Economic benefits from plant species diversity in intensively managed grasslands

Author(s):
Schaub, Sergei; Buchmann, Nina; Lüscher, Andreas; Finger, Robert

Publication date:
2020-02

Permanent link:
https://doi.org/10.3929/ethz-b-000376551

Originally published in:
Ecological Economics 168, https://doi.org/10.1016/j.ecolecon.2019.106488

Title: Economic benefits from plant species diversity in intensively managed grasslands

Authors: Sergei Schaub[a,b,*], Nina Buchmann[b], Andreas Lüscher[c] and Robert Finger[a].

[a] Agricultural Economics and Policy Group, ETH Zürich, Sonneggstrasse 33, CH-8092 Zürich, Switzerland

[b] Grassland Sciences Group, ETH Zürich, Universitätstrasse 2, CH- 8092 Zürich, Switzerland

[c] Forage Production and Grassland Systems, Agroscope, Reckenholzstrasse 191, CH-8046 Zürich, Switzerland

* Corresponding author. E-mail: seschaub@ethz.ch

Abstract

Grasslands cover a major share of the world's agricultural area and are important for global food security. Plant species diversity in grasslands is known to increase and stabilize biomass yields. We economically evaluate these effects, using a rich dataset from 16 intensively managed grassland sites across Europe. We extend earlier research by accounting for plant species diversity effects on both quantity and quality of yields. Consequently, we can express plant species diversity effects in terms of milk production potential yields per hectare and potential revenues thereof. Plant species diversity not only increased milk production potential yields and thus revenues, but also reduced production risks. Thus, increasing plant species diversity resulted in higher certainty equivalents, for example, the certainty equivalent rose by +29% when comparing the average mixture to the average monoculture. For risk averse decision makers, this gain in certainty equivalent was mainly due to the increase in revenues (accounting for 90%) compared to the total insurance value (accounting for 10%). Overall, we show that farmers benefit economically from plant species diversity and that even a moderate increase in this diversity contributes to more stable grassland-based production. Thus, our results are highly relevant for future sustainable intensification of grassland-based production.

Highlights

- We used a rich dataset from 16 intensively managed grassland sites across Europe
- Plant species diversity increased milk production potential yields, and thus revenue
- Production risks decreased in grasslands with higher plant species diversity
- Plant species diversity constituted a significant insurance value for farmers

Keywords: species diversity, insurance value, risk, stability, sustainable intensification, biodiversity

1. Introduction

Grasslands play a central role in global food security. They cover a major share of the world's agricultural area and are the basis for both forage production and a wide range of additional ecosystem services (Sala and Paruelo 1997). Growing population, changes in consumer demand and climatic challenges increase pressure on grassland-based production. Grassland biomass yields and their quality are affected by site-specific characteristics and farmers' management decisions, for example, with respect to land use intensity. Furthermore, plant species diversity also plays an important role in grasslands. This paper provides an economic evaluation of the effects of plant species diversity (henceforth 'diversity effect') on both quantity and quality of yields as well as production risks using an empirical analysis comprising 16 intensively managed grassland sites across Europe.

Plant species diversity can affect grasslands in three ways: Firstly, plant species diversity increases biomass yields (see e.g. Tilman et al. 1996, Marquard et al. 2009, Finn et al. 2013). This effect of plant species diversity is driven mainly by the complementarity and sampling effects (see e.g. Loreau and Hector 2001, Cardinale et al. 2007). The complementarity effect arises either from greater efficiency in acquisition of available resources, as different species have different needs or/and sources of resources, or from positive interactions between different species. One example for this positive interaction is the nitrogen fixing ability of legumes from which other species in the community also benefit (see e.g. Carlsson and Huss-Danell 2003, Lüscher et al. 2014), particularly in grass-legume mixtures. The sampling effect is based on the increased probability that a community includes highly performing species, which then also become dominant. Furthermore, plant species diverse grasslands suffer less from weed invasion (Suter et al. 2017, Connolly et al. 2018), which reduces both the risk of biomass yield loss of the sown species and the costs of weed control.

Secondly, increasing plant species diversity has been found to increase the stability of biomass yields over time (see e.g. Isbell et al. 2009, Hallett et al. 2017, Haughey et al. 2018). Communities with more species are in a better position to guarantee that some species maintain functioning even when others fail. This is called the 'insurance effect' of plant species diversity in the ecological literature (Yachi and Loreau 1999).

Thirdly, plant species diversity can also influence the quality of biomass yields. However, the reported effects on forage quality and quality corrected yields (biomass yields × forage quality) are ambiguous (see e.g. White et al. 2004, Deak et al. 2007, Khalsa et al. 2012, Sturludóttir et al. 2014). Furthermore, studies show that large effect sizes of plant species diversity for amount of biomass yields and quality corrected yields (i.e. nitrogen/protein yields) can be achieved already with an increase of species numbers from one to four species provided that legumes are combined with grasses (Kirwan et al. 2007, Finn et al. 2013, Suter et al. 2015). Implementing such moderate increases in plant species diversity can be considered already now as a ʼready to use' strategy for sustainable intensification.

Existing agricultural and ecological economic research shows that farmers benefit from both higher and more stable biomass yields (Baumgärtner 2007, Baumgärtner and Quaas 2010). However, only a few studies have monetarized the diversity effect on grassland yields and its stability (Schläpfer et al. 2002, Koellner and Schmitz 2006, Finger and Buchmann 2015, Binder et al. 2018). These studies show that grasslands with a higher plant species diversity generate increases in farmers' expected utility. However, current economic literature has four major shortcomings: Firstly, past studies have been restricted to single or a small number of sites. Secondly, earlier studies were usually limited to (very) extensively managed grasslands. Therefore, the implications of these studies for real dairy farm-level decision making remain limited. Thirdly, existing economic literature has not economically evaluated the potential impact of the diversity effect on quality corrected yields. However, accounting for forage quality is crucial for the performance of the production system as it determines the potential to produce meat and milk (e.g. Briner et al. 2015). As an exception, Binder et al. (2018) economically assessed the diversity effect on crude protein contents, in addition to the plant species diversity-biomass yields relationship. However, crude protein alone is insufficient to monetarize the impact of plant species diversity on quality corrected yields because (metabolizable) energy is usually the first restricting factor for ruminant production (Barnes et al. 2003). Fourthly, the effects of plant species diversity on the variability of quality corrected yields have not been addressed so far.

This study contributes to close these gaps by evaluating the diversity effect from an agricultural economic perspective. In this paper, we (1) investigate the mean response of biomass yields, i.e. dry matter of biomass yields (DM yields; kg ha^{-1}), forage quality, i.e. milk production potential per kg of DM yield (DM MPP; kg kg_{DM}^{-1}) and quality corrected yields, i.e. milk production potential yields (MPP yields; kg ha^{-1}), to altered plant species diversity levels in intensively managed grassland sites across a wide range of pedo-climatic conditions; (2) quantify effects of plant species diversity on the variability of biomass yields, forage quality and quality corrected yields. Thus, we quantify effects on production risks; (3) economically evaluate species diversity effects using certainty equivalents and stochastic dominance; (4) test whether the diversity effect is persistent when the best performing monocultures are compared with all mixtures and with the best performing mixtures. For our analysis, we use a dataset that comprises information from 16 experimental sites across Europe on grass monocultures, legume monocultures and grass-legume mixtures with four functionally distinct species (Kirwan et al. 2014).

The remainder of this paper is organized as follows. In Section 2, we develop the ecological-economic and econometric framework. The data and measurement of the data are described in Section 3, followed by results in Section 4. Finally, we discuss and present our conclusions in Section 5.

2. Methodological framework

2.1 Agricultural economic valuation of uncertain outcome

Farmers' production decisions influence forage production. In grassland-based production systems, forage production can be influenced by plant species diversity, D, as well as by other management and environmental factors, X, which have consequences on biomass yields, forage quality and quality corrected yields. In this paper, we use the variable dry matter yields (DM yields) for biomass yields, and milk production potential per kg of dry matter yield (DM MPP) for forage quality as only higher quality can lead to higher milk production per unit dry matter. Finally, we use the variable milk production potential yields (MPP yields) for quality corrected yields ($\mathrm{MPP\ yields}$ (D,X) $= \mathrm{DM\ yields}$ (D,X) × DM MPP (D,X)). The use of MPP yields allows a direct link between production and farm revenues, π:

$$\pi\ (\mathrm{D,X}) = \mathrm{MPP\ yields}\ (\mathrm{D,X}) \times \mathrm{p} \quad (1)$$

where p is the price of milk. The inherent variability of MPP yields causes revenues to be stochastic. This variability is a function of plant species diversity as well as other management and environmental factors.

We use an expected utility framework where plant species diversity is the decision variable. In this framework, the implicit costs arising from risk exposure (i.e. variability of revenues) are captured in the risk premium, RP (see e.g. Chavas 2004). The difference between the expected stochastic revenues E(π(D, X)) and the risk premium RP(D, X) is equal to the utility arising from a deterministic payment, the so called certainty equivalent (CE):

$$CE = E(\pi(D,X)) - RP(D,X) \quad (2)$$

where E(.) is the expectation operator. The risk premium can be approximated as follows (Pratt 1964):

$$RP \approx 0.5\, r\, Var(\pi(D,X)) \quad (3)$$

where r refers to the Arrow-Pratt risk coefficient of absolute risk aversion, which indicates the level of risk aversion of an individual. The coefficient is defined as: r = – U″/U', where U′ and U″ represent first and second derivatives of the utility function U(.), respectively. The relevance of the latter depends on the subjective risk preferences of individual decision makers. We assume farmers are risk averse (see e.g. Maart-Noelck and Musshoff 2014, Meraner and Finger 2018, Iyer et al. 2019). In the following, we use a relative coefficient of risk aversion of 2, if not mentioned otherwise, which represents rather risk averse behavior (Hardaker et al. 2015).

