Science of the Total Environment 735 (2020) 139243

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

Integrating QSAR models predicting acute contact toxicity and mode of action profiling in honey bees (*A. mellifera*): Data curation using open source databases, performance testing and validation

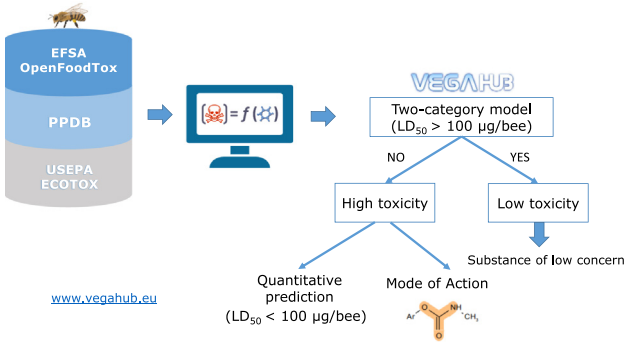
Edoardo Carnesecchi [a,b,*,1], Cosimo Toma [a,b,1], Alessandra Roncaglioni [b], Nynke Kramer [a], Emilio Benfenati [b], Jean Lou C.M. Dorne [c]

[a] Institute for Risk Assessment Sciences (IRAS), Utrecht University, PO Box 80177, 3508 TD Utrecht, the Netherlands
[b] Laboratory of Chemistry and Environmental Toxicology, Department of Environmental Health Sciences, Istituto di Ricerche Farmacologiche Mario Negri IRCCS, Via Mario Negri 2, 20156 Milan, Italy
[c] European Food Safety Authority (EFSA), Scientific Committee and Emerging Risks Unit, Via Carlo Magno 1A, 43126 Parma, Italy

HIGHLIGHTS

- Integrative QSAR models predict acute contact toxicity and profile Mode of Action.
- First harmonised Mode of Action classification scheme for honey bees
- Models were validated using robust internal and external parameters.
- Mode of Action for chemical grouping in component-based mixture risk assessment.
- K-NN algorithm improved the statistical quality of the models and their implementation.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 1 April 2020
Received in revised form 4 May 2020
Accepted 4 May 2020
Available online 17 May 2020

Editor: Yolanda Picó

Keywords:
QSAR models

ABSTRACT

Honey bees (*Apis mellifera*) provide key ecosystem services as pollinators bridging agriculture, the food chain and ecological communities, thereby ensuring food production and security. Ecological risk assessment of single Plant Protection Products (PPPs) requires an understanding of the exposure and toxicity. In silico tools such as QSAR models can play a major role for the prediction of structural, physico-chemical and pharmacokinetic properties of chemicals as well as toxicity of single and multiple chemicals. Here, the first integrative honey bee QSAR model has been developed for PPPs using EFSA's OpenFoodTox, US-EPA ECOTOX and Pesticide Properties DataBase i) to predict acute contact toxicity (LD_{50}) and ii) to profile the Mode of Action (MoA) of pesticides active substances. Three different classification-based and four regression-based models were developed and tested for their performance, thus identifying two models providing the most reliable predictions based on k-NN algorithm.

Abbreviation: 5-fold CV, 5 fold internal cross-validation; AOPs, adverse outcome pathways; AD, applicability domain; ADI, Applicability Domain Index; AGs, assessment groups; DT, decision trees; FRAC, Fungicide Resistance Action Committee; HRAC, Herbicide Resistance Action Committee; IRAC, Insecticide Resistance Action Committee; k-NN, k-nearest neighbor; MAE, mean absolute error; MCC, Matthews correlation coefficient; MRA, mixture risk assessment; MoA, mode of action; MLR, multiple linear regression; PCA, principal component analysis; PLS, partial least squares regression; PPDB, Pesticide Properties DataBase; PPP, plant protection products; QSAR, quantitative structure-activity relationship; RF, random forest; RA, risk assessment; RMSE, root mean square error; SMILES, simplified molecular input line entry system; TS, training set; VS, validation set; VSURF, variable selection (with) random forest.

* Corresponding author at: Laboratory of Chemistry and Environmental Toxicology, Department of Environmental Health Sciences, Istituto di Ricerche Farmacologiche Mario Negri IRCCS, Via Mario Negri 2, 20156 Milan, Italy.
E-mail address: carnesecchiedo@gmail.com (E. Carnesecchi).
[1] Edoardo Carnesecchi and Cosimo Toma equally contributed to this work.

https://doi.org/10.1016/j.scitotenv.2020.139243

Honey bees
Mode of action
Ecological risk assessment
Chemical mixtures

The two-category QSAR model (toxic/non-toxic; $n = 411$) was validated using sensitivity (=0.93), specificity (= 0.85), balanced accuracy (=0.90), and Matthews correlation coefficient (MCC = 0.78) as statistical parameters. The regression-based model ($n = 113$) was validated for its reliability and robustness ($R^2 = 0.74$; MAE = 0.52). Current study proposes the MoA profiling for 113 pesticides active substances and the first harmonised MoA classification scheme for acute contact toxicity in honey bees, including LD_{50s} data points from three different databases. The classification allows to further define MoAs and the target site of PPPs active substances, thus enabling regulators and scientists to refine chemical grouping and toxicity extrapolations for single chemicals and component-based mixture risk assessment of multiple chemicals. Relevant future perspectives are briefly addressed to integrate MoA, adverse outcome pathways (AOPs) and toxicokinetic information for the refinement of single-chemical/combined toxicity predictions and risk estimates at different levels of biological organization in the bee health context.

1. Introduction

The importance of pesticides as plant protection products (PPPs) in agriculture forestry, urban gardens, parks has been recognised worldwide, particularly to protect crops against pests (e.g. insects, weeds), diseases or pathogens (e.g. fungi), which may affect plants health and potentially reduce crop yield, thus potentially threatening food security (FAO, ITPS, 2017; Frische et al., 2018). However, concerns due to the potential harmful effects of PPPs including insecticides on ecosystems, particularly towards non-target species such as pollinators, have raised (Douglas et al., 2020; EFSA, 2018; Tosi and Nieh, 2019; Sanchez-Bayo and Goka, 2016; Simon-Delso et al., 2015; Tosi et al., 2018).

Indeed, pollinators such as honey bees (*Apis mellifera*), bumble bees (*Bombus* spp.) and solitary bees (e.g. *Osmia* spp.) play a key-role as ecosystems service providers (ESP) contributing to the maintenance, reproduction of wild plant communities and biodiversity as well as bridging agriculture, the food chain and the ecological communities, thereby ensuring food production and security (Breeze et al., 2011; Schulp et al., 2014; Rose et al., 2015). Honey bees also represent sentinel species together with their hive products as bioindicators (i.e. honey, pollen, beebread) to monitor environmental contamination by regulated products (e.g. PPPs, veterinary residues), anthropogenic (e.g. persistent organic pollutants, heavy metals, particulate matter) and natural contaminants (mycotoxins, plant alkaloids) (Negri et al., 2015; Bargańska et al., 2016; Tosi et al., 2018).

Moreover, honey bees are employed worldwide as surrogate species for *Apis* and non-*Apis* bees to perform toxicity tests on single pesticides (EFSA, 2013; USEPA, PMRA, CALDPR, 2014). In the EU, pesticide risk assessment (RA) requires the settings of protection goals and the evaluation of the environmental impact associated with the exposure and toxicity of PPPs use (Regulation EC, 1107/2009, 2009). For honey bees, the Regulation lays downs that "*an active substance should only be approved if it results in negligible exposure or has no unacceptable acute or chronic effects on colony survival and development, taking into account effects on honeybee larvae and honeybee behaviour*". In this context, the European Food Safety Authority (EFSA) performs the RA of single active substances based on toxicity data (e.g. LD_{50}) provided with the premarket registration dossiers (e.g. Draft Assessment Report) submitted by applicants (EFSA PPR Panel, 2012). Although there is growing evidence that bees are exposed to a wide range of multiple chemicals "mixtures" (David et al., 2016; Tosi et al., 2018; Prado et al., 2019) which, in some instances, potentially trigger interactions such as synergistic effects (Carnesecchi et al., 2019a, 2019b; Spurgeon et al., 2016), further work is needed to integrate information on such combined toxicity in RA practice (Rortais et al., 2017; Bopp et al., 2019; Topping et al., 2020). EFSA has recently published a MIXTOX guidance document to support harmonised methodologies for ecological RA of combined exposure to multiple chemicals while specifically illustrating the integration of information on combined toxicity for bees (More et al., 2019). In this context, key recommendations include the need to further develop and implement generic in silico models such as quantitative structure–activity relationship (QSAR) to predict combined toxicity for bees and a broader range of species of ecological relevance. These models can support the integration of toxicity and mechanistic data for hazard assessment of single chemicals as well as for component-based approaches for mixture risk assessment (MRA). As a consequence, innovative QSAR models to predict combined toxicity of pesticides active substances in honey bees have been developed recently allowing the identification of structural features that may drive an increase or decrease in combined toxicity (Carnesecchi et al., 2020). In addition, authors have also highlighted that current data gaps regarding information on the mode of action (MoA) of single chemicals is still limiting the development and broader applications of such innovative QSAR models.

In the human-health and animal health areas, MoA refers to the major steps leading to an adverse health effect following interaction of the chemical with biological targets at the sub-cellular level, while not necessarily implying the full understanding of the mechanism of action at the molecular level (WHO, 2009; Boobis et al., 2006; OECD, 2017; EFSA PPR, 2013). Similarly, in ecological RA, MoA has been defined as a functional change at the cellular level triggered by the substance entering the organism which then involves levels of biological organization from organisms, multiple species, to populations all the way to ecosystems (Kienzler et al., 2017, 2019; Segner, 2011). Several different MoA frameworks exist for classifying chemicals (Verhaar et al., 1992; Russom et al., 1997; Kienzler et al., 2017), which allowed developing robust predictive tools for MoA classification such as EnviroTox database (Kienzler et al., 2019; Connors et al., 2019) and TEST software (Martin et al., 2013, 2015). However, such tools are mostly based only on vertebrate information specifically fish toxicity data, thus limiting their application to aquatic environmental RA. Similarly, different QSAR models predicting toxicity for single chemicals in honey bees are available but to date these do not address the challenge of the integration of toxicity prediction together with MoA profiling (Venko et al., 2018; Singh et al., 2014; Hamadache et al., 2018; Como et al., 2017; Toropov and Benfenati, 2007; Devillers et al., 2002). MoA information together with information on adverse outcome pathways (AOPs) in honey bees can provide a sound understanding of the link between molecular targets, as molecular initiating event and key events leading to adverse effects at individual and colony level as recently for neonicotinoids targeting nicotinic acetylcholine receptors in honey bees (LaLone et al., 2017).

The present manuscript aims to address the challenge of integrating MoA information in QSAR models with the first integrative honey bee QSAR models for PPPs using open source databases i) to predict acute contact toxicity (LD_{50}) and ii) to profile the MoA of active substances. In addition, the current study explores the development of harmonised MoA classification schemes to relate and structure toxicological information with target sites of PPPs active substances for a range of applications, including toxicity predictions and refining the grouping of chemicals for component-based RA of multiple chemicals (More et al., 2019).

2. Materials and methods

2.1. QSAR model development

2.1.1. Data curation

Pesticide toxicity data for honey bees (*Apis mellifera*) expressed as LD_{50} μg/bee (acute contact, 48 h) were retrieved in June 2018 from three publicly available databases (1) EFSA's chemical hazards database "OpenFoodTox" (Benfenati et al., 2020; DOI: https://doi.org/10.5281/zenodo.3693783), (2) US-EPA ECOTOXicology knowledgebase (ECOTOX; available at https://cfpub.epa.gov/ecotox/) and (3) Pesticide Properties DataBase (PPDB; available at https://sitem.herts.ac.uk/aeru/ppdb/en/index.htm). Criteria for data pruning were applied following to the official guideline (OECD, 1998) according to which pesticides are administered by contact routes to represent the type of exposure under field conditions. Overall, information on the specific criteria applied for the data pruning within each database are presented in Table 1.

After the creation of a list of unique CAS numbers and names, all the SMILES have been retrieved with a semi-automated workflow (Gadaleta et al., 2018). SMILES provided with the original databases have been used for manual check and no differences have been found.

After an analysis on stereoisomers, it has been found that a high percentage of molecule had chiral points without any specification of chirality. This finding led us to the decision of stripping all the stereoisomer information from SMILES in order to have a more homogenous dataset.

SMILES have then been associated with original values and two different procedures have been adopted whether it was a classification or regression dataset building:

- **Classification-based models**; the threshold used for toxicity classification was 100 μg/bee, which corresponds to the limit test (OECD, 1998). If the values associated to the same SMILES fell under and up this threshold, the relative compound have been excluded for classification modelling. The final dataset is constituted by 413 compounds.
- **Regression-based model**; all compounds presenting the qualifier (>) were excluded. All values were converted in μmol/bee and grouped by SMILES; geometric means for each value associated with the same SMILES were calculated. When values associated with the same SMILES showed a 3-fold difference between the maximum and the minimum, the relative compound was excluded from the regression modelling. When continuous data were available, these were transformed on the logarithmic scale. In addition, compounds excluded from the classification modelling were also excluded from the regression modelling. This is of particular relevance to abamectin and avermectin B1 which were both manually excluded because they are used as a mixture. The final dataset was built from 113 compounds.

2.1.2. Data splitting

Both datasets for classification-based and regression-based models were divided into a Training (TS) and Validation Set (VS) in a ratio of 80:20. The number of compounds in each set is shown in Table 2. In order to ensure a uniform distribution of the endpoint values and to have the widest possible chemical space in the two subsets, we applied an activity/structure sampling method. Pubchem fingerprints are calculated starting from SMILES. In case of regression-based models five equal-sized bins were created based on fixed ranges of experimental values. For classification-based models, only two groups have been considered (high toxicity – low toxicity). For each bin, a deterministic algorithm selected the 80% of compounds starting from a selected group of compounds (the first 5 compounds of the dataset) and looking for the most diverse molecules using as metric the Tanimoto similarity coefficient (Tanimoto, 1958) calculated on fingerprint. The picking algorithm is called MaxMin (Ashton et al., 2002). The resulting 80% of each bin was regrouped in TS and the remaining compounds constituted the VS (Worth et al., 2005; Golbraikh and Tropsha, 2002; Golbraikh et al., 2003).

