Original Article **Integrated Environmental Assessment and Management**

DOI 10.1002/ieam.4078

Evaluating EFSA protection goals for honey bees (*Apis mellifera*): what do they mean for pollination?†

Simon Croft†* , Mike Brown‡, Selwyn Wilkins§, Andy Hart§, Graham C. Smith†

† National Wildlife Management Centre, Animal and Plant Health Agency, Sand Hutton, York, YO41 1LZ.

‡ National Bee Unit, Animal and Plant Health Agency, Sand Hutton, York, YO41 1LZ.

§ Fera Science Limited, National Agri-Food Innovation Campus, Sand Hutton, York, YO41 1LZ.

* Corresponding author (simon.croft@apha.gsi.gov.uk)

All Supplemental Data may be found in the online version of this article at the publisher's website.

Submitted 17 October 2017; Returned for Revision 3 April 2018; Accepted 12 June 2018

Abstract

In recent years there has been growing concern regarding the sudden and unexplained failure of honeybee (*Apis mellifera*) colonies. Several factors have been suggested including pesticides. In an effort to regulate their impact guidance has been published by the European Food Safety Authority (EFSA) recommending that the magnitude of effects on exposed colonies should not exceed 7% reduction in colony size after 2 brood cycles. However, fears have been raised regarding the practicality of measuring such a loss in the field. It is also unclear how this protection goal relates to maintaining the ecosystem services provided by bees, which we argue should be a primary objective for regulators. Here, we evaluate what these protection goals mean in relation to ecosystems performance using a computational colony model incorporating mechanisms to simulate both lethal and sub-lethal pesticide effects. To these simulations we apply a testing regime similar to that commonly used in field trials to produce standard assessment metrics. By relating these measures to losses in forager activity we aim to identify which could be used as effective indicators of reduced ecoservice and to quantify acceptable limits below which performance can be maintained. Our findings show that loss of colony size is the best indicator of reduced ecoservice. Metrics which focus on specific colony functions such as increased brood or forager mortality are ineffective indicators for all types of simulated pesticide effects. At the levels of colony loss recommended by EFSA, using our default parameterisation, we predict a loss of ecosystems performance of 3-4%. However, based on an extensive sensitivity analysis it is clear that this estimate is subject to substantial uncertainty with losses under alternative parameterisations of up to 14%. Nevertheless, our model provides a valuable framework for assessing protection goals, allowing regulators to test relevant impacts and quantify their magnitude. This article is protected by copyright. All rights reserved

Keywords: colony; simulation; pesticide; ecosystems service; regulation

Introduction

Honeybees (*Apis mellifera*) perform vital ecosystems services, such as pollination of crops and honey production, which contribute to biodiversity and the economy (Millennium Ecosystem Assessment, 2005; DEFRA, 2014). Whilst reports suggest a global increase in stocks of managed colonies (Aizen and Harder, 2009) there has been increasing concern regarding instances of sudden losses at a regional scale (Pettis and Delaplane, 2010; Potts et al., 2010; Budge et al., 2015; 2016). Extensive empirical research has been conducted to determine the cause of these events focusing on individual stressors including Varroa mites: *Varroa destructor* (Le Conte et al., 2010; Rosenkranz et al., 2010); pathogens such as *Nosema* (Higes et al., 2009) and bee viruses transmitted by Varroa (Cox-Foster et al., 2007; Genersch and Aubert, 2010); and exposure to pesticides (Henry et al., 2012; Thompson, 2003). As yet the underlying cause of colony collapse remains unclear (Ratnieks and Carreck, 2010; Watson and Stallins, 2016). Nevertheless, regulation is required to ensure that stocks are protected.

Laboratory and field studies assessing the impact of pesticides on honeybees have identified a variety of different effects, both lethal (i.e. direct mortality of brood, bees in the hive and foragers) and sub-lethal (for example, reduced adult efficacy as a result of exposure during brood development or inhibited navigational ability of foragers leading to additional mortality) (Thompson, 2003). To limit the role of pesticides in the decline of bee health the European Food Safety Authority (EFSA) has recently published guidance (European Food Safety Authority, 2013) based on a combination of expert consultation and mathematical modelling (Khoury et al., 2011) which outlines specific protection goals for honeybees to ensure “negligible” impact. This proposes that pesticide effects should not exceed 7% reduction in colony size (assumed to mean adult bees) after two brood cycles and that forager mortality should not be increased compared to controls by a

factor of more than 1.5 for six days, or a factor of more than 2 for three days, or a factor of more than 3 for two days. However, concerns have been raised regarding the practicality of measuring these protection goals in the field given the natural variation in honey bee colonies (Health and Safety Executive (HSE), pers. comm.) with a recent study suggesting that in order to detect changes of this magnitude with an appropriate statistical power, 68 replicates would be required (Woodcock et al., 2016). Naturally, such large field trials would be extremely costly and time consuming. This is where modelling can help providing a platform in which various effects can be explored in isolation.

Several colony models have been developed to assess the impact of pesticides on honeybees (Becher et al., 2014; Henry et al., 2012; Khoury et al., 2011; 2013; Russell et al., 2013). However, all of these models have limited exploration to the impacts of direct forager mortality. Whilst recently models have begun to explore the impact of other pesticide effects (Betti et al., 2017; Thorbek et al., 2017) prior to the work present here only the models by Thompson et al. (2005; 2007) and latterly Rumkee et al. (2015) provided such functionality. The latter provides one of the most comprehensive models to date (the Bee++ model by Betti et al., 2017 now provides similar detail) extending the BEEHAVE multi-stressor model (Becher et al., 2014). They assessed pesticide effects across a range of simulated experiments comparing impact based on a new index, LIS50, which indicated the level of stress on individuals that result in 50% colony mortality. Whilst informative in terms of measuring impact of pesticides on colony survival, the authors noted that this alone may not be an appropriate protection goal endpoint. We argue in particular that protection goals should also demonstrate clear consideration for the impacts on ecosystems services, primarily pollination and honey production, something which is not obviously achieved by the current EFSA guidance.