Var(π(D, X)) is the variance of the revenues and reflects that we expect plant species diversity (D) and other management and environmental factors (X) to affect the variability of quality corrected yields. We focus on deterministic price levels, so that the variance of the revenues can be expressed as Var(y(D, X)) p^2. Thus, the risk premium is defined as follows:

$$RP = 0.5\, r\, Var(\pi(D,X)) = 0.5\, r\, Var(y(D,X))\, p^2 \tag{4}$$

In turn, the calculation of CE is as follows:

$$CE = E(\pi(D,X)) - 0.5\, r\, Var\big(y(D,X)\big)p^2 \tag{5}$$

The insurance value of plant species diversity, IV(D), is the negative of the marginal effect of plant species diversity on the risk premium (Baumgärtner 2007, Finger and Buchmann 2015):

$$IV(D) = -\partial RP/\partial D = -0.5\, r\, p^2\, \partial\, Var\big(y(D,X)\big)/\partial D \tag{6}$$

Thus, IV(D) describes how plant species diversity reduces the cost of risks borne by farmers and allows us to monetarize the risk altering property of plant species diversity from a farmer`s utility perspective. Finally, we derive the total insurance value (total IV) at a certain plant species diversity level by:

$$total\ IV = \int_0^D IV(D)dD = \int_0^D (-(-0.5\, r\, p^2\, \partial\, Var\big(y(D,X)\big)\, /\partial D))dD \tag{7}$$

The CE requires precise information (or assumptions) about the coefficient of risk aversion r. This limitation is overcome by complementing the CE-based analysis with stochastic dominance analysis to obtain a more general basis for comparison without requiring exact information about risk aversion (see e.g. Chavas 2004). The stochastic dominance inference is based on a binary comparison, for example of plant species diversity *a* with plant species diversity *b*. Choice a is preferred over choice b if the utility arising from the respective probability density function of revenues is larger:

$$E_a U(\pi) \geq E_b U(\pi) \tag{8}$$

Applying first order stochastic dominance (see e.g. Chavas 2004), choice a dominates choice b when the underlying cumulative distribution functions, A(π) and B(π) respectively,[1] follow:

$$A(\pi) \leq B(\pi) \text{ for all } \pi \quad (9)$$

If choice a first order dominates choice b, the cumulative distribution function of choice a is always beneath and right of the cumulative distribution function of choice b. Thus, the first order stochastic dominance criterion is independent of risk preferences and only requires that $U' > 0$. However, the discriminatory power of the criterion is often low. If cumulative distribution functions cross, further stochastic dominance criteria can be used, such as the second order stochastic dominance (see e.g. Chavas 2004), which is defined as:

$$\int_{-\infty}^{\pi} A(t)dt \leq \int_{-\infty}^{\pi} B(t)dt \text{ for all } \pi \quad (10)$$

Thus, choice a dominates choice b when the total area beneath the cumulative distribution function of choice a at every level of π is smaller than the total area beneath the cumulative distribution function of choice b. The second order stochastic dominance criterion implies that the decision maker is risk averse, i.e. $r > 0$.

A major limitation of stochastic dominance criteria is that outcomes of these assessments for a specific sample are binary, i.e. that one choice (e.g. a specific level of plant species diversity) either dominates another or not. Thus, this approach remains inconclusive in many applications and cannot account for uncertainties underlying the sample composition and results might be driven by specific observations in the sample. In order to overcome these limitations and allow for statistical inference, we apply a test procedure based on a simulated Kolmogorov Smirnov test (Barrett and Donald 2003).[2] For this procedure, we use the 'simulation method 1' provided by Barrett and Donald (2003). The null-

[1] The cumulative distribution function is defined as the integral of its probability density function, e.g. for a: $A(\pi) = \int_{-\infty}^{\pi} a(t)dt$.

[2] The Kolmogorov-Smirnov simulation allows a comparison of revenues at all revenue values, and can be applied to compare different sample sizes. We can also use it to derive statistical inference, as we can estimate p-values in finite samples. This is also true for second order stochastic dominance, the test statistic of which does not have a closed form solution.

hypothesis that choice a dominates choice b, can be rejected when the p-value$_{a,b}$ < α. α represents significance levels of 0.1, 0.05 and 0.01. To obtain a conclusive inference that choice a dominates choice b, we need simultaneously p-value$_{a,b}$ > α and p-value$_{b,a}$ < α. Note that we do not consider stochastic dominance criteria of higher orders because their additional discriminatory power is expected to be low (Hardaker et al. 2015). Furthermore, it must be borne in mind that empirical stochastic dominance testing does not allow controlling for other influences than plant species diversity, such as for location of the site or year of the experiment. However, to account for key differences in productivity across sites, we only use observations from sites that cover all levels of our diversity gradient in the stochastic dominance analysis.[3]

2.2 Econometric implementation

We use the stochastic production function framework proposed by Just and Pope (1978) to identify the diversity effect on the expected outcome (including biomass yields, forage quality and quality corrected yields) and the variance of outcome.

Firstly, the expected outcome can be estimated by the stochastic production function, which is specified as:

$$y_{i,k} = \alpha + \beta_1 \, \mathrm{D_i^{0.5}} + \beta_2 \, \mathrm{Site_i} + \beta_3 \, \mathrm{Density_i} + \beta_4 \, \mathrm{Year_i} + \beta_5 \, \mathrm{Site_i}{\times}\mathrm{Cuts_i} + \mathrm{e_{1,i,k}} \quad (11)$$

where $y_{i,k}$ represents either annual biomass yields, forage quality or quality corrected yields depending on the index k across the different years for each plot, i. In equation (11) $y_{i,k}$ comprises all monocultures and all mixtures. We selected a square root specification of plant species diversity, $\mathrm{D_i^{0.5}}$, because this allows for a decreasing diversity effects (see e.g. Hooper et al. 2005, Finn et al. 2013) and when compared to other empirical model specifications,[4] it had the lowest Akaike information criterion (AIC) for annual biomass yields and quality corrected yields. To account for farmers' choices for sowing monocultures or mixtures of different plant species diversity levels and to avoid endogeneity problems

[3] Note that results of all sites for the stochastic dominance analysis can be found in Table A.4 and Fig. A.3

[4] Other empirical model specifications include $\mathrm{D_i}$, $\mathrm{D_i} + \mathrm{D_i^2}$ and $\mathrm{D_i} + \mathrm{D_i^{0.5}}$.

in the estimation, we use sown plant species diversity levels to measure $\mathrm{D_i}$. Moreover, we introduce site dummies, $\mathrm{Site_i}$, that account for structural differences in productivity across sites due to the location of the experimental sites. In addition to these site dummies, we include a set of explanatory variables, consisting of a dummy for high/low sown density, $\mathrm{Density_i}$, dummies for the year of the experiment, $\mathrm{Year_i}$, and an interaction term of site and number of cuts per year, $\mathrm{Site_i} \times \mathrm{Cuts_i}$, for those sites with varying cuts per year. The error term, e_1, comprises all uncontrolled factors, such as the variability of weather and pest infestations. Note that due to the experimental nature of our data, these components of the error term are uncorrelated with the explanatory variables. We assume that $E(\mathrm{e_{1,i,k}}) = 0$. Furthermore, we account for structural differences in the variability of the outcome variable across sites, i.e. the standard error is not only influenced by the individual observation when estimating biomass yields, forage quality and quality corrected yields, but also by a common site effect (common cluster effect). To this end, we follow Wooldridge (2003) and model the error term for equation (11) as follows:

$$\mathrm{e_{1,g,i,k}} = \mathrm{v_{1,g,k}} + \mathrm{z_{1,g,i,k}} \quad (12)$$

The subscript g refers to the cluster, i.e. the site, k to either biomass yields, forage quality or quality corrected yields, and i to the single plot. If this clustered structure is ignored in an ordinary least squares (OLS) regression it can lead to a strong bias of standard errors (Moulton 1986). We compute cluster-robust standard errors to correct these biases. This correction also accounts for the expected heteroscedasticity, i.e. the fact that the variance of residuals is bigger with lower levels of plant species diversity.

Secondly, the variance of outcome $\mathrm{y_{i,k}}$ is defined as $\mathrm{Var}(\mathrm{y_{i,k}}) = (\mathrm{y_{i,k}} - \bar{\mathrm{y}}_{\mathrm{i,k}})^2 = \mathrm{e}^2_{1,\mathrm{i,k}}$ (Just and Pope 1978) and econometrically specified as:

$$Var(y_{i,k}) = \alpha + \beta_1 \, \mathrm{D_i^{0.5}} + \beta_2 \, \mathrm{Site_i} + \beta_3 \, \mathrm{Density_i} + \beta_4 \, \mathrm{Year_i} + \beta_5 \, \mathrm{Site_i} \times \mathrm{Cuts_i} + \mathrm{e_{2,i,k}} \quad (13)$$

For this equation, we use again a square root specification of plant species diversity (see e.g. Hooper et al. 2005). The square root specification was chosen since it performed best among the empirical model specifications based on the AIC. We again compute cluster-robust standard errors for this estimation. Note we estimate all models separately using Stata 15.0 for Windows. The Stata code for the econometric estimations as well as the R (R Core Team 2018) code for data preparation are available in the online Appendix.

2.3 Comparison of best performer type I and II

A potential limitation when analyzing experimental data is that comparisons of all mixtures with all monocultures lack practical implications for farmers as they want to select only the best monoculture and mixture. This limitation is addressed in two ways. Firstly, the design of the underlying experiment focusses exclusively on high performing species. This reduces the importance of the sampling effect, which only plays a minor role in intensive agricultural systems. Secondly, in addition to the analysis in Section 2.2, we perform comparisons of all mixtures to the best performing monoculture as well as comparisons of the best mixture to the best monoculture. In the following, the former is referred to as comparison of best performer of *type I* and the latter of *type II*. Both comparisons of best performer assume that farmers know the best performing monoculture and/or mixture in advance.