2.1.3. Calculation of molecular descriptors

Dragon 7.0 was used for the calculation of 2 D molecular descriptors while stereoisomer information was removed. Moreover, descriptors with constant values (standard deviation 0) or correlated over 95% (Pearson correlation coefficient) with another descriptor (stronger correlation with the endpoint) were rejected. Centering and scaling as well as a range of methods of variable selection to fit the algorithm used for the model derivation were applied to all descriptors. Genetic algorithm has been used for Decision Trees (DT), k-nearest neighbors (k-NN), Multiple linear regression (MLR) and Partial least squares regression (PLS) (OECD, 2007), while VSURF (Genuer et al., 2015) has been used for random forest (RF) (Breiman, 2001). Genetic algorithm (OECD, 2007) has been applied with gaselect (Kepplinger et al., 2017) R package implementation, using a custom fitness function. The same user function (Underlying Algorithm for the derivation of the fitness function) is based on the same package of the algorithm used after the descriptor selection and it is the same of the following model derivation. In particular, a custom function has been implemented using cross-validation error as given in the output of DT R implementation and Cohen's Kappa (Cohen, 1960) between experimental and Cross validated predictions for k-NN.

Table 1
Criteria applied for the data pruning of toxicity data (expressed as LD_{50}) as reported in three different databases (OpenFoodTox, US-EPA ECOTOX, PPPDB).

Database	Species	Organism life stage	Exposure duration	Route	Dose unit (LD_{50})	Qualifier (tested chemical)	Chemical purity
OpenFoodTox	Honey bee	Adult	48 h	Dermal	μg/bee μg/piece ng/bee	"As such"	NA
US-EPA ECOTOX	*Apis mellifera* (with all subspecies)	Adult	48 h	Topical, Dermal	μg/piece μg/org μg/bee μg/g org ng/μl ng/org AI ng/org ppm ppb AI mg/org	NA	>80
PPDB	Honey bees (*Apis* spp.)	NA	48 h	Contact	μg/bee	NA	NA

NA = not available.

Table 2
Datasets splitting for classification- and regression-based models. The complete datasets are reported in Tables S1 and S2.

	Classification-based models	Regression-based models
Train	328	88
Test	83	25
Tot	411	113

2.1.4. Learning algorithms

In order to build classification-based models, DT, RF and k-NN were employed. With regard to regression models MLR and PLS were used. All the parameters are reported in Tables S3–S4 for classification- and regression-based models, respectively.

2.1.4.1. Multi linear regression. MLR is the most popular algorithm for QSAR development since it produces a transparent and an algorithm that is easily reproducible (OECD, 2007). MLR describes how a single response variable "Y" depends linearly on a number of predictor variables. A MLR can lead easily to overfitting, especially when dealing with a high number of predictors. In order to avoid overfitting, genetic algorithm has been applied using as fitness function Q^2. The algorithm used is "lm" as implemented in the package caret (Kuhn, 2008).

2.1.4.2. Partial least squares. PLS is a combination of MLR and principal component analysis (PCA). It performs a MLR using as predictors the principal components of the original data matrix. The algorithm used is "pls" as implemented in the package caret. In order to select the best parameters "Caret" hyper parameter tuning grid has been used. For pls implementation only the number of components is tuned.

2.1.4.3. K-nearest neighbor. The k-NN identifies a k number of neighbors for the target compound that will be used to provide a prediction of the endpoint. It is a transparent algorithm widely used for QSAR datasets with different similarity metrics to select neighbors (Manganaro et al., 2016). The algorithm uses a metric to measure distances between molecules, after pruning of molecular descriptors using Genetic Algorithm. The algorithm used is "KKNN" (Samworth, 2012) implemented in the package caret. In order to select the best parameters "Caret" hyper parameter tuning grid has been used. For k-NN implementation these parameters are tuned: the maximum number of neighbors (kmax), Parameter of Minkowski distance (distance) and the type of kernel estimate of the densities, used to weight the mean (kernel).

2.1.4.4. Decision trees. DTs (Quinlan, 1986, 1987) are flowchart-like structures formed by a series of nodes that generates a set of rules that follow a "IF Variable A is X THEN…" pattern. All the rules are hierarchically connected until a terminal node is reached, which assigns the class to the compound. The used algorithm is "RPART" as implemented in the package caret. In order to select the best parameters "Caret" hyper parameter tuning grid has been used. For RPART implementation only the complexity parameter (CP) is optimised, leading to a minimum improvement in the model needed at each node.

2.1.4.5. Random forest. After VSURF variable selection, a RF variant as implemented in the package "Ranger" (Wright and Ziegler, 2017) has been used for model derivation. In order to select the best parameters "Caret" (Kuhn, 2008), hyper parameter tuning grid has been used. Three parameters were tuned by grid search: mtry (number of randomly selected descriptors used in each tree of the RF), splitrule (the rule used to choose descriptors for a single tree, i.e. "gini" or "extratrees" for classification; "variance" or "extra trees" for regression), and min.node.size (minimal node size of trees). The number of trees was left as default value (500).

2.1.5. Statistical criteria

2.1.5.1. Classification-based models. In order to evaluate the performance of the classification-based models for two classes $LD_{50} < 100$ μg/bee (1) and $LD_{50} \geq 100$ (0), the following statistical criteria were: sensitivity, specificity, accuracy (BA) and Matthews correlation coefficient (MCC). Generally, the MCC coefficient is applied in machine learning to measure the quality of binary classifications, particularly when the classes present very different sizes (Dao et al., 2011).

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{1}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{2}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

$$MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{4}$$

TP, TN, FP and FN represent the number of true positives, true negatives, false positives and false negatives, respectively. MCC values range between −1 and +1, while the latter indicates a perfect prediction, a value of 0 indicate a prediction no better than a random one, and a value of −1 show total disagreement between predicted and observed values (Dao et al., 2011).

2.1.5.2. Regression-based models. The determination coefficient (R^2) is the fitness function used to evaluate the goodness of fit and is calculated as shown in Eq. (5).

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y}_i)^2} \tag{5}$$

where yi is the experimental value of the i-th chemical in the dataset; ŷi is the calculated value of the i-th query compound in the dataset for the determination of R^2; $\bar{y}_i$ is the mean of the experimental values of the compounds in the dataset, for all the N compounds. Similarly, RMSE (root mean square error) is an additional parameter used in the evaluation (Eq. (6)) which is calculated as follows:

$$RMSE = \sqrt{\sum \frac{(\hat{y}_i - y_i)^2}{N}} \tag{6}$$

The Cross-validated determination coefficient (Q^2) has been used for the calculation of statistics in cross-validation (Eq. (7)):

$$Q^2 = 1 - \frac{\sum \left(y_k - \acute{y}_k\right)^2}{\sum (y_k - \bar{y}_k)^2} \tag{7}$$

$y_k, \acute{y}_k, \bar{y}$ are observed, cross validated prediction and average values of the dependent variable, respectively (Golbraikh and Tropsha, 2002). For all the models the reported Q^2 is the mean value of the Q^2 of a 5-fold cross-validation repeated 3 times. Similarly, additional statistical parameters such as Q2-F1, Q2-F2, Q2-F3, CCC, r_0^2, r_m^2, $\overline{r^2}$, Δr_m^2, k and k′ are calculated according to Gramatica and Sangion (2016).

2.1.6. Applicability domain

The applicability domain (AD) of a QSAR model is defined as "the physico-chemical, structural, or biological space, knowledge or information on which the TS of the model has been developed, and for which it is applicable to make predictions for new compounds[…]. Ideally, the QSAR should only be used to make predictions within that domain by

interpolation not extrapolation" (Eriksson et al., 2003). The models are specifically designed to deal with pesticides. The model performance is taken into account without considering AD. However, since the set used for modelling contains all the molecules under 800 of molecular weight, this value should be considered as the upper limit to predict compounds in a reliable way. Moreover, the QSAR models here developed will be implemented in the open source platform VEGA-HUB (https://www.vegahub.eu/; Benfenati et al., 2017), therefore the reliability of the prediction will be evaluated using the Applicability Domain Index (ADI), which is an aggregated result taking into account several aspects:

1) Similar molecules with known experimental value and their accuracy (or average error) in their prediction,
2) Concordance among the target and similar molecules for the experimental data,
3) Atom Centered Fragments similarity check,
4) Descriptors noise sensitivity analysis,
5) Model descriptors range check.

As additional value, the information of the MoA of the closest neighbors will be provided in order to better assess the reliability of the prediction.

2.2. Mode of action

Several definitions of MoA are available from the literature, and scientific advisory bodies (WHO, 2009; Boobis et al., 2006; OECD, 2017) although the common goal is facilitating classification of chemicals according to mechanistic information (e.g. chemical class, molecular target) and through robust schemes (Verhaar et al., 1992; Enoch et al., 2008; Carriger et al., 2016; Kienzler et al., 2019). However, since here the authors refer exclusively to pesticides active substances used in PPPs, focus is given to classification schemes which are principally based on MoA as the target site of the substance (Sparks and Nauen, 2015; Casida, 2009; Wing et al., 2005). Hence, the main criterion for the refined classification of the MoA relies on the availability of the information describing the interaction with the receptor of the target species (e.g. sodium channel modulator, Acetylcholinesterase (AChE) inhibitors).

Qualitative information on the MoA of substances ($n = 113$) present in the regression-based QSAR model (see Section 2.1.5.2) were collected from different sources such as publicly available databases and the peer-reviewed scientific literature. Priority was given to well-defined schemes used for the classification of pesticides such as the one proposed by the Insecticide Resistance Action Committee (IRAC; Sparks and Nauen, 2015), Fungicide Resistance Action Committee (FRAC; Hermann and Stenzel, 2019) and Herbicide Resistance Action Committee (HRAC; Beffa et al., 2019). Similarly, MoA information was retrieved from the PPDB (available at https://sitem.herts.ac.uk/aeru/ppdb/en/index.htm). When no data were available on the MoA from the above mentioned sources, the publicly available scientific literature were investigated (Sanchez-Bayo, 2012; Simon-Delso et al., 2015; Johnson et al., 2012, 2013; Leroux et al., 2008; De Castro et al., 2015). In addition, a comparison of MoA nomenclatures (i.e. site of action) reported in the different databases/schemes was carried out taking as main reference the Resistance Action Committee classifications (e.g. IRAC, FRAC, HRAC) in order to provide users with a harmonised classification scheme. Pesticides active substances were classified according to their i) function (e.g. insecticide, fungicide, acaricides, herbicides, etc.), ii) chemical class (e.g. carbamates, pyrethroids, etc.) and iii) site of action (e.g. AChE inhibitors, sodium channel modulators). Finally, chemicals were grouped according to the harmonised MoA (i.e. site of action) to allow an assessment of potential variability in acute contact toxicity

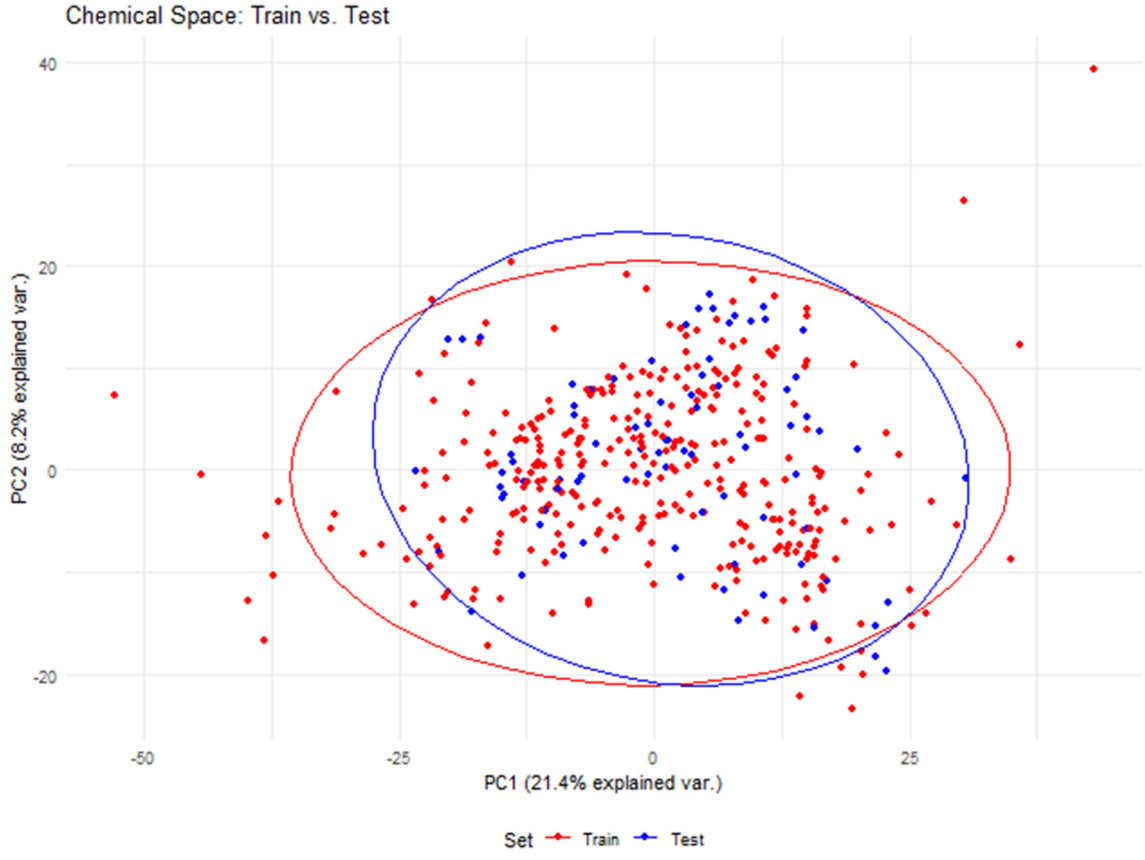


Fig. 1. Principal Component Analysis (PCA) on DRAGON descriptors for classification-based models (ellipsoid calculated at 0.95 probability).

