Finalising a Pesticide Load Indicator for the UK: Phase 4 Report

James Rainford

John Tzilivakis

David Garthwaite

Kathleen Lewis

K. Lewis and J. Tzilivakis, Agriculture and Environment Research Unit, School of Life and Medical Sciences, University of Hertfordshire, Hatfield, Herts, AL10 9AB UK.

J. Rainford and D. Garthwaite, FERA Science Ltd, York Biotech Campus, Sand Hutton, York, YO42 1LZ, UK

Report information

Title:	*Finalising a Pesticide Load Indicator for the UK: Phase 4 Report'*
Subtitle:	Final report
Reference:	PC0116
Date:	23 March 2023
Report to:	Department for Environment, Food and Rural Affairs (Defra)
Written by:	FERA Science Ltd and the Agriculture and Environment Research Unit (AERU), University of Hertfordshire, UK.
Status:	FINAL
Please cite as:	Rainford, J, Tzilivakis, J., Garthwaite, D., Jones, G. and Lewis, K.A. (2023) *Finalising a Pesticide Load Indicator for the UK: Phase 4 Report* Prepared for the Department for Environment, Food and Rural Affairs (Defra), UK, for project PC0116, by FERA Science Ltd and the Agriculture and Environment Research Unit (AERU), University of Hertfordshire, UK.
Related reports:	• Rainford, J, Tzilivakis, J., Garthwaite, D., Jones, G. and Lewis, K.A. (2022a) *A Pesticide Load Indicator for the UK: Final Report*. Prepared for the Department for Environment, Food and Rural Affairs (Defra), UK, for project PC0107, by FERA Science Ltd and the Agriculture and Environment Research Unit (AERU), University of Hertfordshire, UK • Rainford, J, Tzilivakis, J., Garthwaite, D., Jones, G. and Lewis, K.A. (2022b) Finalising a Pesticide Load Indicator for the UK: Technical report. Prepared for the Department for Environment, Food and Rural Affairs (Defra), UK, for project PC0107, by FERA Science Ltd and the Agriculture and Environment Research Unit (AERU), University of Hertfordshire, UK.

Executive summary

The UK Pesticide Load Indicator (PLI) is a multi-component indicator, which combines data on the usage of different pesticide active substances in UK agriculture with information on their propensity to persist, bioaccumulate, or be lost via surface run-off or leaching, as well as information on their relative toxicity to wildlife. Data are derived from the UK Pesticide Usage Survey (PUS) and the Pesticide Properties Database (PPDB). The PLI supplements traditional metrics such as the ‘total mass of pesticides applied’ and the ‘total area treated’ by considering the changing mixture of different substances applied through time and the effect of their varying chemical or biochemical properties. The PLI consists of 4 environmental fate and 16 ecotoxicity metrics. It does not quantify harm or reflect environmental outcomes, as it does not account for any mitigation practices or calculate exposure of real wildlife populations. Instead, the aim of the PLI is to illustrate relative trends in the potential pressure on the environment arising from the use of pesticides, to help inform UK policy decisions and the assessment of policy intervention.

The PLI was originally adapted from an indicator developed by the Danish government, but it has since gone through multiple phases of revision and alignment to the UK policy context. This report outlines the results of Phase 4 (conducted 2022-23), which focused on:

a) Making the scope of the PLI calculation explicit through development of a protocol for assessing a substance’s suitability for inclusion. This permits the inclusion of biopesticides and micro-organisms within the indicator alongside more conventional pesticide treatments.
b) Removal of the previous aggregation step and its replacement with an approach which shows information on relative trends across all 20 individual metrics.
c) Revisions and improvements to the visualisation tool based on the requirements of Defra policy teams.
d) Simplification, streamlining, and documentation of backend calculations and processes, allowing these to be easily maintained on an ongoing basis.
e) Consideration of how the PLI might integrate with other reporting around pesticide usage, including the recently proposed Total Applied Toxicity (TAT).

In addition to highlighting the changes made to the indicator and serving as a revised reference document for the calculation, this report also examines cases where the PLI might be used in practice. The PLI was developed for a wide range of uses including characterising trends in load within the UK landscape and examining the impact of changes in policy such as the approval or withdrawal of active substances or products and initiatives such as those recently introduced under the Sustainable Farming Incentive scheme (SFI).

With respect to general trends in load within arable crops, the overall trend in total mass of active substances applied per hectare of cropped land increased by approximately 16% between 2010 and 2018 and then decreased by 23% between 2018 and 2020. The 20 PLI metrics calculated over the same period show different trends, both to one another and to that suggested by overall mass of pesticide applied. Most of the fate metrics (except for drain flow) track the trend in mass applied between

2010 and 2018. However, the majority of the ecotoxicity metrics show declining trends in load over the same period, which contrasts with the trend in mass applied and may reflect a shift towards the use of less toxic active substances. Whilst the majority of PLI metrics showed a decline between 2018 and 2020, matching the trend in mass applied, the magnitude of the changes varied substantially and there were contrasting trends for some of the ecotoxicity metrics. For example, there was a very large proportional decline in load on bees (which decreased by 77% and 98% for contact and oral exposure respectively); no statistical change for load on acute toxicity for birds and mammals; and increased load on parasitic wasps. The ability of the PLI to characterise these different trends, and to link change to the contribution of specific active substances, greatly enhances the ability of policy makers to understand the impact of actions and to support the development of targeted policy. How the PLI responds to policy change has been explored in detail in this report with the example of the withdrawal of neonicotinoid seed treatments on oilseed rape.

The PLI is one of a broader family of pesticide indicators which explore changes in the mixture of pesticides applied over a given geographic scope and time. One similar indicator is the recently published TAT (a novel international standard for reporting of the potential impacts of pesticides), which although not identical shows sufficient similarity with the PLI to draw comparisons and explore the possibility of alignment. The key difference between the PLI and TAT lies in the underlying data sources for the ecotoxicity assessment and how these are aggregated. In this respect, the PLI has a more transparent link to the underlying measurements, whereas the TAT aggregates across multiple species and study types. Despite these differences, the two indicators are sufficiently similar to be calculated within the same framework, and should a policy need emerge, the TAT could be calculated and displayed alongside the PLI.

The goal of developing the PLI was to provide an exploratory tool that facilitates access to improved information about the potential environmental impacts associated with pesticide use and to provide a tool for exploring relative tends associated with the changing mixture of active substances applied over time. When compared to previous monitoring efforts, which have largely been dependent on the total mass of pesticide applied (irrespective of the properties of active substances), the PLI has been highly successful in adding a greater resolution and ‘colour’ to the discussion around the potential effects of pesticide usage and the potential impacts of policy intervention. Phase 4 marks the transition between the PLI as a research prototype to a stable deployed system. The PLI has undergone a substantial realignment from an initial application of the Danish system to something that is much more targeted and aligned to the needs of UK policy. It fills an important gap in the current UK reporting system with respect to pesticide usage and it is likely to be a useful component of decision making going forwards. While work on the indicator will continue, particularly with the annual inclusion of further PUS surveys, the core processes of the PLI are now established to the point where the indicator is ready for routine operational deployment, as a part of a wider suite of indicators that reflect different elements of the socio-economic context and decision-making processes around pesticide use.

Contents

1.0. Introduction

1.1. Background

Pesticides are a key component of modern agricultural systems, associated with the maintenance of consistent yields, efficiency of farm labour and support for food security (Cooper and Dobson 2007; Oerke 2006). However, there is a wide body of evidence that pesticide usage also has potential negative environmental impacts (Bourguet and Guillemaud 2016; Lee, den Uyl, and Runhaar 2019; Sánchez-Bayo and Wyckhuys 2019; Sud 2020). As a result, there has been interest in the development of indicators from local to national scales to aid decision making (at farm and policy levels) and contribute towards more sustainable use of pesticides. Under the 25-year environment plan (HM Government 2018), the UK government has outlined an aim to minimise the impacts of pesticides in the environment. One of the key challenges in this space is the development of suitable metrics to track progress towards these aims and to identify potential areas for policy intervention (Rainford, Kennedy, and Jones 2021). This report outlines and discusses the UK implementation of the Pesticide Load Indicator (PLI), one of the key measures developed for UK policy (Rainford et al. 2022a) and the revisions and improvements made during the most recent phase of development (Phase 4; conducted 2022-23).

Assessing the environmental pressures, effects and impacts associated with pesticide use is a challenging topic, due to the complexity of the systems involved and the diverse and evolving composition of agrochemicals applied (Milner and Boyd 2017; Schäfer et al. 2019). There are hundreds of different pesticides (some of which breakdown to other compounds, known as metabolites) which have different physical and chemical properties that affect how they behave in the environment and their toxicity to different wildlife species. The scope for capturing this complexity within an assessment is driven by data availability. At the farm level, detailed data on various risk and mitigation factors are theoretically available including the pesticides applied (and their properties), application timing, equipment, crop stage, soil types, soil and weather conditions, buffer zones, surrounding habitats (including water bodies), etc. These factors can be used to determine the likely amounts of pesticides potentially lost from the point of application, to where, and consequent effects and impacts. However, at a regional or national level, much of these data are not systematically collected; hence national/regional datasets are usually limited to amounts of pesticides applied (and their properties) and application timing on different crops, which in the UK are usually based on a sample of holdings.

Agrochemical usage measures fail to account for the shifting composition of pesticides and associated changes in pressure on the environment (Kudsk, Jørgensen, and Ørum 2018; Möhring et al. 2020; Möhring, Gaba, and Finger 2019). As a result, they can present a limited picture if used as a proxy for environmental impact. For example, using quantity applied alone would suggest that the use of a substance with high toxicity in low quantities has a lower impact than a less toxic substance used in large quantities, which may not be the case. As a result, there is a need to develop better indicators which account for the changing use and composition of pesticides. This is the key motivation for the development of the PLI.

1.2. Concept and definition of pesticide load

The Pesticide Load Indicator (PLI) approach draws upon data available at a national scale on usage and properties of pesticides. It aims to reflect the amounts of pesticides used; their potential to cause damage to non-target organisms; and their propensity to persist, bioaccumulate and be lost via run-off or leaching. This approach does not quantify losses to the environment or harm and thus load can only be interpreted as a relative unitless indicator of potential pressure on the environment.

In a wider context, it is important to acknowledge that load is not necessarily a reflection of actual outcomes with respect to the environmental impact of pesticides. It only reflects the potential pressure presented by the inherent properties and amounts used of different pesticides, with reductions in load being desirable (e.g., lower use of substances that persist and/or are toxic to some species). Outcomes will also be determined by how those substances are used, variations in weather, and local mitigation factors and actions (as outlined above, such site-specific data are not collated at the regional/national level).

1.3. Origins and overview of the PLI

There are three core elements that underpin the PLI, which have a background that spans several decades. These are the Danish PLI; the Pesticide Properties Database (PPDB); and the UK Pesticide Usage Survey (PUS). An overview of each is provided below.

Danish PLI: This well characterised national indicator (see e.g. Kudsk, Jørgensen, and Ørum 2018) was the starting point for the development of the method for the UK. It was developed as part of the Danish pesticides strategy and is used to support their pesticide tax system (Finger et al. 2017; Lee, den Uyl, and Runhaar 2019; Miljøstyrelsen 2012; Pendersen, Helle, and Andersen 2015; Sud 2020). It describes the potential impact of pesticides by attributing a score to each pesticide and multiplying this by the amount applied, usually expressed as kilograms per hectare (kg ha^{-1}). It makes use of the enhanced availability of regulatory information around pesticide use and the development of accessible data compilations such as the PPDB. It provides proxies for the potential impacts of pesticides and thus can provide a more precise tool for observing change compared to previous indicators (e.g., the Treatment frequency index which preceded the PLI's use in policy; Miljøstyrelsen 2012). The Danish PLI was the original precursor to the development of an indicator for the UK which has since evolved to have a different scope and form as outlined in Section 2.

PPDB: Created and managed by the Agriculture and Environment Research Unit (AERU), University of Hertfordshire, the PPDB started in 1996 to collate and harmonise datasets from a range of data sources including regulatory dossiers, such as those from EFSA, US EPA and CRD, and peer-reviewed literature (subject to meeting scientific quality standards) (Lewis and Bardon 1998; Lewis et al. 2016). Since 2007, it has been freely available online and, under licence, as an offline MS Access database. It holds over 320 parameters for ~2500 pesticides and ~750 metabolites. Data are supported by an extensive quality management system, which includes scoring the publishing source, data traceability and consistency with other information sources and is subject to internal rolling reviews and periodic external audits. In its 25 years of existence, the PPDB has been utilised in a wide range of policy, commercial and scientific applications. For any given parameter the value reported in the PPDB (and thus used for the PLI) will be the worst-case value reported in underlying literature which meets the required scientific quality standards. In the case where multiple values are available, the value in the PPDB (and PLI) will represent either the highest value recorded in fate context or the lowest concentration which has been shown to be associated with a toxic effect (Lewis et al. 2016).

PUS: This national survey is conducted by Fera Science Ltd, the Science & Advice for Scottish Agriculture (SASA) and the Agri-Food and Biosciences Institute, Northern Ireland (AFBINI), a Non-Departmental Public Body of the Department of Agriculture, Environment and Rural Affairs, Northern Ireland (DAERA). It consists of different surveys including arable, outdoor vegetables, soft fruit, top fruit, and grassland and fodder. Usage is estimated via stratified sampling by region and farm size based on census information taken from the June survey (Defra 2022) and associated census schemes in the devolved administrations. Data are supplied on a voluntary basis using a mix of interview, electronic data collection and farm visits. Most of the component surveys which make up the PLI are conducted on a biennial basis although some, such as the grassland and fodder survey, are only conducted once every four years.

Examples of reports generated can be found on the PUS website (Fera, 2021[1]). Further details on methodology can be found in Thomas (1999) and Garthwaite (2015). Note that the PUS was developed to provide a representative sample of PPP usage in the UK for the purposes of estimating key national (and regional) statistics (area treated, and mass applied), and is not a complete census of total PPP use (particularly at a local or sub-regional scale).

1.4 History of development of the PLI

The development of the UK PLI began in 2019 and was undertaken over three phases. Phase 1 applied the Danish PLI method to the UK using data from the PPDB and the PUS (see Lewis et al. 2021). Phase 2 evolved the method to better cover the UK context and better reflect environmental load, including improvements to the load metrics and how the load score is derived, along with refinements to the utilisation of the PUS data. Phase 3 consisted of a series of workshops and focused on finalising the approach, including refinement of the procedures developed in Phase 1 & 2 and the development of a visualisation tool. Full details of the development can be found in the linked technical report (Rainford et al. 2022b). Finally, Phase 4, which is the focus of this report, has focused on further refinement of the PLI including:

1. The development of a substance inclusion protocol to provide a systematic and transparent approach for deciding which substances are included (or excluded) from the PLI (see Section 2.3).
2. The removal of the aggregation step from the PLI (see Section 2.6).
3. Development of the visualisation tool to enhance functionality and better meet policy needs (see Section 2.7).

[1] https://pusstats.fera.co.uk/home

4. Comparison of the PLI to other related indicators (e.g. the recently developed Total Applied Toxicity indicator (Schulz et al. 2021 see Section 4).

This report also contains reworked examples of the application of the PLI (from Phase 3) using the revised approach. While much of the technical approach remains consistent with that documented in Phase 3, the revisions and the shift in focus necessitate a reinterpretation of previous findings (outlined in Section 3).

2.0. Methods

2.1. Overview of the approach

Figure 2.1 provides an overview of the processes involved in generating the PLI[2]. In summary these are:

1. Load metrics: defining a set of underlying metrics which represent the key factors contributing to the concept of load. They include environmental fate and ecotoxicity properties for each pesticide obtained from the PPDB (see Section 2.2).
2. Standardisation: to ensure that the various underlying metrics (which have different units of measurement) are expressed in a consistent fashion within the PLI, a process of standardisation is needed to normalise the load metrics onto a common scale and to define the structure of relative values between different active substances. The process generates a load score for each substance on a scale of 0 to 1 (low to high load) (see Section 2.4).
3. Multiplication: to account for the amount of active substances used, the standardised load scores are multiplied by the amount (kg) of each active substance used, derived from the PUS (an estimate of the amount applied for a given period). When a pesticide has low persistence (i.e., a half-life <1 day), the principal metabolites and their fate and ecotoxicity properties are used instead. The reported value of each metric is the sum of the estimated load for each active substance (see Section 2.5).
4. Visualisation: the outputs from steps above need to be visualised to aid interpretation and communication. This has been done via the development of a bespoke online application/dashboard (see Section 2.7).

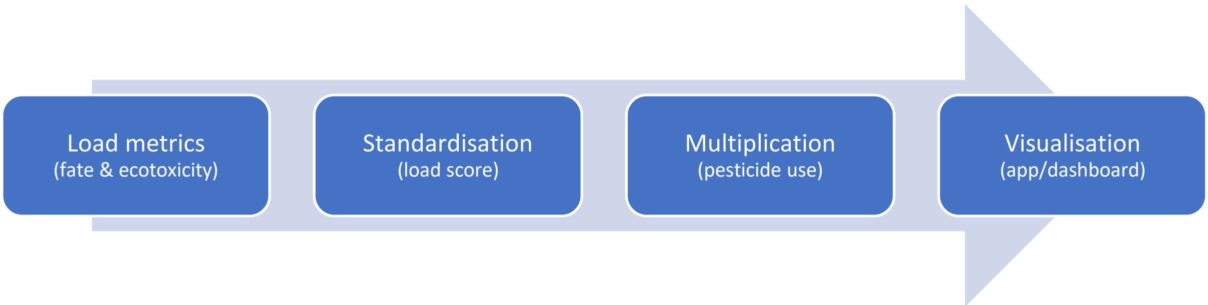

Figure 2.1 Overview of the PLI process

[2] In previous versions of the indicator (e.g., Rainford et al., 2022a), after standardisation, the metrics could be aggregated via the use of aggregation constants to create higher level headline values. The use of aggregated metrics in the PLI has proved to be problematic in terms of development and has been removed from the current version in favour of assessing percentage change in each metric independently (see Section 2.6 for discussion).

2.2. Load metrics

The PLI consists of 4 environmental fate and 16 ecotoxicity load metrics, defined in Tables 2.1 and 2.2 respectively.

Table 2.1: Environmental fate load metrics[3]

Fate metric	Definition
Soil DT_{50} (Persistence)	The propensity for a substance to persist in the environment. This is measured using the DT_{50} value, which is the time required (in days) for the chemical concentration under defined conditions to decline to 50% of the amount at application. DT_{50} values measured under field conditions are the preferred values to use, but when these are not available the DT_{50} values measured in a laboratory are used.
Drain flow (Surface water mobility)	The propensity of a substance to move through the environment and be lost to surface water (e.g., via drain flow). This is measured using the organic carbon sorption coefficient for a substance. The non-linear (K_{foc}) values are the preferred, but when these are not available the linear (K_{oc}) values are used. This load metric is only calculated when a substance is applied during the drain flow period (September to April). Applications outside of this period are given a load score of zero.
Mobility (Groundwater)	The propensity of a substance to move through the environment and be lost to groundwater (e.g., via leaching). This is measured using the Groundwater Ubiquity Score (GUS) (Gustafson, 1989) which is calculated using the DT_{50} and K_{oc} (or K_{foc}) values.
Bio-concentration factor (BCF)	The propensity of a substance to concentrate within the tissues of organisms i.e., the accumulation of pollutants (litres kg^{-1}) through chemical partitioning from the aqueous phase into an organic phase, such as the tissue of a fish. When BCF data do not exist for a substance, it can be calculated using the log 10 value of the octanol-water partition coefficient.

[3] Under the revised data inclusion protocol, the scope of calculated fate metrics has been changed so as to include any substances identified as inorganic (as these are subject to different processes in their movement and persistence in the environment relative to conventional organic based PPP). See Section 2.3 for further discussion.

Table 2.2: Ecotoxicity load metrics

Ecotoxicity metric	Definition
Algae EC_{50} (acute toxicity)	The short-term toxicity to green algae via aqueous exposure. Measured as the EC_{50}[a] (mg L^{-1}) over 72 hours.
Aquatic plants EC_{50} (acute toxicity)	The short-term toxicity to higher aquatic plants (e.g., duckweed *Lemnoideae*) via aqueous exposure. Measured as the EC_{50}[a] (mg L^{-1}) over 7 days.
Daphnia EC_{50} short (acute toxicity)	The short-term toxicity to aquatic invertebrates (e.g., *Daphnia spp*) via aqueous exposure. Measured as the EC_{50}[a] (mg L^{-1}) of mortality over 48 hours.
Daphnia NOEC long (chronic toxicity)	The long-term toxicity to aquatic invertebrates (e.g., *Daphnia spp*) via aqueous exposure. Measured as the NOEC[b] (mg L^{-1}) of the total number of living offspring produced per parent over 21 days post exposure.
Fish EC_{50} short (acute toxicity)	The short-term toxicity to fish (e.g., Rainbow trout *Oncorhynchus mykiss*; Bluegill *Lepomis macrochirus*, Zebrafish *Danio rerio*) via aqueous exposure. Measured as the EC_{50}[a] (mg L^{-1}) over 96 hours.
Fish NOEC long (chronic toxicity)	The long-term toxicity to fish (e.g., Rainbow trout *Oncorhynchus mykiss*; Bluegill *Lepomis macrochirus*, Zebrafish *Danio rerio*). Measured as the NOEC[b] (mg L^{-1}) over 21 days.
Birds LD50 short (acute toxicity)	The short-term toxicity to birds (e.g., Mallard, Northern bobwhite) via oral exposure. Measured as the LD_{50}[c] (mg kg^{-1} bw) usually 72 hours (sometimes 7-14 days).
Birds NOEL long (chronic toxicity)	The long-term toxicity to birds (e.g., Mallard *Anas platyrhynchos*; Northern bobwhite *Colinus virginianus*) via multiple daily oral exposures. Measured as the NOEL[d] (mg kg^{-1} bw d^{-1}) of mortality and egg production over 21 days post exposure.
Earthworms LC_{50} (acute toxicity)	The short-term toxicity to earthworms (e.g., *Lumbricus terrestris*). Measured as the LC_{50}[e] (mg kg^{-1} soil) over 14 days.
Earthworms NOEC reproduction (chronic toxicity)	The long-term toxicity to earthworms (e.g., *Lumbricus terrestris*). Measured as the NOEC[b] (mg kg^{-1} soil) on the number of juveniles successfully hatched after a 4-week period following a 28-day exposure.
Bees contact LD_{50}	The short-term toxicity to bees (*Apis mellifera*) via contact exposure. Measured as the LD_{50}[c] (μg bee^{-1}) over a minimum of 48 hours.
Bees oral LD_{50}	The short-term toxicity to bees (*Apis mellifera*) via oral ingestion. Measured as the LD_{50}[c] (μg bee^{-1}) over a minimum of 48 hours.
Mammals LD_{50} short (acute toxicity)	The short-term toxicity to mammals (e.g., Rat *Rattus norvegicus;* Mice *Mus musculus*) via multiple daily oral, dermal, or inhalation exposures (depending on the substance and/or regulatory system). Measured as the LD_{50}[c] (mg kg^{-1} bw) usually over 48-72 hours (sometimes 7-14 days).
Mammals NOAEL long (chronic toxicity)	The long-term toxicity mammals (e.g., Rat *Rattus norvegicus;* Mice *Mus musculus*), typically via oral exposure. Normally measured as the NOAEL[f] (mg kg^{-1} bw d^{-1}) over 21 days.
Parasitic wasps	The short-term toxicity to parasitic wasps (e.g., *Aphidius rhopalosiphi*) via contact exposure. Measured as the LR_{50}[g] (g ha^{-1}) over a minimum of 48 hours.

Ecotoxicity metric	Definition
Predatory mites	The short-term toxicity to predatory mites (e.g., *Typhlodromus pyri*) via contact exposure. Measured as the LR_{50}[g] ($g\ ha^{-1}$) over 7 days.

a. EC_{50}: The concentration of a substance that can be expected to cause a defined nonlethal effect in 50% of the tested population.
b. NOEC: No Observed Effect Concentration. The greatest concentration of a substance, found by observation or experiment, which causes no detectable effect.
c. LD_{50}: The median lethal dose (required to kill 50% the tested population) of a substance.
d. NOEL: No Observed Effect Level: The greatest level of a substance, found by observation or experiment, which causes no detectable effect.
e. LC_{50}: The median lethal concentration (required to kill 50% the tested population) of a substance.
f. NOAEL: No Observed Adverse Effect Level. The greatest level of a substance, found by observation or experiment, which causes no detectable effect.
g. LR_{50}: The median lethal rate (required to kill 50% the tested population) of a substance.

For the chronic aquatic and terrestrial ecotoxicity load metrics, the load scores are adjusted using the water and soil persistence (DT_{50}) values for each substance respectively. Each DT_{50} is converted to a 0 to 1 coefficient (using equations in Miljøstyrelsen 2012) and the relevant chronic ecotoxicity load scores are then multiplied by this value. Chronic (long-term) exposure will be lower for substances that do not persist, and this is accounted for in the chronic load score using this approach to avoid overestimating chronic load for substances that do not persist (see Rainford *et al.* 2022, Section 2.3 for details).

The calculation of the PLI requires a complete set of data for all pesticides for all load metrics. The PPDB is a substantial data resource; however, it inevitably has some gaps in coverage (especially for some older or historically withdrawn or traditional pesticides (e.g., sulphur, peroxyacetic acid and urea and certain metabolites) and some values are unbounded (e.g., < or > values). With respect to missing data, in most instances, PPDB data coverage (for the parameters used for the 299 active substances currently included) is greater than 80%. Aquatic plants acute, worms chronic, parasitic wasp and predatory mites have the lowest coverage (66 to 70%). With respect to the proportion of PLI substances that have unbounded values for each of the ecotoxicity metrics (there are no unbounded values for fate metrics), there is a low instance of less than (<) values (which could be underestimating load); and there is a higher instance of greater than (>) values (which could be overestimating load) and are more common for acute data (compared to chronic), with birds, mammals, worms, and bees having more than 50% greater than values (see Rainford *et al*., 2022 for full details).

To generate a full set of data to calculate the PLI, a ‘missing data replacement protocol’ has been developed (see Annex 1), which includes using alternative properties where appropriate (e.g., using DT_{50} lab values when field values are missing or K_{oc} for missing K_{foc} values), using values for related substances, and, where none of the above are available, calculating arithmetic mean values for different substance types (i.e., fungicides, herbicides and insecticides) and using these as replacements. For further discussion of this protocol and reasons for its adoption see Rainford et al. (2022b).

2.3. Standards for substance inclusion and data coverage

2.3.1. Background

During Phase 4 of the development of the pesticide load indicator the decision was taken to adopt a formal protocol to determine what substances should be included (or excluded) from the PLI. To date, the PLI has been developed to cover conventional chemical pesticides, but increasingly other substances are being used for crop protection, such as biopesticides. Whether a substance should be included within the PLI depends on several factors including whether it is a single chemical substance, a mixture, or a micro- or macro-organism; and whether there is sufficient coverage of the data required for PLI load metrics (and the metric used is appropriate). To date, this has been done on a case-by-case basis using expert judgement but going forward a more transparent and systematic approach is needed. Hence, the development of this protocol.

Firstly, it is important to acknowledge where this protocol is applied within the PLI creation process, as there are interactions with the missing data replacement protocol (Annex 1). Figure 2.2 shows the process for generating data for the PLI and where the exclusion/inclusion protocol would be applied.