Prior to these best performer analyses, we conduct a pre-test as proposed by Schmid et al. (2008). More specifically, we test if observations within a group are significantly different, i.e. if monocultures differ from each other and if mixtures differ from each other in terms of quality corrected yields. Only if within group differences exists, it makes sense to select the best performer of this group for our analysis. Moreover, to avoid a sampling bias in this analysis, we follow Schmid et al. (2008) and use the mean of all replicates per site and year, with replicates referring to plots with different sown densities and the same sown composition of species (see data description for details). Sown composition includes information about the presence and evenness of each species on a plot. The pre-test commences with a regression analysis, once for all monocultures and once for all mixtures, with quality corrected yields as the dependent variable and with sown compositions, site, year of experiment and cuts per year for sites with varying cuts per year as explanatory variables. Secondly, we conduct Wald tests to analyze whether the coefficients of the sown composition of a regression differ from each other.

Three different rules are applied to identify the best performer per site. Firstly, we assume that farmers will choose those options (monocultures/mixtures) that maximize the average quality corrected yields. Secondly, farmers are assumed to select the monoculture/mixture with the highest minimum quality corrected yields in one year (maximin rule). Thirdly, farmers select the monoculture/mixture with the maximal possible quality corrected yields in one year (maximax rule). The identification of the best performer is based on the average quality corrected yields of the replicates.[5]

The estimation of the diversity effect of the type I comparison differs from equation (11) as we do not control for sown density because the best performer is identified by the mean of the replicates, i.e. plots with same sown composition but different sown densities:

$$y_{i,k} = \alpha + \beta_1 \mathrm{D_i^{0.5}} + \beta_2 \mathrm{Site_i} + \beta_3 \mathrm{Year_i} + \beta_4 \mathrm{Site_i{\times}Cuts_i} + \mathrm{e_{3,i,k}} \quad (14)$$

Note the index k comprises only annual quality corrected yields across different years in this comparison of best performer. The dataset for this analysis includes all mixtures but only the best performing monocultures.

For type II, the model is adjusted as follows:

$$y_{i,k} = \alpha + \beta_1 \mathrm{Mixture_i} + \beta_2 \mathrm{Site_i} + \beta_3 \mathrm{Year_i} + \beta_4 \mathrm{Site_i{\times}Cuts_i} + \mathrm{e_{4,i,k}} \quad (15)$$

In this model, k comprises only annual quality corrected yields across different years. Furthermore, we use a mixture dummy, $\mathrm{Mixture_i}$, as we only have one mixture per site (the best performing mixture), thus, there is no meaningfully gradient of the diversity gradient. The data for the type II comparison

[5] General selection rules determining how farmers choose the best grassland monoculture/mixture comprise opting for the highest profits, the highest yields, the highest utility, or simply copying the status quo (Huber et al. 2018). In practice, decisions about which plants to sow must be taken on the basis of past experience and before the actual yield is known, whereas our selection rules are based on actual yields. Hence, our best performer comparison of type I is more conservative, as it selects the monoculture that really performed the best. Furthermore, seed costs of monocultures and mixtures in the study are very low compared to other costs and management costs for monocultures and mixtures are fairly similar. Thus, these costs do not drive farmers' decisions.

consists exclusively of the best performing mixtures and the best performing monocultures. Both estimations, equations (14) and (15), are corrected for clustered error terms and heteroscedasticity as described above.

3. Data – forage quantity and forage quality

3.1 Experimental setup

The analysis is based on biomass yields and forage quality data retrieved from the database of the *COST Agrodiversity Experiment* (Kirwan et al. 2014), which is available online. A subset of this dataset was used here, including data from 16 intensively managed grassland sites (8 European countries, Figure 1). The coordinated experiment was carried out between 2003 and 2011 and lasted at each site between 2 to 4 years. However, forage quality data was not collected in all years at some sites (see Table A.1 for details). The specific management regimes at each site represented intensive management adopted to the local condition. Over all sites, the management intensity ranged between 2 and 5 cuts and between no fertilizer and 150 kg ha^{-1} nitrogen, 70 kg ha^{-1} phosphorus and 420 kg ha^{-1} potash fertilizer annually. In general, management intensity did not vary within a site, except for three sites at which the numbers of cuts per year varied between years (Table A.1). The experiment comprised four monocultures and eleven mixtures of four functional types of species that produce forage of high quality in intensively managed systems. The eleven mixtures with low species numbers contained four species in different sown proportions (four dominated mixtures with 0.7, 0.1, 0.1, 0.1 in turn, six co-dominated mixtures with 0.4, 0.4, 0.1, 0.1 in turn, and one equiproportional mixture with 0.25, 0.25, 0.25, 0.25; Kirwan et al. 2014). The monoculture and mixtures thus reflect a plant species diversity gradient (*sensu* Isbell et at. 2015, Connolly et al. 2018, Finn et al. 2018). These diversity levels depend on the plant species richness (number of species) and their evenness (relative abundance) in the sward and can be expressed by the Simpson index of diversity (Krebs 1999). The Simpson index is frequently used and considers species richness and evenness:

$$Simpson\ index = 1 - \sum (h_i)^2 \qquad (15)$$

where h_i is the proportion of individuals of species i in the community. The respective values of the Simpson index of sown plant species diversity in the experiment are: 0.48, 0.66 and 0.75 for the different mixtures and 0 for monoculture. The functional types of the four species were selected in order to achieve large functional differences among them, and consisted of a fast-establishing grass, a slow-establishing persistent grass, a fast-establishing legume and a slow-establishing persistent legume. The same four were used in each mixture (Kirwan et al. 2014). Each sward was sown in two density levels. For more information about the experimental design, see Kirwan et al. (2014).

Figure 1: Location of experimental sites used in this study. Numbers indicate the site indications according to Kirwan et al. (2014). See Table A.1 for details on the sites and mixture types.

3.2 Data collection and measurement

DM yields (kg ha^{-1}) were determined for each plot and harvest (Kirwan et al. 2014). For evaluating and monetarizing forage quality, we focused on DM MPP (kg kg_{DM}^{-1}). DM MPP describes the potential milk produced per DM yield and it was derived from metabolizable energy (DM ME; MJ kg_{DM}^{-1}) and ash content (g kg_{DM}^{-1}; Gierus et al. 2012, Jans et al. 2015):

$$DM\ MPP = DM\ ME\ \times [0.46 + 12.38 \times DM\ ME/(1000 - ash)]/3.14 \quad (17)$$

DM ME is the energy content per DM yield available for maintenance, milk production and weight gains and is used to assess overall ruminant-specific nutritive value. (Metabolizable) Energy is usually the factor of forage that is first limiting in ruminant production (Barnes et al. 2003). DM ME and ash content (g kg_{DM}^{-1}; ash consists mainly of minerals) were obtained by using near infrared reflectance spectroscopy (NIRS). DM yields were multiplied with DM MPP to compute MPP yields (kg ha^{-1}):

$$MPP\ yields = DM\ yields\ \times DM\ MPP \quad (18)$$

Moreover, the annual DM yields and MPP yields were calculated by tallying up all harvests per year.[6] The average DM MPP of a year was calculated by taking the mean DM MPP of all harvests weighted by DM yields of the harvests. See Kirwan et al. (2014) for more details about data collection. Finally, the annual revenues from milk sales were calculated by multiplying annual MPP yields with the milk price of 0.35 Euro kg_{milk}^{-1} (EU average of the years 2013 to 2015; Eurostat 2017).[7] Furthermore, results of DM ME and ME yields are available in the online Appendix.

The summary statistics in Table 1 shows that DM yields and MPP yields have a mean of 8748 kg ha^{-1} and 18750 kg ha^{-1} across all observations. The mean for both is lower for monocultures and higher for mixtures. In contrast, mean DM MPP is comparable between monocultures and mixtures. DM yields and MPP yields vary considerably among sites, while DM MPP is more similar across sites (Fig. A.1 to A.3).

[6] Note for site 10 quality was only measured for the first four out of five cuts each year. As this is consistent for all plots and the first cuts are agronomically more important, these observations are included in the analysis.
[7] In addition, we conducted a sensitivity analysis covering different milk prices, ranging from the lowest to the highest EU average milk prices (2013-2015) of a country with a site in our data: 0.27 to 0.36 Euro kg_{milk}^{-1}.

Table 1: Summary statistics: mean (and standard deviation) of annual dry matter yields (DM yields), milk production potentials per kg of DM yield (DM MPP) and milk production potential yields (MPP yields.

	All observations	Monocultures	Mixtures
DM yields (kg ha^{-1})	8748 (3886)	7235 (3519)	9614 (3824)
DM MPP (kg kg_{DM}^{-1})	2.15 (0.09)	2.15 (0.10)	2.14 (0.09)
MPP yields (kg ha^{-1})	18750 (8599)	15500 (7759)	20610 (8512)

The total number of observations in our data is 698, thereof are 254 monocultures and 444 mixtures. Due to the experimental nature of the data, the Simpson index of mixtures is restricted to three levels, i.e. 0.48 (246 observations), 0.66 (132 observations) and 0.75 (66 observations).

4. Results

In this section, we present first the empirical results of the diversity effect on expected outcomes, including dry matter yields (DM yields), milk production potentials per kg of DM yield (DM MPP), and milk production potential yields (MPP yields), and their respective variance (4.1). Next, we use these empirical results for computing the certainty equivalents, expected revenues as well as total insurance values and we perform the stochastic dominance analysis (4.2). In the end of this section, we conduct the comparison of best performer (4.3).

4.1 Diversity effects on DM yields, DM MPP and MPP yields

Our analysis showed a positive but diminishing effect of plant species diversity on DM yields and MPP yields, while the effect on DM MPP was not significant (Table 2).[8] DM yields and MPP yields increased by about 25% when comparing the average mixture (Simpson index = 0.6)[9] with the average monoculture. Furthermore, plant species diversity reduced production risks, as the variance of all three

[8] Note that results of DM ME and ME yield are similar to the results of DM MPP and MPP yield and can be found in Table A.3.