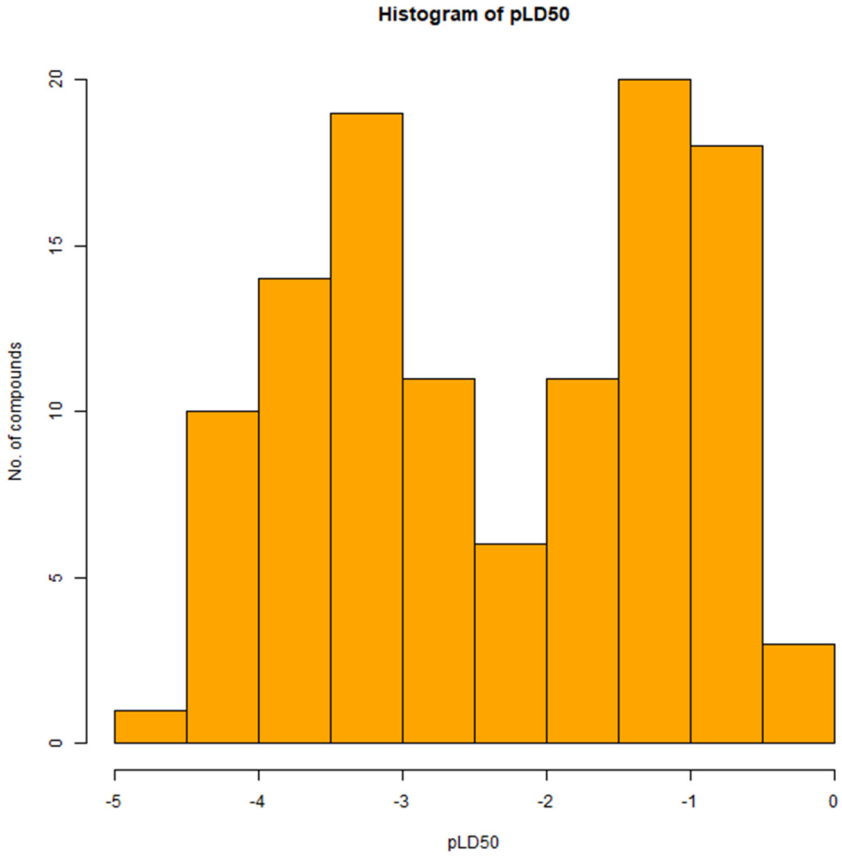


Fig. 2. Histogram of pLD_{50} (pLD_{50} μmol/bee) data used for regression-based model.

(potency expressed as $logLD_{50}$ μmol/bee) in honey bees across and within MoA groups. This exercise provides means for a refined prioritisation and grouping of chemicals for RA of combined exposure to multiple chemicals as recommended by EFSA MIXTOX Guidance, as well as supports to move towards an understanding of mechanisms of toxicity of active substances in PPPs (More et al., 2019).

3. Results and discussions

3.1. QSAR model development

3.1.1. Data collection and analysis

PCA has been performed in order to check if the TS covers the space of the test set (Fig. 1). An ellipsoid has been drawn for both sets in order to check the overlap. The test resulted structurally covered by the TS. As reported in the Materials and methods, 132 substances belong to the minor class (toxic) while 279 substances to the major one (non-toxic), having a ratio higher than 1:2. The dataset is slightly unbalanced towards the low toxicity substances (LD50 ≥ 100 μg/bee), nonetheless there is enough coverage of the minor class, making the dataset suitable for modelling. Similarly, he negative logarithm of LD_{50} (pLD_{50} μmol/bee) have been used for the quantitative modelling. The data distribution curve of pLD_{50} is plotted in Fig. 2.

3.1.2. Chemical descriptors selection

VSURF was used to select best descriptors as input for both classification and regression RF. Each RF used for the feature selection was constituted by 100 trees. Genetic Algorithm used for the other models (MLR, PLS, DT, k-NN) was set up to find the best number of descriptors between 5 and 12. Tables 3 and 4 report the selected descriptors selected for classification and regression models, respectively. Additionally, plots showing the most important variables for each model are reported in the supplementary material (Figs. S1–S2).

3.1.3. Classification-based models

In this study, three classification-based models were developed using different approaches as RF, DT and k-NN. For each model, the performance in TS, 5-Fold cross validation (CV) and VS was evaluated in order to identify the best model (Fig. S3). Results for statistical quality are shown in Table 5. All the models showed an acceptable sensitivity (>0.89), while the performance for RF and DT decreased when evaluating the specificity in CV (0.53 and 0.54, respectively). MCC for VS across the three models is above 0.75, thus showing that the models are able to identify both classes when predicting an external set of compounds (Table 5 and Fig. S3). However, our results show that k-NN is the most robust model when identifying minority class (<100 μg/bee) compounds. Additional Radar plots of the models are reported in

Table 3
Selected descriptors for random forest (RF), decision tree (DT) and k-nearest neighbors (k-NN) classification-based models.

RF	DT	KNN
X0Av	B09[C-O]	ICR
F02[C-P]	F02[C-P]	LOC
ATSC1e	SpMax_A	VE1_B(p)
MATS3v	ChiA_B(m)	MATS8e
SpMax1_Bh(s)	ATS1m	GATS3m
SpMAD_X	SpMax2_Bh(p)	Eta_sh_x
RBF	P_VSA_s_1	CATS2D_04_DA
B05[C-P]	–	CATS2D_00_LL
MATS1s	–	T(Cl..Cl)
–	–	SAdon

Table 4
Selected descriptors for random forest (RF), multi linear regression (MLR), partial least squares (PLS), decision tree (DT) and k-nearest neighbors (k-NN) regression-based models.

RF	MLR	PLS	DT	KNN
CATS2D_07_LL	CATS2D_07_LL	CATS2D_07_LL	piPC06	X4v
MATS8v	MATS4m	F06[N-S]	ChiA_B(p)	GGI10
X3v	GATS8s	MATS4m	MATS8v	SpMin2_Bh(s)
CATS2D_03_DL	GGI8	GATS4e	SpMax2_Bh (p)	Eig02_AEA (dm)
TI2_L	JGI4	GATS8s	P_VSA_m_2	SsOH
IVDE	CATS2D_00_DD	JGI8	SpMaxA_EA (ed)	NssO
SpMax2_Bh(p)	CATS2D_03_DA	JGT	Eig03_EA (bo)	CATS2D_02_DD
CATS2D_09_LL	CATS2D_03_DL	CATS2D_03_DL	Eig05_EA (dm)	CATS2D_07_LL
SaaN	F03[C-N]	F02[N-S]	F04[N-P]	F01[N-O]
SpPosA_B(i)	Psychotic-50	F04[O-O]	F04[O-S]	F05[C-S]
GATS5m	–	–	–	F06[N-S]
MATS4v	–	–	–	F07[C-N]
N-074	–	–	–	–

supplementary materials (Fig. S3), while the predictions and the descriptors are reported in Tables S5–S7.

In the literature, other classification-based QSAR models to predict pesticides toxicity in honey bees are available, and thus a comparison of their predictive performance is illustrated here (Table 6). Overall, the k-NN model reported here shows higher performance with regards to statistical quality of the results compared with those from existing k-NN models (Como et al., 2017; Venko et al., 2018). Similarly, the PNN-QSTR model (Singh et al., 2014) demonstrated a comparable performance compared with the current k-NN model, although based on a probabilistic neural network approach. Here, it is important to highlight that the two-category model presented here is the first classifier built based on the largest available dataset encompassing data for 411 pesticides active substances from three different open access sources (i.e. EFSA's OpenFoodTox, US-EPA ECOTOX and PPDB) and thus it has the advantage to increase the application domain of the model for a large range of pesticides.

3.1.4. Regression-based model

As for the classification-based models, we evaluated the performance in TS, CV and VS. Overall, all models provided good predictivity with regards to goodness of fit (TS) except for PLS and DT, which were associated with significantly R^2 lower than those from other models (Table 7). Statistical quality for DT model was not satisfactory for CV and VS, while RF had an R^2 close to 0.5 in CV. In terms of reproducibility and overall performance, our results show that the best model is represented by k-NN ($R^2 = 0.74$) followed by RF ($R^2 = 0.72$) (Table 7). Additional Scatter plots (predicted vs experimental) of all models here developed are reported in Fig. S4, while the predictions and the descriptors are reported in Tables S8–S12.

The best model developed here (k-NN) was used for comparative purposes with the published QSAR regression models predicting honey bee toxicity. Results are provided in Table 8 and show that is the partial within the current model is related to the VS and the characterisation of the data points. In contrast to Hamadache et al. (2018), our k-NN model is highly curated, and all the data above 100 μg/bee (limit test) were filtered out (OECD, 1998). In fact, these data points, when included in a model, behave as attractors of the regression giving an overestimation of optimistic performances. Hence, after filtering those compounds, the predictivity of the refined QSAR model increased significantly, thus providing a tool to predict honey bee toxicity for substances of concern (i.e. moderate/high toxicity) with high precision. For this reason, the classification- and regression-based QSAR models developed here are not meant to be used as two distinct models but rather as an integrative tool following a specific hierarchical workflow (Fig. 3).

Table 5
Results of the statistical quality for random forest (RF), decision tree (DT) and k-nearest neighbor (k-NN) classification-based models. Test set (TS), cross validation set (CV) and validation set (VS) are reported.

Algorithm	RF			DT			K-NN		
Set	TS	CV	VS	TS	CV	VS	TS	CV	VS
Sensitivity	1.00	0.94	0.96	0.95	0.92	0.89	0.77	0.90	0.93
Specificity	1.00	0.54	0.74	0.58	0.53	0.89	0.96	0.67	0.85
Accuracy	1.00	0.81	0.89	0.83	0.80	0.89	0.90	0.83	0.90
MCC	1.00	0.54	0.75	0.59	0.50	0.76	0.76	0.59	0.78

Table 6
Comparison of the classification-based model (k-NN; best model) developed here with publicly available two-category models for acute toxicity towards honey bees (threshold 100 μg/bee). Results of the statistical quality are reported for training (TS) and validation sets (VS) in each model.

Reference	Set	Compounds (n)	Sensitivity	Specificity	Accuracy
Venko et al., 2018	TS	205	0.88	0.90	0.89
	VS	49	0.75	0.79	0.78
Como et al., 2017	TS	192	0.60	0.88	0.76
	VS	50	0.80	0.86	0.84
Singh et al., 2014	TS	175	1.00	1.00	1.00
	VS	62	0.86	1.00	0.87
Present study (k-NN model)	TS	328	0.77	0.96	0.90
	VS	83	0.93	0.85	0.90

3.1.5. Integrative honey bee QSAR model

As illustrated in the above paragraph, the current study aimed to develop the first integrative honey bee QSAR model to allow predicting the toxicity of unknown compounds following a specific hierarchical workflow (Fig. 3). Here, 12 compounds were tested as independent datasets compared to the training sets of both classification- and regression-based models for the evaluation of the predictivity of the models and the potential application of the integrative tool. Results show that for the majority of compounds ($n = 10$), the quantitative prediction provides satisfactory predictions compared to those from the corresponding experimental data (Table 9) with the exception of two compounds (momfluorothrin, CAS n. 609346-29-4; chloroxuron, CAS n. 1982-47-4) for which the predictions from the classification-based model underestimated the toxicity resulting in an incorrect classification (Table 9).

An example of output from the integrative honey bee QSAR model is shown in Table 9 which reports CAS number, experimental MoA, predicted toxicity class (1 = toxic/0 = non-toxic; threshold 100 μg/bee) and quantitative prediction (experimental LD_{50} μg/org) for the target chemical as well as the five most similar compounds identified by the k-NN model for each target. Hence, one of the novelty of the tool is to provide users with toxicity predictions (LD_{50} μg/org) as well as experimental MoA and LD50 for the five most similar compounds identified by the k-NN model.

3.2. MoA assessment

3.2.1. Harmonised MoA classification scheme

A total of 113 substances were included in the analysis of the MoA nomenclatures in order to provide an harmonised classification of pesticides active substances (used in PPPs) according to i) function (e.g. insecticide, fungicide, etc.), ii) chemical class (e.g. carbamates, organophosphate, etc.), and iii) site of action (e.g. sodium channel modulators) (Table S13).

According to the above-mentioned criteria, an example of different nomenclatures applied in MoA classification for insecticides and

Table 7
Statistical robustness and performance for random forest (RF), multi linear regression (MLR), partial least squares (PLS), decision tree (DT) and k-nearest neighbors (k-NN) regression-based models. Test Set (TS), Cross Validation set (CV) and Validation Set (VS) are reported. Statistical parameters are reported and defined according to Gramatica and Sangion (2016).

Parameter	RF	MLR	PLS	DT	KNN	Acceptability criteria
RMSE - TS	0.41	0.66	0.76	0.71	0.39	
r^{2} - TS	0.88	0.7	0.61	0.66	0.9	0.6
MAE - TS	0.34	0.56	0.61	0.53	0.3	
CCC - TS	0.93	0.82	0.76	0.79	0.94	
$Q^2_{5\text{-fold CV}}$	0.49	0.59	0.53	0.35	0.63	0.5
RMSE - VS	0.8	0.93	0.9	1.01	0.71	
r^2 - VS	0.72	0.55	0.59	0.46	0.74	0.6
MAE - VS	0.63	0.76	0.66	0.79	0.52	
Q2-F1	0.65	0.53	0.56	0.44	0.72	
Q2-F2	0.65	0.52	0.56	0.44	0.72	
Q2-F3	0.57	0.41	0.45	0.31	0.65	
CCC - VS	0.75	0.73	0.71	0.66	0.83	
r^2_0 - VS	0.65	0.53	0.58	0.46	0.73	
r^2_m - VS	0.53	0.48	0.52	0.43	0.66	0.5
$\overline{r^2}$ - VS	0.35	0.42	0.35	0.31	0.54	0.5
Δr^2_m - VS	0.36	0.11	0.34	0.23	0.23	<0.3
k - VS	0.98	0.96	0.93	0.94	0.96	$0.85 < k < 1.15$
k′ - VS	0.93	0.91	0.95	0.91	0.97	$0.85 < k' < 1.15$

fungicides is provided in Tables 10 and 11, respectively. Qualitative information on MoA were extracted from three different sources: PPDB, Resistance Action Committee classifications (i.e. IRAC, FRAC, HRAC), and the peer reviewed scientific literature. Our results suggest that PPDB provides users with additional general and toxicological information on a number of substances (e.g. systemic/non-systemic, authorisation status), although the harmonisation of MoA has not been fully explored yet. Similarly, studies from the peer reviewed scientific literature have often applied a range of criteria (chemical class vs target site) when classifying chemicals, and did not prove univocal MoA nomenclature and classification schemes (Sanchez-Bayo, 2012; Johnson et al., 2012, 2013; Simon-Delso et al., 2015). In contrast, IRAC, FRAC and HRAC provide structured schemes, classifying pesticide active substances according to their target site (and cross-resistance), thus encompassing 32 insecticides, 56 fungicides and 26 herbicides MoAs, respectively (Sparks and Nauen, 2015; Hermann and Stenzel, 2019; Beffa et al., 2019). Therefore, although the different classification schemes analysed here have not been developed for the same purposes, the authors acknowledge that Resistance Action Committee schemes (IRAC, FRAC, HRAC) provide users with a more systematic and sound classification of MoAs (i.e. target site) for pesticide active substances used in PPPs.