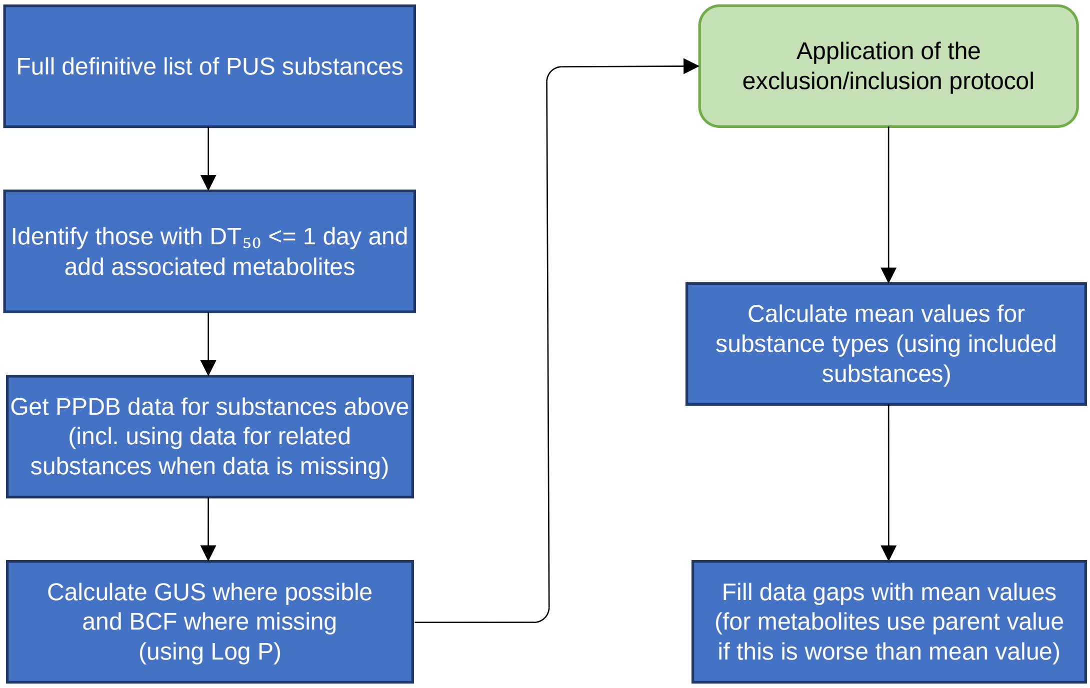

Figure 2.2: PLI data generation flow chart

In summary, a full list of substances in the PUS is collated. This involves matching to corresponding substances in the PPDB, and where a substance has a DT$_{50}$: <=1 day (low persistence), associated metabolites are identified and added to the list of substances (these substances are then used to calculate the PLI, subject to passing the inclusion criteria). GUS (groundwater mobility) is then calculated (using DT$_{50}$ and K$_{oc}$ or K$_{foc}$ values); and where BCF data are missing this is calculated using the Log

P_{ow}[4] value for the substance. This establishes where data are available or missing for each of the PLI load metrics. It is at this point that the exclusion/inclusion protocol is applied. The resulting included substances are used to generate the mean values (for different substance types: fungicides, herbicides, and insecticides), which are then used to fill any remaining data gaps in the PLI load metrics for each substance.

2.3.2. Protocol

The broad processes in the exclusion/inclusion protocol are illustrated in Figure 2.3. In summary, the first steps involve categorising the substance into one of 5 groups:

1. A single chemical substance or a mixture with single dominant substance
2. A mixture with no dominant substance
3. An inorganic substance
4. A micro-organism
5. A macro-organism

It is not possible or relevant to apply the PLI load metrics for substances which are mixtures with no dominant substance[5] or macro-organisms, so these substances are excluded. The remaining substances are then assessed to determine if sufficient data exist across the PLI load metrics. For micro-organisms and inorganic compounds, only the Ecotox load metrics are considered as the Fate metrics do not apply or are less relevant.

In the case of active substances (for single or dominant chemicals only) with a low persistence (DT_{50}: <=1 day), their principal metabolites (and associated properties) are used for the PLI. These metabolites are subject to the same protocol as single chemicals. If all the metabolites are excluded, the parent substance is also excluded. Some inorganic compounds also have DT_{50}: <=1 day, but do not have any metabolites. In these instances, the properties of the parent inorganic compound are used for the exclusion protocol.

The next step of the protocol is to determine if sufficient data are available. This can be done by setting a threshold for data coverage i.e., the number or percent of the load metrics for which data exists for a substance. Setting thresholds for data coverage is an aspect that can be subject to debate. If the coverage thresholds are set very high, then this will significantly limit the number of substances included within the PLI; whereas if the thresholds are too low, then the PLI will contain more substances that use mean data to replace data gaps, thus the PLI may become less meaningful. As a guide, out of 295 substances, if the threshold is set to 100% (i.e., a substance must

[4] Log P_{ow} (or Log P) is the logarithm (base-10) of the partition coefficient between n-octanol and water. It is used in environmental fate studies and large values (+4 or higher) are regarded as an indicator that a substance will bio-accumulate.

[5] These are substances that are a mixture of many (sometimes hundreds) of compounds where the active is not known; the pesticidal properties may be due to multiple compounds; and/or the composition is highly variable. Thus, it is not possible to derive data on fate and ecotoxicity for these substances. Note that this is distinct from the use of the term mixtures to mean formulated products, which are appropriately handled in the PLI by being separated into their constituent active substances during calculation.

have data for all metrics), 208 substances would be excluded; at 90% it is 112; and at 80% it is 79.

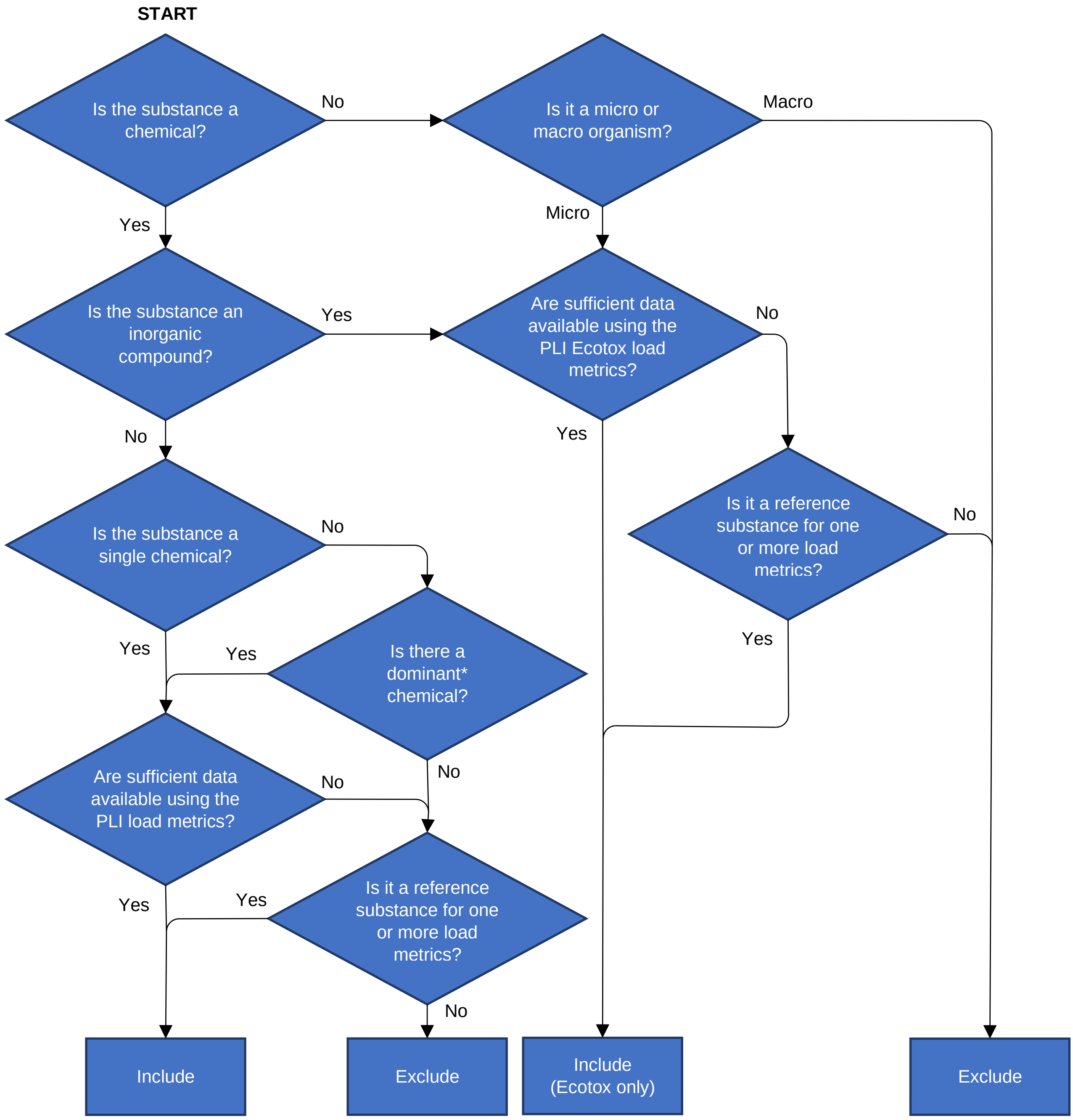

Figure 2.3: Exclusion/inclusion protocol flowchart

* Some substances, such as biopesticides (e.g., essential oils and botanical substances), are not comprised of a single chemical but may be made up of many chemicals. The percentage composition in terms of the individual substances is highly variable depending on the source of the substance, including the crop variety, environmental conditions, and how/where it is grown. However, if one substance within the complex mixture is dominant, it can be used as a surrogate chemical. For example, D-limonene in orange oil accounts for 94-97% of orange oil and so could be used in the PLI as a surrogate (if sufficient data exists). What value (%) defines a 'dominant' substance can be debated, but it proposed that a value of >80% is used for the draft protocol.

Setting the data coverage threshold can simply be based on coverage of the 20 load metrics. However, it could also be set so that there are separate thresholds for Fate and Ecotox metrics; for Ecotox (acute), Ecotox (chronic) metrics; and for any other groups of metrics. An initial set of threshold options (Table 2.3) has been set up to explore how these operate.

Table 2.3: Example data coverage threshold options

Options	Thresholds that can be set
All data	Coverage (%) of the 20 load metrics
Fate and Ecotox	Coverage (%) of the 4 Fate metrics and Coverage (%) of the 16 Ecotox metrics
Fate, Ecotox (acute) and Ecotox (chronic)	Coverage (%) of the 4 Fate metrics; 11 Ecotox (acute) metrics; and 5 Ecotox (chronic) metrics

This facilitates increasing levels of stringency that can be applied to exclude substances. For example, increasing stringency would be achieved ranging from 60% of all data; 60% of both Fate and Ecotox metrics; or 60% of fate, Ecotox (acute) and Ecotox (chronic) metrics. It also facilitates the application of thresholds for different substance types, e.g., for micro-organisms where Fate data are not applicable. It is possible to apply variable thresholds for different metrics. For example, a lower threshold can be set for Ecotox (chronic) metrics (given these are not available for many substances).

Figure 2.4 shows the quantity (tonnes used, summed for all surveys and years) of excluded substances as the data coverage threshold for different metrics (and combinations thereof) increases from 50 to 100%. It shows how as the stringency increases, the mass of substances excluded increases, with notable steps across the criteria. This is usually where a substance with high usage gets excluded. For example, the first large step at 64% is where 2,500 tonnes of Mancozeb gets excluded. This has 3 metabolites ethylenethiourea, ethyleneurea and etem which have Ecotox (acute) data coverage values of 63.6%, 54.5% and 36.4% respectively, thus when this criterion goes above 64% none of the metabolites are included and, thus, the parent gets excluded.

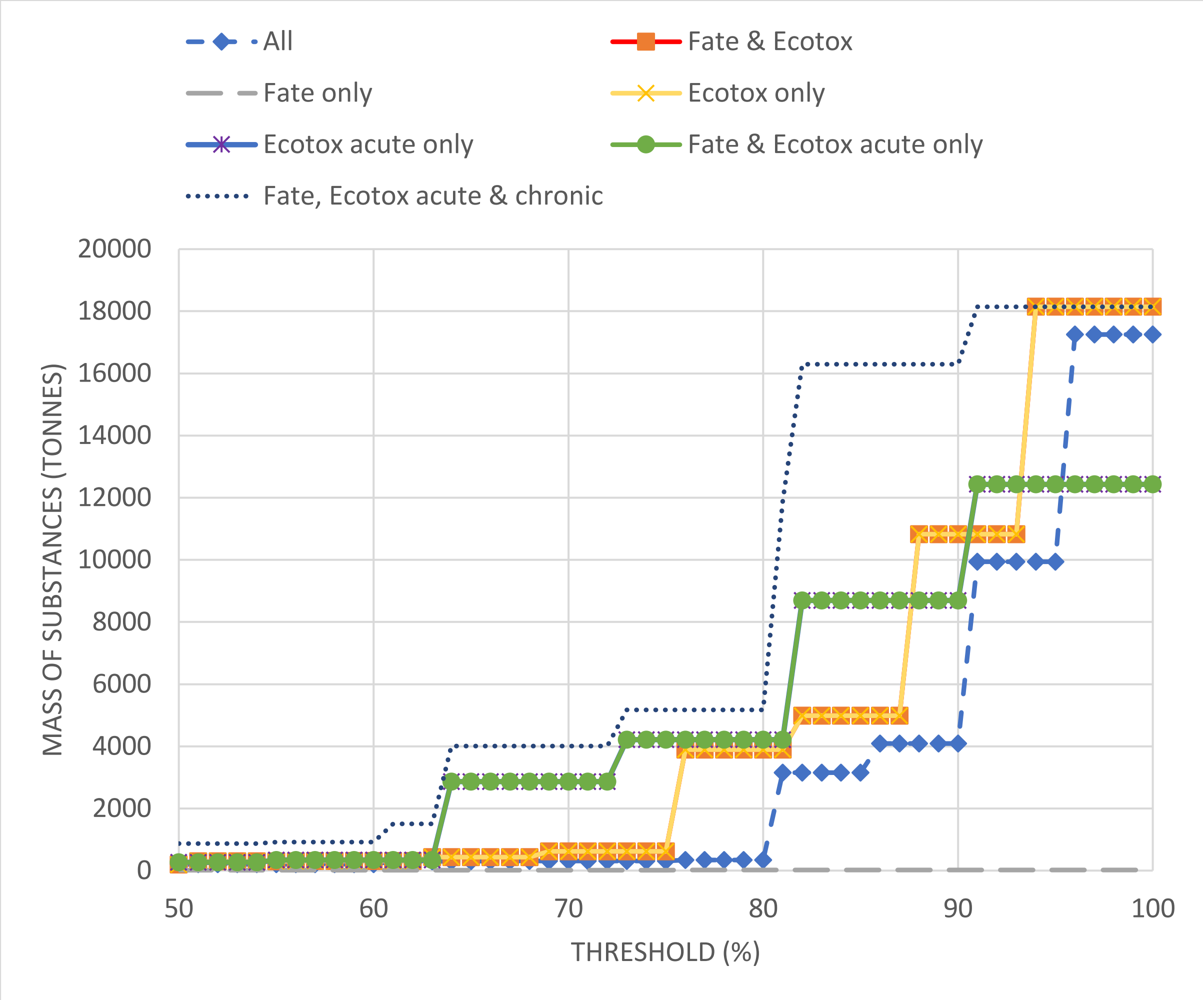

Figure 2.4: Quantity of active substances excluded at different data coverage thresholds

Table 2.4 shows some example threshold options following this approach and the corresponding number and mass of substances that are excluded using each option (for 295 active substances +33 metabolites). Option 1 requires that a substance has data for at least 60% of all load metrics; Option 2 requires data for at least 60% of fate and 60% of Ecotox metrics; Option 3 requires data for at least 60% of Fate and 60% of Ecotox (acute) metrics; and Option 4 requires data for at least 60% of fate; 60% of Ecotox (acute) and 60% of Ecotox (chronic) metrics. In all instances, the values for percent of the total mass applied are very low. This is because the substances that are missing data (below the threshold) typically have low usage.

Table 2.4: Example threshold options and number of substances excluded

Exclusion thresholds	Active substances excluded		
	Number	Mass (tonnes)	Mass (% of total)
1. All (60%)	62	202	0.25
2. Fate (60%) and Ecotox (60%)	71	308	0.37
3. Fate (60%) and Ecotox (acute) (60%) (Ecotox (chronic): 0%)	61	345	0.42
4. Fate (60%), Ecotox (acute) (60%) and Ecotox (chronic) (60%)	88	921	1.11

The Fate criterion seems to have a minimal influence on the results of the exclusion process (probably due to the smaller number of load metrics) with the Ecotox (acute) criterion being the primary influence. Thus Figure 2.4 can be simplified to 3 Ecotox criteria of increasing stringency (Ecotox, Ecotox acute and Ecotox acute and chronic) as shown in Figure 2.5.

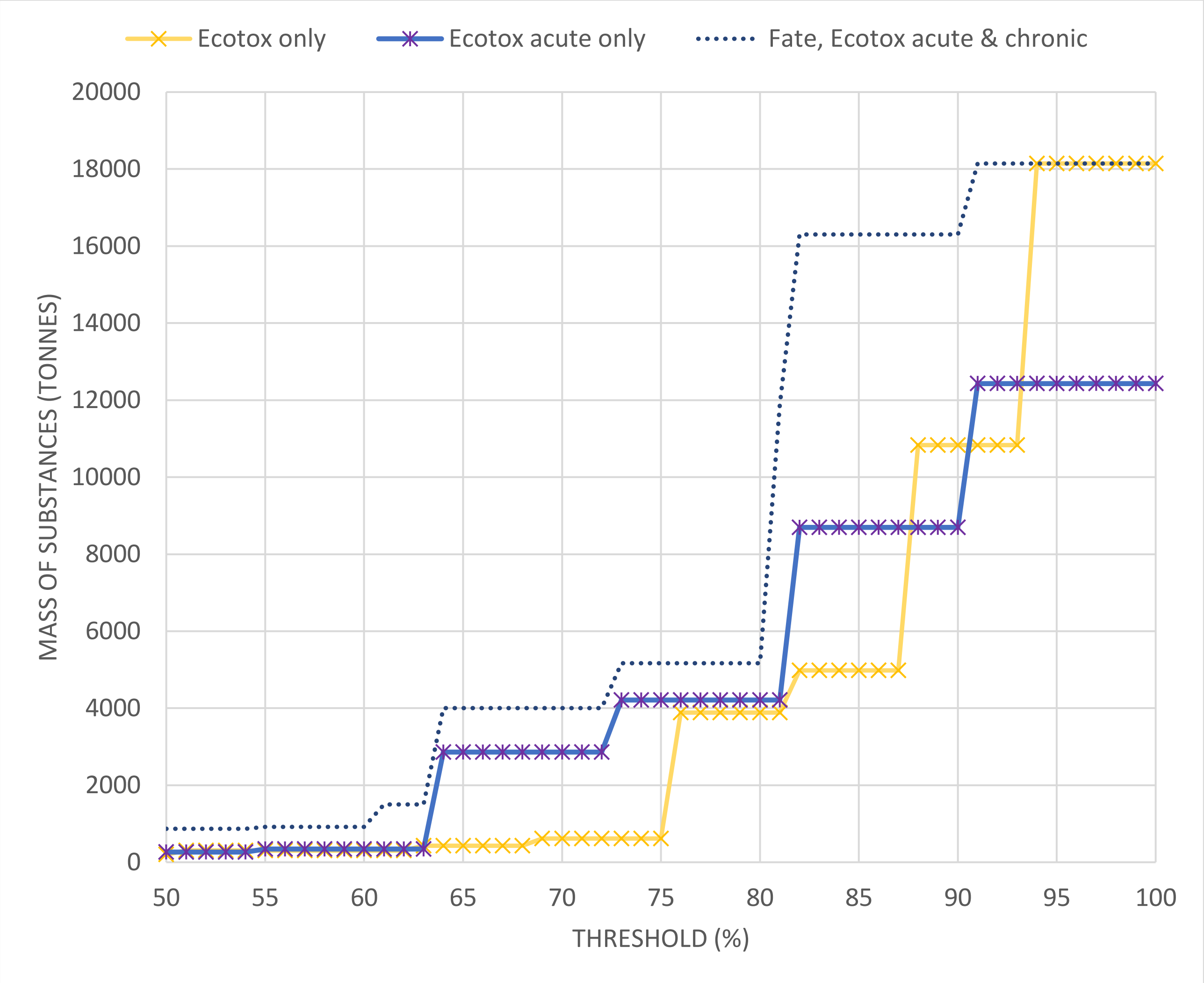

Figure 2.5: Quantity of active substances excluded at different Ecotox data coverage thresholds

Across these criteria, the first significant step of exclusion occurs at 64% where 2500 tonnes of Mancozeb is excluded due to a lack of Ecotox (acute) data (this same step occurs at 76% when using the 'Ecotox only' data coverage threshold). This potentially provides an argument to set an exclusion threshold of 60% data coverage for Ecotox (acute) load metrics. This would result in 60 substances being excluded (345 tonnes) which make up 0.42% of the total mass of all substances applied (across all surveys and years) (see Annex 2 for full list of excluded substances). However, this does not show what impact that would have on the load scores, i.e., are any substances with high load scores being excluded?

To explore the potential impact on the load values of excluded substances, the value for each load metric was calculated (across all surveys and years) and expressed as a percentage of the total load. This showed that most of the excluded substances contribute <0.1% of the total load across all the metrics. The exceptions are:

- Abamectin: 0.14% and 0.31% for Honey bees (contact) and Honey bees (oral)
- Kresoxim-methyl: 0.14 and 0.11% for SW Mobility and GW Mobility
- Potassium phosphonate (phosphite): 0.16% and 0.11% for Persistence and GW Mobility
- Urea: 0.15% for SW Mobility

(Note: for reference, the substances that contribute most to the total load for these metrics are Persistence: Pendimethalin 19%; SW Mobility: Chlormequat 21%; GW Mobility: Chlormequat 23%; Honey bees (contact): Zeta-cypermethrin 30%; and Honey bees (oral): Chlorpyrifos 19%)

These are all relatively low values, thus, the exclusion of these substances (because they have <60% of the data for ecotoxicity acute metrics) should have a minimal impact on the load values for these metrics. However, the exclusion of Abamectin and its metabolite 8a-hydroxyavermectin B1a needs more discussion, as these are reference substances[6] (Abamectin for Honeybees [contact] and 8a-hydroxyavermectin B1a for Honey bees [oral] and Mammals [acute]). Table 2.5 shows the alternative substances that would be used as the reference substance if Abamectin and its metabolites were excluded.

Table 2.5: Alternative reference substances for excluded substances

Metric	Existing ref substance	Value	New ref substance	Value
Mammals (acute) LD_{50} mg kg^{-1} BW	8a-hydroxyavermectin B1a	1.5	Oxamyl	2.5
Honey bees (contact) LD_{50} µg bee^{-1}	Abamectin and its metabolites	0.001	Deltamethrin	0.0015
Honey bees (oral) LD_{50} µg bee^{-1}	8a-hydroxyavermectin B1a	0.001	Imidacloprid	0.0037

[6] Reference substances are used in the Standardisation step of the PLI. Their data is used to define minimum or maximum values to convert the data for each substance for each load metric onto a common 0 to 1 scale. See Section 2.4 for further details.

The biggest impact would be the change of the reference substance for Mammals (acute) (the changes to the Honeybee metrics make a negligible or no difference). The change in the Mammals (acute) reference value from 1.5 to 2.5 mg kg^{-1} bw has the effect of increasing the load value for each substance (as each substance is closer to the reference substance value). Calculations indicate that this would have the effect of increasing load values by 67%.

Where a reference substance is excluded due to a lack of data, it has been decided to create an exception to the data threshold exclusion rule and include the substance within the PLI. This has been done for several reasons including:

1. There is a potential communication issue for the PLI if the reference (worst performing) substance used to calculate any of the load metrics is not included within the load score.
2. Including the worst performing substance from the outset of the PLI is likely to result in a consistent reference substance throughout the lifetime of the PLI (i.e., on the assumption that it is unlikely that new active substances approved in the future are unlikely to be reference (worst performing) substances).
3. As with any other substance lacking data, these data may become available in the future resulting in the substance passing the data coverage threshold and it would then become the reference substance (thus, in alignment with point 2, this would help maintain a consistent reference substance).

This forms one of the final steps in the protocol (see Figure 2.32) where if the substance is a reference substance for one or more of the load metrics, then it is included in the PLI. Thus, in this instance Abamectin and its metabolites would be included within the PLI.

2.3.3. Conclusion

The protocol provides a systematic and transparent mechanism by which to determine whether a substance should be excluded from the PLI. A data coverage value of 60% of the data needed for the Ecotox (acute) load metrics provides an acceptable threshold for excluding substances. Extending the data coverage threshold to include 60% of the Fate metrics or 60% of all Ecotox metrics makes little difference; and extending it to cover 60% of the Ecotox (chronic) metrics would be slightly more stringent.

Those substances excluded (under the 60% of Ecotox [acute] threshold) make up less than 0.5% of the total mass of substances and generally less than 0.1% of any load metric. Abamectin (and its metabolites), Kresoxim-methyl, Potassium phosphonate (phosphite), and Urea are the only exceptions, but still only contribute a maximum 0.1 to 0.3% of any load metric. However, the exclusion of Abamectin and its metabolites needs further consideration, as these are reference substances for 3 load metrics (Mammals [acute], Honeybees [contact], and Honeybees [oral]). The exclusion of Abamectin (and its metabolites) only impacts the results of the Mammals (acute) load calculation, by raising the load values for all substances due to a less toxic reference substance. It has been decided that if a substance is a reference substance for one or more load metrics, then it should be included (overriding the data coverage threshold rule) to help ensure consistency in the reference substances used in the PLI.

All data included in the PLI as of 2023 will be consistent with this protocol. Substances that are excluded from the scope of the PLI at the time of writing (March 2023) are listed in Annex 2 of this document.

2.4. Standardisation

A range of different units of measurement exist across the load metrics. To determine the relative load of each pesticide for each metric, the raw values for each load metric (from the PPDB) need to be standardised to a common scale. This is essentially a normalisation process where the raw value for a pesticide is expressed relative to a reference substance (those pesticides with the highest and lowest values for each load metric amongst those assessed for the PLI) and, for fate metrics only, regulatory interpretation thresholds. The reasons for standardisation are partially a reflection of the previous aggregations applied to combine load metrics (Section 2.6) but also to provide a mechanism that explicitly defines the underlying mathematical relationships that determine the relative 'values' assigned to different active substances. The goal of standardisation is to place all active substances for a given metric on a 0 to 1 scale both as a communication tool for the purposes of visualisation of relative change, but also to reflect the intrinsic differences between how different metrics might be interpreted (which has led to a distinction between how fate metrics and ecotoxicity metrics are treated for the purposes of calculation and limits the misleading tendency to directly compare metrics with different intrinsic scales). An impact of standardisation is to render the metrics in the PLI unitless (because they are expressed relative to the value of the 'worst case' reference substance for each metric). For this reason, it is generally inadvisable to focus communications around the 'absolute' load value, but rather to focus on trends in relative change from one year to another, as the latter are generally more informative regarding the impact of policy and more amenable to targets that incorporate a clearly defined baseline (see Section 2.6, and Section 3).