[9] Diversity effects for the average mixture represents the diversity effects at the mean Simpson index value of all mixtures, which is 0.6.

outcomes decreased (Table 2). More specifically, the diversity effect on the variance of DM yields was significantly negative, while the diversity effect on the variance of DM MPP was negative but not significant. The overall diversity effect on the variance of MPP yields was also negative and significant at the 10% level.

Table 2: Results of the effect of plant species diversity on dry matter yields (DM yields), milk production potentials per kg of DM yield (DM MPP) and milk production potential yields (MPP yields; equation 11) and their variance (equation 13).

	Expected outcome		Variance of outcome
DM yields (kg ha^{-1})		Var(DM yields)	
Simpson index$^{0.5}$ ($D^{0.5}$)	2599 (6.24)***	Simpson index$^{0.5}$ ($D^{0.5}$)	-2.79×10^{6} (-2.41)**
R^2_{adj}	0.766	R^2_{adj}	0.157
DM MPP (kg kg_{DM}^{-1})		Var(DM MPP)	
Simpson index$^{0.5}$ ($D^{0.5}$)	-0.02 (-1.31)	Simpson index$^{0.5}$ ($D^{0.5}$)	-0.011 (-1.16)
R^2_{adj}	0.465	R^2_{adj}	0.025
MPP yields (kg ha^{-1})		Var(MPP yields)	
Simpson index$^{0.5}$ ($D^{0.5}$)	5423.78 (6.32)***	Simpson index$^{0.5}$ ($D^{0.5}$)	-1.13×10^{7} (-1.95)*
R^2_{adj}	0.787	R^2_{adj}	0.161

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. Numbers in parentheses are t-values. T-values are corrected for clustered error terms and heteroscedasticity. Number of observations = 698. Adjusted R^2 refer to the full models and all the coefficients of the models are available in the online Appendix.

4.2 Certainty equivalent and stochastic dominance

The expected revenues, total insurance value, and thus CE increased with plant species diversity (Fig. 2). The rates of increase were lower at higher levels of plant species diversity. Due to the positive effect of plant species diversity on CE, we report a positive effect of plant species diversity on farmers' utility. Overall, the diversity effect increased the CE for the average mixtures (Simpson index = 0.6) by about 1630 Euro ha^{-1} compared to the average monocultures for rather risk averse farmers ($r = 2$). In relative terms, this is a gain of about 29%. To a large extent, the increase was due to a gain in expected revenues of about 1470 Euro ha^{-1} and to a lesser extent to a gain in total insurance value of about 160 Euro ha^{-1}. Re-running the analysis with varying coefficients of risk aversion, from low risk aversion ($r = 0.5$) to very high aversion ($r = 4$; Hardaker et al. 2015), resulted in a total insurance value for the average mixtures between about 40 to 330 Euro ha^{-1} (Fig. 3). The relative values for this range of total insurance values compared to expected revenues of the mean mixture are about 3% and 22%.

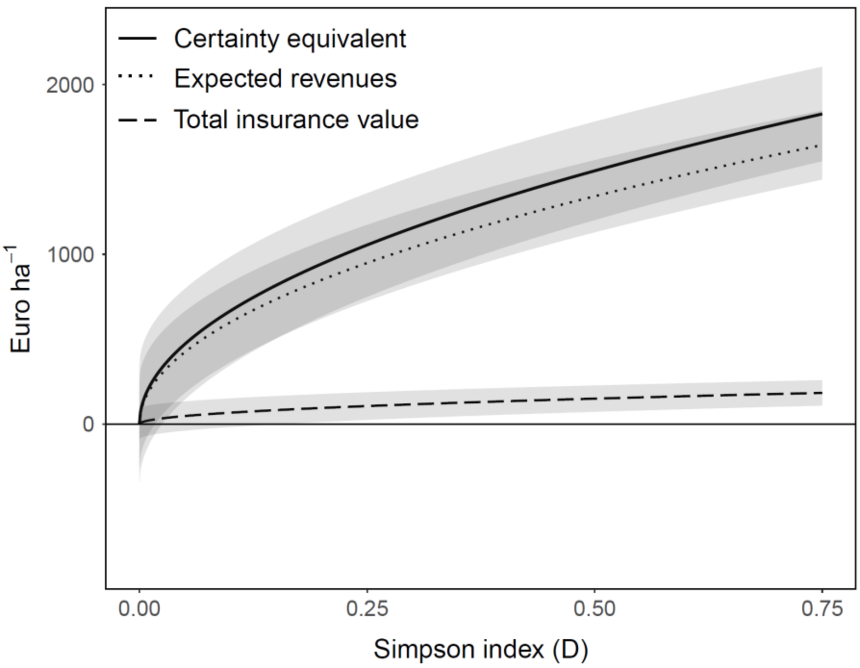

Figure 2: Effect of plant species diversity on the certainty equivalent (CE), expected revenues and total insurance value in Euro ha^{-1} for rather risk averse farmers.[10]

Note: The values show the species diversity effect for the mean site and a coefficient of risk aversion of 2. The diversity effect represents the additional gain from species diversity compared to the average monoculture. CE = expected revenues + total insurance value. The 90-percent confidence intervals are based on standard errors that are corrected for clustered error terms and heteroscedasticity.

[10] Assuming milk prices ranging from 0.27 to 0.36 Euro kg_{milk}^{-1}, the CE, the expected revenues and the total insurance value for the average mixtures compared to a monoculture range from ~1260 to ~1680 Euro ha^{-1}, from ~1130 to ~1510 Euro ha^{-1} and from ~130 to ~170 Euro ha^{-1}, respectively. More detailed results are available upon request from the authors.

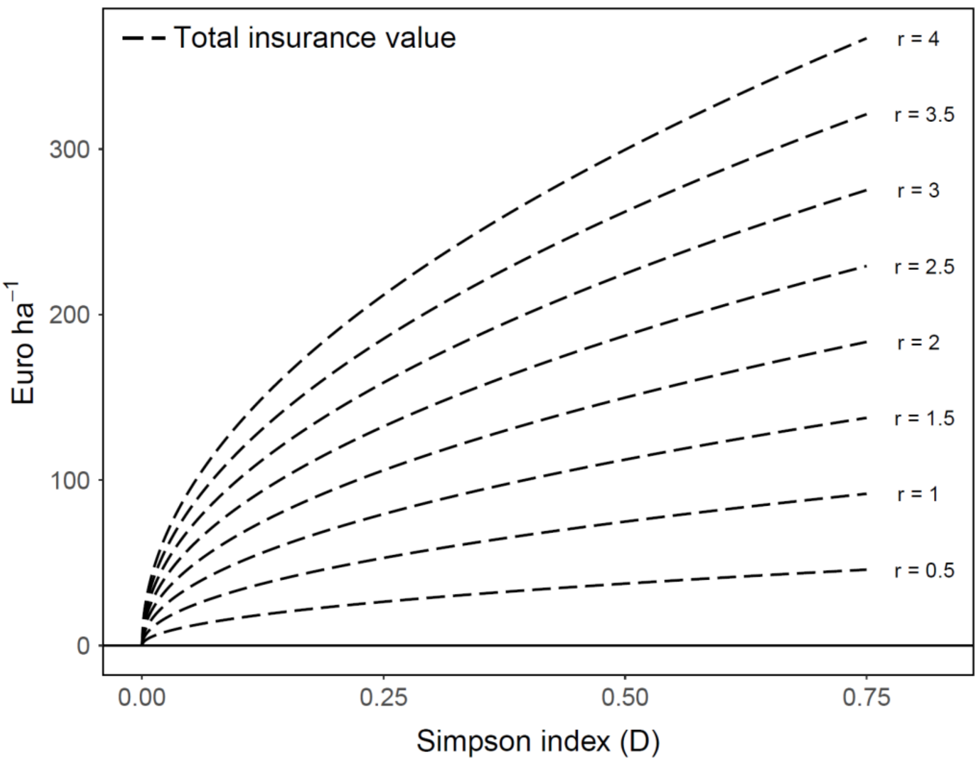

Figure 3: Effect of plant species diversity on total insurance value in Euro ha^{-1} for a range of risk aversions of farmers. Increasing coefficients r represent increasing risk aversion of farmers.

Concerning stochastic dominance, we found that the tested mixtures (Simpson index 0.48, 0.66 and 0.75) dominate the monocultures (Simpson index = 0) in a first and second order stochastic dominance sense. This is illustrated in Figure 4, where the cumulative distribution function of monocultures is always to the left of the cumulative distribution functions of mixtures. This dominance of mixtures was confirmed by the empirical Kolmogorov Smirnov test as for all mixtures the p-value$_{\text{Mixtures, Monocultures}}$ > 0.1 and the p-value$_{\text{Monocultures, Mixtures}}$ < 0.1. Regarding dominance across mixtures, we found no first or second order dominance.

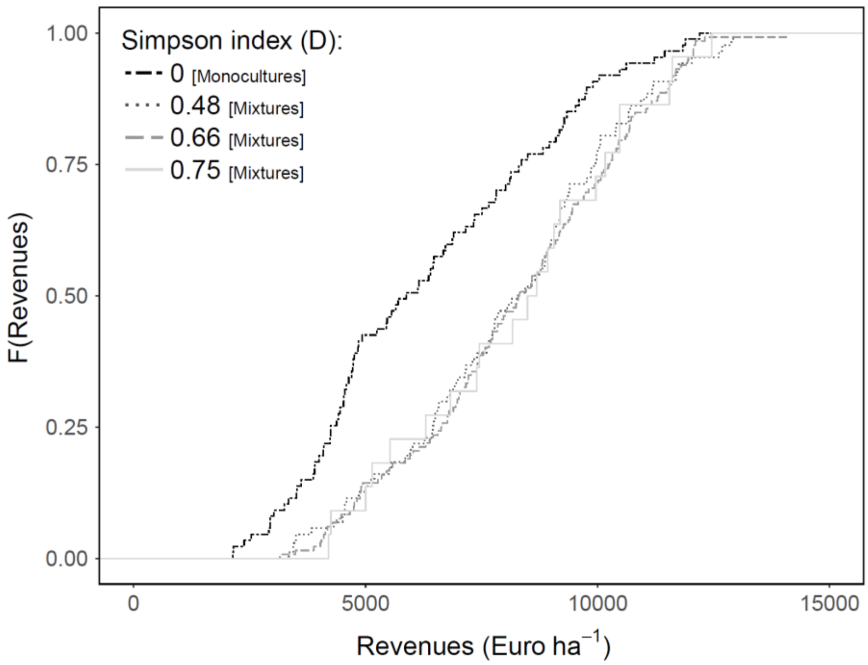

Figure 4: Cumulative distribution function of milk revenues per Simpson index level. F(Revenue) indicates the probability of the observed revenue falling below a certain level.

