/monitoring-analytik/referenzmethoden.html⟩ (2023 07 18).

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

Asseng, S., Ewert, F., Martre, P., Rötter, R.P., Lobell, D.B., Cammarano, D., Kimball, B.A., Ottman, M.J., Wall, G.W., White, J.W., Reynolds, M.P., Alderman, P.D., Prasad, P.V. V., Aggarwal, P.K., Anothai, J., Basso, B., Biernath, C., Challinor, A.J., De Sanctis, G., Doltra, J., Fereres, E., Garcia-Vila, M., Gayler, S., Hoogenboom, G., Hunt, L.A., Izaurralde, R.C., Jabloun, M., Jones, C.D., Kersebaum, K.C., Koehler, A.-K., Müller, C., Naresh Kumar, S., Nendel, C., O'Leary, G., Olesen, J.E., Palosuo, T., Priesack, E., Eyshi Rezaei, E., Ruane, A.C., Semenov, M.A., Shcherbak, I., Stöckle, C., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Thorburn, P.J., Waha, K., Wang, E., Wallach, D., Wolf, J., Zhao, Z., Zhu, Y., 2015. Rising temperatures reduce global wheat production. Nat. Clim. Change 5, 143–147. https://doi.org/10.1038/nclimate2470.

BAFU, 2020. Luftqualität 2019. Messresultate des Nationalen Beobachtungsnetzes für Luftfremdstoffe (NABEL). Bundesamt für Umwelt,, Berne, Switzerland.

Bates, D., Mächler, M., Bolker, B., Walker, S., 2015. Fitting linear mixed-effects models using lme4. J. Stat. Softw. 67 (1), 1–48. https://doi.org/10.18637/jss.v067.i01.

Becker, H.C., Léon, J., 1988. Stability analysis in plant breeding. Plant Breed. 101 (1), 1–23. https://doi.org/10.1111/j.1439-0523.1988.tb00261.x.

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

Berry, P.M., Kindred, D.R., Olesen, J.E., Jorgensen, L.N., Paveley, N.D., 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 Pathol. 59 (4), 753–763. https://doi.org/10.1111/j.1365-3059.2010.02276.x.

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

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).

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

Core Team, R., 2014. R: a language and environment for statistical computing. R. Found. Stat. Comput., Vienna, Austria. ⟨http://www.R-project.org/⟩.

Covarrubias-Pazaran, G., 2016. Genome-assisted prediction of quantitative traits using the R Package sommer. PLOS ONE 11 (6), e0156744. https://doi.org/10.1371/journal.pone.0156744.

Döring, T.F., Neuhoff, D., 2021. Upper limits to sustainable organic wheat yields. Sci. Rep. 11, 12729. https://doi.org/10.1038/s41598-021-91940-7.

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

Erdin, D., 2018. Ertragsentwicklung im Schweizer Ackerbau. AGRISTAT - Stat. Mon. 6, 6–11.

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

Finckh, M.R., Schulte-Geldermann, E., Bruns, C., 2006. Challenges to organic potato farming: disease and nutrient management. Potato Res. 49 (1), 27–42. https://doi.org/10.1007/s11540-006-9004-3.

Gunst, L., Krebs, H., Dubois, D., Mäder, P., 2006. DOK-Versuch: Pilzkrankheiten und Ertrag bei Winterweizen. Agrarforschung 13 (10), 430–435.

Gunst, L., Richner, W., Mäder, P., Mayer, J., 2013. DOK-Versuch: Nährstoffversorgung in Winterweizen – Wo wird es eng? Agrar. Schweiz 4 (2), 74–81.

Hammelehle, A., Oberson, A., Lüscher, A., Mäder, P., Mayer, J., 2018. Above- and belowground nitrogen distribution of a red clover-perennial ryegrass sward along a soil nutrient availability gradient established by organic and conventional cropping systems. Plant Soil 425 (1–2), 507–525. https://doi.org/10.1007/s11104-018-3559-z.

Hejcman, M., Kunzová, E., Šrek, P., 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 Res. 139, 30–38. https://doi.org/10.1016/j.fcr.2012.10.005.

Herrera, J.M., Levy Häner, L., Mascher, F., Hiltbrunner, J., Fossati, D., et al., 2020. Lessons from 20 years of studies of wheat genotypes in multiple environments and under contrasting production systems. Front. Plant Sci. 10. https://doi.org/10.3389/fpls.2019.01745.

Holzkämper, A., Fossati, D., Hiltbrunner, J., Fuhrer, J., 2014. Spatial and temporal trends in agro-climatic limitations to production potentials for grain maize and winter wheat in Switzerland. Reg. Environ. Change 15, 109–122. https://doi.org/10.1007/s10113-014-0627-7.

Hothorn, T., Bretz, F., Westfall, P., 2008. Simultaneous inference in general parametric models. Biom. J. 50 (3), 346–363. https://doi.org/10.1002/bimj.200810425.

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

Hülsbergen, K.J., Schmid, H., Paulsen, H.M., 2022. Steigerung der Ressourceneffizienz durch gesamtbetriebliche Optimierung der Pflanzen- und Milchproduktion unter Einbindung von Tierwohlaspekten – Untersuchungen in einem Netzwerk von Pilotbetrieben. Johann Heinrich von Thünen-Institut,, Braunschweig, p. 540. https://doi.org/10.3220/REP1646034190000.

