Beetles, roots and yields:
Exploring the edge effects of agro-ecosystem diversity on three different scales

M.Sc. Thesis Report
by
Maren Weller

December 2015

Farming Systems Ecology Group

Droevendaalsesteeg 1 - 6708 PB Wageningen - The Netherlands

WAGENINGEN UNIVERSITY
WAGENINGEN UR

Beetles, roots and yields:

Exploring the edge effects of agro-ecosystem diversity on three different scales

Name: Maren Weller

Registration number: 24039007070

Credits: 36 ECTS

Code number: FSE-80436

Supervisor: Dr. ir. Dirk van Apeldoorn

Examiner: Dr. ir. Walter Rossing

Wageningen, December 2015

EXECUTIVE SUMMARY

Edge effects on three different scales were assessed in a diverse strip cropping system at Droevendaal, the organic farm of the Wageningen University and Research Centre in the Netherlands. Research for this thesis was conducted in an experimental field consisting of eight strips subdivided into twelve plots, out of which half consisted of monocultures. The other plots either consisted of different varieties of the respective crop or were intercropped. Statistically significant differences between treatments and rows were detected by fitting generalized linear models to the data and the consecutive post-hoc tests (Tukey's HSD test). For the detection of differences in Nitrogen content, Welch's t-tests were employed.

Using pitfall traps, the activity of ground beetles was monitored in rows between and in the middle of strips over the course of one week. Crop type had significant effects on the activity of ground beetles, with the highest activity in the oilseed rape strip and the lowest in the fallow strip. Depending on crop type, activity in the edge rows was higher (grass-clover, fallow and potato strip) or lower (oilseed rape strip). Differences in activity depended on time of day, with most ground beetles being active by night.

Root samples from four different potato varieties were taken with an auger from the two middle rows of the potato strip. Root density was measured in three different depths (0-15, 15-30 and 30-45 cm). Results showed a significantly higher dry root weight in 30-45 cm depth in the mixed plots. It was also in the mixed plots were the mineral Nitrogen content was significantly higher than in the no-mix. Statistical analysis further showed that root dry weight was a predictor for tuber yield.

The influence of edge effects on yield depended on the crop type. While the potato row bordering on the oilseed rape strip had a significantly higher yield than the middle rows, the yield in the edge rows of the oilseed rape and the wheat were not significantly different from the middle rows. In the mixed plots, the edge rows of grass-clover and wheat neighbouring each other had lower yields than the other rows in those strips, indicating possible competition.

The extent of edge effects on different scales therefore depends on both the mixture and the formation of the crops. The choice of neighbouring crops in a diverse strip cropping system can have a considerable influence on yield and pest control.

ACKNOWLEDGMENTS

I would like to thank everyone who helped and supported me while writing this thesis: my supervisor Dirk van Apeldoorn; my fellow field workers Kirstin Surmann, Rianne Prinsen and Ambar Hernandez Romero; and the providers of free R tutorials, such as the Statistical Consulting Group of the University of California, Los Angeles.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I: Introduction

Considering that farming is extremely vulnerable to the effects of environmental changes, such as climate change, adaptation to those changes will be a key factor in the near future (Altieri et al., 2015). Therefore, the need to consider a different way of intensifying agricultural production is becoming more and more pressing (Tittonell, 2014).
Ecological intensification could increase food production while being sustainable at the same time (Tittonell, 2013). This is achieved by making efficient use of all the natural functionalities which ecosystems have to offer (Tittonell, 2013). Those agro-ecological farming strategies include, amongst others, diversification or the maintenance of local genetic diversity (Altieri et al., 2015). An increase in diversity has been shown to increase the productivity of the farming system (Cardinale et al., 2007; Smith et al., 2008); an effect which has been linked to a complementary pattern of the species' resource use at the edges (Mollison, 1988; Sala, 2001).

Diversity in agro-ecosystems

Diversity not only shows itself in the number of different species, but also in their distribution in time (e.g. crop rotations) and space (arrangement on the field) as well as in their interactions with each other. The diversity in time and space is due to both natural successional processes and human actions, such as the growing of polycultures, crop rotations or harvest (Kremen et al., 2012). Multispecies cropping systems can therefore often be found in agro-ecological or organic farming systems which rely on ecological principles.
Since there are in general more species than functions in an agroecosystem, there is a certain redundancy within the system. However, those components only appear redundant at one point in time – as soon as environmental changes occur, they become important. It is those redundancies that allow the system to continue functioning and to provide ecosystem services. Those agroecosystems would be able to keep up the provision of food production even when they are challenged by drought or extreme rainfall (Altieri, 2015).
Since different species have different traits – such as deep or shallow roots, a preference for sun or shade, a high or low resistance to water stress – farming systems with a large number of different species unite a broader range of traits on their fields (Sala, 2001). Those different traits can thus make the use of resources more efficient by complementing each other.

Resource utilisation at the edges

The interactions between different crops or species can increase resource utilisation and thus productivity (Mollison, 1988). All of those interactions happen at edges – that is, wherever boundaries meet, and those boundaries can consist of different soils, different climates, different species or any kind of natural condition (Mollison, 1988). Since we are human, we see edges from our perspective (Figure 1), but the way in which organisms

perceive and respond to edges will often be different from humans (Lindenmayer et al., 2008).
In ecology, edges are generally defined as boundaries between distinct patch types (Ries et al., 2004). However, this definition depends on how patches are defined within a landscape. Mollison (1988) defined edges as places of varied ecology. Due to ecological flows, the environmental conditions near edges are often intermediate between the prevalent conditions in the neighbouring patches (Ries et al., 2004). With decreasing patch size, the relative amount of edge increases, which exacerbates the influence of the surrounding edges (Fletcher et al., 2007).

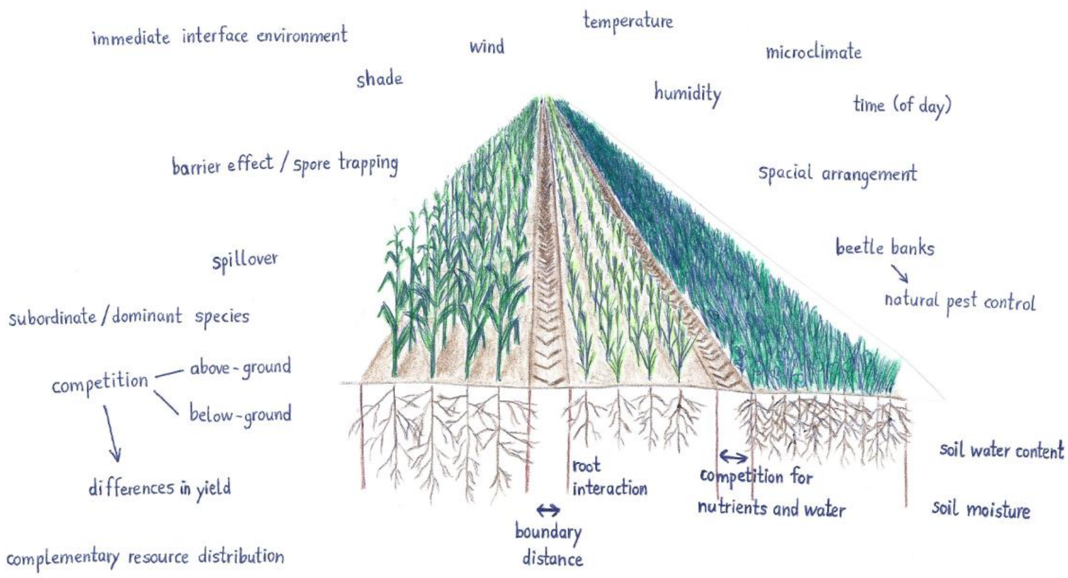


Figure 1: edge effects on different scales in a diverse strip cropping system

Edge effects on crop yields

Edge effects refer to the changes in biological and physical conditions that occur at the boundary (Lindenmayer et al., 2008). They are the result of the interaction between two adjacent ecosystems, when they are separated by an edge (Murcia, 1995). Even though the magnitude of responses to certain edge effects may differ, the nature of the effect (i.e., positive or negative) will often not (Lindenmayer et al., 2008). Murcia (1995), for example, found that some physical edge effects get weaker when there is less exposure to sunlight.
Considering Mollison's (1988) definition of edges, it can be concluded that the more different species a farming system unites, the more edges there are. Strip cropping is therefore also known as edge cropping (Mollison, 1988). A review by Francis et al. (1986) concluded that yields in strip cropping systems are usually higher than in monocultures. However, the surplus of yield depends on the weather: in years with adequate rainfall, the strip cropping systems achieved yields that were up to 10-20 % higher than those of respective monocultures (Francis et al., 1986).
In a field trial, Glowacka (2014) observed changing maize yields in strip cropping, depending on the maize's row position in the strip. The maize yields were significantly higher in the

edge rows, which were neighboured by a lupin strip on one side and an oat strip on the other side. Indeed, maize is often grown in strip cropping systems due to its strong and positive reaction to edge effects (Glowacka, 2014).

In a 3-year-trial, Bouws and Finckh (2008) found that the yields in edge potato rows, which bordered on cereals, were significantly reduced. Those negative effects on the potatoes could have been caused by competition between potatoes and cereals. The highest tuber yields, on the other hand, were obtained in plots that were neighboured by grass-clover. Bouws and Finckh (2008) therefore concluded that grass-clover as a neighbour for potatoes seems to be a better choice than cereals. In their study, Bouws and Finckh (2008) concluded that the choice of neighbouring crops in strip cropping systems will be important.

While strip cropping systems with a large variety of mixed species have been shown to be very productive, they have also been shown to be effective in disease management and the improvement of soil fertility: the larger the number of genotypes and the more random or diverse the mix, the more effective they tend to be (Rämert et al., 2002). However, not all mixtures have the potential to reduce diseases; in order to be functional the resistance genes of the varieties in the mixture must match the avirulence genes that are present in the pathogen population (Mundt, 2002).

Intercrops have the potential to reduce the incidence of pests as well, while producing a greater yield by making more efficient use of resources than a monoculture. This can be achieved by using a mix of crops different in height, canopy structure, rooting structure and nutrient requirements (Lithourgidis et al., 2011).

Ground beetles as natural pest control

Since sustainable agricultural systems also need a functioning self-regulation of predators and pests, beetle banks can be valuable additions to a diversified crop field. Beetle banks usually consist of native grasslands and are maintained at the field margins in order to protect the ground beetle populations (Lin, 2011).

Ground beetles, more specifically carabid beetles, are polyphagous predators and play therefore an important role in natural pest control. Carabid beetles are often night-active, and their relatively typical, dark-coloured body makes it easy to recognize them at the family level. Since they have long legs, these beetles can run quickly over the soil surface (Kromp, 1999).

Carabids respond to microhabitat conditions, such as soil moisture, temperature or light penetration through the canopy - conditions which are influenced by the ground cover (Honek 1997, Kromp 1999).

In field experiments, Allema (2014) found that the motility of ground beetles was higher in crop habitats than at the edges. In fact, at the interface, more ground beetles moved towards the crop than towards the edge, which led Allema (2014) to the conclusion that those edges acted as barriers for ground beetle dispersal.

In northern England, Eyre et al. (2012) monitored the activity of ground beetles in a plot trial system. They found that the crop type had significant effects on the activity and the species richness of ground beetles. The highest activity was in beans and winter barley, while the lowest was in spring barley and vegetable plots mainly consisting of potatoes (Eyre et al., 2012).

In a field trial in Davis, California, where the effects of living mulches on the population dynamics of soil arthropods were tested, Altieri et al. (1985) found significantly higher numbers of ground predators (such as carabid beetles) in pitfalls which were placed in clover plots than in pitfalls which were placed in clean cultivated plots. In the mountain region of North Carolina, carabid beetles seem to prefer systems with ground cover as well (Hummel et al., 2002).

However, high activity does not necessarily mean high efficiency. In a selective review, Sheehan (1986) concluded that natural enemies are not necessarily more effective in a diverse cropping system: locating potential victims could be hindered by patchiness or increased plant densities in the system.

Roots under visible edges

In diverse cropping systems, plants may root to different depths to avoid competition (Kremen and Miles, 2012). For instance, a plant could increase its water uptake by growing deeper reaching roots which can tap a water source that is not available to its more shallow-rooted neighbours (Casper and Jackson, 1997). Indeed, when it comes to belowground competition, the most important factor is the occupation of soil space (Casper and Jackson, 1997), because roots are the only part of a crop that can take up water and nutrients from the soil (Ahmadi et al., 2014).

However, interactions between crop roots do not necessarily have to be competitive. It is also possible that the roots of one crop facilitate the nutrient uptake of another crop (Brooker et al., 2008; Kremen and Miles, 2012). Li et al. (2007), for example, found in a field trial that maize yields were increased by the uptake of phosphorous which had previously been mobilized by intercropped faba beans. Song et al. (2006) found that wheat yields were higher when intercropped with faba beans compared to wheat as a sole crop. The increase in wheat yield had been facilitated by the faba beans via an increase of Phosphorous and Nitrogen in the rhizosphere (Song et al., 2006).

Aim, research questions and hypothesis

Edge effects can take place on many different scales, and the nature of the effects determines if those edge effects contribute to the productivity of a diverse cropping system. While it has been established that ground beetle activity depends on soil cover (Altieri et al., 1985; Hummel et al., 2012) and that the interface between two crops can act as a barrier for ground beetle dispersal (Allema, 2014) the ground beetles' response to edges in a whole strip cropping system has not been investigated yet. Differences in crop yield on the edges of crop strips have been established for certain combinations with maize and potatoes (Bouws and Finckh, 2008; Glowacka, 2014), but the effect of weather on yield (Francis, 1986) makes the occurrence of those edge effects highly context specific. Further research is needed to explore how edge effects influence the performance of a diverse strip cropping system in the Netherlands.

The aim of this study is therefore to explore how edge effects on three different scales influence the performance of a diverse agro-ecosystem in a trial that is experimenting with diversity in space, time and genes. The three scales are strip-strip interactions through ground beetle activity, root-root interactions through resource use of different potato varieties and crop-crop interactions through yield differences at the strip edges.
The research questions following from the aim are the following:

- Are ground beetles more active in edges?
  - Is there a difference within and between strips?
  - Is there a difference between diurnal and nocturnal activities?
- Are yields different at crop edges?
  - Is there a difference within and between treatments?
- Is the root dry weight different in potato mixtures?
  - Does the dry root weight of potatoes in the no-mix plots differ from the dry root weight of four different potato varieties in the mixed plots?

Based on the findings in the literature mentioned above, it can be hypothesised that ground beetle activity is different between crops and time of day. It can further be expected that crop yields are different at the strip edges and that a difference between and within treatments exists. Lastly, it can be assumed that the root density is different in potato mixtures, and that the root dry weight of potatoes in the no-mix plots differs from the root dry weight of four different potato varieties in the mixed plots.

Outline of this thesis report

The materials and methods of both the field experiments and the statistical analysis are explained in the following chapter; it is also there where a detailed description of the experiment is provided. R-scripts of the statistical analysis can be found in the supplementary material. The results are shown in Chapter 3. In Chapter 4, the limitations of the approach and the results are discussed. The results are being compared to what could have been expected based on the literature review in the introduction. Finally, the conclusion provides an overview of the results of both the experiment and the discussion and provides recommendations for further research.

CHAPTER II: Material and Methods

The experimental site is part of the Droevendaal experimental and training farm and belongs to the Wageningen University's Unifarm in the Netherlands. The Droevendaal farm is located north of the campus (51°59'28"N, 5°39'42"E) on sandy soil and is subjected to an average annual rainfall of 829 mm and an average temperature of 11°C. The systems trial with the strip intercropping experiment was started in 2014 and is managed by the Farming Systems Ecology group (FSE) of the Wageningen University.

2.1 Design of the experimental field

The experimental field consisted of eight strips that were subdivided into 12 plots. Every plot had two buffer zones, one before the plot began and one after the plot. In total, the field was 250 m long. Every strip was 3 m wide and consisted of plots divided into "mix" and "no-mix". In the no-mix plots, the respective crop grew by itself, while the mixed plots either consisted of different varieties of one crop or were intercropped (Figure 2). Every plot was 10 m long and 3 m wide, with buffer zones of the same width and 5 m length. The no-mix plots were plots 1, 3, 7, 8, 9 and 12. The mixed plots were plots 2, 4, 5, 6, 10 and 11.

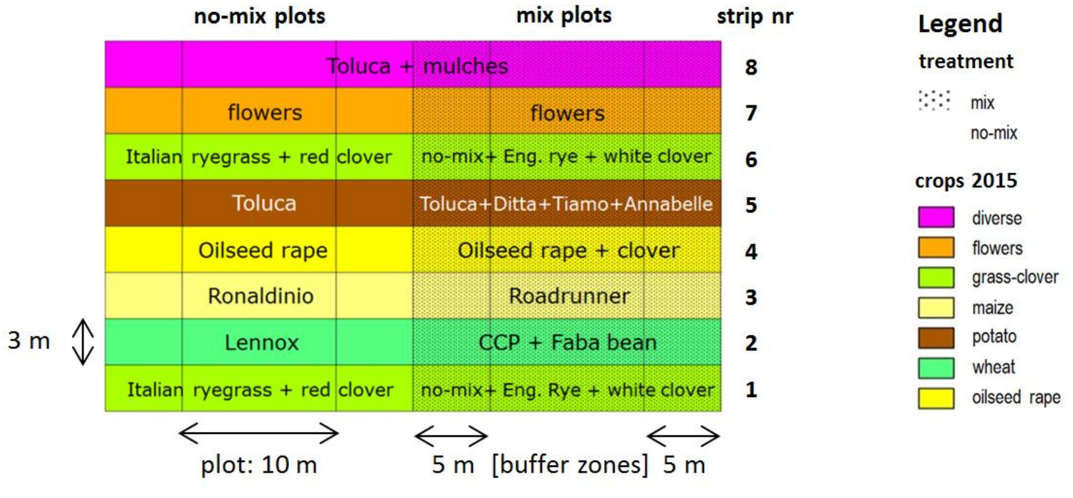


Figure 2: design of the strips and plots in the experimental field

In **strip 1**, the grass-clover had been sown in May 2014. The no-mix plots consisted of Italian ryegrass (*Lolium multiflorum* LAM. *var. Sultano*), which was sown at a density of 35 kg/ha, and red clover (*Trifolium pratense* L. *var. Lucrum*), which was sown at a density of 5 kg/ha. The mixed plot consisted of Italian ryegrass (*var. Tetra*), English ryegrass (*Lolium perenne* L. *var. Country Balance*), red clover (*var. Lucrum*) and white clover (*Trifolium repens* L. *var. Alice* and *var. Riesling*). The mixture was sown in a proportion of 20:14:2:2:2 kg/ha. A row distance of 12.5 cm was kept in both treatments.

In the no-mix plots in **strip 2**, summer wheat (*Triticum* L. *var. Lennox*) was sown at a density of 140 kg/ha. In total, there were 12 rows of wheat with a distance of 25 cm between each other. In the mixed plots, the wheat variety *CCP* was intercropped with white flowering faba beans (*Vicia faba* L. *var. Nile*) (Figure 3). The wheat was sown at a target density of 160 seeds per m², with an additional 40 beans per m². The same row distance of 25 cm was kept in the mixed plots, where the outer and inner rows consisted of wheat and were then followed by one row of faba beans, two rows of wheat and another row of faba beans. All crops in strip 2 had been sown on the 15th of April, 2015.

Figure 3: mixed plots of wheat and faba bean in strip 2; order of plants

In **strip 3**, maize (*Zea mays* L. *var. Ronaldinio*) had been sown in the no-mix plots and the maize variety *Roadrunner* had been sown in the mixed plots. However, crows picked so many seeds that the maize strip was marked as a failure. It was tilled and a grass-clover mix was sown later in the season.