A similar research effort to harmonise MoA classification schemes has been proposed by Kienzler et al. (2017, 2019). However, authors focused only on aquatic toxicity (e.g. fish) while taking into account several different definitions of MoA having degree of specificity based on fish behavioural responses, toxicological responses or weight of evidence classification. Therefore, this manuscript provides the first harmonised MoA classification for terrestrial non-target species and has been applied to honey bees, taking into account existing knowledge on the specific target site of pesticide active substances.

Overall, such MoA harmonised classification is valuable for the development and the testing of the validity of QSAR model and other in silico tools, thereby contributing to a sound mechanistic interpretation of the model (OECD, 2007). A specific application of this MoA analysis is in the refinement of Applicability Domain Index (ADI) definitions for k-NN models in the open source VEGA platform (see Section 2.1.6). In a similar fashion, k-NN QSAR models can be used to identify similar compounds "neighbors" and their structural features responsible for the toxicological mechanism(s) of the active substance (as illustrated in Sections 3.1.3 and 3.1.4), thus allowing prediction of potential target sites in terrestrial organisms such as earthworms and honey bees (Roy et al., 2020; Ghosh et al., 2020).

It should be noted that the harmonised classification proposed here (Table S13) has been carried out for 113 chemicals for which data were applied for the development of the regression-based QSAR model (Section 3.1.4), thereby the full list of substances provided by the Resistance Action Committee classifications was not included (i.e. IRAC, FRAC, HRAC). However, although a smaller number of chemicals was included in the MoA analysis, the substances underwent data curation through a structured workflow in order to avoid ambiguous structures, and thus providing high-quality and curated datasets (Gadaleta et al., 2018).

Table 8
Comparison of regression-based model here developed with publicly available models for acute toxicity towards honey bees. Determination coefficient (R^2) and root-mean-square error (RMSE) are reported.

Reference	Set	Compounds (n)	R^2	RMSE
Devillers et al., 2002	TS	86	0.82	0.430
	VS	11	0.94	0.39
Toropov and Benfenati, 2007	TS	85	0.68	0.82
	VS	20	0.72	0.68
Dulin et al., 2012	TS	39	0.81	0.350
	VS	6	0.85	0.218
Singh et al., 2014	TS	190	0.85	0.50
	VS	47	0.86	0.33
Hamadache et al., 2018	TS	95	0.98	0.36
	VS	16	0.96	0.71
Present study	TS	88	0.90	0.39
	VS	25	0.74	0.71

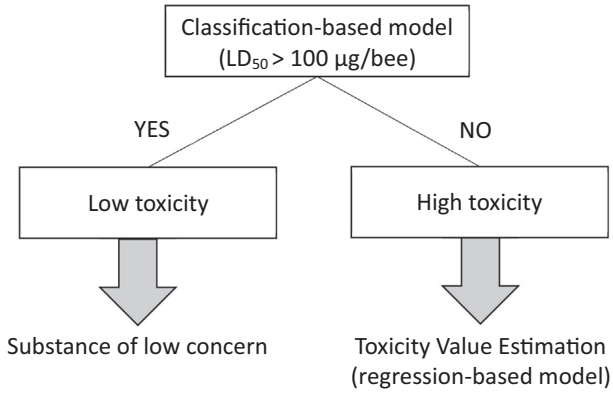


Fig. 3. Hierarchical workflow to apply classification- and regression-based models for hazard assessment of pesticides active substances in honey bees.

Table 9
Results of the validation of the integrative honey bee QSAR model for 12 target chemicals that are independent from the training sets of both classification- and regression-based models. The output illustrates the CAS number, experimental MoA, predicted toxicity class (1 = toxic/0 = non-toxic; threshold 100 μg/bee), experimental LD_{50} (μg/org), and quantitative prediction (pred LD_{50} μg/org) of the target substance. In addition, the table provides the five most similar compounds identified through the k-NN model for each target chemical.

Target	*Chemical#1*	*Chemical#2*	*Chemical#3*	*Chemical#4*	*Chemical#5*
Momfluorothrin **CAS: 609346-29-4** **MoA: Sodium channel modulators_Na channel(+)** **Pred Class: 0** **Exp (μg/org): 0.2** **Pred (μg/org): 0.22**	4-(2,4-Dichlorophenoxy)butanoic acid CAS: 94-82-6 MoA: Synthetic auxins (action like indole acetic acid) Exp (μg/org): 14.5	Chlorpropham CAS: 101-21-3 MoA: Inhibition of mitosis/microtubule organization Exp (μg/org): 96.1	2,2-Dimethyl-1,3-benzodioxol-4-ol 4-(N-methylcarbamate) CAS: 22781-23-3 MoA: Acetylcholinesterase (AChE) inhibitors_AChE(-) Exp (μg/org): 0.43	methomyl CAS: 16752-77-5 MoA: Acetylcholinesterase (AChE) inhibitors_AChE(-) Exp (μg/org): 0.16	alachlor CAS: 15972-60-8 MoA: Inhibition of very-long-chain fatty acid synthesis (VLCFAs) Exp (μg/org): 16.0
Terbufos **CAS: 13071-79-9** **MoA: Acetylcholinesterase (AChE) inhibitors_AChE(-)** **Pred Class: 1** **Exp (μg/org): 4.1** **Pred (μg/org): 2.79**	Tefluthrin CAS: 79538-32-2 MoA: Sodium channel modulators_Na channel(+) Exp (μg/org): 0.28	Resmethrin CAS: 10453-86-8 MoA: Sodium channel modulators_Na channel(+) Exp (μg/org): 0.06	Imiprothrin CAS: 72963-72-5 MoA: Sodium channel modulators_Na channel(+) Exp (μg/org): 0.4	5-Amino-1-[2,6-dichloro-4-(trifluoromethyl)phenyl]-4-[(trifluoromethyl)sulfinyl]-1H-pyrazole-3-carbonitrile CAS: 120068-37-3 MoA: GABA-	(1,3,4,5,6,7-Hexahydro-1,3-dioxo-2H-isoindol-2-yl)methyl ester 2,2-dimethyl-3-(2-methyl-1-propenyl)cyclopropane carboxylic acid CAS: 7696-12-0 MoA: Sodium channel modulators_Na channel(+)

				gated chloride channel blockers_GABA -R(-) Exp (μg/org): 0.01	Exp (μg/org): 0.16
Chloroxuron **CAS: 1982-47-4** **MoA: Inhibition of photosynthesis at PS II** **Pred Class: 0** **Exp (μg/org): 16.0** **Pred (μg/org): 19.88**	Alachlor CAS: 15972-60-8 MoA: Inhibition of very-long-chain fatty acid synthesis (VLCFAs) Exp (μg/org): 16.0	Methiocarb CAS: 2032-65-7 MoA: Acetylcholinesterase (AChE) inhibitors_AChE(-) Exp (μg/org): 0.29	Chlorpropham CAS: 101-21-3 MoA: Inhibition of mitosis/microtubule organization Exp (μg/org): 96.1	4-Bromo-2-(4-chlorophenyl)-1-(ethoxymethyl)-5-(trifluoromethyl)-1H-pyrrole-3-carbonitrile CAS: 122453-73-0 MoA: Uncoupler of oxidative phosphorylation Exp (μg/org): 0.12	2,2-Dimethyl-1,3-benzodioxol-4-ol 4-(N-methylcarbamate) CAS: 22781-23-3 MoA: Acetylcholinesterase (AChE) inhibitors_AChE(-) Exp (μg/org): 0.43

3.2.2. Pesticide active substances toxicity in relation to their MoA

While focus is given to the application of MoA for the grouping of multiple chemicals for combined exposure RA using component-based approaches in the ecological area, the honey bee example provides an application for non-target terrestrial organisms as illustrated in the EFSA MIXTOX guidance (More et al., 2019).

Pyrethrins (cinerin II)	Methomyl	Thiodicarb	Chlorbromuron	Chlorpropham	2,2-Dimethyl-1,3-
CAS: 121-20-0					benzodioxol-4-ol 4-(N-
MoA: Sodium channel	CAS: 16752-77-5	CAS: 59669-26-0	CAS: 13360-45-7	CAS: 101-21-3	methylcarbamate)
modulators_Na					
channel(+)	MoA:	MoA:	MoA: Inhibition of	MoA: Inhibition	CAS: 22781-23-3
Pred Class: 1	Acetylcholinesteras	Acetylcholinesteras	photosynthesis at	of	
Exp (μg/org): 0.01	e (AChE)	e (AChE)	PS II	mitosis/microtu	MoA:
Pred (μg/org): 0.02	inhibitors_AChE(-)	inhibitors_AChE(-)		bule	Acetylcholinesterase
			Exp (μg/org): 16.0	organization	(AChE)
	Exp (μg/org): 0.16	Exp (μg/org): 3.1			inhibitors_AChE(-)
				Exp (μg/org):	
				96.1	Exp (μg/org): 0.43
Pyrethrins (jasmolin	Methomyl	Thiodicarb	Chlorbromuron	Chlorpropham	2,2-Dimethyl-1,3-
II)					benzodioxol-4-ol 4-(N-
CAS: 1172-63-0	CAS: 16752-77-5	CAS: 59669-26-0	CAS: 13360-45-7	CAS: 101-21-3	methylcarbamate)
MoA: Sodium channel					
modulators_Na	MoA:	MoA:	MoA: Inhibition of	MoA: Inhibition	CAS: 22781-23-3
channel(+)	Acetylcholinesteras	Acetylcholinesteras	photosynthesis at	of	
Pred Class: 1	e (AChE)	e (AChE)	PS II	mitosis/microtu	MoA:
Exp (μg/org): 0.01	inhibitors_AChE(-)	inhibitors_AChE(-)		bule	Acetylcholinesterase
Pred (μg/org): 0.02			Exp (μg/org): 16.0	organization	(AChE)
	Exp (μg/org): 0.16	Exp (μg/org): 3.1			inhibitors_AChE(-)
				Exp (μg/org):	
				96.1	Exp (μg/org): 0.43
Pyrethrins (cinerin I)	Methomyl	Thiodicarb	Chlorbromuron	Chlorpropham	2,2-Dimethyl-1,3-
CAS: 25402-06-6					benzodioxol-4-ol 4-(N-
MoA: Sodium channel	CAS: 16752-77-5	CAS: 59669-26-0	CAS: 13360-45-7	CAS: 101-21-3	methylcarbamate)
modulators_Na					
channel(+)	MoA:	MoA:	MoA: Inhibition of	MoA: Inhibition	CAS: 22781-23-3

Pred Class: 1	Acetylcholinesteras	Acetylcholinesteras	Photosynthesis at	of	
Exp (µg/org): 0.01	e (AChE)	e (AChE)	PS II	mitosis/microtu	MoA:
Pred (µg/org): 0.07	inhibitors_AChE(-)	inhibitors_AChE(-)		bule	Acetylcholinesterase
			Exp (µg/org): 16.0	organization	(AChE)
	Exp (µg/org): 0.16	Exp (µg/org): 3.1			inhibitors_AChE(-)
				Exp (µg/org):	
				96.1	Exp (µg/org): 0.43
Pyrethrins (pyrethrin	Methomyl	Thiodicarb	Chlorbromuron	Chlorpropham	2,2-Dimethyl-1,3-
I)					benzodioxol-4-ol 4-(N-
CAS: 121-21-1	CAS: 16752-77-5	CAS: 59669-26-0	CAS: 13360-45-7	CAS: 101-21-3	methylcarbamate)
MoA: Sodium channel					
modulators_Na	MoA:	MoA:	MoA: Inhibition of	MoA: Inhibition	CAS: 22781-23-3
channel(+)	Acetylcholinesteras	Acetylcholinesteras	photosynthesis at	of	
Pred Class: 1	e (AChE)	e (AChE)	PS II	mitosis/microtu	MoA:
Exp (µg/org): 0.01	inhibitors_AChE(-)	inhibitors_AChE(-)		bule	Acetylcholinesterase
Pred (µg/org): 0.06			Exp (µg/org): 16.0	organization	(AChE)
	Exp (µg/org): 0.16	Exp (µg/org): 3.1			inhibitors_AChE(-)
				Exp (µg/org):	
				96.1	Exp (µg/org): 0.43
2,2-Dimethyl-3-(2-	Methomyl	Thiodicarb	Chlorbromuron	Oryzalin	Chlorpropham
methyl-1-					
propenyl)cyclopropan	CAS: 16752-77-5	CAS: 59669-26-0	CAS: 13360-45-7	CAS: 19044-	CAS: 101-21-3
ecarboxylic acid (3-				88-3	
phenoxyphenyl)methy	MoA:	MoA:	MoA: Inhibition of		MoA: Inhibition of
l ester	Acetylcholinesteras	Acetylcholinesteras	photosynthesis at	MoA: Inhibition	mitosis/microtubule
CAS: 26002-80-2	e (AChE)	e (AChE)	PS II	of microtubule	organization
MoA: Sodium channel	inhibitors_AChE(-)	inhibitors_AChE(-)		assembly	
modulators_Na			Exp (µg/org): 16.0		Exp (µg/org): 96.1
channel(+)	Exp (µg/org): 0.16	Exp (µg/org): 3.1		Exp (µg/org):	