We aim to address this issue by assessing what the EFSA protection goals mean in terms of ecosystem performance. In order to achieve this we developed our own computational model and applied it to simulate a regulatory field trial exploring various potential pesticide exposure scenarios, including both lethal and sub-lethal effects on the colony. Using the outcome from these simulations we assessed the relationship between common field metrics routinely collected during assessments, including reduction in colony size, and loss of ecosystems performance; firstly, to identify appropriate indicators of reduced ecosystem performance and secondly, to inform the setting by regulatory authorities of limits below which impacts can be considered "negligible".

Methods and materials

Modelling framework

Our model was based on that described by Thompson et al. (2005; 2007) incorporating additional mechanisms and revised parameterisation from more recent studies (Becher et al., 2014; Russell et al., 2013; Schmickl and Crailsheim, 2007). It was written and implemented in Python™ version 2.7. All source files and operating instructions are provided in the supplementary material (Files S2 & S3). An overview of the model framework is provided in Figure 1. It considered a colony comprising of an egg-laying queen, age-structured brood cohorts, and an adult population of younger nurse bees and older forager bees. Updates were applied on a daily time-step by removing individuals according to computed mortality rates then ageing the remaining colony and simulating queen egg-laying. At the end of each time step summary statistics were produced detailing the current state of the colony, for example, the number of individuals at each life stage.

Colony processes

Mortality was applied using fixed baseline rates for each life stage (eggs, young larvae, old larvae, pupae, nurses and foragers). For some life stages these rates were subject to increase in

response to particular stresses within the colony. The mortality of adults was considered task-dependant (Schmickl and Crailsheim, 2007) with daily rates for nurses and foragers computed as a weighted average between the effective proportion of each available workforce performing at maximum capacity (referred to as effort), and therefore subject to a higher mortality, against the proportion at rest which experience baseline mortality. Nursing effort was computed as the number of nurses required to care for the existing larvae population assuming a fixed maximum larvae per nurse (BC_{max}) divided by the total number of nurses. Foraging effort was more complex and relied on external factors such as weather and resource availability (flowers). For simplicity we applied a fixed annual pattern derived from the mortality rates presented by Russell et al. (2013) using the daily mortality of a foraging adult (assumed as 0.161 from Visscher and Dukas 1997) to calculate the equivalent proportion which must be active. Consequently the resulting mortality of foragers followed the seasonal trends expressed in Russell et al. (2013). The mortality of brood can be affected by a number of factors including insufficient brood care and starvation. Our model did not consider resource dynamics within the colony, the justification for which is discussed elsewhere, and consequently the latter was not considered. However, a feedback mechanism relating to the effects of limited brood rearing capacity was incorporated. The model assumed that brood care was only required by larvae and that, in accordance with the observations of Schmickl and Crailsheim (2001) and the implementation in the BEEHAVE model (Becher et al., 2014), younger larvae (0-3 days) are most affected by reduced brood care (either as a result of cannibalisation or termination) relative to older larvae (4-5 days) which are preferentially protected. Consequently, if the maximum amounts of brood care available within the colony, calculated as the maximum larvae per nurse multiplied by the number of potential nurses, falls below that required (the total number of larvae) the survival of young larvae was reduced by a factor proportional to the relative deficiency.

Ageing of the colony follows standard timings for worker brood development with eggs hatching after 3 days, capping larvae after a further 8 days and emergence of pupae into adults after 21 days (Fukuda and Sakagami, 1968). In order to implement this process a cohort based mechanism was imposed. At the end of day the oldest cohort was added to the adult nursing subpopulation to be replaced by the next youngest and so on until all cohorts have been migrated. The youngest cohort was populated according to the queen's egg laying rate which was determined by a seasonal function. For consistency with the seasonal variation in forager mortality we used the seasonal egg laying presented in Russell et al. (2013) derived from the empirical measurements of Harbo (1986). An additional feedback mechanism was included to reduce egg laying if available brood care is limited. This represents the notion that nurse bees will remove eggs to avoid potential termination at a later brood stage (Becher et al., 2014).

The transition from nurse bees to foragers is perhaps the most difficult to simulate and there were significant differences in the way previous models represented this mechanism. Thompson et al. (2007) suggested a variable rate of transition dependant on the number of pupae emerging with a fixed ratio of nurses ageing per emergence. This method maintained an "ideal" ratio of nurses per larvae for brood care, and as a result of the seasonal variation in brood production, captured the change in ageing of normal adult honey bees, at around three weeks (Graham, 1992), to over-wintering bees which remained able to rear brood in the following spring. Whilst this process reproduces realistic colony dynamics under no stress it does not allow for potential feedback mechanisms to adjust the rate of transition as a result of perturbations in the ideal numbers of both nurses for brood rearing or foragers for resource acquisition. To include such feedback mechanisms we adopted a method similar to that outlined in BEEHAVE (Becher et al., 2014) where the age of first flight (AFF) was adjusted depending on the state of the colony. This approach used observed

threshold limits for "ideal" division of labour within a stable colony to maintain the levels of brood care, ratio of foragers to hive bees and resource stores by moving the AFF accordingly within a fixed range (3-65 days). AFF was increased if there was not enough brood care, the ratio of foragers to hive bees was too high (simulating social inhibition as introduced by Khoury et al. 2011) or the current value was greater than 7 days above the base AFF. Conversely, it was decreased if the ratio of foragers to hive bees was too low or the current value was less than 7 days below the base AFF. Rather than a constant base AFF as used in BEEHAVE we applied a variable value dependant on the same rate of transition implemented in Thompson et al. (2007) with a nurse to brood ratio of 1, meaning that in a stable colony a nurse will on average spend 21 days nursing.

Pesticide exposure

To explore the impact of exposure to various pesticides we incorporate three mechanisms simulating commonly reported effects: (i) increased forager mortality applied as a reduction in the survival of active foragers only, effectively an increase in the task-specific forager mortality (this includes directly lethal and sub-lethal effects affecting forager navigation which reduce the probability of successful return to the hive); (ii) increased brood mortality applied as a reduction in the survival of uncapped brood, eggs and larvae; (iii) reduced efficacy of adult workers represented by a reduction in survival of adults exposed during development, i.e. a reduced lifespan due to increased stress associated with performing at the same level as healthy adults. The latter required affected individuals to be monitored through their development and into adulthood. This was achieved by creating additional adult subpopulations of nurses and foragers for affected individuals. An individual was deemed to have been affected if laid during pesticide exposure. Therefore any individuals emerging from pupation within 21 days of pesticide exposure were considered affected and were diverted to the affected adult population where reduced survival may then be applied. Other than mortality rates individuals in affected subgroups were considered to behave identically

to those which are healthy. However, this method of separation within the adult population could be used in future to simulate differences linked to other pesticide effects, for example, reduced brood caring capacity similarly as a result of interrupted brood development.