In **strip 4**, summer oilseed rape (*Brassica napus* L.) was sown in the no-mix plots, whereas the mixed plots consisted of summer oilseed rape and a white clover mixture of the varieties *Jura*, *Riesling* and *Alice* at a ratio of 2:2:2. A row distance of 25 cm was kept in both treatments, so there were 12 rows in total. All crops in strip 4 were sown on the 14th of April, 2015.

In **strip 5**, the potato (*Solanum tuberosum* L.) variety *Toluca* was planted in the no-mix plots with 270 plants per plot (30 m^2). The mixed plots consisted of four different potato varieties, which were planted in the following order (from West to East): *Annabelle* – *Toluca* – *Ditta* – *Tiamo*. In order to keep the biggest distance possible between the varieties and thus to slow down the spread of *Phytophthora infestans*, the potatoes were planted in a check board pattern (Figure 4). Each potato variety in the mixed plots was represented by 67 plants. In total, there were four rows of potatoes with a distance of 75 cm between each. The distance between each individual potato plant was 30 cm. All potatoes in strip 5 had been pre-germinated and were planted on the 23rd of April, 2015, when soil temperatures reached between 7 and 10°C.

Annabelle	*Toluca*	*Ditta*	*Tiamo*
Toluca	*Ditta*	*Tiamo*	*Annabelle*
Ditta	*Tiamo*	*Annabelle*	*Toluca*
Tiamo	*Annabelle*	*Toluca*	*Ditta*

Figure 4: planting pattern of potato varieties in mixed plots, starting from the Western plot border (left)

In **strip 6**, grass-clover was growing for the first year. Like in strip 1, the no-mix plots consisted of Italian ryegrass (*var. Tetra*), which was sown at a density of 35 kg/ha, and red clover (*Lucrum*), which was sown at a density of 5 kg/ha. The mixed plots consisted of English ryegrass (*var. Country Balance*), Italian ryegrass (*var. Tetra*), red clover (*var. Lucrum*), white clover (*var. Alice* and *var. Riesling*), Alexandrian clover (*Trifolium alexandrinum* L.), crimson clover (*Trifolium incarnatum* L.), chicory (*Cichorium intybus* L.) and plantains (*Plantago* L.). The mixture was sown in a proportion of 10:7:2:2:2:2:2:1:3:5 kg/ha. In all plots, a row distance of 12.5 cm was kept.

Strip 7 consisted of the flower mix "mengsel van Dijke" which contained, amongst others, fennel, buckwheat, cornflowers and corn daisies. It was sown at a density of 20 kg/ha with a row distance of 12.5 cm. **Strip 8** consisted of an experimental potato strip, where the potato variety Toluca was treated with different kinds of mulch. The same row distances and planting densities as in strip 5 were applied. Neither one of those strips were used for this study.

Table 1: sowing / planting dates of crops, seed density and number of rows per strip

strip	**Crop**	**sowing / planting date**	**seed density**	**row distance**
1	grass-clover	May 2014	40 kg/ha	12.5 cm
2	wheat and faba bean	15th of April, 2015	no-mix: 140 kg/ha mix: 80:210 kg/ha	25 cm
4	oilseed rape	14th of April, 2015		25 cm
5	potatoes	23rd of April, 2015	4/m²	75 cm
6	grass-clover	end of February 2015	40 kg/ha	12.5 cm

2.2 Field sampling

During the growing season, the weather was rather unstable, with heavy rainfalls in July and August and unseasonal cold during the last week of July (Figure 5).

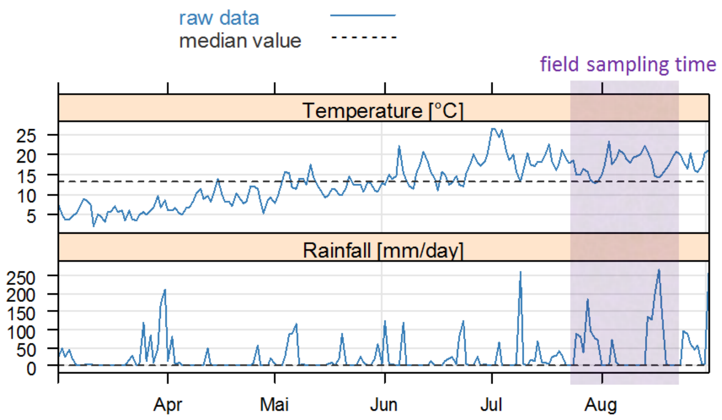


Figure 5: weather at the experimental farm during the growing season and the sampling time

All samples were taken between the 20th of July and the 19th of August, 2015. Due to frequent rainfalls, the wheat and the oilseed rape were harvested later than initially intended, because the crops needed to be dry for the harvest. For the same reason, the potato plants were burned and harvested later than initially planned.

2.2.1 Life trapping of ground beetles

For the duration of one week (27th - 31st of July, 2015), pitfall traps were placed in the middle and side tractor tracks of strips 1-6. There were 144 pitfall traps in total: two pitfall traps per plot, one in the middle and one in the side furrow. The pitfall traps consisted of empty plastic cups which were put into pre-dug holes in the ground, so that the rim of the cup was level with the surrounding ground. The cups were 14 cm deep and had a diameter of 9 cm. To prevent the cups from getting flooded, roofs with a diameter of 12 cm were placed at a height of 5 cm over the cups. The roofs consisted of a round, saucer-shaped plastic shell, with metal pillars which were stuck into the ground to keep it in position. The pitfall traps were emptied every 12 hours for 5 days; once in the morning between 7:30 and 8:30 and once in the evening. The insects which had fallen into the trap were identified to family level and counted before they were released again outside a two meter radius from the pitfall location. Since the pitfall traps were put into place on a Monday morning, they were emptied on four mornings and on five evenings.

2.2.2 Collection of potato root samples

The samples for the potato roots were taken between the 20th and the 22nd of July, 2015, in rows 2 and 3 of strip 5. In the no-mix plots, where only one variety (Toluca) grew, two plants were sampled per row. In the mixed plots, where four different varieties grew (Annabelle, Ditta, Tiamo and Toluca), one plant of each variety was sampled: Tiamo and Toluca in row 2, and Annabelle and Ditta in row 3. Every sample was taken with an auger at a distance of 5 cm from the potato plant stem, on both sides of one stem (Figure 6). All sampling locations were in the Western buffer zones of each plot, with the first potato plant to be sampled right next to the border between buffer zone and plot.

In addition to the root samples, mineral Nitrogen samples were taken as well. For the mineral Nitrogen, composite samples were taken from every Western buffer zone of each plot with a narrow auger (2.8 cm in diameter). The auger was positioned between the two potato plants from which the root samples had been taken, at a distance of 15 cm from the stem. The soil samples were put into plastic bags and put into a cooling room until they were processed further.

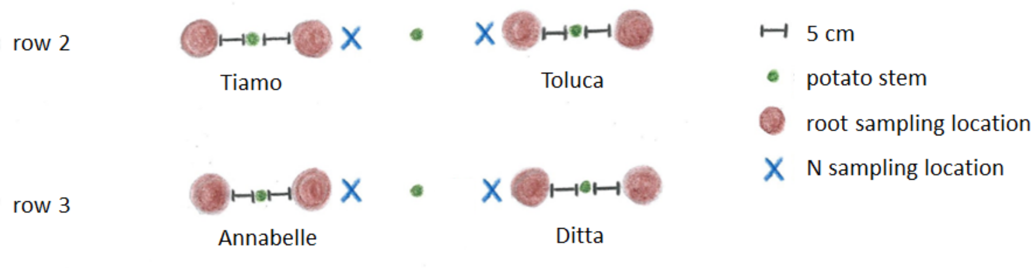


Figure 6: potato root sampling locations in western buffer zones of each plot in strip 5

In order to take samples of the potato roots, holes were dug with an auger that consisted of a cylindrical part with a jagged rim on the bottom side, which was attached to a long handle (Figure 7). The cylindrical part was 7.5 cm in diameter and 15 cm long. In total, every hole dug was 7.5 cm wide and 45 cm deep. The samples consisted of three parts per hole: the first part was from a depth of 0-15 cm, the second from a depth of 15-30 cm and the third from a depth of 30-45 cm. Each sample was put into a plastic bag and stored in a cooling room until it could be processed further.

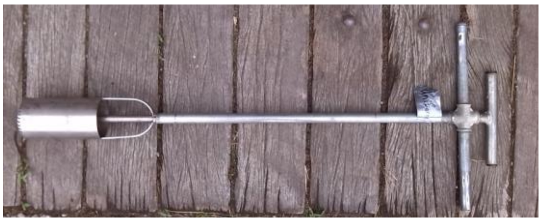


Figure 7: auger used to take potato root samples

In order to separate the roots from the soil, every sample was washed in a 2 mm sieve. The roots were then put into aluminium trays and dried in ovens set to 105°C for a total of 24 hours. Afterwards, the dry roots were weighed on a precision scale.
In order to determine the mineral Nitrogen, the soil samples were filled into aluminium trays and put into an oven set to 40°C for 48 hours. The soil was then sieved through 1.5 mm sieves and filled into small plastic jars which were given to the laboratory of the Farming Systems Ecology group for further analysis. There, the dried soil samples were extracted at 20 °C in a 1:10 (w/v) ratio with a 0.01 M $CaCl_2$ solution of 20 °C. After reaching equilibrium, which was achieved in two hours shaking, the pH was measured in the settling suspension and part of the suspension was centrifuged for the manual or automated determination of all nutrients, metals and other elements. These methods have been adapted from Houba and Novozamsky, 1998.

3.2.3 Yield sampling of crops

Grass-clover

Yield samples for the grass-clover (only strip 1) were taken in squares of 25x25 cm on the 6th on August, 2015, 37 days after the strip had been mown. Every plot was sub-divided into 6 rows, each 25 cm wide. The samples were taken in random locations, one per row per plot (Figure 8). In order to define the sampling area, a frame made of bamboo sticks was placed in the sampling location before the plants were cut. The grass-clover was harvested by hand and cut with garden scissors at 3 cm above the ground. Every sample was weighed fresh before it was put into an aluminium tray and dried in an oven which was set to 70°C for 48 hours. Afterwards, the dry weight was determined.

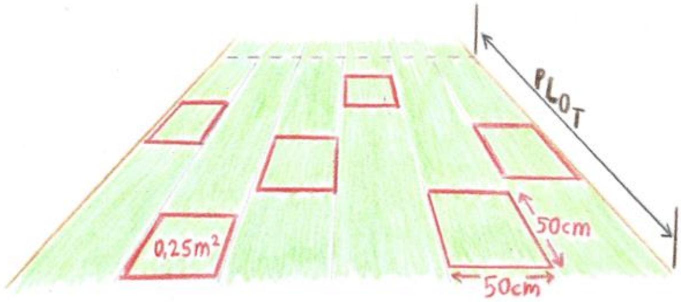


Figure 8: random yield sampling locations per plot (strip 1)

Wheat

The wheat was harvested by hand on the 19th of August, 2015, 126 days after sowing. For the sampling, the wheat strip was subdivided into 4 rows, each 75 cm wide. The sampling locations were squares of 75x75 cm, one in every row in every plot. The wheat was cut with garden scissors at 5 cm above the ground. In order to define the sampling area, a frame made of bamboo sticks was placed in the random sampling location before the plants were cut. In total, 48 samples were taken: one per row in four rows in 12 plots. The weeds growing within that frame were cut as well, but bagged separately.

Both the wheat and the weeds were weighed fresh. After that, the weeds were discarded. The wheat was put into aluminium trays and oven-dried at 40°C for eight hours and at 70°C for the following 40 hours. Once the wheat was dry, the dry weight was determined. The weighed samples were threshed mechanically. The kernels were then cleaned by hand through sieving under an air exhaust. Once the kernels were clean, they were counted by a seed counting machine and weighed by hand.

In order to determine the Nitrogen content of the wheat stems and the kernels, subsamples were taken from every row and combined by plot. The composite samples of stems and kernels were ground mechanically and then given to the laboratory of the Farming Systems Ecology group for further analysis.

The faba beans which were intercropped with the wheat were harvested separately on the 21st of August, 2015, 128 days after sowing. They were cut in the same locations as previously the wheat had been harvested. The faba beans were put into aluminium trays in which they were oven-dried at 70°C for 48 hours. Afterwards, the dry weight was determined and the beans were separated from their shells by hand. The beans were then counted by a seed counting machine and weighed. In order to determine the Nitrogen content, composite samples of beans and stems were made in a similar fashion as for the wheat. The composite samples were ground mechanically before their Nitrogen content was determined. For this procedure, the samples were digested with a mixture of H_2SO_4–Se and salicylic acid (Novozamski et al., 1983). The actual digestion was started by H_2O_2. In this step, most of the organic matter was oxidized. After decomposition of the excess H_2O_2 and the evaporation of water, the digestion was completed by concentrated H_2SO_4 at an elevated temperature (330°C) under the influence of Se as a catalyst. In these digestions, the total Nitrogen was measured spectrophotometrically with a segmented-flow system (autoanalyzer II, Technicon). Remark: Salicylicacid is added to prevent loss of nitrate-N. This is done by coupling the Nitrate to salicylic acid, a reaction which proceeds easily in the acid medium. In this way, 3-nitrosalicylic acid and/or 4-nitrosalicylic acid are formed. These compounds are reduced to their corresponding amino forms by the plant organic matter.

Oilseed Rape

Yield samples for the oilseed rape were taken in 75x75 cm squares on the 21st of August, 2015, 129 days after sowing. The strip was subdivided into four rows which were each 75 cm wide. The stems were cut with secateurs at 5 cm above the ground. The sampling area was delimited by a wooden frame made of bamboo sticks. For every plot, four samples were taken in random locations, one in every row. Within the bamboo frame, the number of plants was counted, as well as the number of stems. The harvested plants were then put into plastic bags and stored in a cool storage room until they could be processed further after the weekend.

The oilseed rape was put into aluminium trays and oven-dried at 70°C for the following 48 hours before the dry weight was determined. The plants were threshed by hand and the seeds were cleaned manually as well, using sieves of different sizes (2-3 mm). The seeds were then weighed and counted by a seed counting machine.

In order to determine the Nitrogen, composite samples were made of every plot in a similar way as for the wheat. The stems were ground mechanically; the seeds were left as they were. The procedure for the analysis was the same as described above for the wheat.

Potatoes

In order to take yield samples from the potatoes (strip 5), four neighbouring plants were harvested by hand in every row in every plot on the 11th of August, 2015, 110 days after planting. The location of those plants was chosen at random. The potatoes were put in net bags; every row in every plot was bagged separately. The potatoes were then washed mechanically and weighed under water. Before the fresh weight was determined, the potatoes were manually sorted into three categories: cull, small (less than 4 cm in diameter) and big (more than or equal to 4 cm in diameter). For every category, the fresh weight was determined.

In order to harvest potatoes that could later be used to relate the tuber yield to the root dry weight, another hand harvest took place on the 20th of August, 2015, 119 days after planting. This time, only potato plants in the mixed plots were harvested: four neighbouring plants per row per plot; separately bagged by variety. In every row, the potato plants harvested were the ones closest to the western buffer zone of each plot. The potatoes were cleaned mechanically and sorted into the three categories mentioned above before the fresh weight was measured. The underwater weight was measured for every variety per plot.

On the 24th of September, 154 days after planting, the potato strip was harvested mechanically. The potatoes from every row in every plot were bagged separately. Before the fresh weight was determined, the potatoes were roughly cleaned from soil by hand.

2.3 Statistical analysis

All statistical analyses were conducted using the program R (version 3.2.2). Since generalized linear models (GLMs) allow for response variables which have non-normal error distributions, which is often the case in biological data (Fox et al., 2015), GLMs were used to assess the effects of variables on ground beetle numbers and crop yields. Opposed to multiple regression models, GLMs further allow linear combinations of multiple dependent variables (Park et al., 2005). Since the covariates for the response variable were similarly correlated with each other as those used to fit the model, all variables were assumed to be independent.

For the selection of the best model, the “dredge” function of R’s MuMIn-package was employed to rank possible models according to the Akaike Information Criterion (AIC) based on the variables in the full model. The AIC is a measure for the relative quality of statistical models for a given set of data. Out of the models in the ranking, the preferred model was the one with the lowest AIC value (see Appendix A for model ranking tables).

R’s default option for GLMs with a Gaussian distribution is the link function “identity”. For GLMs with a Poisson distribution, the default link function is “log”. Unless indicated otherwise, those link functions have been used in the respective GLMs.

2.3.1 Ground beetles

Since the data set of the ground beetles consisted of count data for the outcome variable (i.e., the number of ground beetles found in the pitfall traps), the data set contained many zeros. Therefore, the data set would have been unlikely to follow a normal distribution. Instead, count data sets usually follow a Poisson distribution (Ats.ucla.edu, 2015). To model the ground beetle count data, a series of zero-inflated Poisson models was fit to account for the many zeros. The models were compared using Vuong tests. Based on those comparisons, the best model was chosen. As a direct comparison had previously shown, the best zero-inflated Poisson model had a lower AIC than the best GLM which the dredge function of R’s MuMIn-package could find.

In order to make pairwise comparisons (i.e. to compare beetle activity during the night with the activity during the day), small GLMs were fitted containing only the additional or the interaction effect of time with treatment as a predictor. This allowed the performance of Tukey’s HSD test to look for statistically significant differences. T-tests could not be used because the assumption of normal data distribution was not met, and even the Wilcoxon-Mann-Whitney test could have led to wrong conclusions (McElduff et at., 2010). Regression modelling was therefore a valid alternative (McElduff et at., 2010).

2.3.2 Potato roots

Before a GLM was fit to look for effects of root density on potato yield, several t-tests were performed in order to compare the root density of different potato varieties in different locations. T-tests are rather robust, even if not all assumptions are met (Kang and Harring, 2012). The default setting in R’s stats-package assumes unequal variances between the two groups in an unpaired t-test and therefore applies the Welch modification. Welch’s test estimates the variances of the two groups and, based on this, adjusts the degrees of freedom which are used in the test (Spector, 2014). If the variances between the two groups are equal, Welch's *t*-test returns the same result as Student's *t*-test. Therefore, all t-tests were performed using the default setting in R's stats-package.

A two-sample Welch's *t*-test assuming unequal variances was performed to test the hypotheses that

- ✓ the dry root weight of the potato variety Toluca in both treatments is equal.
- ✓ the dry root weight in both treatments is equal.
- ✓ the mineral Nitrogen content of the soil in both treatments is equal.
- ✓ the root dry weight of the different potato varieties in the mixed plots is equal.

A t-test was statistically significant when $p < 0.05$. For all p-values higher than 0.05, the test results were statistically not significant and not written in detail in the results.

In order to test if the Nitrogen content had influence on the root dry weight, another GLM was fit with root dry weight as the outcome variable and the mineral Nitrogen content in three different depths as predictors.

The mean (M) and standard deviation (SD) to describe differences between treatments were obtained using the "describe" function of R's psych-package.

2.3.3 Crop yields

GLMs were used to assess the effects of variables on outcome variables relevant for the particular crop (Table 2). In order to test the absolute quality of a model, the fit of the GLMs was determined using the estimated deviance D^2, which is a measure of the variance reduction and the equivalent to R^2 in linear models (Guisan and Zimmermann, 2000). D^2 was calculated using the R-package modEva by Barbosa et al. (2014).

For the rows and the plots, the variables were used both as factors and as continuous variables. The row coordinates (Field Y) and the plot coordinates (Field X) as continuous variables took into account that the rows and plots next to each other may have had more in common than the rows and plots which were further apart.

For the grass-clover, the outcome variable was the dry weight (kg/m^2). For the wheat, the grain yield is the most important indicator. Usually, the grain dry weight constitutes about half of the final crop dry weight. The harvest index is the ratio between the dry grain yield and the total crop dry weight (Sylvester-Bradley et al., 2015). Therefore, two GLMs were fit; one model with dry weight in kg/m^2 as outcome variable and one with kernel weight per m^2. The harvest index was compared between treatments using Welch's t-test. Due to the intercropped faba beans in the mixed plots of strip 2, there were more rows of wheat in the no-mix than in the mix. To account for this difference, all values of the variables for the mix plots were multiplied by two prior to the application of any statistical tests. The only exception was the comparison of biomass between rows and treatments in strip 2. The biomass was calculated as the sum of the dry weight of wheat, weeds and faba beans.

