Quantity based indicators fail to identify extreme pesticide risks

Journal Article

Author(s):
Möhring, Niklas; Gaba, Sabrina; Finger, Robert

Publication date:
2019-01-01

Permanent link:
https://doi.org/10.3929/ethz-b-000288799

Originally published in:
Science of The Total Environment 646, https://doi.org/10.1016/j.scitotenv.2018.07.287

Quantity based indicators fail to identify extreme pesticide risks

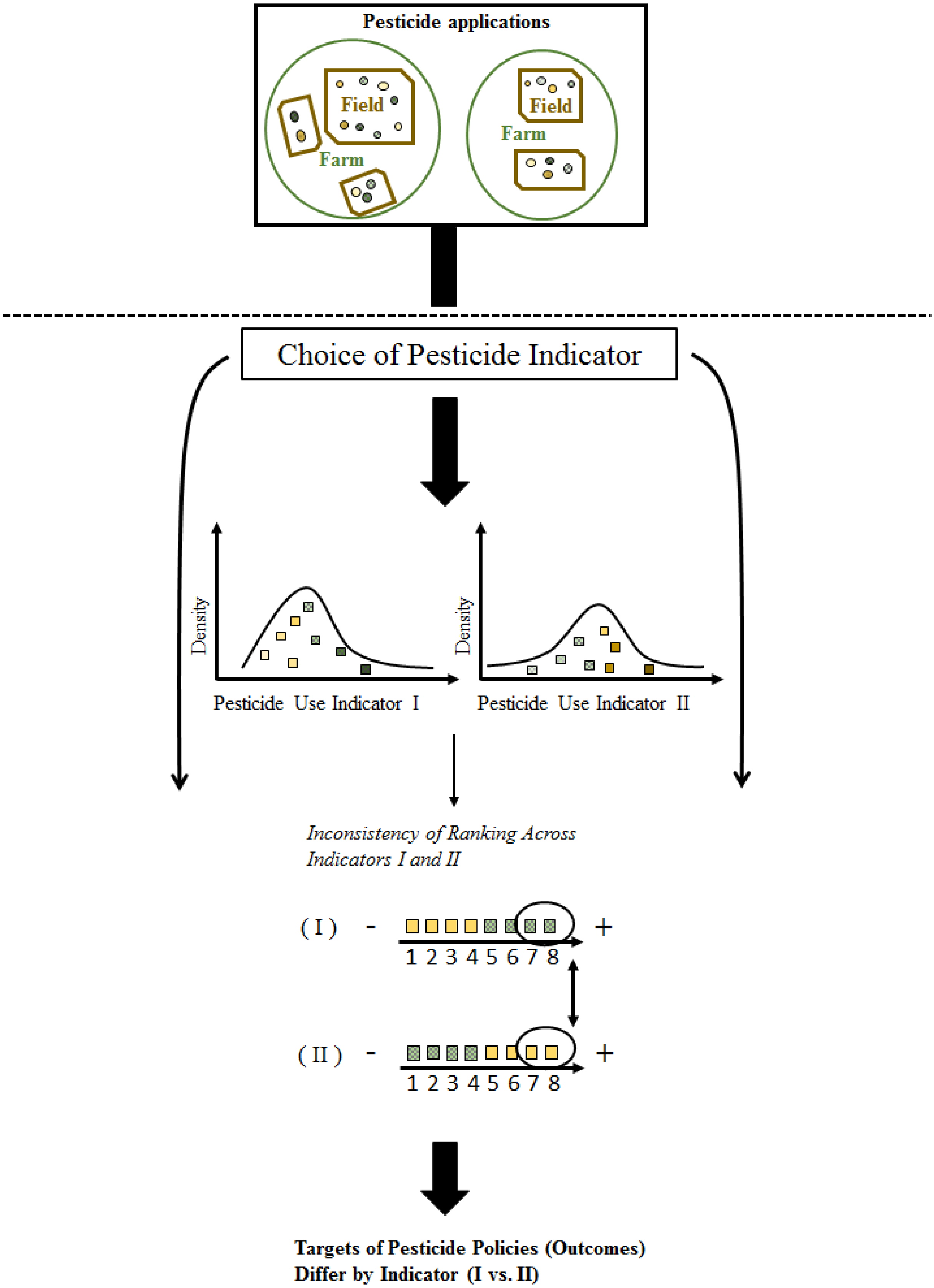
Authors: Niklas Möhring[a*], Sabrina Gaba[b], Robert Finger[a]

[a] Agricultural Economics and Policy Group, ETH Zürich

[b] USC 1339, Centre d'Etudes Biologiques de Chizé, INRA, 79360 Villiers en Bois, France

*Corresponding author: nmoehring@ethz.ch

Graphical Abstract



Abstract.

As a matter of policy, minimizing human health and environmental risks associated with pesticide use is a major challenge but necessary for improving agricultural sustainability. Efficient and effective policies that encourage the use of less risky pesticides, such as pesticide taxes, necessitate a precise and realistic quantification of potential adverse effects. Various indicators are currently utilized in policies and they focus mainly on a purely quantitative dimension of the pesticides used, which can lead potentially to unfavorable outcomes of pesticide policies. A unique dataset applied to pesticide use by Swiss farmers in winter wheat and potato production, demonstrates that on average the two most important quantitative indicators show a significant correlation with pesticide risks as expressed by the Danish Load Indicator. However, they have almost no explanatory power for extreme risks (e.g. most intensive use patterns for pesticides with unfavorable toxicity profiles). Results remain stable over a range of aggregation levels, from application- to farm-level indicators of pesticide use. These findings render the commonly used, quantitative indicators ineffective to reduce potential environmental and human health risks of pesticides and, in the worst case, lead to misinformed market-based pesticide policies consequential to National Action Plans.

Keywords: Pesticides, Pesticide risks, Pesticide policies, Pesticide indicators

1. Introduction

Current agricultural production systems often rely on an intensive use of pesticides and other agrochemical inputs. Pesticides are tightly regulated in many countries, subject to rigorous testing and highly conservative risk assessment paradigms. However, the use of pesticides can still present potential risks to human and environmental health (Strange and Scott 2005; Damalas 2009; Beketov et al. 2013; Malaj et al. 2014; Stehle and Schulz 2015). The introduction of effective policies to reduce adverse effects of pesticides, while maintaining production levels, is a major challenge on the way to achieving improved sustainability in agriculture (Tilman et al. 2002). In the European Union, the US and China, public policies have been established to address pesticide risks and stricter pesticide policies are also being implemented in several other countries (Lefebvre et al. 2015; Pimentel and Burgess 2014; Osteen and Fernandez-Cornejo 2013; Zhang and Wen 2008; Sun et al. 2012, MAAF 2015, Bundesrat 2017, Böcker et al., 2018). However, the effectiveness of current policies has been questioned recently (e.g. Hossard et al. 2017, Finger, 2018). The focus of this research was to evaluate pesticide use/risk indicators utilized for the purposes of informing market-based pesticide policy and high-level reduction targets related to National Action Plans.

Setting up policies which promote a reduction in the impacts of pesticides on the environment and human health is far from straightforward. Pesticides are highly heterogeneous with respect to properties, application regimes and their potential impact on the environment and human health. For instance, in the EU alone, a range of 494 active substances for pesticides, with potentially different adverse effects, are currently authorized (EU 2017). The choice of suitable pesticide indicators to quantify pesticide use is therefore essential to define efficient and effective policy measures.

Currently, implemented indicators differ significantly. For example, there are simple, purely quantitative indicators like the Quantity of Active Ingredients (QA) and the Treatment Frequency Index (TFI), which abstract from inherent pesticide properties, to very detailed, risk-adjusted indicators such as the Load Index (LI). QA is a simple measure of kilograms of pesticides used per area. TFI measures the intensity of applications, i.e. quantity applied per unit of cropped area in relation to the recommended dosage (Coll and Wajnberg 2017). The LI indicator accounts for application intensity as well as a broad range of potential environmental and health risks for each pesticide (Miljøministeriet 2012; Kudsk et al. 2018). The indicator chosen differs across countries and institutions. For example, France uses the QA and TFI indicators to set targets for pesticide policies (MAA 2017). Furthermore, QA and TFI are applied as key indicators for pesticide use statistics by institutions worldwide (Eurostat 2017; USDA 2017; JKI 2017; MAA 2017) and are standard indicators for studies on the economics of pesticide use (Ghimire and Woodward 2013; Hossard et al. 2014; Gaba et al. 2016; Perry et al. 2016; Lechenet et al. 2017). The Danish Load Index (LI) is currently the only risk-based indicator implemented in European pesticide policies which holistically assesses potential environmental and health risks of pesticides on a product level. As with the purely quantitative QA and TFI indicators, this allows pesticide risks to be upscaled along a gradient of temporal and spatial resolution (Kudsk et al., 2018). Since 2013, it has been used in Denmark for the assessment of policy targets and at the same time as the basis for pesticide taxation (Böcker and Finger 2016; Kudsk et al. 2018).

However, it is hypothesized that large quantities of pesticides, as indicated by high QA or TFI indicator values, may not inherently mean higher risks for human health and the environment (e.g. high LI indicator values). For example, Kniss (2017) recently showed that herbicide use trends for major crops in the US were reversed when the assessment was switched from quantity-based to toxicity-based indicators. More importantly, the use of quantity-based indicators compared to risk-adjusted indicators may lead to major shifts in policy targets if indicators rank pesticide use inconsistently. These policy targets include, for example, the reduction of temporal or spatial “hotspots” and extreme application regimes (over a given cropping season), as extreme applications are major contributors to the negative effects of pesticides on human health (Larsen et al. 2017) and the environment (Releya and Hovermann 2006; Gordon et al. 2012; Bundschuh et al. 2013, Topping and Elmeros 2016). Along these lines, Larsen et al. (2017) conclude that there is a need for the implementation of pesticide policies that tackle extreme applications. Market-based policy measures such as specific taxes, quotas or subsidies can complement regulatory frameworks and admission procedures in achieving this target (Baumol and Oates, 1988). However, a misspecification of policy targets may result in biased policy incentives and finally, adverse policy outcomes. Depending on the degree of inconsistencies between indicators, indicator choice may therefore have severe implications for policy outcomes. Although a well-informed policy discussion is of vital importance, there is a lack of studies which quantify the extent of inconsistencies between pesticide use indicators in a common, robust framework.

This research gap was addressed by investigating the consistency of pesticide use rankings between the purely quantitative, but widely used QA and TFI indicators and the LI indicator. The focus goes beyond average consistency by analyzing consistency for “extreme” application regimes as well as temporal and spatial “hotspots”. Throughout the article, we refer to extreme applications as the most intensive use patterns and highest risk scenarios and profiles compared to all other applications, i.e. the upper tails of the observed distributions of pesticide applications. In the analysis, across-indicator consistency was tested using a unique panel dataset of pesticide applications in real farming conditions in Switzerland for two major crops, potatoes and winter wheat. These crops were chosen because potatoes are characterized by the highest average pesticide use and winter wheat is the most abundant crop in European (and Swiss) arable crop production. Pesticide application patterns of farmers, including choices made regarding the products used, their concentration, spatial distribution of application, and their timing, are strongly heterogeneous across farmers. A comparison of indicators on “real” application data was therefore necessary to derive meaningful policy recommendations, especially regarding risks from the most intensive pesticide use patterns. The analysis started off by comparing the structure of the indicators. Then correlation coefficients were used to test the consistency of indicator rankings over the whole distribution. Secondly, copulas were used to analyze tail dependence between pesticide use indicators. This meant focus could be placed on across-indicator consistency for extremes, i.e. observations in the tails of the distribution. The detailed dataset allowed the robustness of results to be assessed on different aggregation levels and for different pesticide types (i.e. all pesticides, herbicides and fungicides).

Current pesticide policies aim to reduce negative environmental and health effects of pesticides. Market-based policy measures achieve this target by incentivizing a change in the farmers’ application behavior. The objective of this paper is to show that the choice of underlying pesticide indicators can crucially shift policy targets and incentives, and in the worst-case lead to adverse outcomes of pesticide policies. The analysis emphasizes that the comparison of indicators should not merely be based on their average fit. In fact, quantity-based indicators are found to be unsuitable to proxy high-risk situations.

2. Background

2.1. Current pesticide policies

Since 2012, European Union member states have to draw up National Action Plans (NAPs) for a “Sustainable use of pesticides” with the goal of reducing “risks and impacts of pesticide use on human health and the environment” (Directive 2009/128/EC). The revision of existing, and implementation of new pesticide policies is an ongoing process. In the EU, Directive 2009/128/EC demands that EU member states review NAPs every five years and the EU commission has recently announced a “REFIT” evaluation of all EU pesticide legislation (EC 2015). Furthermore, adverse effects from pesticide use are at the top of the policy agenda in other countries like Switzerland, the US or China (Bundesrat 2017, Pimentel and Burgess 2014; Osteen and Fernandez-Cornejo 2013; Zhang and Wen 2008; Sun et al. 2012). Policy measures within the framework of the NAPs include training and extension services, the promotion of Integrated Pest Management, inspection and renewal of equipment, information and awareness raising, as well as market-based policy measures like pesticide taxes (Böcker and Finger 2016). Market-based policies complement existing regulatory frameworks and pesticide admission procedures (e.g. see regulations EC No. 1107/2009 and EC No. 396/2005 in the EU and the Federal Insecticide, Fungicide and Rodenticide Act and the Federal Food, Drug and Cosmetic Act in the USA). Their mix provides an “optimal regulatory strategy” (Skevas et al. 2013, p.97) to reduce externalities from pesticide use on the environment and human health. The analysis in this article focuses on the incentives of market-based pesticide policies and the definition of pesticide reduction targets. See e.g. Storck et al. (2017), Coll and Wajnberg (2017) and EC SAM (2018) for a discussion of current pesticide regulations and admission procedures.

2.2. Utilized pesticide indicators

The design of pesticide policies, i.e. setting and reviewing targets and defining pesticide indicators, is complex due to the heterogeneity of available pesticide products, production environments and farming practices. Worldwide, many countries have reported national targets for the reduction of impacts from pesticide use. In Europe, the majority of EU member states as well as Switzerland have defined risk reduction targets - yet, nine EU countries also report pesticide use reduction targets. However, a clear definition of measurable targets is less common. For example, only five EU member states have so far set “high level, measurable reduction targets” according to EU DGHFS (2017). Out of these five member states, four have defined risk reduction targets (Belgium, Denmark, Greece and Germany) and one has defined a use reduction target (France). Risk indicators utilized in these member states include the POCER indicator in Belgium (Vercruysse and Steurbaut 2002, Claeys et al. 2005, Houbraken et al. 2016), the Load Indicator in Denmark (Kudsk et al. 2018), an environmental and human risk assessment in Greece (Tsaboula et al. 2016) and the SYNOPS indicator in Germany (Gutsche and Rossberg 1997, Strassemeyer et al. 2017). A detailed spatial mapping of pesticide use and associated risks, using pesticide reports, is for example already possible in California based on the CalPip database (http://calpip.cdpr.ca.gov/) or in Denmark (Kudsk et al. 2018). An extensive discussion of existing pesticide risk indicators does not lie within the scope of this article, but we refer to Labite et al. (2011), OECD (2016) and Yang et al. (2017) for an overview and a discussion. The analysis in this article is restricted to three indicators. More specifically, the widely used, quantity-based indicators QA and TFI are compared to the Danish, risk-based LI indicator. QA, TFI and LI are chosen because they are i) currently implemented in the framework of the NAP; ii) allow pesticide use to be scaled to different aggregation levels (product, application, farm, landscape); and iii) can be computed based on pesticide administration data. They are therefore policy-relevant, can be used as a foundation for market-based policy instruments (such as taxes) and can be easily implemented in other countries. Note that easy scalability and low data requirements are accompanied by the drawback that the indicators utilized do not consider the application context (e.g. timing, weather conditions, soil, and distance to bodies of water) and different risk components are aggregated (in the case of the LI indicator).

3. Methodology and Data

3.1 Indicator calculation

The pesticide use records are based on observations of single applications from various farmers. Farmers may grow the same crop on several fields of varying size. The observations include a wide range of different pesticides and span across different years (see Data, Section 3.5 below for a detailed overview). Thus, the data exhibit a high level of spatial and temporal heterogeneity of pesticide use. This also implies that pesticide use indicators cannot be simply transformed into each other without knowledge of field level information. Figure 1 gives an overview of the procedure used for computation of the QA, TFI and LI indicators.