The load scores for fate measures are derived from a combination of regulatory interpretation thresholds and reference substance values (see Table 2.8 in Rainford *et al*., 2022b[7]) to generate standardisation curves shown in Figures 2.6 to 2.9. These standardisation curves are used to convert the raw value (x axis) from the PPDB into a load score (y axis) for each pesticide for each of the fate load metrics.

This approach (using regulatory threshold values to introduce reference points) helps reduce the skewing effect that can result from extreme reference substance values. For example, for the surface water mobility metric, the lowest value (best performer) is deltamethrin with a K_{foc} of 10,240,000 L kg^{-1}, which is so high that it cannot be shown in Figure 2.9 (the chart is truncated to 10,000 L kg^{-1} be able to see the other reference points on the curve). If a straight line were drawn between 0 and 10,240,000 L kg^{-1}, most pesticides would get a high load score, but the introduction of the

[7] To summarise, values for reference points were selected from regulatory interpretation thresholds known to the PPDB (for example representing the transition between where an active substance would be considered Low, Intermediate, or High risk respectively). The position of the reference values within the 0 to 1 standardised scale was intended to assure coverage of these risk categories across this scale, e.g., in the previous example any substances which exceed for being classed as 'High' risk for a given metric are plotted in the upper third of the standardised distribution of that metric, see Table 2.8 in Rainford *et al*., 2022b for discussion.

reference points reduces this skewing effect, with only those with a K_{oc} (or K_{foc}) >4,000 L kg^{-1} (<0.167 load score) being affected by the skewing.

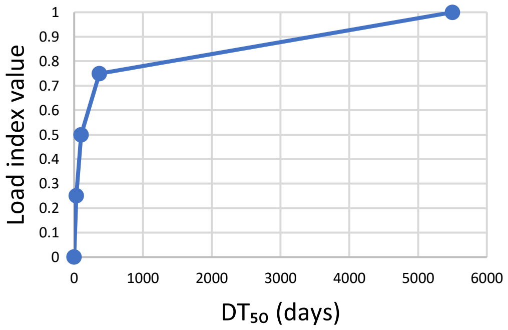

Figure 2.6: Standardisation curve: Persistence

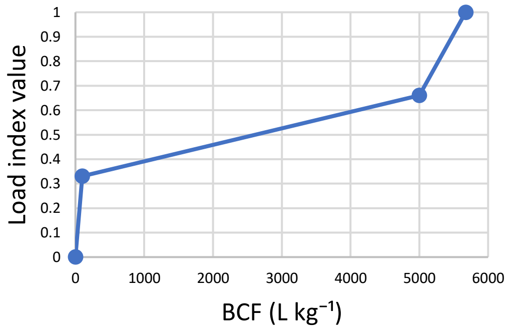

Figure 2.7: Standardisation curve: Bio-concentration factor

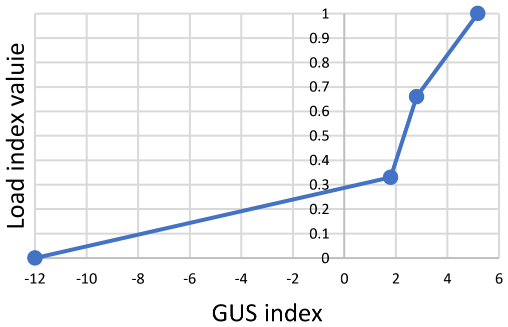

Figure 2.8: Standardisation curve: Groundwater mobility

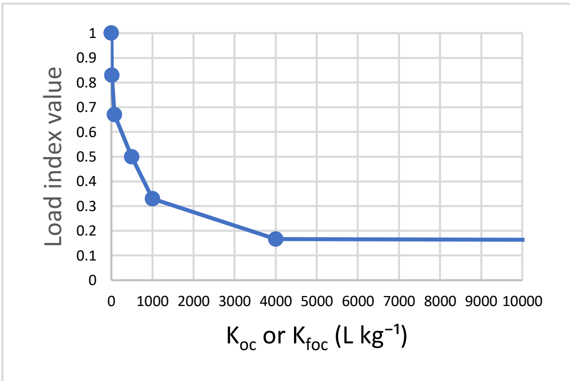

Figure 2.9: Standardisation curve: Surface water mobility

Plots show the relationship between the the raw metric value (x axis) and the standardised value (y axis) used in calculating the PLI. Points represent the regulatory threshold values and the lines between each point cover the values of each active substances (including the reference substance for each load metric as the highest value, assigned a standardised value of 1). The chart for surface water mobility is truncated at 10,000 L kg^{-1} to ensure all reference points are visible.

For ecotoxicity load metrics, the pesticide property value (see Table 2.2) of a given active substance is expressed relative to the reference substance (most toxic pesticide amongst those assessed for the PLI) using Equation 2.1.

$$\text{Load metric (lm) index}_{as} = \frac{1}{\text{Active substance value}_{lm}/\text{Reference substance value}_{lm}} \quad \text{(Equation 2.1)}$$

Figure 2.10 shows an example of the standardisation curve that emerges from Equation 2.1 for the Mammal ecotoxicity (acute) load metric.

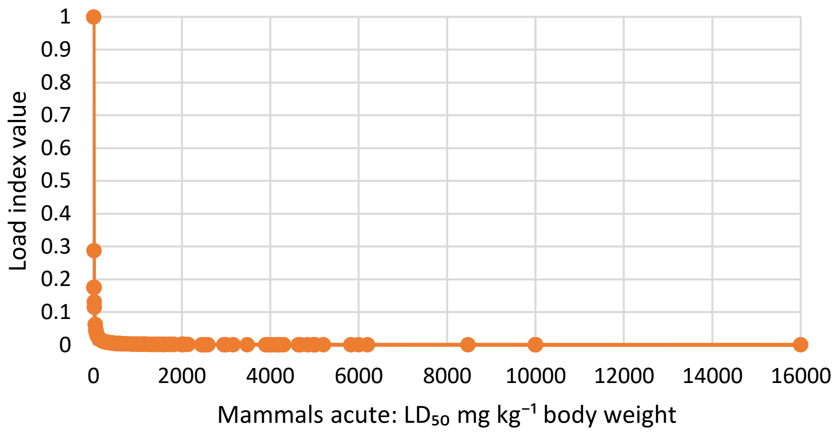

Figure 2.10: Example Standardisation curve: Acute toxicity to mammals (LD_{50})

The use of regulatory thresholds for standardising ecotoxicity metrics was also explored (in Phases 2 & 3). However, this resulted in a loss of natural scaling with respect to relative toxicity. For example, a substance with an LD_{50} of 8.7 mg kg^{-1} body weight compared to one with an LD_{50} of 2000 mg kg^{-1} body weight would be considered as having ~200 times greater relative toxicity per kg of active substance applied. Without regulatory thresholds, they would have a load index value of 0.28 and 0.001 respectively (maintaining the ~200 multiple), whereas with regulatory thresholds they would have index values of 0.97 and 0.33 respectively, which is a multiple of ~3. Following discussions during Phase 3 it was determined that the need to maintain natural scaling for ecotoxicity metrics was a greater concern than the potential skewing associated with extreme reference substance values; thus, regulatory thresholds have not been implemented for ecotoxicity load metrics.

2.5. Multiplication

The final step in calculating the PLI is to multiply the load scores with rates of usage for individual active substances, using data derived from the PUS (Garthwaite et al. 2019) and crop areas derived from the June surveys (Defra 2022). An estimate of usage (here defined as the mass in kilograms of active substances applied) is calculated by extrapolating data from records provided by growers for a sample of surveyed holdings[8]. The rates of application for each pesticide were estimated using a combination of the grower supplied rates and the LIASON pesticide registration database[9] (Fera, 2023). Pesticides applied as seed treatments are treated separately for the purposes of calculation, and any subsidiary co-formulants (e.g., surfactants) or

[8] Data on pesticide applications on holdings in the UK is legally required to be maintained for a period of up to 3 years. However, other than for those holdings voluntarily in the PUS these data are not currently nationally collected or aggregated.
[9] The LIAISON database (https://liaison.fera.co.uk/) is a commercial product used by agronomists and research organisations. It contains all products registered with CRD and uses the product label rate to show minimum and maximum rates for all crops mentioned on the product label. In the PUS and PLI, it is used to fill in rates of application where these have not or cannot be provided by the grower e.g., in the case of seed treatments on purchased seed.

treatments which could not be assigned to a specific product/active substance (predominantly unknown seed treatments) are excluded. Likewise, applications made during or outside the drain flow period (Sep-Apr) were labelled as such and estimated separately (applications outside of the drain flow period are assigned a value of 0 on the surface water load metric). Note that with the exception of the split between treatments inside and outside the drain flow period, the PLI only considers the total mass of PPP applied throughout the growing season (i.e., regardless of how many individual treatments / spray passes this might represent in practice). This is one of several assumptions associated with the calculation of the PLI, which are outlined in detail Rainford et al (2022b) and briefly discussed in Section 5.2. Pesticides that (within a specific survey) are uniquely associated with specific crops are not estimated for regions where that crop is not grown (e.g., pesticides associated with sugar beet are not estimated for Scotland, where this crop is not grown). Only crops where suitable stratification is provided by the June survey(s) (primarily arable and some soft fruit crops) are associated with individual estimates of usage.

All estimates are based on summed annual usage over the period covered by each survey (typically applications over a 12-month growing season, estimated regionally or nationally). The variation in the sampled and extrapolated data are used via a conservative bootstrap procedure (Efron and Raoul 1992) to calculate the confidence intervals shown in the PLI outputs. Conceptually, the intervals for a single active substance are drawn by bootstrapped replication of the observed rates for a given active substance on a given crop among the holdings observed in a given region and size class of holdings, with this population being substituted for a more general one (e.g. the overall regional / national population) in those cases where the local data provides insufficient information to reliably approximate the unvisited holdings (e.g., in the case where no local holding was observed to apply a given active substance). Confidence intervals on the overall PLI metrics are calculated using the sum of these underlying intervals on the assumption that estimated rates are independent of one another. All confidence intervals presented in both the PLI tool (Section 2.7) and this report represent the 90% coverage envelope on the estimated values given the assumptions outlined in Section 5.2. The full details of the calculation and assumptions made during the process are fully documented in Section 2.6 of Rainford et al. (2022b). The current implementation of summaries for load metrics in the PLI is expressed in terms of relative change in the ‘absolute’ value of load (this is the load based on the estimated total mass of PPP applied). In other options within the visualisation tool, load may be expressed per hectare of cropping area or per tonne of production. In these cases, the data used are based on the national estimates / census provided by the June survey. The absolute value of load is preferred as the most appropriate proxy for potential impacts on the environment that includes the collective decisions made by growers related to what crops to grow and in what quantities (i.e., where there is a choice between intensively treated and less intensively treated break crops, see Section 3). However, for some of the more subtle and targeted questions around grower behaviour, accounting for changes in cropping area and or production can be more appropriate, as these influence the economic incentives around the use of PPP.

2.6. Removal of the aggregation step

At the outset of the development of the PLI, it was envisaged that it would result in a single combined environmental load metric (like the Danish PLI); thus, the aggregation step was retained and developed through Phases 2 and 3 (Rainford et al. 2022b). However, as it developed, the vision, role, and application of the UK PLI evolved and broadened compared to that used in Denmark, where it has a specific role in supporting their pesticides strategy and taxation system. Through Phases 2 and 3, it became apparent that the UK PLI needs to support a wider range of policy contexts and narratives. In addition, the exploration of different policy narratives (during workshops in Phase 3) highlighted that, in many instances, more interesting and valuable narratives could be derived from exploring the trends associated with individual load metrics (e.g. has the use of pesticides with high persistence decreased or increased, has the use of pesticides that are toxic to bees decreased or increased, etc.) rather than a combined metric (aggregated using aggregation constants that are inherently subjective and/or set for specific policy contexts). Consequently, the development of an aggregation technique became problematic, with a lack of consensus on an approach that would work for all contexts (see Annex 3 for a detailed description of the challenges associated with aggregation).

The lack of an acceptable aggregation methodology led to a decision in Phase 4 to discard the aggregation step and re-orientate the core visualisation of the indicator to targets for relative change rather than a headline value. One of the main strengths of the PLI framework is its capacity to express change through time in a consistent way which is largely independent of the issues created by aggregation. Viewed in isolation, relative change in each metric (particularly through time) can be intuitively interpreted according to the expectations of stakeholders. In place of a single headline aggregated value, the decision was, therefore, made to stress relative change in the visualisations. For example, in place of the previous aggregated headline values, which expressed the summed change across different metrics, the revised approach asks whether, when considered in isolation, each metric has undergone a sufficiently large relative change relative to a user defined threshold.

This change has important implications for how to approach measurement of progress in the context of the PLI in that it replaces the increasingly problematic change in the headline value, with a series of compliance statements which express that a given metric has (based on a best estimate of uncertainty) undergone a reduction of at least 'X' magnitude relative to its value in a given reference year. The advantage of this fully relative approach is that it can be extended naturally to give an overall criterion for 'success' (for example that ALL metrics should have shown a reduction in load of at least 'X' magnitude relative to a reference year). It also avoids the problem that the indicator is unitless (see Section 1) by emphasising that interpretation should always be based on patterns of relative change rather than absolute values (see Section 3). Hence, while not representing a major innovation in terms of the technical underpinnings of the calculation for individual metrics, this is a substantial shift in the way the results are interpreted and presented, which sidesteps the issues outlined above. Section 3 re-examines some of the key findings from the PLI considering this revised focus. It is hoped that by doing so, the strengths of the chosen approach can be emphasised in terms of providing an in-depth tool for policy, while also showing that a threshold-driven approach still provides a convenient and simple tool for describing overall performance going forwards.

2.7. Outputs and visualisation

A key requirement is the need for accessible visualisations to denote trends in the indicator from a range of different perspectives. The PLI was developed with several different policy audiences in mind, including users with differing levels of experience and information needs, particularly the need to often link change to the use of specific active substances. To address these requirements, an R shiny application (RStudio, 2022[10]) has been developed for Defra users to create standard visualizations of trends in individual metrics, as well as the contribution of individual pesticides, and appropriate estimates of uncertainty. At present the application covers (biennial) arable surveys between 2010 and 2020, as well as the orchard and soft fruit surveys[11] of 2016, 2018 and 2020, the outdoor vegetable survey for 2015, 2017 and 2019 and the grassland and fodder survey for 2013 and 2017.

During Phase 4, several refinements and revisions were made to the developed visualisation tool to attempt to better align to the policy needs of Defra. The most consequential of these changes is the removal of the aggregation step and consequent reorganisation of the visualisations. To reflect the new focus on progress measurement, the visualisations on the tab **Summary of Included metrics** now explicitly include a plot which categorises percentage change in each metric between any given pair of years based on user defined threshold (for example, identifying all metrics which show at least a 10% reduction between a specific pair of years; see example shown in Figure 3.5). The aim here was to provide a rapid visual overview of the direction of travel for different metrics, which would be useful in directing the user to cases requiring further investigation. Similarly, visualisations that were related to the aggregation step, e.g., those showing the contribution of metrics to the overall headline score, have been removed, as have all references to the former aggregated statistics (including the previous option to select alternative aggregation approaches).

The visuals under the panel **'View a metric in detail'** have also been revised to include both total load and load per hectare of cropping area and their respective percentage change relative to the earliest year in the series. An optional download feature has been added, which allows users with an appropriate level of access to download data directly from the application. Again, reflecting the more threshold-driven view a new panel '**View importance of actives'** has been added which ranks active substances in terms of their average percentage contribution to the overall value across each of the 20 PLI metrics in a given year location and crop group. This should make it easier for users to focus in on the key active substances that are driving load across many metrics and so identify potential areas for policy intervention. There have also been various small-scale graphical improvements made to various figures and headings which will hopefully improve the user experience and improve the transparency of findings. A revised User Guide for internal Defra users has been

[10] https://shiny.rstudio.com/
[11] Including crops grown under permanent protection in Scotland and Northern Ireland.

developed, which includes updated instructions for how the tool is used as well as screen shots and technical details for how the app will be maintained going forward (see Section 2.8).

At the backend, the visualisation tool has been modified to work with the revised standards to substance inclusion which are now maintained as part of the associated technical infrastructure. This includes maintaining a list of micro-organism-based products which can now be viewed under the **‘Select Chemical Group’** options throughout the tool. As noted above, background data filters are now maintained to exclude substances which fail to meet the data coverage criteria outlined in Section 2.3 and to remove inorganic substances from consideration in the calculation of fate metrics.

2.8. Ongoing delivery

As part of Phase 4, the PLI delivery team have drafted an explicit protocol and contracting structure to facilitate the ongoing delivery of the PLI as a national tool for understanding change in the composition of PPPs applied in the UK. As of 2023, the protocol for calculating the PLI is now fully documented and available to the delivery team and Defra including description of all steps and processes involved. This is intended to provide a robust pipeline for calculating the indicator and providing the supporting datasets required by the visualisation tool. The key constraint identified in ongoing delivery was the frequency and maintenance of the PUS which represents the key data source for usage estimates (under the current model, most surveys used in the PLI are conducted on a biennial basis, with the exception of the grassland and fodder survey, which are conducted only once every four years). Defra and the delivery team have agreed a process of annual review to ensure that appropriate surveys within the PUS for use within the PLI are identified and processed promptly after publication. At present, the PLI infrastructure will remain separate from the publicly accessible infrastructure around reporting of usage (Fera 2021; PUSSTATS[12]) reflecting its focus as an internally facing Defra tool. This may, however, be subject to review in future. As currently scoped, the PLI and visualisation tool are intended to be updated on a rolling basis for each of the PUS surveys currently considered within scope (arable, orchards, soft fruit, outdoor vegetables, and grassland and fodder) although this may be subject to revisions.

[12] https://pusstats.fera.co.uk/home

3.0. Results and discussion

3.1. Introduction

One of the key goals for the PLI was to make the indicator accessible to a wide audience and to develop a tool that allows users to explore policy relevant case studies in a way that is informative and consistently structured. This section provides a brief and illustrative summary of some of the key features developed for the application focused on understanding high level trends in the sum of applications made in the arable setting, as well as a more focused case study on one of the key policy interventions of the last decade, the withdrawal of neonicotinoid seed treatments, specifically focusing on winter-sown oilseed rape (OSR). The intention is to provide a guide for the types of analyses, visualisations, and topics that users may wish to explore with the PLI, as opposed to an exhaustive examination of trends in environmental load within the UK landscape.

3.2. Trends in metrics of the PLI (summed arable crops 2010-2020)

In Section 3 of the Phase 3 report (Rainford et al. 2022a) the trends in the then defined aggregated headline value of the PLI are discussed. Given the recent revisions to the indicator structure and the increasing focus on relative change in individual metrics (Section 2.6), a re-examination of these findings is provided here, with a focus on how the tool might be used in practice to identify key concerns. All visualizations presented are taken from the developed tool accompanying the PLI (see Section 2.7) and reflect the range of potential options available to users for exploring changes in pesticide policy and application behaviour on an ongoing basis (see Section 2.8).

Arable crops (includes potatoes and sugar beet, alongside cereals, oilseed and pulse break crops) is by far the largest crop group in terms of pesticide usage, accounting for approximately 90% of the area treated and weight applied of UK agricultural and horticultural PPP applications. One of the weaknesses that the PLI inherits from the PUS is that because of the biennial nature of surveys it is inadvisable to try to combine data from different crop groups, which have been assessed in different calendar years, and thus may have been subject to different regulations relating to which products and active substances were authorised for use. For this reason, it is suggested that crop groups be examined individually, with arable applications serving as an effective proxy for national PPP applications, although other crops groups can also represent important contributions particularly at a regional scale.

Currently within the PLI, arable data are available for surveys conducted every two years between and including 2010 and 2020. Between these years, the overall trend in mass of PPP applied (Figure 3.1) has been one of a general increase, peaking at 14,902 tonnes in 2018 (or between 14.7k and 15.1k tonnes; based on the 90% confidence intervals estimated on the mass of application) or 3.57 kg of pesticide applied per ha grown (or between 3.53 and 3.61 kg ha^{-1}). This was followed by a substantial (25%) decline in mass applied in 2020 relative to 2018, with 11,208 tonnes (between 11-11.3k) or 2.67 kg ha^{-1} (between 2.63 and 2.71 kg ha^{-1}) being applied in 2020.

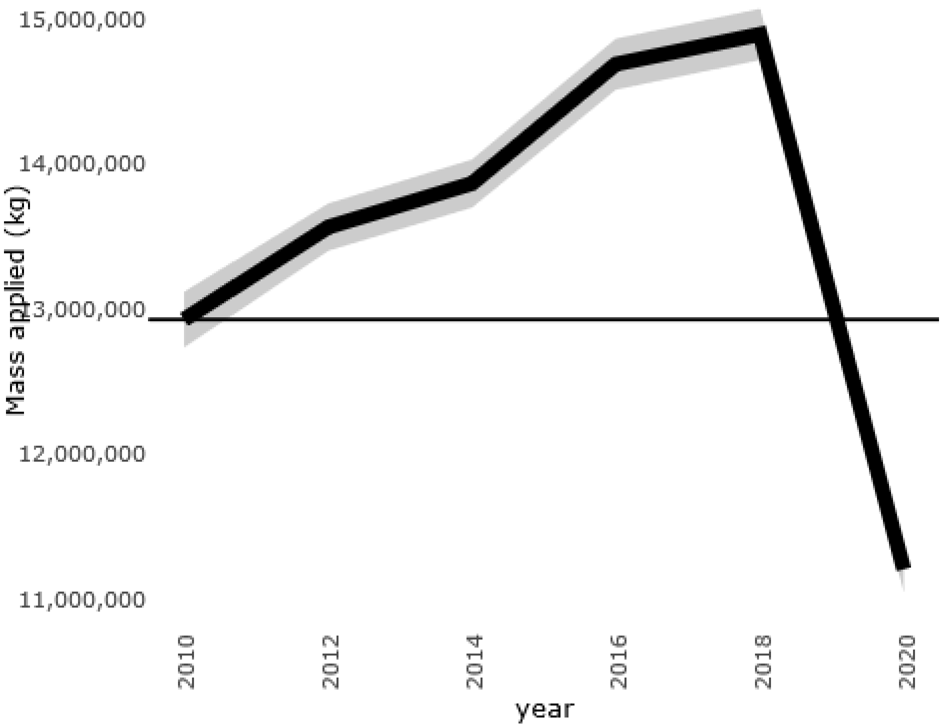

Figure 3.1: Trend in total mass of active substance applied to UK arable crops (2010-20)

As discussed in more detail in Rainford et al. (2022a), the decline in 2020 is not meaningfully associated with declines in overall cropping area (which only reduced by 0.5% between 2018 and 2020), but rather reflects a shift in choice of crop and associated application patterns in use of PPP. The national cropping area(s) of wheat (mainly winter sown and the principal arable crop grown in the UK) declined by 23.6%, as did winter barley by 31%, between 2018 and 2020[13]. In contrast, cropping areas of spring barley increased by 52%, and peas and beans (a combination of spring and winter sown crops) increased by 26% and 32% respectively. OSR showed a 32% decline in the winter sown area and 22% increase in the spring sown area respectively.

All these changes are suspected to be linked to extreme weather conditions and flooding in the autumn of 2019, which may have substantially impacted winter sown crops and reduced the incentives and ability of growers to carry out spray applications. Associated crop losses are also suspected to be associated with greater uptake of spring sown crops and varieties. Unfortunately, this period also corresponds with the height of the COVID 19 pandemic in the UK which both disrupted agricultural practice, as well as the data collection on which national estimates of cropping area are based (e.g., regional and size group level estimates of the areas for specific crops[14]). This

[13] Values taken from the 'June' agricultural surveys conducted by Defra and the various devolved authorities.

[14] For a number of specific crops, regional level data was not complied in 2020 due to pressures caused by the pandemic. As a result, many values used in both the PUS and PLI for cropping areas in 2020 are based on regional cropping data from the previous year (2019) (see Section 2.6.2 in Rainford *et al.*, 2022b). UK national estimates of usage in the PLI are derived from the sum of regional estimates.

had the net effect of increasing uncertainty in the drivers of trends as it was much more difficult to reliably assess the economic conditions for UK growers. At present, 2020 is considered an atypical year in terms of pesticide usage statistics and data for the 2022-2023 growing season (ongoing at the time of writing) will be used to assess these changes in the context of longer-term trends in usage.

Despite the uncertainty associated with the 2020 usage data it is, nevertheless, possible to use this trend as an illustration of how the metrics contributing to the PLI can be visualised and how the different visualisations available within the tool inform one another to help users with key decisions. Figure 3.2 shows the trends in the calculated metrics between 2010 and 2020, expressed as a percentage of the 2010 value. Different metrics show very different trends, some of which diverge substantially from the trend in overall mass applied. In particular, while the Bioconcentration factor, Soil DT_{50}, and Mobility metrics show (to varying degrees) the same broad trends as the overall mass applied (gradual increase over the period 2010 to 2018, followed by rapid decline in 2020), the Drain flow metric (which is heavily influenced by the timings of respective applications), peaks much earlier (around 2014) and has been undergoing a longer term decline in recent years that is much more substantial in percentage terms than that observed in mass applied. This is likely as a result of changes in the timing of key applications, particularly in the shift towards increased areas of spring sown crops, which are often most intensively treated after the end of the winter drain flow period.

The trends in ecotoxicity metrics are even more complex and varied, with some, such as Bird LD_{50} (acute/short-term) and Mammal LD_{50} (acute/short-term), plateauing in recent years, while others such as Bees contact LD_{50}, Parasitic wasps, Daphnia EC_{50} (acute/short-term), Daphnia NOEC (chronic/long-term) and Fish NOEC (chronic/long-term) show pronounced declines over the period when the total mass of pesticides applied was increasing. Other metrics, notably Bees oral LD_{50} and Algae EC_{50}, trend upwards prior to 2018 before declining in 2020. The individual trends in specific metrics are examined in more detail below. However, this diversity of trends between metrics, when expressed in their own scale, serves to illustrate the limitations of the aggregated indicators that were previously defined under the PLI (Rainford et al. 2022a) and highlights the strength of the revised single metric focus approach adopted in Phase 4 (see Section 2.6).

Given that the key function of the PLI is to provide information and 'colour' to the trends arising from the changing mix of active substances applied (Section 1), it makes sense to structure the calculation and presentation to provide the widest possible range of information to decision makers seeking to prioritise alternative interventions in PPP. The following section examines how the visualisation tool handles this complexity by providing targeted visualisation for a given policy need and how the various approaches to visualisation can be used together to provide a joined up and highly resolved image of the drivers of change in UK PPP applications and the potential priorities for further monitoring (see Section 3.3).