Note: Observations are only from sites that comprise the entire range of the Simpson index. Number of sites = 6. Number of observations = 328.

Table 3: Results of the empirical Kolmogorov Smirnov test for stochastic dominance of first and second order. Binary comparison of swards with different Simpson indices (0 = monocultures; 0.48, 0.66 and 0.75 = mixtures).

Simpson index (D)		First order stochastic dominance		Second order stochastic dominance	
A	B	p-value$_{a,b}$	p-value$_{b,a}$	p-value$_{a,b}$	p-value$_{b,a}$
0.48	0	1	<0.001	0.8	<0.001
0.66	0	1	<0.001	0.4	<0.001
0.75	0	1	0.018	0.8	<0.001
0.66	0.48	0.857	0.293	0.4	0.2
0.75	0.48	0.936	0.481	1	0.2
0.75	0.66	0.732	0.77	0.6	0.4

Note: The null-hypothesis, that for example choice a dominates choice b, cannot be rejected if the p-value$_{a,b}$ > α. Observations are only from sites that comprise the entire range of the Simpson index. Number of sites = 6. Number of observations = 328.

4.3 Comparison of best performer type I and II

Additional to the above analysis, i.e. all mixtures vs. all monocultures, we conducted a comparison of best performer of type I, i.e. all mixtures vs. the best monocultures, and type II, i.e. the best mixtures vs. the best monocultures. We did this because the pre-test revealed significant differences within groups (Wald tests for differences within monocultures: p-value=0.02 and for differences within mixtures: p-value<0.001).

The comparison of best performer of type I showed that, regardless of the rule defining how the best monocultures were selected, plant species diversity had a positive effect on MPP yields (Table 4). Moreover, MPP yields of mixtures were also significantly higher in the comparison of best performer of type II for all rules (Table 4). Here, a mixture dummy instead of the square root of the Simpson index captures the diversity effect. The diversity effect was lowest when employing the maximize average MPP yields rule and the maximax rule, and was highest when using the maximin rule. Overall, the effect of plant species diversity was smaller in the best performer comparisons of type I and II than it was when all monocultures and all mixtures were compared (see Table 2).

Table 4: Results of the comparison of best performer of type I and type II with milk production potential yields (MPP yields; kg ha^{-1}) as the dependent variable.

Rule	Type I		Type II	
Maximize average MPP yields	Simpson index$^{0.5}$ ($D^{0.5}$)	1068.36 (2.06)*	Mixture dummy	2324.81 (4.84)***
	Adjusted R^2	0.849	Adjusted R^2	0.759
Maximin	Simpson index$^{0.5}$ ($D^{0.5}$)	1891.79 (2.37)**	Mixture dummy	2648.18 (3.20)***

rule	Adjusted R^2	0.843	Adjusted R^2	0.755
Maximax	Simpson index$^{0.5}$ ($D^{0.5}$)	1068.36 (2.06)*	Mixture dummy	2291.50 (4.57)***
rule	Adjusted R^2	0.849	Adjusted R^2	0.753

* p < 0.1; ** p < 0.05; *** p < 0.01. Numbers in parentheses are t-values. T-values are corrected for clustered error terms and heteroscedasticity. The rules maximize average MPP yields and maximax lead to the selection of the same best monocultures. Number of observations: type I = 473, type II = 64.

5. Discussion and conclusion

We found that plant species diversity as a production factor in intensively managed grasslands was beneficial for milk production per area of land as already a moderate increase in plant species diversity (up to 4 species) increased and stabilized quality corrected yields, i.e. MPP yields. Khalsa et al. (2012) also showed a positive impact of plant species diversity on energy related yields, i.e. gross energy, in extensively managed grasslands. In the case of biomass yields, i.e. DM yields, our results that grasslands with higher plant species diversity produce more DM yields support findings from earlier studies on both intensively managed grasslands (e.g. Finn et al. 2013[11]) and extensively managed grasslands (up to 60 plant species; e.g. Marquard et al. 2009; Tilman et al. 1996). Plant species diversity did not affect forage quality, i.e. DM MPP, which is similar to results of earlier studies that found insignificant to small effects (Deak et al. 2007, Khalsa et al. 2012, Sturludóttir et al. 2014). However, our results on forage quality differed from White et al. (2004), who found a negative relationship between plant species diversity and forage quality. The latter study, however, did not control for differing environmental conditions of grasslands with low and high plant species diversity. We can also confirm former findings concerning the positive stability impact of plant species diversity on DM yields for intensively managed grasslands (e.g. Isbell et al. 2009, Hallett et al. 2017, Haughey et al. 2018, Wang et al. 2019). Moreover, we found no significant diversity effects on the variability of DM MPP.

[11] Note that we employed a subset (16 out of 31 sites) of the biomass data used by Finn et al. (2013).

As a result, MPP yields were more stable at higher plant species diversity, which consequently implies lower production risks for farmers. Therefore, we can show a significant insurance effect of plant species diversity not only for DM yields but also for MPP yields, even when considering a moderate increase in plant species diversity.

We used two complementary approaches to assess the economic value of plant species diversity from the perspective of a risk averse farmer, i.e. certainty equivalents and stochastic dominance. Increased plant species diversity implies considerable monetary benefits for farmers in terms of CE. The CE benefits amounted to a large part from an increase in the expected revenues and to a lesser degree to an increase in the total insurance value of plant species diversity. In the case of rather risk averse farmers ($\mathrm{r} = 2$), the respective values were about 1470 and 160 Euro ha^{-1} for the average mixtures compared to the average monoculture. Thus, our results show that farmers gain from plant species diversity already at low levels of species diversity. Therefore, farmers have an incentive to increase plant species diversity until the resulting costs exceed the expected increase in CE. The optimal level of plant species diversity increases with farmers' risk aversion. Costs of plant species diversity were not considered in our analysis as the costs of experiments cannot serve as a proxy for farm settings. Stochastic dominance also supported the CE inference that mixtures were preferred over monocultures. However, we could not determine that mixtures with higher levels of plant species diversity dominated mixtures with lower levels. This might be due to the fact that stochastic dominance analysis does not allow controlling for any independent variables other than plant species diversity or to the diminishment of the diversity effect and that our mixtures only differ in evenness. The latter would indicate that in agricultural practice, the presence of different species is more important than the exact distribution of each.

Finally, the comparison of best performer of type I and II also showed that grasslands with greater plant species diversity generated higher MPP yields. These additional comparisons were more conservative tests than comparing all mixtures with all monocultures as these comparisons assume that farmers know swards' performances. However, farmers might only be able to select good, but not the best species, as Finn et al. (2013) showed that the best species often changes across years (in 26 out of 54 cases). Thus, our comparison of all mixtures with all monocultures from a pool of four high performing species, reflects decision making in agricultural practice. In addition, the uncertainty about which species will produce the best performance in a certain year is a strong argument in favor of mixtures. This is because the increased number of species in the sward ultimately increases the probability of it containing the best performing species, leading to higher and more stable quality corrected yields.

Our findings on risk reducing properties and possible economic benefits of plant species diversity in grasslands are in line with earlier research (Schläpfer et al. 2002, Koellner and Schmitz 2006, Finger and Buchmann 2015, Binder et al. 2018). Our results go beyond these earlier findings and show that this also holds for quality corrected yields, i.e. MPP yields, a critical variable for farmers' decision making, in intensively managed grasslands. Moreover, our results are more robust and transferable into real world settings because we use data from a wide range of pedo-climatic conditions across Europe. In addition, the plant species in the experiment are suitable for intensive management and they meet quality demands of lactating cows. Thus, our findings are highly important for farmers, extension services and policy makers. Policy makers and extension services for example can use our results about the plant species diversity benefits, expressed in economic terms, to encourage farmers to increase plant species diversity in grasslands. Furthermore, the mixtures used in our analysis are mixtures with low species numbers readily available for application in agricultural practice. There are several ways to increase plant species diversity, such as over-sowing existing swards with additional species, drilling of seed mixtures into weed free plots, or sowing ex-arable fields with mixtures (see e.g. Walker et al. 2004). Furthermore, the risk reduction property of plant species diversity, even at low plant species diversity and when considering quality corrected yields, supports that plant species diversity can substitute financial insurances (Baumgärtner 2007). This property is crucial, especially because insurance mechanisms are often not available for grasslands (see e.g. Vroege et al. 2019). Furthermore, the positive insurance value is likely to increase in importance under more variable and uncertain future environmental conditions. Thus, our findings contribute more general to better risk management in dairy production (e.g. Finger et al. 2018). There are other positive effects of plant species diversity which are not included in our valuation, such as supporting pollinators, providing weed control and nitrogen fixation (Potts et al. 2009, Suter et al. 2015, Suter et al. 2017, Connolly et al. 2018). Our results show that private and public goal functions are not in conflict to each other, but increased plant species diversity may induce private and public welfare gains. Policy makers could further promote these positive diversity effects by providing incentives for more plant species diverse temporal grasslands. Baumgärtner and Quaas (2010) show for example that subsidies can be efficient policy measures to increase plant species diversity and provide private and public benefits.