The yield of oilseed rape is determined by the number of seeds per m^2 and the weight of each individual seed, with a particular focus on the number of seeds (Berry et al., 2012). For the oilseed rape, the dependent variables for the models therefore were the dry weight in kg/m², the seed weight in kg/m², the number of seeds per m² and the thousand kernel weight (TKW).

For the potatoes, two models were used, one for the fresh weight in kg/m² as dependent variable and one for the yield defined as fresh weight minus cull in kg/m² as dependent variable.

Table 2: crops and their dependent variables in the generalized linear models

crop	**dependent variable(s) in GLM**
grass clover	dry weight in kg/m²
wheat	dry weight in kg/m², kernel weight in kg/m², biomass in kg/m² (including faba beans and weeds)
oilseed rape	dry weight in kg/m², number of seeds per m², seed weight in kg/m², TKW
potatoes	fresh weight in kg/m², yield (fresh weight – cull) in kg/m²

After fitting a full model with all predictors relevant for the outcome variable, the best models were selected using the "dredge" function of R's MuMIn-package. However, not all models with the lowest AICs contained the categorical variables "row" and "treatment". Since the goal was to find differences between rows in different treatments, additional GLMs were written which only contained the additional or the interaction effect between row and treatment as predictor. Those additional GLMs served the sole purpose of making pairwise comparisons of rows and treatments and thus to see if there were statistically significant differences between those.

The Nitrogen contents of dry matter (grass-clover), stems and kernels (wheat) as well as stems and seed (OSR) was only compared on a treatment basis using two-sample Welch's t-tests assuming unequal variances. The t-tests' null hypothesis was that there was no difference in Nitrogen content between treatments.

In order to compare the potato variety Toluca in the mixed plots with Toluca in the no-mix, several two-sample Welch's *t*-tests, assuming unequal variances, were performed. A t-test was statistically significant when $p < 0.05$. The t-tests tested the hypotheses that

- ✓ the fresh weight of Toluca in the mix is equal to the fresh weight in the no-mix.
- ✓ the weight of the different categories (less than 4 cm in diameter, greater than or equal to 4 cm in diameter, cull) of Toluca in the mix is equal to the weight of the different categories in the no-mix.
- ✓ the specific gravity of Toluca in the mix is equal to the specific gravity in the no-mix.

The specific gravity is an expression of density. Since there is a very high correlation between the starch content and the specific gravity of the tuber, which is an indicator for potato quality in processing, the specific gravity is the most widely accepted measurement of potato quality. The specific gravity equals the weight of tubers in the air divided by the difference between the weight in air and the weight in water (Agriculture.vic.gov.au, 2015).

The mean (M) and standard deviation (SD) to describe differences between rows and treatments were obtained using the "describe" function of R's psych-package.

CHAPTER III: Results

3.1 Ground beetle activity

In total, 1024 ground beetles were captured during the sampling week (M = 0.79, SD = 1.37, N = 1296). Out of those, 479 ground beetles fell into the pitfalls in the middles of the strips, while 545 ground beetles were captured along the strip edges. In the pitfalls positioned in the middles and edges of the mixed plots, 497 ground beetles were caught, while 527 were caught in the no-mix plots. Most ground beetles were caught in the middle of the oilseed rape strip (n = 201). The lowest number of ground beetles was caught in the middle of the fallow strip (n = 30) (Figure 9).

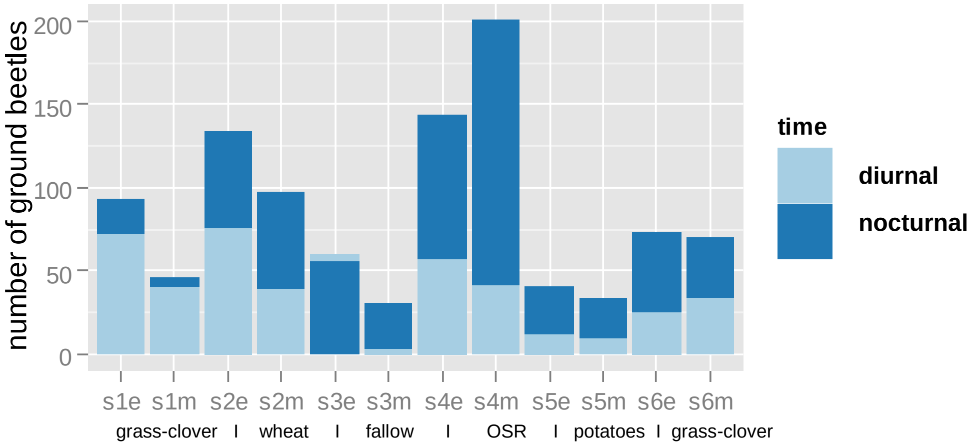


Figure 9: number of ground beetles trapped during the sampling week; s = strip, e = edge, m = middle

A generalized linear model (GLM) was fit to assess the effects of treatment, time of day, location and edge on the activity of ground beetles. The model of choice was zero-inflated and Poisson-distributed and contained the interaction effect between time and edge as the only predictors for the number of ground beetles: *zeroinfl (ground beetles = (time-1)*edge)* (Table 3).

Table 3: Model coefficients and 95% confidence intervals of the best model predicting ground beetle numbers based on variables defining edges and time of day. The model is Poisson distributed with a log-link function. For the edges, only significant predictors for the outcome variable are shown.

	estimate	**lower 95% CI**	**upper 95% CI**
diurnal	-2.9957	-4.1303	-1.8611
nocturnal	-0.5754	-0.9526	-0.1982
[grass]I[grass-clover]	3.6481	2.479	4.8173
grass-clover (strip 1)	3.204	1.9865	4.4215
[grass-clover]I[wheat]	3.4173	2.2446	4.5899
wheat	3.0583	1.8289	4.2876
[fallow]I[OSR]	3.3726	2.1859	4.5592

OSR	3.2166	2.0037	4.4295
[OSR]I[potatoes]	2.0186	0.2574	3.7798
[potatoes]I[grass-clover]	3.2566	1.9968	4.5164
grass-clover (strip 6)	3.6591	2.453	4.8652
nocturnal * [grass]I[grass-clover]	-3.4076	-4.853	-1.9623
nocturnal * grass-clover (strip 1)	-4.7081	-6.213	-3.2032
nocturnal * [grass-clover]I[wheat]	-2.3759	-3.6522	-1.0996
nocturnal * [wheat]	-1.9315	-3.2571	-0.6059
nocturnal * [fallow]I[OSR]	-2.1576	-3.4292	-0.8861
nocturnal * OSR	-1.3613	-2.6421	-0.0806
nocturnal * [potato]I[grass-clover]	-2.1809	-3.5442	-0.8177
nocturnal * grass-clover (strip 6)	-2.6177	-3.9507	-1.2848

Tukey's HSD test for the best model

Significant differences in ground beetle activity between day and night were found in the middles of the fallow strip ($p < 0.01$) and the oilseed rape strip ($p < 0.01$). In both strips, significantly higher amounts of ground beetles had fallen into the pitfall traps during the night.

In the edges of the strips, significant differences in ground beetle activity between day and night were found at the edge between grass and grass-clover in strip 1 ($p < 0.01$), between the wheat and the fallow ($p < 0.001$), between the fallow and the oilseed rape ($p < 0.05$) and between the potato strip and the grass-clover strip ($p < 0.05$).

Tukey's HSD test for the best model without the interaction factor

The results of Tukey's HSD test for the best model with the additional factor of time and edge showed that overall, the activity of ground beetles during the night was significantly higher than during the day ($p < 0.001$). Overall, the highest activity of ground beetles was observed in the middle of strip 4 (OSR). With $p < 0.001$, the activity of ground beetles there was significantly higher than in the middles of all the other strips.

Altogether, it depended on the crops whether there was a significant difference in ground beetle activity between the edges and the middle of a strip (Table 4). Strip 1 (grass-clover) was the only strip where the activity of ground beetles along both edges was considerably higher than in the middle.

Table 4: results of Tukey's HSD test on the Poisson distributed GLM *ground beetles = time + edge* to compare the activity of ground beetles between the edges and the middle of a strip; gc = grass-clover and OSR = oilseed rape.

strip	**[edge] - [middle] == 0**	**estimate**	**std. error**	**z value**	**Pr(>|z|)**	**95% CI lwr**	**upr**	**sign.**
1	[grass I gc] - [gc]	0.704	0.18	3.905	<0.01	0.119	1.289	**
	[gc I wheat] - [gc]	1.069	0.171	6.257	<0.01	0.515	1.623	***

2	[gc I wheat] - [wheat]	0.313	0.133	2.354	0.417	-0.118	0.744	
	[wheat I fallow] - [wheat]	-0.491	0.164	-2.993	0.102	-0.041	1.022	
3	[wheat I fallow] - [fallow]	0.693	0.224	3.1	0.076	-0.032	1.418	.
	[fallow I OSR] - [fallow]	1.569	0.201	7.816	<0.01	0.918	2.22	***
4	[fallow I OSR] - [OSR]	-0.333	0.109	-3.055	0.087	-0.688	0.021	.
	[OSR I potatoes] - [OSR]	-1.614	0.173	-9.325	<0.01	-2.176	-1.053	***
5	[OSR I potatoes] - [potatoes]	0.163	0.233	0.697	1	-0.594	0.919	
	[potatoes I gc] - [potatoes]	0.778	0.207	3.754	<0.01	0.106	1.45	**
6	[potatoes I gc] - [gc]	0.056	0.167	0.333	1	-0.485	0.596	

Significance codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1 (Adjusted p-values reported – single-step method). CI = confidence interval.

Furthermore, the activity of the ground beetles was dependent on the edge itself. With the exception of strips 1 and 5, there were significant differences in ground beetle activity between the edges of a strip (Table 5).

Table 5: results of Tukey's HSD on the Poisson distributed GLM *ground beetles = time + edge* to compare the activity of ground beetles along the edges; gc = grass-clover and OSR = oilseed rape.

strip	**[edge] - [edge] == 0**	**estimate**	**std. error**	**z value**	**Pr(>|z|)**	**95% CI lwr**	**upr**	**sign.**
1	[grass I gc] - [gc I wheat]	-0.365	0.135	-2.706	0.209	-0.803	0.073	
2	[gc I wheat] - [wheat I fallow]	0.804	0.155	5.173	<0.01	0.3	1.307	***
3	[wheat I fallow] - [fallow I OSR]	-0.875	0.154	-5.697	<0.01	-1.374	-0.377	***
4	[OSR I fallow] - [OSR I potatoes]	1.281	0.179	7.167	<0.01	0.701	1.861	***
5	[potatoes I OSR] - [potatoes I gc]	-0.615	0.196	-3.135	0.068	-1.252	0.021	.

Significance codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1 (Adjusted p-values reported – single-step method). CI = confidence interval.

Differences in time and location

A GLM including only the interaction term of time and location as a predictor was used to look for statistically significant differences between edges and middle rows in general. Tukey’s HSD test showed that the activity of the ground beetles was significantly affected by the time of day and the location. In the middle of the strips, the diurnal activity of the ground beetles was significantly lower ($p < 0.001$) than the nocturnal. On the strip edges, this difference was not as highly significant with $p < 0.01$.

The overall difference between the middles and the edges of the strips was non-significant, both during the day ($p = 0.982$) and during the night ($p = 0.134$).

Differences between treatments

The result of Tukey’s HSD test for a GLM including only the interaction term of time and treatment as a predictor showed that the differences in ground beetle activity between the mix and no-mix were not significant, with $p = 0.989$ in the morning and $p = 0.945$ in the evening.

There were also no statistically significant differences between treatments along the edges and the middles of the strips.

3.2 Dry root weight of potato varieties

Results of Welch's t-tests for Toluca: mix vs. no-mix

The mean dry root weight (mg/cm³) of Toluca in the mixed plots (M = 0.05, SD = 0.03, N = 36) was not significantly different from the mean dry root weight of Toluca in the no-mix plots (M = 0.06, SD = 0.05, N = 72), t (94.358) = -1.190, p = 0.236. T-tests for the difference in mean dry root weight in different depths (0-15, 15-30 and 30-45 cm) showed no statistically significant differences between Toluca in different treatments.

Results of Welch's t-tests for treatments: mix vs. no-mix

Overall, the mean dry root weight (mg/cm³) of the potato varieties in the mixed plots was not significantly different from the mean dry root weight of Toluca in the no-mix plots. There were further no significant differences at 0-15 and 15-30 cm depth. At 30-45 cm, the mean dry root weight of the potato varieties in the mixed plots (M = 0.03, SD = 0.02, N = 48) was significantly higher than the mean dry root weight of Toluca in the no-mix plots (M = 0.02, SD = 0.01, N = 24), t (69.982) = 3.833, $p < 0.001$.

Results of Welch's t-tests for mineral Nitrogen content in treatments: mix vs. no-mix

Overall, the NO_3 content (kg/ha) in the mixed plots (M = 16.24, SD = 7.44, N = 18) was significantly higher than in the no-mix plots (M = 9.81, SD = 2.81, N = 18), t (21.746) = 3.426, p = 0.002 (Figure 10). In 0-15 cm and 15-30 cm depth, the NO_3 content between treatments was not significantly different. In 30-45 cm depth, the NO_3 content in the mixed plots (M = 12.4, SD = 2.42, N = 6) was significantly higher than in the no-mix plots (M = 7.64, SD = 0.88, N = 6), t (6.29) = 4.5257, p = 0.003. The overall NH_4 content (kg/ha) in the mixed plots was significantly higher (M = 6.06, SD = 1.93, N = 18) than in the no-mix plots (M = 4.75, SD = 1.52, N = 18), t (32.248) = 2.254, p = 0.031 (Figure 10). In 0-15 cm and 15-30 cm depth, there was no statistically significant difference in NH_4 content between treatments. In 30-45 cm depth, the NH_4 content in the mixed plots (M = 5.59, SD = 1.37, N = 6) was significantly higher than in the no-mix plots (M = 3.54, SD = 0.68, N = 6), t (7.343) = 3.280, p = 0.012.

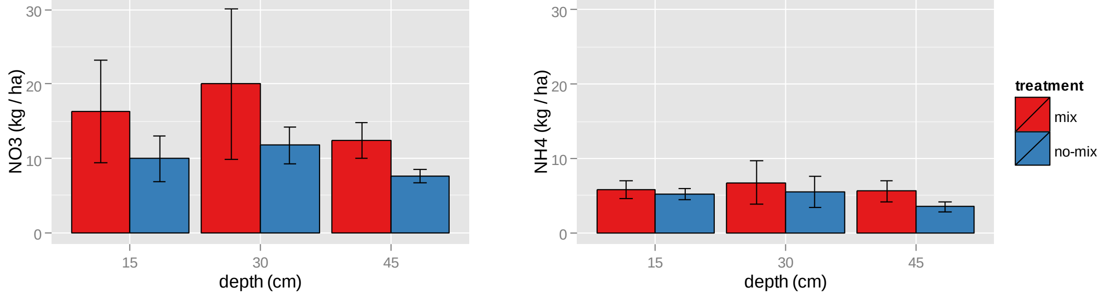


Figure 10: mean mineral Nitrogen content (NO_3 and NH_4) in the potato strip

The total mineral Nitrogen content ($N_{total} = NO_3 + NH_4$) was significantly higher in the mix (M = 22.30, SD = 7.38, N = 18) than in the no-mix (M = 14.57, SD = 3.52, N = 18), t (24.347) = 4.014, $p < 0.001$. The differences between treatments in 0-15 cm depth were not statistically significant. In 15-30 cm depth, the total N content in the mix (M = 26.77, SD = 8.11, N = 6) was significantly higher than in the no-mix (M = 17.29, SD = 2.03, N = 6), t (5.626) = 2.774, p = 0.034. This was also the case in 30-45 cm depth, where the mix (M = 17.99, SD = 3.55, N = 6) had again a significantly higher Nitrogen content than the no-mix (M = 11.18, SD = 1.39, N = 6), t (6.495) = 4.383, p = 0.003 (Figure 11).

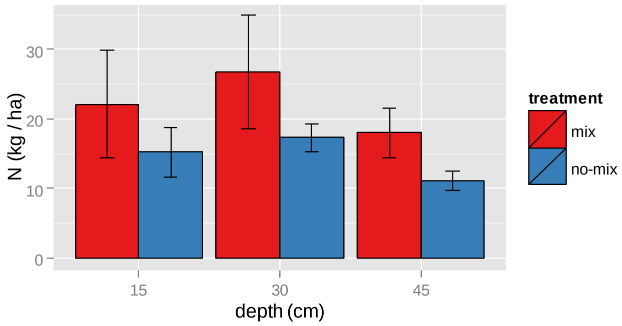


Figure 11: mean mineral Nitrogen content ($NO_3 + NH_4$) in the potato strip

Results of Welch's t-tests for varieties

The potato variety Annabelle had the highest mean root dry weight (mg/cm³), while Tiamo had the lowest (Figure 12). However, none of the differences in root dry weight between the potato varieties in the no-mix plots were statistically significant.

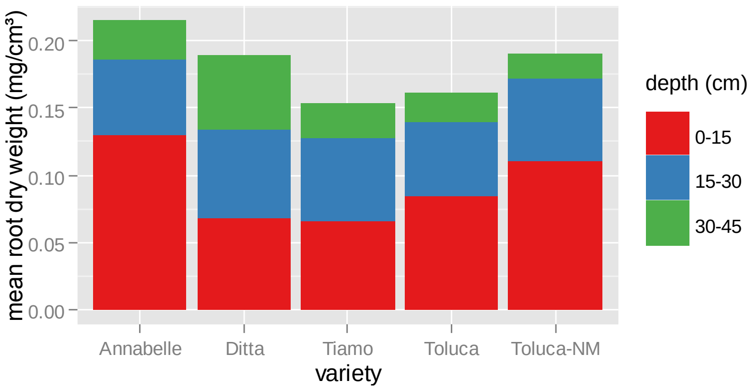


Figure 12: mean root dry weight of potato samples in three different depths;
Toluca-NM = Toluca no-mix

Root dry weight as predictor for tuber yield

A Gaussian distributed GLM containing all variables for root dry weight in different depths in interaction with the potato variety was fit to determine if those variables could predict the tuber yield: *yield = (variety-1)*(East15 + West15 + East30 + West30 + East45 + West45)*. Yield was defined as total fresh weight minus cull. There was evidence that the variety of which the root samples were taken from ($F = 197.491$, $p < 0.001$) was a highly significant predictor of the yield. The amount of deviance accounted for by the GLM was $D^2 = 0.96$.

A linear model fit to determine if the predicted values were equal to the observed values resulted in $R^2 = 0.95$ with $F = 674.3$ and $p < 0.001$. The graph plotting the observed yield versus the yield predicted by the GLM shows that the predicted values were close to the observed (Figure 13).

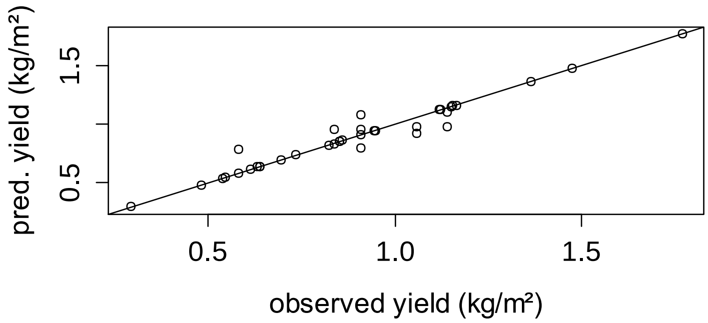


Figure 13: potato yield as predicted by the GLM compared to the observed yield

Mineral Nitrogen as predictor for root dry weight

A GLM containing the mineral Nitrogen content in 0-15 cm (N15), 15-30 cm (N30) and 30-45 cm (N45) depth as predictors was fit to determine if those variables could predict the root dry weight (RDW): *RDW = (treatment - 1) * (N15 + N30 + N45).* There was evidence that treatment was a highly significant predictor for the root dry weight ($p > 0.001$). The amount of deviance accounted for by the model was $D^2 = 0.95$. A linear model fit to determine if the predicted values of the GLM match the actual values resulted in $R^2 = 0.25$ with $p = 0.001$. The graph (Figure 14) shows that the predicted mineral Nitrogen content is not congruent with the observed.