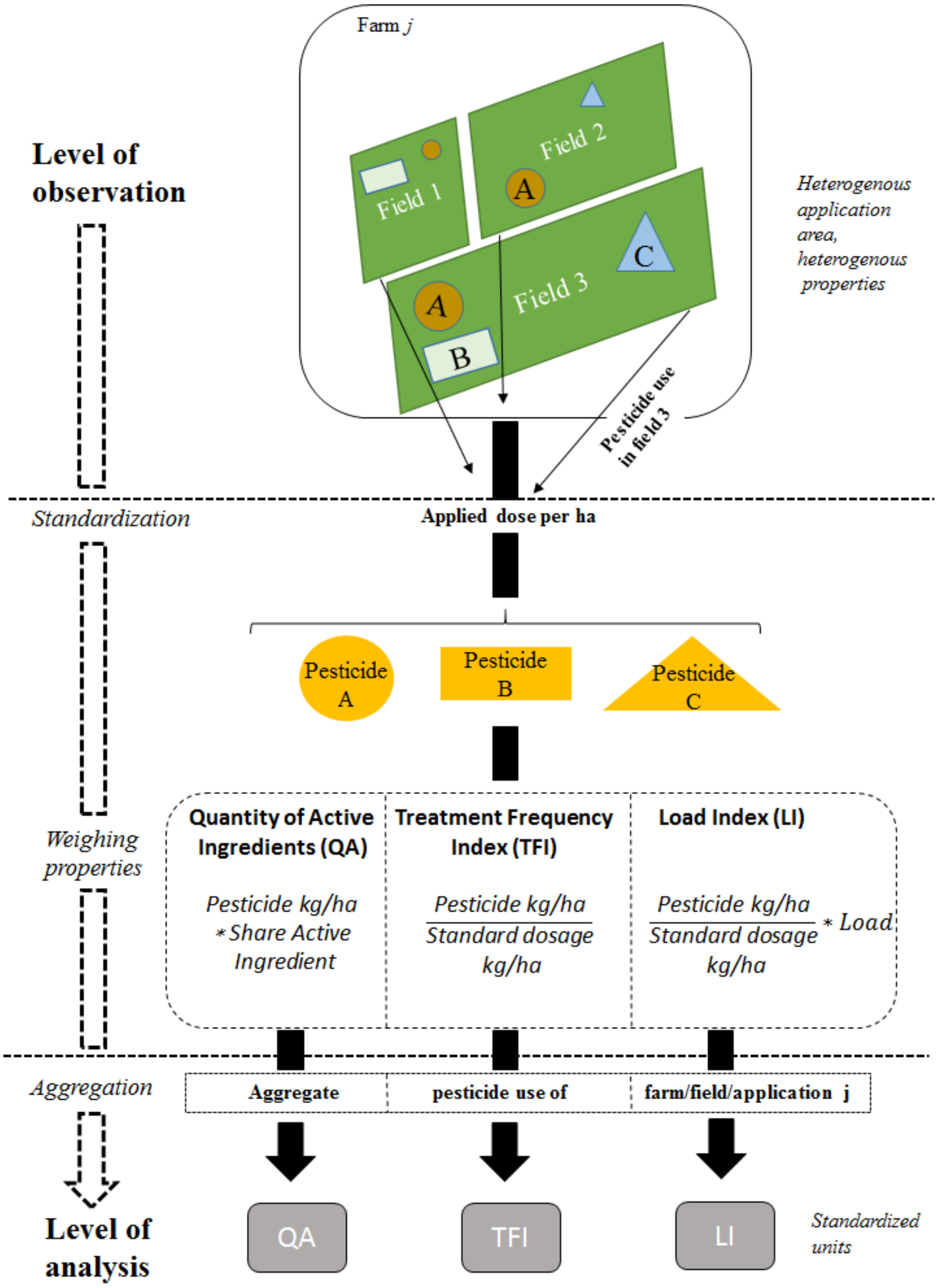


Fig. 1. Intermediate steps in the computation of the QA, TFI and LI indicators

The pesticide products applied (i.e. A, B and C) may contain one or several active ingredients in different concentrations. It is possible that some products contain similar, or identical active ingredients. Products are applied in different quantities and on different fields (or parts of fields). The computation procedure for the QA, TFI and LI indicators is described in detail below. All indicators are computed on different aggregation levels.

Firstly, all applications were standardized to obtain per ha measures of pesticide use per product.

$$(1)\ \ Pesticide\ Use_{ijk}\ per\ ha = \sum_{l=1}^{n} Pesticide\ Use_{ijkl} in\ kg * \frac{application\ surface\ in\ ha}{surface\ field\ in\ ha}$$

where, for a given crop and year (indicators omitted for clarity), i denotes the farm, j the field, k the pesticide product and l a single application of the product.

The Quantity of Active Ingredient (QA) indicator is a simple indicator of the quantity of active substances used per ha (Eurostat 2017; USDA 2017). It was first computed on a product level:

$$(2)\ \ QA_{ijk} = Pesticide\ Use_{ijk} * concentration\ of\ active\ ingredient_k$$

where, for a given crop and year, i denotes the farm, j the field and k the pesticide product.

Field-level pesticide use was computed in a second step. Finally, farm-level pesticide use was computed as the area weighed mean over all fields per farm, crop and year.

$$(3)\ \ QA_{ij} = \sum_{k=1}^{m} QA_{ijk}$$

$$(4)\ \ QA_i = \sum_{j=1}^{n} QA_{ij} * \frac{surface_j}{surface_i}$$

where, for a given crop and year, i denotes the farm, j the field, k the pesticide product and *n* the number of fields per farmer, crop and year.

The Treatment Frequency Index (TFI) indicator is an indicator of pesticide use intensity (Coll and Wajnberg 2017). It was first computed on a product level:

$$(5)\ \ TFI_{ijk} = \frac{Pesticide\ Use_{ijk}}{Recommended\ (standard)\ dosage\ for\ product\ k\ per\ ha\ in\ the\ given\ culture}$$

where, for a given crop and year, i denotes the farm, j the field and k the pesticide product. If a range of possible dosages was indicated the maximum dosage was taken, respectively.

Field-level pesticide use was computed in a second step. Finally, farm-level pesticide use was computed as the weighed mean over all fields per farmer, crop and year.

$$(6)\ \ TFI_{ij} = \sum_{k=1}^{m} TFI_{ijk}$$

$$(7)\ \ TFI_i = \sum_{j=1}^{n} TFI_{ij} * \frac{surface_j}{surface_i}$$

where, for a given crop and year, i denotes the farm, j the field, k the pesticide product and *n* the number of fields per farmer, crop and year.

The Danish Load Index (LI) indicator (Miljøministeriet 2012; Kudsk et al. 2018) is an indicator which takes into account health, fate and toxicity properties of pesticides. It was first computed on a product level:

$$(8)\ \ LI_{ijk} = \frac{Pesticide\ Use_{ijk}}{Recommended\ (standard)\ dosage\ of\ product\ k\ per\ ha\ in\ the\ given\ culture} * Load_k$$

where, for a given crop and year, i denotes the farm, j the field and k the pesticide product. $Load_k$ denotes the potential, product specific environmental and health impacts of product *k*. More precisely, Load is the sum of the Human Health Load, Environmental Fate Load and Environmental Toxicity Load. The Human Health Load is based on Hazard and Precautionary Statements; the Environmental Fate Load on half-time in soil, half-time in water, bio-concentration factors and the SCI-GROW (Screening Concentration in Ground Water) index; the Environmental Toxicity Load on long- term toxicity for fish, daphnia and earthworms and shortterm toxicity for birds, mammals, daphnia, algae, aquatic plants, earthworms and bees (see e.g. Kudks et al., 2018). Human Health Load, Environmental Fate Load and Environmental Toxicity Load was first computed separately. More specifically, each product was compared to all other products for each sub-indicator. The relative scores obtained for all sub-indicators were then summed up in the respective categories Human Health, Environmental Fate and Environmental Toxicity. Thus, a single value per product was obtained for each of the three Load categories. Load values in the three categories were finally added up to a total Load, using equal weights for each category (sub-indicator scores were weighted according to the scheme in Miljøministeriet 2012).

Field-level pesticide use was computed in a second step. Finally, farm-level LI of pesticide use was computed as the weighed mean over all fields per farmer, crop and year:

(9) $LI_{ij} = \sum_{k=1}^{m} LI_{ijk}$

(10) $LI_i = \sum_{j=1}^{n} LI_{ij} * \frac{surface_j}{surface_i}$

where, for a given crop and year, i denotes the farm, j the field, k the pesticide product and *n* the number of fields per farmer, crop and year.

3.2 Confidence intervals and correlation coefficients

Confidence intervals for the respective moments were computed to assess the statistical significance of differences in the distribution moments. It was then checked to see if confidence intervals overlap. If confidence intervals do not overlap, the hypotheses of significantly different parameters can be accepted (given the chosen error rate). In this way it can, for example, be shown that two distributions have significantly different means, variances or skewness. Payton et al. (2003) show that 85% of the confidence intervals should be used to check for overlaps, given a maximum error rate of 5% to falsely reject the null hypothesis. To compute confidence intervals, 10 000 non-parametric bootstrap replications were drawn from the empirical distribution of the data, from which a bootstrap cumulative distribution function and confidence intervals were then calculated (e.g. Davison and Hinkley 1997).

Correlation coefficients were computed first to assess the dependence between indicators. As dependence between indicators was highly non-linear, a measure of ordinal dependence was chosen. Kendall's tau was selected due to its direct connection to the concept of copulas (see Section 3.3) and its better efficiency properties and robustness compared to Spearman's rho (Croux and Dehon 2010). Given two sets of observations of the joint random variables X and Y, Kendall's tau is composed of the difference between concordant and disconcordant pairs of values x_i and y_i , i.e. pairs where rankings "agree" or "disagree". See Croux and Dehon (2010) for a formal definition.

3.3 Copulas theoretical background

(Bivariate) copulas are functions that allow (two) distribution functions to be "coupled" to a multivariate distribution function. Formally, consider the two (continuous) distribution functions $F(x) = P[X \leq x]$ and $G(y) = P[Y \leq y]$ of the random variables X and Y. Then the joint distribution function $H(x, y)$ can be expressed by the copula function $C(.)$ as follows:

(11) $H(x, y) = C[F(x), G(y)]$

where the copula function $C(.)$ captures the dependence structure between the two distribution functions. Empirical applications of the copula concept can be found in various disciplines such as finance and insurance, as well as in applied agriculture. Copulas can also be used to model joint occurrence of unfavorable weather conditions for agricultural production. These models then serve to improve insurances against weather shocks, where the tail of the loss distribution is particularly relevant (Xu et al. 2010, Okhrin et al. 2013). Copulas are linked directly to concordant measures of dependence, such as Kendall's tau (Nelsen 2006). Indeed, such measures of dependence are functions of the copula value (see Nelsen (2006) for an overview). This can be illustrated for Kendall's tau as follows (where $F(x) = u$ and $G(y) = v$):

$$(12)\ \tau = 4 * E[C(u, v)] - 1$$

The upper and lower tail dependence coefficients describe dependence in the upper and lower tails of bivariate distributions. Given the copula function defined in equation (11), the upper and lower tail dependence coefficients (TDCs), λ_U and λ_L , are defined as:

$$(13)\ \lambda_U = 2 - \lim_{t \to 1} \frac{1 - C(t,t)}{1 - t} \text{ and}$$

$$(14)\ \lambda_L = \lim_{t \to o} \frac{C(t,t)}{t}$$

where λ_U and λ_L lie in the range of [0,1] and t denotes the $100t$-th percentile of F and G, respectively. Therefore, TDCs depend solely on the copula of the (continuous) random variables and "theoretical" TDCs can be computed directly from copula parameters (for an overview see Joe (1997) and Nelsen (2006)). The values obtained for TDCs can be interpreted in such a way that larger values indicate a higher probability of joint extremes in the respective upper or lower tail than lower values of the respective TDC. A TDC of zero implies asymptotic (tail) independence. In the context of this article, a larger "upper TDC" implied a higher probability of jointly capturing major application events with the two indicators compared, while a smaller "upper TDC" signified that indicator choice had a large influence on the analysis of pesticide applications, as major application events were not captured jointly. Differences in upper and lower tail dependence further indicated an asymmetric dependence structure.

3.4 Copula selection and estimation procedure

The following estimation procedure and robustness checks were applied in this article for the estimation of copulas and TDCs[1]:

(i) *Test independence of indicator pairs analytically with independence tests*. Before the dependence structure of indicator pairs was analyzed, independence of all indicator pairs was tested analytically with a bivariate, asymptotic independence test based on Kendall´s tau (Genest and Favre 2007). Independence was rejected for all indicator pairs, which led to the next step, the estimation of copulas and TDCs as indicated above.

[1] The described estimation and testing procedure was implemented using the R statistical software (R Core Team 2013); especially the VineCopula (Schepsmeier et al. 2017), CopBasic (Asquith 2017) and gofCopula (Okhrin et al. 2016) packages. All R-codes used for the implementation are documented in the online Appendix to provide full transparency of the procedure.

ii) *Estimate copula parameters and TDCs for a wide range of copula families with a semi-parametric estimation procedure with empirical margins*. Given the sample size, Frahm et al. (2005) find semi-parametric estimation procedures perform best for the estimation of TDCs. The semi-parametric estimation consisted of two steps: firstly, parametric copula families with empirical margins were fitted to the data with a maximum likelihood estimation (Joe 1997). Bivariate Gaussian-, t-, Archimedean- and two parameter Archimedean copulas were considered, therefore covering a wide range of possible dependence structures between indicator pairs and all of the most important and most used copula classes (compare Joe (1997) and Nelsen (2006)). More precisely, 39 different parametric copula types were fitted to the data with the VineCopula package in R. See Schnepsmeier et al. (2017) for a full list of these copula types. Secondly, theoretical TDCs were computed on the basis of the copula parameters estimated in step one.

iii) *Choose the three copula families which fit best overall indicator pairs*. It was then analytically and graphically evaluated which of the copula families computed in step ii) provided the best fit for the data at hand. Firstly, all the copula families tested were ranked for each indicator pair according to Akaike Information criteria (AIC), Bayesian Information Criteria (BIC) and maximum likelihood values. Copula fit was likewise checked graphically with density plots in which the observed data were plotted against 5000 simulated observations (based on the copula parameters estimated in step ii)).The three copula families with the best fit for each crop were selected to provide a broad and robust overview.

iv) *Identify copula family with best goodness-of-fit*. The copula family with best goodness-of-fit compared to all other copula families considered in step ii) was identified out of the three copula families selected in step iii). Goodness-of-fit was tested with Vuong and Clarke tests (Vuong 1989, Clarke 2007), and the copula family with the highest score among all copula families was chosen. In addition, Akaike and Schwarz corrections (Vuong 1989, Clarke 2007) were applied for the number of estimated parameters and then compared to the original results as a robustness check.

v) *Check robustness of obtained copula results with a parametric bootstrap test and the fully non-parametric stable tail dependence function approach*. Firstly, to support the copula selection and the findings on tail dependence obtained with the semi-parametric approach described in steps ii) - iv), a parametric bootstrap test for copula goodness-of-fit (Genest et al. 2006) was performed, based on Kendall's process (using Cramer-von-Mises test statistics (Anderson, 1962)). Finally, a second, fully non-parametric robustness check was carried out using the stable tail dependence function (Kiriliouk et al. 2016), which is an approach for inference on tail dependence.

3.5 Data description and validity checks

The dataset used here was provided by the Swiss Central Evaluation of Agri-Environmental Indicators (CE-AEI). The data is obtained from Agroscope, the Swiss Federal center of excellence for agricultural research, that supervises the collection of the here used data. Farmers participate voluntarily in the data collection scheme and receive compensations for participation (for detailed information on payment schemes and confidentiality agreements see www.agrarmonitoring.ch). Data collection is handled by farmers with the AGRO-TECH software, which is also used to document compliance with regulations and direct payment schemes, and therefore integrated in daily working routines. Primary data were collected and anonymized by intermediaries to guarantee anonymity.

The dataset used for the analysis is unique in its level of detailed information on farmers' management practices. It consisted of an unbalanced panel (i.e. not all entities of farmers appear in the sample in every year) of non-organically producing Swiss farmers reporting on their input use (e.g. fertilizer, pesticides, machinery) from 2009 to 2013. While the absolute number of observations on a farm-level was comparatively small (in total 462 farm-level observations), the dataset consisted of around 7500 and 6000 field-level management measures reported with daily resolution for winter wheat and potatoes, respectively. Detailed information on the properties of the pesticides used was available on a product level for each pesticide application. This allowed the robustness of results from the analysis to be tested over different years, during single years, over different aggregation levels and for the most important subgroups of pesticides (herbicides and fungicides). Detailed tables indicating the number of farmers and fields per crop and year and the types of pesticides used per field are documented in the Appendix (Tables A1 and A2). The analysis considered a wide range of application regimes, including 118 different pesticides for potatoes and 131 different pesticides for winter wheat.

As data were self-reported by farmers, double entries and typos might appear in the data set. These entries had to be identified meticulously and removed before the dataset could be used. Therefore, a transparent procedure was set up to remove such observations. The procedure is summarized in the Appendix, Section B1. To calculate pesticide indicators, CE-AEI data were merged on a product level with information on fate, toxicity and formulation of products from the Pesticide Properties Database (Lewis et al. 2016), which collects publically available information on pesticides (e.g. from pesticide admission). Information on recommended standard application dosages was obtained from the Swiss pesticide register (e.g. BLW 2013). A detailed description of the databases used and data sources, as well as their linkages, can be found in Figure A1 in the Appendix.

Sample farms differed with regard to farm types, climate conditions, topographic and geographic regions: thirty-six percent of sample observations consisted of purely crop producing farms, whereas sixty-four percent of the farms kept at least one animal unit per ha. Seventy-five percent of observations were located in valleys or hilly regions and twenty-five percent in mountain regions. The altitude of farm locations extended from 360m to 950m above sea level. Farm locations were distributed over the most important crop production regions of Switzerland (see Figure 2).

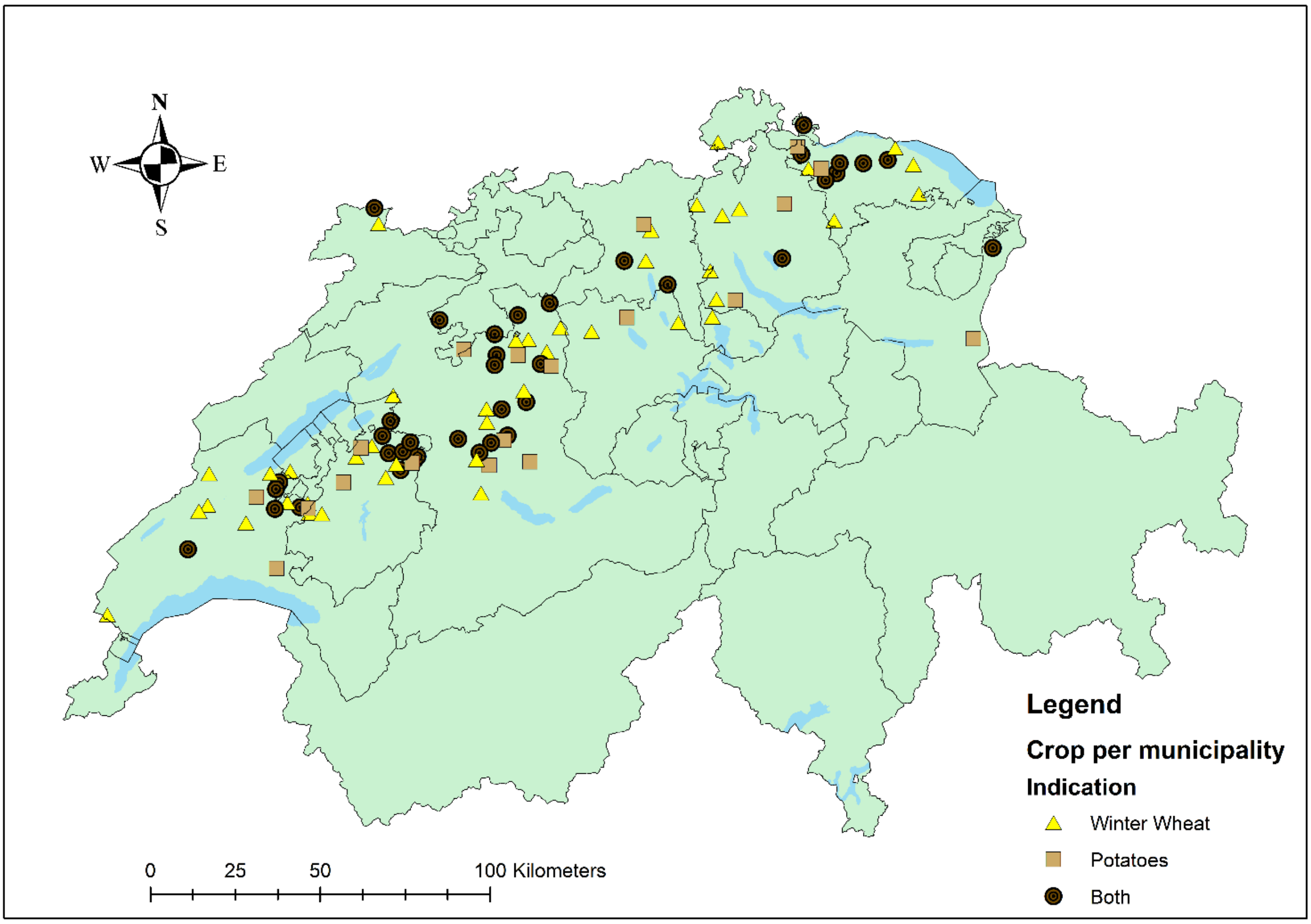


Figure 2. Location of sample farms in Switzerland

Source: Own depiction. Shapefiles from Bundesamt für Statistik (BFS), GEOSTAT (2017). Points show farm locations per municipality.