Pred Class: 1 **Exp (μg/org): 0.07** **Pred (μg/org): 0.02**				40.8	
Carbosulfan| **Carbosulfan** **CAS: 55285-14-8** **MoA:** **Acetylcholinesterase** **(AChE)** **inhibitors_AChE(-)** **Pred Class: 1** **Exp (μg/org): 0.18** **Pred (μg/org): 0.26**	2-Methyl-2-(methylthio)propanol O-[(methylamino)carbonyl]oxime CAS: 116-06-3 MoA: Acetylcholinesterase (AChE) inhibitors_AChE(-) Exp (μg/org): 0.29	Thiofanox CAS: 39196-18-4 MoA: Acetylcholinesterase (AChE) inhibitors_AChE(-) Exp (μg/org): 0.06	Methomyl CAS: 16752-77-5 MoA: Acetylcholinesterase (AChE) inhibitors_AChE(-) Exp (μg/org): 0.16	2,2-Dimethyl-1,3-benzodioxol-4-ol 4-(N-methylcarbamate) CAS: 22781-23-3 MoA: Acetylcholinesterase (AChE) inhibitors_AChE(-) Exp (μg/org): 0.43	Imiprothrin CAS: 72963-72-5 MoA: Sodium channel modulators_Na channel(+) Exp (μg/org): 0.4
Ethoprophos **CAS: 13194-48-4** **MoA:** **Acetylcholinesterase** **(AChE)** **inhibitors_AChE(-)** **Pred Class: 1**	Tefluthrin CAS: 79538-32-2 MoA: Sodium channel modulators_Na	Imiprothrin CAS: 72963-72-5 MoA: Sodium channel modulators_Na	Resmethrin CAS: 10453-86-8 MoA: Sodium channel modulators_Na	5-Amino-1-[2,6-dichloro-4-(trifluoromethyl)phenyl]-4-[(trifluoromethyl)sulfinyl]-1H-	(1,3,4,5,6,7-Hexahydro-1,3-dioxo-2H-isoindol-2-yl)methyl ester 2,2-dimethyl-3-(2-methyl-1-propenyl)cyclopropane carboxylic acid

Exp (μg/org): 5.56	channel(+)	channel(+)	channel(+)	pyrazole-3-	
Pred (μg/org): 0.56				carbonitrile	CAS: 7696-12-0
	Exp (μg/org): 0.28	Exp (μg/org): 0.4	Exp (μg/org): 0.06		
				CAS: 120068-	MoA: Sodium channel
				37-3	modulators_Na
					channel(+)
				MoA: GABA-	
				gated chloride	Exp (μg/org): 0.16
				channel	
				blockers_GABA	
				-R(-)	
				Exp (μg/org):	
				0.01	
Fenitrothion | Phosphorothioic acid O,O-dimethyl O-(3-methyl-4-nitrophenyl)ester	N,N-Bis(2-methylpropyl)carbamothioic acid S-ethyl ester	Triazamate	Chlorpyrifos-methyl	6-Methyl-1,3-dithiolo[4,5-b]quinoxalin-2-one	Phosphoramidothioic acid, O,S-Dimethyl ester
CAS: 122-14-5	CAS: 2008-41-5	CAS: 112143-82-5	CAS: 5598-13-0		CAS: 10265-92-6
MoA: Acetylcholinesterase (AChE) inhibitors_AChE(-)		MoA: Acetylcholinesterase (AChE) inhibitors_AChE(-)	MoA: Acetylcholinesterase (AChE) inhibitors_AChE(-)	CAS: 2439-01-2	MoA: Acetylcholinesterase (AChE) inhibitors_AChE(-)
Pred Class: 1	MoA: Inhibition of lipid synthesis – not ACCase			MoA: N/A	
Exp (μg/org): 0.25		Exp (μg/org): 27.0	Exp (μg/org): 0.15		
Pred (μg/org): 3.0	Exp (μg/org): 29.0			Exp (μg/org): 66.47	Exp (μg/org): 1.37

Following the results of the critical analysis of MoA nomenclatures (Section 3.2.1), 17 classes of MoAs have been defined for pesticide active substances ($n = 113$) according their specific target site (Sparks and Nauen, 2015; Hermann and Stenzel, 2019; Beffa et al., 2019) (Fig. 4 and Table S13). Results show that 38% of chemicals were classified as insecticide/acaricides "*Acetylcholinesterase (AChE) inhibitors*", 18% as

Phosphorothioic acid, O,O-Diethyl-O-(4-nitrophenyl)ester CAS: 56-38-2 MoA: Acetylcholinesterase (AChE) inhibitors_AChE(-) Pred Class: 1 Exp (μg/org): 0.18 Pred (μg/org): 0.19	triazamate CAS: 112143-82-5 MoA: Acetylcholinesterase (AChE) inhibitors_AChE(-) Exp (μg/org): 27.0	N,N-Bis(2-methylpropyl)carbamothioic acid S-ethyl ester CAS: 2008-41-5 MoA: Inhibition of lipid synthesis – not ACCase Exp (μg/org): 29.0	6-Methyl-1,3-dithiolo[4,5-b]quinoxalin-2-one CAS: 2439-01-2 MoA: N/A Exp (μg/org): 66.47	methomyl CAS: 16752-77-5 MoA: Acetylcholinesterase (AChE) inhibitors_AChE(-) Exp (μg/org): 0.16	chlorpyrifos-methyl CAS: 5598-13-0 MoA: Acetylcholinesterase (AChE) inhibitors_AChE(-) Exp (μg/org): 0.15

insecticide/acaricide "*Sodium channel modulators*", 6% as herbicides "*Inhibitor of photosynthesis at PS II*", 4% as insecticides "*Nicotinic acetylcholine receptor (nAChR) competitive modulators*", 4% as acaricide "*Mitochondrial complex I-II electron transport inhibitors*" and 4% as fungicides "*Sterol Biosynthesis Inhibiting (SBI) class I-II (erg11/cyp51)*" (Fig. 4).

Overall, the method for grouping chemicals into assessment groups (AGs) proposed here has been carried out for chemicals (n = 113) used in the development of the regression-based QSAR model, thus including substances with moderate/high acute (contact) toxicity in honey bees (LD_{50} < 100 μg/bee) (Fig. 5 and Table S13).

3.2.2.1. Insecticides/acaricides. According to our results, pyrethroid/pyrethrin insecticides and/or acaricides belonging to the MoA group of "sodium channel modulators", showed the highest acute contact toxicity in honey bees (LD_{50} = 0.013–23.57 μg/bee), among which 66% with LD_{50} = 0.013–0.10 μg/bee (Fig. 5 and Table S13). However, the most potent toxic insecticide was by far fipronil (CAS. 120068-37-3) which reported LD_{50} = 0.00389–0.00593 μg/bee (EFSA, 2006) and it is classified as a GABA-gated chloride channel blocker (Sanchez-Bayo, 2012). Similarly, insecticides/acaricides within the MoA group "Acetylcholinesterase (AChE) inhibitors", showed high (contact) toxicity in honey bees (LD_{50} = 0.0049–59.8 μg/bee), 69% of which presenting LD_{50} < 1 μg/bee (Fig. 5 and Table S13). Interestingly, all AChE inhibitors insecticides/acaricides were classified as carbamate or organophosphate substances (Table S13). By definition, insecticides and acaricides are generally harmful to non-target terrestrial organisms such as bees (Douglas et al., 2020; Sanchez-Bayo and Goka, 2016). However, differences in insecticide toxicity within the same MoA appear to be driven by their structural features and reactivity, as demonstrated for cyano-substituted neonicotinoids (e.g. thiacloprid and acetamiprid) which are three orders of magnitude less toxic to honey bees compared with other compounds of the same MoA group (Iwasa et al., 2004; Sanchez-Bayo, 2012; Carnesecchi et al., 2019a, 2019b).

3.2.2.2. Herbicides and fungicides. The majority of triazine- and urea-derived herbicides (acting as "inhibitors of photosynthesis at PS II")

Table 10
Example of different MoA nomenclatures used for the classification of substances according to different databases/schemes (PPDB; IRAC (Sparks and Nauen, 2015); Sanchez-Bayo, 2012; Johnson et al., 2012, 2013).

Substance name	CAS n.	MoA/site of action			
		PPDB	IRAC	Others (Sanchez-Bayo, 2012; Johnson et al., 2012, 2013)	Harmonised classification
Triazamate	112143-82-5	Systemic with contact and stomach action.	Acetylcholinesterase (AChE) inhibitors Nerve action (Strong evidence that action at this protein is responsible for insecticidal effects)	Neurotoxic AChE(−)	Acetylcholinesterase inhibitors [AChE(−)]
Thiodicarb	59669-26-0	Mainly stomach action but some contact effects. Cholinesterase inhibitor.			
Chlorpyrifos	2921-88-2	Non-systemic with contact, inhalation and stomach action. Acetylcholinesterase (AChE) inhibitor.			

Table 11
Example of different MoA nomenclatures used for the classification of fungicides according to different databases/schemes (PPDB; FRAC (Hermann and Stenzel, 2019)).

Substance name	CAS n.	MoA/site of action		
		PPDB	FRAC	Harmonised classification
Tetraconazole	112281-77-3	Systemic with protectant, eradicant and curative properties. Sterol biosynthesis inhibitor, acts mainly on the vegetative stages of fungi by blocking the mycelial growth either inside or on the surface of the host plant.	DMI-fungicides (DeMethylation Inhibitors) (SBI: Class I). TARGET: C14-demethylase in sterol biosynthesis (erg11/cyp51)	SBI: Class I_C14-demethylase in sterol biosynthesis (erg11/cyp51)
Spiroxamine	118134-30-8	Systemic with protective, curative and eradicative action. Disrupts membrane function. Inhibits sterol biosynthesis in membranes.	Amines ("morpholines") (SBI: Class II). TARGET: Δ14-reductase and Δ8 to Δ7-isomerase in sterol biosynthesis (erg24, erg2)	SBI: Class II_Δ14-reductase and Δ8 to Δ7-isomerase in sterol biosynthesis (erg24, erg2)

presented $LD_{50s} \leq 20$ μg/bee (Fig. 5 and Table S13). However, such active substances have nor to date been authorised on the EU market (EU Pesticide Database, 2019). Interestingly, oryzalin (dinitroaniline herbicide acting as inhibitor of microtubule assembly in weeds; CAS n. 19044-88-3), is currently authorised in some EU countries (ES, FR, IT, PT) and has its high toxicity in honey bees with $LD_{50} = 40.8$ μg/bee (Table S13). Oryzalin also has $LD_{50} = 32$ μg/bee following oral exposure to honey bees (EFSA, 2010). Similarly, sulfonylurea herbicides such as nicosulfuron (CAS n. 111991-09-4) acting as inhibitor of acetolactate synthase (currently authorised in several EU member states) exhibits moderate contact toxicity towards honey bees ($LD_{50} = 76$ μg/bee) (EFSA, 2008). In contrast, Sterol Biosynthesis Inhibiting (SBI class I) fungicides such as triazoles and pyrimidine have acute contact LD_{50} in honey bees ranging from 20 to 69 μg/bee with spiroxamine (SBI class II; CAS n. 118134-30-8) as the most potent active substance ($LD_{50} = 4.22$ μg/bee).

Overall, our results suggest that, although the majority of active substances (60%) were classified as insecticides and/or acaricides with MoA groups including AChE inhibitors, sodium channel modulators and nAChR competitive modulators, some herbicides and fungicides (authorised in the EU) exhibit moderate to high toxicity ($LD_{50} =$ 0.8–69 μg/bee; Fig. 5) in honey bees. Similarly, high variability in the acute contact toxicity of PPPs in honey bees is often reported within the same MoA as shown for insecticides classified under the MoA AChE inhibitors, "GABA-gated chloride channel blockers", acaricides as mitochondrial complex I electron transport inhibitors (METI) (Fig. 5). However, looking at the size of the database, more toxicity data points would be needed to allow a more robust statistical analysis of LD_{50} variability in honey bees.

3.2.3. MoA as tool for grouping chemicals into assessment groups for ecological risk assessment of multiple chemicals

Notwithstanding that several definitions of MoA have been proposed, a range of applications of MoA schemes have also been reported for chemical RA. In the human health area, scientific advisory bodies such as the US-EPA, the WHO, the OECD and EFSA have proposed the application of MoA when defining AGs in component-based approaches for the identification and characterisation of combined toxicity of pesticide active substances as for example potential neurotoxicity in the thyroid to set cumulative AGs (Meek et al., 2011; EFSA PPR, 2013; OECD, 2018; More et al., 2019). Similarly, MoA can be applied as a grouping tool for ecological and human-health RA to group chemicals when applying mathematical models such as concentration addition or independent action for predicting mixtures toxicity (Kienzler et al., 2016; More et al., 2019; Carnesecchi et al., 2019a). Furthermore, MoA has been used to investigate potential pesticide resistance in target organisms (i.e. field population), thus representing a key-aspect of pest management worldwide (Sparks and Nauen, 2015; Hermann and Stenzel, 2019).