Assumption and limitations

As already mentioned, unlike some colony models (Becher et al., 2014; Russell et al., 2013; Schmickl and Crailsheim, 2007), explicit dynamics of hive stores such as pollen and nectar were not considered as it is assumed that good beekeeping practices will be observed during regulatory field trials which includes food supplementation (S. Wilkins pers. comm.) to prevent starvation and hive management to ensure continued production of honey i.e. no curtailment of foraging due to lack of hive space; thereby removing the need for the additional complexity of associated feedback mechanisms. The consequences of this choice are discussed further in the supplementary material (File S1). The effects of site based factors such as weather are also omitted as, in accordance with good experimental design which would seek to minimise differences between sites, we can assume identical environmental conditions. Any general changes in environment can be reflected implicitly in the choice of parameterisation, for example the choice of seasonal cycle controlling active foraging patterns. Similarly, we do not consider a spatially explicit landscape as developed in BEEHAVE (Becher et al., 2014). We recognise that such functionality is necessary to explore specific mechanisms, for example exposure routes into the hive. However, we argue that the additional complexity is unnecessary in this case; specifically as sites are assumed identical thus factors such as the distance to foraging patches can be ignored and since regulators cannot legislate to restrict application spatially. Consequently any evaluation should be based on a "worst case" scenario which assumes 100% uptake of the pesticide by growers within a local area.

Parameterisation

One of the specific failings of the original Thompson model (Thompson et al., 2005; 2007) was a lack of clear referenced parameterisation, which is a requirement of good modelling guidelines (EFSA Panel on PPR, 2014; HM Treasury, 2015). We therefore sought to provide this by re-parameterising the model based on updated values obtained from existing literature. Where appropriate, parameter values were selected to produce realistic colony development (for details see supplementary material; File S1). A detailed summary of each parameter together with the potential range of values is given in Table 1.

In the absence of biologically realistic probability distributions for each parameter, stochastic variation was not included in the model. As we are primarily interested in the relative difference between colonies in the presence of a pesticide stochastic approaches may obscure the underlying effects.

To quantify the uncertainty in parameter values associated with the model outputs we performed an extensive sensitivity analysis conducting simulated experiments (see Simulations) using all possible combinations of parameter values identified from the available literature (Table 1: over 3000 distinct permutations). In addition we also conducted a more standard analysis varying default parameter values in turn by an arbitrary factor and observing the change in output metric. The results of this analysis are presented in the supplementary material (File S1) and show, in agreement with similar studies (Becher et al., 2014), that parameters controlling egg-laying (i.e. the ability of a colony to replace lost workers) and mortality, specifically that of foragers, were the most sensitive.

Simulations

We performed simulations considering each specific pesticide effect one-at-a-time and then an extreme case applying all three together. For each, a range of scenarios were modelled considering different timings, durations, and sizes of effects: the timings simulated were exposure start-dates of 1st April, 1st May, and 1st June; the durations of effects simulated were 1, 3, 7, 14 and 28 days; the sizes of effects simulated were 20%, 40%, 60% and 80% effects. Each simulation was run from 1st January with an initial adult population of 20000 bees (13333 nurses and 6667 foragers) for 3 years with pesticide exposure implemented in year 2. A range of model outputs were produced for each treatment and compared with a control simulation to quantify the effects of pesticides in line with the reviews of honey bee ecosystems service and existing field measurement techniques.

To measure ecosystem performance (pollination and honey production), we computed the total number of forager days (collective time in days foragers spend outside the hive computed daily according to available workforce and seasonally dependant conditions) across the key foraging season, 1st April to 31st August, each year post exposure. The resulting quantity was compared with that of the control and expressed as a percentage loss.

To reproduce potential field metrics we compared the outputs from each treatment with that of the control and calculated: (i) the maximum percentage loss of brood and adults for one year post exposure using a typical testing schedule of weekly for 6 weeks and then monthly on the last day of subsequent months (S. Wilkins pers. comm.); (ii) the percentage loss of overwintering adult population tested on 31st October each year post exposure; (iii) the maximum relative increase in forager mortality using a 2 day, 3 day and 6 day moving average within a daily testing period of 28 days post exposure as described by EFSA guidance; and (iv) the relative increase in brood mortality

of a single cohort over one brood cycle in accordance with typical laboratory and photo-capture brood studies. The reader should note that the testing schedule used to monitor losses of brood and adults is different to that recommended by EFSA. However, additional analysis presented in the supplementary material (File S1) showed good comparability between metrics recorded using either schedule.

Results and analysis

Ecosystems service can be challenging to measure in the field. Recent technological advances have the capability to measure potential proxies such as forager activity but can be prohibitively expensive. To provide a simple alternative we explored the relationship between various standard measures which are routinely used in field trials and foraging activity. For each measure we plotted the relative losses observed across the full range of pesticide scenarios to determine whether the metric could provide a suitable basis for protection goals.

Figure 2(a) shows the relationship between the maximum observed losses of adults observed in the year of exposure, assumed to be EFSA's definition of colony size, against loss of forager days. This shows a strong correlation which would suggest loss of adults would be a good indicator of a loss in ecosystems service. The relationship is generally linear for all single effect simulations with a gradient of approximately 0.5, i.e. given a 7% loss of adult population we would anticipate on average a 3.5% loss in ecoservice.

The relationship becomes more non-linear for simulations applying a combination of all pesticide effects, particularly those which with more intense prolonged exposures. However when combined, the magnitude of impact does not appear to be additive, i.e. the effects are not as great as the sum of their parts. Exploring the effects individually we noted that sub-lethal pesticide effects

(in this case applied as a delayed increase in mortality simulating impaired performance) appear to have a lower relative effect on colony performance compared to direct mortality.