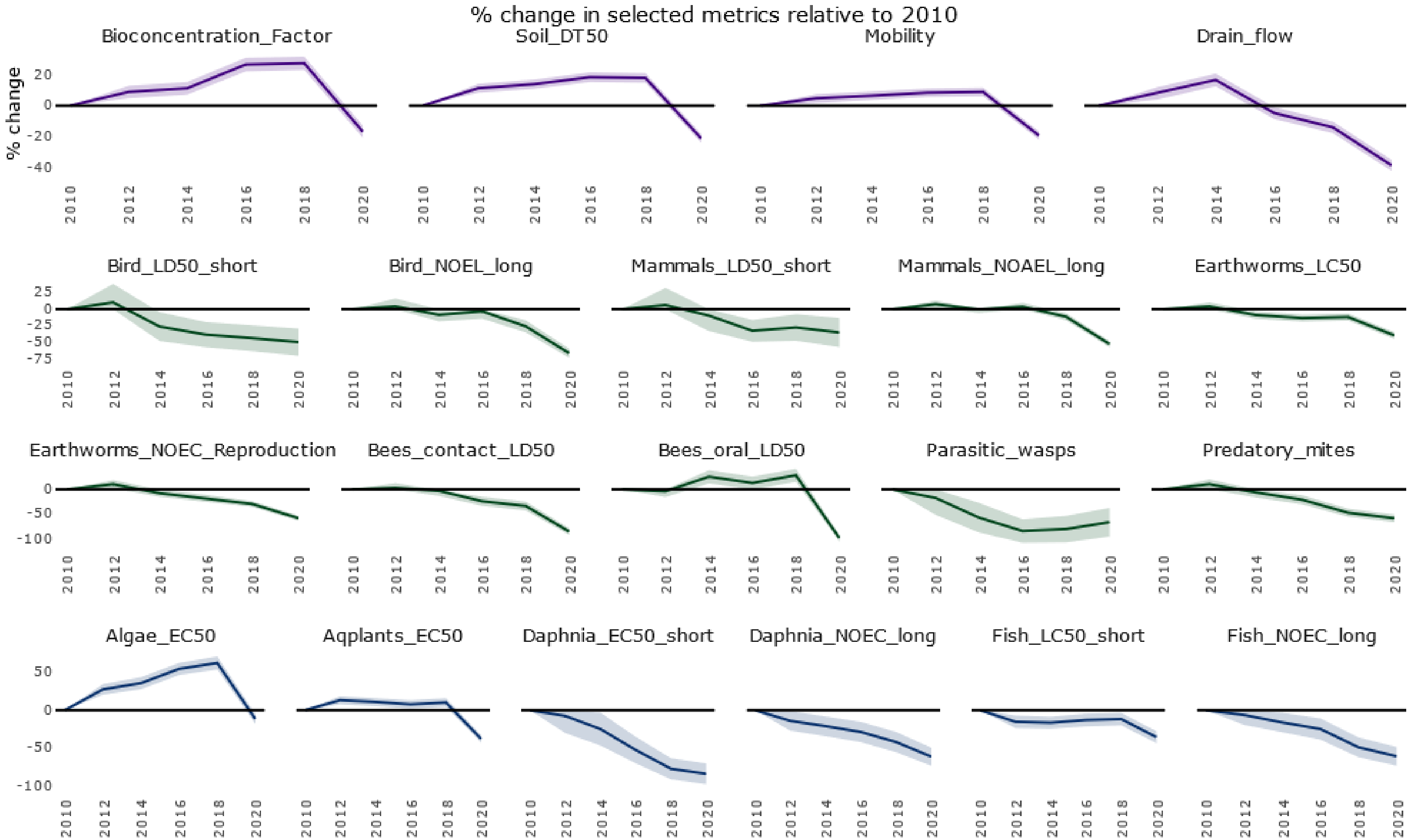

Figure 3.2: Trends in metrics comprising the PLI for the sum of all arable cropping 2010 to 2020

Values are expressed as % change relative to 2010. Shading around the trend lines reflect the 90% confidence interval. Solid horizontal reference line denotes no change relative to 2010.

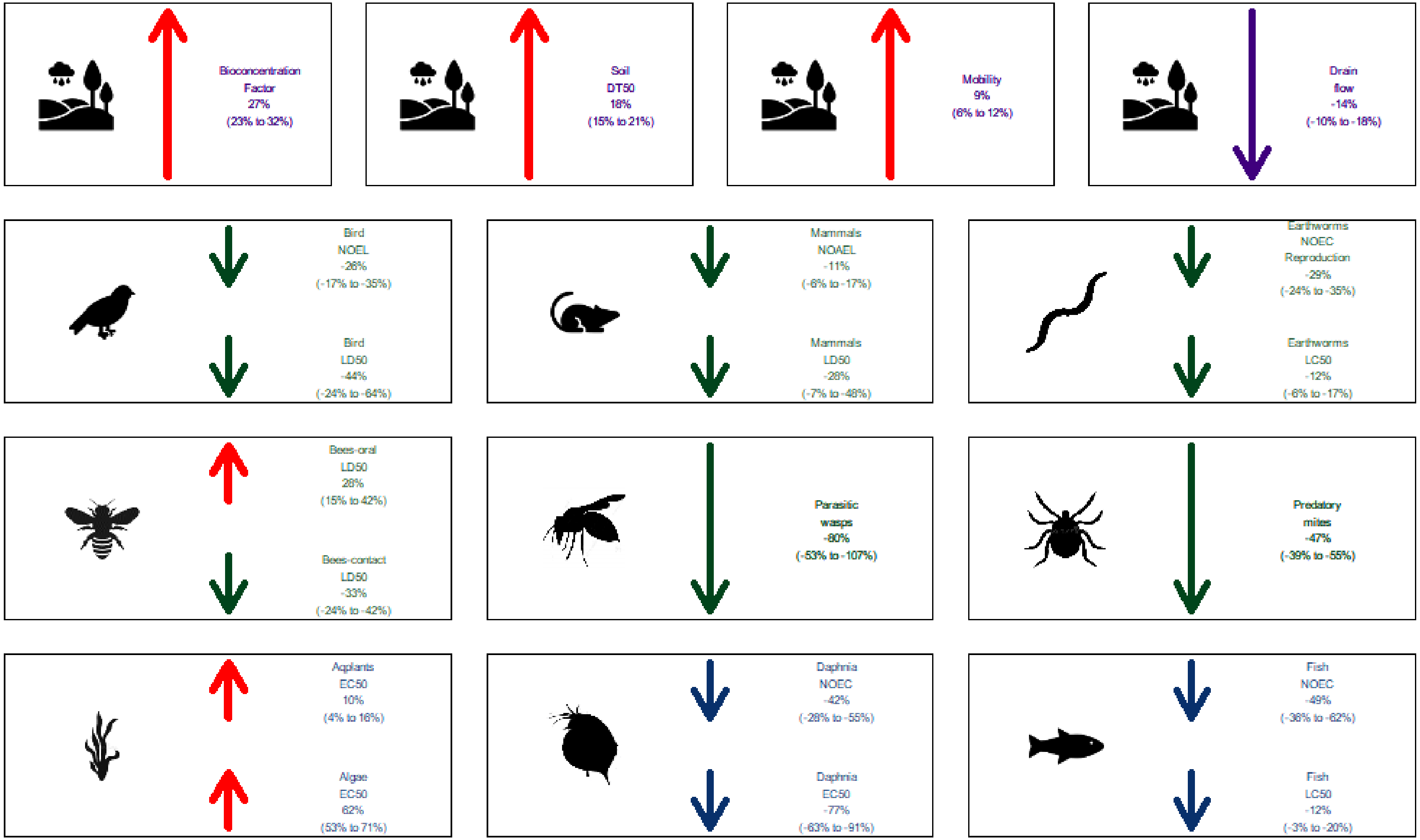

Figure 3.3: Change (%) in load for selected metrics for summed arable cropping between 2010 and 2018

Values are expressed as percentage change in the total value of the indicator relative to the value in 2010. Confidence estimates are based on 90% intervals around the mean. Arrows are used to denote direction of net change for the named metric. No net change considering the range of uncertainty around the estimate is denoted by circles. Metrics related to the same taxnomic group are shown together.

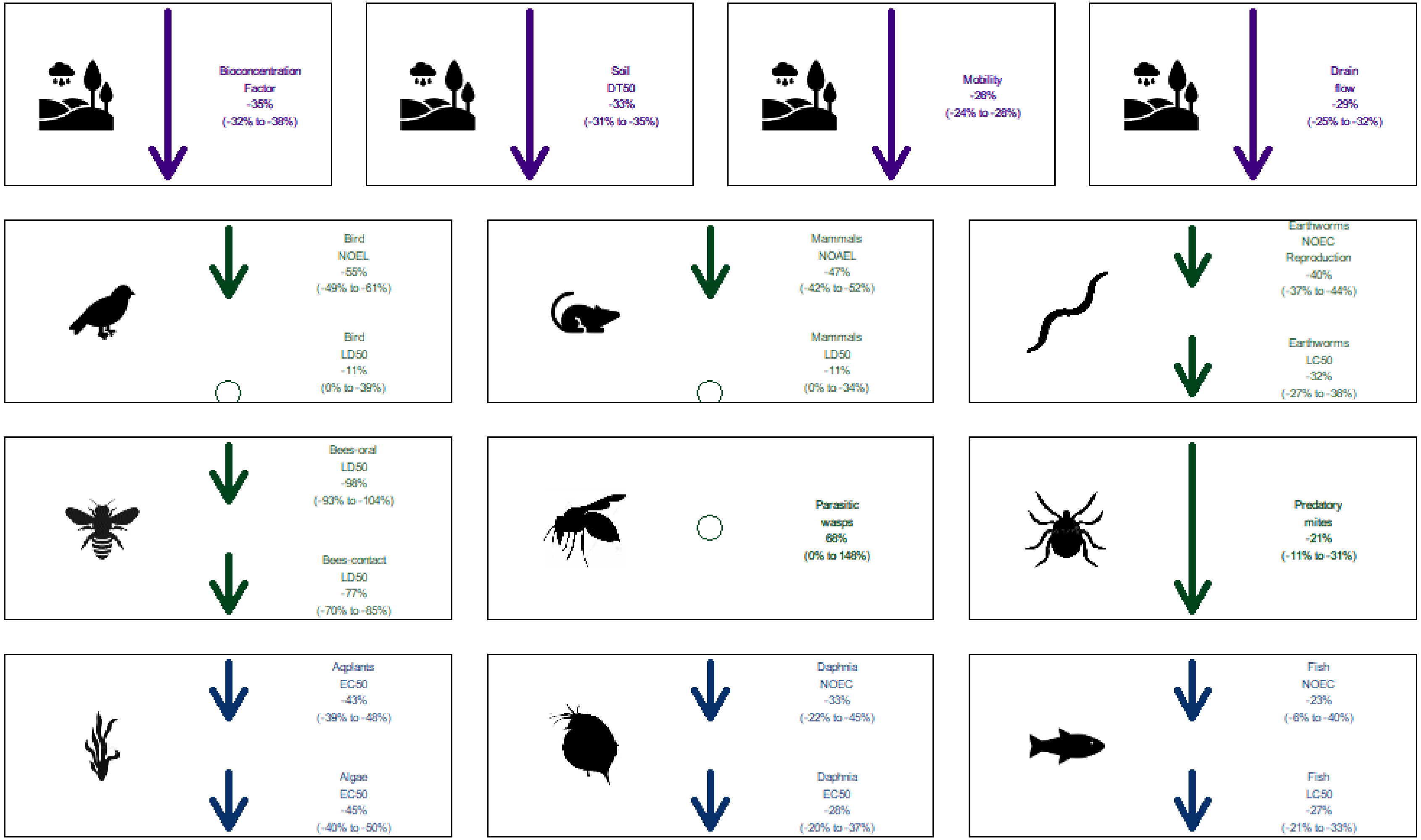

Figure 3.4: Change (%) in load for selected metrics for summed arable cropping between 2018 and 2020

Values are expressed as percentage change in the total value of the indicator relative to the value in 2018. Confidence estimates are based on 90% intervals around the mean. Arrows are used to denote direction of net change for the named metric. No net change considering the range of uncertainty around the estimate is denoted by circles. Metrics related to the same taxonomic group are shown together.

Figures 3.3 and 3.4 show the extent and direction of change of individual metrics as an infographic developed for the visualisation tool over the period 2010 to 2018 and 2018 to 2020 respectively Arrows show the direction of inferred change while the values list the extent of percentage change (relative to the earlier year) with the associated confidence interval. Arrows associated with a net increase in load are denoted in red, while a circle is used to denote where there has been no net statistical change in a given indicator (or that the minimum estimate of any observed change between the two years is within the uncertainty estimated for the earlier year). Viewed in this way allows the user to highlight the general trends up to 2018 (when the overall mass of active substance applied was increasing) and contrast this with the recent changes between 2018 and 2020.

As a benchmark for comparisons between 2010 and 2018, the overall mass of active substances applied to all arable crops in the UK increased by approximately 16% (±3%). Looking at Figure 3.3 it is immediately apparent that this increase is unevenly distributed across different metrics, meaning that there are different components of load that are responding differently to the changes in the mix of PPP actives applied. This is unsurprising, as it is change in the *composition* of active substances that is the major source of variation over this period, and this is what indicators like the PLI seek to understand. Focusing initially on those metrics which have shown increases in net load over this period, it can be seen that none of them align precisely with the trend in mass applied, with the closest being the Soil DT_{50} (increases by 18% ±3%), Mobility (increased by 9% ±3%), and Aquatic plants EC_{50} (increased by 10% ±6%). There are also metrics where the proportional increase is larger than that of mass applied, notably Bees oral LD_{50} (increases by 28% ±13%) and Algae EC_{50} (62% ±10%), reflecting the fact that some of the substances applied are more potent in terms of their toxic effect for the same mass applied with reference to these specific taxonomic groups. However, for the majority of metrics considered by the PLI, the trend over the period 2010 to 2018 is towards decreasing load, i.e., for the vast majority of metrics the average kilogram of active substance applied is associated with a lower toxic effect (although this statement must be caveated with various observations that are outside of the scope of the PLI such as the potential for synergistic effects arising from the mixtures of substances applied; see Section 5.2). This may reflect the results of successful policy intervention and the ongoing withdrawal of some of the most harmful active substances (most notably neonicotinoid insecticide seed treatments; see Section 3.3).

The changes from 2018 to 2020 can be contrasted with this ongoing trend. The total mass of active substance applied to all arable crops declined by 25% between 2018 to 2020. This decline was matched by the majority of the metrics (Figure 3.4) except for Bird LD_{50} (acute/short-term) and Mammals LD_{50} (acute/short-term), which show only very slight declines of approximately 11% on average, while the mean estimate for load on parasitic wasps shows an overall increase. Notably, however, in these three cases, when appropriate uncertainties on the estimated values are taken into account it is not possible to discount the possibility of no net change between the two years (see confidence envelopes on Figure 3.2 and the values in brackets in Figure 3.3 & 3.4). The largest declines from 2018 to 2020 are associated with load on bees, with the two metrics calculated on oral and contact toxicity declining by 98% (±5%) and 77% (±7%) respectively. This implies a substantial shift in the average toxicity per kilogram applied with respect to this key taxonomic group. Looked at in context (Figure 3.2), for contact toxicity these declines are part of an ongoing trend stretching back to at least 2014, whereas for oral toxicity the decline is more recent and associated with the period 2018 to 2020. Some of the drivers behind these trends are explored further in Section 3.3. Other large declines are observed in load associated with long-term toxicity for both birds and mammals, as well as load on algae and aquatic plants.

A further simplification of the same information is shown in Figure 3.5, which is intended as a high-level summary tool for policy to assess progress towards potential target thresholds (which are assumed to be expressed as a net percentage change relative to a prespecified baseline). In this case the metrics are classified into a concise format based on whether the minimum observed change (after uncertainty in the estimation is taken into consideration) exceeds a predefined value (in this example 10% change relative to the values in the earlier year; Figures 3.3 & 3.4). This allows users to identify at-a-glance which metrics may: a) be successfully responding to any interventions (dark green); b) require further intervention to reach the target level (light green); c) be subject to too much uncertainty to confidently assess the direction of change (yellow); or d) be trending in the wrong direction (red). Hence, it is possible to reinforce the fact that the key metrics of interest for 2010 to 2018 are the Bioconcentration factor, Soil DT_{50}, Mobility, Bees oral LD_{50} Algae EC_{50} and Aquatic Plants EC_{50} (all of which are coded red to signify an increase in load), while for 2018 to 2020 the metrics of interest are acute toxicity to birds and mammals and load on parasitic wasps, which are coded yellow to emphasise that these are within the observed margin of uncertainty (at 90% confidence). Tools like this give a quick overview of the key information for a given context and so guide the user in what they may wish to further explore.

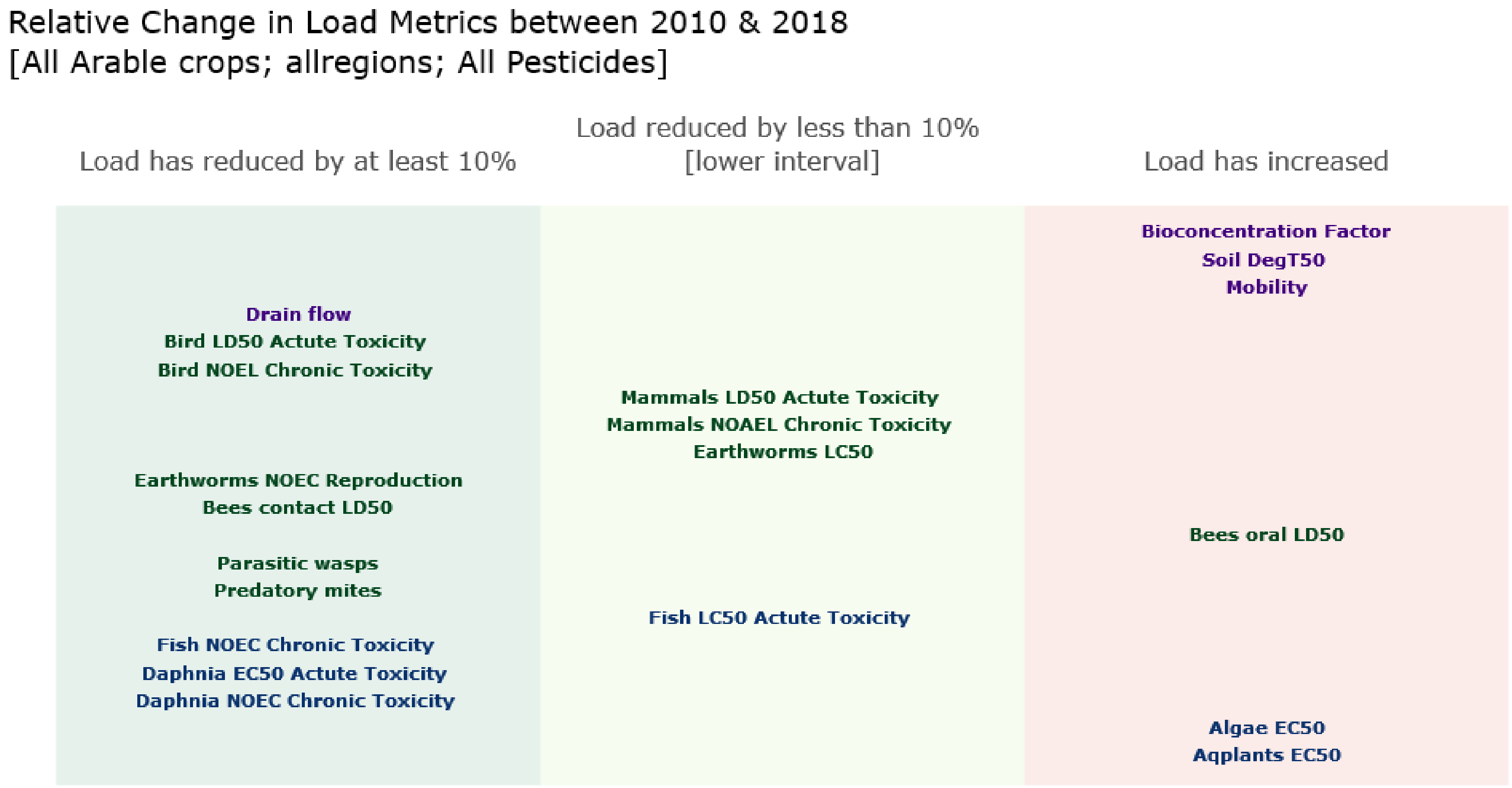

Figure 3.5a: Categorised metrics from the PLI based on percentage change between 2010 and 2018 for the sum of applications made on all arable crops

Underlying values are expressed as percentage change in the total value of the indicator relative to the value in 2010. A threshold target of 10% change has been used in the classification.

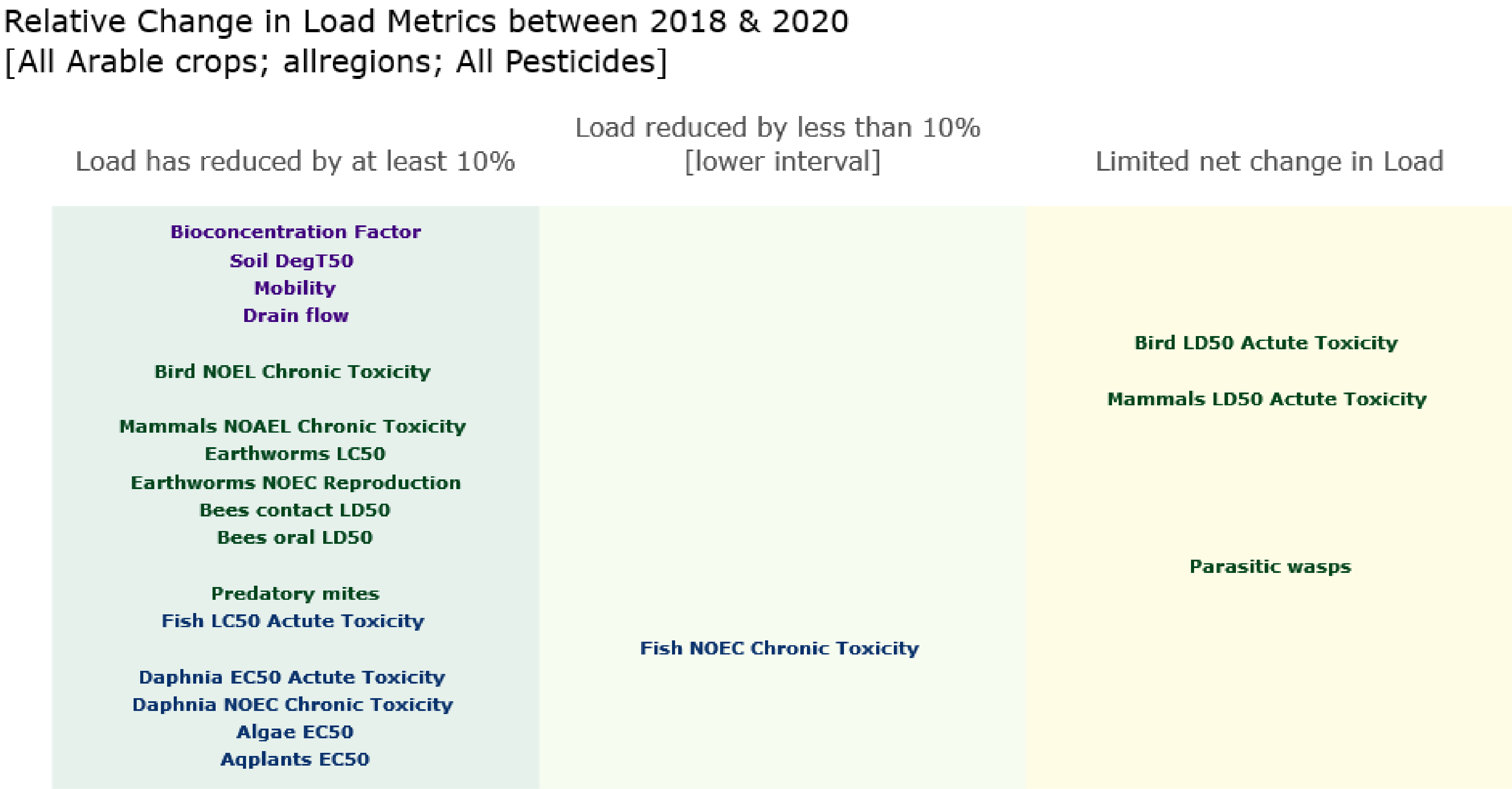

Figure 3.5b: Categorised metrics from the PLI based on percentage change between 2018 and 2020 for the sum of applications made on all arable crops

Underlying values are expressed as percentage change in the total value of the indicator relative to the value in 2018. A threshold target of 10% change has been used in the classification.

As a note when interpreting the PLI it is important to recall that (by design) the indicator is only able to characterise load in terms of the potential pressure placed on the environment by the combination of PPPs applied (see Section 1), and that other lines of evidence, for example the monitoring of real populations (including field level mitigation measures), would be necessary to extrapolate how these pressures might be realised in real ecosystems.

This concludes the overview of the high-level visualisations in the PLI and how these suggest topics for further investigation. Section 3.3 explores how the PLI might be applied to a very specific historic scenario and how it provides additional higher-resolution information to users interested in exploring trends in each metric and linking these trends to specific active substances.

3.3. Case study: Understanding load on insects associated with winter sown oil seed rape 2010-2022

One of the most notable changes in agronomic practice around PPP use in the UK over recent years has been related to the use of neonicotinoid seed treatments and their potential impacts on pollinator populations (Budge et al. 2015; Woodcock et al. 2016). While neonicotinoids have been used across several crops in the UK, discussions of pressure on the environment often centre on OSR, as the major flowering crop that is attractive to pollinators (Goulson 2013; Goulson, Thompson, and Croombs 2018). As well as being an important commercial product, OSR often serves as a break crop for wheat and other cereals. It is affected by several key pests of high concern, most notably cabbage stem flea beetles (CSFB; *Psylliodes chrysocephala*) and peach-potato aphid (*Myzus persicae*) for which neonicotinoid seed treatments were historically recognised as the major chemical control (Budge et al. 2015; Coston et al. 2016; Dewar 2017; Kathage et al. 2018; Scott and Bilsborrow 2019; Lundin et al. 2020).

As background to the trends in environmental load, the total cropping area of winter sown OSR over the studied period peaked in 2012 and has since declined, in part due to shifts in commercial viability in parts of eastern England impacted by CSFB (Figure 3.6). Note that these estimates are based on the official government values published from the June survey(s) in the years for which the arable PUS was conducted. These may show minor differences from other datasets such as the AHDB's Planting and Variety survey and the Defra Basic Payments, all of which have known limitations related to their selected methodologies (AHDB 2020; Defra and RPA 2018). Total use of PPP, on a per hectare basis, has increased in recent years, particularly during the period from 2014 to 2018, which reflects a longer-term trend of increased intensification and management during a period of relatively high prices (Figure 3.7). As with many winter-sown crops (see above), the total mass of applications in 2020 on a per hectare basis is noticeably reduced relative to the historic trend (being roughly comparable with 2010 levels).