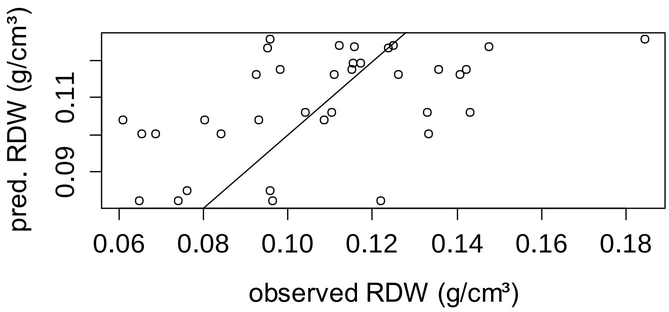


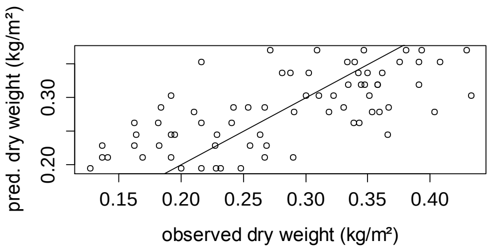
Figure 14: root dry weight as observed compared to root dry weight as predicted by the GLM

3.3 Yield differences between rows and treatments

3.3.2 Grass-clover (strip 1)

Dry weight

Overall, the mean dry weight per sample was highest in row 1 (0.34 kg/m²) and lowest in row 6 (0.24 kg/m²). In the no-mix plots, the samples had a higher mean dry weight (0.33 kg/m²) than in the mixed plots (0.24 kg/m²) (Figure 15). The results of the statistical analyses confirm both hypotheses stating that the yield is expected to be different at the strip edges and between treatments.



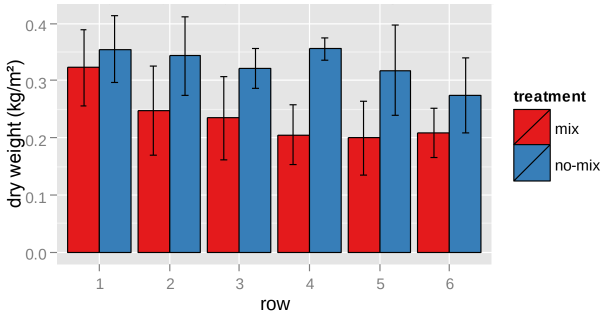


Figure 15: mean dry weight of grass-clover samples compared to those predicted by the model (left); mean dry weight of grass-clover samples as measured in strip 1 (right)

A generalized linear model (GLM) was fit to assess the effects of treatment, row, plot, the row coordinates (Field Y) and the plot coordinates (Field X) on the dry weight of the grass-clover samples. The model with the lowest AIC was *dry weight = (treatment-1) + Field Y* (Table 6).

Table 6: Model estimates and 95% confidence intervals of the best model predicting grass-clover dry weight. The GLM has an AIC of -194.12 (AIC_c = -193.52) and a Gaussian distribution.

	estimate	lower 95% CI	upper 95% CI
mix	0.2789	0.2503	0.3074
no-mix	0.3704	0.3419	0.3990
Field Y	-0.0340	-0.0504	-0.0175

There was evidence that both the treatment ($F = 797.614$, $p < 0.001$) and the row coordinates ($F = 16.431$, $p < 0.001$) were highly significant predictors of the dry weight. The amount of deviance accounted for by that model was $D^2 = 0.96$. A linear model fit to determine if the predicted values matched the observed dry weight resulted in $R^2 = 0.44$ with $F = 58.14$ and $p < 0.001$.
Since the row coordinates in the model were numeric variables, pairwise comparisons were not possible. For the purpose of pairwise comparisons, a GLM containing only the interaction effects between the categorical variables of row and treatment was fit. Tukey's HSD test for that model showed significant differences between rows, both between and within treatments. Within the mixed plots, the dry weight of the samples from row 1 was significantly higher than the dry weight of the samples from rows 4, 5 and 6. The differences between the edge rows in the different treatments turned out to be statistically non-significant. Tukey's HSD test on a GLM consisting of the additional effect of row and treatment as predictor showed highly significant differences between treatments. Also, row 1 had a significantly higher dry weight than row 6 (Table 7).

Table 7: results of Tukey's HSD test on the GLMs *dry weight = row * treatment* and *dry weight = row + treatment*

[row] - [row] == 0	**estimate**	**std. error**	**z value**	**Pr(>|z|)**	**95% CI lwr**	**upr**	**sign.**
no-mix 1 - no-mix 6	0.081	0.035	2.302	0.475	-0.034	0.196	
mix 1 - mix 4	0.117	0.035	3.329	0.042	0.002	0.232	*
mix 1 - mix 5	0.123	0.035	3.502	0.023	0.008	0.238	*
mix 1 - mix 6	0.114	0.035	3.239	0.055	-0.001	0.229	.
mix 1 - no-mix 1	-0.032	0.035	-0.922	0.999	-0.147	0.083	
mix 6 - no-mix 6	-0.065	0.035	-1.858	0.785	-0.180	0.050	
row 1 - row 6	0.097	0.025	3.866	0.002	0.026	0.169	**
[treatment] - [treatment] == 0							
mix - no-mix	-0.092	0.015	-6.289	< 0.001	-0.120	-0.063	***

Significance codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1 (Adjusted p-values reported – single-step method)

Nitrogen content

A t-test was performed to look for significant differences in Nitrogen content (% N per plot) in the grass-clover between treatments. The result of Welch’s t-test showed that the **mean N content** in the mix (M = 2.86, SD = 0.18, N = 6) was not significantly different from the mean N content in the no-mix (M = 2.90, SD = 0.15, N = 6), t (68.241) = -1.046, p = 0.299.

3.3.3 Wheat (strip 2)

Differences in dry weight and the kernel weight between the edges and the middle rows of the wheat strip depended on treatment. For the biomass, there were no significant differences between rows and treatments.

The **mean dry weight** of all samples was slightly higher in the mix (0.62 kg/m²) than in the no-mix (0.61 kg/m²). Irrespective of treatment, row 4 had the highest mean dry weight of all samples (0.66 kg/m²), while row 1 had the lowest (0.57 kg) (Figure 16).

The **mean kernel weight** in the mixed plots was also lower (0.25 kg/m²) than in the no-mix plots (0.30 kg/m²). Likewise, row 4 had the highest mean kernel weight (0.32 kg) and row 1 the lowest (0.23 kg) (Figure 17).

The **mean biomass** in the mixed plots was lower (0.61 kg/m²) than in the no-mix plots (0.69 kg/m²). Overall, row 3 had the highest mean biomass (0.71 kg/m²) and the edge rows 1 and 4 the lowest (0.62 kg/m²).

GLM for dry weight (kg/m²)

A generalized linear model (GLM) was fit to assess the effects of treatment, row, plot, weed dry weight, the coordinates of the sampling locations and the Nitrogen content of stems and kernels on the **dry weight** of the wheat samples. The GLM with the lowest AIC was Gaussian distributed and contained the interaction effect of treatment with the row coordinates (Field Y) in addition to the weed dry weight in the sampling area as predictors: *wheat dry weight = (treatment-1)*(Field Y) + weed dry weight* (Table 8).

Table 8: Model estimates and 95% confidence intervals of the best model predicting wheat dry weight. The model has an AIC of -68.20 (AIC_c = -66.15) and a Gaussian distribution.

	estimate	lower 95% CI	upper 95% CI
mix	0.2181	-0.0165	0.4527
no-mix	0.7286	0.5058	0.9514
Field Y	0.1113	0.0584	0.1642
weeds	-0.3055	-0.7376	0.1266
no-mix * Field Y	-0.1343	-0.2094	-0.0591

The amount of deviance accounted for by this model was $D^2 = 0.97$. There was evidence that treatment was a significant predictor of the dry weight (F = 741.686, $p < 0.001$), as well as the interaction effect between treatment and the no-mix (F = 12.268, $p = 0.001$).
A linear regression model that set the predicted values in relation to the observed values resulted in an R^2 value of 0.30 with F = 21.63 and $p < 0.001$ (Figure 16).

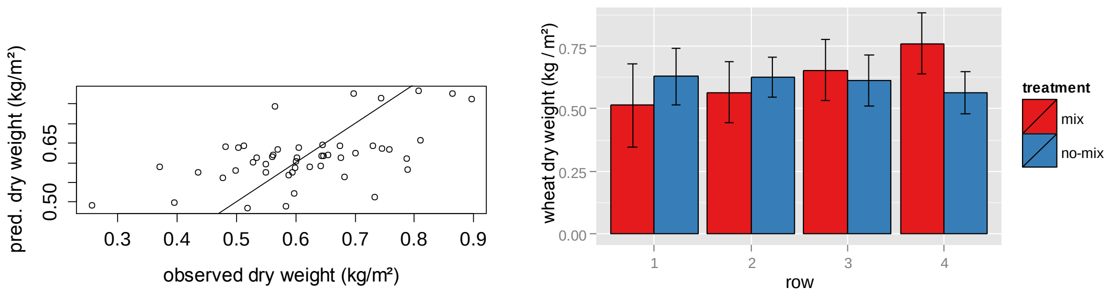


Figure 16: comparison of observed wheat dry weight with dry weight predicted by the GLM (left) and the distribution of dry weight between rows and treatments (right).

Concerning pairwise comparisons, Tukey's HSD test for the interaction effect between rows and treatments showed a significant difference in wheat dry weight between rows 1 and 4 of the mixed plots (p = 0.005); row 4 had a higher dry weight than row 1. Further, there was a significant difference between rows 2 and 4 of the mixed plots (p = 0.072); again, row 4 had a higher dry weight than row 2.
Between treatments, row 4 in the mixed plots had a significantly higher dry weight than row 4 in the no-mix plots (p = 0.065). 95% confidence intervals on the difference between the rows mentioned above indicated that the only significant evidence for a difference in wheat dry weight was between row 1 and row 4 in the mixed plots (0.044, 0.452).

GLM for kernel weight (kg/m²)

A generalized linear model (GLM) was fit to assess the effects of treatment, row, plot, weed dry weight, the coordinates of the sampling locations and the Nitrogen content of stems and kernels on the **kernel weight (kg/m²)** of the wheat samples. The GLM with the lowest AIC was *kernel weight = (treatment-1)*Field Y + plot* (Table 9).

Table 9: Model estimates and 95% confidence intervals of the best model predicting wheat kernel weight. The GLM has an AIC of -106.02 (AIC_c = -91.01) and a Gaussian distribution. For the plots, only significant predictors for the outcome variable are shown.

	estimate	lower 95% CI	upper 95% CI
mix	-0.0581	-0.2113	0.0952
no-mix	0.3219	0.1687	0.4752
Field Y	0.0963	0.0630	0.1296
plot 4	-0.1786	-0.2752	-0.0819
plot 5	-0.1245	-0.2211	-0.0279
plot 6	-0.1555	-0.2521	-0.0588
plot 10	-0.0996	-0.1962	-0.0030
no-mix * Field Y	-0.1089	-0.1560	-0.0619

The amount of deviance accounted for by this model was $D^2 = 0.96$. There was evidence that treatment (F = 378.813, $p < 0.001$), Field Y (F = 12.160, $p = 0.001$), the interaction between the two (F = 20.598, $p < 0.001$) and plot (F = 3.242, $p = 0.004$) were significant predictors for the kernel weight. A linear model to determine if the predicted values were equal to the observed values resulted in $R^2 = 0.67$ with F = 95.58 and $p < 0.001$. The graph plotting the observed kernel weight versus the kernel weight as predicted by the GLM shows that the values are close to each other (Figure 17).

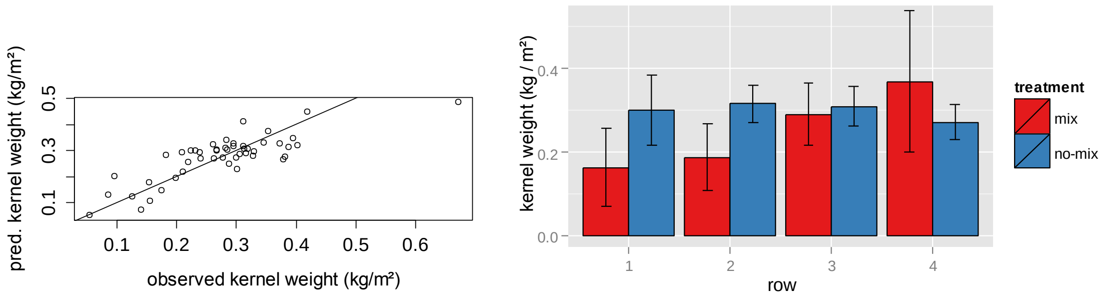


Figure 17: comparison of observed wheat kernel weight with kernel weight as predicted by the GLM (left) and the distribution of kernel weight between rows and treatments (right).

A GLM containing only the interaction effect between the categorical variables of row and treatment was fit for the purpose of pairwise comparisons. Tukey's HSD test for the interaction effect between rows and treatments showed a significant difference in **kernel weight (kg/m²)** between rows 1 and 4 of the mixed plots (p = 0.001); row 4 had a higher kernel weight than row 1. Further, there was a significant difference between rows 2 and 4 of the mixed plots (p =0.008); again, row 4 had a higher kernel weight than row 2.

Tukey's HSD test for a GLM with the additional effects of row and treatment revealed no statistically significant differences between the overall kernel weights in the different rows. The **kernel weight (kg/m²)** in the no-mix plots, however, was significantly higher than in the mixed plots ($p = 0.099$). Yet, a 95 % confidence interval indicated no significant evidence of a difference (-0.008, 0.102).

Harvest index

The harvest index is the ratio between kernel weight and total dry weight. The result of Welch's t-test showed that the **mean harvest index** in the mixed plots (M = 0.39, SD = 0.13, N = 24) was significantly lower than the mean harvest index in the no-mix plots (M = 0.49, SD = 0.04, N = 24), t (26.838) = -3.6759, $p = 0.001$.

Biomass

Welch's t-test to test the hypothesis that the biomass differed between treatments showed that there was no significant difference between the mix (M = 0.61, SD = 0.14, N = 24) and the no-mix (M = 0.69, SD = 0.09, N = 24), t (-2.46), $p = 0.018$. In order to make pairwise comparisons between rows in the different treatments, a GLM containing only the interaction effect between rows and treatments was fit. Tukey's HSD test showed no significant differences in biomass between the rows within their respective treatments.

Nitrogen content

A t-test was performed to look for significant differences in the Nitrogen content of stems and grains between treatments. The result of Welch's t-test showed that the mean **Nitrogen content of the grains** in the mixed plots (M = 2.17, SD = 0.09, N = 24) was significantly higher than the mean Nitrogen content of the grains in the no-mix plots (M = 1.91, SD = 0.13, N = 24), t (40.055), $p < 0.001$. The results for the mean **Nitrogen content of the stems** were similar. The mean Nitrogen content of the stems was significantly higher in the mixed plots (M = 0.42, SD = 0.05, N = 24) than in the no-mix plots (M = 0.32, SD = 0.02, N = 24), t (27.982), $p < 0.001$.

3.3.4 Oilseed rape (strip 4)

For the oilseed rape, the initial hypothesis that the yields would be different between rows and treatments has not been confirmed.

Dry weight

The mean dry weight of the samples taken in the mixed plots (0.54 kg/m²) was lower than in the no-mix plots (0.57 kg/m²). Overall, the row with the highest mean dry weight was row 4 with 0.60 kg/m², while the row with the lowest was row 1 with 0.51 kg/m² (Figure 18).
A generalized linear model (GLM) was fit to assess the effects of treatment, row, plot, the coordinates of the sampling locations and the Nitrogen content of stems and kernels on the **dry weight** of the oilseed rape samples. The GLM with the lowest AIC contained the Nitrogen content of the seeds as the only predictor (Table 10).

Table 10: Parameter estimates and 95% confidence intervals of the best model predicting OSR dry weight. The model has an AIC of -30.76 (AIC_c = -30.22) and a Gaussian distribution.

	estimate	**lower 95% CI**	**upper 95% CI**
Intercept	-0.0015	-0.8271	0.8240
N seeds	0.1362	-0.0657	0.3381

The amount of deviance accounted for by the model was D^2= 0.03. Nevertheless, there was evidence that the Nitrogen content of the seeds was not a significant predictor for the dry weight (F = 1.747, *p* = 0.192). A linear model fit to compare the predicted with the observed values resulted in R^2 = 0.02 with F = 1.747 and *p* = 0.192. The predicted values were not congruent with the observed (Figure 18).

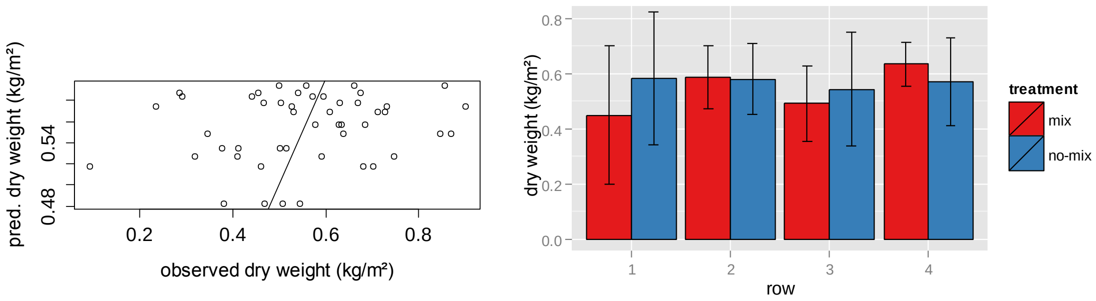


Figure 18: comparison of the observed dry weight of OSR with the dry weight predicted by the GLM (left) and the distribution of dry weight between rows and treatments (right).

In order to look for differences in **dry weight** between rows and within treatments, a GLM containing only the interaction effect between rows and treatments was fit. Tukey's HSD test for that interaction effect showed no statistically significant differences in dry weight between rows and treatments.

Number of seeds per m²

The effects of treatment, row, plot, the coordinates of the sampling locations and the Nitrogen content of stems and kernels on the **number of seeds per m²** were assessed with another GLM. The GLM with the lowest AIC was Gaussian distributed and contained the coordinates for the rows (Field Y) in addition to the Nitrogen content of the stems: *number of seeds = Field Y + N stems* (Table 11).

Table 11: Parameter estimates and 95% confidence intervals of the best model predicting OSR number of seeds. The model has an AIC of 1018.30 (AIC_c = 1019.23) and a Gaussian distribution.

	estimate	**lower 95% CI**	**upper 95% CI**
Intercept	2304	-47804.2725	52412.3900
Field Y	4074	-630.6227	8777.6040
N stems	-34575	-53201.7231	-15947.7100

While there was evidence that the row coordinates were not a very significant predictor for the seed number (F = 2.880, *p* = 0.096), the Nitrogen content of the stems was a highly significant predictor (F = 13.235, $p < 0.001$). The amount of deviance accounted for by the model was $D^2 = 0.21$. The linear model fit to test if the values predicted by the GLM matched the observed resulted in $R^2 = 0.25$ with F = 16.47 and $p < 0.001$ (Figure 19).