Figures 2(b), brood loss, and 2(c), forager mortality, plot alternative field measures similar to that outlined by EFSA. These showed no correlation with the loss of forager days and therefore should not be considered suitable measures to indicate reduced ecosystems service. Each measure only captured effects directly affecting its associated life-stages, i.e. measuring brood loss only captured direct mortality of brood. No measurable level of brood loss occurred even with high levels of adult mortality, whereas loss of adults captured the impact from all pesticide effects. In the case of forager mortality the highly stratified pattern may be due to a disparity between monitoring periods with forager days reflecting a long-term impact and increased forager mortality reflecting only a short-term impact which does not fully capture all aspects of an exposure; e.g. differences in the duration of exposure. Additional results shown in the supplementary material (File S1) suggest that increasing the period over which increases are monitored did reduce stratification. However, the period would need to be extended substantially beyond the 6 day maximum considered by EFSA to provide a meaningful relationship which could be used to predict loss of ecosystems performance.

We also considered increased brood mortality over a cycle which is often used in laboratory studies of pesticides. However, the practicality of monitoring this metric under field conditions may be more limited than those above. As for previous metrics Figure 2(d) plots recorded increases against loss of foraging. Despite a longer monitoring period than that used to compute forager mortality we observed a similarly stratified relationship likely reflecting the model input. This may be as a consequence of the assumption that pesticides only affected uncapped brood (eggs and larvae) thereby limiting any impacts across all brood and potentially reducing the effective

measurement period closer to 8 days. Similar to loss of brood and increase in forager mortality, brood mortality only increased when targeted directly. There were no indirect increases as a result of adult loss suggesting substantial redundancy in the brood care capacity of the colony which would explain how colonies are generally able to recover so rapidly following pesticide exposure.

All of the impacts discussed above relate to observations made during the year of exposure. For all treatments, recovery from pesticide exposure was sufficiently rapid so as not to impact colony survival or have any lasting impact on colony performance in subsequent years, i.e. no loss was observed in the year after exposure. This may in part be due to interventions from good management practices as was assumed in the model. If this were not the case then the loss of foragers may impact the colonies ability to rear brood and hence slow recovery (at least for the scenarios explored here further analysis in the supplementary material provides no evidence to support this showing that when the condition is removed results remained unchanged). Additionally, the choice of exposure timings may also be a contributing factor as the latest date of exposure is June which leaves a sizeable proportion of the season for recovery to occur.

Figure 3 illustrates the results of the sensitivity analysis based on values from published literature (see Parameterisation). As described previously we performed pesticide scenarios using all possible combinations of parameter values extracted from other modelling studies. The aim was to provide an indication of parameter uncertainty associated with any potential indicators of forager loss and to use this information to suggest appropriate quantitative limits for protection. Shaded regions are combined ellipses centred on the sample mean and containing 50%, 75%, 90%, 95% and 99% of results for each pesticide scenario. These regions are not equivalent to confidence or probability regions as we have not taken into account the relative likelihood of different parameter values. The results may underestimate uncertainty, because only reported values were considered

and only a single value was available for 11 of the 19 parameters (see Table 1). However, the results could also overestimate uncertainty, because errors would cancel out between parameters to some degree. Nevertheless, it is clear that the degree of uncertainty may be substantial.

The plot (Figure 3) shows losses in forager days are highly uncertain for any given loss of adults. Importantly, there is no threshold value for loss of adults below which we can be certain ecosystems performance is unaffected. However, irrespective of parameterisation the trend remains approximately linear. There is some indication that the default parameterisation (shown by the points in Figure 3) may tend to underestimate effects when adult losses are high.

Discussion

In this study we outlined a model to assess the impact of pesticides on honeybees and used it to conduct a series of simulated field trials applying pesticide exposures with different effects, intensities, durations and seasonal timings. Using the outcome from these simulations we explored the relationship between typical field metrics, including those used to define protection goals published by EFSA, and ecosystems performance which we argue is an important endpoint for regulation but can be difficult to measure empirically. The output (Figure 2) suggests that of the field metrics tested, a reduction in maximum colony size (adult bees) is the best indicator of reduced ecosystems performance. Metrics which focus on specific colony function such as increased brood or forager mortality are ineffective indicators for all types of simulated pesticide effects. This is perhaps intuitive if we consider that losses in brood will inherently impact future adult populations whereas losses of adults will not necessarily be sufficient to impact brood care and therefore produce any losses in brood.

Assessing the suitability of the 7% limit on adult bee losses outlined by EFSA guidance in terms of potential impact on ecosystem services the model under the default parameterisation

predicts a 3-4% loss of forager days (Figure 2a). However, sensitivity analysis shows that these results may be subject to substantial uncertainty (Figure 3). This uncertainty is problematic in determining a threshold for colony loss below which impacts could be regarded as "negligible". Even as observed losses of adults approach zero there remains a possibility of reduced foraging (perhaps up to 5% loss, or even more), caused as a result of responsive changes in the colony's population structure; the transition of foragers is slowed reducing the average mortality of adults, mitigating any increases from pesticides, but in turn also reducing foraging. On the other hand, it is also possible that adult losses could reach 35% or more before there will be any impact on the annual total of forager days (Figure 3).

Regulatory authorities may wish to set a threshold for adult losses that would ensure, with some stated level of confidence, that impacts on ecosystem services would not exceed a specified level. Our sensitivity analysis provides only an indication of the possible range of uncertainty around our default estimates. To provide confidence levels it would be necessary to quantify uncertainty probabilistically. One example of such an approach is outlined by Holland et al. (2009), who use values from published studies to fit either a beta-binomial or Poisson-gamma distributions from which values can be randomly generated according to their probability.