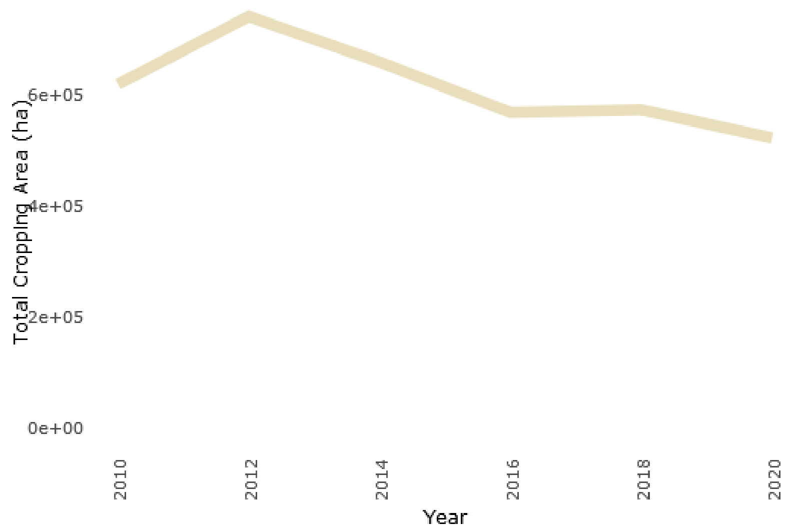

Figure 3.6: Total cropping area of winter sown OSR (2010-20)

Based on the June survey for the years where an arable PUS survey was conducted. Values include summed areas for 2019 for regions of England and Wales (see Section 3.2).

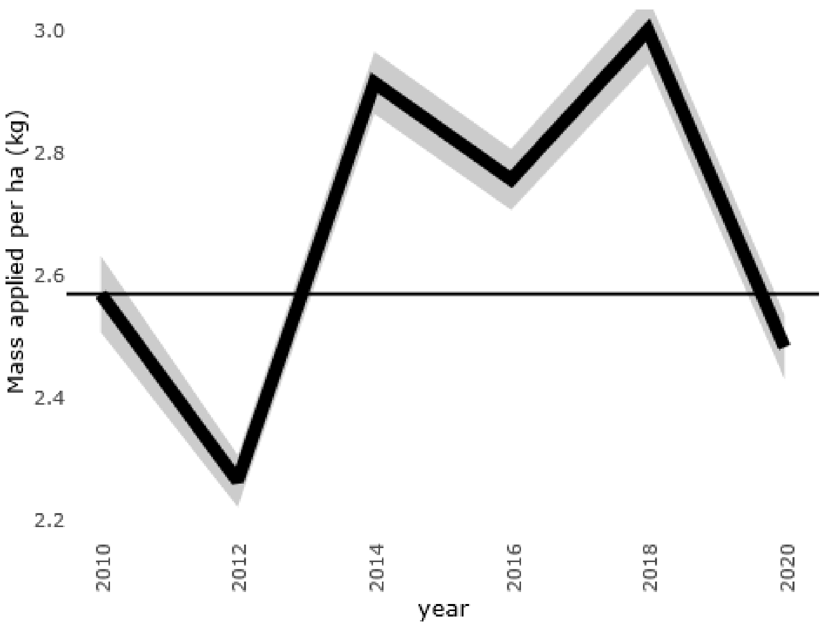

Figure 3.7: Trend in pesticides applied (kg ha⁻¹) to winter sown OSR (2010-20)

Shading around the trend line reflects the 90% confidence interval on the mean. Horizontal reference line shows the mean in 2010.

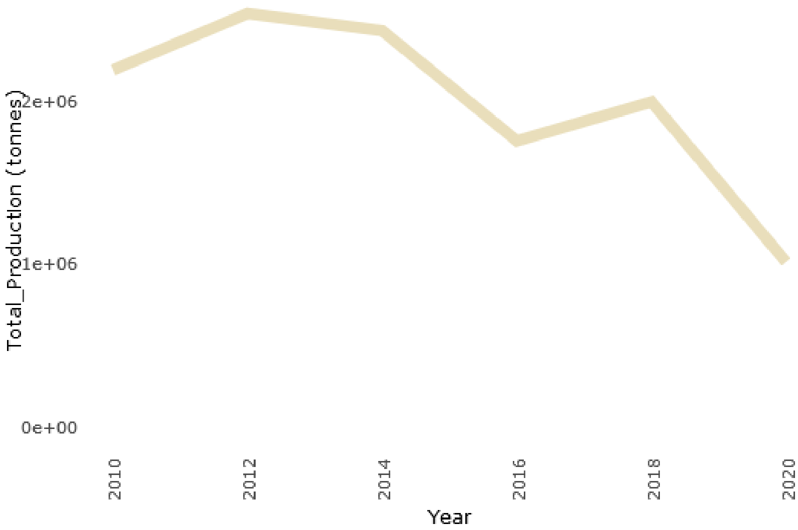

Figure 3.8: Total UK production (tonnes) of winter sown OSR (2010-20)

Prior to 2014, the neonicotinoid seed treatments, clothianidin, thiamethoxam and imidacloprid were widely used on winter sown OSR, but subject to a phased withdrawal prior to the end of the 2016 harvest season. At the time, following neonicotinoid withdrawal, concerns were raised that control of the key pests would necessitate increased foliar applications of pyrethroids and potentially counteract reductions in total environmental load (Zhang *et al*., 2017). Looking at the tends in metrics from around this period (Figure 3.9), it can be seen that relative to their previous level both metrics relating to bees (Bees contact LD_{50} and Bees oral LD_{50}) have shown substantial declines relative to their previous values. However, load on parasitic wasps, particularly during the most recent years between 2018 and 2020, has increased to the point where the metric classification flags this metric as being of concern and trending in the wrong direction (Figure 3.9). When focusing only on the period 2018 to 2020 (Figure 3.10) it can be highlighted that, relative to a notional target of at least 10% reduction relative to 2018 values, load on parasitic wasps has increased (suggesting some change in the mixture of substances applied which may have a negative impact on this taxon), while the recent declines in load on predatory mites, fish (both in terms of acute toxicity and chronic toxicity) and acute toxicity to *Daphnia* are all less than the 10% relative target defined (and indeed most are within the estimated margin of uncertainty).

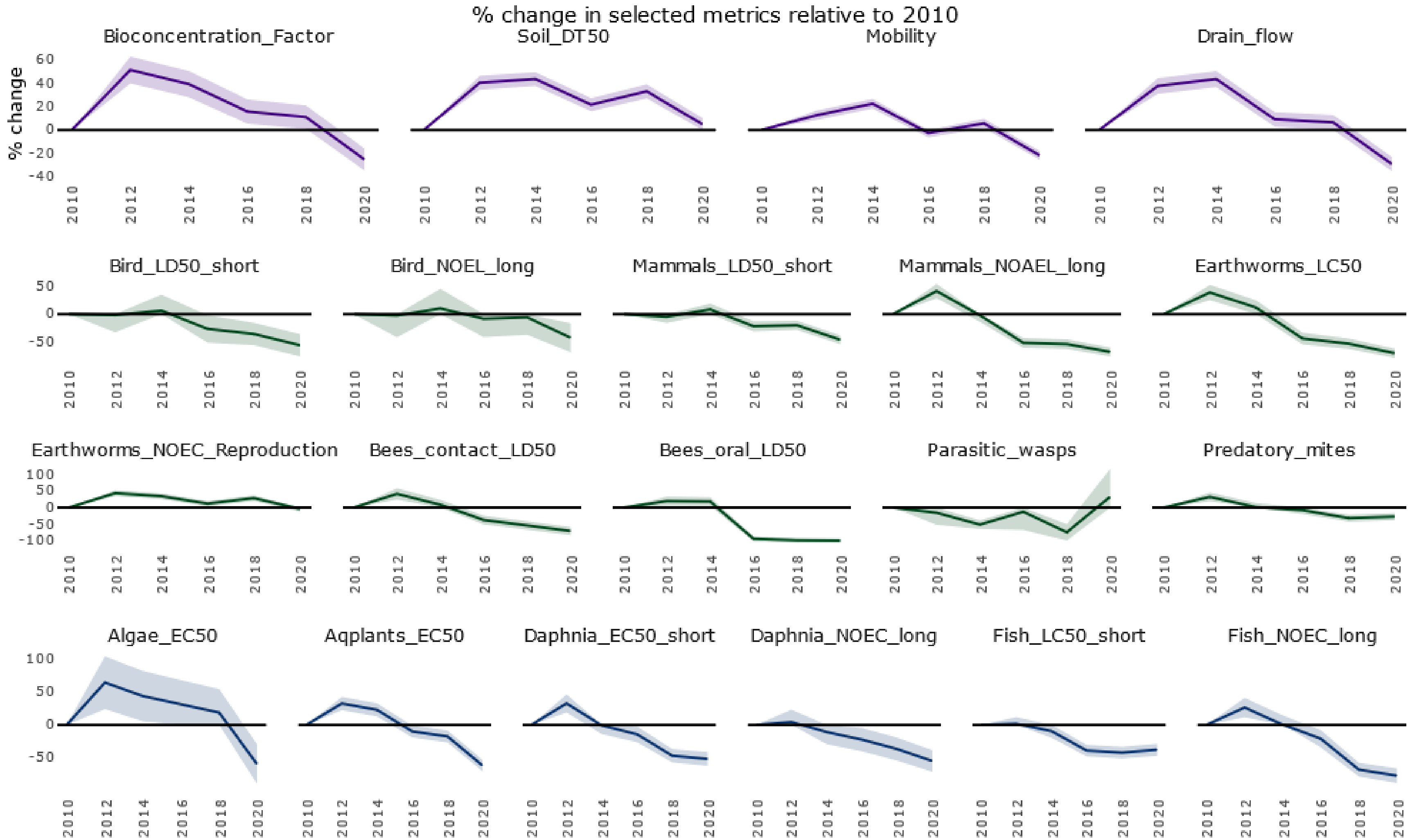

Figure 3.9: National trends in metrics for winter sown OSR (2010-20)

Environmental load is shown with the solid coloured lines. Values are expressed as % change in total PLI units relative to 2010. Shading around the trend lines reflect the 90% confidence interval. Solid horizontal reference line denotes no change relative to 2010.

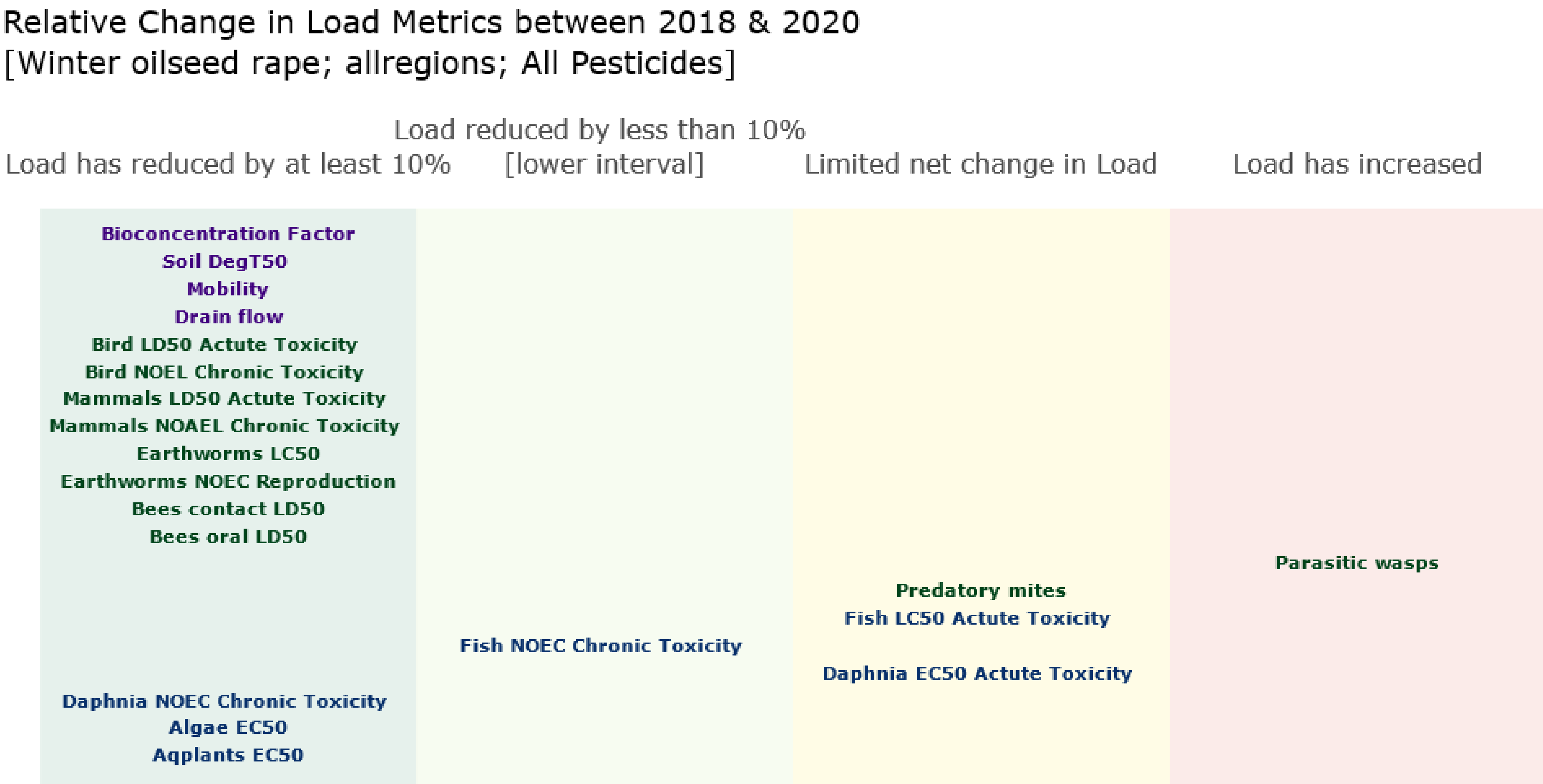

Figure 3.10: Categorised metrics from the PLI based on percentage change between 2018 and 2020 for winter sown OSR

Underlying values are expressed as percentage change in the total value of the indicator relative to the value in 2018. A threshold target of 10% change has been used in the classification.

Examining which specific changes have driven the decline in individual metrics, looking at load on bees specifically, the large declines observed between 2014 and 2016 (Figures 3.9 & 3.11), particularly with respect to oral toxicity, can be almost completely attributed to the withdrawal of the key neonicotinoid seed treatments clothianidin, thiamethoxam and imidacloprid.

While some limited use of clothianidin did persist in 2016 as part of the process of the phased withdrawal, the absolute quantity used in that year was well below previous recorded usage (exemptions were granted only for four counties in Eastern England) and consequently its impact on load was greatly reduced. The more recent changes in load, including the further decline in 2020 from the already low baseline in 2018 (Figure 3.11) are associated with further shifts in the contribution of different pyrethroids. This included a reduction in the use of cypermethrin and zeta cypermethrin in favour of lambda-cyhalothrin, which saw increased usage following the withdrawal of the neonicotinoids as an alternative control for CSFB. From discussion with growers (see also Kathage *et al*., 2018), it seems that following the loss of neonicotinoids as a control option, there has been a general trend towards reduced planting in the areas most afflicted by CSFB (see Coston *et al*., 2016; Scott and Bilsborrow, 2019), as well as a gradual transition towards earlier sowing dates, which may reduce the impact by the pests (Kathage *et al*., 2018). Average yield per unit cropping area for winter sown OSR in the UK has declined since the withdrawal of neonicotinoids (Lundin et al., 2020), although this can still be visualised as a substantial decline in load per tonne of production (total load in 2016 is estimated to be -92% ±8% lower than in 2010 after correcting for difference in output).

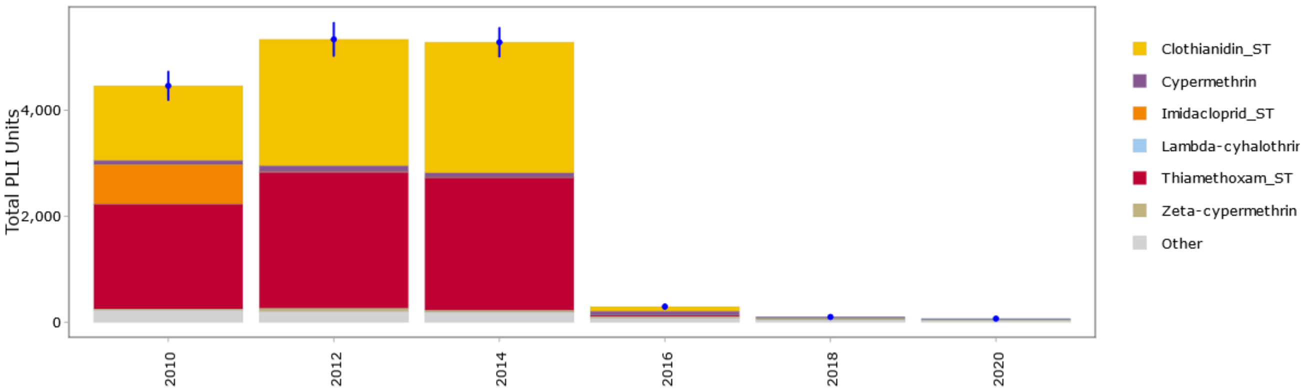

Figure 3.11: Contribution of major pesticides to load on bees oral for winter sown OSR (2010-20)

Values are expressed in total PLI units. Major pesticides are those representing at least 10% of the total value in any year. Blue points denote the mean and the associated 90% confidence intervals.

The fact that a change to authorisation of PPP results in behavioural change beyond the scope of individual active substances, highlights the importance of viewing the use of all pesticides as part of the wider agricultural system and the many different decisions that could be associated with changes in PPP use. Pesticides are only one way in which growers mitigate against risk of loss (albeit one which is often favoured because of their perceived reliability and low costs (Möhring et al. 2019). Hence, if frameworks like Integrated Pest Management (IPM) are to have their expected impact, PPP usage needs to be viewed within the context of growers (and agronomists) risk avoidance behaviour and a systematic view needs to be taken of how to best manage pest related risk within the landscape (Deguine et al. 2021).

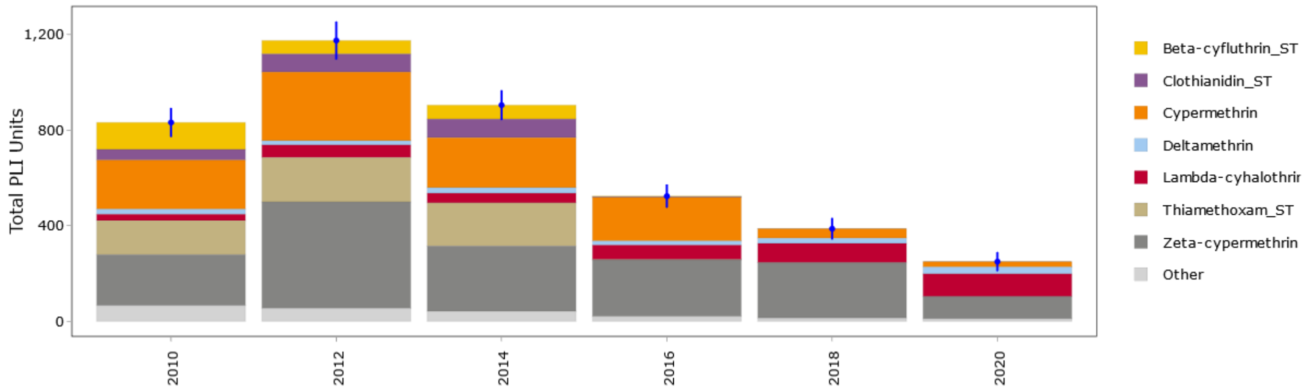

Figure 3.12: Contribution of major pesticides to load on bees contact for winter sown OSR (2010-20)

Values are expressed in total PLI units. Major pesticides are those representing at least 5% of the total value in any year. Blue points denote the mean and the associated 90% confidence intervals.

Examining the load from Bee contact toxicity (Figure 3.12), many of the same patterns emerge. Compared to oral toxic load (for bees) the major neonicotinoid seed treatments clothianidin and thiamethoxam make up a lower proportion of the overall load value, although again, with their common co-formulant beta-cyfluthrin, they make up the predominant drivers of change between years[15]. Here, the transition to the use of pyrethroids can be seen more clearly, with cypermethrin (and zeta cypermethrin) becoming much less important between 2018 and 2020 with small compensatory increases in deltamethrin and lambda-cyhalothrin. Cypermethrin has been the target of a number of recent campaigns aimed at reducing usage (e.g., Environment Agency 2019) and mesocosm studies have shown that '*cypermethrin application had a somewhat greater impact than the lower lambda-cyhalothrin treatment rate (due to effects on peracarid crustaceans)*' (Farmer, Hill, and Maund 1995). Cypermethrin has also been identified as 1.16 and 3.02 times more toxic to fish compared to deltamethrin and lambda-cyhalothrin respectively (Farmer, Hill, and Maund 1995). Cypermethrin has also been identified as 1.16 and 3.02 times more toxic to fish compared to deltamethrin and lambda-cyhalothrin respectively (Salako et al. 2020). However, it should be noted that all three pyrethroids are considered dangerous to aquatic and terrestrial life and are subject to strict controls regarding treatment rates and the maximum number of applications in a season.

Breaking down the metrics in this way, so that each draws on (as far as possible) directly comparable datasets, highlights one of the conceptual strengths of the revised PLI approach (particularly when compared to previous results which used an opaque

[15] It is worth noting that the increased use of pyrethroids is likely a response to the need to control CSFB in the absence of neonicotinoid seed treatments. The differences in mode of delivery, presence of other management practices, and the different regulations which surround insecticide applications mean that this change is unlikely to be a one-to-one replacement in many cases.

aggregated metric; see Rainford et al., 2022, Section 3). Whether it is most appropriate over the longer term to focus discussion of trend around individual metrics (and the PLI), or taxonomic groups (as in the recently developed TAT indicator (Schulz et al. 2021, see Section 4.3) is a topic for ongoing discussion. However, having both options is considered an important first step towards the long-term future of how indicators of this type may be presented and the implications for their use in policy decisions. However, it should be noted that all three pyrethroids are considered dangerous to aquatic and terrestrial life and are subject to strict controls regarding treatment rates and the maximum number of applications in a season. Breaking down the metrics in this way, so that each draws on (as far as possible) directly comparable datasets, highlights one of the conceptual strengths of the revised PLI approach (particularly when compared to previous results which used an opaque aggregated metric; see Rainford et al., 2022, Section 3). Whether it is most appropriate over the longer term to focus discussion of trend around individual metrics (and the PLI), or taxonomic groups (as in the recently developed TAT indicator (Schulz et al. 2021), see Section 4.3) is a topic for ongoing discussion. However, having both options is considered an important first step towards the long-term future of how indicators of this type may be presented and the implications for their use in policy decisions.

With respect to the potentially worrying apparent increase in load observed on parasitic wasps (Figure 3.13), while the trends for load prior to 2014 show strong impacts of some of the substances mentioned above (notably imidacloprid seed treatments and alpha cypermethrin), the most recent trends, including an increase in load between 2018 and 2020, are largely attributable to the use of the substance acetamiprid (Figure 3.13). The metric 'parasitic wasps' here represents toxicity studies conducted on *Aphidius rhopalosiphi* (Hyemoptera Braconidae: Aphidiinae), which parasitize specific aphids, and which together with the predatory mite *Typhlodromus pyri* (Acari, Phytoseiidae; commonly found in orchards and predates a number of species of mite including Fruit Tree Spider mite *Panonychus ulmi*), is one of the two principal first tier models for risk assessment with respect to 'non target arthropods'. These species were selected due to a combination of ease of rearing and high sensitivity to pesticide impacts (see e.g. EFSA Panel on Plant Protection Products and their Residues (PPR) 2015). It is worth noting that, while load does increase over this period, looked at over a longer period (for example in comparison to 2016), the situation is much less clear cut and the very wide estimated confidence intervals imply a large degree of heterogeneity in usage between individual holdings and regions, particularly in 2020 where, for reasons outlined above, there are also concerns about the wider data quality.

Acetamiprid is a neonicotinoid used principally to target insects with sucking mouth parts including aphids, Thysanoptera and Lepidoptera[16], and since 2021 is the only neonicotinoid that can be applied in open field cultivations in the EU and UK (Varga-Szilay and Tóth 2022; applications restricted to a single treatment per field per year prior to the end of flowering). Unlike other neonicotinoids, acetamiprid is primarily associated with foliar or ground sprays and not applied as a seed treatment in the UK. In terms of its history of use on OSR, in the autumn of 2014 and 2015, an emergency authorization was granted for the product InSyst (acetamiprid) for use against CSFB,

[16] http://www.agchemaccess.com/Acetamiprid

but there has been no further authorization for use to control this pest in subsequent years (White et al. 2020). Anecdotal information is skeptical as to its efficacy as a means of CSFB control (White et al. 2020). Acetamiprid (and indoxacarb) are also cited as the key control agents for pollen beetle *Meligethes aeneus* (Denholm 2011; Burnett et al. 2020), for which InSyst has approval up to 2025, although the extent to which this drives usage relative to pyrethroids is unknown. In discussions of resistance management (particularly where pyrethroids are ineffective), acetamiprid is often cited, alongside indoxacarb, as a potential component of a joint spray programme, but the extent to which this represents current practice and how this contributes to the overall usage statistics is unclear (Insecticide Resistance Action Group 2022).

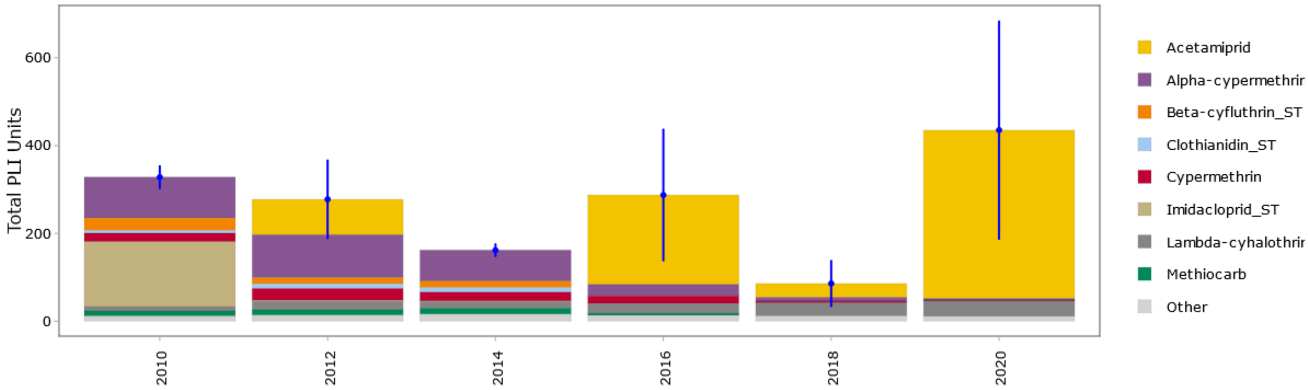

Figure 3.13: Contribution of major pesticides to load on parasitic wasps for winter sown OSR (2010-20)

Values are expressed in total PLI units. Major pesticides are those representing at least 5% of the total value in any year. Blue points denote the mean and the associated 90% confidence intervals.