Spycher et al. (2013) find that average pesticide use in the here used database is well in line with national pesticide use statistics for Switzerland, which also addresses general concerns regarding self-reporting and self-selection of farmers. They state that the dataset covers field crops, i.e. winter wheat and potatoes and the most common pesticide types, i.e. herbicides and fungicides quite well. However, Spycher et al. (2013) find a slightly lower average pesticide use than in national sales statistics, as commonly reported for similar datasets (e.g. Krujine et al. 2012). Furthermore, a good fit of average sample statistics and population averages was found (see Appendix, Section B2). Any concerns regarding biased self-reporting of extreme pesticide use were also addressed by analyzing the compliance of Swiss farmers with pesticide regulations, i.e. pesticide overdosing (see Appendix, Section B2). Results suggest that farmers have reported actual management decisions (i.e. extreme applications), and were not concerned about anonymity issues.

Possible concerns about the unbalanced structure of the sample and pooling of years were addressed by testing for significant differences in skewness of pesticide use distribution across single years (Tables A3 and A4 in the Appendix). In general, none of the five years was found to differ significantly in skewness from all other years (at the 5% significance level), except for LI for winter wheat in 2010. As 2010 was the year with the highest number of participants with winter wheat, this could indicate that skewness of the LI indicator in winter wheat was biased downwards by pooling.

The opposite was observed for the QA indicator in winter wheat, where the skewness was found to be significantly smaller in 2011 and 2012, compared to 2009 and 2010.

4. Results

4.1 Pesticide use trends shift by indicator

Yearly averages of QA, TFI and LI indicators for winter wheat and potato production in Switzerland (Figure A2, Appendix) showed that pesticide use trends reversed when using different indicators. More specifically, the TFI and LI indicators pointed towards a positive trend in pesticide use, while the QA indicator pointed towards a negative trend for the period (2009-2013) under consideration.

Data for all years were then pooled and focus was placed on observations at farm and crop levels, as the crucial pest management decisions are generally taken at the farm-level (Waterfield and Zilberman 2012). Histograms of pesticide use by indicator and crop are shown in Figures A3 and A4 in the Appendix. Results showed that QA and LI indicators had a significantly larger positive skewness than the TFI indicator (Tables A5 and A6, Appendix), suggesting that QA and LI were more likely to indicate extreme applications than TFI. Median, 90% and 99% quantiles of indicators were likewise compared, leading to the same conclusions (Tables A7 and A8, Appendix).

4.2 Analysis of the dependence structure

Significant, positive correlations were found between the QA and LI (in potatoes n=192, Kendall's tau= 0.27 $P< 0.01$; in winter wheat n=270, Kendall's tau= 0.48, $P<0.01$) and between TFI and LI (in potatoes n=192, Kendall's tau= 0.41, $P< 0.01$ and in winter wheat n=270, Kendall's tau= 0.59, $P< 0.01$). Correlation for TFI and LI indicators was higher than for QA and LI indicators, and higher for winter wheat than for potatoes (Table 1). Copulas and tail dependence coefficients (TDCs) were then used to reveal possible asymmetric dependence structures and provide information on the degree of dependence in the tails of the distributions, i.e. between extreme observations. Firstly, the best fitting copulas for the indicator pairs (see Methods, Section 3.4) were identified. Then copula density was plotted. Figure 3 visualizes how the dependence structure shown in the scatterplot (Fig. 3a) is captured by the estimated copula function (Fig. 3b). The figure focuses on the comparison of QA and LI indicators for farm-level pesticide use in potatoes and shows a relatively high dependence between the lower tails of QA and LI and almost no dependence in the upper tails. Respective figures for the LI and TFI indicators, as well as farm-level pesticide use in winter wheat can be found in the Appendix (Figures A5, A6 and A7).

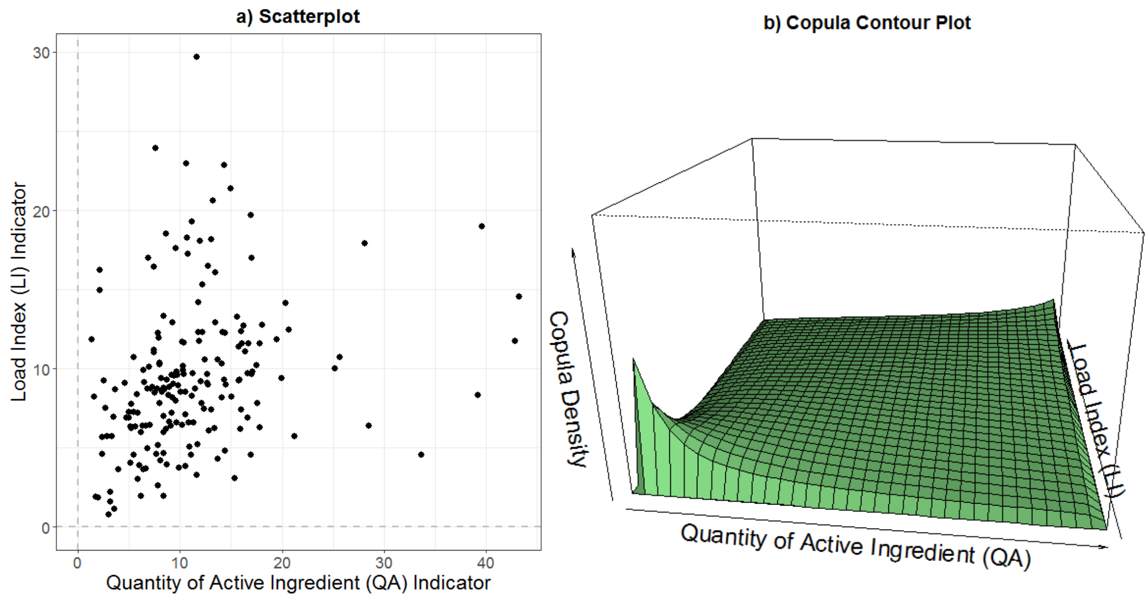


Fig. 3. Copulas reveal differences in the dependence structure of indicators between upper and lower tails

The scatterplot in Figure 3a) shows the relationship between QA and LI values for farm-level pesticide use in potatoes. Pesticide use in terms of the QA and LI indicators is plotted on the X- and Y- axis, respectively. Extremely high values of the QA and LI indicator, which do not correspond with extremely high values of the other indicator, are thus located in the bottom right and top left of the plot respectively. Figure 3b) shows a contour plot of the fitted copula model for the same observations and indicators (Survival Gumbel copula, see Methods, Section 2.4 for selection and estimation procedure). X- and Y- axis in the contour plot show distribution functions for QA and LI values, respectively (see Methods, Section 3.3). The Z-Axis ('Copula Density') shows the probability density of the fitted Survival Gumbel copula. The copula captures the dependence structure between the distributions of the two indicators, not only on average, but over the whole distributions. The plot shows a high copula density for joint low values of QA and LI indicator (bottom left corner) and a low density of the copula for jointly high values of the QA and LI indicator (top right corner).

Upper and lower TDCs of the estimated copula models were then computed to quantify the degree of dependence in the tails of the distributions. TDCs are a measure for dependence in the upper and lower tails of distributions and lie in the range from 0 to 1, where larger values indicate a higher probability of joint extremes. Table 1 shows that all copulas reveal a clearly asymmetric dependence structure between all indicator pairs. While they show a moderate to strong dependence in the lower tails, they exhibit very weak to no dependence in the upper tails. Thus, the analysis suggested that high quantities do not necessarily correspond to high risks. Results were found to be robust over the three best fitting copula types (see Methods, Section 3.4 for selection procedure) and copula goodness-of-fit was confirmed (see Appendix, Section B3 on robustness checks for copulas). TDCs were larger between TFI and LI than between QA and LI, in line with results from the correlation coefficients. Similarly, a stronger dependence was observed for winter wheat than for potatoes. In addition, the robustness of the results was checked further by performing the same analysis again separately for herbicides and fungicides. These are the most relevant pesticide types in the sample and in European agriculture in general (e.g. Eurostat, 2017). Results were robust for both indicator pairs over all analyzed crops, pesticide types and copula families, leading to a total of 36 separately estimated copulas and TDCs (see Appendix Table A9).

Table 1. Significant correlation of quantitative indicators and farm-level pesticide risks – but very low explanatory power for extreme risks

Potatoes

Comparing LI Indicator to:	Kendall's τ correlation	Survival Gumbel Copula		Survival BB1 Copula		Survival BB7 Copula	
		L_{TDC}	U_{TDC}	L_{TDC}	U_{TDC}	L_{TDC}	U_{TDC}
QA Indicator	0.27	**0.32**	**0.00**	0.32	0.00	0.36	0.02
TFI Indicator	0.41	0.49	0.00	**0.45**	**0.06**	0.50	0.28

Winter Wheat

Comparing LI Indicator to:	Kendall's τ correlation	Survival Gumbel Copula		Survival BB1 Copula		BB7 Copula	
		L_{TDC}	U_{TDC}	L_{TDC}	U_{TDC}	L_{TDC}	U_{TDC}
QA Indicator	0.48	**0.56**	**0.00**	0.56	0.00	0.62	0.16
TFI Indicator	0.59	**0.63**	**0.00**	0.61	0.01	0.65	0.36

Level of dependence between farm-level indicators of pesticide use in potato (n=190) and winter wheat (n=270) production. QA, TFI and LI denote the Quantity of Active Ingredient, Treatment Frequency Index and Load Index Indicators, respectively. L_{TDC} and U_{TDC} are in the range of [0,1] and indicate lower and upper tail dependence coefficients, i.e. the dependence for extremely low and extremely high observations between two indicator distributions, respectively. The three copula families analyzed were chosen per crop for the farm-level observations, according to the six-step selection and testing procedure described in Section 3.4. Survival Gumbel, BB1 and BB7 copulas refer to the 180 degrees rotated Gumbel, Clayton-Gumbel and Joe-Clayton copulas, which are one parametric and two parametric Archimedean copulas, respectively. The L_{TDC} of all three copula families lies in a positive range. The U_{TDC} of the Survival Gumbel copula is zero by definition and in a positive range for Survival BB1 and BB7 copulas. See Nelsen (1997) for exact definitions of the selected copula families. Bold numbers indicate the best fitting copula type per indicator pair, according to Vuong and Clarke goodness-of-fit tests (Vuong 1989, Clarke 2007). Kendall correlation coefficients measured the (average) ordinal dependence between two indicators and all were significant at the 1% level. Detailed robustness and goodness-of-fit tests for copulas and TDCs can be found in Appendix, Section B3.

Results continued to remain robust when pesticide use was aggregated on a field level instead of a farm level (Appendix, Table A10).

4.3 Temporal hotspots shift by indicator

So far, extreme pesticide use was analyzed as "one-year application regimes", located in the upper tail of observed pesticide use distributions, without considering the spatial and time dimension of applications. It was then analyzed if results remained consistent when defining extreme pesticide use as temporal "hotspots", i.e. single, daily applications with high risks for the environment and human health compared to the other applications. Analysis with copulas and TDCs was again applied on a level of non-aggregated, single pesticide applications. In line with field- and farm-level results, these results showed that the purely quantitative QA and TFI indicators cannot capture hotspots in terms of environmental and health risks, as indicated by the LI indicator. Although goodness-of-fit of the copula models is weaker at the application level than at the farm-level, results remained qualitatively consistent for both crops, all copula families and different pesticide types tested (Appendix, Table A11). Single applications, pooled for all farmers and years, were plotted on a daily scale (Figure 4) to illustrate differences between QA, TFI and LI. Temporal hotspots as indicated by QA, TFI and LI, respectively (the highest 1% of all applications), were then highlighted. The graph for potatoes illustrates how indicator choice may shift policy targets. While hotspots for the QA occurred between May and July, the TFI identifies June to August as the relevant timeframe. Most importantly, the risk-adjusted LI identified another timeframe (July-August) and other applications as relevant hotspots. This was similarly observed for winter wheat (see Appendix, Figure A8).

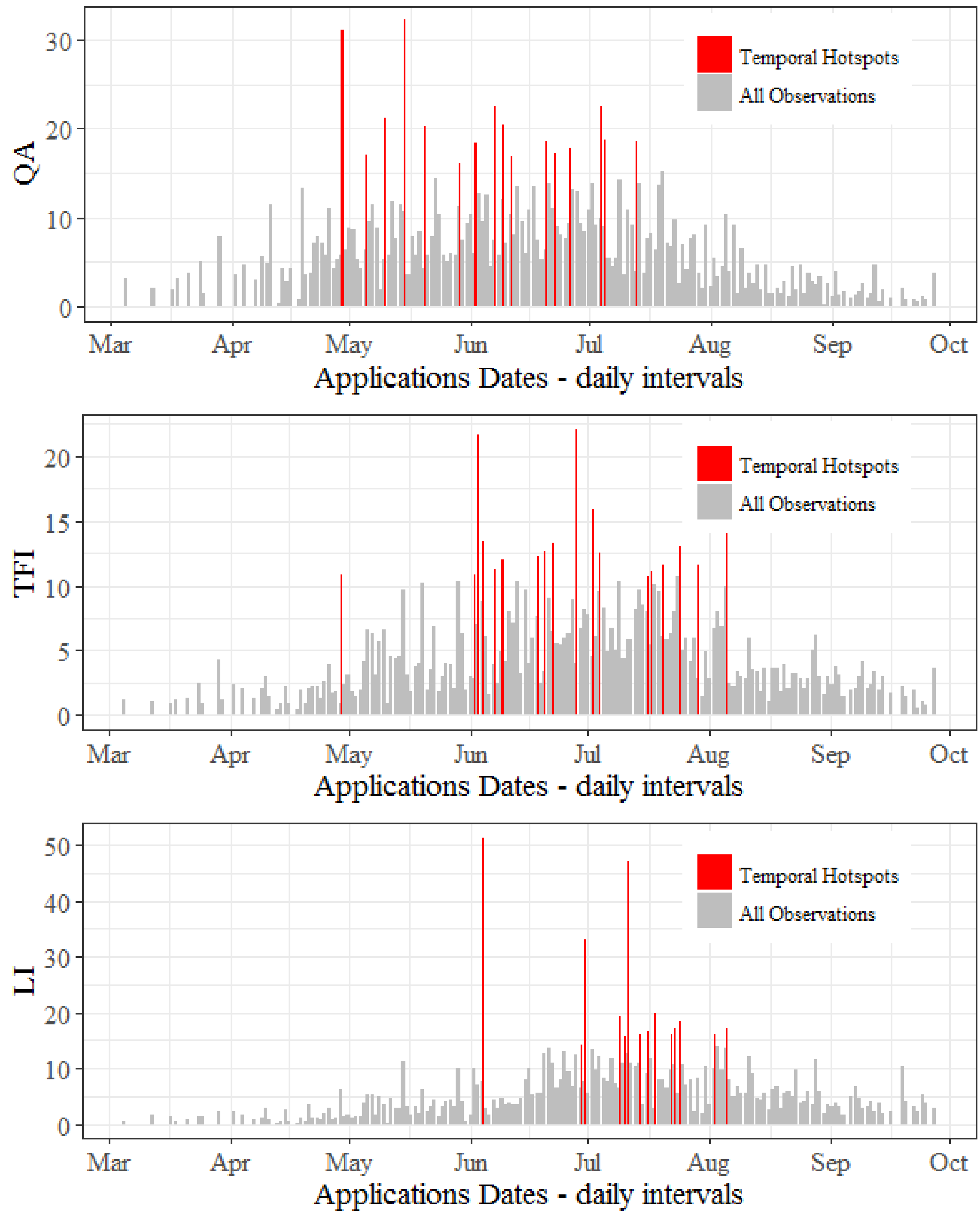


Fig. 4. Temporal hotspots of pesticide use in potatoes shift with indicator choice

Bars represent single pesticide applications in potatoes, plotted on a daily scale. Plotted observations include application data for the years 2009-2013 and all sample farmers. Pesticide use is expressed in Quantity of Active ingredient (QA), Treatment Frequency Index (TFI) and Load Index (LI) indicators, respectively. “Temporal Hotspots” indicate the Top 1% of all applications per indicator, respectively.