We also recognise that the model includes a number of simplifications. In particular we assumed good beekeeping management which mitigates the effects from depletion of resources, primarily a loss in foraging capability. To examine the validity of this assumption further testing was conducted comparing results against an extended model with explicit simulation of food stores (pollen and honey). Whilst this suggested a difference in colony dynamics, for example good management allows the colony to persist without foragers; the relationships between relative metrics describing the impact of treatments were identical. This was likely because the duration of

pesticide exposure simulated, regardless of intensity or timing, was not sustained long enough for stores to become fully depleted causing starvation. We therefore conclude that the omission of food stores was justified for the scenarios explored in this study. However, the reader should be aware that for more extreme scenarios the validation does not necessarily hold and further assessment would need to be conducted. Additional simplifications included a lack of consideration for specific regional weather dependence, as recommended by EFSA modelling guidance (EFSA PPR Panel, 2014), and no consideration for drone production which limit the general applicability of the model. Without food stores the latter may appear unnecessary as drones do not contribute to the colony but recent observations by Henry et al. (2015) suggest honeybees reduce drone production and increase worker production, to cope with increased forager mortality. Whilst explicit drone production may not be necessary the model could be developed to account for this by adding a forager-dependant egg laying rate.

Notwithstanding the above suggestions, this work has demonstrated that bee colony modelling may usefully be used to investigate standard field measures of colony success, against which protection goals are defined, to regulate effects on ecosystem performance. We found that the total number of adults appears to be the best measure of colony function within the ecosystem and thus the best measure to help define protection goals.

Acknowledgements

This work was conducted in collaboration with Fera Science Ltd. as part of a review of EFSA guidance on protection goals for bee commissioned by the Health and Safety Executive (HSE). We would like to acknowledge colleagues at HSE for their helpful feedback throughout and to DEFRA for funding the project (PS2377).

Data accessibility

All data used and generated as part of this study are included within the main text or supplementary material. For specific queries and requests readers are welcome to contact the corresponding contact author directly.

Supplementary material

S1 File: additional model description and analysis.

S2 File: description of computation resources and model operation.

S3 File: repository containing model source files (standard model and model with explicit resource dynamics).

References

Aizen, M.A., Harder, L.D., 2009. The global stock of domesticated honey bees is growing slower than agricultural demand for pollination. Curr Biol 19, 915-918.

Becher, M.A., Grimm, V., Thorbek, P., Horn, J., Kennedy, P.J., Osborne, J.L., 2014. BEEHAVE: a systems model of honeybee colony dynamics and foraging to explore multifactorial causes of colony failure. J Appl Ecol 51, 470-482.

Becher, M.A., Hildenbrandt, H., Hemelrijk, C.K., Moritz, R.F., 2010. Brood temperature, task division and colony survival in honeybees: a model. Ecol Model 221, 769-776.

Beshers, S.N., Fewell, J.H., 2001. Models of division of labor in social insects. Annu Rev Entomol 46, 413-440.

Betti, M., LeClair, J., Wahl, L., Zamir, M., 2017. Bee++: An object-oriented, agent-based simulator for honey bee colonies. Insects 8, 31.

Bodenheimer, F., 1937. Studies in animal populations. II. Seasonal population-trends of the honey-bee. Q Rev Biol 12, 406-425.

Budge, G.E., Pietravalle, S., Brown, M., Laurenson, L., Jones, B., Tomkies, V., Delaplane, K.S., 2015. Pathogens as predictors of honey bee colony strength in England and Wales. PLoS ONE 10, e0133228.

Budge, G.E., Adams, I., Thwaites, R., Pietravalle, S., Drew, G.C., Hurst, G.D.D., Tomkies V., Boonham, N., Brown, M., 2016. Identifying bacterial predictors of honey bee health. J Invertebr Pathol 141, 41–44.

Cox-Foster, D.L., Conlan, S., Holmes, E.C., Palacios, G., Evans, J.D., Moran, N.A., Quan, P.L., Briese, T., Hornig, M., Geiser, D.M., Martinson, V., 2007. A metagenomic survey of microbes in honey bee colony collapse disorder. Science 318, 283-287.

DEFRA (Department for Environment, Food and Rural Affairs)., 2014. National pollinator strategy: for bees and other pollinators in England: www.gov.uk/government/publications/national-pollinator-strategy-for-bees-and-other-pollinators-in-england

DeGrandi-Hoffman, G., Roth, S.A., Loper, G., Erickson, E.H., 1989. Beepop: a honeybee population dynamics simulation model. Ecol Model 45, 133-150.

European Food Safety Authority, 2013. Guidance on the risk assessment of plant protection products on bees (*Apis mellifera*, *Bombus spp.* and solitary bees). EFSA Journal 11, 3295.

EFSA Panel on PPR (Plant Protection Products and their Residues), 2014. Scientific opinion on good modelling practice in the context of mechanistic effect models for risk assessment of plant protection products. EFSA Journal 12, 3589.

Fukuda, H., Sakagami, S.F., 1968. Worker brood survival in honeybees. Res Popul Ecol 10, 31-39.

Genersch, E., Aubert, M., 2010. Emerging and re-emerging viruses of the honey bee (*Apis mellifera L.*). Vet Res 41, 54.

Graham, J.M., 1992. The hive and the honey bee. Dadant & Sons, Hamilton, IL, USA. 1324 pp.

Harbo, J.R., 1986. Effect of population size on brood production, worker survival and honey gain in colonies of honeybees. J Apic Res 25, 22-29.

Henry, M., Beguin, M., Requier, F., Rollin, O., Odoux, J.-F., Aupinel, P., Aptel, J., Tchamitchian, S., Decourtye, A., 2012. A common pesticide decreases foraging success and survival in honey bees. Science 336, 348-350.

Henry, M., Cerrutti, N., Aupinel, P., Decourtye, A., Gayrard, M., Odoux, J.-F., Pissard, A., Rüger, C., Bretagnolle, V., 2015. Reconciling laboratory and field assessments of neonicotinoid toxicity to honeybees. Proc R Soc Biol Sci Ser B 282(1819), 20152110.

Higes, M., Martín-Hernández, R., Garrido-Bailón, E., González-Porto, A.V., García-Palencia, P., Meana, A., Del Nozal, M.J., Mayo, R., Bernal, J.L., 2009. Honeybee colony collapse due to *Nosema ceranae* in professional apiaries. Environ Microbiol Rep 1, 110-113.

HM Treasury, 2015. The Aqua Book: guidance on producing quality analysis for government. ISBN 978-1-910337-67-7.