Acetamiprid has been suggested to have a relatively low toxicity to honey bees (at least when compared to the other neonicotinoids (Zhu et al. 2015), although see Varga-Szilay and Tóth (2022) for a dissenting opinion). In a recent EFSA review of the state of evidence around acetamiprid (Hernandez Jerez et al. 2022), its impacts on survival, reproduction, growth and behavior of bees are rated low to moderate (with low confidence in the latter), with the evidence for the potentially higher sensitivity of *Megachile rotundata* (as a proxy for wild bees, see also Camp et al. 2020; Varga-Szilay and Tóth 2022) being highlighted as an area of concern for further investigation. In their overall conclusions Hernandez Jerez et al. (2022) stated that *'no conclusive, robust evidence of higher hazards compared to the previous assessment was found for birds, aquatic organisms, honeybees and soil organisms'*. No explicit discussion is made in this document for data pertaining to beneficial invertebrates (the category to which parasitic wasps belong), although in the proposed interim registration review decision for acetamiprid by the US EPA '*Commenters also stressed the relative safety of acetamiprid to workers and to beneficial insects and pollinator species, as compared to other pesticides, including in comparison to other neonicotinoid pesticides'* (US EPA 2020; see also Smitley et al. 2019; Ambrose 2003). The product label for Insyst[17] lists

[17] https://www.pcs.agriculture.gov.ie/media/pesticides/content/products/labels/03249%20-%20Insyst%20-%202019%20to%20date.pdf

this product as ‘slightly toxic to predatory mites and generally slightly toxic to other beneficials’ (which would include parasitic wasps).

Arguably this lack of weight in the decision making process reflects a systematic reliance in regulatory practice on data from bees (and honey bees in particular) as (in effect) proxies for all invertebrates, a topic that has been raised as a concern by a number of authors (Franklin and Raine 2019; Siviter and Muth 2020). Despite their importance as pollinators, the highly colonial life style of some bee species may result in different susceptibility when compared to other arthropods (Franklin and Raine 2019). The question of what taxa to incorporate into the context of risk assessment is a complex one, involving a number of trade-offs between tractability and cost vs. ensuring that risk assessment is protective with respect to the wider community (Schäfer et al. 2019). As was previously the case with bumblebees and solitary bees (Lewis and Tzilivakis 2019) the growing volume of toxicity data relating to parasitic wasps, and to beneficial arthropods generally, is slowly enhancing tools for the assessment of PPP but has not yet fully filtered through to some of the widely implemented tools and indicators developed in this space (see Rainford, Kennedy, and Jones 2021).

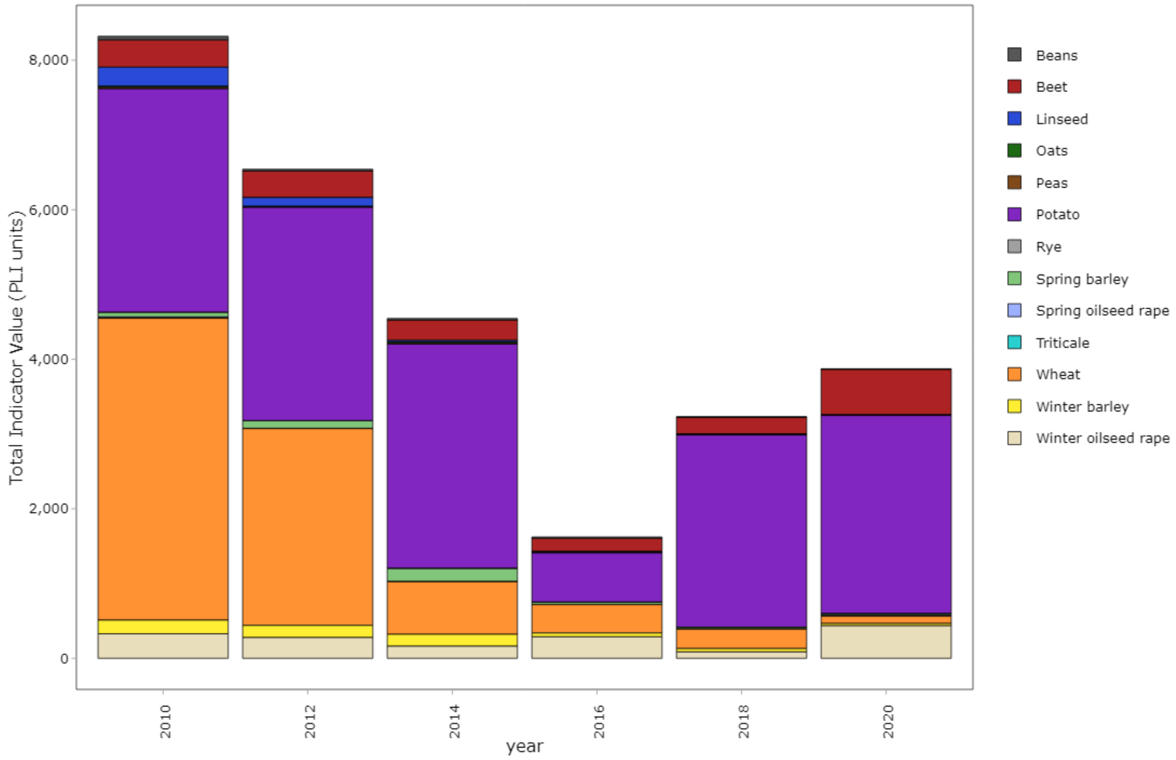

Figure 3.14: Contribution of major crops to load on parasitic wasps (arable crops 2010-20)

Values are expressed in total PLI units.

Viewed in the wider context of UK arable cropping, the load on parasitic wasps from winter sown OSR is only a part of the wider picture associated with this metric. The most important individual component of overall load is generated by applications made on potato, generally the most intensively treated crop in UK arable farming, and again a role for acetamiprid in the most recent data (Figure 3.14), alongside the dominant role played by oxamyl, is seen. Similarly, when the same metric for applications made on sugar beet is examined (Figure 3.15), acetamiprid is again seen as an increasingly important component in 2020 (note that while emergency authorisations of neonicotinoid seed treatments on sugar beet were approved in principle for 2020, they were not required because the predicted aphid threshold, based on winter temperature needed for approval, was not reached and hence these are not considered here).

Pre-2014, applications on wheat (Figure 3.16) and, in particular, the use of dimethoate (now withdrawn), was an important driver of trends in load with respect to parasitic wasps. In 2020, across all arable applications (Figure 3.17), acetamiprid totals 30% of the total load on parasitic wasps, which together with oxamyl is the one of the principal substances that is responsible for the majority of load, and which is most linked to recent change. This serves to highlight how having a combined tool that can examine data from many angles helps to support decision makers in pulling together priorities for future investigation. When ranked alongside other insecticides and nematicides across all 20 included metrics in the PLI (mean average of ranks on individual metrics by relative percentage contribution), acetamiprid is ranked 14 out of 24, well behind such important compounds as lambda-cyhalothrin, oxamyl, cypermethrin and tefluthrin. This indicates that while acetamiprid may be a compound to watch when it comes to load on parasitic wasps, it still needs to be viewed in the context of the wider hazard profile of PPP applications and judged accordingly in comparison to potential alternatives.

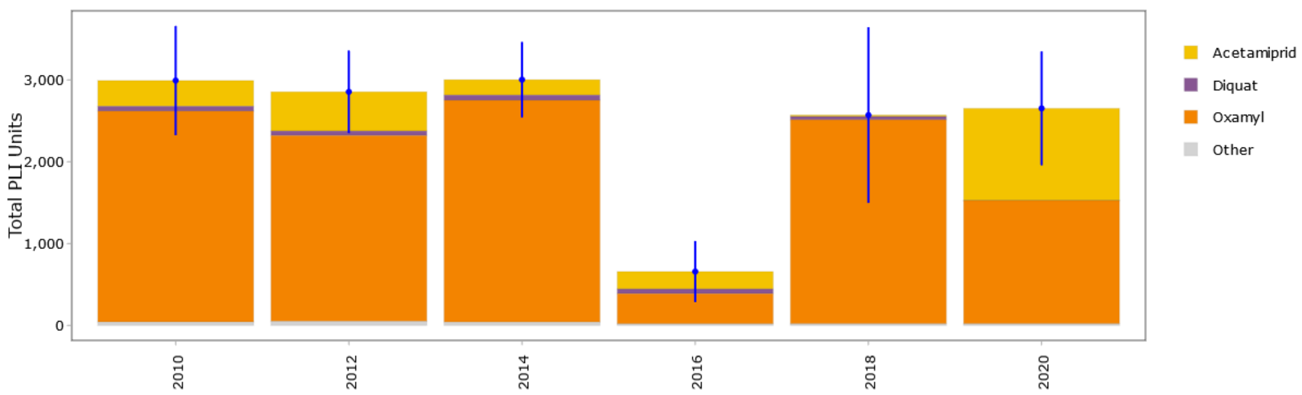

Figure 3.15: Contribution of major pesticides to load on parasitic wasps (Potato 2010-20)

Values are expressed in total PLI units. Major pesticides are those representing at least 5% of the total load value in any year. Blue points denote the mean and the associated 90% confidence intervals.

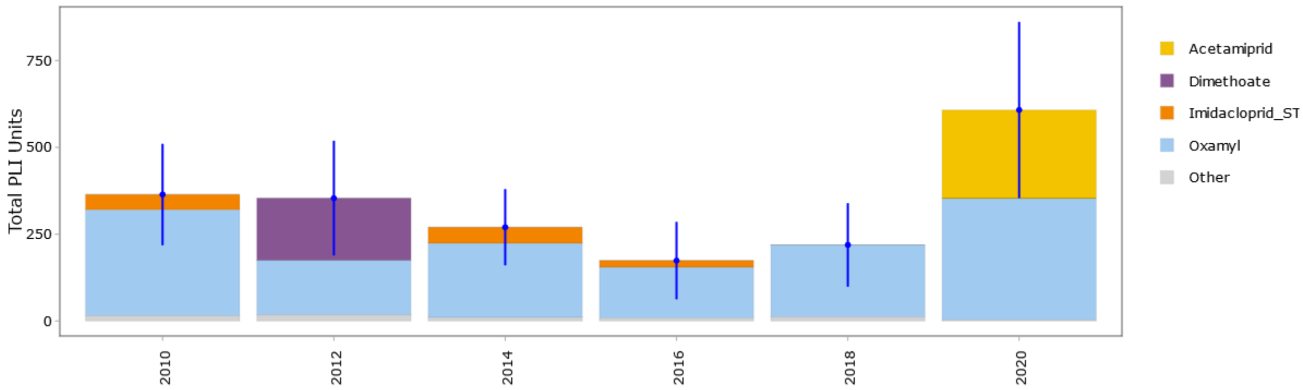

Figure 3.16: Contribution of major pesticides to load on parasitic wasps (Beet 2010-20; GB only)

Values are expressed in total PLI units. Major pesticides are those representing at least 5% of the total load value in any year. Blue points denote the mean and the associated 90% confidence intervals.

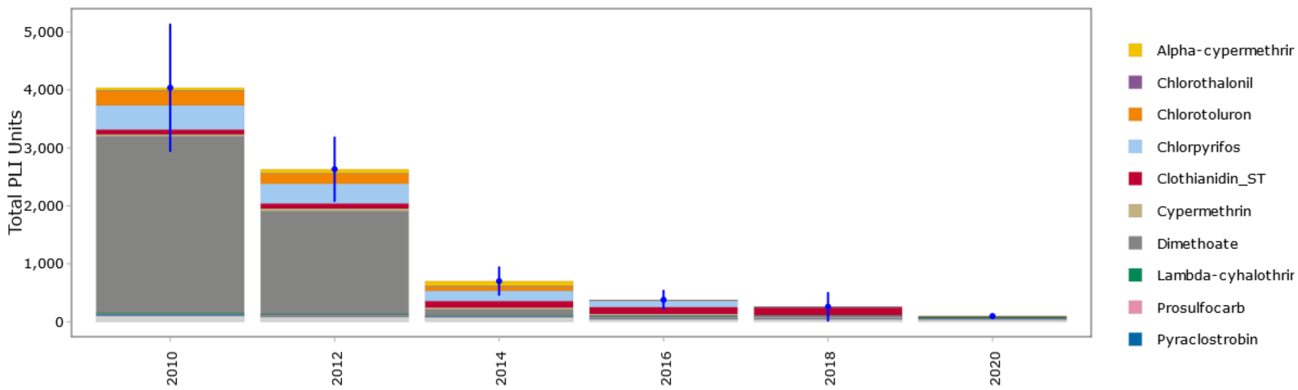

Figure 3.17: Contribution of major pesticides to load on parasitic wasps (Wheat 2010-20)

Values are expressed in total PLI units. Major pesticides are those representing at least 5% of the total load value in any year. Blue points denote the mean and the associated 90% confidence intervals.

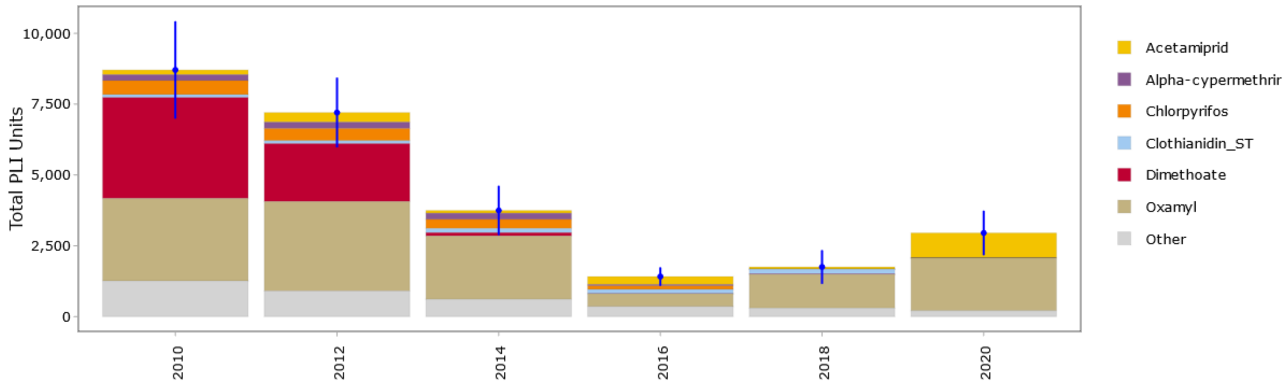

Figure 3.18: Contribution of major pesticides to load on parasitic wasps (all arable crops; 2010-20)

Values are expressed in total PLI units. Major pesticides are those representing at least 5% of the total load value in any year. Blue points denote the mean and the associated 90% confidence intervals.

Given the fact that acetamiprid appears the be one of the key substances associated with the trend towards increased load on parasitic wasps, and their various uncertainties surrounding usage, it is recommended that use, and the potential impact of this substance, continue to be monitored in future. To be clear, there is currently insufficient evidence to indicate that acetamiprid is a concern for parasitic wasps or indeed any other taxonomic group, but it does highlight the role which tools like the PLI can have in prioritisation and identifying substances that may warrant further investigation.

This role as a tool for prioritisation is closely linked to the wider role of the PLI as a tool for post-authorization monitoring of PPP to help policy and decision makers understand the changing composition of substances used and their implications for environmental impacts. Since the publication of the concept of 'pesticidovigilance' (Milner and Boyd 2017), there has been increasing discussion around the fact that the current authorization process around PPP struggles to be reactive and informed by changes in agricultural practice. While limited by the inherent delays associated with collecting and aggregating usage data via the PUS, the PLI does help to contextualize current and historic practice to help the user to understand the potential direction of change within the landscape.

4.0. Comparisons between the PLI and the Total Applied Toxicity indicator

The PLI is one of a broader family of pesticide indicators which explore changes in the mixture of pesticides applied over a given geographic scope and time, by summation of the mass applied scaled by some physical or biochemical property of the substances involved. Such indicators have a long history as tools for agricultural policy and decision making and have been developed by a wide range of authors for a wide range of different purposes. Readers interested in review of the various classes of indicators as well as their use in policy and decision support are directed to discussion in Rainford, Kennedy, and Jones (2021) which includes a comparison of the Danish PLI with a number of similar indicators with emphasis on the different scope and approaches to calculation used in different tools.

The Total Applied Toxicity (TAT) is an indicator initially developed by Schulz et al. (2021) for use in the USA and a similar indicator has also recently been applied in Germany (Bub et al. 2023). TAT has recently been discussed as a novel international standard for reporting of the potential impacts of pesticides, for example under the Convention on Biological Diversity (Open-ended working group on the post-2020 global biodiversity framework 2022) and Beyond 2020[18]. While not identical in either form or function, the scope of the PLI in its revised context shows sufficient similarity with the TAT to allow comparisons and to explore the possibility of aligning UK reporting around PPP to this emerging international standard. This section summarises the key conceptual similarities and differences between the PLI as described above and the TAT. It is not a full protocol for how the TAT might be calculated in a UK context (which may follow depending on agreement with UK stakeholders and policy) but rather a targeted discussion of the key similarities and differences between the approach adopted for the PLI (see Section 2) and what might be required for the TAT to be calculated in a UK context.

As a note, while based on the same principles, the two implementations of TAT differ sufficiently in their precise approach (see below) that they are henceforth denoted as TAT_{USA} and $TAT_{Germany}$. At the highest level, both TAT implementations share a common structure with the PLI. Like the PLI, the TAT serves to provide a proxy for the potential environmental impact of PPP use by combining information on usage rates (represented by the estimated mass of application of different active substances) with data on ecotoxicity, represented by Regulatory Threshold Levels (RTLs) in the TAT. The PLI also includes fate elements serving as proxies for the persistence, mobility and bioaccumulation of PPP which have no equivalent scope within the TAT and will be excluded from this discussion.

The key differences between the PLI and TAT approaches lie in a) the underlying data sources that provide the raw values for ecotoxicity assessment; b) how data from different regulatory assessments conducted on the same taxonomic group are combined; and c) the way in which usage data is collected. This section briefly

[18] https://beyond2020.se/

discusses how these differences affect the scope of the different indicators and the implications this has for comparing trends in the PLI and the TAT in a UK context.

4.1. Core principles

The core structure of the PLI and TAT indicators is identical and common across each of the implementations considered here (the Danish PLI has the same underlying structure but is aggregated across multiple measurements and is not discussed in detail here, see Rainford et al., 2022b). For both the (UK) PLI and TAT the overall value of interest for a given component (a metric, or taxonomic group) across a set of applications involving multiple active substances can be expressed using Equation 4.1:

$$\text{Total Indicator score of a component} = \sum \text{Mass}_{as} \times \frac{1}{\text{Ecotoxicty}_{as}} \qquad \text{(Equation 4.1)}$$

Where:

Mass_{as} and the measure of Ecotoxicity_{as} are values specific to a given active substance.

In practice, a measure of ecotoxicity is itself a mass (or dose) of the active substance which is associated with some measured toxic effect (e.g., decreased survival, growth, or reproduction) in a laboratory setting. For both the PLI and TAT, such values may be first standardised for example by scaling relative to the worse case substance (see Rainford et al., 2022). The individual components of both the PLI (i.e., the 'load metrics') and TAT are thus in practice unitless indicators intended to express relative change within their respective contexts.

How components are defined marks the first key difference between the PLI and TAT. In the PLI, individual components (load metrics) are representations of a specific measurement made during the first-tier authorisation of PPP. Each component thus has a specific corresponding study type and methodology and (ignoring for the moment the treatment of missing data) all the values under that metric are derived from the same or very similar methods and are directly comparable. For example, the PLI metric 'Daphnia acute' represents the results of studies conducted on the mortality of the crustacean *Daphnia* spp (usually *D. magna* conducted over 48 hours and measured as the EC_{50} (mg L^{-1}) of a given active substance[19]). Thus, the components of the PLI can be thought of as being a direct link to the results of the underlying laboratory studies maintained in a common framework which is the scope and intention of the PPDB as a data source (Lewis et al. 2016). As a reminder in the rare case of duplicated values matching the same scope the data reported in the PPDB and PLI represent the worst-case value subject to scientific quality standards.

The way in which components of the TAT (in both its German and US implementations) are defined is subtly different. Specifically, the TAT aims to capture the worst-case value for an active substance from any regulatory test/study that has been performed on the named group associated with a particular sub component (e.g., birds, aquatic invertebrates, or pollinators). To continue the analogy above, 'Daphnia

[19]The vast majority of 'Daphnia acute' studies will correspond to the following guidance published by the OECD https://read.oecd-ilibrary.org/environment/test-no-202-daphnia-sp-acute-immobilisation-test_9789264069947-en#page1

acute' is a proxy for risks to all aquatic invertebrates (and represents one of the primary measurements taken in the first tier of risk assessment process). However, in most regulatory regimes (including the UK) during risk assessment, measurements of mortality would in practice be taken from a wider array of different aquatic invertebrate species and over different time periods and conditions to those outlined above (for example during higher tier assessments). Under the metrics defined in the PLI, this supplementary information is discarded because it is often not directly comparable between different active substances (which may have been based using different methodologies and model taxa). The TAT takes a different view in that it first collates all the regulatory values that pertain to aquatic invertebrates in the source databases (regardless of which specific species were used or the length of the assessment) and then selects from among them a 'worst case' value to represent the specific active substance. The logic here is that there may be substances which for one reason or another mortality of *D. magna* conducted over 48 hours is a poor proxy for the overall potential impact on aquatic invertebrates. By considering a wider array of data, this is potentially being more 'realistic' in terms of the potential relative risks (which also aligns to the way in which risk assessment is conducted in that the decision is based on the reasonable worst-case value across a range of studies).

4.2. How active substance values are generated in the TAT

In almost all cases, the worst-case value for a given taxonomic group will be some sort of minimum amount of substance that affects the taxon concerned (typically expressed as a concentration). However, taking this view potentially means (in some cases) that data may be compared that are derived from different kinds of studies (for example those considering acute vs. chronic effects). The way the TAT deals with this is defining what they term 'regulatory threshold levels' or RTLs. An RTL is a combination of a measurement for a given end point (e.g., acute or chronic mortality) with some predefined (and regulatory relevant) constant (termed the assessment/ adjustment factors), which controls how different studies are to be compared when selecting which value represents the 'worst case' value for a given active substance. This is closely analogous the way in which a *Predicted No-Effect Concentration* (*PNEC*) might be calculated in risk assessment (although not necessarily identical due to the reduced consideration of factors which might influence localised factors linked to exposure through time). From the $TAT_{Germany}$, when generating a worst-case value for 'Aquatic invertebrates', measurements of acute toxicity are first divided by a factor of 100, while measurements of chronic toxicity are only divided by a factor of 10, before calculating the minimum that is treated as representative of the active substance going forward.

These risk assessment factors are, as the name implies, taken from the risk assessment process of the given regulatory regime (and notably differ between the given values for TAT_{USA} and $TAT_{Germany}$) and are intended to 'represent the uncertainty in the representativeness of the considered endpoints for other non-target taxa' (i.e., to ensure that risk assessment decisions are protective of species other than those directly used in the underlying measurements). Similar assessment factors are also used to distinguish between studies conducted on different model species in the $TAT_{Germany}$ for the 'Terrestrial Vertebrates' and 'Soil organisms' sub-components. In the case of ‘Terrestrial Vertebrates’ a further scaling (termed adjustment factors) has been applied to studies based on feed vs. those based on body weight (values from feed studies are multiplied by either 5 or 10 prior to calculating the ‘worst-case’ value depending on whether the measurement is acute or chronic). It is worth noting that because the TAT is unitless and best expressed as relative change, in cases where assessment factors are defined as constants across all study types, they can in effect be ignored (as they will have no impact on the way relative values change over time).

This kind of scaling of values has a long history in risk assessment around PPPs and represents a clear attempt to align the indicator to different risk assessment regimes in the US and Germany. The fact that (as documented) the assessment and adjustment factors used are inconsistent represents an issue in the international comparability of trends in TAT across different nations and potentially is something that may need to be taken into consideration if the approach were to be adapted to the UK. Given the fundamentally similar underpinnings of EU and UK risk assessment protocols it is likely that the approach outlined for the $\text{TAT}_{\text{Germany}}$ could be followed, although this is a topic that needs wider consideration if the indicator is to be used for comparison at an international level.

Equations 4.2 and 4.3 summarise the differences between the indicators:

$$\text{PLI}_{\text{single metric}} = \sum \text{Mass}_{\text{as}} \times \frac{1}{\text{Toxicty}_{\text{as}}} \qquad \text{(Equation 4.2)}$$

Where:
$\text{Toxicty}_{\text{as}}$ is measured by a **specific** corresponding methodology

$$\text{TAT}_{\text{focal taxonomic group}} = \sum \text{Mass}_{as} \times \frac{1}{\text{Minimum Toxicty}_{as}\ or\ RTL_{as}} \qquad \text{(Equation 4.3)}$$

Where:
Minimum Adjusted $\text{Toxicty}_{\text{as}}$ or RTL_{as} is measured by **ANY** regulatory study on the focal group

Hence, the TAT can be thought of as a somewhat more broadly scoped and more aggregated approach compared to the PLI, which uses the same information within the PLI metrics in combination with any other regulatory studies for that active substance on that taxonomic group to try be conservative about what its potential risks might be. The chief advantage of this broader scope is that it may (in some cases) reveal risk associated with active substances that are not captured within the specific metrics used in the PLI, and that the approach overall is more closely aligned to existing risk assessment practice. The primary drawback is a loss of transparency in linking values back to specific measurements and ambiguity as to the best treatment of missing data (see below). Which of these should be prioritised is partially a question about the scope and usage of the indicator and as such is difficult to define without discussion with policy makers and other key stakeholders.

4.3. Differences in data sources

While the text above captures the primary differences between the subcomponents of the PLI and each of the TAT implementations that currently exist, it is worth noting that the most important difference between TAT_{USA} and $TAT_{Germany}$ is their choice of backend database from which they draw data from regulatory studies. In both cases, the choice of backend database has been aligned to the national context for the country in which the indicator has been calculated. Thus, the TAT_{USA} prioritises data sources held by the US EPA including the ECOTOXicology knowledgebase[20] and the OPP Pesticide Ecotoxicity Database[21] (with the EFSA OpenFoodtox database[22] performing a supporting role as a source of supplementary data for "mammals and terrestrial arthropods for which no other relevant study endpoint from US databases was available"). By contrast, $TAT_{Germany}$ list the EFSA OpenFoodtox database as the primary data source, supplemented by the Pesticides Properties DataBase (PPDB) and the US EPA Office of Pesticide Programs Pesticide Ecotoxicity Database. Despite the apparent overlap it becomes apparent that when reported values for the same set of active substances are compared (e.g., between those reported in the published TAT_{USA} vs. the PPDB) there can be large discrepancies in the exact value reported. Tracing the source of these issues without detailed examination of the source databases is challenging and highlights that if the TAT is to be adopted more widely a degree of consensus as to the appropriate data sources and how they are prioritised would be required. Given the work around the PLI, the PPDB is naturally favoured as the primary information source in a UK context (potentially supplemented by the EFSA OpenFoodtox database[23]). However, it is difficult to judge which databases may be most suitable in general for an international comparison and it is difficult to comment on the potential role that US data sources might be expected to play. Were the TAT to be adopted in a UK context it is recommended that the protocols used are modelled closely on those outlined for the $TAT_{Germany}$ (as this has the greatest alignment to regulatory practice in the UK), potentially with a wider role for the PPDB as a way of resolving missing data.