5. Discussion

Throughout the world, current pesticide policies have been implemented using diverse indicators for the quantification of pesticide use. Inconsistencies between different pesticide indicators were analyzed in this paper to aid the identification of targets for pesticide policies, i.e. use patterns and profiles associated with the greatest risks for the environment and human health. Results show that the most important, commonly utilized, quantitative pesticide indicators fail to identify the use patterns with the greatest risks for the environment and human health from pesticide use, as indicated by the LI indicator. Therefore, reliance on these indicators for the implementation of high-level pesticide reduction targets and market-based pesticide policies designed to incentivize a reduction in environmental and health impacts from pesticide use, can lead to a misspecification of policy targets and, consequently, to biased policy incentives and outcomes. The goal of such policies is a reduction of risks from pesticide use relative to current levels. It is therefore important to identify the use patterns associated with the greatest risks, even if they are deemed acceptable in terms of potential risks by regulatory authorities (e.g. the United States Environmental Protection Agency and the European Food Safety Authority).

More specifically, the analysis showed that the quantitative indicators QA and TFI have a good average fit to proxy potential environmental and human health risks of pesticides indicated by the LI indicator but are unsuitable for identifying the use patterns associated with the greatest risks, i.e. extreme risks. Differences between indicators may occur due to variations in standard dosages (i.e. between QA and both TFI and LI), inherent pesticide properties (i.e. between TFI and LI) or combinations of both (i.e. between QA or TFI and LI). Positive, significant correlations between both QA and TFI with LI were found. This is in line with Kudsk et al. (2018), who find a good average fit of TFI and LI indicators. The analysis of the dependence structure was then extended with copulas and tail dependence coefficients. Very small coefficients of upper tail dependence were found between both the QA and TFI with the LI. The finding of small “upper TDCs” suggested that indicator choice has a great influence on the analysis of pesticide applications, as extremely high values are not captured jointly. These findings show that the two quantitative indicators (QA, TFI) have almost no explanatory power for the pesticide use observations with the greatest risks for the environment and health as indicated by the LI indicator.

The findings relating to the very low explanatory power of QA and TFI for extreme pesticide risks remained robust over different crops (winter wheat and potatoes), when separately testing the main pesticides types in potato and winter wheat (herbicides and fungicides) and for all tested aggregation levels of pesticide use (single applications, fields, farms). The existence of higher correlation coefficients and TDCs for herbicides compared to fungicides as well as for potatoes compared to winter wheat highlights that the degree of dependence is determined by registered (available) products and farmers’ application patterns in the analyzed cropping system. Differences in results between pesticide groups may be an indication of different levels of heterogeneity with regard to pesticide properties (e.g. toxicity) in the respective group of products. Differences between crops (winter wheat < potatoes) may be an indication that the more input intensive a crop is, the greater the inconsistencies between indicators, as one would expect. The comparison of different aggregation levels is especially important, as the definition of “extreme pesticide use” (i.e. the greatest observations, located in the upper tail of the distribution) is central for the analysis in this article. Two policy-relevant definitions of “extreme pesticide use” were covered in the analysis. Extremes might be relevant in a context of a high risk of aggregated, farm-level pesticide use over the cropping year (“extreme application regimes”), but also in a context of high temporal and spatial concentrations of pesticide risks from single applications (“extreme application events”). Peaks, or “temporal hotspots”, of pesticide use are especially relevant for their effects on ecosystems, like bodies of water. Results revealed that in both cases, quantitative indicators are unsuitable for pinpointing the pesticide applications and application regimes associated with the greatest risks (as shown by the LI indicator), highlighting the broad relevance of the findings.

In addition, results show that trends in national pesticide use may shift with the choice of pesticide indicators, as different indicators capture different properties of the pesticides used. To this end, trends of average, yearly pesticide use were computed and compared for the QA, TFI and LI indicators. The findings confirmed results of Kniss (2017) for the US, who has observed a reversal in pesticide use with different indicators. National pesticide use trends serve to evaluate the efficiency of public policies to regulate pesticide utilization. It is therefore important, from a regulatory perspective, that the representation of pesticide trends reflects potential pesticide impacts in a realistic manner. The scope of our study was not to infer on long-term pesticide use trends in Switzerland, but to identify the differences between indicators. Inference on long-term pesticide use trends demands a longer time series, which would be more robust to yearly shocks, e.g. caused by adverse weather events.

The above findings are transferable to other countries and regions with a similar or even greater heterogeneity of pesticide products in use. The pesticides used by farmers in the sample are similar to those used by producers in the EU. 97.5% and 95% of the 40 most used herbicides and fungicides in the non-EU country Switzerland were also approved in Germany (BLV 2017) and France (Anses 2017), and pesticide use patterns are similar between these countries (CH: Baan et al. 2015, GER: JKI 2017, FR: Agreste 2017). Structural differences in the computation of the three indicators were compared to facilitate the transfer of results to other countries in a more general way. It was found that in comparison to the QA indicator, the lower skewness and shorter right tail of the TFI indicator (Tables A5 and A6 and Figures A3 and A4 in the Appendix) are based on the normalization of applied dosages used to compute it. The greater skewness and longer tail of the LI indicator when compared to the TFI indicator (see Methods, Section on “Indicator calculation”; Tables A5 and A6 and Figures A3 and A4 in the Appendix) are based on the heterogeneity of qualitative pesticide properties (i.e. toxicity) that are also considered. It can therefore be argued that the identified inconsistencies between indicators translate to countries with similar application patterns and registered products. Along these lines, inconsistencies should be even more pronounced for countries with a potentially greater range of qualitative properties (e.g. toxicity) across the pesticides used.

6. Conclusions

In conclusion, results suggest that when implementing pesticide policies, policymakers should consider choice and design of underlying pesticide indicators with great care, as they have significant impacts on the incentives of policies. This is highly relevant in the context of the present implementation and revision of NAPs in Europe, which aim to reduce “risks and impacts of pesticide use on human health and the environment” (Directive 2009/128/EC). More specifically, given the results, market-based policies with underlying quantitative indicators currently in force might create mis-incentives for farmers. Policy measures, such as pesticide taxes (Finger et al. 2017) based on quantitative indicators, could even encourage the use of more risky products if these were to become cheaper than products with low inherent risks, but high standard dosages or a large number of applications. Similarly, the adoption of high-level pesticide reduction targets using quantitative indicators might lead to the implementation of policies which incentivize the use of more risky pesticides. Countries which plan to introduce new or revise existing market-based pesticide policies or reduction targets should effectively target the “risk components” of pesticides (e.g. risks for the environment and human health, etc.), to prevent possible mis-incentives. Such “risk components” might vary in their importance depending on the country and should be adapted to country specific circumstances. The results of the study are especially relevant in the light of recent evidence on the effects of extreme pesticide use on the health of residents in agricultural areas (Larsen et al. 2017) and on the environment (Releya and Hovermann 2006; Gordon et al. 2012, Bundschuh et al. 2013 and Topping and Elmeros 2016). Larsen et al. (2017) conclude that there is a need for the implementation of pesticide policies that tackle extreme applications. Focusing on incentives of policies, this study shows that the choice of a non-target-specific indicator can render such policies inefficient or even lead to detrimental effects. Given the large number of policies currently introduced to reduce adverse effects from pesticide use and achieve greater sustainability in agriculture, the results potentially have large-scale implications for farmers’ pesticide use worldwide.

The results of this study also have implications for the analysis of the economic productivity of pesticides (e.g. as in Hossard et al. 2014 and Lechenet et al. 2017). More specifically, they emphasize the need to investigate the lack of precision in the assessment of negative external effects with purely quantitative indicators. Further research is needed in this field to compare the commonly used quantitative indicators to actual indicators of pesticide efficacy.

The study can be extended in various directions. Firstly, other crops should be considered. This is particularly important for special crops, vegetables and fruits that are characterized by very high absolute levels of pesticide use. Secondly, an analysis of determinants of extreme pesticide use is necessary to design efficient and effective policies for the reduction of this practice. Thirdly, other indicators might be included to quantify environmental and health risks (e.g. Labite et al. 2011; OECD 2016; Kudsk et al. 2018). These indicators could explicitly include information on the application context, thus improving the quality of risk assessment (Pierlot et al. 2017). In addition, indicators based on more complex, site-specific risk assessment models might become more widely available in the future, given the increasing collection of timely, detailed and spatially explicit data in agriculture. Finally, the creation of a harmonized risk indicator for pesticides in Europe could further improve the coherence of policies in the NAP and help to monitor pesticide risks on a detailed level across time and space in Europe.

Acknowledgements

We thank Agroscope for the provision of the used dataset. We further thank the five anonymous reviewers for their constructive feedback on earlier versions of this article.

7. References

Agreste. Pratiques Culturales sur les grandes cultures et prairies. http://agreste.agriculture.gouv.fr/enquetes/pratiques-culturales/grandes-cultures-prairies. Accessed November 2017 (2017).

Anderson, T.W. On the distribution of the two-sample Cramer-von Mises criterion. *The Annals of Mathematical Statistics*, pp.1148-1159 (1962).

Anses. Le catalogue des produits phytopharmaceutiques et de leurs usages, des matières fertilisantes et des supports de culture autorisés en France. https://ephy.anses.fr. Accessed December 2017 (2017).

Asquith, W. Package 'copBasic'. R package version 2.0.5 (2017).

Baan, de L., Spycher, S., & Daniel, O. Einsatz von Pflanzenschutzmitteln in der Schweiz von 2009 bis 2012. *Agrarforschung Schweiz* 6, 48-55 (2015).

Baumol, W.J., Oates, W.E., Bawa, V.S., Bawa, W.S. and Bradford, D.F. *The theory of environmental policy*. Cambridge university press (1988).

Beketov, M.A., Kefford, B.J., Schäfer, R.B. & Liess, M. Pesticides reduce regional biodiversity of stream invertebrates. *Proceedings of the National Academy of Sciences*, *110*(27), pp.11039-11043 (2013).

Böcker, T. & Finger, R. European Pesticide Tax Schemes in Comparison: An Analysis of Experiences and Developments. *Sustainability 8*(4), 378 (2016).

Böcker, T., Britz, W., Finger, R. (2018). Modelling the effects of a glyphosate ban on weed management in silage maize production. *Ecological Economics* 145: 182–193.

Bundesrat der Schweizerischen Eidgenossenschaft. *Aktionsplan zur Risikoreduktion und nachhaltigen Anwendung von Pflanzenschutzmitteln*. Bern, Schweiz (2017).

Bundesamt für Statistik (BFS), GEOSTAT 2017. BFS GEOSTAT Gemeindegrenzen. https://www.bfs.admin.ch/bfs/de/home/dienstleistungen/geostat/geodaten-bundesstatistik/administrative-grenzen/generalisierte-gemeindegrenzen.html. Accessed March 2017 (2017).

Bundesamt für Statistik (BFS). BFS – Landwirtschaftliche Strukturerhebung 2016. https://www.bfs.admin.ch/bfs/de/home/statistiken/land-forstwirtschaft/landwirtschaft/strukturen.html. Accessed October 2017 (2017).

Bundesamt für Lebensmittel und Verbraucherschutz (BLV). Zugelassene Pflanzenschutzmittel. https://www.bvl.bund.de/DE/04_Pflanzenschutzmittel/psm_node.html. Accessed November 2017 (2017).

Bundschuh, M., Zubrod, J.P., Klemm, P., Elsaesser, D., Stang, C. & Schulz, R. Effects of peak exposure scenarios on *Gammarus fossarum* using field relevant pesticide mixtures. *Ecotoxicology and environmental safety*, *95*, pp.137-143 (2013).

Bundesamt für Landwirtschaft (BLW). BLW Pflanzenschutzmittelverzeichnis. http://www.psm.admin.ch/psm/produkte. Accessed December 2013 (2013).

Clarke, K. A Simple Distribution-Free Test for Nonnested Model Selection. *Political Analysis*, 15, 347-363 (2007).

Claeys, S., Vagenende, B., De Smet, B., Lelieur, L. and Steurbaut, W. The POCER indicator: a decision tool for non-agricultural pesticide use. *Pest management science*, *61*(8), pp.779-786 (2005).

Coll, M., & Wajnberg, E (eds.). *Environmental Pest Management: Challenges for Agronomists, Ecologists, Economists and Policymakers*. Hoboken, NJ, USA: Wiley (2017).

Croux, C., & Dehon, C. Influence functions of the Spearman and Kendall correlation measures. Statistical methods & applic*ations*, *19*(4), 497-515 (2010).

Damalas, C. A. Understanding benefits and risks of pesticide use. *Scientific Research and Essays 4*(10), 945-949 (2009).

Davison, A. C., & Hinkley, D. V. Bootstrap methods and their application (Vol. 1). Cambridge University press (1997).

European Commission (EC) (ed.).*Annex II: Refit Initiatives to the Commission Work Programme 2016*. European Commission, Strasbourg, France (2015).

European Commission Scientific Advice Mechanism (EC SAM) (ed.). *EU authorization processes of Plant Protection Products from a scientific point of view*. Group of Chief Scientific Advisors, Brussels, Belgium (2018).

European Union (EU). EU Pesticides database. http://ec.europa.eu/food/plant/pesticides/eu-pesticides-database/public. Accessed December 2017 (2017).

European Union, DG Health and Food Safety (ed.). "*Overview Report – Sustainable Use of Pesticides*". Luxembourg: Publications Office of the European Union (2017).

Eurostat. Agri-environmental indicator – consumption of pesticides. http://ec.europa.eu/eurostat/statistics-explained/index.php/Agri-environmental_indicator_-_consumption_of_pesticides#Data_used_and_methodology. Accessed April 2017 (2017).

Finger, R., Möhring, N., Dalhaus, T. & Böcker, T. Revisiting pesticide taxation schemes. Ecological Economics, 134, pp.263-266 (2017).

Finger, R. Take a holistic view when making pesticide policies stricter. Nature 556, 174 (2018).

Frahm, G., Junker, M. & Schmidt, R. Estimating the tail-dependence coefficient: properties and pitfalls. Insurance: mathematics and Economics, 37(1), pp.80-100 (2005).

Gaba, S., Gabriel, E., Chadœuf, J., Bonneu, F. & Bretagnolle, V. Herbicides do not ensure for higher wheat yield, but eliminate rare plant species. Scientific Reports, 6 (2016).

Genest, C., Quessy, J. F., & Rémillard, B. Goodness-of-fit procedures for copula models based on the probability integral transformation. Scandinavian Journal of Statistics, 33(2), 337-366 (2006).

Genest, C., & Favre, A. C. Everything you always wanted to know about copula modeling but were afraid to ask. Journal of hydrologic engineering, 12(4), 347-368 (2007).

Genest, C., & Rémillard, B. Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models. In Annales de l'Institut Henri Poincaré, Probabilités et Statistiques (Vol. 44, No. 6, pp. 1096-1127). Institut Henri Poincaré (2008).

Ghimire, N., & Woodward, R. T. Under-and over-use of pesticides: An international analysis. Ecological Economics, 89, 73-81 (2013).

Gordon, A.K., Mantel, S.K. & Muller, N.W. Review of toxicological effects caused by episodic stressor exposure. Environmental toxicology and chemistry, 31(5), pp.1169-1174 (2012).

Gutsche, V. and Rossberg, D. SYNOPS 1.1: a model to assess and to compare the environmental risk potential of active ingredients in plant protection products. *Agriculture, ecosystems & environment*, *64*(2), pp.181-188 (1997).

Hoop, D. & Schmid, D. Grundlagenbericht 2013: Zentrale Auswertung von Buchhaltungsdaten. Ettenhausen: Agroscope INH (2014).

Hossard, L., Philibert, A., Bertrand, M., Colnenne-David, C., Debaeke, P., Munier-Jolain, N., Jeuffroy, M.H., Richard, G. & Makowski, D. Effects of halving pesticide use on wheat production. *Scientific reports*, 4 (2017).

Hossard, L., Guichard, L., Pelosi, C. and Makowski, D., 2017. Lack of evidence for a decrease in synthetic pesticide use on the main arable crops in France. *Science of the Total Environment*, 575, pp.152-161 (2014).

Houbraken, M., Bauweraerts, I., Fevery, D., Van Labeke, M.C. and Spanoghe, P. Pesticide knowledge and practice among horticultural workers in the Lâm Đồng region, Vietnam: A case study of chrysanthemum and strawberries. *Science of the Total Environment*, *550*, pp.1001-1009 (2016).

Julius Kühn-Institut (JKI). Statistische Erhebungen zur Anwendung von Pflanzenschutzmitteln in der Praxis. Ergebnisse. http://papa.jki.bund.de. Accessed November 2017 (2017).

Joe, H. Multivariate models and multivariate dependence concepts. CRC Press (1997).

Kiriliouk, A., Segers, J. & Warchoł, M. Nonparameteric estimation of extremal dependence: in Extreme Value Modeling and Risk Analysis, D.K. Dey and Jun Yan (eds.), Boca Raton, FL, CRC Press (2016).

Kniss, A. R. Long-term trends in the intensity and relative toxicity of herbicide use. Nature communications, 8 (2017).

Kruijne R., van der Linden A.M.A., Deneer J.W., Groenwold J. G. & Wipfler E.L. *Dutch Environmental Risk Indicator for Plant Protection Products*. Wageningen, Netherlands, Alterra-Report 2250.1. (2012).

Kudsk, P., Jørgensen, L.N. & Ørum, J.E. Pesticide Load—A new Danish pesticide risk indicator with multiple applications. Land Use Policy, 70, pp.384-393 (2018).

Labite, H., Butler, F. and Cummins, E. A review and evaluation of plant protection product ranking tools used in agriculture. Human and Ecological Risk Assessment, 17(2), pp.300-327 (2011).

Larsen, A. E., Gaines, S. D., & Deschênes, O. Agricultural pesticide use and adverse birth outcomes in the San Joaquin Valley of California. Nature Communications, 8(1), 302 (2017).

Lechenet, M., Dessaint, F., Py, G., Makowski, D. & Munier-Jolain, N. Reducing pesticide use while preserving crop productivity and profitability on arable farms. Nature Plants, 3, p.17008 (2017).

Lefebvre, M.; Langrell, S.R.H. & Gomez-y-Paloma, S. Incentives and policies for integrated pest management in Europe: A review. Agron. Sustain. Dev., 35, 27–45 (2015).

Lewis, K. A., Tzilivakis, J., Warner, D. J. & Green, A. An international database for pesticide risk assessments and management. Human and Ecological Risk Assessment: An International Journal 22(4), 1050-1064 (2016).