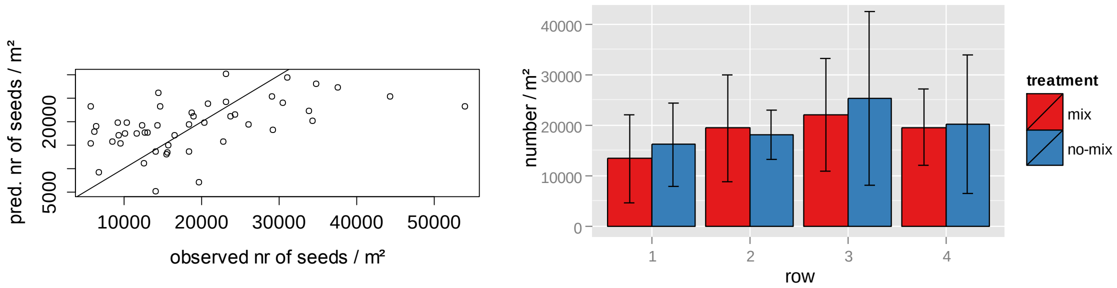


Figure 19: comparison of the counted number of seeds with the number of seeds as predicted by the GLM (left) and the counted number of seeds as distributed between rows and treatments (right).

In order to look for row effects within treatments, a GLM containing only the interaction effects between the categorical variables of row and treatment was fit. However, Tukey's HSD test showed no statistically significant differences between rows within the different treatments.

Seed weight

The **mean seed weight** of the samples in the mixed plots was 0.058 kg/m² (SD = 0.03, N = 24), which was almost similar to the mean seed weight in the no-mix plots (M = 0.062, SD = 0.04, N = 24). Overall, row 3 had the highest mean seed weight (0.073 kg/m²), while row 1 had the lowest (0.045 kg/m²).

A GLM was fit to assess the effects of treatment, row, plot, the coordinates of the sampling locations and the Nitrogen content of stems and kernels on the seed weight. The GLM with the lowest AIC was Gaussian distributed and contained the coordinates for the rows (Field Y) in addition to the Nitrogen content of the stems as the only predictors: *seed weight = Field Y + N stems* (Table 12).

Table 12: Model estimates and 95% confidence intervals of the best model predicting OSR seed weight. The model has an AIC of -195.96 (AIC_c = -195.03) and a Gaussian distribution.

	estimate	**lower 95% CI**	**upper 95% CI**
Intercept	-0.0059	-0.1668	0.1551
Field Y	0.0163	0.0012	0.0315
N stems	-0.1407	-0.2005	-0.0808

There was evidence that both the row coordinates (F = 4.492, $p = 0.039$) and the Nitrogen content of the stems (F = 21.233, $p < 0.001$) were significant predictors for the seed weight. The amount of deviance accounted for by the GLM was $D^2 = 0.36$. With $R^2 = 0.35$, F = 26.30 and p < 0.001, the predicted values were not congruent with the observed (Figure 20).

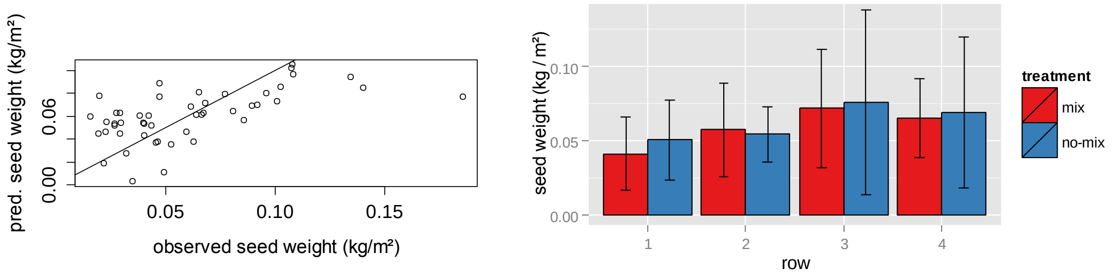


Figure 20: comparison of the measured seed weight (kg/m²) with the seed weight as predicted by the GLM (left) and the mean seed weight as measured (right).

Tukey's HSD test for a GLM containing only the interaction effect between treatment and rows showed no significant differences between rows and treatments.

Thousand kernel weight (TKW)

The **mean thousand kernel weight (TKW)** in the mixed plots was 3.16 g (SD = 0.36, N = 24), while the mean TKW in the no-mix plots was 3.06 g (SD = 0.56, N = 24). Overall, row 4 had the highest TKW with 3.25 g, while row 2 had the lowest with 3.00 g (Figure 21).
In order to assess the effects of treatment, row, plot, the coordinates of the sampling locations and the Nitrogen content of stems and kernels on the **TKW**, another GLM was fit. The GLM with the lowest AIC was Gaussian distributed and contained the coordinates for the plots (Field X) in addition to the treatment as the only predictors: *TKW = Field X + treatment* (Table 13).

Table 13: Parameter estimates and 95% confidence intervals of the best model predicting OSR thousand seed weight. The model has an AIC of 48.47 (AIC_c = 49.39) and a Gaussian distribution.

	estimate	lower 95% CI	upper 95% CI
Field X	0.0041	0.0025	0.0057
mix	2.6796	2.4409	2.9182
no-mix	2.5520	2.3057	2.7983

The amount of deviance accounted for by the model was $D^2 = 0.99$. There was evidence that both Field X (F = 2659.68, $p < 0.001$) and treatment (F = 280.76, $p < 0.001$) were highly significant predictors for the thousand kernel weight. A linear model fit to check if the predicted values match the observed (Figure 21) resulted in $R^2 = 0.36$ with F = 27.58 and $p < 0.001$. Tukey's HSD test for treatment showed no statistically significant differences in TKW between mix and no-mix ($p = 0.461$).

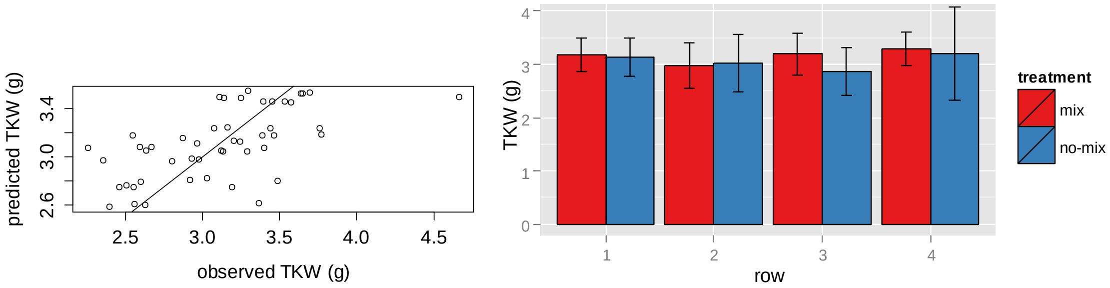


Figure 21: comparison of the measured TKW with the TKW as predicted by the GLM (left) and the mean TKW as measured (right).

In terms of pairwise comparisons, Tukey's HSD test for the interaction effect of rows and treatments showed no statistically significant differences between rows within the different treatments.

Nitrogen

The differences in mean Nitrogen content between treatments, both for the stems and the seeds, were statistically not significant.

3.3.5 Potatoes (strip 5)

In fresh weight and yield, there were no statistically significant differences between treatments. However, row 1 was significantly different from row 2, which partially confirmed the initial hypothesis that yields in the edge rows would be different. In specific gravity, there were no statistically significant differences between treatments and varieties.

Fresh weight

The **mean fresh weight** of potatoes in the mixed plots (4.557 kg/m²) was almost the same as in the no-mix plots (4.555 kg/m²). With 5.057 kg/m², row 1 had the highest fresh weight, while row 2 had the lowest (4.267 kg/m²) (Figure 22).

In order to determine which factors influenced the **fresh weight** of the potatoes, a GLM was fit. The GLM with the lowest AIC was Gaussian distributed: *fresh weight = Field X + row + (treatment-1)*NO_3* (Table 14).

Table 14: Parameter estimates and 95% confidence intervals of the best model predicting potato fresh weight. The model has an AIC of 91.34 (AIC_c = 96.08) and a Gaussian distribution.

	estimate	lower 95% CI	upper 95% CI
Field X	0.0674	0.0142	0.1207
row 1	4.4354	3.5037	5.3671
row 2	3.6447	2.7130	4.5764
row 3	3.7675	2.8358	4.6992
row 4	3.8875	2.9558	4.8192
no-mix	-1.8555	-3.6484	-0.0627
NO_3	0.0290	-0.1045	0.1625

no-mix * NO_3	0.4668	0.0610	0.8726

There was evidence that the coordinates for the plots ($F = 2586.184$, $p < 0.001$) and the row factors ($F = 130.037$, $p < 0.001$) were highly significant predictors for the fresh weight, while the interaction effect between treatment and NO_3 was not as significant a predictor for the fresh weight ($F = 5.084$, $p = 0.029$).
The amount of deviance accounted for by this GLM was $D^2 = 0.98$. A linear model fit to test if the values predicted by the GLM were similar to the observed values resulted in $R^2 = 0.45$ with $F = 40.55$ and $p < 0.001$ (Figure 22).

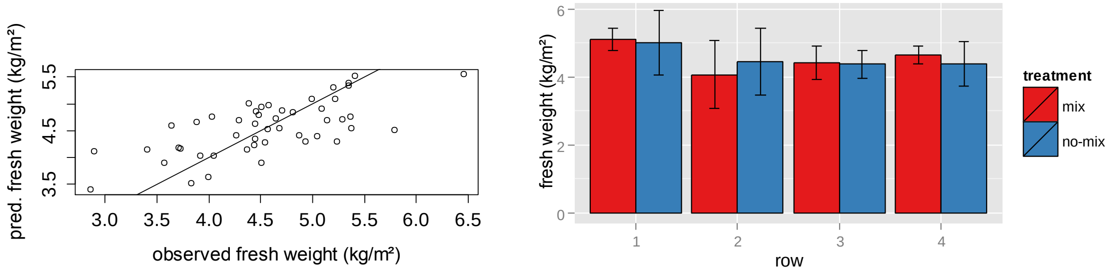


Figure 22: comparison of the fresh weight as measured with the fresh weight as predicted by the GLM (left); measured fresh weight as distributed between rows and treatments (right).

The output of Tukey's HSD test of a GLM consisting only of the additional effect of rows and treatments showed no statistically significant difference between treatments ($p = 0.992$). Altogether, the fresh weight in row 1 was significantly higher than in row 2 ($p = 0.023$) and in row 3 ($p = 0.077$). However, a 95 % confidence interval on the difference between row 1 and 3 showed evidence that this difference may not be significant (-1.383, 0.047). The output of Tukey's HSD test on a GLM consisting of the interaction effect of rows and treatments showed no statistically significant differences between rows within and between treatments.

Fresh weight of different varieties in the mix

Between the varieties, Ditta had the highest **fresh weight** with 5.610 kg/m², while Tiamo had the lowest with 3.739 kg/m² (Figure 23).
For the **fresh weight in the mixed plots**, a GLM based on variables defining the position in the field, the mineral Nitrogen content and the different potato varieties was fit. The best GLM had a Gaussian distribution and was composed of the additional effect of the coordinates for the plots, the row factors and the variety: *fresh weight = Field Y + row + variety* (Table 15).

Table 15: Parameter estimates and 95% confidence intervals of the best model predicting the potato fresh weight of different varieties. The GLM has an AIC of 107.36 ($AIC_c = 108.78$) and a Gaussian distribution. For rows and varieties, only significant predictors for the outcome variable are shown.

	estimate	**lower 95% CI**	**upper 95% CI**
Intercept	5.2103	4.3122	6.1083
Field Y	-0.3358	-0.5964	-0.0752
row 2	-1.0475	-1.7370	-0.3580

Ditta	1.5722	0.7904	2.3540

There was evidence that the variety (F = 8.894, $p < 0.001$) was a highly significant predictor for the fresh weight. The amount of deviance accounted for by this GLM was $D^2 = 0.30$. A linear model fit to check for consistency of the observed with the predicted values resulted in $R^2 = 0.30$ with F = 42.02 and $p < 0.001$.

Tukey's HSD test on the rows showed that row 1 had a significantly higher fresh weight than row 2 (p = 0.003), row 3 (p = 0.077) and row 4 (p = 0.055). However, 95 % confidence intervals only showed evidence for statistically significant differences between rows 1 and 2. Between the varieties, Tukey's HSD test showed that Ditta had a significantly higher fresh weight than all the other varieties ($p < 0.001$).

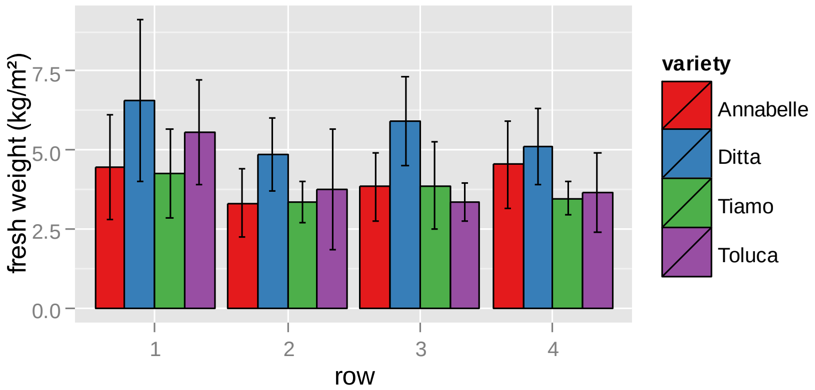


Figure 23: Mean fresh weight of four different potato varieties in the mixed plots, compared by rows

Fresh weight of Toluca

Overall, the **mean fresh weight (kg/m²)** of Toluca in the mixed plots (M = 4.09, SD = 1.62, N = 24) was not significantly different from the mean fresh weight of Toluca in the no-mix plots (M = 4.56, SD = 0.78, N = 24), t (33.125) = -1.2666, p = 0.214. The mean amount of **potatoes smaller than 4 cm** was slightly higher in the mix (M = 0.10, SD = 0.13, N = 24) than in the no-mix (M = 0.09, SD = 0.05, N = 24). However, this difference was statistically not significant. The mean amount of **potatoes bigger than or equal to 4 cm** was higher in the no-mix (M = 3.99, SD = 0.74, N = 24) than in the mix (M = 3.53, SD = 1.50, N = 24). Again, this difference was statistically not significant. The mean amount of **cull** was almost the same in the mix (M = 0.46, SD = 0.68, N = 24) as in the in the no-mix (M = 0.47, SD = 0.30, N = 24).

Fresh weight – machine harvest

The fresh weight in the no-mix plots (4.578 kg/m²) was higher than in the mixed plots (4.308 kg/m²). With 5.176 kg/m², row 1 had the highest fresh weight, while row 2 had the lowest (3.929 kg/m²) (Figure 24).

Based on variables defining treatment, position in the field and mineral Nitrogen content, a GLM was fit. The best GLM to predict the fresh weight had a Gaussian distribution: *fresh weight = Field Y * (treatment-1) +* NO_3 *+ plot + row* (Table 16).

Table 16: Parameter estimates and 95% confidence intervals of the best model predicting potato fresh weight (machine harvest). The GLM has an AIC of 21.35 (AIC_c = 41.75) and a Gaussian distribution. For the plots and rows, only significant predictors for the outcome variable are shown.

	estimate	lower 95% CI	upper 95% CI
Field Y	-0.2826	-0.3781	-0.1871
mix	6.3973	-5.4170	18.2116
no-mix	6.1883	-2.9256	15.3022
NO3	-0.1486	-2.3088	2.0115
plot 7	0.5590	0.1581	0.9599
plot 9	0.5381	0.2081	0.8680
row 2	-0.9166	-1.0999	-0.7333
Field Y * no-mix	-0.0967	-0.2282	0.0347

There was evidence that all variables were highly significant predictors for the fresh weight ($p < 0.001$). The amount of deviance accounted for by the model was $D^2 = 0.99$. In order to determine if the predicted values by the GLM correspond to the observed values, a linear model was fit resulting in $R^2 = 0.88$ with F = 367.40 and $p < 0.001$.

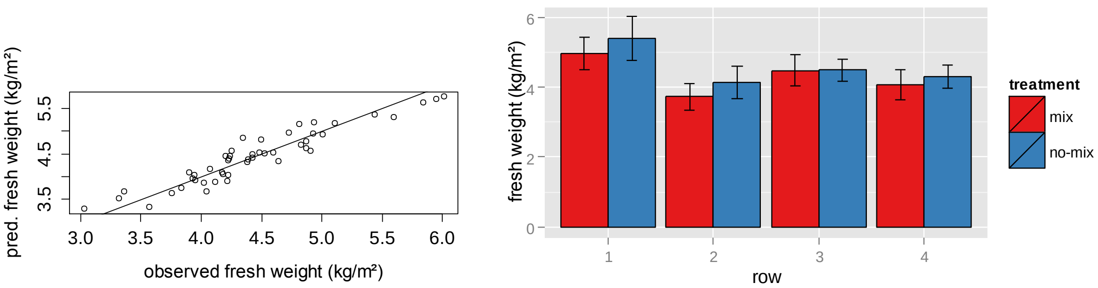


Figure 24: comparison of the fresh weight as measured with the fresh weight as predicted by the GLM (left); measured fresh weight as distributed between rows and treatments (right).

For the purpose of pairwise comparisons, a GLM consisting only of the interaction effect between row and treatment as predictor was fit. The output showed that in the mixed plots, the fresh weight of row 1 was significantly higher than in row 2 ($p < 0.001$) and in row 4 ($p < 0.01$). In the no-mix, the fresh weight of row 1 was significantly higher than in all the other rows ($p < 0.001$ for rows 2 and 4, $p < 0.01$ for row 3). The output of Tukey's HSD for a GLM with the additional effect of row and treatment as predictors showed a significant difference between treatments ($p = 0.031$), with a higher fresh weight in the no-mix. Overall, row 1 had a significantly higher fresh weight than all the other rows ($p < 0.001$). Row 2 had a significantly lower yield than row 3 ($p = 0.009$).

Yield

The mean **yield** (defined as total fresh weight minus cull) was higher in the no-mix plots (4.087 kg/m²) than in the mixed plots (3.998 kg/m²). Row 1 had the highest collective mean yield with 4.345 kg/m², while row 2 had the lowest (3.824 kg/m²).

Based on variables defining treatment, position in the field and mineral Nitrogen content, a GLM was fit to determine which factors influence the **yield**. The best GLM was Gaussian distributed and consisted of the interaction effect between the treatments and the plot coordinates: *Yield = Field X*(treatment-1)* (Table 17).

Table 17: Parameter estimates and 95% confidence intervals of the best model predicting potato yield. The model has an AIC of 107.36 (AIC_c = 108.78) and a Gaussian distribution.

	estimate	**lower 95% CI**	**upper 95% CI**
Field X	-0.0012	-0.0884	0.0860
mix	4.0065	3.3880	4.6251
no-mix	3.3946	2.8180	3.9712
Field X * no-mix	0.1052	-0.0103	0.2207

There was evidence that both the coordinates for the plots ($F = 1329.502$, $p < 0.001$) and the treatment ($F = 145.605$, $p < 0.001$) were highly significant predictors for the yield. The amount of deviance accounted for by that GLM was $D^2 = 0.97$. The linear model fit to compare the predicted with the actual values resulted in $R^2 = 0.12$ with $F = 7.783$ and $p = 0.007$. Tukey's HSD test on a GLM containing only the additional effect between rows and treatments showed no statistically significant differences in **yield**, neither between rows nor between treatments. Tukey's HSD test on a GLM containing the interaction effect between rows and treatments showed no statistically significant differences either.

Yield of different varieties in the mix

Similar to the fresh weight, Ditta was also the variety with the highest yield (5.086 kg/m²). The variety with the lowest yield was Tiamo with 3.407 kg/m². The **yield in the mixed plots** was best predicted by a GLM with a Gaussian distribution consisting of the additional effect of variety and NO_3 content: *yield = variety +* NO_3 (Table 18).