One way in which the difference in information sources manifests between the different indicators is in the handling of missing information. From the supporting dataset for the TAT_{USA} and the descriptions in the $TAT_{Germany}$, it can be seen that both indicators incorporate extensive amounts of missing data for individual active substances, and unlike the PLI, the TAT lacks any explicit protocol for how missing values might be inferred. Given the additive nature of the overall indicators, the impact of these missing data is equivalent to treating these values as 0's (i.e., for the reported value of the TAT change in substances with missing data for a given taxonomic group will have no effect on the trend in the overall indicator for that group). As noted in previous PLI reports, the assumption that a lack of a reported regulatory value is equivalent to no load being generated by the use of a substance is potentially highly flawed (Rainford et al. 2022b), and could lead to the overall indicator underrepresenting pressure on the landscape if

[20] https://cfpub.epa.gov/ecotox/
[21] https://ecotox.ipmcenters.org
[22] https://www.efsa.europa.eu/en/data/chemical-hazards-data
[23] https://www.efsa.europa.eu/en/data-report/chemical-hazards-database-openfoodtox

unaddressed. We, the current authors, remain of the opinion that the treatment of missing data is a key concern for the roles which indicators like the TAT and PLI are expected to play in policy development. Likewise, the scope of the TAT (i.e., which substances contribute to the overall value) is set by occurrence of appropriate data in the underlying databases examined, rather than being subject to an explicit protocol as outlined in Section 2.3. This potentially means that unless adapted to their explicit inclusion the contribution of non-traditional substances such as biopesticides and botanicals may not be reflected in the calculated value for the TAT.

If the TAT were to be adopted in its current form alongside the PLI there is the potential to create inconsistency in the PLI approach to missing data which would need to be addressed (as the PLI explicitly assigns nonzero values to substances with missing data for specific metrics). This is potentially not a serious issue, but it is one which warrants discussion in terms of how the indicators are presented.

4.4. Differences in estimating usage

The final and least important difference between the TAT implementations and the PLI is the source of the estimates of usage. In the TAT (both in the USA and Germany) these are based on national sales data aggregated in various ways across different active substances. For the PLI, equivalent values are based on estimates of usage based on the sample of holdings assessed during the UK PUS. In practice, if the TAT were to be adapted to a UK context it is likely it would be based on the PUS data (which provides a richer resource than the often highly aggregated sales records) but this is something that needs to be considered, e.g., in the treatment of uncertainty. There is no easy comparison to be made between information on sales and information of usage (given that there are very few cases where both kinds of data have been collected over the same period and scope). However, because long term storage of PPP by growers appears to be rare it is considered reasonable that usage in any given year should correlate reasonably well with sales over that same period in the majority of cases (Thomas et al 1999).

4.5. Summary and Discussion

In the opinion of the authors, it would be both feasible and relatively straightforward to implement something close to the methodology outlined for the $TAT_{Germany}$ to the UK given the information in the PPDB, EFSA OpenFoodtox database and infrastructure created for the PLI. Key decisions that would need to be taken include: a) confirming the prioritisation of data sources (PPDB vs. EFSA OpenFoodtox); b) confirming the scope of the metrics to be included (the PLI has traditionally not included any metrics relating to terrestrial plants which are listed as a sub component in $TAT_{Germany}$); c) ensuring agreement on the assessment/adjustment factors and other components of how different studies should be compared when generating a ‘worst-case’ value; and d) agreeing a protocol for the treatment of missing data and ensuring this is aligned and consistent with the approach presented in the PLI.

5.0. Reflections and conclusions

5.1. Overview

The development of a national or regional indicator that aims to account for both the amounts applied and properties of pesticides (with respect to characterising their potential environmental impacts) is a challenging undertaking. As described in Section 1, there are inherent limits on what can be developed, such as not being able to account for site-specific factors (e.g., mitigation practices such as the use of buffer strips, low-drift nozzles, avoiding headlands, etc.). Due to the significant amount of work undertaken in the past developing environmental indicators for different contexts, there is also scope for different interpretations of what is meant by a 'load indicator'. Thus, it is important to remember that the PLI is a tool that aims to reflect the amounts used of a range of pesticides; their potential to cause harm to end receptors; and their propensity to persist, bioaccumulate and be lost via run-off or leaching (i.e., the PLI does not quantify harm; it is a relative unitless indicator of potential pressure on the environment and does not account for any local factors and actions).

This section reflects on the strengths, weaknesses, and limitations of the PLI, outlines a plan for its future maintenance, and discusses options for future development. In addition, the key contributions of Phase 4 to the ongoing delivery of the PLI are outlined and the implications for how the indicator might be taken forward as a tool for national monitoring and policy setting, including review of the case studies outlined in Sections 3 and 4, are discussed.

5.2. Strengths, weaknesses, and limitations

A key strength of the PLI is bringing together two complex datasets (the PPDB and PUS) into a flexible and dynamic tool that facilitates the visualisation of different outputs for different end user needs. In so doing, it potentially provides a powerful way to support examination of a range of relevant policy questions related to the impacts of pesticide use, as well as ensuring that the user experience and documentation are as transparent as possible.

The examples given in Section 3 demonstrate that the ability to explore change in load at the level of individual metrics, and to associate these changes to individual pesticides, is central to the utility of the PLI for decision making and policy assessment. By being a multi-component indicator with an explicit framework for how different pesticides and metrics can be compared across time and space, the PLI represents a significant advance in terms of the questions that can be explored around pesticide use in the UK, greatly widening the scope of investigation and potential input into policy development.

The limitations of the PLI include:

- Missing data: The PPDB is one of the most comprehensive datasets for pesticide properties currently available. However, while data coverage of the load metrics (across the 299 pesticides in the PLI) is generally more than 80 or 90%, some metrics (e.g., aquatic plants acute, earthworms chronic, parasitic wasp and predatory mites) have lower data coverage (66-70%). Ongoing maintenance of the

PPDB will plug some gaps, but inevitably some gaps in data will remain. An approach to generate replacement data is therefore still needed. The approach in the PLI includes the use of arithmetic mean values by substance type as replacements. This potentially means that for some pesticides, some load metrics could be over or underestimated. The alternative approach would be to use reasonable worst-case values as replacements, thus only overestimating load, but it was considered that the use of arithmetic means was a better approach for observing trends over time. In Phase 4, the role of missing data has been re-examined from the perspective of substance inclusion (Section 5.3.1), which further makes explicit the way in which missing data will be treated.

- Unbounded data: Similarly, the use of unbounded data (data which in the PPDB is preceded by a greater than or less than sign) for ecotoxicity load metrics could also potentially overestimate load. The significance of this is difficult to judge but following discussion and consideration of the resourcing costs of the additional data collection that would be required to resolve this problem, the most parsimonious and transparent approach was to use these values without qualification. It is acknowledged that this creates an indicator that is, in some ways, over conservative (as most unbounded data will be an overestimate of real toxic effect). However, this is the only pragmatic approach that can be applied at the scale of the PLI as currently defined.
- Standardisation: The use of a reference substance to set the maximum or minimum values of the standardisation curve (to convert the raw metric to a 0 to 1 load score) can result in skewed load scores when the pesticide has extreme values. This has been partly addressed for the fate load metrics with the introduction of additional reference points based on regulatory interpretation thresholds (Gustafson 1989; Hollis 1991; Kerle, Jenkins, and Vogue 1996; Rao and Hornsby 2004; USEPA 2011). However, for the ecotoxicity load metrics the issue of skewing remains relevant. For example, imidacloprid has a LD_{50} value of 31 mg kg^{-1} for acute ecotoxicity to birds, but this results in a load score 0.1 due to the reference substance (oxamyl) having an LD_{50} value of 3.16 mg kg^{-1}, yet it is interpreted as being of high toxicity within the PPDB (AERU, 2022). This effect is notably less influential given the adoption of the individual metric focused approach in Phase 4 (which helps to ensure that all metrics are presented on their own intrinsic scale and avoids issues associated with aggregation experienced in the previous version of the PLI). The issues of skewing, and the approach taken to standardisation are dealt with in more detail in Section 2.6.2 in Rainford et al. (2022b). On balance, it is considered that the approach taken, and the different standardisations applied to fate and ecotoxicity metrics respectively, represent a good compromise between competing views of what the indicator needs to present. As a result, it closely aligns to the needs of UK decision makers, who are the key users of the PLI and the associated visualisation tool.
- Estimation of mass applied: For the UK PLI, mass applied is derived from a statistical approximation of use. In Denmark, there is a legal obligation for growers to submit annual records of pesticide applications, meaning that over 80% of usage can be accurately measured based on values from a centralised government repository, supplemented where needed by information on sales (Kudsk *et al.*, 2018). In the UK no such requirement exists, and estimations made from the PUS are based on an entirely voluntary sample which, by necessity, is of finite size when compared to the full population. Hence, where the Danes can simply express their indicator based on the usage data that has been recorded (with relatively little

uncertainty on the values), the UK is reliant on the use of a statistical approximation which carries several assumptions (see Section 2.6.2 in Rainford *et al*., 2022b) and an intrinsic level of uncertainty.

Implicit assumptions: The use of the PLI as a national estimate of the average potential pressure experienced by UK cropping areas rests on a series of implicit assumptions, most of which are inherited from the PUS. These include that:

- The PUS is a representative stratified sample of relevant holdings within the UK (in particular the likelihood that a holding appears in the survey is independent of the quantity of pesticide applied on that holding).
- Records collected on behalf of the PUS are complete with reference to substances included in the PLI used on all included holdings.
- The UK population of total cropping area of holdings (by crop) is known without error and accurately reflected in the values reported by the June survey(s).
- There is at least one record within the associated PUS of all active substances used in the UK within the period and cropping area represented by a given survey.
- All active substances are potentially available to any holding in any region within the UK, (with the specific exception of those substances not estimated due to the absence of associated crops, for example the absence of sugar beet in Scotland).
- The relative cropping area of different crops among sampled holdings in the PUS within a specified stratum is approximately representative of that of the true population of holdings.
- There is full statistical independence between the use of individual active substances with one another on a given holding, and within and between regions.
- The impact of all active substances on load are cumulative and additive. Also covered by this assumption is that the load contributed by a given active substance across multiple real applications is assumed to be proportional to the overall total mass applied per unit area over the entire period (regardless of the number of true applications represented).
- The 'average' load experienced by areas covered by a survey (or selected subset) is proportional to that generated by the 'average' mixture of in scope active substances applied across that area over the entirety of the associated growing season (see Section 2.3; note no real field is likely to experience the exact mixture of substances estimated and applications are unlikely to be spread evenly over the course of the growing season).

Several of these assumptions are challenging to verify or known to be oversimplifications of more complex processes, most notably the assumption that use of specific actives is independent and that the effect of these actives is cumulative and additive within the context of the survey period. However, they reflect simplifications needed to make the estimation of the indicator computationally tractable, as well as to transparently reflect the current state of knowledge regarding PPP usage. The decision to tie estimated load to the average mixture of substances applied is imposed by the structure of the PUS, and the difficulty of using a limited sample of holdings to infer tank mixtures and the number/timing of applications made on the unsampled holdings (i.e., the real localised application rates experienced by the landscape). The consequent reliance on averages is one of the principal reasons why the PLI is most informative when expressed in terms of relative change between a pair of years, as there is a reasonable assumption of comparability in agronomic practice through time (see Section 3.2). While it is conceptually possible to resolve some of these issues by 'scoring' individual applications made on real holdings prior to estimating national trends, to do so undermines the ability of the indicator to clearly visualise the contribution of individual active substances to total load[24], and thus would limit the utility of the PLI as a national policy tool. Again, it must be emphasised that the PLI is intended as a unitless index of potential pressure generated by PPP used in the environment based on applications and should not be interpreted as a measurement of the actual outcomes of PPP use.

5.3. Developments made in Phase 4 (2022-2023)

5.3.1 The development of a substance inclusion protocol

One of the key enhancements made to the PLI in Phase 4 is the development of an explicit protocol for the assessment of which substances are in scope for the indicator. This work was motivated by the desire to expand the scope of the PLI to include biopesticides (that are becoming an increasingly important part of pest and disease control in UK agriculture) but evolved into more general guidelines aimed at a) ensuring transparency and consistency in the way in which the PLI is presented, and b) making it easier to expand the indicator when calculated in a novel context (whether that is forward looking in the incorporation of additional future surveys or backwards looking based on historical data[25]). A strength of the PLI, relative to other frameworks such as the TAT, is its focus on specific identifiable measurements as a basis for metrics. This facilitates like for like comparison of different active substances and a conservative approach to missing data (in terms of not underestimating potential load generated; see Section 2.2). A risk of this implementation is the potential to undermine confidence in the indicator by over-reliance on inferred values for active substances with missing data (see Section 2.3; note in practice the contributions of substances with inferred missing data to any given metric are likely to be very small in all but the most extreme cases). By having an explicit protocol that determines if a given substance is included in the PLI, the aim is to circumvent this concern and so improve user confidence in the trends identified by ensuring consistency in the scope of the indicator across time and space. The revised protocol also has a role in streamlining the process of calculating the indicator, making it less dependent on expert opinion and so easier to maintain going forward.

5.3.2 The removal of the aggregation step from the PLI

One of the most controversial and far-reaching changes made to the PLI during Phase 4 was the transition from presentation of the PLI as a single headline value (calculated by aggregating the values of underlying metrics, as in the Danish implementation; Lewis et al. 2021) to an array of 20 metrics each representing a single identifiable measurement (usually made during lower tier authorisation studies). As outlined in Section 2.6 and Annex 3, aggregation has historically been the most problematic step within the PLI calculation due to the lack of objectivity in how the 'aggregation

[24]As the contribution of active substance would no longer be additive at the national scale (see Section 3.3)

[25] The PLI is potentially compatible with any PUS survey which follows a consistent data structure, of which the earliest was in 1997.

constants' involved are defined. The resulting lack of transparency in the overall calculation has been consistently highlighted as a concern and a potential block to widespread adoption of the PLI as a policy support tool.

The previous case for a headline value rested on the analogy between the UK PLI and its Danish precursor (which relies on a single headline value to support its role in pesticide taxation), as well as a potential role for the headline value in target setting and proving an overall focus to monitoring efforts. It has become increasingly clear from discussion with Defra and others, the potential role of the PLI in UK policy decisions is very distinct from the Danish case. The obscuring effect that aggregation can have on identification of trade-offs from policy intervention was therefore judged as inappropriate for a UK context (Annex 3). As a result, a decision was made to remove the previous aggregation step in favour of a more concise reporting of the underlying data and directing the user towards key information (see example in Section 3.3). The hope is that by doing so the resulting calculation will be more transparent and clearly linked to the underlying data. The previous role of sub-aggregations, such as the overall load on a given taxonomic group has also been removed from the PLI, as it relied on the same subjective determination of aggregation constants as the overall headline value.

Looking ahead, it is expected that the integration of the TAT within the PLI framework (see Section 5.3.4) could potentially solve the issue of multiple metrics for a single taxonomic group (at least for ecotoxicological metrics) by providing a revised structure for understanding how multiple measurements might be combined in processing. In practice, the key change here (in addition to new data resources; see below) is a shift away from weighted summation of underlying metrics (with the associated subjectivity in selection of aggregation constants) towards an approach which takes the regulatory data from multiple sources and selects a 'worst-case' value (typically the minimum concentration associated with any toxic effect) as reflective for the purposes of higher level comparison between substances (i.e., for example, the TAT for an active substance with respect to pollinators might use whichever is the lowest value of the metrics Bees contact LD_{50} and Bees oral LD_{50} metrics plus any other measurements that are judged to be within scope). This has the advantage of being more in line with the underlying logic of the risk assessment process for pesticides, which uses the 'worst-case' value regardless of origin (under the precautionary principle)[26] and helps to resolve some of the perceived weaknesses in the PLI with reference to missing data. How this will work precisely is still to be determined (see Section 5.3.4) but is expected to integrate with existing visualisations, as presented in the developed tool and Section 3.

5.3.3. Development of the visualisation tool to implement the revisions and to refine the tool to better meet the needs of Defra

The fundamentals for the visualisation tool associated with the PLI were created in Phase 3 and have since been refined and adjusted as new user requirements have come to light. The primary function of changes made during Phase 4 was the removal

[26] https://www.hse.gov.uk/pesticides/pesticides-registration/index.htm

of the aggregation step (see above) and the consequent need for a reorganisation of the tool to highlight relative change in individual metrics and support threshold setting (see Section 5.3.2). New visuals were developed (as showcased in Section 3) and some existing panels were restructured to fit with the revised paradigm. Alongside the increased focus on target thresholds and the removal of visualisations that directly related to aggregation, the major new feature added in Phase 4 is the panel entitled '**View importance of actives',** which is intended as a tool to help rank different active substances based on their percentage contribution (including rates of application) across multiple metrics, and so provide an overview and prioritisation of which active substances might be the target of policy interventions.

The PLI is intended as a tool to help inform the user about the potential impacts of pesticide applications made in a particular context and time. It is not directly a decision support platform for policy intervention. In particular, one of the major gaps in the tool is knowledge of which active substances are potential alternatives for the same pest issue (based on their efficacy, availability etc), which can be an important component of understanding the potential side effects of product withdrawal. This was briefly investigated in Phase 4 but ultimately failed to proceed due to the absence of officially recognised sources from which such information could be routinely collected. The related activity of developing counterfactuals or alternative scenarios for usage (to judge the impact of e.g., withdrawal of authorisation from an active substance on the trend in the indicator) was likewise deemed over-reliant on assumptions and expert judgement which could not be easily validated in the majority of cases. While this remains an area of interest for the development of PLI, discussion with relevant policy teams has so far failed to identify a way in which this could practically be implemented and, as a result, it was deprioritised as a part of Phase 4.

Alongside the structural changes to the visualisation tool there have also been several cosmetic changes made to various visualisations to improve readability and signposting to key datasets. The addition of a data download option for authorised users also allows the tool to better integrate into existing reporting around PPP usage further cementing the role of the PLI as a tool for ongoing reporting (see Section 5.4).

The backend for the visualisation tool has been streamlined to work with the revised protocol for substance inclusion and appropriate labelling of substances.This is intended to reduce dependence on experts when calculating the PLI and to improve consistency in its ongoing delivery. One of the priorities for Phase 4 was to ensure that the protocol for calculating the PLI was properly documented to help support the transition from a research tool to deployed system. From 2023 onwards the PLI is anticipated to be delivered on an ongoing basis to internal Defra stakeholders and improved documentation (for both the tool and the associated process) is seen as critical in supporting this transition and for providing quality assurance. The delivery team will be working closely with Defra to ensure that new information is integrated correctly into the tool and backend data files and will review, on an ongoing basis, any perceived issues in the revised data delivery protocol.

The development of the visualization tool for the PLI is an ongoing process that will adapt to changing policy and stakeholder needs. While the current version of the tool is designed for an internal Defra policy audience, some of the features may (over the longer term) be made available to other key stakeholders in the pesticide policy space. The PLI is anticipated as becoming increasingly integrated into the national reporting framework of indicators relating to PPP usage in the UK and will hopefully become an important source of information for the changes experienced by the UK landscape with reference to this key area of policy concern (see Section 5.4).

5.3.4 Comparison of the PLI to Total Applied Toxicity

The TAT (Schulz et al. 2021) is a relatively new indicator that has recently gained prominence as a potential international standard for how the relative potential impacts of PPP are explored between countries (for example it is cited as such in recent discussion around the Convention on Biological Diversity; see Open-ended working group on the post-2020 global biodiversity framework 2022). As outlined in Section 4, the TAT is conceptually similar to the PLI with significant overlap in the infrastructure required to calculate both values. It therefore makes sense to integrate these into a common pipeline (for example, both are dependent on having national and regional estimates of the total mass of different active substances applied). As also discussed in Section 4, the two indicators differ in their approach to combining different sources of regulatory data and the implications this has for their calculation. It is likely that any UK implementation of the TAT will most closely resemble what has been termed $TAT_{Germany}$, as the UK regulatory and risk assessment regime remains (at present) closely aligned to the process used by EU member states (although this may change depending on the trajectory of international discussions around baselining and data usage). Both existing TAT approaches take a different view to missing data to that implemented in the PLI (Section 4.3) and this may need to be harmonised if the two indicators are to be presented together. As noted above (Section 4.5), the calculation of the TAT is considered feasible in a UK context, but actual implementation lay outside the scope of resources available in Phase 4. Integration of the TAT or a similar indicator alongside the visualizations developed for the PLI is a potential option for future development.

Alongside integration of the TAT (possibly over a slightly longer term), there is growing interest in the alignment of the PLI to the results of routine PPP monitoring of surface water conducted by the Environment Agency. Historically, the structure of the PUS, as randomly sampled representative holdings, has inhibited widespread discussion of the relationship between what is applied to crops and what is detected in watercourses in a UK context. However, assuming suitable datasets can be identified, the PLI may have a role to play in the exploration of these relationships, which would also provide a useful benchmark of the relevance of various PLI metrics to environmental outcomes. This work remains at early stage of discussion but is one of the more promising avenues for future development of indicators in this policy area and is likely to be explored in more detail in a subsequent phase of work.

5.4. Conclusions

The goal of developing the PLI was to provide an exploratory tool that facilitates access to improved information about the potential environmental impacts associated with pesticide use and to provide a tool for exploring relative tends associated with the changing mixture of active substances applied. When compared to previous monitoring efforts, which have largely been dependent on the total mass of PPP applied, the PLI has been highly successful in adding a greater resolution and 'colour' to the discussion around the potential effects of PPP usage and the potential impacts of policy intervention (Section 3). As the UK continues to develop its pesticide policy post-EU exit, the role and relevance of tools like PLI will increase in importance as indicators of progress. A key objective of Phase 4 was to further enhance this role by:

a) Improving the transparency with which the indicator is calculated both in terms of the scope of substances considered and documenting the process by which the indicator is calculated on an ongoing basis.
b) Improving the linkage between the indicator and the underlying datasets by removing the previously problematic aggregation step in the calculation.
c) Improving the utility of the indicator as a communication tool via redesign of the visualisation tool to be more aligned with the needs of Defra policy.
d) Considering how the PLI might be used alongside other related indicators such as the TAT, and the practical issues this might present in a UK context.

All these elements are intended to work towards the wider goal of providing decision makers in the UK with access to transparent and high-quality information around pesticide load that can identify where progress has been made and where more targeted policy intervention may be required. The revised visualisation tool, with its emphasis on relative trends for different metrics through time is intended, first and foremost, as a platform for internal Defra users to investigate the impact of historic change in policy to help improve future decision making (see examples in Section 3). The reliance of the PLI on the PUS and time lags involved in data collection mean that the tool will never (at least in current form) be a real time monitoring system for UK agriculture. Its primary function will, therefore, always be mainly in linking post hoc analysis to inform future decision making.

As primarily a monitoring tool it is critical that users trust in the consistency of the calculation and underlying data. This is the main motivation behind the documentation of the explicit protocol for how substances are included (Section 2.3) and the steps involved in calculating the indicator for a new set of survey data (Rainford et al 2022b). This additional documentation, which emphasises the link between load values and the underlying data, aims to support users in communicating the indicator and providing confidence in the presentation of documented trends. The structure of the PLI makes it relatively straight forward for a user to drill down into effects at an increasingly fine scale, including the impact of an individual active substance which may be the target of future policy intervention (see Section 3.3). The goal in developing the visualisation tool is to guide the user through various ways of viewing the data that allow increasing focus on those trends that are most important (Section 3.2).

Phase 4 is intended to mark the transition between the PLI as a research tool, that is periodically redefined based on the shifting needs of UK policy, to a stable deployed system that will be made accessible to a wider array of internal Defra policy actors. While work on developing the PLI will continue, for example via the integration of the TAT or similar metrics, it is expected that the core system and visualisation tool will continue to be supported on an ongoing basis via the inclusion of further PUS surveys.

With the upcoming publication of the UK's revised National Action Plan for pesticides this increased role as a potential national indicator becomes timely (Defra 2020), although the details of how the PLI will integrate into the revised NAP have not yet been fully agreed at the time of writing (March 2023). The PLI has, through its multiple phases of development, undergone a substantial realignment from an initial reimplementation of the Danish system (Lewis et al., 2021) to something that is much more targeted and aligned to the needs of UK policy (see above and Rainford et al. 2022b). Through this redevelopment it is hoped that users have come to understand the role and limitations of the indicator, so that it can be used as an effective tool in future pesticide policy development and a significant advance for exploring trends in environmental pressures associated with pesticide use in the UK at a national and regional scale.

The need for post authorisation monitoring of PPP usage in the UK is an area of increasing concern for policy, for example Milner and Boyd (2017) and Walker et al. (2021), and part of this process is to get a better grasp on what PPP are applied and the implications for natural systems. The PLI thus fills an important gap in the current UK reporting system with respect to pesticide usage and it is likely to be a useful component of decision making going forwards. While by no means a perfect indicator, in particular due to limitations arising from the lack of universal and spatially explicit means of reporting applications made in the UK, the PLI is nonetheless a substantial improvement over what has come before and is now well suited to the monitoring needs of a UK policy audience. While work on the indicator will continue, particularly with the integration of the TAT, the core processes of the PLI are now established to the point where the indicator is ready for routine operational deployment as a part of a wider suite of indicators that reflect different elements of the socio-economic context and decision-making processes around PPP application.

References

AHDB. 2020. “Analyst Insight: BPS or June Survey – What Should I Believe? Market Commentary.” https://ahdb.org.uk/news/analyst-insight-bps-or-june-survey-what-should-i-believe.

Ambrose, Margery Lee. 2003. “Characterization of the Insecticidal Properties of Acetamiprid under Field and Laboratory Conditions.” Masters of Science, North Carolina State University. https://repository.lib.ncsu.edu/bitstream/handle/1840.16/624/etd.pdf?sequence=1.

Bourguet, Denis, and Thomas Guillemaud. 2016. “The Hidden and External Costs of Pesticide Use.” In *Sustainable Agriculture Reviews: Volume 19*, edited by Eric Lichtfouse, 35–120. Sustainable Agriculture Reviews. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-26777-7_2.

Bub, Sascha, Jakob Wolfram, Lara L. Petschick, Sebastian Stehle, and Ralf Schulz. 2023. “Trends of Total Applied Pesticide Toxicity in German Agriculture.” *Environmental Science & Technology* 57 (1): 852–61. https://doi.org/10.1021/acs.est.2c07251.

Budge, G. E., D. Garthwaite, A. Crowe, N. D. Boatman, K. S. Delaplane, M. A. Brown, H. H. Thygesen, and S. Pietravalle. 2015. “Evidence for Pollinator Cost and Farming Benefits of Neonicotinoid Seed Coatings on Oilseed Rape.” *Scientific Reports* 5 (1): 12574. https://doi.org/10.1038/srep12574.

Burnett, F., T. Davey, A.C. Newton, T.D. Heilbronn, N. Havis, K Davie, H. Coutts, and A. Lees. 2020. “Proceedings Crop Production in Northern Britain 2020.” Dundee: Bayer. https://www.cpnb.org/sites/www.cpnb.org/files/inline-files/Index_and_Proceedings_2020.pdf.

Camp, Allison A., Michael A. Batres, Wanda C. Williams, Robert W. Koethe, Kimberly A. Stoner, and David M. Lehmann. 2020. “Effects of the Neonicotinoid Acetamiprid in Pollen on Bombus Impatiens Microcolony Development.” *Environmental Toxicology and Chemistry* 39 (12): 2560–69. https://doi.org/10.1002/etc.4886.

Cooper, Jerry, and Hans Dobson. 2007. “The Benefits of Pesticides to Mankind and the Environment.” *Crop Protection* 26 (9): 1337–48. https://doi.org/10.1016/j.cropro.2007.03.022.