Malaj, E., Peter, C., Grote, M., Kühne, R., Mondy, C.P., Usseglio-Polatera, P., Brack, W. & Schäfer, R.B.. Organic chemicals jeopardize the health of freshwater ecosystems on the continental scale. Proceedings of the National Academy of Sciences, 111(26), pp.9549-9554 (2014).

Miljøministeriet. The Agricultural Pesticide Load in Denmark 2007-2010. Environmental review no. 2, Miljøstyrelsen, Kopenhagen (2012).

Ministère de l'Agriculture, de l'Agroalimentaire et de la Forêt (MAAF). Plan Ecophyto II. http://agriculture.gouv.fr/sites/minagri/files/151022_ecophyto.pdf. Accessed January 2018 (2015).

Ministère de l'Agriculture et de l'Alimentation (MAA). Indicateur de Fréquence de Traitements phytosanitaires (IFT). http://agriculture.gouv.fr/indicateur-de-frequence-de-traitements-phytosanitaires-ift. Accessed November 2017 (2017).

Nelsen, R. B. An introduction to copulas, 2nd. New York: SpringerScience Business Media (2006).

Organisation for Economic Co-operation and Development (OECD). Report on the online pesticide risk indicators database (PRIER). Environment Directorate, OECD, Paris (2016).

Okhrin, O., Odening, M. and Xu, W. Systemic weather risk and crop insurance: the case of China. *Journal of Risk and Insurance*, *80*(2), pp.351-372 (2013).

Okhrin, O., Trimborn, S., Zhang, S. & Zhou, Q. M. gof-Copula: Goodness-of-Fit Tests for Copulae. R package version 0.2.3 (2016).

Osteen, C. D., & Fernandez-Cornejo, J. Economic and policy issues of US agricultural pesticide use trends. Pest management science, 69(9), 1001-1025 (2013).

Payton, M. E., Greenstone, M. H., & Schenker, N. Overlapping confidence intervals or standard error intervals: what do they mean in terms of statistical significance? Journal of Insect Science, 3(1), 34 (2003).

Perry, E. D., Ciliberto, F., Hennessy, D. A., & Moschini, G. Genetically engineered crops and pesticide use in US maize and soybeans. Science advances, 2(8), e1600850 (2016).

Pierlot, F., Marks-Perreau, J., Real, B., Carluer, N., Constant, T., Lioeddine, A., Van Dijk, P., Villerd, J., Keichinger, O., Cherrier, R. and Bockstaller, C. Predictive quality of 26 pesticide risk indicators and one flow model: A multisite assessment for water contamination. *Science of the Total Environment*, *605*, pp.655-665 (2017).

Pimentel, D. and Burgess, M. “Environmental and economic costs of the application of pesticides primarily in the United States.” In Pimentel, D, Peshin, R., ed. Integrated pest management Pesticide Problems Vol. 3. Dordrecht, Netherlands: Springer, pp. 47-71 (2014).

R Core Team. R: A language and environment for statistical computing (2013).

Relyea, R. & Hoverman, J. Assessing the ecology in ecotoxicology: a review and synthesis in freshwater systems. Ecology Letters, 9(10), pp.1157-1171 (2006).

Schepsmeier, U., Stoeber, J., Brechmann, E. C., & Graeler, B. VineCopula: statistical inference of vine copulas. R package version 2.1.3 (2017).

Skevas, T., Lansink, A.O. and Stefanou, S.E. Designing the emerging EU pesticide policy: A literature review. *NJAS-Wageningen Journal of Life Sciences*, *64*, pp.95-103 (2013).

Spycher, S., Badertscher, R., & Daniel, O. Indikatoren für den Einsatz von Pflanzenschutzmitteln in der Schweiz. Agrarforschung Schweiz 4(4), 192–199 (2013).

Stehle, S. & Schulz, R. Agricultural insecticides threaten surface waters at the global scale. Proceedings of the National Academy of Sciences, 112(18), pp.5750-5755 (2015).

Storck, V., Karpouzas, D.G. and Martin-Laurent, F. Towards a better pesticide policy for the European Union. *Science of the Total Environment*, *575*, pp.1027-1033 (2017).

Strange, R.N. & Scott, P.R. Plant disease: a threat to global food security. Annu. Rev. Phytopathol., 43, pp.83-116 (2005).

Strassemeyer, J., Daehmlow, D., Dominic, A.R., Lorenz, S. and Golla, B. SYNOPS-WEB, an online tool for environmental risk assessment to evaluate pesticide strategies on field level. *Crop protection*, *97*, pp.28-44. (2017)

Sun, B., Zhang, L., Yang, L., Zhang, F., Norse, D., & Zhu, Z. Agricultural non-point source pollution in China: causes and mitigation measures. Ambio, 41(4), 370-379 (2012).

Tilman, D., Cassman, K. G., Matson, P. A., Naylor, R., & Polasky, S. Agricultural sustainability and intensive production practices. Nature, 418(6898), 671 (2002).

Topping, C.J. and Elmeros, M. Modeling Exposure of Mammalian Predators to Anticoagulant Rodenticides. *Frontiers in Environmental Science*, *4*, p.80 (2016).

Tsaboula, A., Papadakis, E.N., Vryzas, Z., Kotopoulou, A., Kintzikoglou, K. and Papadopoulou-Mourkidou, E. Environmental and human risk hierarchy of pesticides: A prioritization method, based on monitoring, hazard assessment and environmental fate. *Environment international*, *91*, pp.78-93 (2016).

United States Department of Agriculture (USDA). USDA/NASS QuickStats Ad-hoc Query Tool. https://quickstats.nass.usda.gov. Accessed July 2017 (2017).

Vercruysse, F. and Steurbaut, W. POCER, the pesticide occupational and environmental risk indicator. *Crop Protection*, *21*(4), pp.307-315 (2002).

Vuong, Q. H. Ratio tests for model selection and non-nested hypotheses. Econometrica 57 (2), 307-333 (1989).

Waterfield, G. & Zilberman, D. "Pest management in food systems: an economic perspective." Annual review of environment and resources 37: 223-245 (2012).

Xu, W., Filler, G., Odening, M. and Okhrin, O. On the systemic nature of weather risk. *Agricultural Finance Review*, *70*(2), pp.267-284 (2010).

Yang, Y., Lian, X.Y., Jiang, Y.H., Xi, B.D. and He, X.S. Risk-based prioritization method for the classification of groundwater pesticide pollution from agricultural regions. *Integrated environmental assessment and management*, *13*(6), pp.1052-1059 (2017).

Zhang, K. M., & Wen, Z. G. Review and challenges of policies of environmental protection and sustainable development in China. Journal of environmental management, 88(4), 1249-1261 (2008).

Zhang, S., Okhrin, O., Zhou, Q.M. and Song, P.X.K. Goodness-of-fit test for specification of semiparametric copula dependence models. Journal of Econometrics, 193(1), pp.215-233 (2016).

Appendices

Table A1. Sample size

Year	Winter Wheat	Potatoes
2009	61 (195)	44 (101)
2010	66 (193)	38 (103)
2011	52 (158)	40 (108)
2012	48 (137)	32 (82)
2013	43 (134)	38 (88)
Total	270 (817)	192 (482)

Number of farms (fields) per year and crop

Table A2. Share of treated fields per pesticide type

	Winter Wheat			Potatoes		
Year	Herbicide	Fungicide	Insecticide	Herbicide	Fungicide	Insecticide
2009	0.92	0.76	0.07	0.88	0.97	0.25
2010	0.94	0.84	0.08	0.93	0.99	0.27
2011	0.96	0.76	0.25	0.92	0.97	0.31
2012	0.96	0.85	0.12	0.96	1.00	0.35
2013	0.97	0.80	0.04	0.93	0.98	0.39
Total	0.95	0.80	0.11	0.92	0.98	0.31

Share of fields treated with at least one pesticide, per pesticide type and year

Table A3. 85% confidence intervals for skewness of yearly pesticide use distributions in potatoes

Pesticide Type	Year	QA		TFI		LI	
		Lower	*Upper*	*Lower*	*Upper*	*Lower*	*Upper*
All Pesticides	2009	1.9	4.5	-0.0	0.6	0.3	1.7
	2010	0.9	2.1	-0.1	0.8	0.9	2.9
	2011	1.8	4.4	-0.4	0.3	0.8	1.9
	2012	1.4	3.6	0.1	1.1	0.3	1.3
	2013	2.2	5.2	-0.5	0.2	0.1	1.0

Skewness is calculated for each year (2009-2013) from farm and culture level pesticide use distributions. Lower and Upper indicate the lower and upper bound of the 85% confidence interval for each statistic, respectively. Payton et al. (2003) show that, given a maximum error rate of 5%, 85% confidence intervals should be used to check for overlaps (see Methods, Section 2.2). Confidence intervals are computed with non-parametric bootstrapping techniques with 10000 repetitions, respectively. QA, TFI and LI denote the Quantity of Active Ingredient, Treatment Frequency Index and Load Index Indicators, respectively.

Table A4. 85% confidence intervals for skewness of yearly pesticide use distributions in winter wheat

Pesticide Type	Year	QA		TFI		LI	
		Lower	*Upper*	*Lower*	*Upper*	*Lower*	*Upper*
All Pesticides	2009	1.1	3.0	-0.3	0.3	0.9	1.7
	2010	1.3	3.1	-0.1	0.5	-0.0	0.7
	2011	0.2	0.8	0.0	0.6	2.0	3.7
	2012	-0.1	0.6	-0.3	0.3	1.2	2.9
	2013	0.5	1.3	0.3	1.8	1.6	2.9

Skewness is calculated for each year (2009-2013) from farm and culture level pesticide use distributions. Lower and Upper indicate the lower and upper bound of the 85% confidence interval for each statistic, respectively. Payton et al. (2003) show that, given a maximum error rate of 5%, 85% confidence intervals should be used to check for overlaps (see Methods, Section 2.2). Confidence intervals are computed with non-parametric bootstrapping techniques with 10000 repetitions, respectively. QA, TFI and LI denote the Quantity of Active Ingredient, Treatment Frequency Index and Load Index Indicators, respectively.

Table A5. 85% confidence intervals of summary statistics for farm-level pesticide use in potatoes

Pesticide Type	Confidence Interval of	QA		TFI		LI	
		Lower	*Upper*	*Lower*	*Upper*	*Lower*	*Upper*
Herbicides	Mean	2.46	2.82	1.99	2.23	3.92	4.65
	Sd	1.63	1.82	1.05	1.24	3.09	4.00
	Skewness	0.12	0.50	0.33	0.83	1.23	2.20
Fungicides	Mean	5.98	6.66	5.54	6.19	4.41	5.08
	Sd	3.10	3.47	2.87	3.31	2.99	3.67
	Skewness	-0.01	0.30	0.27	0.66	0.90	1.58
All Pesticides	Mean	10.43	11.87	8.48	9.22	8.86	9.84
	Sd	5.87	7.94	3.31	3.82	4.27	5.23
	Skewness	1.80	2.60	0.05	0.53	0.79	1.47

Summary statistics are calculated for the pooled data (2009-2013) on a farm and culture level. Lower and Upper indicate the lower and upper bound of the 85% confidence interval for each statistic, respectively. Payton et al. (2003) show that, given a maximum error rate of 5%, 85% confidence intervals should be used to check for overlaps (see Methods, Section 2.2). Confidence intervals are computed with non-parametric bootstrapping techniques with 10000 repetitions, respectively. QA, TFI and LI denote the Quantity of Active Ingredient, Treatment Frequency Index and Load Index Indicators, respectively.

Table A6. 85% confidence intervals of summary statistics for farm-level pesticide use in winter wheat

Pesticide Type	Confidence Interval of	QA		TFI		LI	
		Lower	*Upper*	*Lower*	*Upper*	*Lower*	*Upper*
Herbicides	Mean	0.76	0.93	1.23	1.34	0.79	1.03
	Sd	0.89	1.05	0.57	0.65	1.08	1.62
	Skewness	1.25	1.65	0.35	0.82	3.39	4.47
Fungicides	Mean	0.71	0.84	1.25	1.42	1.32	1.55
	Sd	0.67	0.86	0.92	1.01	1.22	1.4
	Skewness	1.29	2.71	-0.11	0.2	0.76	1.07
All Pesticides	Mean	1.9	2.15	3.13	3.37	2.32	2.73
	Sd	1.28	1.55	1.33	1.5	1.99	2.6
	Skewness	0.84	1.84	0.08	0.56	1.75	2.74

Summary statistics are calculated for the pooled data (2009-2013) on a farm and culture level. Lower and Upper indicate the lower and upper bound of the 85% confidence interval for each statistic, respectively. Payton et al. (2003) show that, given a maximum error rate of 5%, 85% confidence intervals should be used to check for overlaps (see Methods, Section 2.2). Confidence intervals are computed with non-parametric bootstrapping techniques with 10000 repetitions, respectively. QA, TFI and LI denote the Quantity of Active Ingredient, Treatment Frequency Index and Load Index Indicators, respectively.

Table A7. Median, 90% decile and 99% percentiles of farm-level pesticide use in potatoes

Indicator	Median	90% decile	99% percentile
QA	9.85	17.32	41.92
TFI	8.65	14.00	17.55
LI	8.75	16.44	23.66

Parameters are calculated for all years pooled (2009-2013) on a farm and culture level. QA, TFI and LI denote the Quantity of Active Ingredient, Treatment Frequency Index and Load Index Indicators, respectively. The ratio of 50% to 90% and 99% percentiles give information about the skewness of the distribution and the weight of the distribution on upper tails.

Table A8. Median, 90% decile and 99% percentiles of farm-level pesticide use in winter wheat

Indicator	Median	90% decile	99% percentile
QA	1.78	3.9	6.1
TFI	3.17	5.14	6.53
LI	2.13	4.71	12.63

Parameters are calculated for all years pooled (2009-2013) on a farm and culture level. QA, TFI and LI denote the Quantity of Active Ingredient, Treatment Frequency Index and Load Index Indicators, respectively. The ratio of 50% to 90% and 99% percentiles give information about the skewness of the distribution and the weight of the distribution on upper tails.

Table A9. Significant correlation of quantitative indicators and farm-level herbicide and fungicide risks – but very low explanatory power for extreme risks

Potatoes

Pesticide Type	Comparing LI Indicator to:	Kendall's τ correlation	Survival Gumbel Copula L_{TDC}	Survival Gumbel Copula U_{TDC}	Survival BB1 Copula L_{TDC}	Survival BB1 Copula U_{TDC}	Survival BB7 Copula L_{TDC}	Survival BB7 Copula U_{TDC}
Herbicides	QA Indicator	0.17	0.32	0.00	0.32	0.00	**0.46**	**0.00**
	TFI Indicator	0.46	0.58	0.00	0.58	0.00	**0.67**	**0.20**
Fungicides	QA Indicator	0.36	**0.45**	**0.00**	0.42	0.01	0.48	0.18
	TFI Indicator	0.62	**0.70**	**0.00**	0.68	0.10	0.74	0.52

Winter Wheat

Pesticide Type	Comparing LI Indicator to:	Kendall's τ correlation	Survival Gumbel Copula L_{TDC}	Survival Gumbel Copula U_{TDC}	Survival BB1 Copula L_{TDC}	Survival BB1 Copula U_{TDC}	BB7 Copula L_{TDC}	BB7 Copula U_{TDC}
Herbicides	QA Indicator	0.54	**0.55**	**0.00**	0.64	0.00	0.71	0.25
	TFI Indicator	0.48	0.57	0.00	**0.50**	**0.24**	0.57	0.47
Fungicides	QA Indicator	0.58	0.71	0.00	0.71	0.00	**0.85**	**0.06**
	TFI Indicator	0.66	0.77	0.00	0.77	0.00	**0.89**	**0.28**

Level of dependence between farm-level indicators of pesticide use in potato and winter wheat production. QA, TFI and LI denote the Quantity of Active Ingredient, Treatment Frequency Index and Load Index Indicators, respectively. LTDC and UTDC are in the range of [0,1] and indicate lower and upper tail dependence coefficients, i.e. the dependence for extremely low and extremely high observations, respectively. The three analyzed copula families are chosen per crop for farm-level observations, according to the six-step selection and testing procedure described in the Methodology, Section 2.4. Bold numbers indicate the best fitting copula type per indicator pair, according to Vuong and Clarke goodness-of-fit tests (Vuong 1989, Clarke 2007). The UTDC of the Survival Gumbel copula is zero by definition. Kendall correlation coefficients measure the (average) ordinal dependence between two indicators and are all significant at the 1% level. Detailed robustness and goodness-of-fit tests for copulas and TDCs can be found in the Appendix, Section B3.

Table A10. Significant correlation of quantitative indicators and field-level pesticide risks – but very low explanatory power for extreme risks

Potatoes

Comparing LI Indicator to:	Kendall's τ correlation	Survival Gumbel Copula		Survival BB1 Copula		Survival BB6 Copula	
		L_{TDC}	U_{TDC}	L_{TDC}	U_{TDC}	L_{TDC}	U_{TDC}
QA Indicator	0.47	**0.28***	**0.00***	0.28	0.00	0.28*	0.00*
TFI Indicator	0.58	0.42	0.00	**0.44**	**0.00**	0.44	0.00

Winter Wheat

Comparing LI Indicator to:	Kendall's τ correlation	Survival Gumbel Copula		Survival BB1 Copula		Survival BB8 Copula	
		L_{TDC}	U_{TDC}	L_{TDC}	U_{TDC}	L_{TDC}	U_{TDC}
QA Indicator	0.22	0.57	0.00	0.57	0.00	**0.00***	**0.00***
TFI Indicator	0.37	0.63	0.00	0.62	0.00	**0.00**	**0.00**

Level of dependence between field-level indicators of pesticide use in potato and winter wheat production. QA, TFI and LI denote the Quantity of Active Ingredient, Treatment Frequency Index and Load Index Indicators, respectively. LTDC and UTDC are in the range of [0,1] and indicate lower and upper tail dependence coefficients, i.e. the dependence for extremely low and extremely high observations, respectively. LTDC and UTDC are in the range of [0,1] and indicate lower and upper tail dependence coefficients, respectively. The three analyzed copula families are chosen per crop for field-level pesticide use observations, according to the six-step selection and testing procedure described in the Methodology, Section 2.4. Bold numbers indicate the best fitting copula type per indicator pair, according to Vuong and Clarke goodness-of-fit tests (Vuong 1989, Clarke 2007). The UTDC of the Survival Gumbel, Survival BB6 and Survival BB8 copulas is zero by definition. Kendall correlation coefficients measure the (average) ordinal dependence between two indicators and are all significant at the 1% level. * denotes acceptance of Kendall copula-goodness-of-fit tests (Genest et al. 2006). Cramer-von Mises test statistics were used to accept/reject the null-hypothesis of a matching copula family at the 5% level. P-values were computed according to the parametric bootstrap procedure described in Genest et al. (2006).