Holland, E.P., Burrow, J.F., Dytham, C., Aegerter, J.N., 2009. Modelling with uncertainty: Introducing a probabilistic framework to predict animal population dynamics. Ecol Model 220, 1203-1217.

Khoury, D.S., Barron, A.B., Myerscough, M.R., 2013. Modelling food and population dynamics in honey bee colonies. PloS ONE 8, e59084.

Khoury, D.S., Myerscough, M.R., Barron, A.B., 2011. A quantitative model of honey bee colony population dynamics. PLoS ONE 6, e18491.

Le Conte, Y., Ellis, M., Ritter, W., 2010. Varroa mites and honey bee health: can Varroa explain part of the colony losses? Apidologie 41, 353-363.

Millennium Ecosystem Assessment, 2005. Millennium ecosystem assessment, ecosystems and human well-being: synthesis report. Island Press, Washington, DC.

Nolan, W.J., 1925. The brood-rearing cycle of the honeybee. Bull U.S. Dep Agric 1349, 1-56.

Pettis, J.S., Delaplane, K.S., 2010. Coordinated responses to honey bee decline in the USA. Apidologie 41, 256-263.

Potts, S.G., Biesmeijer, J.C., Kremen, C., Neumann, P., Schweiger, O., Kunin, W.E., 2010. Global pollinator declines: trends, impacts and drivers. Trends Ecol Evol 25, 345-353.

Ratnieks, F.L.W., Carreck, N.L., 2010. Clarity on honey bee collapse? Science 327, 152-153.

Rosenkranz, P., Aumeier, P., Ziegelmann, B., 2010. Biology and control of *Varroa destructor*. J Invertebr Pathol 103, Supplement, S96-S119.

Rumkee, J.C., Becher, M.A., Thorbek, P., Kennedy, P.J., Osborne, J.L., 2015. Predicting honeybee colony failure: using the BEEHAVE model to simulate colony responses to pesticides. Environ Sci Technol 49, 12879-12887.

Russell, S., Barron, A.B., Harris, D., 2013. Dynamic modelling of honey bee (*Apis mellifera*) colony growth and failure. Ecol Model 265, 158-169.

Schmickl, T., Crailsheim, K., 2001. Survival of honeybee larvae in times of pollen stress. Apidologie 32, 496-498.

Schmickl, T., Crailsheim, K., 2007. HoPoMo: a model of honeybee intracolonial population dynamics and resource management. Ecol Model 204, 219-245.

Thompson, H.M., 2003. Behavioural effects of pesticides in bees–their potential for use in risk assessment. Ecotoxicology 12, 317-330.

Thompson, H.M., Wilkins, S., Battersby, A.H., Waite, R.J., Wilkinson, D., 2005. The Effects of Four Insect Growth-Regulating (IGR) Insecticides on Honeybee (*Apis mellifera L.*) Colony Development, Queen Rearing and Drone Sperm Production. Ecotoxicology 14, 757-769.

Thompson, H.M., Wilkins, S., Battersby, A.H., Waite, R.J., Wilkinson, D., 2007. Modelling long-term effects of IGRs on honey bee colonies. Pest Manag Sci 63, 1081-1084.

Thorbek, P., Campbell, P.J., Thompson, H.M. (2017) Colony impact of pesticide-induced sublethal effects on honeybee workers: a simulation study using BEEHAVE Environ Toxicol Chem 36, 831-840.

Visscher, P.K., Dukas, R., 1997. Survivorship of foraging honey bees. Insectes Soc 44, 1-5.

Watson, K., and Stallins, J.A., 2016. Honey bees and colony collapse disorder: a pluralistic reframing. Geogr Compass 10, 222–236.

Winston, M.L., 1987. The biology of the honey bee. Harvard University Press, Cambridge, MA, USA. 296 pp.

Woodcock, B.A., Heard, M.S., Jitlal, M.S., Rundlöf, M., Bullock, J.M., Shore, R.F., Pywell, R.F., 2016. Replication, effect sizes and identifying the biological impacts of pesticides on bees under field conditions. J Appl Ecol 53, 1358-1362.

Figure 1: Diagram outlining the model structure. The model consisted of cohort based brood development with seasonal egg laying (Harbo, 1986; Russell et al., 2013) and background age related mortality. Larval mortality was increased depending on available brood care and pesticide exposure (brood mortality switch). After 21 days, brood hatched and became young hive workers (or nurses). Hatching workers were added to either the healthy or affected population depending on pesticide exposure (development switch). Affected populations were subjected to greater mortality to simulate a reduced adult longevity. Each time step an equal proportion of in-hive population was aged to become foragers. The proportion making this transition was dependent upon the state of the hive: numbers of brood hatching, the nurse to brood ratio and the ratio of foragers to nurses (these describe the relative size of the colony and ensure that optimal division of labour was maintained). Forager populations were further divided into active and resting bees by applying a seasonal factor representing environmental conditions; derived from Russell et al. (2013). The size of the active forager population represented an estimate of the number of forager days or pollination effort. Mortality of active foragers was applied at a greater level and this could be increased depending on pesticide toxicity. Resting foragers were considered identical to hive bees.

Figure 2: Relationship between standard field measures and protection goal metric. The colour of symbols indicates the type of pesticide effect (red, green, blue and black denoting forager mortality, brood mortality, adult longevity and all respectively), the shape indicates the timing of exposure (circle, square, triangle denoting April, May and June respectively) and the intensity indicates the duration of the effect (1, 3, 7, 14 and 28 days). Plots display: (a) maximum observed percentage loss of adults; (b) maximum observed percentage loss of brood; (c) maximum relative increase (multiplicative scale factor) in forager mortality over 3 days; (d) relative increase in brood mortality after 1 brood cycle; against the percentage loss of forager days. These suggest that loss of adults (colony size) is the only suitable indicator of reduced performance across all potential pesticide effects.

Figure 3: Effect of alternative parameter values on model estimates for the most appropriate field measure. Plot displays maximum loss of adults against loss of forager days with shaded regions to indicate the impact of parameter uncertainty. From darkest to lightest shading denotes 50%, 75%, 90%, 95% and 99% regions within which the corresponding percentages of simulation results lie. Black circles mark output using default parameter values. Importantly, this shows that there is no threshold value for loss of adults below which we can be certain performance is unaffected.