Coston, Duncan, Simon Potts, Tom Breexe, Linda Field, and Sam Cook. 2016. *Implications of the Neonicotinoid Restriction on Oilseed Rape Pest Control, Pollination and Productivity.*

Defra. 2020. “Revised National Action Plan for the Sustainable Use of Pesticides (Plant Protection Products).” https://consult.defra.gov.uk/pesticides-future-strategy/sustainable-use-of-pesticides-national-action-plan/supporting_documents/NAPConsultationDocument.pdf.

Defra. 2022. “Structure of the Agricultural Industry in England and the UK at June. Department for Environment, Food & Rural Affairs (Defra).” https://www.gov.uk/government/statistical-data-sets/structure-of-the-agricultural-industry-in-england-and-the-uk-at-june.

Defra, and RPA. 2018. “Crop Areas in England Sourced from the Basic Payment Scheme and the June Survey of Agriculture 2015-2017 (Experimental Statistics).” Department for Environment, Food & Rural Affairs and the Rural Payments Agency (RPA). https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/703934/Crop_areas_England_from_BPS_and_June_Survey2015-2017_release_01may2018b.pdf.

Deguine, Jean-Philippe, Jean-Noël Aubertot, Rica Joy Flor, Françoise Lescourret, Kris A.G. Wyckhuys, and Alain Ratnadass. 2021. “Integrated Pest Management: Good Intentions, Hard Realities. A Review.” *Agronomy for Sustainable Development* 41 (3): 38. https://doi.org/10.1007/s13593-021-00689-w.

Denholm, Ian. 2011. “Controlling Pollen Beetle and Combating Insecticide Resistance in Oilseed Rape.” Information Sheet 12. HGCA. https://irac-online.org/documents/hgca-uk-information-sheet/.

Dewar, Alan M. 2017. “The Adverse Impact of the Neonicotinoid Seed Treatment Ban on Crop Protection in Oilseed Rape in the United Kingdom.” *Pest Management Science* 73 (7): 1305–9. https://doi.org/10.1002/ps.4511.

Efron, Bradley, and LePage Raoul. 1992. *An Introduction to the Bootstrap*. 1st ed. New York: Wiley & Sons. http://www.ru.ac.bd/stat/wp-content/uploads/sites/25/2019/03/501_02_Efron_Introduction-to-the-Bootstrap.pdf.

EFSA Panel on Plant Protection Products and their Residues (PPR). 2015. “Scientific Opinion Addressing the State of the Science on Risk Assessment of Plant Protection Products for Non-Target Arthropods.” *EFSA Journal* 13 (2): 3996. https://doi.org/10.2903/j.efsa.2015.3996.

Environment Agency. 2019. “Cypermethrin: Sources, Pathways and Environmental Data.” https://consult.environment-agency.gov.uk/++preview++/environment-and-business/challenges-and-choices/user_uploads/cypermethrin-pressure-rbmp-2021.pdf.

Farmer, Deborah, Ian R. Hill, and Stephen J. Maund. 1995. “A Comparison of the Fate and Effects of Two Pyrethroid Insecticides (Lambda-Cyhalothrin and Cypermethrin) in Pond Mesocosms.” *Ecotoxicology* 4 (4): 219–44. https://doi.org/10.1007/BF00116342.

Finger, Robert, Niklas Möhring, Tobias Dalhaus, and Thomas Böcker. 2017. “Revisiting Pesticide Taxation Schemes.” *Ecological Economics* 134 (April): 263–66. https://doi.org/10.1016/j.ecolecon.2016.12.001.

Franklin, Elizabeth L., and Nigel E. Raine. 2019. “Moving beyond Honeybee-Centric Pesticide Risk Assessments to Protect All Pollinators.” *Nature Ecology & Evolution* 3 (10): 1373–75. https://doi.org/10.1038/s41559-019-0987-y.

Garthwaite, David G. 2015. “The potential for using pesticide usage surveys to monitor plant pests.” *EPPO Bulletin* 45 (2): 294–98. https://doi.org/10.1111/epp.12210.

Garthwaite, David G., Lee Ridley, Abigale Mace, G. Parrish, Ian Barker, James L. Rainford, and Roy MacArthur. 2019. “Pesticide Usage Survey Report 284 Arable Crops in the United Kingdom 2018.” Pesticide Usage Survey. FERA Science Limited on behalf of DEFRA. https://secure.fera.defra.gov.uk/pusstats/surveys/documents/arable2018.pdf.

Goulson, Dave. 2013. “REVIEW: An Overview of the Environmental Risks Posed by Neonicotinoid Insecticides.” *Journal of Applied Ecology* 50 (4): 977–87. https://doi.org/10.1111/1365-2664.12111.

Goulson, Dave, Jack Thompson, and Amy Croombs. 2018. “Rapid Rise in Toxic Load for Bees Revealed by Analysis of Pesticide Use in Great Britain.” *PeerJ* 6 (July): e5255. https://doi.org/10.7717/peerj.5255.

Gustafson, David. 1989. “Groundwater Ubiquity Score: A Simple Method for Assessing Pesticide Leachability.” *Environmental Toxicology and Chemistry* 8 (April): 339–57. https://doi.org/10.1002/etc.5620080411.

Hernandez Jerez, Antonio, Paulien Adriaanse, Philippe Berny, Tamara Coja, Sabine Duquesne, Andreas Focks, Marina Marinovich, et al. 2022. “Statement on the Active Substance Acetamiprid.” *EFSA Journal* 20 (1): e07031. https://doi.org/10.2903/j.efsa.2022.7031.

HM Government. 2018. “A Green Future: Our 25 Year Plan to Improve the Environment.” https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/693158/25-year-environment-plan.pdf.

Hollis, J. M. 1991. “Mapping the Vulnerability of Aquifers and Surface Waters to Pesticide Contamination at the National/Regional Scale.” In *Pesticides in Soils and Water, BCPC Monograph*, edited by A. Walker, 47:165–74. Cranfield University, Silsoe: National Soil Resources Institute.

Insecticide Resistance Action Group. 2022. “Insecticide Resistance Status in UK Oilseed Rape Crops.” IRAG. https://projectblue.blob.core.windows.net/media/Default/Imported%20Publication%20Docs/AHDB%20Cereals%20&%20Oilseeds/Pests/IRAG/IRAG3%20Insecticide%20resistance%20status%20in%20UK%20oilseed%20rape%20crops%20(2022).pdf.

Kathage, Jonas, Pedro Castañera, José Luis Alonso-Prados, Manuel Gómez-Barbero, and Emilio Rodríguez-Cerezo. 2018. “The Impact of Restrictions on Neonicotinoid and Fipronil Insecticides on Pest Management in Maize, Oilseed Rape and Sunflower in Eight European Union Regions.” *Pest Management Science* 74 (1): 88–99. https://doi.org/10.1002/ps.4715.

Kerle, E.A., J.J. Jenkins, and P.A. Vogue. 1996. “Understanding Pesticide Persistence and Mobility for Groundwater and Surface Water Protection.” Oregon State University. EM 8561.

Kudsk, Per, Lise Nistrup Jørgensen, and Jens Erik Ørum. 2018. “Pesticide Load—A New Danish Pesticide Risk Indicator with Multiple Applications.” *Land Use Policy* 70 (January): 384–93. https://doi.org/10.1016/j.landusepol.2017.11.010.

Lee, Rhiannon, Roos den Uyl, and Hens Runhaar. 2019. “Assessment of Policy Instruments for Pesticide Use Reduction in Europe; Learning from a Systematic Literature Review.” *Crop Protection* 126 (December): 104929. https://doi.org/10.1016/j.cropro.2019.104929.

Lewis, Kathleen A., and K.S. Bardon. 1998. “A Computer-Based Informal Environmental Management System for Agriculture.” *Environmental Modelling & Software* 13 (2): 123–37. https://doi.org/10.1016/S1364-8152(98)00010-3.

Lewis, Kathleen A., James L. Rainford, John Tzilivakis, and David G. Garthwaite. 2021. “Application of the Danish Pesticide Load Indicator to UK Arable Agriculture.” *Journal of Environmental Quality*. https://doi.org/10.1002/jeq2.20262.

Lewis, Kathleen A., and John Tzilivakis. 2019. “Wild Bee Toxicity Data for Pesticide Risk Assessments.” *Data* 4 (3): 98. https://doi.org/10.3390/data4030098.

Lewis, Kathleen A., John Tzilivakis, Douglas J. Warner, and Andrew Green. 2016. “An International Database for Pesticide Risk Assessments and Management.” *Human and Ecological Risk Assessment: An International Journal* 22 (4): 1050–64. https://doi.org/10.1080/10807039.2015.1133242.

Lundin, Ola, Gerard Malsher, Carol Högfeldt, and Riccardo Bommarco. 2020. “Pest Management and Yield in Spring Oilseed Rape without Neonicotinoid Seed Treatments.” *Crop Protection* 137 (November): 105261. https://doi.org/10.1016/j.cropro.2020.105261.

Miljøstyrelsen. 2012. “The Agricultural Pesticide Load in Denmark 2007–2010. Environmental Review 2, 2012.” København. https://www2.mst.dk/Udgiv/publikationer/2012/03/978-87-92779-96-0.pdf.

Milner, Alice M., and Ian L. Boyd. 2017. “Toward Pesticidovigilance.” *Science* 357 (6357): 1232–34. https://doi.org/10.1126/science.aan2683.

Möhring, Niklas, Martina Bozzola, Stefan Hirsch, and Robert Finger. 2020. “Are Pesticides Risk Decreasing? The Relevance of Pesticide Indicator Choice in Empirical Analysis.” *Agricultural Economics* 51 (3): 429–44. https://doi.org/10.1111/agec.12563.

Möhring, Niklas, Sabrina Gaba, and Robert Finger. 2019. “Quantity Based Indicators Fail to Identify Extreme Pesticide Risks.” *Science of The Total Environment* 646 (January): 503–23. https://doi.org/10.1016/j.scitotenv.2018.07.287.

Möhring, Niklas, David Wuepper, Tomke Musa, and Robert Finger. 2019. “Why Farmers Deviate from Recommended Pesticide Timing: The Role of Uncertainty and Information.” *Pest Management Science* n/a (n/a). https://doi.org/10.1002/ps.5826.

Oerke, E.-C. 2006. “Crop Losses to Pests.” *The Journal of Agricultural Science* 144 (1): 31–43. https://doi.org/10.1017/S0021859605005708.

Open-ended working group on the post-2020 global biodiversity framework. 2022. “Science Briefs on Targets, Goals and Monitoring in Support of the Post-2020 Global Biodiversity Framework Negotiations.” Convention on Biological Diversity; https://www.cbd.int/doc/c/6053/38a4/3710d6e83f5b006ef774607d/wg2020-04-inf-02-rev-01-en.pdf.

Pendersen, Anders, Nielson Helle, and Mikael Andersen. 2015. “The Danish Pesticide Tax.” In *Use of Economic Instruments in Water Policy*, 14:73–87. Global Issues in Water Policy. Springer, Cham. https://doi.org/10.1007/978-3-319-18287-2_6.

Rainford, James L., Marc Kennedy, and Glyn Jones. 2021. “Pesticide Risk Indices: A Review of Practical Implications for Policy and Interpretation.” Critical Review. FERA Science Limited on behalf of the [UK] National Farmers Union. https://www.nfuonline.com/archive?treeid=151098.

Rainford, James L., John Tzilivakis, David G. Garthwaite, Glyn Jones, and Kathleen A. Lewis. 2022a. “A Pesticide Load Indicator for the UK: Final Report. Prepared for the Department for Environment, Food and Rural Affairs (Defra), UK, for Project PC0107, by FERA Science Ltd and the Agriculture and Environment Research Unit (AERU), University of Hertfordshire, UK.”

Rainford, James L., John Tzilivakis, David G. Garthwaite, Glyn Jones, and Kathleen A. Lewis 2022b. “Finalising a Pesticide Load Indicator for the UK: Technical Report. Prepared for the Department for Environment, Food and Rural Affairs (Defra), UK, for Project PC0107, by FERA Science Ltd and the Agriculture and Environment Research Unit (AERU), University of Hertfordshire, UK.”

Rao, P.S.C., and A.G Hornsby. 2004. *Behaviour of Pesticides in Soils and Water.* University of Florida.

Salako, Abdulbashir Femi, Nnamdi Henry Amaeze, Hannah Motunrayo Shobajo, and Fidelia Ijeoma Osuala. 2020. “Comparative Acute Toxicity of Three Pyrethroids (Deltamethrin, Cypermethrin and Lambda-Cyhalothrin) on Guppy Fish (Poecilia Reticulata Peters, 1859).” *Scientific African* 9 (September): e00504. https://doi.org/10.1016/j.sciaf.2020.e00504.

Sánchez-Bayo, Francisco, and Kris A. G. Wyckhuys. 2019. “Worldwide Decline of the Entomofauna: A Review of Its Drivers.” *Biological Conservation* 232 (April): 8–27. https://doi.org/10.1016/j.biocon.2019.01.020.

Schäfer, Ralf Bernhard, Matthias Liess, Rolf Altenburger, Juliane Filser, Henner Hollert, Martina Roß-Nickoll, Andreas Schäffer, and Martin Scheringer. 2019. “Future Pesticide Risk Assessment: Narrowing the Gap between Intention and Reality.” *Environmental Sciences Europe* 31 (1): 21. https://doi.org/10.1186/s12302-019-0203-3.

Schulz, Ralf, Sascha Bub, Lara L. Petschick, Sebastian Stehle, and Jakob Wolfram. 2021. “Applied Pesticide Toxicity Shifts toward Plants and Invertebrates, Even in GM Crops.” *Science* 372 (6537): 81–84. https://doi.org/10.1126/science.abe1148.

Scott, Charles, and Paul E. Bilsborrow. 2019. “The Impact of the EU Neonicotinoid Seed-Dressing Ban on Oilseed Rape Production in England.” *Pest Management Science* 75 (1): 125–33. https://doi.org/10.1002/ps.5189.

Siviter, Harry, and Felicity Muth. 2020. “Do Novel Insecticides Pose a Threat to Beneficial Insects?” *Proceedings of the Royal Society B: Biological Sciences* 287 (1935): 20201265. https://doi.org/10.1098/rspb.2020.1265.

Smitley, David, Diane Brown, Rebecca Finneran, and Erwin Elsner. 2019. “How to Control Invasive Pests While Protecting Pollinators and Other Beneficial Insects.” Michigan State University. https://www.canr.msu.edu/news/how-to-control-invasive-pests-while-protecting-pollinators-and-other-beneficial-insects.

Sud, Megha. 2020. “Managing the Biodiversity Impacts of Fertiliser and Pesticide Use: Overview and Insights from Trends and Policies across Selected OECD Countries.” 1555. OECD Environment Working Papers. Paris, France. http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV/WKP(2020)2&docLanguage=En.

Thomas, Miles R. 1999. “Guidelines for the Collection of Pesticide Usage Statistics within Agriculture and Horticulture.” OECD, Eurostat. https://secure.fera.defra.gov.uk/pusstats/surveys/documents/guide.pdf.

US EPA. 2020. “Acetamiprid Proposed Interim Registration Review Decision Case Number 7617.” Docket Number EPA-HQ-OPP-2012-0329. US EPA. https://www.epa.gov/sites/default/files/2020-01/documents/acetamiprid_pid.pdf.

USEPA. 2011. “Design for the Environment Program Alternative Assessment Criteria for Hazard Evaluation, v 2.0.” Washington DC, USA: Office of Pollution Prevention &

Toxics, United States Environmental Protection Agency (USEPA). https://www.epa.gov/saferchoice/alternatives-assessment-criteria-hazard-evaluation.

Varga-Szilay, Zsófia, and Zoltán Tóth. 2022. "Is Acetamiprid Really Not That Harmful to Bumblebees (Apidae: Bombus Spp.)?" *Apidologie* 53 (1): 2. https://doi.org/10.1007/s13592-022-00909-6.

Walker, Lee A., Richard F. Shore, Jacquline S Chaplow, Elizabeth A. Barnett, and David J. Spurgeon. 2021. "A Proposal for Terrestrial Environmental Monitoring of Plant Protection Products." Report by CEH to Natural England.

White, Sacha, Steve Ellis, Fran Pickering, Daniel Leybourne, Isabel Corkley, Sarah Kendall, Lariss Collins, Max Newbert, Luke Cotton, and Richard Phillips. 2020. "Project Report No. 623 Integrated Pest Management of Cabbage Stem Flea Beetle in Oilseed Rape." 623. AHDB Cereals & Oilseeds. https://farmpep.net/sites/default/files/2022-05/PR623%20Final%20Project%20Report%20CSFB.pdf.

Woodcock, Ben A., Nicholas J. B. Isaac, James M. Bullock, David B. Roy, David G. Garthwaite, Andrew Crowe, and Richard F. Pywell. 2016. "Impacts of Neonicotinoid Use on Long-Term Population Changes in Wild Bees in England." *Nature Communications* 7 (1): 12459. https://doi.org/10.1038/ncomms12459.

Zhu, Yu Cheng, John Adamczyk, Thomas Rinderer, Jianxiu Yao, Robert Danka, Randall Luttrell, and Jeff Gore. 2015. "Spray Toxicity and Risk Potential of 42 Commonly Used Formulations of Row Crop Pesticides to Adult Honey Bees (Hymenoptera: Apidae)." *Journal of Economic Entomology* 108 (6): 2640–47. https://doi.org/10.1093/jee/tov269.

Annex 1: Protocol for addressing missing or unbounded data

Step	Description
1	Load the raw data from the PPDB. Where data are missing for a parameter for a substance, if that substance has related compounds, these are used to populate the data for that parameter (if it exists). Some pesticides in the PUS contain a mix of different compounds and/or may be a variant (such as salt), but this is not known due to commercial confidentiality. In the absence of the specific compound being listed in the PUS the following approach has been developed. Where data for the PUS substance exists in the PPDB, these data are used. If data for that substance are not available, data for related compounds are used (if they exist) (e.g., if data do not exist for a parameter for chlormequat but they do exist for chlormequat chloride the latter is used, and vice versa). For glyphosate (specifically), the data used are an average for the 8 compounds (incl. salts) of glyphosate that exist.
2	Determine which DT_{50} and K_{oc} (or K_{foc}) values to use for the groundwater mobility metric: • If DT_{50} field is available use this, else use DT_{50} lab if available. • If K_{foc} is available, use this, else use K_{oc} if available.
3	Calculate GUS where DT_{50} and K_{oc} (or K_{foc}) are available from the process above.
4	For the BCF, where this is not available within the PPDB it is calculated using Log P_{ow} (if available) (see Equations 2.3 and 2.4 above).
5	If the value is unbounded, use the value stated for the substance ignoring the qualifier (following a precautionary principal approach).
6	Where data are missing (after the steps above), these are determined by calculating the arithmetic mean value across the data that does exist for each parameter for all substances, herbicides, insecticides, and fungicides (a mean value cannot be calculated for all the PUS substance types due to the limited number of substances in some groups). The respective arithmetic mean values are used to plug gaps in the following: • DT_{50} field: to plug gaps when field and lab values are missing, for the persistence metric. • BCF: to plug gaps when BCF and logP are missing. • GUS: to plug gaps when DT_{50} and/or K_{oc} (or K_{foc}) are not available to calculate GUS. • Kfoc: to plug gaps when K_{oc} and Kf_{oc} are missing, for the surface water mobility metric. • All the ecotoxicity metrics. • For metabolites, for all the above, when the parent substance has worse values than the arithmetic mean (for the parent substance type) then the parent value is used.
7	Insert the arithmetic mean values for missing data for each type of substance in the PUS following Table 2.7.

Annex 2: Substances excluded from the scope of the PLI by the revised data inclusion protocol (March 2023)

Substance name	PUS Type	Substance Type	Exclude
(E,E)-8,10-dodecadien-1-ol/n-tetradecyl acetate/Z-11 tetradecenyl acetate	Attractant	Single chemical	Y
Acequinocyl	Insecticide	Single chemical	Y
Ampelomyces quisqualis strain AQ 10	Fungicide	Micro-organism	Y
Aureobasidium pullulans	Fungicide	Micro-organism	Y
Bacillus amyloliquefaciens strain MBI 600	Fungicide	Micro-organism	Y
Bacillus amyloliquefaciens subsp. plantarum strain D747	Fungicide	Micro-organism	Y
Bacillus subtilis	Fungicide	Micro-organism	Y
Beauveria bassiana ATCC-74040	Insecticide	Micro-organism	Y
Beauveria bassiana GHA	Insecticide	Micro-organism	Y
Cerevisane (saccharomyces cerevisiae strain LAS 117)	Fungicide	Single chemical	Y
Chitosan hydrochloride	Fungicide	Single chemical	Y
Coniothyrium minitans	Fungicide	Micro-organism	Y
COS-OGA	Fungicide	Single chemical	Y
Dodecylphenol ethoxylate	Insecticide	Single chemical	Y
Fatty acids	Insecticide	Mixture: no dominant substance	Y
Fatty acids C7-C20	Insecticide	Mixture: no dominant substance	Y
Fenhexamid	Fungicide	Single chemical	Y
Fenoxaprop-P-ethyl	Herbicide	Single chemical	Y
Gibberellic acid	Growth regulator	Single chemical	Y
Gibberellins	Growth regulator	Single chemical	Y
Gliocladium catenulatum strain J1446	Fungicide	Micro-organism	Y
Kresoxim-methyl	Fungicide	Single chemical	Y
Lecanicillium muscarium strain Ve6	Insecticide	Micro-organism	Y
Metarhizium anisopliae	Fungicide	Micro-organism	Y
Peroxyacetic acid	Disinfectant	Inorganic compound	Y
Pinoxaden	Herbicide	Single chemical	Y
Potassium phosphonate (phosphite)	Fungicide	Inorganic compound	Y
Sodium chloride	Herbicide	Inorganic compound	Y
Spirotetramat	Insecticide	Single chemical	Y
Sugar	Attractant	Single chemical	Y
Trichoderma asperellum strain T34	Fungicide	Micro-organism	Y
Trichoderma harzianum	Fungicide	Micro-organism	Y
Urea	Fungicide	Inorganic compound	Y

Annex 3: The challenges of aggregation

Aggregation of environmental metrics can be problematic, as doing so inherently tends to result in a loss of detail and has implications for interpretation and communication. Both the Danish precursor of the PLI and previous iterations of the UK indicator to date (see Rainford et al., 2022) have included a headline aggregated value which was intended to reflect the overall trend in load across the various subcomponents and metrics which make up the indicator. The structure of this aggregation and how to best reflect the needs and requirements of decision makers has been an ongoing source on contention throughout the development of the PLI.

The central challenge for aggregation is that the fundamental measurements which underly different metrics are not mutually consistent (see Tables 2.1 and 2.2) and thus any attempt to combine them will always involve some degree of interpretation and subjectivity. The previous conceptions of the PLI suffer particularly acutely from this issue as they were conceived as 'additive' in how they combined metrics, i.e. the value of the headline aggregated value presented for any given active substance is expressed as the (weighted) sum of the values calculated for each of the contributing metrics. Such a structure is very appealing for an indicator like the PLI as it naturally reflects the potential independence of different components of load (e.g. load generated by fate metrics can vary independent from load generated by ecotoxicity), however in the context of pesticides (which by their very nature tend to be targeted at a specific biological group) it also has the potential create problems of scaling and differentiating between substances. For example, it might be the case that substance A might be highly toxic to birds but low toxicity to bees, while substance B may be the opposite, and yet under an additive framework both might achieve a similar intermediate overall score. This issue is further complicated by the fact that the absolute quantities of different active substances used in UK vary across five orders of magnitude and that many of the substances of greatest policy concern are used in small absolute quantities. The issue of how to combine different load metrics thus has major consequences for how the indicator is to be interpreted and used in policy and requires careful consideration and engagement with stakeholders and decision makers.

Within the additive framework the key decisions in calculating the final, value relate to the potential use of 'aggregation constants'. An aggregation constant is defined as a fixed value associated with a given metric which is used to scale the standardised values for that metric when calculating higher level headline values. There is no truly 'objective' way in which aggregation constants can be defined from a scientific perspective, as they inherently carry subjective value judgements about the relative 'importance' of different load metrics and the implicit goal of presenting an aggregated value (which will tend to vary based on the purpose for which the indicator is being calculated). One of the consistent challenges with the PLI has been to identify appropriate aggregation constants that are acceptable to a wide array of stakeholders within government with different views on how the indicator might be used in practice. What this has mean in practice has been that throughout Phases 1-3 the definitions and basis of such constants has changed repeatedly resulting in confusion around how the indicator is to be used. To summarise:

- The earliest implementations of the PLI to the UK (Lewis et al 2020) used aggregation constants that were identical to those used by the Danish implementation. This had the advantage of directly tying the UK PLI to its Danish equivalent allowing for a like for like comparison of trends in the overall headline value. However, in the documentation provided by the Danish implementation there is no explicit basis for how these values were derived or guidance on how they might be modified if additional metrics were included. As the scope of the UK version of the PLI moved away from its Danish predecessor it was necessary to abandon direct alignment in favour a UK specific alternative.
- The second major phase of development defined aggregation constants as the maximum values that can be obtained on any given metric (or equivalent the value that is assigned to a 1kg application of the reference substance for each metric). This approach adapted the same approach to calculation used in the Danish implementation but tailored it to a UK focus. Originally these values were to be set by consensus of the stakeholders (see Rainford et al 2021) however in practice a 'policy neutral' compromise was established which formed the basis of values used during Phases 2 and 3 and which are referenced as the 'fixed' approach in Rainford et al. (2022). The key drawbacks to this approach are related to the fact that by setting the aggregation constant as the maximum value for a metric and given that the majority of metrics have highly skewed distributions active substances, the average value of a 'typical' substance varied widely between different metrics. This means that when they are summed the resulting headline indicator tended to be dominated by a small fraction of specific metrics which had less skewed distribution (particularly those describing Fate, see Rainford et al., 2022). Hence the headline indicator thus defined which was insensitive to changes in relative toxicity (which show large variation between difference substances) but strongly tied to fate metrics and the relative mass of active substance applied. In turn this resulted in a headline value which closely resembled the total mass applied significantly undermining the power of the PLI as an independent tool for policy determination.

All subsequent experiments with the redefinition of aggregation constants during Phases 3 and 4 were an attempt to resolve these issues with the maximum value approach. The first attempts were to set all aggregation constants to a fixed value (e.g. 1). This was insufficient to solve the issues around the dominance of metrics with a narrow intrinsic range and produced results that were nearly indistinguishable from those of the 'policy neutral' maximum values. Another attempt defined the aggregation constants such that as opposed to being a 'score' for the worse case substance, they were instead interpreted as the 'score' of the 'median' substance (i.e. they were placed in the centre of the overall distribution rather than at the extremes). This change has the effect of inverting the relative impact of skew on how different metrics contribute to the headline value (i.e., under this approach, the most skewed metrics, which tend to be derived from ecotoxicity have the largest impact on the overall value). This fundamental re-orientates the indicator to be in terms of sensitivity to change in relative toxicity but was perceived as harder to communicate and subject to comparable 'biases' as the maximum score approach. Finally, an attempt was made which defined the value of aggregation constants based on an external standard, e.g. to set the values of all constants such that when calculated over the in 2010 survey of all arable crops each metric made up some predefined percentage of the total (referred to as the variable approach in Rainford et al., 2022). While this last approach is (by design) perhaps the most intuitively balanced way to construct a headline value it was found to be too counterintuitive and open to misinterpretation by stakeholders and was ultimately withdrawn.