In the present manuscript, we provided the first MoA harmonised classification for pesticide active substances used in PPPs. However, a broad question remains: can one use MoA information to group chemicals in the broad context of ecological RA of multiple chemicals? In order to answer this question, we should clarify that two fundamental aspects intrinsically characterise a given pesticide active substance: its toxicity and MoA at target site (i.e. site of action). As in the case of insecticides, their toxicity and specificity are a consequence of the MoA at the cellular or physiological level in the organism (Simon-Delso et al.,

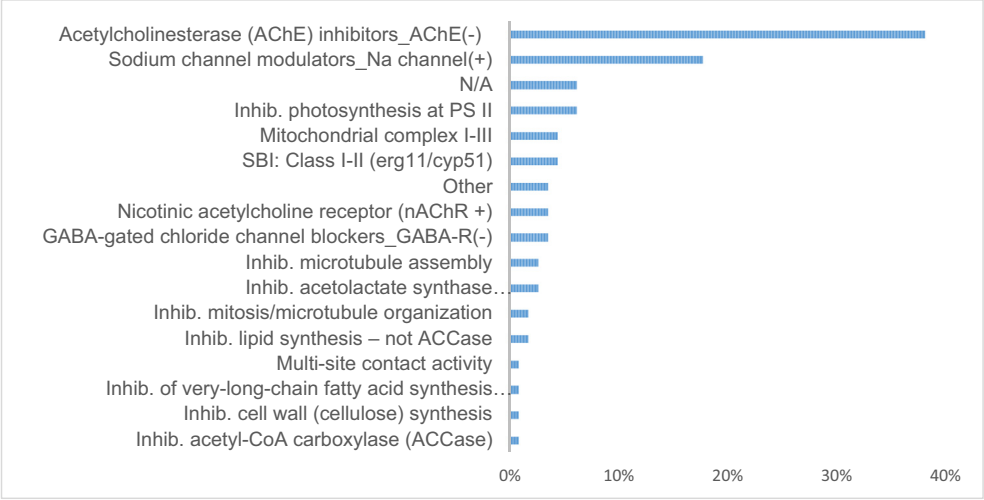


Fig. 4. Percentage of pesticides active substances ($n = 113$) that are classified into each harmonised MoA group.

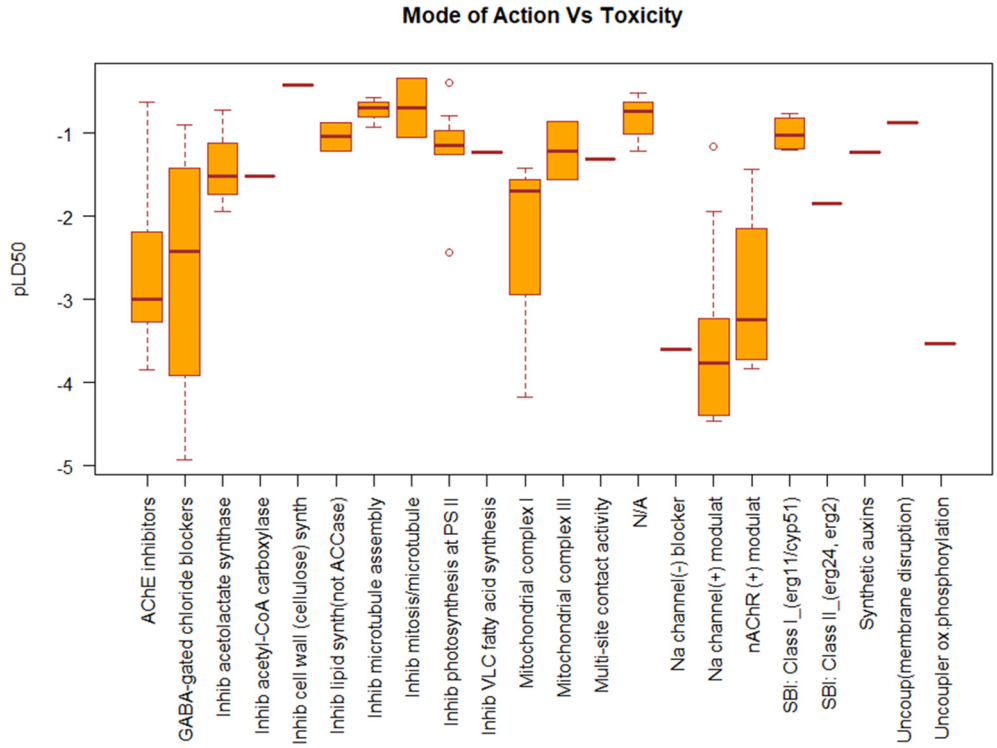


Fig. 5. Distribution of honey bee acute effects (pLD_{50} μmol/bee) for 22 different classes of MoA i.e. target site for pesticide active substances ($n = 113$).

2015). Similarly, while toxicity is triggered by the internal dose causing the adverse effect (e.g. death of the organism), the specificity depends on the key events leading to the adverse outcome as biochemical or physiological mechanisms targeted by the insecticide in the specie (s) of interest (Sanchez-Bayo, 2012; Meek et al. 2011). Examples of how pesticide specificity vary among taxa (as in the case of selective insecticides), or are conserved across taxa (e.g. broad-spectrum insecticides) are available in the literature (Sanchez-Bayo, 2012). The author collected publicly available LD_{50s} of insecticides in honey bees while reporting avermectins as the most toxic insecticide class (LD_{50} 0.04 μg/bee) > neonicotinoids (typical LD_{50} 0.03–3.6 μg/bee) > pyrethroids (typical LD_{50} 0.07–1.3 μg/bee). Although our results are in line with Sanchez-Bayo (2012), we further demonstrated that high variability of pesticide toxicity (LD_{50}) among the same MoA group is often reported (Fig. 5). Therefore, in order to reply to the question above, we suggest that further develop and test the proposed approach for grouping using the MoA harmonised classification in bees, while an additional analysis of the variability of single pesticide toxicity within the same MoA group (e.g. Sodium channel modulator) to integrate the potency aspect when prioritising chemicals of concern in component-based approaches. Nonetheless, besides the intrinsic characteristics of each pesticide active substance (i.e. toxicity and MoA), it is noteworthy that TK-TD variability and inter-colonies variability also play a key role when assessing pesticide (hazard) toxicity (Medrzycki et al., 2013; Wang et al., 2020; Heard et al., 2017; Chmiel et al., 2020).

4. Conclusions

This manuscript has explored the application of EFSA's OpenFoodTox, the US-EPA ECOTOX database and PPDB to develop the first integrative honey bee QSAR models i) to predict acute contact toxicity (LD50) and ii) to profile the MoA of pesticides active substances. Here, seven different QSAR models have been developed, tested and validated for their performance. Two models which can be applied as integrative tools according to a specific hierarchical workflow provided the best predictions:

- A k-NN classification-based model ($LD_{50} \geq 100$ μg/bee (non-toxic) vs. $LD_{50} < 100$ μg/bee (toxic)) which has been validated using statistical parameters i.e. sensitivity (=0.93), specificity (=0.85), accuracy (=0.90), and Matthews correlation coefficient (MCC = 0.78);
- A k-NN Regression-based model ($LD_{50} < 100$ μg/bee; only continuous data) validated with statistical parameters, which were demonstrated to be reliable and robust ($R^2 = 0.72$; MAE = 0.52; in validation test).

These models are currently being implemented within the VEGA-HUB platform (https://www.vegahub.eu/) and all supplementary materials will be available open source on EFSA's Knowledge Junction platform (DOI: https://doi.org/10.5281/zenodo.3755675). The authors acknowledge that k-NN based models present advantages for their implementation, such as qualitative/quantitative predictions based on the most similar compounds i.e. "neighbor" (Mansouri et al., 2018; ECHA, 2016). Similarly, k-NN models also offer further advantages for the implementation of MoA profiling in VEGA models such as visualization of the neighbors used to predict the target compound as well as reliability of the prediction while assessing their known MoA.

Similarly, the current study also proposes the first harmonised MoA classification scheme for 113 pesticides active substances (used as insecticides, acaricides, herbicides, fungicides and plant growth regulator), including potency of the chemicals as acute contact toxicity as LD_{50s} values from three different sources (EFSA's OpenFoodTox, US-EPA ECOTOX and PPDB). Hence, this exercise allowed further defining toxicological MoAs and the target site of PPPs active substances in honey bees, thus enabling regulators and scientists to refine chemical grouping and toxicity extrapolations for component-based RA of multiple chemicals (More et al., 2019; Carnesecchi et al., 2020). In addition, this approach can be of value for research and development to design new active substances for which potency in non-target species such as honey bees is known and controlled, by relying more on the "a priori" knowledge of the

pesticide chemical structure and its potential target site (Commission Regulation, 283/2013, 2013). In a broader context, this manuscript also highlights how New Approach Methodologies (NAMs) such as in silico tools can shift RA to a more mechanistically based understanding of the MoA and mechanism of action of chemicals, thus reducing traditional in vivo experiments and providing alternative testing methods.

Finally, many data gaps remain and still limit the development and broader applications of such QSAR models in honey bees:

- Acute contact toxicity data (e.g. mortality) are the major available datasets for a significant number of compounds and their mixtures but acute/chronic toxicity oral data and toxicity data for sub-lethal effects of pesticides and contaminants in honey bees and wild bees are still lacking (Carnesecchi et al., 2019a, 2019b, 2020). Hence, further work is needed to generate acute and chronic toxicity oral data and sub-lethal toxicity data as the basis to develop QSAR models to predict these effects and integrate such quantitative metrics in RA of single and multiple chemicals (Carnesecchi et al., 2019a; Toma et al., in preparation).
- Statistical variability for single toxicity tests (e.g. LD_{50}) as well as sample size are not usually reported in toxicological studies on bees, thus increasing uncertainty for hazard and RA of chemicals towards non-target organisms (Denton et al., 2003). Such data would be of value when predicting the toxicity of unknown compounds, thereby enhancing their reliability.
- Experimental toxicokinetic data (e.g. absorption, distribution, metabolism and excretion including elimination rate, half-life and bioaccumulation) are also lacking and are an entire part of the MoA of chemicals. This type of information would further support the development of ad-hoc QSAR models to further characterise the impact of fast elimination or persistence of chemicals on the toxicity of single substances as well as on combined toxicity of multiple chemicals in bees (including interactions). Availability of such datasets will allow the further development of QSAR models combined with biometric and life cycle information to generate the next generation of Dynamic Energy Budget (DEB) models for honey bees and wild bees. Finally, since the understanding of single AOPs and their networks is unfolding and, new tools are being developed to monitor species of ecological relevance at the landscape level; integration of mechanistic understanding at different levels of biological organization is foreseen as a mid-term perspective. Ultimately, it will allow risk assessors to tackle toxicity of single chemicals, combined toxicity for multiple chemicals and the resulting risk for honey bees and wild bees at the individual, colony, population and landscape level (Topping et al., 2020; Carnesecchi et al., 2019a, 2020; Baas et al., 2018; LaLone et al., 2017; Spurgeon et al., 2017; Hesketh et al., 2016).

CRediT authorship contribution statement

Edoardo Carnesecchi: Supervision, Project administration, Investigation, Conceptualization, Methodology, Data curation, Writing - original draft, Writing - review & editing, Visualization. **Cosimo Toma:** Conceptualization, Data curation, Software, Methodology, Validation, Formal analysis, Writing - original draft. **Alessandra Roncaglioni:** Writing - review & editing, Supervision, Validation, Funding acquisition. **Nynke Kramer:** Writing - review & editing. **Emilio Benfenati:** Writing - review & editing, Supervision, Funding acquisition. **Jean Lou C.M. Dorne:** Writing - review & editing, Visualization.

Declaration of competing interest

Authors have no competing interests to declare.

Acknowledgments

The views expressed in this manuscript do not reflect the views of the European Food Safety Authority and are the authors only. This work was supported by the European Food Safety Authority (EFSA) [contract number: OC/EFSA/SCER/2018/01 and NP/EFSA/AFSCO/2016/02 (Edoardo Carnesecchi)]; and LIFE-VERMEER (LIFE16 ENV/IT/000167).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2020.139243.

References

Ashton, M., Barnard, J., Casset, F., Charlton, M., Downs, G., Gorse, D., Holliday, J., Lahana, R., Willett, P., 2002. Identification of diverse database subsets using property-based and fragment-based molecular descriptions. Quant. Struct. Relation. https://doi.org/10.1002/qsar.200290002.

Baas, J., Augustine, S., Marques, G.M., Dorne, J. Lou, 2018. Dynamic energy budget models in ecological risk assessment: from principles to applications. Sci. Total Environ. https://doi.org/10.1016/j.scitotenv.2018.02.058.

Bargańska, Ż., Ślebioda, M., Namieśnik, J., 2016. Honey bees and their products: bioindicators of environmental contamination. Crit. Rev. Environ. Sci. Technol. 46, 235–248. https://doi.org/10.1080/10643389.2015.1078220.

Beffa, R., Menne, H., Köcher, H., 2019. Herbicide resistance action committee (HRAC): herbicide classification, resistance evolution, survey, and resistance mitigation activities. Modern Crop Protection Compounds. https://doi.org/10.1002/9783527699261.ch1.

Benfenati, E., Como, F., Manzo, M., Gadaleta, D., Toropov, A., Toropova, A., 2017. Developing innovative in silico models with EFSA's OpenFoodTox database. EFSA Support. Publ. https://doi.org/10.2903/sp.efsa.2017.en-1206.

Benfenati, E., Carnesecchi, E., Roncaglioni, A., Baldin, R., Ceriani, L., Ciacci, A., Kovarich, S., Sartori, L., Mostrag, A., Magdziarz, T., Yang, C., 2020. Maintenance, Update and Further Development of EFSA's Chemical Hazards: OpenFoodTox 2.0. EFSA Supporting Publication 2020:EN-1822. 36 pp. https://doi.org/10.2903/sp.efsa.2020.EN-1822.

Boobis, A.R., Cohen, S.M., Dellarco, V., McGregor, D., Meek, M.E., Vickers, C., Willcocks, D., Farland, W., 2006. IPCS framework for analyzing the relevance of a cancer mode of action for humans. Crit. Rev. Toxicol. https://doi.org/10.1080/10408440600977677.

Bopp, S.K., Kienzler, A., Richarz, A.N., van der Linden, S.C., Paini, A., Parissis, N., Worth, A.P., 2019. Regulatory assessment and risk management of chemical mixtures: challenges and ways forward. Crit. Rev. Toxicol. https://doi.org/10.1080/10408444.2019.1579169.

Breeze, T.D., Bailey, A.P., Balcombe, K.G., Potts, S.G., 2011. Pollination services in the UK: how important are honeybees? Agric. Ecosyst. Environ. https://doi.org/10.1016/j.agee.2011.03.020.

Breiman, L., 2001. Mach. Learn. 45 (1), 5–32 (Stat. Dep. Univ. California, Berkeley, CA 94720). https://doi.org/10.1023/A:1010933404324.