In conclusion, our results show that already a moderate increase in plant species diversity is beneficial for the quantity and stability of quality corrected yields, i.e. MPP yields, in intensively managed grasslands and across a wide range of pedo-climatic conditions. By expressing these results in terms of DM MPP and MPP yields and their variability, this study is the first which draws the direct link to secondary production and thus to revenue streams from such. The positive economic diversity effect has important implications for farmers' long-term perspective. Firstly, the direct insurance value of plant species diversity reduces temporal risk and thus increases farmers' robustness to shocks. Secondly, gains from quality corrected yields and revenues due to plant species diversity increase revenues and flexibility and thus contribute to higher adaptability of farm management. It follows that plant species diversity can contribute to overall more resilient and stable agricultural systems (see e.g. Meuwissen et al. 2019).

6. Acknowledgements

This study was supported by the Mercator Foundation Switzerland within a Zürich-Basel Plant Science Center PhD Fellowship program and the project BU1080/4-1 from the Deutsche Forschungsgemeinschaft.

7. References

Barrett, G. F., Donald, S. G., 2003. Consistent tests for stochastic dominance. Econometrica 71(1):71-104.

Barnes, R. F., Nelson, C. J., Collins, M., Moore, K. J., 2003. Forages. An introduction to grassland agriculture. Sixth ed. Iowa State Press, Iowa, USA.

Baumgärtner, S., 2007. The insurance value of biodiversity in the provision of ecosystem services. Nat. Resour. Model 20(1), 87-127.

Baumgärtner, S., Quaas M. F., 2010. Managing increasing environmental risks through agrobiodiversity and agrienvironmental policies. Agric. Econ. 41(5), 483-496.

Binder, S., Isbell, F., Polasky, S., Catford, J. A., Tilman, D. 2018. Grassland biodiversity can pay. PNAS 115(15), 3876-3881.

Briner, S., Lehmann, N., Finger, R., 2015. Bio-economic modeling of decisions under yield and price risk for suckler cow farms. Anim. Prod. Sci. 55(1), 64-73

Cardinale, B.J., Wright, J.P., Cadotte, M.W., Carroll, I.T., Hector, A., Srivastava, D.S. et al., 2007. Impacts of plant diversity on biomass production increase through time because of species complementarity. PNAS 104(46), 18123-18128.

Carlsson, G., Huss-Danell, K., 2003. Nitrogen fixation in perennial forage legumes in the field. Plant Soil 253(2), 353-372.

Chavas, J. P., 2004. Risk Analysis in Theory and Practice. Elsevier Academic Press, San Diego, USA.

Connolly, J., Sebastià, M. T., Kirwan, L., Finn, J. A., Llurba, R., Suter, M., et al., 2018. Weed suppression greatly increased by plant diversity in intensively managed grasslands: A continental-scale experiment. J. Appl. Ecol. 55(2), 852-862.

Deak, A., Hall, M. H., Sanderson, M. A., & Archibald, D. D. 2007. Production and nutritive value of grazed simple and complex forage mixtures. Agron J. 99(3), 814-821.

[dataset] Eurostat, 2017. Selling prices of raw cow's milk.
http://ec.europa.eu/eurostat/tgm/table.do?tab=table&init=1&language=en&pcode=tag00070&plugin=1 (accessed 13 June 2017).

Finger, R., Buchmann, N., 2015. An ecological economic assessment of risk-reducing effects of species diversity in managed grasslands. Ecol. Econ. 110, 89-97.

Finger, R., Dalhaus, T., Allendorf, J., Hirsch, S., 2018. Determinants of downside risk exposure of dairy farms. Eur. Rev. Agric. Econ. *45*(4), 641-674.

Finn, J. A., Kirwan, L., Connolly, J., Sebastià, M. T., Helgadottir, A., Baadshaug, O. H., 2013. Ecosystem function enhanced by combining four functional types of plant species in intensively managed grassland mixtures: a 3-year continental-scale field experiment. J. Appl. Ecol. 50(2), 365-375.

Gierus, M., Kleen, J., Loges, R., Taube, F., 2012. Forage legume species determine the nutritional quality of binary mixtures with perennial ryegrass in the first production year. Anim. Feed Sci. Tech. 172(3-4), 150-161.

Hallett, L. M., Stein, C., Suding, K. N., 2017. Functional diversity increases ecological stability in a grazed grassland. Oecologia 183(3), 831-840.

Hardaker, J. B., Lien, G., Anderson, J. R., Huirne, R. B M., 2015. Coping with risk in agriculture: applied decision analysis. Third ed. CAB International, Wallingford, UK.

Haughey, E., Suter, M., Hofer, D., Hoekstra, N. J., McElwain, J. C., Lüscher, A., Finn, J. A., 2018. Higher species richness enhances yield stability in intensively managed grasslands with experimental disturbance. Sci. Rep. 8, 15047.

Hooper, D.U., Chapin Iii, F.S., Ewel, J.J., Hector, A., Inchausti, P., Lavorel, S., et al., 2005. Effects of biodiversity on ecosystem functioning: a consensus of current knowledge. Ecol. Monogr. 75(1), 3–35.

Huber, R., Bakker, M., Balmann, A., Berger, T., Bithell, M., Brown, C., et al, 2018. Representation of decision-making in European agricultural agent-based models. Agric. Syst. 167, 143-160.

Just, R., E., Pope, R., D., 1978. Stochastic specification of production functions and economic implications. J. Econom. 7(1), 67-86.

Isbell, F., Polley, H. W., Wilsey, B. J., 2009. Biodiversity, productivity and the temporal stability of productivity: patterns and processes. Ecol. Lett. 12(5), 443-451.

Isbell, F., Craven, D., Connolly, J., Loreau, M., Schmid, B., Beierkuhnlein, C., et al., 2015. Biodiversity increases the resistance of ecosystem productivity to climate extremes. Nature 526(7574), 574-577.

Iyer, P., Bozzola, M., Hirsch, S., Meraner, M., Finger, R. 2019. Measuring Farmer Risk Preferences in Europe: A Systematic Review. J. Agric. Econ. in Press

Jans, F., Kessler, J., Münger, A., Schlegel, P., 2015. Fütterungsempfehlungen für die Milchkuh, in: Agroscope (Eds.), Grünes Buch. Agroscope. Posieux, Switzerland.

Khalsa, J., Fricke, T., Weisser, W. W., Weigelt, A., Wachendorf, M., 2012. Effects of functional groups and species richness on biomass constituents relevant for combustion: results from a grassland diversity experiment. Grass Forage Sci. 67(4), 569-588.

Kirwan, L., Lüscher, A., Sebastià, M. T., Finn, J. A., Collins, R. P., Porqueddu, C., et al., 2007. Evenness drives consistent diversity effects in intensive grassland systems across 28 European sites. J. Appl. Ecol. 95(3), 530-539.

[dataset] Kirwan, L., Connolly, J., Brophy, C., Baadshaug, O. H., Belanger, G., Black, A. , et al., 2014. The Agrodiversity Experiment: three years of data from a multisite study in intensively managed grasslands. Ecological Archives, E095-232-D1. http://esapubs.org/archive/ecol/E095/232/#data

Koellner, T., Schmitz, O. J., 2006. Biodiversity, ecosystem function, and investment risk. BioScience 56(12), 977-985.

Krebs, C.J., 1999. Ecological methodology. Second ed. Addison Wesley, Menlo Park, USA.

Loreau, M., Hector, A., 2001. Partitioning selection and complementarity in biodiversity experiments. Nature 412(6842), 72-76.

Lüscher, A., Mueller-Harvey, I., Soussana, J. F., Rees, R. M., Peyraud, J. L., 2014. Potential of legume-based grassland–livestock systems in Europe: a review. Grass Forage Sci. 69(2), 206-228.

Maart-Noelck, S. C., Musshoff, O., 2014. Measuring the risk attitude of decision-makers: are there differences between groups of methods and persons? Aust. J. Agr Resour. Ec. 58(3), 336-352.

Marquard, E., Weigelt, A., Temperton, V. M., Roscher, C., Schumacher, J., Buchmann, N., et al., 2009. Plant species richness and functional composition drive overyielding in a six-year grassland experiment. Ecology 90(12), 3290-3302.

Meraner, M., Finger, R., 2018. Risk perceptions, preferences and management strategies: Evidence from a case study using German livestock farmers. J. Risk Res. 22(1), 110-135.

Meuwissen, M, Balmann, A, Spiegel, A, Ciechomska, A, Kopainsky, B, Soriano, B, et al., 2019. A framework to assess the resilience of farming systems. Agric. Syst. in Press.

Moulton, B. R., 1986. Random group effects and the precision of regression estimates. Journal of Econometrics 32(3), 385-397.

Potts, S. G., Woodcock, B. A., Roberts, S. P. M., Tscheulin, T., Pilgrim, E. S., Brown, V. K.,Tallowin, J. R., 2009. Enhancing pollinator biodiversity in intensive grasslands. J. Appl. Ecol. 46(2), 369-379.

Pratt, J. W., 1964. Risk aversion in the small and in the large. Econometrica 32, 122-136.

Sala, O. E., Paruelo, J. M., 1997. Ecosystems services in grasslands, in: Daily, G. C. (Eds.), Nature's Services: Societal Dependence on Natural Ecosystems. Island Press, Washington DC, USA.

Schläpfer, F., Tucker, M., Seidl, I., 2002. Returns from hay cultivation in fertilized low diversity and non-fertilized high diversity grassland. Environ. Resour. Econ. 21(1), 89-100.

Schmid, B., Hector, A., Saha, P., Loreau, M., 2008. Biodiversity effects and transgressive overyielding. J. Plant Ecol. 1(2), 95-102.

Sturludóttir, E., Brophy, C., Bélanger, G., Gustavsson, A. M., Jørgensen, M., Lunnan, T., Helgadóttir, Á., 2014. Benefits of mixing grasses and legumes for herbage yield and nutritive value in Northern Europe and Canada. Grass Forage Sci. 69(2), 229-240.