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

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

Kühne, S., Reelfs, T., Ellmer, F., Moll, E., Kleinhenz, B., et al., 2008. Efficacy of biological insecticides to control the Colorado potato beetle (Leptinotasara decemlineata) in organic farming. Cultiv. Future Based Sci. Modena, Italy 480–483.

Laidig, F., Piepho, H.-P., Drobek, T., Meyer, U., 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. https://doi.org/10.1007/s00122-014-2402-z.

Leifeld, J., Reiser, R., Oberholzer, H.-R., 2009. Consequences of Conventional versus Organic farming on Soil Carbon: Results from a 27-Year Field Experiment. Agron. J. 101 (5), 1204–1218. https://doi.org/10.2134/agronj2009.0002.

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

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

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

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

Möller, K., Habermeyer, J., Zinkernagel, V., Reents, H.-J., 2007. Impact and Interaction of Nitrogen and Phytophthora infestans as Yield-limiting and Yield-reducing Factors in Organic Potato (Solanum tuberosum L.) Crops. Potato Res. 49, 281–301. https://doi.org/10.1007/s11540-007-9024-7.

Oberson, A., Nanzer, S., Bosshard, C., Dubois, D., Mäder, P., et al., 2007. Symbiotic N2 fixation by soybean in organic and conventional cropping systems estimated by 15N dilution and 15N natural abundance. Plant Soil 290 (1–2), 69–83. https://doi.org/10.1007/s11104-006-9122-3.

Oberson, A., Frossard, E., Bühlmann, C., Mayer, J., Mäder, P., et al., 2013. Nitrogen fixation and transfer in grass-clover leys under organic and conventional cropping systems. Plant Soil 371 (1–2), 237–255. https://doi.org/10.1007/s11104-013-1666-4.

Palmer, M.W., Cooper, J., Tétard-Jones, C., rednicka-Tober, D., Barański, M., Eyre, M., Shotton, P.N., Volakakis, N., Cakmak, I., Ozturk, L., Leifert, C., Wilcockson, S.J., Bilsborrow, P.E., 2013. The influence of organic and conventional fertilisation and crop protection practices, preceding crop, harvest year and weather conditions on yield and quality of potato (Solanum tuberosum) in a long-term management trial. Eur. J. Agron. 49, 83–92. https://doi.org/10.1016/j.eja.2013.03.004.

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

Piepho, H.-P., 1999. Stability analysis using the SAS system. Agron. J. 91 (1), 154. https://doi.org/10.2134/agronj1999.00021962009100010024x.

Piepho, H.-P., 2004. An algorithm for a letter-based representation of all-pairwise comparisons. J. Comput. Graph. Stat. 13 (2), 456–466. https://doi.org/10.1198/1061860043515.

Ponisio, L.C., M'Gonigle, L.K., Mace, K.C., Palomino, J., de Valpine, P., et al., 2015. Diversification practices reduce organic to conventional yield gap. Proc. R. Soc. Lond. B: Biol. Sci. 282 (1799), 20141396 https://doi.org/10.1098/rspb.2014.1396.

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

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

Reckling, M., Döring, T.F., Bergkvist, G., Stoddard, F.L., Watson, C.A., 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. https://doi.org/10.1007/s13593-018-0541-3.

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

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

Rotchés-Ribalta, R., Armengot, L., Mäder, P., Mayer, J., Sans, F., 2017. Long-Term Management Affects the Community Composition of Arable Soil Seedbanks. Weed Sci. 65 (1), 73–82. https://doi.org/10.1614/WS-D-16-00072.1.

Rotchés-Ribalta, R., Sans, F.X., Mayer, J., Mäder, P., 2020. Long-term farming systems and last crop sown shape the species and functional composition of the arable weed seed bank. Appl. Veg. Sci. 23, 428–440. https://doi.org/10.1111/avsc.12496.

Rudmann, C., Willer, H., 2005. Jahrbuch Biolandbau Schweiz 2005. FiBL, Frick, Switz.

Rueda-Ayala, V., Ahrends, H.E., Siebert, S., Gaiser, T., Hüging, H., 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. Eur. J. Agron. 96, 34–46. https://doi.org/10.1016/j.eja.2018.03.002.

Runno-Paurson, E., Hansen, M., Tein, B., Loit, K., Jõgi, K., et al., 2014. Cultivation technology influences the occurrence of potato early blight (Alternaria solani) in an organic farming system. Zemdirb. -Agric. 101 (2), 199–204. https://doi.org/10.13080/z-a.2014.101.026.

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

Smith, O.M., Cohen, A.L., Rieser, C.J., Davis, A.G., Taylor, J.M., et al., 2019. Organic farming provides reliable environmental benefits but increases variability in crop yields: a global meta-analysis. Front. Sustain. Food Syst. 3. https://doi.org/10.3389/fsufs.2019.00082.

Smith, R.G., Menalled, F.D., Robertson, G.P., 2007. Temporal yield variability under conventional and alternative management systems. Agron. J. 99 (6), 1629–1634. https://doi.org/10.2134/agronj2007.0096.

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

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

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

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

Webb, J., Jephcote, C., Fraser, A., Wiltshire, J., Aston, S., 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 Manag. 32 (1), 3–16. https://doi.org/10.1111/sum.12250.