Carnesecchi, E., Svendsen, C., Lasagni, S., Grech, A., Quignot, N., Amzal, B., Toma, C., Tosi, S., Rortais, A., Cortinas-Abrahantes, J., Capri, E., Kramer, N., Benfenati, E., Spurgeon, D., Guillot, G., Dorne, J.L.C.M., 2019a. Investigating combined toxicity of binary mixtures in bees: meta-analysis of laboratory tests, modelling, mechanistic basis and implications for risk assessment. Environ. Int., 133 https://doi.org/10.1016/j.envint.2019.105256.

Carnesecchi, E., Svendsen, C., Lasagni, S., Grech, A., Quignot, N., Amzal, B., Toma, C., Benfenati, E., Guillot, G., Dorne, J.L.C.M., 2019b. Individual datasets investigating combined toxicity of binary mixtures in bees from laboratory tests [data set]. Zenodo https://doi.org/10.5281/zenodo.3383713 Available at.

Carnesecchi, E., Toropov, A.A., Toropova, A.P., Kramer, N., Svendsen, C., Dorne, J. Lou, Benfenati, E., 2020. Predicting acute contact toxicity of organic binary mixtures in honey bees (*A. mellifera*) through innovative QSAR models. Sci. Total Environ. https://doi.org/10.1016/j.scitotenv.2019.135302.

Carriger, J.F., Martin, T.M., Barron, M.G., 2016. A Bayesian network model for predicting aquatic toxicity mode of action using two dimensional theoretical molecular descriptors. Aquat. Toxicol. https://doi.org/10.1016/j.aquatox.2016.09.006.

Casida, J.E., 2009. Pest toxicology: the primary mechanisms of pesticide action. Chem. Res. Toxicol. https://doi.org/10.1021/tx8004949.

Chmiel, J.A., Daisley, B.A., Pitek, A.P., Thompson, G.J., Reid, G., 2020. Understanding the effects of sublethal pesticide exposure on honey bees: a role for probiotics as mediators of environmental stress. Front. Ecol. Evol. https://doi.org/10.3389/fevo.2020.00022.

Cohen, J., 1960. A coefficient of agreement for nominal scales. Educ. Psychol. Meas. https://doi.org/10.1177/001316446002000104.

Commission Regulation (EU) No 283/2013 of 1 March 2013 Setting Out the Data Requirements for Active Substances, in Accordance With Regulation (EC) No 1107/2009 of the European Parliament and of the Council Concerning the Placing of Plant Protection Products on the Market.

Como, F., Carnesecchi, E., Volani, S., Dorne, J.L., Richardson, J., Bassan, A., Pavan, M., Benfenati, E., 2017. Predicting acute contact toxicity of pesticides in honeybees (Apis mellifera) through a k-nearest neighbor model. Chemosphere https://doi.org/10.1016/j.chemosphere.2016.09.092.

Connors, K.A., Beasley, A., Barron, M.G., Belanger, S.E., Bonnell, M., Brill, J.L., de Zwart, D., Kienzler, A., Krailler, J., Otter, R., Phillips, J.L., Embry, M.R., 2019. Creation of a curated

aquatic toxicology database: EnviroTox. Environ. Toxicol. Chem. https://doi.org/10.1002/etc.4382.
Dao, P., Wang, K., Collins, C., Ester, M., Lapuk, A., Sahinalp, S.C., 2011. Optimally discriminative subnetwork markers predict response to chemotherapy. Bioinformatics https://doi.org/10.1093/bioinformatics/btr245.
David, A., Botías, C., Abdul-Sada, A., Nicholls, E., Rotheray, E.L., Hill, E.M., Goulson, D., 2016. Widespread contamination of wildflower and bee-collected pollen with complex mixtures of neonicotinoids and fungicides commonly applied to crops. Environ. Int. 88, 169–178. https://doi.org/10.1016/j.envint.2015.12.011.
De Castro, R.D., de Souza, T.M.P.A., Bezerra, L.M.D., Ferreira, G.L.S., de Brito Costa, E.M.M., Cavalcanti, A.L., 2015. Antifungal activity and mode of action of thymol and its synergism with nystatin against Candida species involved with infections in the oral cavity: an in vitro study. BMC Complement. Altern. Med. 15, 417. https://doi.org/10.1186/s12906-015-0947-2.
Denton, D.L., Fox, J.F., Fulk, F.A., 2003. Enhancing toxicity test performance by using a statistical criterion. Environ. Toxicol. Chem. https://doi.org/10.1897/02-384.
Devillers, J., Pham-Delègue, M.H., Decourtye, A., Budzinski, H., Cluzeau, S., Maurin, G., 2002. Structure-toxicity modeling of pesticides to honey bees. SAR QSAR Environ. Res. https://doi.org/10.1080/1062936021000043391.
Douglas, M.R., Sponsler, D.B., Lonsdorf, E.V., Grozinger, C.M., 2020. County-level analysis reveals a rapidly shifting landscape of insecticide hazard to honey bees (Apis mellifera) on US farmland. Sci. Rep. https://doi.org/10.1038/s41598-019-57225-w.
Dulin, F., Halm-Lemeille, M.P., Lozano, S., Lepailleur, A., Sopkova-de Oliveira Santos, J., Rault, S., Bureau, R., 2012. Interpretation of honeybees contact toxicity associated to acetylcholinesterase inhibitors. Ecotoxicol. Environ. Saf. https://doi.org/10.1016/j.ecoenv.2012.01.007.
ECHA (European Chemicals Agency), 2016. Practical guide - how to use and report (Q) SARs. ECHA-16-B-09-EN. Available at. https://echa.europa.eu/documents/10162/13655/pg_report_qsars_en.pdf.
EFSA (European Food Safety Authority), 2006. Conclusion regarding the peer review of the pesticide risk assessment of the active substance fipronil, 2006. EFSA J. https://doi.org/10.2903/j.efsa.2006.65r.
EFSA (European Food Safety Authority), 2008. Conclusion regarding the peer review of the pesticide risk assessment of the active substance nicosulfuron. EFSA J. https://doi.org/10.2903/j.efsa.2008.120r.
EFSA (European Food Safety Authority), 2010. Conclusion on the peer review of the pesticide risk assessment of the active substance oryzalin. EFSA Journal 8 (9), 1707 2010. [59 pp.]. https://doi.org/10.2903/j.efsa.2010.1707 Available online. www.efsa.europa.eu.
EFSA (European Food Safety Authority), 2013. Guidance on the risk assessment of plant protection products on bees (Apis mellifera, Bombus spp. and solitary bees). EFSA J. 11. https://doi.org/10.2903/j.efsa.2013.3295.
EFSA (European Food Safety Authority), 2018. Evaluation of the Data on Clothianidin, Imidacloprid and Thiamethoxam for the Updated Risk Assessment to Bees for Seed Treatments and Granules in the EU. EFSA Supporting Publication 2018:EN-1378. 31pp. https://doi.org/10.2903/sp.efsa.2018.EN-1378.
EFSA PPR Panel (EFSA Panel on Plant Protection Products and their Residues), 2012. Scientific opinion on the science behind the development of a risk assessment of plant protection products on bees (Apis mellifera, Bombus spp. and solitary bees). EFSA J. 10, 2668. https://doi.org/10.2903/j.efsa.2012.2668.
EFSA PPR Panel (EFSA Panel on Plant Protection Products and their Residues), 2013. Scientific opinion on the identification of pesticides to be included in cumulative assessment groups on the basis of their toxicological profile (2014 update). EFSA J. 11 (7), 3293 2013. 131 pp. https://doi.org/10.2903/j.efsa.2013.3293.
Enoch, S.J., Hewitt, M., Cronin, M.T.D., Azam, S., Madden, J.C., 2008. Classification of chemicals according to mechanism of aquatic toxicity: an evaluation of the implementation of the Verhaar scheme in Toxtree. Chemosphere https://doi.org/10.1016/j.chemosphere.2008.06.052.
Eriksson, L., Jaworska, J., Worth, A.P., Cronin, M.T.D., McDowell, R.M., Gramatica, P., 2003. Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-based QSARs. Environ. Health Perspect. https://doi.org/10.1289/ehp.5758.
European Pesticide Database, 2019. . Available at. https://ec.europa.eu/food/plant/pesticides/eu-pesticides-atabase/public/?event=activesubstance.selection&language=EN, Accessed date: 23 March 2020.
FAO, ITPS, 2017. Global Assessment of the Impact of Plant Protection Products on Soil Functions and Soil Ecosystems. FAO, Rome (40 pp).
Frische, T., Egerer, S., Matezki, S., Pickl, C., Wogram, J., 2018. 5-point programme for sustainable plant protection. Environ. Sci. Eur. https://doi.org/10.1186/s12302-018-0136-2.
Gadaleta, D., Lombardo, A., Toma, C., Benfenati, E., 2018. A new semi-automated workflow for chemical data retrieval and quality checking for modeling applications. J. Cheminform. https://doi.org/10.1186/s13321-018-0315-6.
Genuer, R., Poggi, J.M., Tuleau-Malot, C., 2015. VSURF: an R package for variable selection using random forests. R. J. https://doi.org/10.32614/rj-2015-018.
Ghosh, S., Ojha, P.K., Carnesecchi, E., Lombardo, A., Roy, K., Benfenati, E., 2020. Exploring QSAR modeling of toxicity of chemicals on earthworm. Ecotoxicol. Environ. Saf. https://doi.org/10.1016/j.ecoenv.2019.110067.
Golbraikh, A., Tropsha, A., 2002. Predictive QSAR modeling based on diversity sampling of experimental datasets for the training and test set selection. J. Comput. Aided Mol. Des. https://doi.org/10.1023/A:1020869118689.
Golbraikh, A., Shen, M., Xiao, Z., Xiao, Y. De, Lee, K.H., Tropsha, A., 2003. Rational selection of training and test sets for the development of validated QSAR models. J. Comput. Aided Mol. Des. https://doi.org/10.1023/A:1025386326946.
Gramatica, P., Sangion, A., 2016. A historical excursus on the statistical validation parameters for QSAR models: a clarification concerning metrics and terminology. J. Chem. Inf. Model. https://doi.org/10.1021/acs.jcim.6b00088.
Hamadache, M., Benkortbi, O., Hanini, S., Amrane, A., 2018. QSAR modeling in ecotoxicological risk assessment: application to the prediction of acute contact toxicity of pesticides on bees (Apis mellifera L.). Environ. Sci. Pollut. Res. https://doi.org/10.1007/s11356-017-0498-9.
Heard, M.S., Baas, J., Dorne, J. Lou, Lahive, E., Robinson, A.G., Rortais, A., Spurgeon, D.J., Svendsen, C., Hesketh, H., 2017. Comparative toxicity of pesticides and environmental contaminants in bees: are honey bees a useful proxy for wild bee species? Sci. Total Environ. https://doi.org/10.1016/j.scitotenv.2016.10.180.
Hermann, D., Stenzel, K., 2019. FRAC mode-of-action classification and resistance risk of fungicides. Modern Crop Protection Compounds. https://doi.org/10.1002/9783527699261.ch14.
Hesketh, H., Lahive, E., Horton, A.A., Robinson, A.G., Svendsen, C., Rortais, A., Dorne, J.L., Baas, J., Spurgeon, D.J., Heard, M.S., 2016. Extending standard testing period in honeybees to predict lifespan impacts of pesticides and heavy metals using dynamic energy budget modelling. Sci. Rep. https://doi.org/10.1038/srep37655.
Iwasa, T., Motoyama, N., Ambrose, J.T., Roe, R.M., 2004. Mechanism for the differential toxicity of neonicotinoid insecticides in the honey bee, Apis mellifera. Crop Prot. 23, 371–378. https://doi.org/10.1016/j.cropro.2003.08.018.
Johnson, R.M., Mao, W., Pollock, H.S., Niu, G., Schuler, M.A., Berenbaum, M.R., 2012. Ecologically appropriate xenobiotics induce cytochrome P450s in Apis mellifera. PLoS One 7, 1–9. https://doi.org/10.1371/journal.pone.0031051.
Johnson, R.M., Dahlgren, L., Siegfried, B.D., Ellis, M.D., 2013. Acaricide, fungicide and drug interactions in honey bees (Apis mellifera). PLoS One 8, e54092. https://doi.org/10.1371/journal.pone.0054092.
Kepplinger, D., Filzmoser, P., Varmuza, K., 2017. Variable Selection With Genetic Algorithms Using Repeated Cross-Validation of PLS Regression Models as Fitness Measure (arXiv preprint arXiv:1711.06695).
Kienzler, A., Bopp, S.K., van der Linden, S., Berggren, E., Worth, A., 2016. Regulatory assessment of chemical mixtures: requirements, current approaches and future perspectives. Regul. Toxicol. Pharmacol. 80, 321–334. https://doi.org/10.1016/j.yrtph.2016.05.020.
Kienzler, A., Barron, M.G., Belanger, S.E., Beasley, A., Embry, M.R., 2017. Mode of action (MOA) assignment classifications for ecotoxicology: an evaluation of approaches. Environ. Sci. Technol. https://doi.org/10.1021/acs.est.7b02337.
Kienzler, A., Connors, K.A., Bonnell, M., Barron, M.G., Beasley, A., Inglis, C.G., Norberg-King, T.J., Martin, T., Sanderson, H., Vallotton, N., Wilson, P., Embry, M.R., 2019. Mode of action classifications in the EnviroTox database: development and implementation of a consensus MOA classification. Environ. Toxicol. Chem. https://doi.org/10.1002/etc.4531.
Kuhn, M., 2008. Building predictive models in R using the caret package. J. Stat. Softw. https://doi.org/10.18637/jss.v028.i05.
LaLone, C.A., Villeneuve, D.L., Wu-Smart, J., Milsk, R.Y., Sappington, K., Garber, K.V., Housenger, J., Ankley, G.T., 2017. Weight of evidence evaluation of a network of adverse outcome pathways linking activation of the nicotinic acetylcholine receptor in honey bees to colony death. Sci. Total Environ. https://doi.org/10.1016/j.scitotenv.2017.01.113.
Leroux, P., Bach, J., Debieu, D., Fillinger, S., Fritz, R., Walker, A.S., 2008. Mode of action of sterol biosynthesis inhibitors and resistance phenomena in fungi. Mod. Fungic. Antifung. Compd. V 15th Int. Reinhardsbrunn Symp. Friedrichroda, Ger, pp. 85–92 May 6–10.
Manganaro, A., Pizzo, F., Lombardo, A., Pogliaghi, A., Benfenati, E., 2016. Predicting persistence in the sediment compartment with a new automatic software based on the k-Nearest Neighbor (k-NN) algorithm. Chemosphere https://doi.org/10.1016/j.chemosphere.2015.10.054.
Mansouri, K., Grulke, C.M., Judson, R.S., Williams, A.J., 2018. OPERA models for predicting physicochemical properties and environmental fate endpoints. J. Cheminform. https://doi.org/10.1186/s13321-018-0263-1.
Martin, T.M., Grulke, C.M., Young, D.M., Russom, C.L., Wang, N.Y., Jackson, C.R., Barron, M.G., 2013. Prediction of aquatic toxicity mode of action using linear discriminant and random forest models. J. Chem. Inf. Model. https://doi.org/10.1021/ci400267h.
Martin, T.M., Young, D.M., Lilavois, C.R., Barron, M.G., 2015. Comparison of global and mode of action-based models for aquatic toxicity. SAR QSAR Environ. Res. https://doi.org/10.1080/1062936X.2015.1018939.
Medrzycki, P., Giffard, H., Aupinel, P., Belzunces, L.P., Chauzat, M.P., Claßen, C., Colin, M.E., Dupont, T., Girolami, V., Johnson, R., Conte, Y. Le, Lückmann, J., Marzaro, M., Pistorius, J., Porrini, C., Schur, A., Sgolastra, F., Delso, N.S., Van Der Steen, J.J.M., Wallner, K., Alaux, C., Biron, D.G., Blot, N., Bogo, G., Brunet, J.L., Delbac, F., Diogon, M., El Alaoui, H., Provost, B., Tosi, S., Vidau, C., 2013. Standard methods for toxicology research in Apis mellifera. J. Apic. Res. https://doi.org/10.3896/IBRA.1.52.4.14.
Meek, M.E., Boobis, A.R., Crofton, K.M., Heinemeyer, G., Van Raaij, M., Vickers, C., 2011. Risk assessment of combined exposure to multiple chemicals: a WHO/IPCS framework. Regul. Toxicol. Pharmacol. 60, S1–S14. https://doi.org/10.1016/j.yrtph.2011.03.010.
More, S.J., Bampidis, V., Benford, D., Bennekou, S.H., Bragard, C., Halldorsson, T.I., Hernández-Jerez, A.F., Koutsoumanis, K., Naegeli, H., Schlatter, J.R., Silano, V., Nielsen, S.S., Schrenk, D., Turck, D., Younes, M., Benfenati, E., Castle, L., Cedergreen, N., Hardy, A., Laskowski, R., Leblanc, J.C., Kortenkamp, A., Ragas, A., Posthuma, L., Svendsen, C., Solecki, R., Testai, E., Dujardin, B., Kass, G.E., Manini, P., Jeddi, M.Z., Dorne, J.C., Hogstrand, C., 2019. Guidance on harmonised methodologies for human health, animal health and ecological risk assessment of combined exposure to multiple chemicals. EFSA J. 17. https://doi.org/10.2903/j.efsa.2019.5634.
Negri, I., Mavris, C., Di Prisco, G., Caprio, E., Pellecchia, M., 2015. Honey bees (Apis mellifera, L.) as active samplers of airborne particulate matter. PLoS One https://doi.org/10.1371/journal.pone.0132491.
OECD (Organisation for Economic Co-operation and Development), 1998. Test No. 214: Honeybees, Acute Contact Toxicity Test, OECD Guidelines for the Testing of