Table 18: Parameter estimates and 95% confidence intervals of the best model predicting potato yield of different varieties. The model has an AIC of 331.99 (AIC_c = 332.93) and a Gaussian distribution. For the varieties, only significant predictors for the outcome variable are shown.

	estimate	**lower 95% CI**	**upper 95% CI**
Annabelle	2.7841	1.6615	3.9067
Ditta	4.1974	3.0748	5.3200
Tiamo	2.5187	1.3961	3.6413
Toluca	2.7404	1.6177	3.8630
NO_3	0.1314	-0.0151	0.2780

There was evidence that variety was a highly significant predictor for yield ($F = 222.350$, $p < 0.001$), while the NO_3 content was not such a significant predictor ($F = 3.089$, $p = 0.082$). The amount of deviance accounted for by the model was $D^2 = 0.90$. A linear model fit to check for congruence between the observed and the predicted values resulted in $R^2 = 0.22$ with $F = 28.47$ and $p < 0.001$.

Tukey's HSD test showed that Ditta had a significantly higher yield than Annabelle ($p = 0.001$) and both Toluca and Tiamo ($p < 0.001$). Since the best model did not contain the factor row, an additional GLM was fit to make pairwise comparisons. This GLM contained the additional effect of variety and row as the only predictors. However, the output of Tukey's HSD test showed no statistically significant differences between the rows.

Potato sizes – results of Welch's t-test to compare between treatments

The mean yield of potatoes less than 4 cm in diameter in the mixed plots (M = 0.18, SD = 0.06, N = 24) was significantly higher than in the no-mix plots (M = 0.09, SD = 0.05, N = 24), t (44.90) = 4.738, $p < 0.001$. Between treatments, there was no statistically significant difference in the yield of potatoes equal to or bigger than 4 cm in diameter ($p = 0.423$). There was also no statistically significant difference in cull between treatments ($p = 0.484$).

Specific gravity

There were no significant differences in specific gravity between treatments (p = 1). Within the mixed plots, there were no significant differences between the different potato varieties (p = 1). The mean specific gravity of Toluca in the mix (M = 1.07, SD = 0.02, N = 6) was the same as in the no-mix (M = 1.07, SD = 0, N = 24).

CHAPTER IV: Discussion

4.1 Limitations of the approach

4.1.1 Ground beetle sampling

All results of the experiment have been potentially influenced by the weather. Especially during the sampling week for the ground beetles, heavy rainfalls created conditions on the field that are usually not encountered during this time of the year. Since the ground beetles were trapped alive, they were released after they had been caught. The point of release was outside a two meter radius from the pitfall location. However, it cannot be ruled out that the same ground beetles fell into the same pitfalls over and over again. Since the goal was to compare the ground beetle activity in different locations, the absolute numbers were secondary. Considering that the whole field was subjected to the same weather conditions, the relative numbers were assumed to be representative.

4.1.2 Root sampling

During the root sampling of the potatoes, it happened that potatoes were in the way of the auger, taking up space that could have been filled by roots. The potato pieces were weighed and then discarded; however, this proved to be insufficient to calculate the volume they filled in the core. Since there were four different varieties, it was impossible to determine their individual mass density later because there is no literature about the mass density of the potato tubers of different varieties. Nonetheless, in the soil, space taken up by potatoes is not filled by roots. Therefore, the amount of roots per cm^3 depends on their location in the soil. Since all samples were taken at the same distance from the stem of the potato plant, they all reflect the conditions in that precise location, and thus the amount of roots per cm^3 that can be found there.

4.1.3 Harvesting

It was again the wet weather that delayed the harvest time. This should have no influence on the comparison between rows and treatments though, because all strips were exposed to the same conditions. Since the Unifarm does not have equipment for processing harvested oilseed rape, the same equipment as for wheat was used wherever possible. Especially the air exhaust that was employed for cleaning the seeds proved to be not suitable for the cleaning of oilseed rape seeds. This may have lowered the total amount of seeds; however, the relation between rows and treatments is assumed to be correct in any case because all samples had been subjected to the same procedure.

Further, the automatic seed counting machines worked with a light barrier, which also counted dust particles as seeds. Since it would have taken too long to count a thousand seeds for each sample by hand, the seed counting machines were used in spite of this flaw. Again, every sample passed the same procedure, so the proportions are assumed to be correct.

For the potato samples, the sampling area was 0.75 x 1.2 m, containing four potato plants. This area represented 12% of a row, while the machine later harvested all the potatoes that were still left in the rows, so 88% in the no-mix and 76% in the mix. When comparing the results of the machine harvest with the results of the hand harvests, it became obvious that the variance within the harvests was not equal. During the hand harvests, every single potato was dug up, and the potatoes were washed mechanically before they were strictly sorted into categories. During the machine harvest, small potatoes fell through the conveyor belt before they could make it into the net bag, and rotten potatoes (cull) were manually removed from the conveyor belt before they could be bagged. Later, the dirt and the weeds in the net bags were removed manually before the potatoes were weighed. Due to the rainy weather, the machine harvest took place later in the season than originally planned. Many potatoes were green because the soil had been washed away from the potato ridges. If the same sorting criteria had applied to the machine harvested potatoes as to the potatoes harvested by hand, approximately 75 % of the machine harvested potatoes would have been cull. Therefore, only rotten potatoes were removed before the potatoes were weighed. While these inaccuracies could have caused unequal variances, it is also possible that the sampling area of 1.2 x 0.75 cm was too small for a representative picture.

4.2 Comparison of results to other studies

4.2.1 Ground beetles

The activity of the ground beetles along the edges and in the middle of a strip was significantly affected by the crop type, which confirmed the initial hypothesis. The outcome provides similar results to those reported by Eyre et al. (2012) when comparing ground beetle species activity in relation to crop type in northern England. Apart from the fallow, the lowest activity of ground beetles was found in the potato strip, which was, compared to the other strips, clean cultivated. This is consistent with the results of Eyre et al. (2012), who found the lowest ground beetle activities in vegetables plots mainly consisting of potatoes. The results reported by Altieri et al. (1985), who found significantly higher numbers of ground predators (such as carabid beetles) in pitfalls which were placed in clover plots than in pitfalls which were placed in clean cultivated plots, are consistent with this finding as well. A possible reason for this is the lack of shelter for ground beetles in plots without ground cover. In the potato strip, the micro topography of the rows may have influenced ground beetle activity as well.

It therefore appears that ground cover plays an important role in ground beetle activity. The highest number of ground beetles was found in the oilseed rape strip, which had an uneven ground cover consisting of grass and weeds. This might have been a good food source, while being not too dense to considerably slow down the ground beetles' movements.

In a field experiment studying the movement behaviour of the carabid beetle *Pterostichus melanarius*, Allema et al. (2014) found that significantly more ground beetles moved from rye to oilseed radish than the other way around. The ground beetles were further more reluctant to leave the oilseed radish than the rye. Allema et al. (2014) explained this

preference with two possible mechanisms: the ground beetles were both attracted to oilseed radish and had a greater tendency to stay there. It is possible that those reasons also apply for strip 4, since significantly more ground beetles were captured in the middle of the strip than along the edges. The results of an experiment by Goltermann (1994) showed that in spring and early summer, the epigaeic predator community in oilseed rape fields is dominated by rove beetles and ground beetles. This is the also the main time of the year when the larvae of many OSR pests fall from the flower stands and drop on the ground (Goltermann, 1994), which could explain the ground beetle's attraction to the OSR strip.
Furthermore, when interpreting the results, the weather may need to be taken into account. Ground beetles respond to microhabitat conditions, such as soil moisture (Honek, 1997). Considering the rainfall during the sampling week, it is possible that many ground beetles sought shelter from the rain, which was also provided by the oilseed rape.

Higher activity along the edges

In strip 1 (grass-clover), 3 (fallow) and 5 (potatoes), ground beetles were significantly more active along the edges than in the middles of the strips.
Since the vegetation in strip 1 was very dense, the ground beetles' locomotion may have been slowed down, which lowered the amount of ground beetles that could have fallen into the traps. Along the edges, however, it was easier to move, and this higher activity may have increased the amount of ground beetles in the pitfall traps. Furthermore, the neighbouring wheat strip provided a food source, and the only way to get there was to cross the edge between the grass-clover and the wheat strip. On the other side of the grass clover strip was a grass strip neighboured by a hedge, which provided undisturbed habitat. Indeed, the abundance of ground beetles within a field is often higher close to uncultivated field edges (Sotherton, 1985).
The fallow strip had almost no ground cover, and thus provided neither shelter nor food. The significantly higher numbers of ground beetles among the edges reflect those circumstances; the ground beetles probably stayed close to more attractive strips. The same reasons are likely to apply in the potato strip, where the middle of the strip was almost bare, whereas the edges provided access to shelter and food supply.
These results confirm previously reported results by Altieri et al. (1985), Hummel et al. (2002) and Eyre et al. (2012), where ground beetles preferred plots with dense ground cover over plots with sparse or no ground cover.

No difference in activity between middle and edge

In strips 2 (wheat) and 6 (grass-clover), there was no significant difference between the amount of ground beetles captured along the edges and the amount captured in the middle.
In the wheat strip, the tractor track in the middle was very well established, with almost no ground cover. These conditions are similar to the ones along the edge towards the fallow. In direct comparison, there were significantly more ground beetles along the edge towards the grass-clover than along the fallow edge. The edge towards the grass-clover was possibly

crossed many times due to the reasons already mentioned above, while the edge towards the fallow was likely not as attractive to the ground beetles.
Since strip 6 (grass-clover) had a very dense ground cover, the locomotion of the ground beetles was likely slowed down. The potato strip offered little shelter and food, which may have led the ground beetles to crossing the grass-clover strip towards the other side, where the flower strip was.

No overall difference between middle and edge

The fact that there was no overall significant difference in ground beetle activity between the edges and the middles of the strips can be explained by the large differences in ground beetle numbers captured in the middle and the edge of certain crops; those numbers could have levelled each other out and made the overall mean numbers similar to each other.

Ground beetle activity during night and day

The activity of the ground beetles was further significantly affected by the time of day. In the middles and along the edges of strips 3 (fallow) and 4 (oilseed rape), ground beetle activity was significantly higher during the night than during the day. Since most ground beetle species are night active (Kromp, 1999), it is likely that the night active species waited for darkness before they crossed the fallow to reach the food supply in strip 4.
At the edge between the grass and strip 1 (grass-clover), the activity pattern was reversed: significantly more ground beetles had fallen into the pitfall traps during the day than during the night. One possible explanation is that there were more day active species than night active species along that edge.

4.2.2 Potato root relation to yield

There was an obvious relation between the potato root dry weight and the tuber yield. This result was consistent with last year’s results, when a clear relation between the tuber yield and the root dry weight had been found in the same experiment, albeit with different potato varieties (Yang, 2015). As reported by Iwama (2008), there generally is a positive correlation between the root mass and the final yield of potatoes. A study by Wishart et al. (2012) about the variation in potato roots found a weak, but positive correlation between root weight and final tuber yield.
Between the different potato varieties in the mixed plots, there were no statistically significant differences in root dry weight. This stands in contrast to the findings reported by Puértolas et al. (2014), who concluded that there are large differences in root dry weight and root length between different potato varieties, especially within the first 30 cm of the soil layer. Further, there was no difference in root dry weight between the Toluca plants growing in the no-mix and the Toluca plants growing in the mixed plots. Iwama (2008) suggested taking as many samples as possible of one potato plant; therefore, the number of samples taken in this experiment might not be high enough to get a clear picture of the different varieties’ root dry weight.

There was, however, one statistically significant difference between treatments: at 30-45 cm depth, the mean dry root weight of the potato plants in the mixed plots was significantly higher than the mean dry root weight of Toluca in the no-mix plots. At this depth, only roots, but no tubers could be found. Since roots are the only part of a crop that can take up water and nutrients from the soil (Ahmadi et al., 2014), plants with a higher rooting density have broader access to water and nutrients, which makes them more resistant to drought. The higher root density in this depth in the mixed plots could have resulted from competition between the different varieties. Considering that plants acquire nutrients from the soil via their roots, the mineral Nitrogen content of that soil layer would be expected to be lower than in the no-mix plots. This, however, was not the case. Both the NO_3 and the NH_4 content at 30-45 cm depth were significantly higher in the mixed plots than in the no-mix plots. Furthermore, there was no relation between the mineral Nitrogen content and the root dry weight.

Comparing the overall mineral Nitrogen content between treatments, both the NO_3 and the NH_4 content were significantly higher in the mixed plots, even though there was no overall difference in root dry weight between treatments. This may be due to the preceding crops: in the previous year, wheat had been growing in the potato strip, which was intercropped with faba beans in the mixed plots. Because of the faba beans' Nitrogen fixation, large quantities of Nitrogen were contained in the plant parts left behind on the field after harvest (e.g. in the roots, leaves or stems) (Butz et al., 2015; Wichmann et al., 2006). During the decomposition of the organic matter, NO_3 and NH_4 were released (Butz et al., 2015). The catch crop growing there over the winter prevented leaching (Butz et al., 2015), so the mineral Nitrogen was available for the potato mix in the summer. Even though there were more roots in 30-45 cm depth in the mixed plots, the amount of Nitrogen might have been so high that the roots did not take up enough to make the difference in mineral Nitrogen between treatments statistically non-significant.

4.2.3 Crop yields

The initial hypothesis that yields would be different in the edge rows has only partially been confirmed, as well as the hypothesis that there would be differences in yield between treatments. Reactions to row position or treatment were dependent on the crop type; while there were strong edge effects throughout the whole potato strip, there were none in the oilseed rape strip. In the wheat and the grass-clover strip, the occurrence of edge effects depended on treatment.

Grass-clover

The best model to predict the yield contained the additional effect of treatment and the row coordinates as predictors, and there was evidence that both were highly significant predictors for the dry weight. The row coordinates in the model were continuous variables, taking account of the fact that adjoining rows were more closely related than rows that were further apart. It becomes obvious in Figure 15 that this was an important aspect, especially for the mix. A t-test for the N content in the different treatments showed no significant differences, however. This may be due to the GLM being the best in relation to any other options, but not the best in an absolute sense. A linear regression model that set the predicted values in relation to the observed values only resulted in an R^2 value of 0.49, indicating that the GLM explains less than half of the variance. Therefore, other factors not measured may have influenced the grass-clover yield.

The significant difference in dry weight between treatments is somewhat astonishing, as both treatments consisted of ryegrass and clover, albeit different varieties. In an experiment in the UK, Stopes et al. (1995) compared the dry matter of legumes grown at an organic farm over different periods of time and found that white clover yielded a slightly higher dry matter than red clover. However, white clover was part of the mixed plots, while red clover was part of the no-mix plots. Therefore, it is unlikely that the difference in dry weight between treatments is caused by their composition.

In a field trial over the course of three years, Huguenin-Elie et al. (2009) found that grass-clover mixtures with a clover share of 60-70% had the highest yields. However, big advantages in yield could already be achieved with a relatively small number of different grass and clover varieties. Amongst others, the exploitation of different root horizons as well as complementing growth rhythms during the vegetative period were factors which allowed a more efficient use of resources (Huguenin-Elie et al., 2009). While English ryegrass and clover root up to 1 metre deep, Italian ryegrass can root even deeper (Lenuweit and Gharadjedaghi, 2002). In addition to that, it has numerous rootlets, which allow a quick uptake of nutrients (Lenuweit and Gharadjedaghi, 2002). Therefore, it is possible that the resource utilization in the no-mix was more efficient than in the mix, where the share of Italian ryegrass was much lower.

In a field trial, Boller and Nösberger (1987) found that the percentage of Nitrogen derived from the rhizobium-legume symbiosis was significantly higher for red than for white clover. Since white clover has a more shallow root system, it is better able to compete for soil mineral Nitrogen and thus to derive more of its Nitrogen from non-symbiotic sources (Boller and Nösberger, 1987). Those findings could offer another possible explanation for the difference in dry weight between treatments: the ryegrass in the no-mix was provided with more Nitrogen and had at the same time less competition for mineral Nitrogen than the ryegrass in the mix.

Since there was no discernible border between the grasses growing next to strip 1, the first row of the strip was partially overgrown. This factor could explain the highest yields in that edge row, and the consecutive decline in rows 2 and 3. While this is clearly visible in the mixed plots, there is no such pattern in the no-mix plots. However, the grass-clover strip was patchy and some random sampling locations were in spots with almost no ground cover. The outcome was possibly influenced by those patches.

In both treatments, the lowest yields were obtained in row 6, the edge row bordering on the wheat strip. Here, a tractor track along the edge was clearly visible, unlike on the other side of the grass-clover strip. The furrow repressed the growth of grass here, which had a negative effect on the sample weights in the edge row. Even though wheat roots can be about 200 cm deep (Canadell, 1996), 70% of the root length can be found in the top 30 cm

(Sylvester-Bradley et al., 2015). Therefore, in addition to the compaction caused by the tractor track, competition between wheat and grass-clover along the edge may be another cause for that difference.

Wheat

Both GLMs to predict the performance of the wheat strip contained the interaction effect between treatment and row coordinates as a predictor, and in both GLMs there was evidence that those predictors were significant for the outcome variable. Especially in the mix, it was obvious that the rows next to each other had more in common than the rows which were further apart (see Figures 16 and 17), which explains why the row coordinates were part of the model, but the rows as factors were not. However, there was a large difference in the goodness of fit of those GLMs. Therefore, other factors which had not been measured may have had influence on wheat performance as well.

While there was no significant difference in wheat dry weight, the number of kernels per m² and the total biomass between treatments, the harvest index of the wheat was significantly lower in the mix. In a field experiment about intercropping wheat with white clover, Thorsted et al. (2006) found that the harvest index of wheat was reduced by intercropping. However, the harvest index increased with increasing sowing width (Thorsted et al., 2006). In a field trial, Burke et al. (2013) found that intercropping wheat and peas led to significantly lower grain yields than in mono-cropped treatments. Furthermore, the harvest index was reduced; however, this reduction was not statistically significant, which led Burke et al. (2013) to the conclusion that the peas did not affect the resource allocation in the wheat. It is therefore possible that in the mix, both the row spacing and the presence of the faba beans led to a lower harvest index of the wheat.

In an experiment by Xiao et al. (2004), root contact between wheat and faba beans enhanced the N acquisition of wheat. This confirms the results of the t-tests showing that the N content of both stems and grains was significantly higher in the mix than in the no-mix. Furthermore, Bedoussac et al. (2015) concluded that the protein concentration of wheat intercropped with legumes is almost always greater than that of wheat as a sole crop. Since the weight of grain protein is directly related to the weight of Nitrogen (Sylvester-Bradley et al., 2015), this conclusion is consistent with the results of this experiment.

In the mixed plots, there was a significant difference in dry weight and number of kernels between the edge rows. One possible explanation is competition with the grass-clover along the edge of row 1. Thorsted et al. (2006) reported a decrease in wheat grain yield when intercropped with white clover, compared to wheat as a sole crop. Amongst others, Thorsted et al. (2006) assumed that the yield reduction was caused by competition for Nitrogen during vegetative growth and for soil water during grain filling. This assumption is also possible to hold here. In the mixed plots, the wheat dry weight in row 1 was significantly lower than on the other side, where the wheat was bordering on the fallow strip. Similarly, the kernel weight in row 4 was significantly higher than in row 1, where wheat was bordering on the grass-clover strip. Another possible reason for this difference between the edge rows is that strip 3 had been manured, but the maize that had been sown did not grow there in the end. Therefore, more nutrients may have been available for the wheat row bordering on the fallow strip than for the wheat bordering on the grass-clover strip.