Table 1: List of model parameters and values. Where a range of values is available in the published literature details are given in brackets in the References column.

Parameter	Description	Value	References (Published values if different)
Max. egg laying rate (L_{max})	Limit describing the maximum number of eggs the queen can lay per day.	1308	derived as in Russell et al. (2013), Nolan (1925): 1500, Bodenheimer (1937): 1600, Khoury et al. (2011): 2000, DeGrandi-Hoffman et al. (1989): 3000
Seasonal egg laying factor (L_S)	Seasonal function describing the proportion of eggs laid relative to the maximum.	0 - 1	Russell et al. (2013)
Egg mortality (M_E)	Daily baseline mortality of eggs.	0.03	Fukuda and Sakagami (1968)
Larvae mortality (M_L)	Daily baseline mortality of larvae.	0.01	Fukuda and Sakagami (1968)
Pupae mortality (M_P)	Daily baseline mortality of pupae.	0.001	Fukuda and Sakagami (1968)
Adult mortality (M_H)	Daily baseline mortality of adults.	0.01	Schmickl and Crailsheim (2007), Becher et al. (2014): 0.004, Russell et al. (2013): 0.007
Additional nursing mortality (M_N)	Additional task specific mortality for active nurses.	0.005	Schmickl and Crailsheim (2007)
Additional forager mortality (M_F)	Additional task specific mortality for active foragers.	0.15	Visscher and Dukas (1997), Schmickl and Crailsheim (2007): 0.045
Forager activity (F_S)	Seasonal function describing the proportion of foragers actively forging on any given day.	0.16 - 0.71	derived from Russell et al. (2013)
Max. larvae per nurse (BC_{max})	Maximum number of larvae each nurse (hive worker) can care for on any given day.	3	Becher et al. (2014), Thompson et al. (2007): 1.5, Becher et al. (2010): 2.65
Larvae Per Nurse Threshold ($BC_{threshold}$)	Ideal ratio of larvae to nurse within the colony.	1	Becher et al. (2014): 0.3, Schmickl and Crailsheim (2007): 0.5 - 1, Thompson et al. (2007): 1.5

Min. forager to hive ratio (F_{min})	Minimum ratio of foragers per hive bee.	0.3	Beshers and Fewell (2001)
Max. forager to hive ratio (F_{max})	Maximum ratio of foragers per hive bee.	0.5	Schmickl and Crailsheim (2007)
Min. AFF (AFF_{min})	Minimum age at which a nurse can transition to a forager.	3	Winston (1987): 3 - 20, Becher et al. (2014): 7
Max. AFF (AFF_{max})	Maximum age by which a nurse must transition to a forager.	65	Winston (1987): 27 - 65, Becher et al. (2014): 50
ΔAFF brood care (AFF_{BC})	Increase in AFF if available brood care is below ideal threshold.	2	Becher et al. (2014)
ΔAFF foragers (AFF_{F})	Change in AFF if forager to hive ratio is above or below ideal threshold (+ & - change respectively).	1	Becher et al. (2014)
ΔAFF base (AFF_{base})	Change in AFF towards base AFF if greater than specified days difference.	1	Becher et al. (2014)
AFF base threshold ($AFF_{threshold}$)	Threshold difference in days between current AFF and base AFF.	7	Becher et al. (2014)