Table A11. Significant correlation of quantitative indicators and application-level pesticide risks – but very low explanatory power for extreme risks

Potatoes

Comparing LI Indicator to:	Kendall's τ correlation	Student t Copula		Survival BB6 Copula		Survival BB8 Copula	
		L_{TDC}	U_{TDC}	L_{TDC}	U_{TDC}	L_{TDC}	U_{TDC}
QA Indicator	0.10	0.06	0.06	0.28	0.00	**0.00**	**0.00**
TFI Indicator	0.23	0.07	0.07	0.25	0.00	**0.00**	**0.00**

Winter Wheat

Comparing LI Indicator to:	Kendall's τ correlation	Student t Copula		Frank Copula		Survival BB8 Copula	
		L_{TDC}	U_{TDC}	L_{TDC}	U_{TDC}	L_{TDC}	U_{TDC}
QA Indicator	0.33	0.13	0.13	**0.00**	**0.00**	0.00	0.00
TFI Indicator	0.36	0.07	0.07	**0.00**	**0.00**	0.00	0.00

Level of dependence between application-level indicators of pesticide use in potato and winter wheat production. QA, TFI and LI denote the Quantity of Active Ingredient, Treatment Frequency Index and Load Index Indicators, respectively. LTDC and UTDC are in the range of [0,1] and indicate lower and upper tail dependence coefficients, i.e. the dependence for extremely low and extremely high observations, respectively. LTDC and UTDC are in the range of [0,1] and indicate lower and upper tail dependence coefficients, respectively. The three analyzed copula families are chosen per crop for application-level pesticide use observations. They represent the best fitting copulas according to the six-step selection and testing procedure described in the Methodology, Section 2.4. Bold numbers indicate the best fitting copula type per indicator pair, according to Vuong and Clarke goodness-of-fit tests (Vuong 1989, Clarke 2007). The UTDC of the Frank, Survival BB6 and Survival BB8 copulas is zero by definition. All estimates of Kendall correlation coefficients are significant at the 1% level. Kendall (Genest et al. 2006) goodness-of-fit tests with Cramer-von Mises test statistics were used to test the null-hypothesis of matching Survival BB6, Survival BB8 and Frank copulas. P-values were computed according to the parametric bootstrap procedure described in Genest et al. (2006). PIOS goodness-of-fit tests were used for the Student t copula (Zhang et al. 2016). Copula goodness-of fit could not be confirmed at the 5% level with the above tests.

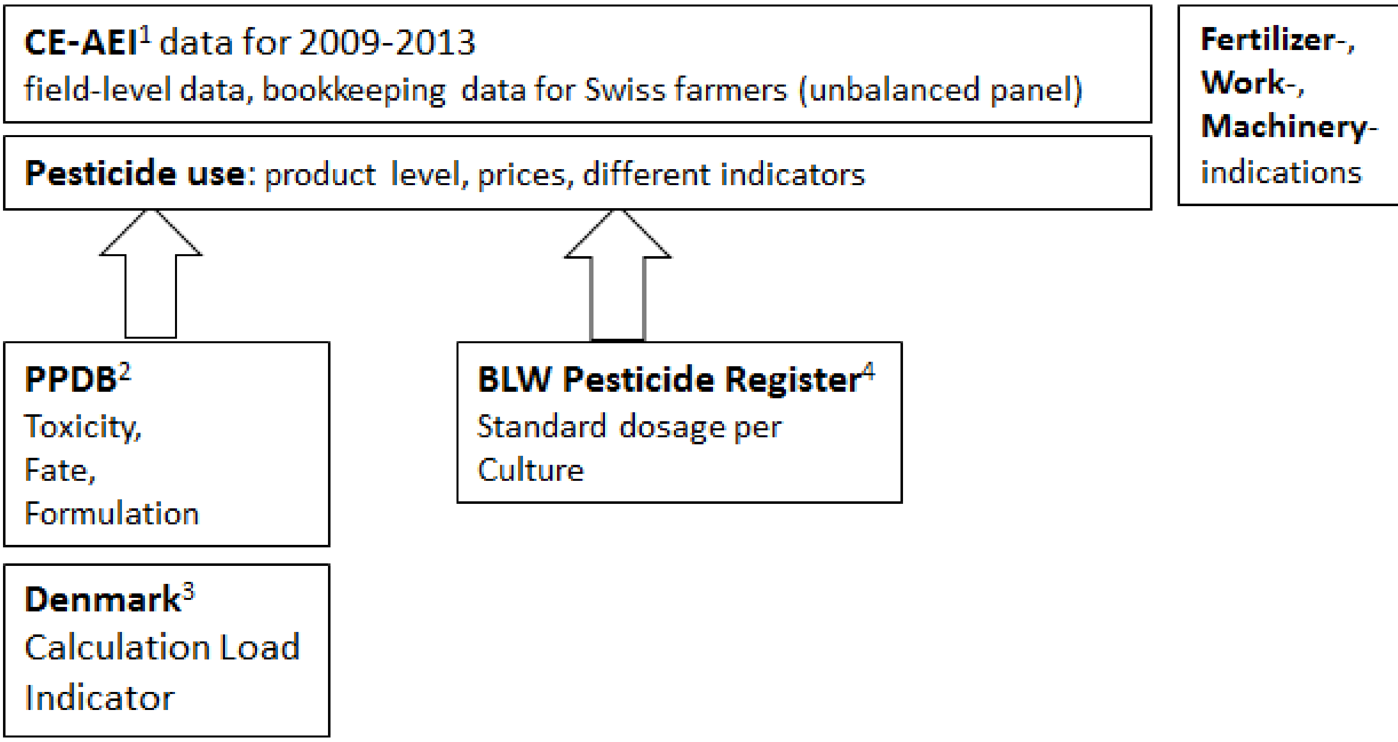


Figure A1. Linkages of used databases

Source: Own depiction. 1 : Central Evaluation of Agri-Environmental Indicators, Switzerland (2009-2013, compare: Spycher et al. 2013); 2 : Pesticides Properties Database, March, 2016 (compare: Lewis et al. 2016); 3: Pesticide Load Indicator (compare Miljøministeriet 2012); 4: BLW Pflanzenschutzmittelverzeichniss (2009-2013, compare BLW 2013).

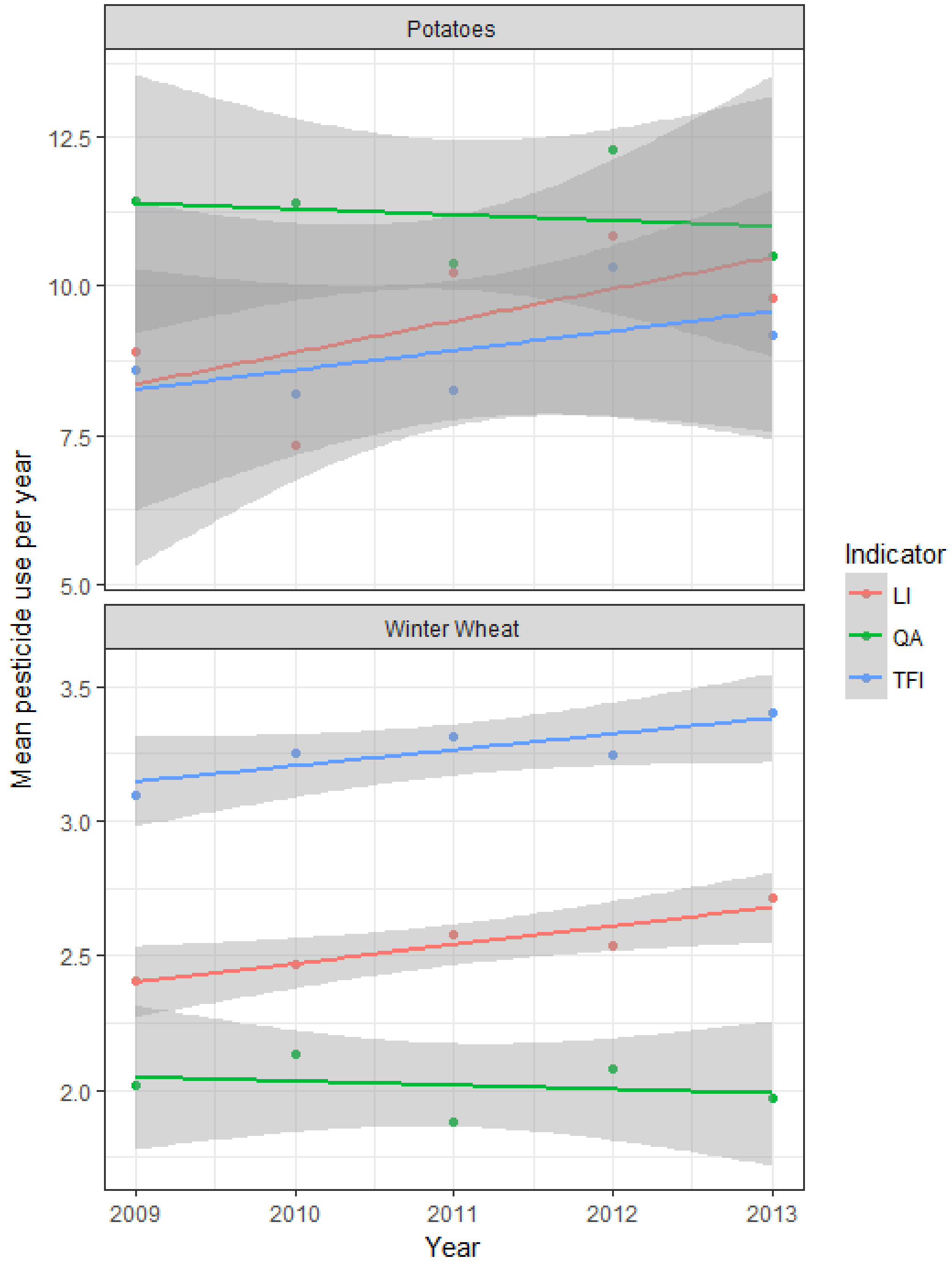


Fig. A2. Pesticide use trends shift by indicator

Linear pesticide use trends in winter wheat and potatoes from 2009-2013 with the QA, TFI and LI indicators. Linear trends are significant for TFI and LI indicators in winter wheat at the 10% and 5% level, respectively. No significant trend was found for potatoes. Shaded areas indicate 95% confidence intervals of linear trends.

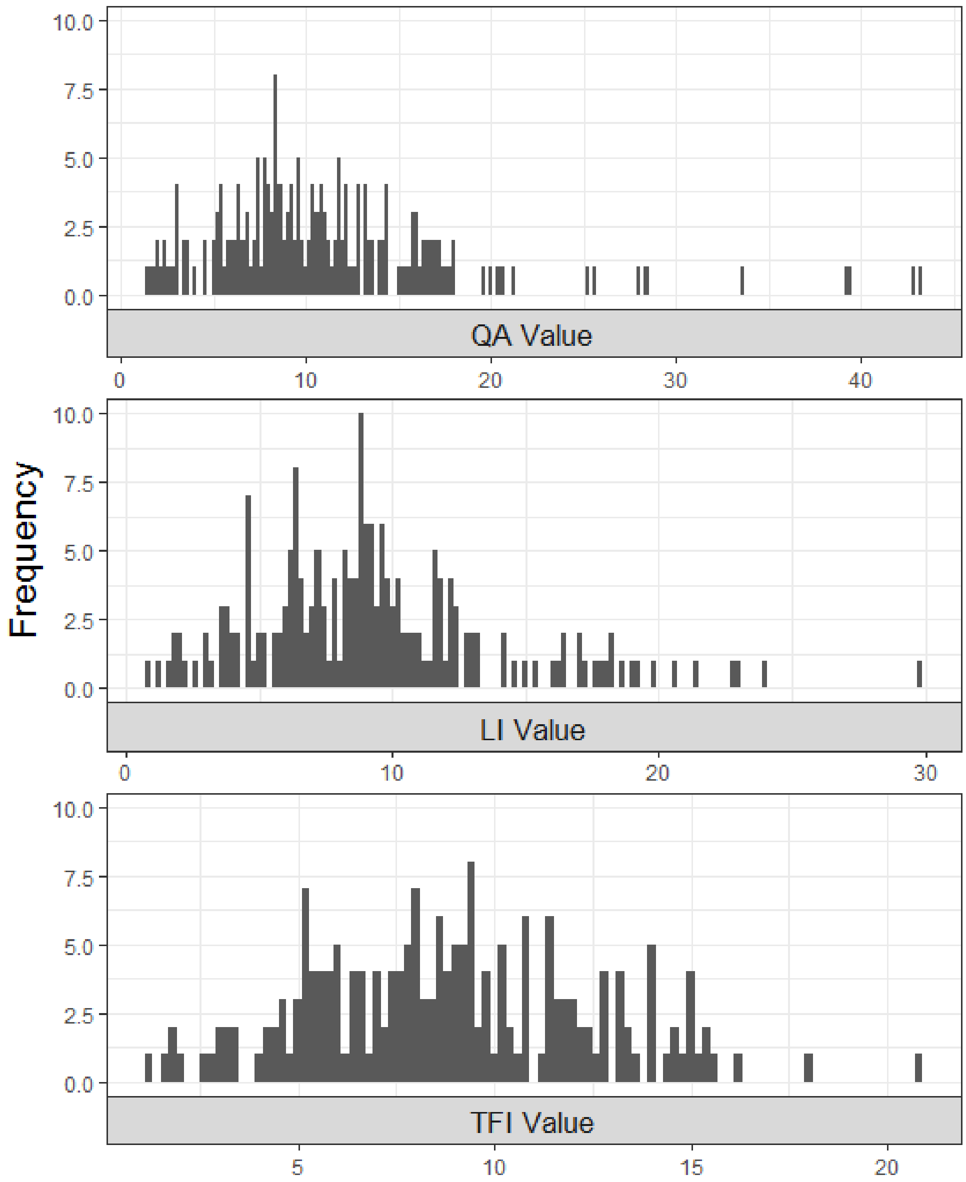


Fig. A3. Distributions of LI and QA indicators show more skewness than the TFI

Histograms of farm-level pesticide use in potato production (pooled observations from 2009-2013). QA, TFI and LI denote the Quantity of Active Ingredient, Treatment Frequency Index and Load Index Indicators, respectively.

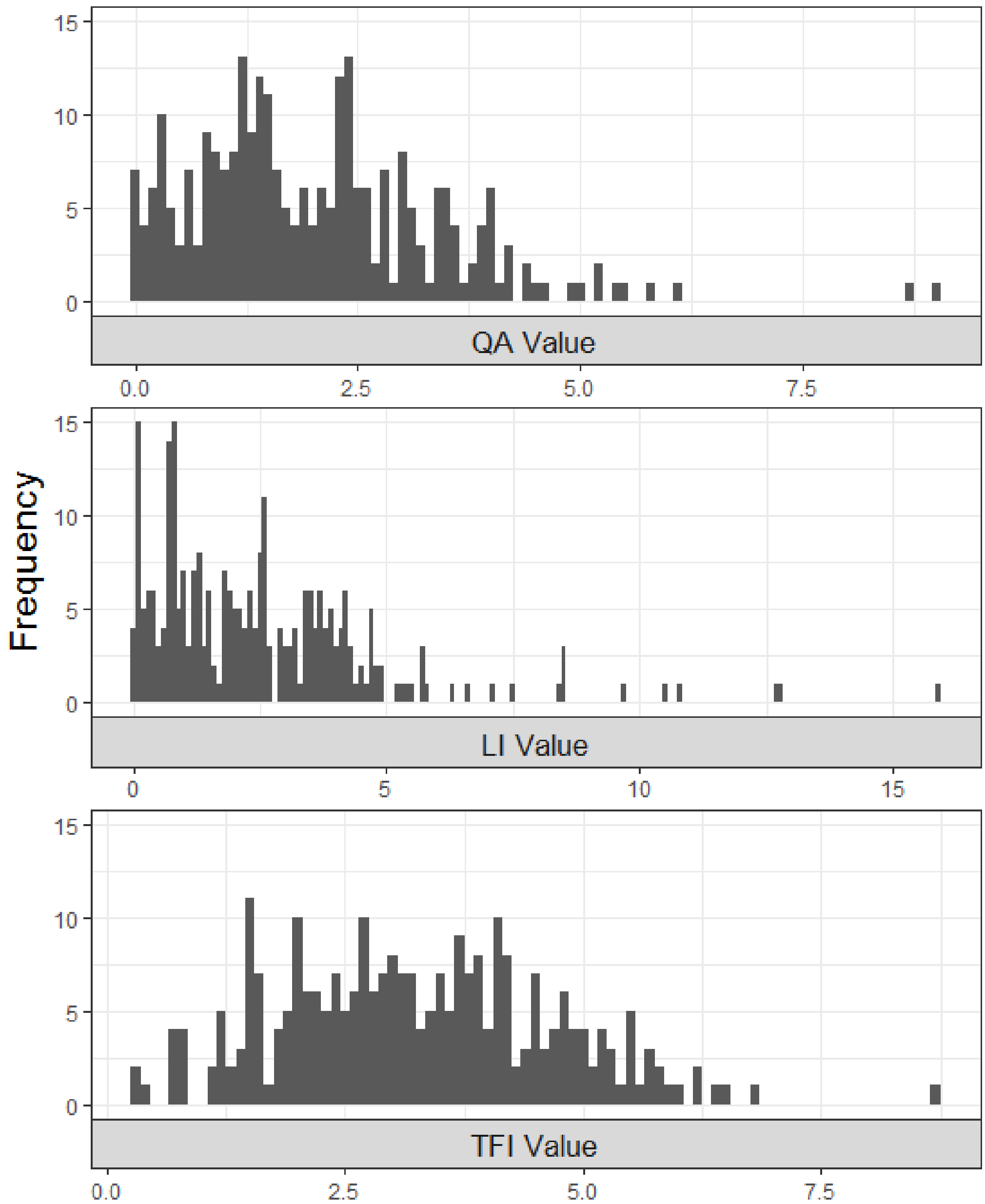


Fig. A4. Distributions of LI and QA indicators show more skewness than the TFI

Histograms of farm-level pesticide use in winter wheat production (pooled observations from 2009-2013). QA, TFI and LI denote the Quantity of Active Ingredient, Treatment Frequency Index and Load Index Indicators, respectively.