Compared to the wheat yields on other organic farms, the wheat yield from strip 2 was higher than the average wheat yields in organic agriculture. In Eastern Germany, the average yield from 2000 ha of organic wheat was 2.25 t/ha in 2013 (Böhm, 2013). A conversion of the kernel weight (kg/m²) from strip 2 results in yields of 2.5 t/ha for the mix and 3.0 t/ha for the no-mix. Those above-average yields indicate that the diverse cropping system as a whole worked well for the wheat. In the previous year, mustard had been growing in the wheat strip and was followed by a catch crop over the winter. Mustard is often used as green manure and has been shown to be a beneficial rotational crop prior to winter wheat (Guy and Gareau, 1998).

Oilseed rape

Even though in absolute numbers, there were differences in dry weight and the number of seeds between rows and treatments, those were not statistically significant. However, none of the GLMs proved to be a good fit, which points to the fact that there must have been factors not measured that played an important part in OSR performance. It is also possible that the power of the statistical tests was not sufficient to detect those differences.

Since oilseed rape roots go almost two metres deep (Barraclough, 1989), they are likely to have encountered roughly the same resources within all rows, since there were no crops growing in strip 3 and the potatoes in strip 5 rooted at more shallow depths.

The thousand seed weight in the mixed plots, where oilseed rape was intercropped with clover, was slightly higher. The output of the GLM showed that both the treatments and the coordinates of the plots were highly significant predictors for the thousand seed weight. The seed weight of oilseed rape has been shown to be related to the Nitrogen supply of the plants (Kullmann et at., 2010; Lunn et al., 2003). In a greenhouse experiment, Kullmann et al. (2010) found that an increase in Nitrogen supply increased the seed weight of oilseed rape. In a field trial, Lunn et al. (2003) associated a greater thousand seed weight and fraction of seeds greater than 2 mm with greater Nitrogen fertilisation.

Moreover, Lunn et al. (2003) found a relationship between plant population, seed size distribution and the thousand seed weight of oilseed rape. This may be another possible explanation for the difference in thousand seed weight between treatments. The whole oilseed rape strip was very patchy due to birds picking the seeds after sowing. Consequently, some samples were taken in areas with a very sparse plant population, and it is possible that most of the affected sampling areas were in the no-mix plots, while the sampling areas with a more dense plant population were in the mixed plots. Lunn et al. (2003) concluded that an increase in plant population leads to an increase in thousand seed weight, seed maturity and the fraction of seeds greater than 2 mm.

A conversion of the seed weight (kg/m²) to the oilseed rape yield in t/ha shows that with 0.58 t/ha in the mix and 0.62 t/ha in the no-mix, the yields stayed far behind average yields on organic farms. In Eastern Germany, the average yield on 370 ha of organic oilseed rape was 1.3 t/ha in 2013 (Böhm, 2013). Reasons for the low yield of strip 4 could be seed-picking birds; some of the random sampling locations had considerably less plants in them than others. Since the Unifarm did not have equipment for threshing oilseed rape, and thus the same equipment as for the wheat had been used, it is also possible that the seed losses during handling and measuring were higher than they usually are. Furthermore, the oilseed rape variety which has been growing in strip 4 was not optimized for high yields; it had originally been developed as green manure.

Potatoes

For the hand harvest of the whole strip, both GLMs (for fresh weight and yield) contained treatment as a predictor. This indicates that treatment played an important role. However, there were no statistically significant differences between treatments, except for one: the amount of potatoes less than 4 cm in diameter was significantly higher in the mix; in fact, it was almost twice as high. During various field studies, Westermann and Kleinkopf (1985) found that Nitrogen fertilization generally decreased the amount of undersized tubers, while increasing the marketable yield. Considering that the mineral Nitrogen content in the mixed plots was significantly higher than in the no-mix, their finding stands in contrast to the results of this experiment. A possible explanation may therefore be a difference in tuber size between the varieties. In the course of a four year tuber bulking rate study, Schliephake and Trautz (2014) found that Ditta had a high amount of undersized tubers in comparison to other potato varieties used in organic farming. Also Annabelle and Tiamo had a higher amount of undersized tubers than Toluca (Figure 26).

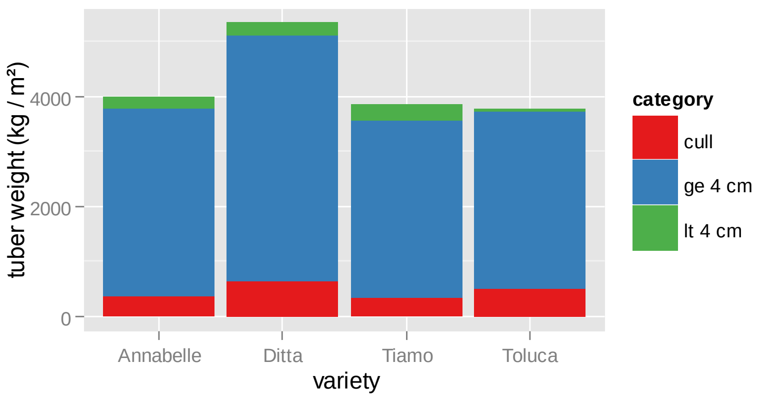


Figure 25: tuber weight of different potato varieties as sorted into three different categories: cull, greater/equal 4 cm (ge 4 cm), less than 4 cm (lt 4 cm)

This difference in potato sizes may also be the reason why the no-mix plots gave a significantly higher yield than the mixed plots during the machine harvest. Since the machine only harvested potatoes that were bigger than the slots in the conveyor belt, it is possible that the loss rate of small potatoes in the mix was much higher than in the no-mix. This could have ultimately caused the statistically significant difference in fresh weight between treatments.

For the mix, both GLMs (with fresh weight and yield as outcome variables) contained variety as a predictor, indicating that this was an important factor for all measurements. However, none of the GLMs fit the data particularly well. Therefore, it is possible that other factors which had not been measured were important predictors for the performance of the different varieties in the mixed plots.

The fresh weight and the yield of Ditta were significantly higher than those of all other varieties. Among the potato varieties used in organic agriculture, Ditta is known to have an average to high yield, while Annabelle is known to have a below-average yield (Schliephake and Trautz, 2014). Even though Tiamo is also known to be a high to very high yielding potato variety (Semagri.nl, 2015), its yield was far behind that of Ditta. Since potato tuber yields depend on local conditions, such as the soil properties, the nutrient supply or the preceding crop on the field (Berner et al., 2010), it is not possible to make a general statement about those yields.

Comparing Toluca in the mix with Toluca in the no-mix, there were no statistically significant differences. The fresh weight in the no-mix was higher than in the mix, but the specific gravity was the same for both treatments, which indicates that the potato quality was even.

For all three harvests, the potato yield in row 1 was significantly higher than the yields in the other rows. Row position had a significant effect here, and it was the position at the edge towards the oilseed rape that gave the highest yields, which partially confirms the initial hypothesis that the yield would be different in the edge rows. A possible explanation is the difference in rooting depth between potatoes and oilseed rape. In a field trial by Barraclough (1989), oilseed rape plants reached a rooting depth of at least 180 cm. Potatoes, on the other hand, reach a rooting depth of about 100 cm, depending on the water supply (Stalham and Allen, 2001). This difference in rooting depth might have allowed the potato plants in the edge row access to more resources than the potato varieties in the other rows, which were surrounded by plants of the same rooting depth. Oilseed rape is further known to improve the soil structure (Williams, 2010); a trait the neighbouring potatoes could have benefitted from as well.

In the edge row next to the grass-clover strip, potato yields were not significantly different from the two middle rows of the strip. Since grass-clover mixtures root in similar depths as potato plants (Høgh-Jensen and Schjoerring, 2001), there would have been no advantage in resource utilisation for the edge row, like there was along the edge towards the oilseed rape. This stands in contrast to the findings reported by Bouws and Finck (2008), who obtained the highest tuber yields in potato plots that were neighboured by grass-clover. However, Bouws and Finck (2008) made that comparison on a plot basis, comparing the tuber yields of potato plots neighboured by grass-clover to those neighboured by cereals.

While on average, organic farms in the Netherlands obtain yields up to 30 t/ha (Tamm et al., 2004), a conversion from kg/m² to t/ha of the yields in strip 5 resulted in yields around 40 t/ha. In general, the variation in potato yields is much higher than in cereal yields, which is why the site selection is very important (Kolbe et al., 2012). Influences on growth, development and potato quality can be traced back to differences between sites, such as differences between type of soil, weather conditions, nutrient supply or pest infestation

(Kolbe et al., 2012). In Germany, potato yields on organic farms in the year 2000 varied between 5 t/ha and 35 t/ha, which reflects the large variability in yields between different farms (Tamm et al., 2004). Therefore, it is likely that the circumstances in strip 5 were very favourable for potato cultivation.

Diversity – mix compared to no-mix

For the grass-clover and the wheat, there were clear edge effects in the mix, whereas there were none for the no-mix. In the grass-clover strip, the yield of the mixed plots was highest in the row bordering on the grass, and then gradually declined towards the row neighboured by the wheat. In the wheat strip, the yield in the mixed plots was lowest in the row next to the grass-clover strip, and then gradually increased until the row bordering on the fallow. In both strips, the yields of the no-mix were considerably less affected by possible edge effects (Figure 26).

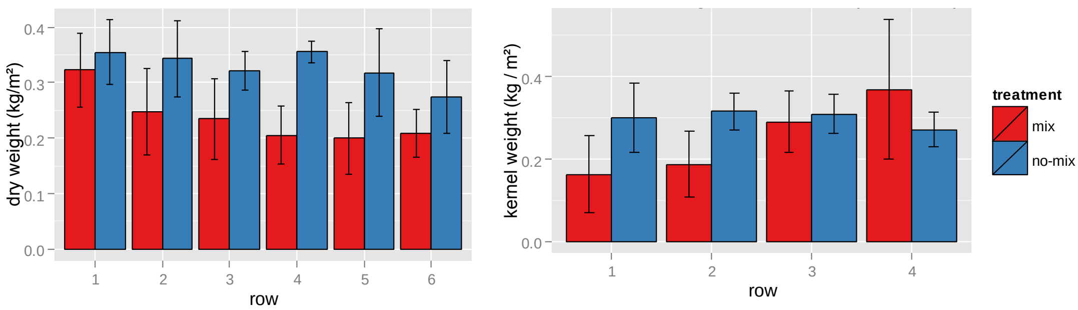


Figure 26: yields of grass-clover (left) and wheat (right) and the response of the different treatments to edge effects

However, a similar pattern was not visible in the oilseed rape strip and neither in the potato strip. It therefore seems that the combination of plants in the mix of the grass-clover and the wheat strip was not as productive as the no-mix. Considering that the difference in grass-clover yield between treatments was relatively small in the first row, it can be assumed that the reason for this similarity in yields lies in the overgrowth of the adjoining grass strip. In the no-mix, there would not have been much vegetation to overgrow, whereas in the less productive mix, the overgrowth could have had a stronger impact. For the wheat, the same assumption could be true: considering that the difference in yield in the fourth row was relatively small, it is possible that the mixed yields were positively influenced by the fertilized fallow, while the other rows in the mix did not have this benefit. The no-mix would not have been impacted as much by the fertilized fallow, because it was performing well already.

However, it cannot be ruled out that this pattern was created by chance: the grass-clover strip was very patchy, and the random sampling locations inevitably included patches with very little vegetation. In the wheat strip, the amount of weeds in the mix was much higher than in the no-mix. Furthermore, the wheat plants along the tractor tracks were only sporadically fully grown. Those factors could have contributed to the distinct pattern in dry weight and kernel weight.

CHAPTER V: Conclusion and Recommendations

It can be concluded that the extent of edge effects on different scales depends on both the mixture and the formation of the crops. Therefore, edge effects can play a considerable role in diverse agro-ecosystems.

If ecosystem services are required from ground beetles, such as pest control, the composition of the strips needs to be taken into account. While the oilseed rape strip attracted more ground beetles than any other strip, the activity of the ground beetles in the grass-clover strips was unexpectedly low. Since this may have been due to lower motility or the weather during the sampling week, conducting the pitfall trap experiment over a longer period of time could answer the question if the preferences expressed by the ground beetles were indeed only a result of the weather conditions. The influence of highly attractive strips on ground beetles could also be counterproductive for natural pest control in a diverse strip cropping system, as it affects the dispersion of the beetles over the field. It would therefore be beneficial to find out for which reasons ground beetles prefer one strip over another. Ground beetle activity depended not only on the crops, but also on the time of day, indicating that most ground beetles which had fallen into the pitfalls were night-active. Considering the high activities along the edges, especially between grass-clover and wheat as well as between oilseed rape and the fallow strip, a follow-up of the experiment with directional pitfalls could shed more light on ground beetle locomotion.

For the root dry weight, it can be concluded that a more diverse composition of potato varieties results in a higher root weight in 30-45 cm depth. A higher root density makes plants more resistant, increasing the amount of water and nutrients that can be taken up by the plants. Growing different potato varieties together in one plot may therefore be an option to consider when growing potatoes in areas that are affected by drought due to climate change. The combination of potato varieties also greatly affects the yield; out of the four varieties in this experiment, Ditta had the highest yield by far, even though other varieties were supposed to be high-yielding as well (e.g. Tiamo). However, there may have been other influences on yield (such as the planting density), so the interactions between different potato varieties in a mixed plot may be a subject for further research.

It was shown that edge effects can have both positive and negative effects on crop yields, which resulted in yield differences between rows and treatments. The most predominant edge effects were the positive effect of the oilseed rape strip on the edge row of the potatoes and the negative effect of the grass-clover strip on the edge row of the wheat and vice versa, especially in the mixed plots. It may be interesting to explore those edge effects further to see exactly how beneficial or detrimental they really are. Nonetheless, it can be concluded that the choice of the neighbouring crops in a diverse strip cropping system has an effect on crop yields.

REFERENCES

Agriculture.vic.gov.au, (2015). *Potatoes: Measurement of specific gravity | Potatoes | Vegetables | Horticulture | Agriculture | Agriculture*. [online] Available at: http://agriculture.vic.gov.au/agriculture/horticulture/vegetables/potatoes/potatoes-measurement-of-specific-gravity [Accessed 23 Nov. 2015].

Ahmadi, S., Sepaskhah, A., Andersen, M., Plauborg, F., Jensen, C. and Hansen, S. (2014). Modeling root length density of field grown potatoes under different irrigation strategies and soil textures using artificial neural networks. *Field Crops Research*, 162, pp. 99-107.

Allema, B. (2014). *Quantifying and simulating movement of the predator carabid beetle Pterostichus melanarius in arable land*. PhD thesis report. Wageningen University, Wageningen.

Allema, B., van der Werf, W., van Lenteren, J., Hemerik, L. and Rossing, W. (2014). Movement Behaviour of the Carabid Beetle Pterostichus melanarius in Crops and at a Habitat Interface Explains Patterns of Population Redistribution in the Field. *PLoS ONE*, 9(12), e115751.

Altieri, M., Nicholls, C., Henao, A. and Lana, M. (2015). Agroecology and the design of climate change-resilient farming systems. *Agronomy for Sustainable Development*, 35(3), pp. 869-890.

Altieri, M., Wilson, R. and Schmidt, L. (1985). The effects of living mulches and weed cover on the dynamics of foliage- and soil-arthropod communities in three crop systems. *Crop Protection*, 4(2), pp. 201-213.

Ats.ucla.edu, (2015). *SAS Data Analysis Examples: Zero-inflated Poisson Regression*. [online] Available at: http://www.ats.ucla.edu/stat/sas/dae/zipreg.htm [Accessed 30 Oct. 2015].

Barbosa A.M., Brown J.A. & Real R. (2014) modEvA - an R package for model evaluation and analysis. R package, version 0.1.

Barraclough, P. (1989). Root growth, macro-nutrient uptake dynamics and soil fertility requirements of a high-yielding winter oilseed rape crop. *Plant Soil*, 119(1), pp. 59-70.

Bedoussac, L., Journet, E., Hauggaard-Nielsen, H., Naudin, C., Corre-Hellou, G., Jensen, E., Prieur, L. and Justes, E. (2015). Ecological principles underlying the increase of productivity achieved by cereal-grain legume intercrops in organic farming. A review. *Agronomy for Sustainable Development*, 35(3), pp. 911-935.

Berner, A., Böhm, H., Buchecker, K., Dierauer, H., Dresow, J., Dreyer, W., Finckh, M., Fuchs, A., Keil, S., Keiser, A., Kühne, S., Landzettel, C., Mahnke-Plesker, S., Six, R., Speiser, B., Tamm, L., Völkel, G. (2010). *Biokartoffeln - Qualität mit jedem Anbauschritt*. [pdf] Bioland Beratung GmbH, Kompetenzzentrum Ökolandbau Niedersachsen, Forschungsinstitut für biologischen Landbau (FibL). Available at:

https://www.fibl.org/fileadmin/documents/shop/1404-biokartoffel.pdf [Accessed 10 Nov. 2015].

Berry, P., Cook, S., Ellis, S., Gladders, P. and Roques, S. (2012). *HGCA Oilseed Rape Guide*. Warwickshire: HGCA, p.8. Available at: http://cereals.ahdb.org.uk/media/305093/g55-oilseed-rape-guide-jan-2014-update.pdf [Accessed 17 Nov. 2015].

Böhm, U. (2013). *Abschluss der Getreideernte im Land Brandenburg*. [online] Lbv-brandenburg.de. Available at: http://www.lbv-brandenburg.de/index.php?option=com_content&view=article&id=1416:abschlus-der-getreideernte-im-land-brandenburg&catid=87:aktuelles&Itemid=63 [Accessed 4 Dec. 2015].

Boller, B. and Nösberger, J. (1987). Symbiotically fixed nitrogen from field- grown white and red clover mixed with ryegrasses at low levels of15N-fertilization. *Plant Soil*, 104(2), pp. 219-226.

Bouws, H. and Finckh, M. (2008). Effects of strip-intercropping of potatoes with non-hosts on late blight severity and tuber yield in organic production. *Plant Pathology* (2008) 57, pp. 916-927.

Brooker, R.W., Maestre, F.T., Callaway, R.M., Lortie, C.L., Cavieres, L.A., Kunstler, G., Liancourt, P., Tielbörger, K., Travis, J.M., Anthelme, F. (2008). Facilitation in plant communities: the past, the present, and the future. *J. Ecol.* (96), pp. 18-34.

Burke, I., Tautges, N. and Pittmann, D. (2013). *Interrow cultivation and intercropping for organic transition in dryland crop production systems*. [pdf] Center for Sustaining Agriculture and Natural Resources (CSANR), Washington State University. Available at: http://csanr.wsu.edu/wp-content/uploads/2013/06/120Burke.pdf [Accessed 13 Dec. 2015].

Butz, A., Jung, F., Mastel, K., Michelsburg, S., Raupp, J. and Schaufelberger, R. (2015). *Ackerbohne (Vicia faba L.) - Hinweise zum Pflanzenbau*. [pdf] Karlsruhe, Germany: Landwirtschaftliches Technologiezentrum Augustenberg (LTZ). Available at: http://www.ltz-bw.de/pb/site/pbs-bw-new/get/documents/MLR.LEL/PB5Documents/ltz_ka/Service/Schriftenreihen/Hinwe ise%20zum%20Pflanzenbau/Ackerbohne.pdf [Accessed 23 Nov. 2015].

Canadell, J., Jackson, R., Ehleringer, J., Mooney, H., Sala, O. and Schulze, E. (1996). Maximum rooting depth of vegetation types at the global scale. *Oecologia*, 108(4), pp. 583-595.

Cardinale, B., Wright, J., Cadotte, M., Carroll, I., Hector, A., Srivastava, D., Loreau, M. and Weis, J. (2007). Impacts of plant diversity on biomass production increase through time because of species complementarity. *Proceedings of the National Academy of Sciences*, 104(46), pp. 18123-18128.

Casper, B. and Jackson, R. (1997). Plant competition underground. *Annu. Rev. Ecol. Syst.*, 28(1), pp. 545-570.

Extranet.agrico.nl, (2015). *Potato varieties - Agrico*. [online] Available at: http://extranet.agrico.nl/index.php?a=82&rassenID=160 [Accessed 9 Nov. 2015].