Suter, M., Connolly, J., Finn, J. A., Loges, R., Kirwan, L., Sebastià, M. T., Lüscher, A., 2015. Nitrogen yield advantage from grass–legume mixtures is robust over a wide range of legume proportions and environmental conditions. Glob. Chang. Biol. 21(6), 2424-2438.

Suter, M., Hofer, D., Lüscher, A., 2017. Weed suppression enhanced by increasing functional trait dispersion and resource capture in forage ley mixtures. Agric. Ecosyst. Environ. 240, 329-339.

R Core Team, 2018. R: A language and environment for statistical computing.

Tilman, D., Wedin D., Knops, J., 1996. Productivity and sustainability influenced by biodiversity in grassland ecosystems. Nature 379(6567), 718.

Vroege, W., Dalhaus, T., Finger, R., 2019. Index insurances for grasslands – A review for Europe and North-America. Agric. Syst. 168, 101-111.

Walker, K. J., Stevens, P. A., Stevens, D. P., Mountford, J. O., Manchester, S. J., Pywell, R. F., 2004. The restoration and re-creation of species-rich lowland grassland on land formerly managed for intensive agriculture in the UK. Biol. Conserv. 119(1), 1-18.

Wang, Y., Cadotte, M. W., Chen, Y., Fraser, L. H., Zhang, Y., Huang, F., et al., 2019. Global evidence of positive biodiversity effects on spatial ecosystem stability in natural grasslands. Nat. Commun. 10, 3207.White, T. A., Barker, D. J., Moore, K. J., 2004. Vegetation diversity, growth, quality and decomposition in managed grasslands. Agric. Ecosyst. Environ. 101(1), 73-84.

Wooldridge, J. M., 2003. Cluster-sample methods in applied econometrics. Am. Econ. Rev. 93(2), 133-138.

Yachi, S., Loreau, M., 1999. Biodiversity and ecosystem productivity in a fluctuating environment: the insurance hypothesis. PNAS 96(4), 1463-1468.

8. Online appendix

8.1 Tables

Table A.1: Site information

Mixture type	Site indication	Location	Period of the experiment	Years of yield forage quality data	Years of the experiment	Cuts per year
Mid-European[1]	10	Germany_a	2004-2006	1	2	5
	11	Germany_b	2005-2006	2	1-2	4/4
	15	Ireland_a	2004-2006	2	1,3	5/5
	18	Lithuania_a	2003-2005	3	1-3	3/3/3
	20	Lithuania_c	2004-2006	3	1-3	3/3/2
	22	Norway_a	2004-2006	2	1-2	3/3
	24	Norway_c	2003-2005	2	1-2	3/3
	26	Poland_a	2004-2006	3	1-3	4/4/3
	27	Poland_b	2005-2006	2	1-2	4/3
	35	Wales_a	2003-2006	2	2-3	4/4
	36	Wales_b	2004-2006	1	1	4
North-European[2]	13	Iceland_a	2003-2005	3	1-3	2/2/2
	14	Iceland_b	2004-2006	3	1-3	2/2/2
	23	Norway_b	2003-2006	1	1	2
Other[3]	1	Belgium	2003-2005	1	1	4
	45	Ireland_d	2010-2011	1	1	5

Classification of clusters and site indication are based on Kirwan et al. (2014). [1]The Mid-European mixture type includes *Lolium perenne*, *Dactylis glomerata*, *Trifolium pratense* and *Trifolium repens*. [2]The North-European mixture type includes *Phleum pratense*, *Poa pratensis*, *Trifolium pratense* and *Trifolium repens*. [3]The other mixture type includes *Lolium perenne*, *Phleum pratense*, *Trifolium pratense* and *Trifolium repens*.

Table A.2: Estimation results of expected outcome and variance of expected outcome (Equations 11 and 13). Results include expected outcome and variance of dry matter yields (DM yields), milk production potentials per kg of DM yield (DM MPP) and milk production potential yields (MPP yields).

	DM yields	Var (DM yields)	DM MPP	Var(DM MPP)	MPP yields	Var (MPP yields)
Intercept	13105.82 (69.83)***	6547610.3 (13.34)***	2.11 (393.56)***	0.004 (1.24)	27514.77 (70.72)***	30662372 (12.69)***
Simpson index$^{0.5}$ ($D^{0.5}$)	2599 (6.24)***	-2791192.6 (-2.41)**	-0.02 (-1.31)	-0.011 (-1.16)	5423.78 (6.32)***	-11349391 (-1.95)*
Sown density (Density; (0=low, 1=high)	190.42 (3.19)***	219783.45 (0.88)	-0.01 (-0.7)	0.007 (1.58)	375.45 (2.92)**	757824.46 (0.69)
Year of the experiment (Year)						
2	-836.29 (-0.86)	-1731787.7 (-1.13)	-0.02 (-0.92)	-0.001 (-0.5)	-2055.06 (-0.98)	-8805375.7 (-1.24)
3	-2274.51 (-1.73)	-1784775 (-1.01)	-0.06 (-3.62)***	0.01 (0.86)	-5205.87 (-1.85)*	-7292065.8 (-0.92)
Site (Site)						
10	-1170.28 (-1.24)	-2266710.2 (-1.36)	0.17 (8.52)***	0.002 (1.21)	-310.36 (-0.15)	-11180737 (-1.43)
11	-1492.92 (-3.28)***	-1974259.1 (-2.18)**	0.21 (21.11)***	0.003 (1.91)*	-454.79 (-0.47)	-7660550.8 (-1.78)*
13	-9183.59 (-13.95)***	-3240717.4 (-3.88)***	-0.06 (-5.39)***	0.018 (4.72)***	-19405.67 (-13.72)***	-16064889 (-4.19)***
14	-10096.03 (-15.34)***	247982.79 (0.3)	0.04 (3.35)***	0 (0.04)	-20941.05 (-14.81)***	-486319.72 (-0.13)
15	-398.03 (-0.59)	299600.27 (0.33)	0.09 (10.39)***	-0.005 (-0.82)	51.85 (0.04)	-2268889 (-0.55)
18	-6697.65 (-10.17)***	1026133.2 (1.23)	-0.01 (-1.12)	-0.002 (-0.51)	-14073.46 (-9.95)***	2402855.4 (0.63)
20	-11248.17 (-3.14)***	-5464436.8 (-1.03)	0.2 (5.55)***	-0.037 (-1.08)	-22162.88 (-2.89)**	-24382916 (-1.03)
22	-4031.78 (-8.87)***	3900508.5 (4.31)***	0.1 (9.84)***	0.002 (1.04)	-7500.47 (-7.7)***	17429519 (4.05)***
23	-4452.89 (-59.71)***	-2037647.9 (-9.84)***	-0.05 (-18.58)***	0.001 (0.54)	-9919.71 (-64.57)***	-12766270 (-12.27)***
24	-2931.08 (-5.82)***	-2209741 (-2.81)**	-0.02 (-1.4)	0.002 (2.88)**	-6330.62 (-5.85)***	-11749507 (-3.25)***
26	-9023.92 (-1.91)*	-8038533.9 (-1.13)	0.27 (5.98)***	-0.039 (-0.85)	-17877.46 (-1.77)*	-39619601 (-1.25)
27	-25572.13 (-6.6)***	-2411366.8 (-0.38)	-0.11 (-1.28)	0.183 (32.9)***	-52716.8 (-6.34)***	-2259037.1 (-0.08)
35	-3646.97 (-3.69)***	-702928.68 (-0.56)	0.12 (6.52)***	-0.004 (-0.61)	-6627.04 (-3.12)***	-3873599 (-0.67)
36	-5833.31 (-2.9e+14) ***	-4163286.7 (-3.7e+14) ***	0.07 (1.9e+13) ***	-0.001 (-2.7e+12) ***	-11669.97 (-1.0e+14) ***	-19730794 (-5.3e+13) ***

45	-5010.18 (-130.07)***	-4542317.3 (-25.63)***	0.11 (21.73)***	-0.007 (-2.53)**	-9517.8 (-120.86)***	-20847751 (-24.16)***
Site × Cuts per year (Site×Cuts)						
20	1910.78 (1.64)	1301170.7 (0.71)	-0.07 (-6.75)***	0.013 (1.11)	3480.83 (1.4)	4680635.1 (0.58)
26	928.73 (0.8)	1949525.5 (1.07)	-0.04 (-4.02)***	0.01 (0.89)	1893.56 (0.76)	9008473.3 (1.11)
27	5853.8 (5.99)***	183860.65 (0.12)	0.02 (1.05)	-0.045 (-33.84)***	11953.31 (5.69)***	-2342880.2 (-0.33)
Number of observations	698	698	698	698	698	698
Adjusted R^2	0.766	0.157	0.465	0.025	0.787	0.161

Table includes DM yields, DM MPP and MPP yields. Numbers in parentheses are t-values. * p < 0.1; ** p < 0.05; *** p < 0.01. T-values are corrected for clustered error terms and heteroscedasticity.

Table A.3: Estimation results for expected outcome and variance of expected outcome (Equations 11 and 13). Results include expected outcome and variance of metabolizable energy per kg of DM yield (DM ME) and metabolizable energy yields (ME yields).