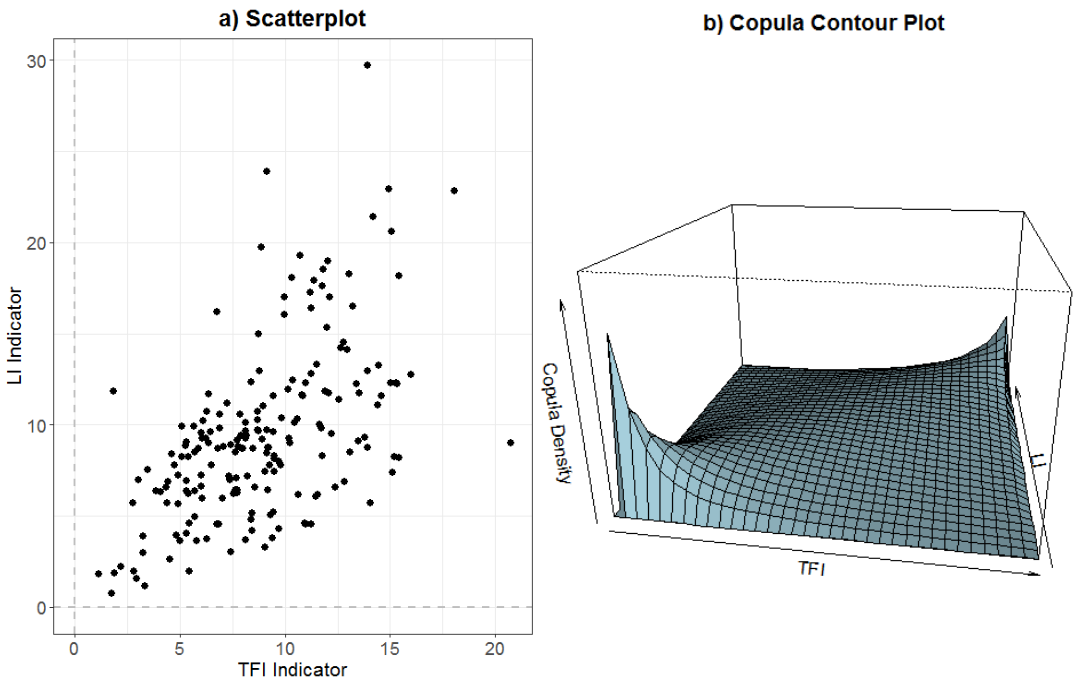


Fig. A5. Copulas reveal differences in the dependence structure of indicators between upper and lower tails (TFI and LI for potatoes).

Comparing farm-level pesticide use in potatoes measured with the TFI and LI indicators. Figure a) shows a scatterplot of TFI and LI values. Figure b) shows a contour plot of the fitted copula model (Survival BB1 copula, see Methods, Section 2.4 for selection and estimation procedure). TFI and LI denote the distribution functions for TFI and LI values, respectively (see Methods, Section 2.3). The copula captures the dependence structure between the distributions of the two indicators, not only on average, but over the whole distribution.

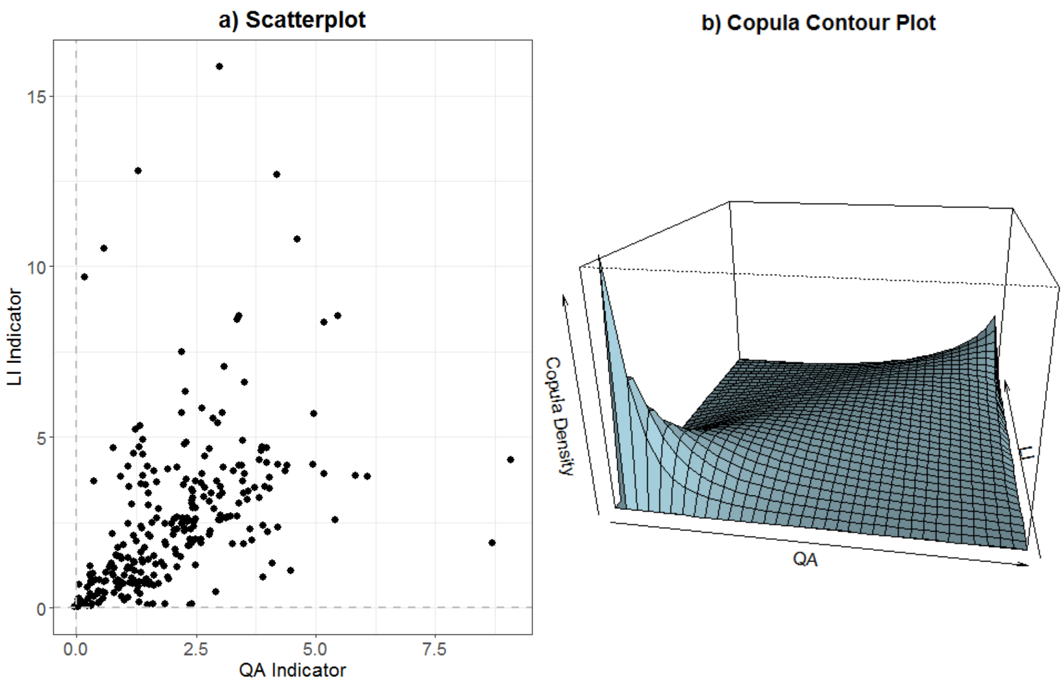


Fig. A6. Copulas reveal differences in the dependence structure of indicators between upper and lower tails (QA and LI for winter wheat).

Comparing farm-level pesticide use in winter wheat measured with the QA and LI indicators. Figure a) shows a scatterplot of QA and LI values. Figure b) shows a contour plot of the fitted copula model (Survival Gumbel copula, see Methods, Section 2.4 for selection and estimation procedure). QA and LI denote the distribution functions for QA and LI values, respectively (see Methods, Section 2.3). The copula captures the dependence structure between the distributions of the two indicators, not only on average, but over the whole distribution.

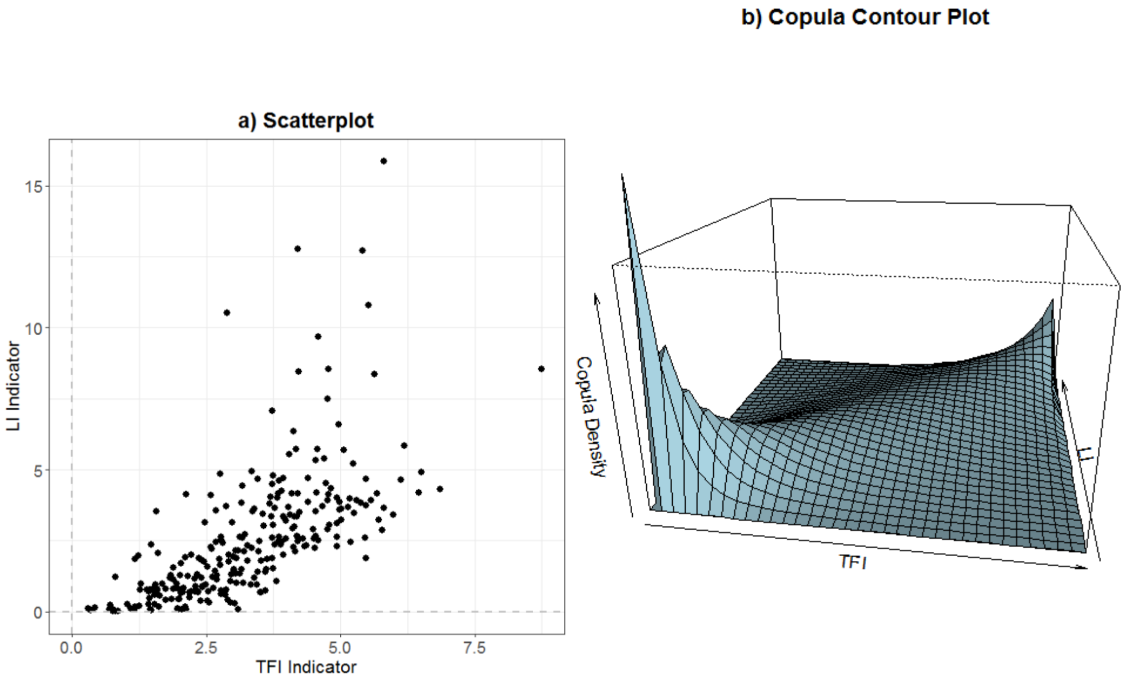


Fig. A7. Copulas reveal differences in the dependence structure of indicators between upper and lower tails (TFI and LI for winter wheat).

Comparing farm-level pesticide use in winter wheat measured with the TFI and LI indicators. Figure a) shows a scatterplot of TFI and LI values. Figure b) shows a contour plot of the fitted copula model (Survival Gumbel copula, see Methods, Section 2.4 for selection and estimation procedure). TFI and LI denote the distribution functions for TFI and LI values, respectively (see Methods, Section 2.3). The copula captures the dependence structure between the distributions of the two indicators, not only on average, but over the whole distribution.

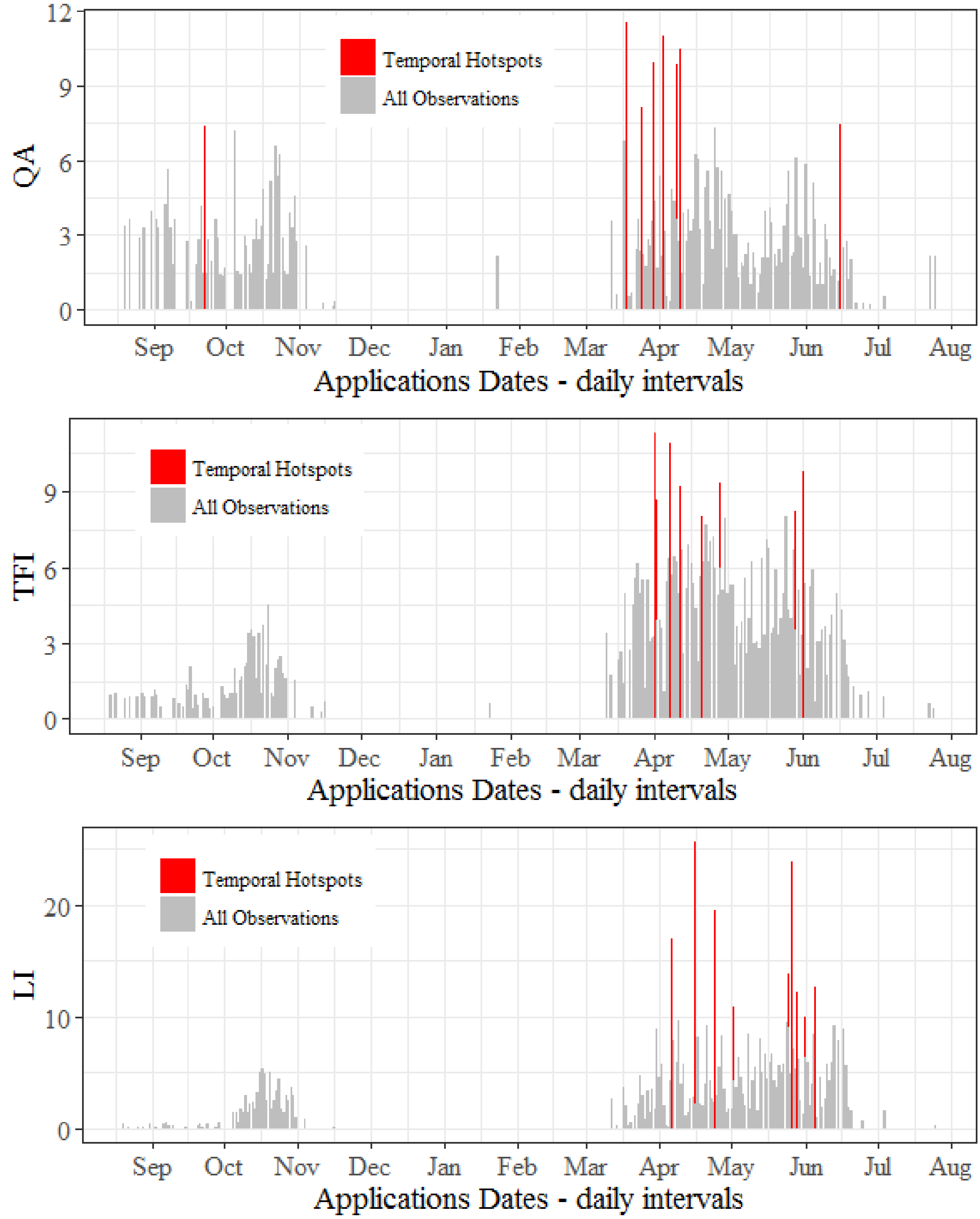


Fig. A8. Temporal hotspots of pesticide use shift with indicator choice (winter wheat).

Single pesticide applications in winter wheat, plotted on a daily scale. Observations include application data for the years 2009-2013 and all sample farmers. Pesticide use is expressed in Quantity of Active ingredient (QA), Treatment Frequency Index (TFI) and Load Index (LI) indicators respectively. “Temporal Hotspots” indicate the Top 1% of all applications per indicator, respectively.

B1. Documentation of data cleaning procedure

The following detailed procedure was used to clean data entry mistakes from the raw data:

1) *Selection*. Only farmers with complete observations (field calendar and socio-economic/bookkeeping data), only relevant cultures, only pesticide applications where all indicators are available, only farmers with a focus on arable farming (exclude those which have an area share of more than 5% wine or fruit).
2) *Remove outlier fields*. Only keep those fields where seeding <u>and</u> harvesting activity (in this order) is reported. Eliminate fields where a yield of zero is indicated or no seeding and or harvesting takes place.
3) *Remove outlier applications*. Remove applications with double accounting (same product, same day, same amount).
4) *Correct farmers' typos*. According to the following procedure: Identify applications where the product is applied with an amount of (mean +- 1*sd). For applications (mean + 1*sd) also check if the amount exceeds the recommended standard dosage. For those outliers, check every individual application in the context of the whole field calendar for the given field/year and correct if necessary. In addition, compare those applications to applications by other farmers and applications of the same farmer in other years. In particular, consider the possibility of split applications/packages/mixtures for those outliers.

Each application was tested according to the above procedure with the software package R (R Core Team 2013) to avoid removing or changing actual documented behavior and only remove individual mistakes and thus guarantee a high data quality. The procedure was designed to test the applications in the context of the given regulations, of applications of the similar products on other Swiss farms, other applications on the same farm and the data quality of information on other management measures provided by the farmer. Given the richness of the data, applications could therefore be subject to context-specific evaluation thus guaranteeing a higher reliability than simple outlier identification algorithms.

B2. Validity checks for self-reporting and self-selection

Swiss pesticide regulations explicitly forbid pesticide overdosing (application dosages higher than recommended dosages). Therefore, to address any concerns regarding self-reporting by farmers checks were carried out to verify that they complied with these regulations according to the reported data, or reported any violations of these regulations. The latter would be a sign for a high degree of trust in anonymity assurances given to farmers. Applications were rated as overdosed when the maximum allowed amount in the respective culture was exceeded by more than 5%. 5-year average values of the share of overdosed applications were then computed per culture and pesticide type. The share of overdosed applications in potato and winter wheat production is 2% and 4.8% for herbicides and 5.6% and 3.4% for fungicides, respectively. This suggests that farmers were not concerned about reporting and anonymity issues during data collection. We also compared important socio-economic characteristics of our sample to the population averages, as reported by Agroscope for 2013 (Hoop and Schmid 2014) and the Swiss Institute for Statistics for 2016 (BFS 2017). The average farm size in the sample is 26.8 ha, which is representative of Swiss crop producers with an average farm size of 26.5 ha. We also find a representative age structure in our sample. The percentage of farmers in a given age group (population averages in parenthesis) is as follows: <39 years: 28(21)%; 40-49 years: 30(30)%; 50-59 years: 35(37)% and >60 years 7(13)%. Finally, we found the ratio of farmers with no education lies at 15% compared to 2% as the population average. The higher rate of educated farmers, compared to population averages, might be explained by the focus on crop producing farms in our sample, compared to a high share of milk producing farms in the population. Moreover, our dataset does not cover very remote regions, as crop farms are located in more accessible regions (compare Figure 2).

B3. Assessing robustness of farm-level results for copulas and tail dependence

As described in the Methodology Section, step v) of the article, robustness was also checked regarding the results obtained for copula estimation and tail dependence with a parametric bootstrap test and the stable tail dependence function, respectively. The first robustness check was a goodness-of-fit test based on Kendall's process as described in Genest et al. (2006). Cramer-von-Mises test statistics were used to assess if the null-hypothesis, that the chosen copula family matches the empirical copula, is rejected or accepted at the 5% level (Genest et al. (2006) report a higher test power for Cramer-von-Mises test statistics than for Kolmogorov-Smirnov test statistics). Genest et al. (2006) and Genest and Rémillard (2008) reported that test power for some of the copula families is low. Furthermore, it was not possible to identify a power study for the two-parameter copulas used to assess tail dependence in this article. However, the results of the test still give some indication of whether the distributional assumptions of the selected copulas are met, or if further analysis is needed.