Chemicals, Section 2. OECD Publishing, Paris https://doi.org/10.1787/9789264070189-en Available at.
OECD (Organisation for Economic Co-operation and Development), 2007. Guidance document on the validation of (quantitative) structure-activity relationship [(Q)SAR] models, ENV/JM/MONO(2007)2. OECD Publishing, Paris Available at. https://www.oecd.org/env/guidance-document-on-the-validation-of-quantitative-structure-activity-relationship-q-sar-models-9789264085442-en.htm.
OECD (Organisation for Economic Co-operation and Development), 2017. Revised guidance document on developing and assessing adverse outcome pathways. Ser. Test. Assessment. OECD Publ. Off., Paris-France Available at:. http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2013)6&doclanguage=en.
OECD (Organisation for Economic Co-operation and Development), 2018. Considerations for assessing the risks of combined exposure to multiple chemicals, series on testing and assessment no. 296, environment, health and safety division, environment directorate. Available at. https://www.oecd.org/chemicalsafety/risk-assessment/considerations-for-assessing-the-risks-of-combined-exposure-to-multiple-chemicals.pdf.
Prado, A., Pioz, M., Vidau, C., Requier, F., Jury, M., Crauser, D., Brunet, J.-L., Le Conte, Y., Alaux, C., 2019. Exposure to pollen-bound pesticide mixtures induces longer-lived but less efficient honey bees. Sci. Total Environ. 650, 1250–1260. https://doi.org/10.1016/j.scitotenv.2018.09.102.
Quinlan, J.R., 1986. Induction of decision trees. Mach. Learn. https://doi.org/10.1023/A:1022643204877.
Quinlan, J.R., 1987. Simplifying decision trees. Int. J. Man. Mach. Stud. https://doi.org/10.1016/S0020-7373(87)80053-6.
Regulation (EC), 2009. No 1107/2009 of the European Parliament and of the council of 21 October 2009 concerning the placing of plant protection products on the market and repealing council directives 79/117/EEC and 91/414/EEC. OJ L 309 (1), 1–50 24.11.2009.
Rortais, A., Arnold, G., Dorne, J.-L., More, S.J., Sperandio, G., Streissl, F., Szentes, C., Verdonck, F., 2017. Risk assessment of pesticides and other stressors in bees: principles, data gaps and perspectives from the European Food Safety Authority. Sci. Total Environ. 587–588, 524–537. https://doi.org/10.1016/j.scitotenv.2016.09.127.
Rose, T., Kremen, C., Thrupp, A., Gemmill-Herren, B., Graub, B., Azzu, N., Antunes, V., Bruteig, I.E., Buchori, D., Donaldson, J., Dhyani, P.P., Garibaldi, L., Getz, A., Goss, M., Iqbal, J., Kasina, M., Kinuthia, W., Kwapong, P., Manetto, S., Martins, D., Nyamasio, G., Nyamongo, D.O., Odhiambo, C., América Suarez De Oliveira, D., Owusu, E.H., Pandey, B., Poole, C., Roubik, D.W., Roy, P., Waghchoure, E., Lusike, W., 2015. Policy analysis paper: mainstreaming of biodiversity and ecosystem services with a focus on pollination. United Nations Food and Agriculture Organization (FAO) with the contribution of participants at the "Policies for Pollination Management" Worksh. Available at. http://www.fao.org/3/a-i4242e.pdf.
Roy, J., Kumar Ojha, P., Carnesecchi, E., Lombardo, A., Roy, K., Benfenati, E., 2020. First report on a classification-based QSAR model for chemical toxicity to earthworm. J. Hazard. Mater. https://doi.org/10.1016/j.jhazmat.2019.121660.
Russom, C.L., Bradbury, S.P., Broderius, S.J., Hammermeister, D.E., Drummond, R.A., 1997. Predicting modes of toxic action from chemical structure: acute toxicity in the fathead minnow (pimephales promelas). Environ. Toxicol. Chem. https://doi.org/10.1897/1551-5028(1997)016<0948:pmotaf>2.3.co;2.
Samworth, R.J., 2012. Optimal weighted nearest neighbour classifiers. Ann. Stat. https://doi.org/10.1214/12-AOS1049.
Sanchez-Bayo, F., 2012. Insecticides mode of action in relation to their toxicity to non-target organisms. J. Environ. Anal. Toxicol. https://doi.org/10.4172/2161-0525.s4-002.
Sanchez-Bayo, F., Goka, K., 2016. Impacts of pesticides on honey bees. Beekeeping and Bee Conservation - Advances in Research. https://doi.org/10.5772/62487.
Schulp, C.J.E., Lautenbach, S., Verburg, P.H., 2014. Quantifying and mapping ecosystem services: demand and supply of pollination in the European Union. Ecol. Indic. https://doi.org/10.1016/j.ecolind.2013.07.014.
Segner, H., 2011. Moving beyond a descriptive aquatic toxicology: the value of biological process and trait information. Aquat. Toxicol. 105, 50–55 Available online. http://ec.europa.eu/health/scientific_committees/docs/challenges_mandate_en.pdf.
Simon-Delso, N., Amaral-Rogers, V., Belzunces, L.P., Bonmatin, J.M., Chagnon, M., Downs, C., Furlan, L., Gibbons, D.W., Giorio, C., Girolami, V., Goulson, D., Kreutzweiser, D.P., Krupke, C.H., Liess, M., Long, E., Mcfield, M., Mineau, P., Mitchell, E.A., Morrissey, C.A., Noome, D.A., Pisa, L., Settele, J., Stark, J.D., Tapparo, A., Van Dyck, H., Van Praagh, J., Van Der Sluijs, J.P., Whitehorn, P.R., Wiemers, M., 2015. Systemic insecticides (Neonicotinoids and fipronil): trends, uses, mode of action and metabolites. Environ. Sci. Pollut. Res. https://doi.org/10.1007/s11356-014-3470-y.
Singh, K.P., Gupta, S., Basant, N., Mohan, D., 2014. QSTR modeling for qualitative and quantitative toxicity predictions of diverse chemical pesticides in honey bee for regulatory purposes. Chem. Res. Toxicol. https://doi.org/10.1021/tx500100m.
Sparks, T.C., Nauen, R., 2015. IRAC: mode of action classification and insecticide resistance management. Pestic. Biochem. Physiol. https://doi.org/10.1016/j.pestbp.2014.11.014.
Spurgeon, D., Hesketh, H., Lahive, E., Svendsen, C., Baas, J., Robinson, A., Horton, A., Heard, M., 2016. Chronic oral lethal and sub-lethal toxicities of different binary mixtures of pesticides and contaminants in bees (Apis mellifera, Osmia bicornis and Bombus terrestris). EFSA Support. Publ. 13. https://doi.org/10.2903/sp.efsa.2016.en-1076.
Spurgeon, D., Hesketh, H., Lahive, E., Svendsen, C., Baas, J., Robinson, A., Horton, A., Heard, M., 2017. Chronic oral lethal and sub-lethal toxicities of different binary mixtures of pesticides and contaminants in bees (Apis mellifera, Osmia bicornis and Bombus terrestris). EFSA Support. Publ. 13. https://doi.org/10.2903/sp.efsa.2016.en-1076.
Tanimoto, T.T., 1958. An Elementary Mathematical Theory of Classification and Prediction in IBM Internal Report.
Toma, C., Carnesecchi, E., Benfenati, E., Dorne, J.L., 2020. First report on regression based- and classification-based QSAR models for predicting oral acute toxicity of pesticides active substances towards honey bee (*Apis mellifera*). Toxicol. Lett. J. (in preparation).
Topping, C.J., Aldrich, A., Berny, P., 2020. Overhaul environmental risk assessment for pesticides. Science (80-.) https://doi.org/10.1126/science.aay1144.
Toropov, A.A., Benfenati, E., 2007. SMILES as an alternative to the graph in QSAR modelling of bee toxicity. Comput. Biol. Chem. https://doi.org/10.1016/j.compbiolchem.2007.01.003.
Tosi, S., Nieh, J.C., 2019. Lethal and sublethal synergistic effects of a new systemic pesticide, flupyradifurone (Sivanto®), on honeybees. Proc. R. Soc. B Biol. Sci. 286, 20190433. https://doi.org/10.1098/rspb.2019.0433.
Tosi, S., Costa, C., Vesco, U., Quaglia, G., Guido, G., 2018. A 3-year survey of Italian honey bee collected pollen reveals widespread contamination by agricultural pesticides. Sci. Total Environ. https://doi.org/10.1016/j.scitotenv.2017.09.226.
USEPA, PMRA, CALDPR, 2014. Guidance for assessing pesticide risks to bees. Available at. U.S. Environmental Protection Agency (USEPA), Health Canada Pest Management Regulatory Agency (PMRA) and California Departament of Pesticide Regulation, Office of Pesticide Programs USEPA, Health Canada PMRA, CDPRhttps://www.epa.gov/pollinator-protection/pollinator-risk-assessment-guidance.
Venko, K., Drgan, V., Novič, M., 2018. Classification models for identifying substances exhibiting acute contact toxicity in honeybees (Apis mellifera)$. SAR QSAR Environ. Res. https://doi.org/10.1080/1062936X.2018.1513953.
Verhaar, H.J.M., Van Leeuwen, C.J., Hermens, J.L.M., 1992. Classifying environmental pollutants. 1: structure-activity-relationships for prediction of aquatic toxicity. Chemosphere https://doi.org/10.1016/0045-6535(92)90280-5.
Wang, M., Braasch, T., Dietrich, C., 2020. Reduction of variability for the assessment of side effects of toxicants on honeybees and understanding drivers for colony development. PLoS One https://doi.org/10.1371/journal.pone.0229295.
WHO (World Health Organization), 2009. Environmental health criteria 240: principles and methods for the risk assessment of Chemicals in Food. WHO, Geneva (Definitions page A-25). Available at. http://www.who.int/foodsafety/publications/chemical-food/en.
Wing, K.D., Andaloro, J.T., McCann, S.F., Salgado, V.L., 2005. Indoxacarb and the sodium channel blocker insecticides: chemistry, physiology, and biology in insects. Comprehensive Molecular Insect Science. https://doi.org/10.1016/B0-44-451924-6/00070-3.
Worth, A.P., Bassan, A., Gallegos, A., Netzeva, T.I., Patlewicz, G., Pavan, M., Vračko, M., 2005. The characterisation of (quantitative) structure-activity relationships: preliminary guidance. Available at. Institute for Health and Consumer Protection, Toxicology and Chemical Substances Unit, European Chemical Bureauhttps://core.ac.uk/download/pdf/93511332.pdf.
Wright, M.N., Ziegler, A., 2017. Ranger: a fast implementation of random forests for high dimensional data in C++ and R. J. Stat. Softw. https://doi.org/10.18637/jss.v077.i01.