Eyre, M., Luff, M., Atlihan, R. and Leifert, C. (2012). Ground beetle species (Carabidae, Coleoptera) activity and richness in relation to crop type, fertility management and crop protection in a farm management comparison trial. *Ann Appl Biol*, 161(2), pp. 169-179.

Fox, G., Negrete-Yankelevich, S. and Sosa, V. (2015). *Ecological Statistics: Contemporary theory and application*. Oxford University Press.

Fletcher, R. J., Ries, L., Battin, J. and Chalfoun A. D. (2007). The role of habitat area and edge in fragmented landscapes: definitively distinct or inevitably intertwined? *Canadian Journal of Zoology*, 85, pp. 1017-1030.

Francis, C., Jones, A., Crookston, K., Wittler, K. and Goodman, S. (1986). Strip cropping corn and grain legumes: A review. *American Journal of Alternative Agriculture,* 1, pp. 159-164.

Glowacka (2014). The influence of strip cropping and adjacent plant species on the content and uptake of N, P, K, Mg and Ca by maize (*Zea Mays* L.).

Goltermann, S. (1994). Das Auftreten von Laufkäfern (Col., Carabidae) auf Winterrapsfeldern und deren Einfluss auf den Massenwechsel von Meligethes aeneus F. (Col. Nitidulidae). PhD thesis report. University of Rostock, Germany.

Guisan, A. and Zimmermann, N. (2000). Predictive habitat distribution models in ecology. *Ecological Modelling*, 135(2-3), pp. 147-186.

Guy, S. and Gareau, R. (1998). Crop Rotation, Residue Durability, and Nitrogen Fertilizer Effects on Winter Wheat Production. *jpa*, 11(4), p. 457.

Houba, V. and Novozamsky, I. (1998). Influence of storage time and temperature of air-dried soils on pH and extractable nutrients using 0.01 mol/L CaCl 2. *Fresenius' Journal of Analytical Chemistry*, 360(3-4), pp. 362-365.

Høgh-Jensen, H. and Schjoerring, J. (2001). Rhizodeposition of nitrogen by red clover, white clover and ryegrass leys. *Soil Biology and Biochemistry*, 33(4-5), pp. 439-448.

Hummel, R., Walgenbach, J., Hoyt, G. and Kennedy, G. (2002). Effects of vegetable production system on epigeal arthropod populations. *Agriculture, Ecosystems & Environment*, 93(1-3), pp. 177-188.

Honek, A. (1997). The Effect of Plant Cover and Weather on the Activity Density of Ground Surface Arthropods in a Fallow Field. *Biological Agriculture & Horticulture*, 15(1-4), pp. 203-210.

Iwama, K. (2008). Physiology of the potato: New insights into root system and repercussions for crop management. *Potato Research*, 51, pp. 333–353.

Kang, Y. and Harring, J. (2012). *Investigating the Impact of Non-Normality, Effect Size, and Sample Size on Two-Group Comparison Procedures: An Empirical Study*. [online] Available at: http://www.education.umd.edu/EDMS/fac/Harring/MIsc/Kang&H-2012.pdf [Accessed 30 Oct. 2015].

Kolbe, H., Karalus, W., Schuster, M., Hänsel, M., Schaerf, A. and Pölitz, B. (2012). *Kartoffeln im Ökolandbau - Informationen für Praxis und Beratung*. [pdf] Sächsisches Landesamt für Umwelt, Landwirtschaft und Geologie. Available at: http://orgprints.org/15102/16/OekoKartoffeln_Broschuere12.pdf [Accessed 4 Dec. 2015].

Kremen, C., Iles, A. and Bacon, C. (2012). Diversified Farming Systems: An Agroecological, Systems-based Alternative to Modern Industrial Agriculture. *Ecology and Society*, 17(4).

Kremen, C. and Miles, A. (2012). Ecosystem services in biologically diversified versus conventional farming systems: benefits, externalities, and trade-offs. *Ecology and Society* 17(4).

Kromp, B. (1999). Carabid beetles in sustainable agriculture: a review on pest control efficacy, cultivation impacts and enhancement. *Agriculture, Ecosystems and Environment* 74 (1999). pp 187-228.

Kullmann, A., Ogunlela, V. and Geisler, G. (1990). Seed characters in oilseed rape (Brassica napus) in relation to nitrogen nutrition. In: M. L. van Beusichem (Ed.) *Plant nutrition - physiology and applications*. Dordrecht: Kluwer Academic Publishers, pp. 569-575.

Lenuweit, U. and Gharadjedaghi, B. (2002). *Biologische Basisdaten zu Lolium perenne, Lolium multiflorum, Festuca pratensis und Trifolium repens*. [pdf] Berlin: Umweltbundesamt. Available at: http://www.gfn-umwelt.de/publikation/2103.pdf [Accessed 7 Dec. 2015].

Li, L., Li, S., Sun, J., Zhou, L., Bao, X., Zhang, H. and Zhang, F. (2007). Diversity enhances agricultural productivity via rhizosphere phosphorus facilitation on phosphorus-deficient soils. *Proceedings of the National Academy of Sciences*, 104(27), pp. 11192-11196.

Lin, B. (2011). Resilience in Agriculture through Crop Diversification: Adaptive Management for Environmental Change. *BioScience*, 61(3), pp. 183-193.

Lindenmayer, D., Hobbs, R., Montague-Drake, R., Alexandra, J., Bennett, B., Burgman, M., Cale, P., Calhoun, A., Cramer, V., Cullen, P., Driscol, D., Fahrig, L., Fischer, J., Franklin, J., Haila, Y., Hunter, M., Gibbons, P., Lake, S., Luck, G., MacGregor, C., McIntyre, S., Mac Nally, R., Manning, A., Miller, J., Mooney, H., Noss, R., Possingham, H., Saunders, D., Schmiegelow, F., Scott, M., Simberloff, D., Sisk, T., Tabor, G., Walker, B., Wiens, J., Woinarski, J., Zavaleta, E. (2008). A checklist for ecological management of landscapes for conservation. *Ecol. Lett. (11), pp. 78–91.*

Lithourgidis, A.S., Dordas, C.A., Damalas, C.A., Vlachostergios, D.N., 2011. Annual intercrops: an alternative pathway for sustainable agriculture. *Aust. J. Crop Sci.(5), pp. 396-410.*

Lunn, G., Bullard, M., Holme, S. and Blunt, R. (2003). *Improvement of oilseed rape establishment by seed selection or seed treatment*. Project Report No. 313. [pdf] HGCA. Available at: http://cereals.ahdb.org.uk/media/341541/pr313-final-project-report.pdf [Accessed 20 Nov. 2015].

McElduff, F., Cortina-Borja, M., Chan, S. and Wade, A. (2010). When t-tests or Wilcoxon-Mann-Whitney tests won't do. *AJP: Advances in Physiology Education*, 34(3), pp. 128-133.

Mollison, B. (1988). *Permaculture. A designers' manual*. Washington, D.C.: Island Press, pp. 26+29.

Mundt, C. (2002). Use of multiline cultivars and cultivar mixtures for disease management. *Annual review of phytopathology*, 40(1), pp. 381-410.

Murcia, C. (1995). *Edge effects on fragmented forests: implications for conservation. Trends Ecol. Evol., 10, 58–62.*

Novozamsky, J., Houba, V.J.G. , van Eck, R., van Vark., W. (1983). A novel digestion technique for multielement plant analysis. Communications in Soil Science and Plant Analysis (14), pp. 239-248.

Park, S., Hwang, C. and Vlek, P. (2005). Comparison of adaptive techniques to predict crop yield response under varying soil and land management conditions. *Agricultural Systems*, 85(1), pp. 59-81.

Puértolas, J., Ballester, C., Elphinstone, E. and Dodd, I. (2014). Two potato (Solanum tuberosum) varieties differ in drought tolerance due to differences in root growth at depth. *Functional Plant Biology*, 41(11), p. 1107.

Rämert, B., Lennartsson, M., Davies, G. (2002). The use of mixed species cropping to manage pests and diseases – theory and practice. [online] Orgprints.org. Available at: http://orgprints.org/8289 [Accessed 22 Oct. 2015].

Ries, L., Fletcher, R.J., Battin, J. & Sisk, T.D. (2004). *Ecological responses to habitat edges: mechanisms, models, and variability explained. Annu. Rev. Ecol. Syst., 35, 491–522.*

Sala, O.E. (2001). Price put on biodiversity. *Nature*, [online] 412, pp. 34-36. Available at: http://www.nature.com/nature/journal/v412/n6842/pdf/412034a0.pdf [Accessed 7 Jun. 2015].

Semagri.nl, (2015). *Semagri | Tiamo*. [online] Available at: http://www.semagri.nl/semagri/en/Fresh-potatoes/tiamo.html [Accessed 9 Nov. 2015].

Sheehan, W. (1986). Response by Specialist and Generalist Natural Enemies to Agroecosystem Diversification: A Selective Review. *Environmental Entomology*, 15(3), pp. 456-461.

Smith, R., Gross, K. and Robertson, G. (2008). Effects of Crop Diversity on Agroecosystem Function: Crop Yield Response. *Ecosystems*, 11(3), pp. 355-366.

Song, Y., Zhang, F., Marschner, P., Fan, F., Gao, H., Bao, X., Sun, J. and Li, L. (2006). Effect of intercropping on crop yield and chemical and microbiological properties in rhizosphere of wheat (Triticum aestivum L.), maize (Zea mays L.), and faba bean (Vicia faba L.). *Biol Fertil Soils*, 43(5), pp. 565-574.

Sotherton, N. W. (1985). The distribution and abundance of predatory arthropods in field boundaries. *Ann. Appl. Biol.* 106, pp. 17-21.

Spector, P. (2014). *Using t-tests in R | Department of Statistics*. [online] Statistics.berkeley.edu. Available at: http://statistics.berkeley.edu/computing/r-t-tests [Accessed 30 Oct. 2015].

Stopes, C., Millington, S. and Woodward, L. (1996). Dry matter and nitrogen accumulation by three leguminous green manure species and the yield of a following wheat crop in an organic production system. *Agriculture, Ecosystems & Environment*, 57(2-3), pp. 189-196.

Sylvester-Bradley, R., Berry, P., Blake, J., Kindred, D., Spink, J., Bingham, I., McVittie, J. and Foulkes, J. (2015). *Wheat Growth Guide*. Warwickshire: AHDB. Available at:

http://cereals.ahdb.org.uk/media/185687/g66-wheat-growth-guide.pdf [Accessed 17 Nov. 2015].

Tamm, L., Smit, A., Hospers, M., Janssens, S., Buurma, J., Molgaard, J., Laerke, P., Hansen, H., Hermans, A., Bodker, L., Bertrand, C., Lambion, J., Finckh, M., Schüler, C., Lammerts van Bueren, E., Ruissen, T., Nielsen, B., Solberg, S., Speiser, B., Wolfe, M., Phillips, S., Wilcoxon, S. and Leifert, C. (2004). *Assessment of the Socio-Economic Impact of Late Blight and State-of-the-Art Management in European Organic Potato Production Systems*. [pdf] Research Institute of Organic Agriculture, Frick, Switzerland. Available at: https://www.fibl.org/fileadmin/documents/shop/1340-late-blight.pdf [Accessed 7 Dec. 2015].

Tittonell, P. (2013). Farming systems ecology: towards ecological intensification of world agriculture. Inaugural lecture upon taking up the position of Chair in Farming Systems Ecology at Wageningen University on 16 May 2013. Wageningen, NL: Wageningen Universiteit.

Tittonell, P. (2014). Ecological intensification of agriculture - sustainable by nature. *Current Opinion in Environmental Sustainability*, 8, pp. 53-61.

Thorsted, M., Olesen, J. and Weiner, J. (2006). Width of clover strips and wheat rows influence grain yield in winter wheat/white clover intercropping. *Field Crops Research*, 95(2-3), pp. 280-290.

Westermann, D. and Kleinkopf, G. (1985). Nitrogen Requirements of Potatoes. *Agronomy Journal*, 77(July-August 1985).

Wichmann, S., Loges, R. and Taube, F. (2006). Kornerträge, N2-Fixierungsleistung und N-Flächenbilanz von Erbsen, Ackerbohnen und Schmalblättrigen Lupinen in Reinsaat und im Gemenge mit Getreide. *Pflanzenbauwissenschaften*, 10(1), pp. 2-15.

Williams, I. (2010). *Biocontrol-based integrated management of oilseed rape pests*. Dordrecht: Springer. p 306.

Wishart, J., George, T., Brown, L., Ramsay, G., Bradshaw, J., White, P. and Gregory, P. (2012). Measuring variation in potato roots in both field and glasshouse: the search for useful yield predictors and a simple screen for root traits. *Plant Soil*, 368 (1-2), pp. 231-249.

Xiao, Y., Li, L. and Zhang, F. (2004). Effect of root contact on interspecific competition and N transfer between wheat and fababean using direct and indirect 15 N techniques. *Plant and Soil*, 262(1/2), pp. 45-54.

Yang, Z. (2015). Performance of two potato planting structures within a diverse cropping system in the Netherlands. MSc Thesis report, Wageningen University, Wageningen.

APPENDIX

The model rankings show the best model (with the lowest AIC according to the "dredge"-function of R's MuMIn-package), the full model based on which the "dredge"-function has been employed, the null model, a model only containing the predictor accounting for multimodality (M2) and the model fit for the purpose of pairwise comparisons (M3). The models have been ranked according to their Akaike weights, which indicate the weight of evidence in the model's favour.

The abbreviations used in the model ranking tables are as follows:

df	degrees of freedom
AIC_c	corrected AIC for finite sample sizes
Δ AIC_c	difference in AICc to the model with the lowest AICc value
weight	the Akaike weight indicating the weight of evidence in favour of the model

A 1: model ranking for the prediction of grass-clover dry weight (kg/m²)

GLM	treatment	plot	Field X	row	Field Y	df	AIC_c	Δ AIC_c	weight
best	X				X	69	-193.5221	0.0000	0.4996
M3	X			X		69	-193.5221	0.0000	0.4996
M2	X					70	-180.3866	13.1355	0.0007
full	X	X	X	X	X	48	-157.0237	36.4984	0.0000
null						71	-154.4130	39.1091	0.0000

A 2: model ranking for the prediction of wheat dry weight (kg/m²)

GLM	treatment	plot	Field X	row	Field Y	weeds	N_{stems}	N_{grains}	df	AIC_c	Δ AIC_c	weight
best	X				X	X			43	-66.1590	0.0000	0.9769
null									47	-57.4406	8.7184	0.0125
M3	X			X					40	-56.0436	10.1154	0.0062
M2	X								46	-55.3609	10.7981	0.0044
full	X	X	X	X	X	X	X	X	26	-12.3216	53.8374	0.0000

A 3: model ranking for the prediction of wheat kernel weight (kg/m²)

GLM	treatment	plot	Field X	row	Field Y	weeds	N_{stems}	N_{grains}	df	AIC_c	Δ AIC_c	weight
best	X	X			X				34	-91.0155	0.0000	0.9798
M3	X			X					40	-82.9514	8.0642	0.0174
M2	X								46	-78.0790	12.9366	0.0015
null									47	-77.7875	13.2281	0.0013
full	X	X	X	X	X	X	X	X	26	-51.6517	39.3638	0.0000

A 4: model ranking for the prediction of OSR dry weight (kg/m²)

GLM	treatment	plot	Field X	row	Field Y	N_{stems}	N_{seeds}	df	AIC_c	Δ AIC_c	weight
null*								47	-30.7107	0.0000	0.4620
best							X	46	-30.2214	0.4893	0.3617
M2	X							46	-28.7752	1.9355	0.1755
M3	X			X				40	-17.7287	12.9820	0.0007
full	X	X	X	X	X	X	X	28	23.3065	54.0172	0.0000

* Here, the null model was the best of this set of models. The fact that the null model leads the ranking means that a difference in the outcome variable cannot be explained by the variables in the alternative hypothesis, i.e. the "best model".

A 5: model ranking for the prediction of the number of OSR seeds (nr/m²)

GLM	treatment	plot	Field X	row	Field Y	N_{stems}	N_{seeds}	df	AIC_c	Δ AIC_c	weight
best					X	X		45	1019.2310	0.0000	0.9911
null								47	1029.2600	10.0295	0.0066
M2	X							46	1031.3550	12.1242	0.0023
M3	X			X				40	1042.5470	23.3168	0.0000
full	X	X	X	X	X	X	X	28	1056.4370	37.2070	0.0000

A 6: model ranking for the prediction of OSR seed weight (kg/m²)

GLM	treatment	plot	Field X	row	Field Y	N_{stems}	N_{seeds}	df	AIC_c	Δ AIC_c	weight
best					X	X		45	-195.0342	0.0000	0.9997
null								47	-177.9946	17.0396	0.0002
M2	X							46	-175.8334	19.2008	0.0001
M3	X			X				40	-164.2316	30.8027	0.0000
full	X	X	X	X	X	X	X	28	-154.0782	40.9561	0.0000

A 7: model ranking for the prediction of the OSR thousand seed weight (g)

GLM	treatment	plot	Field X	row	Field Y	N_{stems}	N_{seeds}	df	AIC_c	Δ AIC_c	weight
best	X		X					45	49.3998	0.0000	0.9998
null								47	67.2806	17.8808	0.0001
M2	X							46	68.9846	19.5847	0.0001
full	X	X	X	X	X	X	X	28	81.6147	32.2148	0.0000
M3	X			X				40	98.2798	48.8800	0.0000

A 8: model ranking for the prediction of potato fresh weight (kg/m²)

GLM	treatment	plot	Field X	row	Field Y	NO_3	NH_4	df	AIC_c	Δ AIC_c	weight
best	X		X	X		X		40	96.0833	0.0000	0.9965

null								47	107.9546	11.8713	0.0026
M2	X							46	110.2333	14.1500	0.0008
M3	X			X				40	115.2446	19.1614	0.0001
full	X	X	X	X	X	X	X	30	127.8343	31.7510	0.0000

A 9: model ranking for the prediction of potato fresh weight (kg/m²) of different varieties

GLM	variety	plot	Field X	row	Field Y	NO_3	NH_4	df	AIC_c	Δ AIC_c	weight
best	X			X	X			89	344.9146	0.0000	0.5000
M3	X			X				89	344.9146	0.0000	0.5000
M2				X				92	363.1004	18.1858	0.0001
null								95	366.8612	21.9466	0.0000
full	X	X	X	X	X	X	X	75	374.7102	29.7956	0.0000

A 10: model ranking for the prediction of potato fresh weight (kg/m²) from the machine harvest

GLM	treatment	plot	Field X	row	Field Y	NO_3	NH_4	df	AIC_c	Δ AIC_c	weight
best	X	X		X	X	X		32	41.7548	0.0000	0.9495
full	X	X	X	X	X	X	X	30	47.6212	5.8665	0.0505
M3	X			X				40	71.3244	29.5697	0.0000
null								47	97.0231	55.2683	0.0000
M2	X							46	97.0905	55.3358	0.0000

A 11: model ranking for the prediction of potato yield (kg/m²)

GLM	treatment	plot	Field X	row	Field Y	NO_3	NH_4	df	AIC_c	Δ AIC_c	weight
best	X		X					44	108.7850	0.0000	0.4674
null								47	109.1265	0.3415	0.3940
M2	X							46	111.2223	2.4373	0.1382
M3	X			X				40	122.8697	14.0846	0.0004
full	X	X	X	X	X	X	X	30	148.1749	39.3899	0.0000

A 12: model ranking for the prediction of the potato yield (kg/m²) of different varieties

GLM	variety	plot	Field X	row	Field Y	NO_3	NH_4	df	AIC_c	Δ AIC_c	weight
best	X					X		91	332.9318	0.0000	0.8418
M3	X			X				89	336.2781	3.3463	0.1580
null								95	349.5134	16.5816	0.0002
M2				X				92	352.4417	19.5099	0.0000
full	X	X	X	X	X	X	X	75	366.2866	33.3547	0.0000