Table B1. Results of copula goodness-of-fit tests

Potatoes

		Survival Gumbel	Survival BB1	Survival BB7
Herbicides	QA-LI	Reject	Reject	Reject
	TFI-LI	Accept	Accept	Accept
Fungicides	QA-LI	Accept	Accept	Accept
	TFI-LI	Accept	Accept	Accept
All Pesticides	QA-LI	Accept	Accept	Accept
	TFI-LI	Reject	Accept	Reject

Winter Wheat

		Survival Gumbel	Survival BB1	BB7
Herbicides	QA-LI	Accept	Accept	Reject
	TFI-LI	Reject	Accept	Accept
Fungicides	QA-LI	Reject	Reject	Reject
	TFI-LI	Reject	Reject	Reject
All Pesticides	QA-LI	Accept	Accept	Reject
	TFI-LI	Reject	Reject	Reject

Note: The goodness-of-fit test described in Genest et al. (2006) was used. Cramer-von Mises test statistics were used to accept/reject the null-hypothesis of a matching copula family at the 5% level. P-values were computed according to the parametric bootstrap procedure described in Genest et al. (2006).

Table B1 reports outcomes of the goodness-of-fit test described by Genest et al. (2006) for all indicator pairs and copula families for farm-level pesticide use. With the exception of four copula pairs (QA-LI for herbicides in potatoes, fungicides in winter wheat, TFI-LI for all pesticides in winter wheat) goodness-of-fit for the best fitting copula is confirmed. The semi-parametric approach used to assess tail dependence in this article is expected to show a bad performance if distributional assumptions are not met (Frahm et al. 2005). In addition, the fully non-parametric stable tail dependence function (Kiriliouk et al. 2016) was used to assess tail dependence and test if the results on tail dependence remain valid for those four indicator pairs.

Computation of the stable tail dependence function was implemented in accordance with Kiriliouk et al. (2016) and using the R package copBasic (Asquith 2017). The function was estimated for the following indicator pairs:

i) QA and LI indicator for herbicides in potato production
ii) QA and LI indicator for fungicide use in winter wheat production
iii) TFI and LI indicator for fungicide use in winter wheat production
iv) TFI and LI indicator for overall pesticide use in winter wheat production

More specifically, it was checked to verify that the statement of a low degree of upper tail dependence for i) to iv) respectively, can be supported by the results of the stable tail dependence function method. To this end, the stable tail dependence function was plotted for all indicator pairs, as suggested by Kiriliouk et al. (2016). This allows qualitative confirmation or rejection of findings on tail dependence. The approach consists of plotting several levels of the stable tail dependence function, where levels are ordered according to their relative distance to the upper endpoint of the joint distribution of the indicator pairs. The stable tail dependence function $l(x)$ is defined as follows.

$$l(x) = \lim_{t \downarrow 0} t^{-1}\, \mathbb{P}[F_1(X_1) > 1 - tx_1, \dots, F_d(X_d) > 1 - tx_d]$$

where $X = (X_1, \dots, X_d)$ is a random vector with continuous marginal distribution functions and a joint distribution function. $\mathbb{P}\,(X_1 \le x_1, \dots, X_d \le x_d),\ t > 0$ is small and the numbers $x_1, \dots, x_d \in [0, \infty)$ parametrize the relative distances to the upper endpoints of the d variables (in this case two, namely the respective indicator pair). Tail dependence is therefore investigated close to the upper endpoints (1,1) of the joint (bivariate) distribution. An example of how to interpret plots of the stable tail dependence function is given in Figure B1 and Figure B2. These illustrate plots of the stable tail dependence function for several levels of two Gumbel-Hougaard copulas with theoretical upper TDCs of 0.78 and 0.01 indicating strong upper tail dependence and almost no upper tail dependence, respectively.

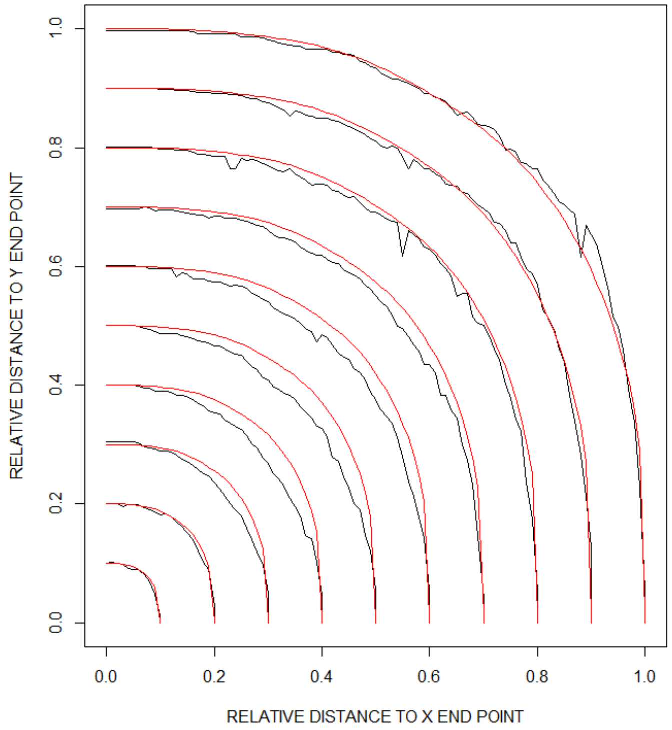


Figure B1. Example plot of the stable tail dependence function for a Gumbel-Hougaard copula with a TDC of 0.78

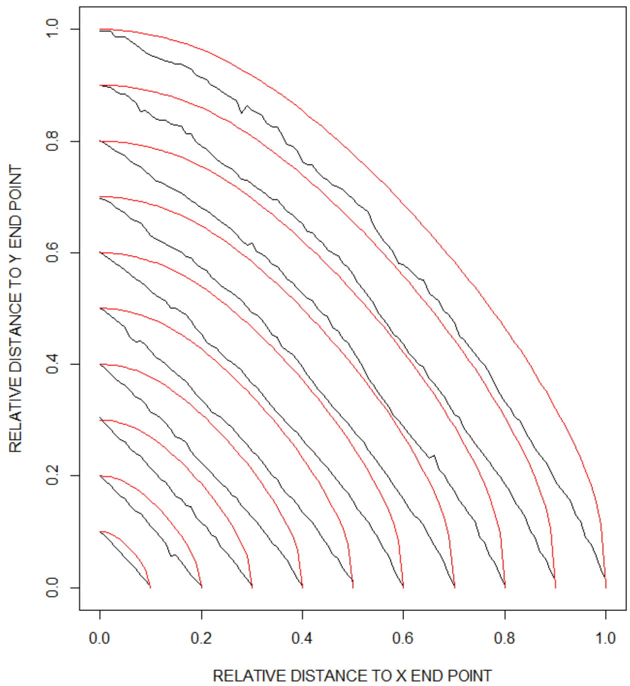


Figure B2. Example plot of the stable tail dependence function for a Gumbel-Hougaard copula with a TDC of 0.01

Level sets would show 90° bends and straight lines, respectively for the two extreme cases of complete asymptotic dependence and asymptotic independence. Given this indication, the upper figure clearly depicts the high degree of tail dependence present, whereas the lower figure points towards (asymptotic) independence, as expected. Figures B3, B4, B5 and B6 show plots of the stable tail dependence function for the indicator pairs i) - iv) , respectively.

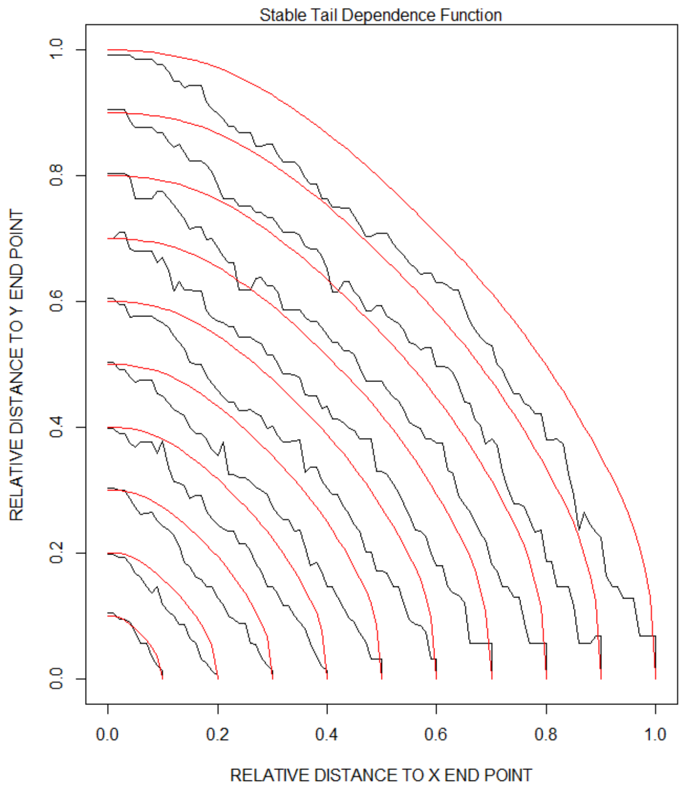


Figure B3. Plot of the stable tail dependence function for indicator pair i)

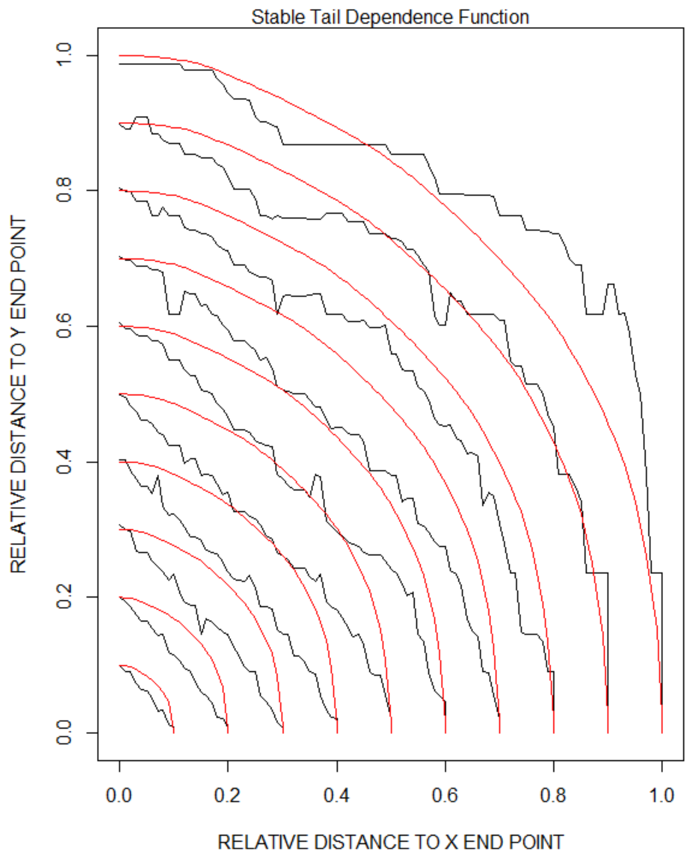


Figure B4. Plot of the stable tail dependence function for indicator pair ii)

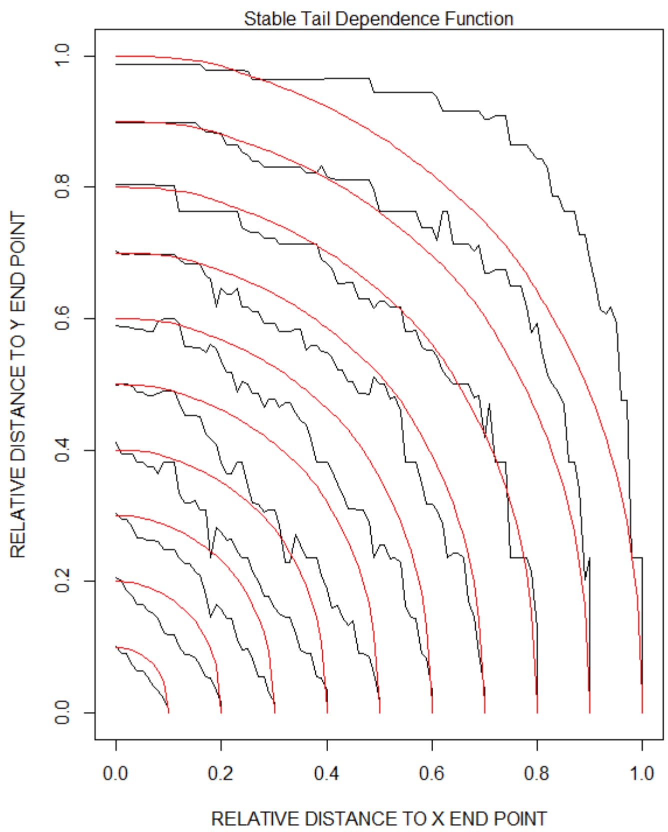


Figure B5. Plot of the stable tail dependence function for indicator pair iii)

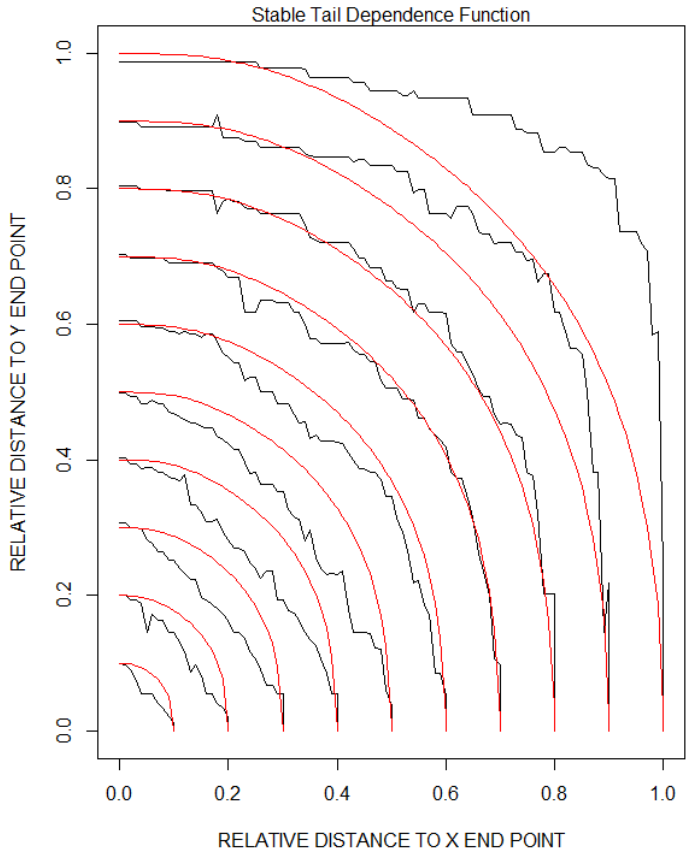


Figure B6. Plot of the stable tail dependence function for indicator pair iv)

The plots clearly indicate a low level of tail dependence for indicator pairs i) to iv) as the level sets are close to straight lines. However, level sets of indicator pair iv) get closer to the curves shown in Figure A10 with a higher relative distance to the end point which indicates a higher degree of asymptotic dependence than indicator pairs i) to iii). These findings are in line with the values reported for the upper TDCs in Tables 1 and A5, where the estimated upper TDCs for indicator pairs i) to iii) were {0.00; 0.00;0.00}, {0.00;0.00;0.06} and {0.00;0.00;0.28} and {0.00;0.01;0.32} for indicator pair iv), respectively. The findings for the stable tail dependence function method are therefore in line with the findings reported in Tables 1 and A9 and qualitatively support the result indicating only a weak upper tail dependence.

Online Appenix: R-Code

```
# Niklas Möhring, 13.06.2018
# Appendix to the paper "Quantity based indicators fail to identify extreme
pesticide risks" by Möhring, Gaba and Finger
# Documentation of the computation procedure of tail dependence
coefficients and correlation coefficients in R
# A detailed description of the procedure and sources can be found in
section 3.4 of the paper


library(dplyr)
library(VineCopula)
library(copBasic)


# i) Test independence of indicator pairs analytically with independence
tests ####


# Independence tests
# Repeat for all indicator combinations, crops and aggregation levels

BiCopIndTest(dat[1]][,1], dat[[1]][,2])   # Convert observations to pseudo
observations with pobs() before testing




# ii) Estimate copula parameters and TDCs for a wide range of copula
families with a semi-parametric estimation procedure with empirical margins
####


# Repeat for all indicator combinations, crops, aggregation levels and
copula families
summary(BiCopSelect(dat[[1]],dat[[2]],familyset = xy,rotations = FALSE)) #
Convert observations to pseudo observations with pobs() before testing,
insert specific copula in familyset




# iii) Choose the three best fitting copula families over all indicator
pairs ####


# Choose the three best fitting copula families according to AIC and BIC
values, as well as graphical fit with simulated copula distributions
# Repeat for all indicator combinations, crops and aggregation levels

BiCopCompare(pobs(dat$ind1), pobs(dat$ind2))





# iv) Identify copula family with best goodness-of-fit ####

# Rank copula families according to Vuong and Clarke tests (with and
without corrections)
# Repeat for all indicator combinations, crops and aggregation levels
```

```
BiCopVuongClarke(pobs(dat$ind1), pobs(dat$ind2))# 20 - 16 -19
BiCopVuongClarke(pobs(dat$ind1), pobs(dat$ind2),correction = "Akaike")
BiCopVuongClarke(pobs(dat$ind1), pobs(dat$ind2),correction = "Schwarz")  # Based on the results of all tests identify the copula with best goodness of fit out of the three families in step iii)




# v) Check robustness of obtained copula results with a parametric bootstrap test and the fully non-parametric stable tail dependence function approach ####



# Semi-parametric bootstrap test for copula goodness of fit  based on Kendalls process (and CvM statistics)
# Repeat for all indicator combinations, crops and aggregation levels

BiCopGofTest(dat[[1]], dat[[2]], family = XY,method = "Kendall",B=100) # Test if the above chosen copula families resemble the empirical copula well based on CvM results



# Fully non-parametric robustness checks for copulas using the stable tail dependence function
# Repeat for all indicator combinations, crops and aggregation levels


stabtaildepf(dat[[1]])  # the data must contain two transformed (pobs()) columns with the indicators to compare
stabtaildepf(dat[[1]], smooth=TRUE, ploton=F, col=2)




# Estimate Kendalls tau correlation coefficients and p-values ####
# Repeat for all indicator combinations, crops and aggregation levels

cor.test( dat[[1]], dat[[2]], method = "kendall",exact = FALSE)
```