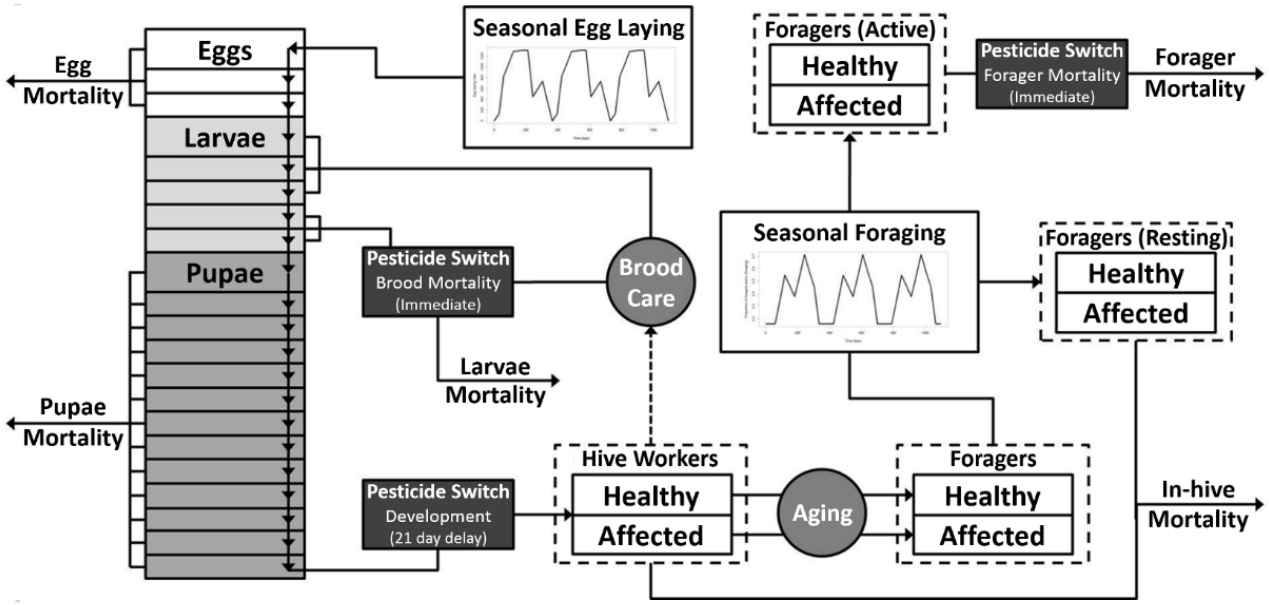

Figure 1

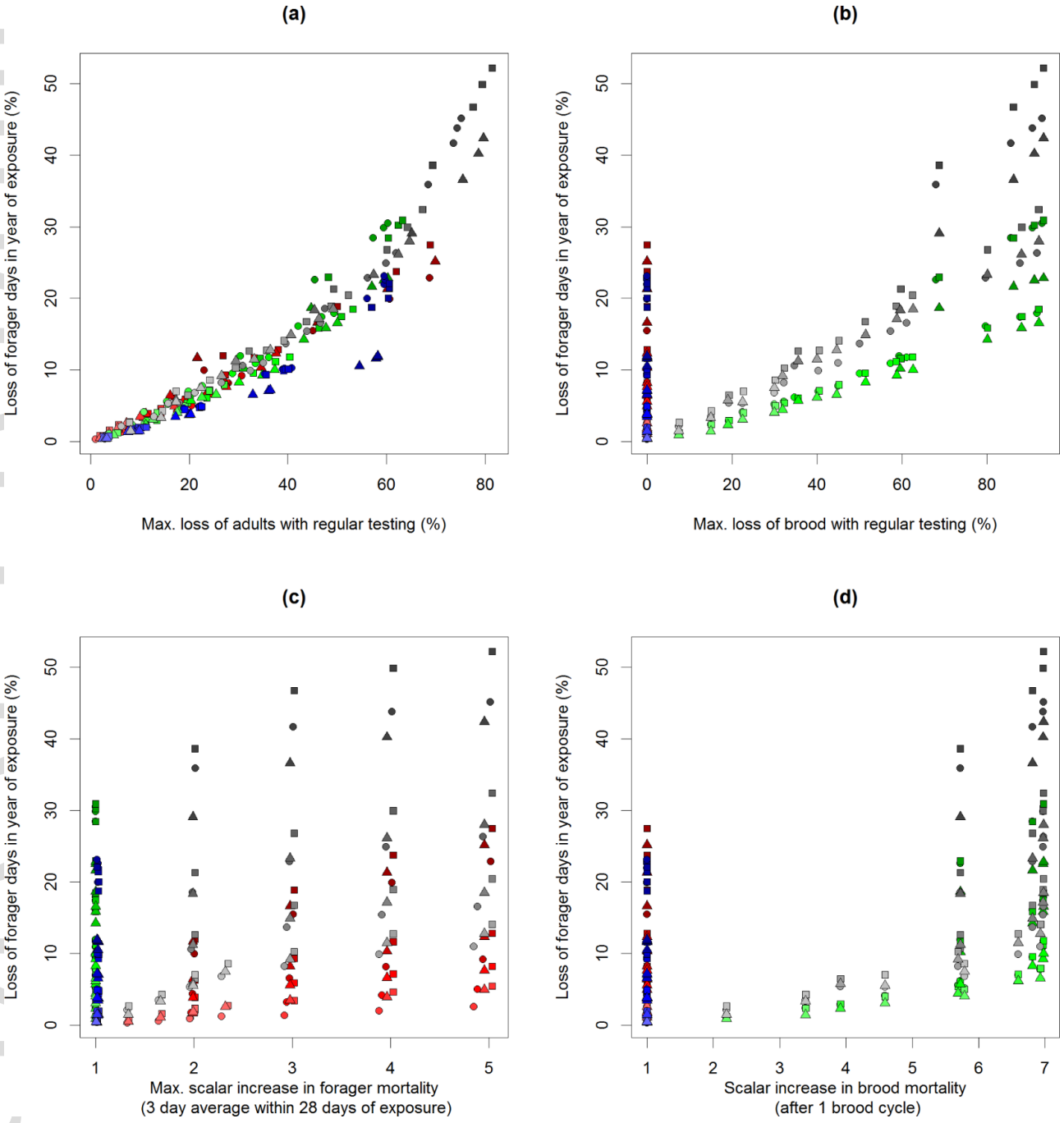

Figure 2

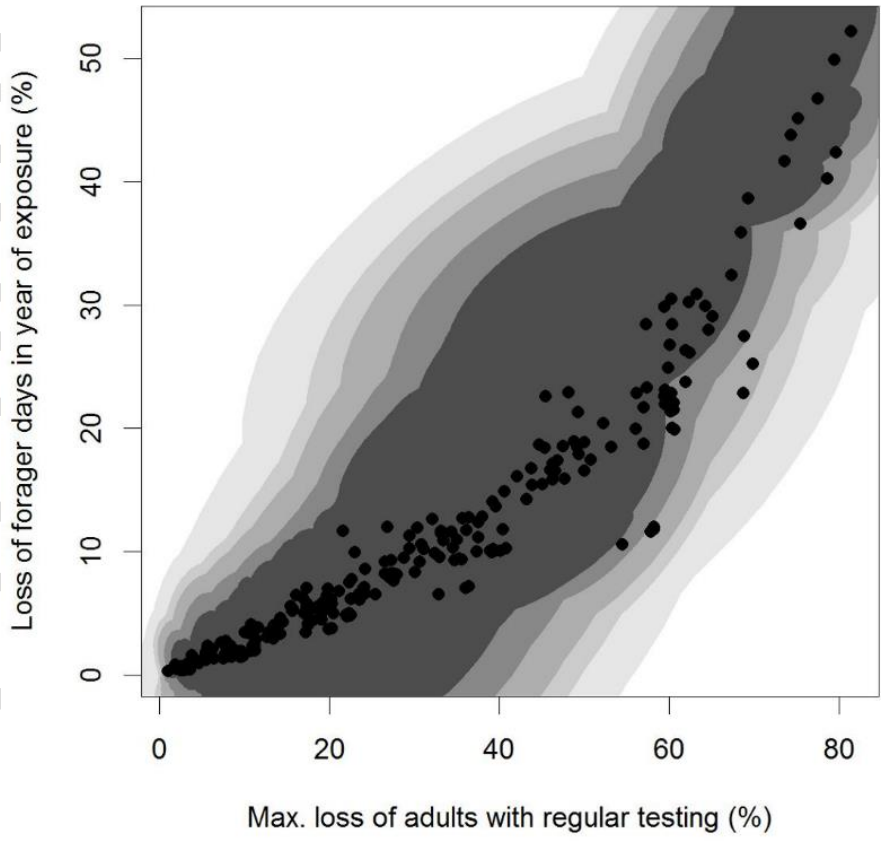

Figure 3