	DM ME	Var(DM ME)	ME yields	Var(ME yields)
Intercept	10.87 (471.29)***	0.05 (0.55)	142103.81 (70.76)***	799100000 (12.81)***
Simpson index$^{0.5}$ ($D^{0.5}$)	-0.06 (-0.94)	-0.23 (-0.92)	28153.8 (6.35)***	-306000000 (-2.05)*
Sown density (Density; (0=low, 1=high)	-0.03 (-0.79)	0.19 (1.56)	1924.59 (2.93)**	20946406 (0.72)
Year of the experiment (Year)				
2	-0.07 (-0.82)	-0.01 (-0.3)	-10240.97 (-0.95)	-226600000 (-1.22)
3	-0.26 (-3.66)***	0.3 (0.97)	-26418.83 (-1.83)*	-196100000 (-0.93)
Site (Site)				
10	0.67 (7.82)***	0.05 (1.01)	-4343.41 (-0.42)	-287800000 (-1.41)
11	0.83 (19.73)***	0.07 (1.52)	-5551.3 (-1.11)	-210500000 (-1.88)*
13	-0.27 (-5.7)***	0.47 (4.46)***	-100169.49 (-13.83)***	-410500000 (-4.07)***
14	0.2 (4.17)***	-0.04 (-0.41)	-108356.38 (-14.96)***	3479964.6 (0.03)
15	0.38 (10.64)***	-0.14 (-0.85)	-504.1 (-0.07)	-39965016 (-0.37)
18	-0.05 (-1.14)	-0.07 (-0.68)	-72676.81 (-10.03)***	86094466 (0.85)
20	0.8 (4.79)***	-1.03 (-1.1)	-116339.38 (-2.96)***	-649900000 (-1.03)
22	0.39 (9.2)***	0.04 (1.02)	-39921.26 (-7.99)***	464300000 (4.14)***
23	-0.15 (-12.57)***	0.03 (0.67)	-50198.55 (-63.23)***	-315600000 (-11.82)***
24	-0.07 (-1.46)	0.04 (2.56)**	-32534.36 (-5.86)***	-297700000 (-3.13)***
26	1.19 (5.47)***	-1.19 (-0.95)	-93218.01 (-1.8)*	-1017000000 (-1.2)
27	-0.76 (-2.16)**	4.92 (39.83)***	-275139.73 (-6.46)***	-99873246 (-0.13)
35	0.49 (6.8)***	-0.12 (-0.76)	-35261.56 (-3.25)***	-95900942 (-0.63)
36	0.3	-0.01	-60707.94	-510900000

	(1.3e+14)***	(-1.1e+12)***	(-3.7e+13)***	(-1.8e+14)***
	0.48	-0.16	-49891.05	-542200000
45	(19.76)***	(-2.02)*	(-122.69)***	(-24.3)***
Site × Cuts per year (Site×Cuts)				
	-0.29	0.35	18702.6	134500000
20	(-5.32)***	(1.12)	(1.46)	(0.62)
	-0.2	0.31	9756.04	234500000
26	(-3.7)***	(0.99)	(0.76)	(1.08)
	0.17	-1.22	62544.38	-44962444
27	(1.91)*	(-42.07)***	(5.81)***	(-0.24)
Number of observations	698	698	698	698
Adjusted R^2	0.398	0.023	0.783	0.16

Numbers in parentheses are t-values. * p < 0.1; ** p < 0.05; *** p < 0.01. T-values are corrected for clustered error terms and heteroscedasticity.

8.2 Figures

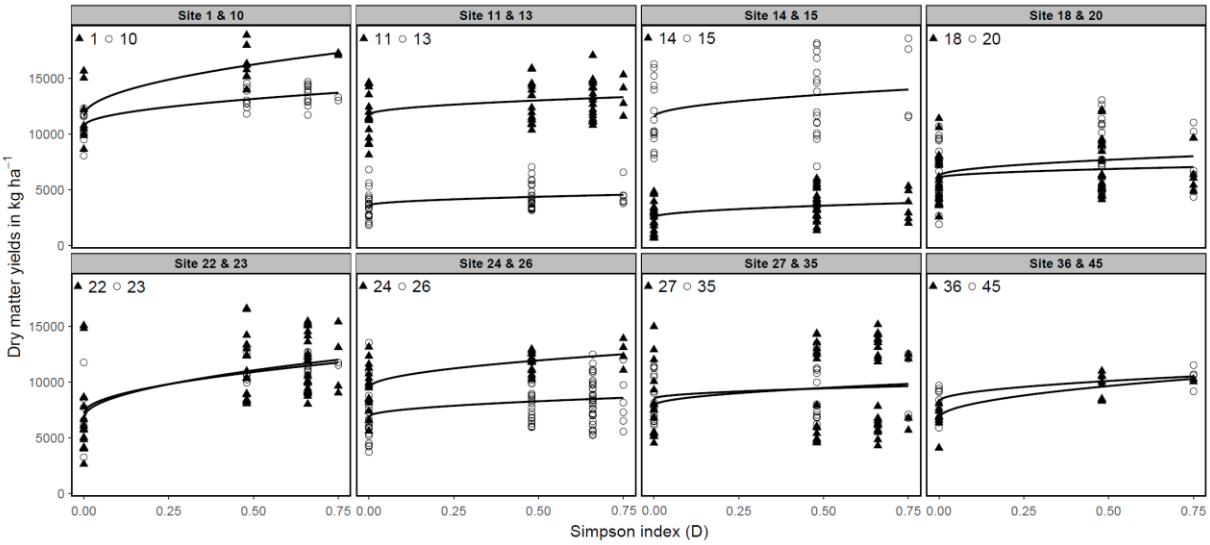

Figure A.1: Dry matter yields in kg ha^{-1} per site. Note for site 10 that forage quality was only measured for the first four out of five cuts each year. These observations (of the four cuts) are included as this is consistent for all plots and the first cuts are agronomically more important than later ones.

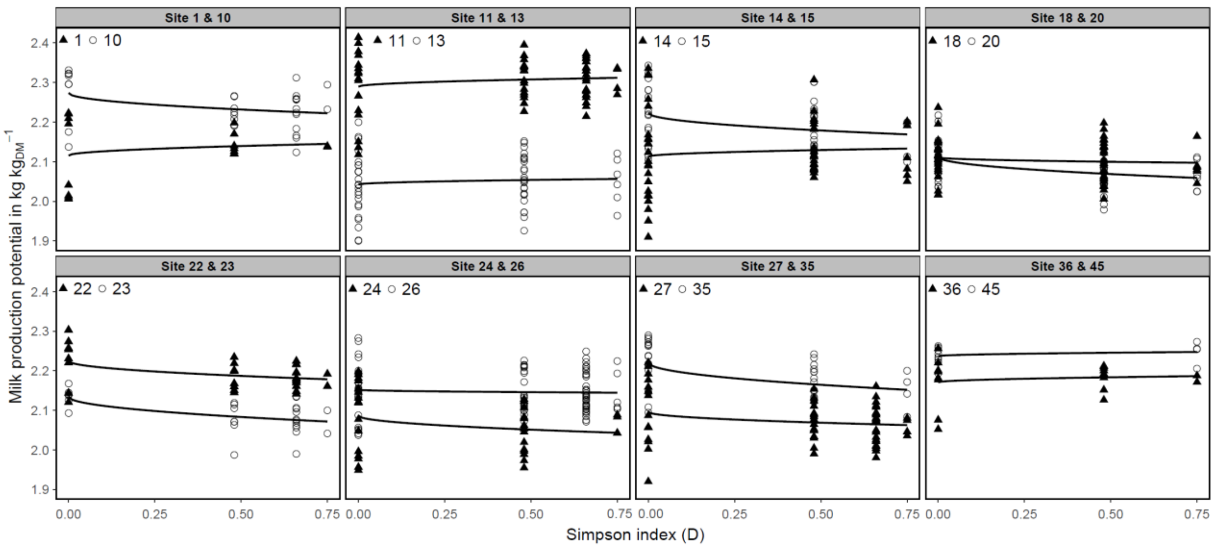

Figure A.2: Milk production potential in kg kg_{DM}^{-1} per site. Note for site 10 that forage quality was only measured for the first four out of five cuts each year. These observations (of the four cuts) are included as this is consistent for all plots and the first cuts are agronomically more important than later ones.

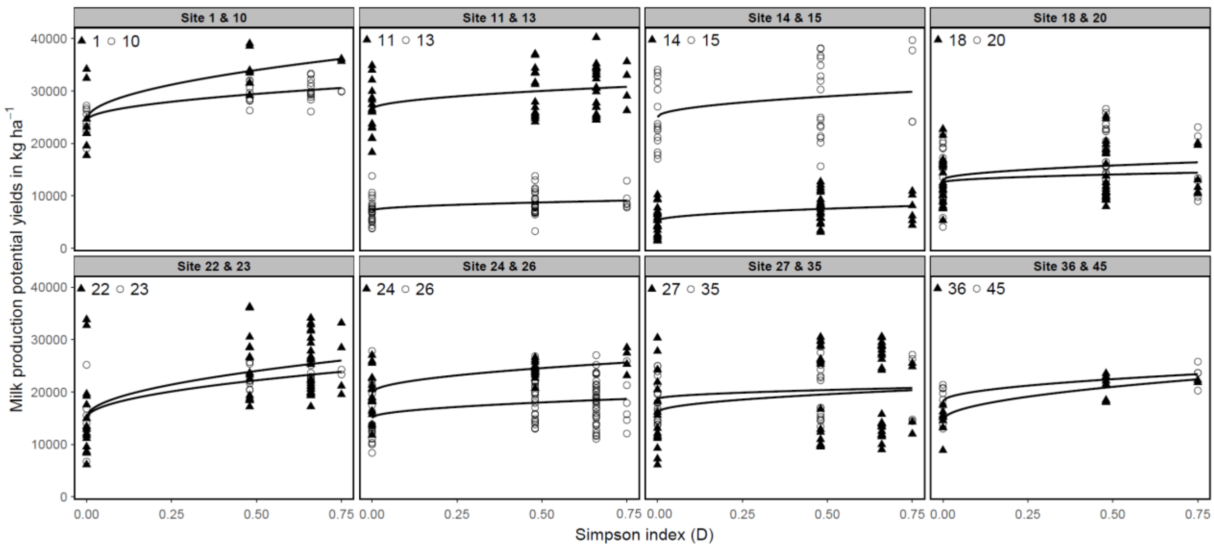

Figure A.3: Milk production potential yields in kg ha^{-1} per site. Note for site 10 that forage quality was only measured for the first four out of five cuts each year. These observations (of the four cuts) are included as this is consistent for all plots and the first cuts are agronomically more important than later ones.