Annex to: EFSA (European Food Safety Authority), Adriaanse P, Arce A, Focks A, Ingels B, Jölli D, Lambin S, Rundlof M, Süßenbach D, Del Auguila M, Ercolano V, Ferilli F, Ippolito A, Szentes C, Neri F M, Padovani L, Rortais A, Wassenberg J and Auteri D, 2023. Revised guidance on the risk assessment of plant protection products on bees (Apis mellifera, Bombus spp. and solitary bees). EFSA Journal 2023;21(5):7989, 27 pp. doi:10.2903/j.efsa.2023.7989

Annex A – Refinement of residue dissipation in the ERA of pesticides for terrestrial organisms

Table of Contents

1. Introduction

Pesticide residues generally decrease over time in all environmental matrices. Dissipation is especially relevant from items that are eaten or used as shelter by non-target organisms, because it has an obvious direct influence on the exposure levels they experience. Dissipation and degradation of residues from plant material including pollen and nectar, seeds and arthropods is often more rapid than in other environmental media (e.g., soil). The speed of dissipation is generally expressed as the 'half-life', and this parameter is used to:

determine the peak concentration following successive applications of the substance, in comparison with a single application, i.e., the so-called MAF (Multiple Application Factor);

determine the time-weighted average exposure level when this is prolonged in time. Often this is also expressed as a factor (fTWA).

Especially for chronic exposure these aspects should be viewed as complementary in describing the residue dynamics. For birds and mammals, a distinction is made between MAF_{acute}[1] and MAF_{repro}[1] × fTWA. For bees, these aspects are combined into the parameter '*Conc*' (for either pollen or nectar).

Reasonable worst-case default half-life values are used in Tier 1 of the exposure assessment. However, these default value can be adjusted if reliable substance-specific data are available.

This document discusses several aspects which should be considered when designing/performing and assessing studies aimed at refining the decline of residues in environmental matrices other than soil, water and sediment – with a specific focus on food items for birds, wild mammals, and bees.

Within this Annex, the decline of residues is – unless otherwise specified – due to dissipation of the chemical. The processes involved in the dissipation of residues may differ between different matrices (hereafter referred to as 'items'). Physical-chemical degradation[2] processes such as photolysis, abiotic chemical degradation and biotic metabolisation, as well as some physical processes such as wash-off and volatilisation from the matrix on which the residue is present are likely relevant for all items, but they may contribute to a different extent to the overall fate of the substance. Moreover, some other physical and/or biological processes differ significantly between items.

Residue decline for some plant materials (i.e., green parts including seedlings, flower buds, and fruits) can be substantially influenced by plant growth. Especially for fast-growing species and varieties, the increase of plant mass can determine a decrease of the residue concentration, expressed as the ratio between mass of active substance and mass of matrix, via increasing the denominator. On the contrary, this aspect is unlikely to be relevant for other items such as pollen, nectar, (ungerminated) seeds, arthropods, and earthworms. Their mass either does not increase with time or does not increase significantly in a time-frame comparable to the dissipation time of a pesticide.

The exposure assessment focuses on the quantification of residues in items that are present in a certain area during a specified time-window, i.e., between the first application and a specified number of days after the last application, depending on the non-target organism (21 days for terrestrial vertebrates, 5 to 27 days for bees). Any plant material collected in the area of interest at any time point is the same that was initially contaminated following the application of the pesticide. Hence, the overall residue decline is driven uniquely by the decline in/on each of the relevant items. The only exception to this is represented by pollen and nectar: for those items, a substantial depletion of the initially contaminated material may be caused by active sampling from pollinators, with material replacement due to plant production.

[1] Please note that MAF_{acute} and MAF_{repro} are the same as MAF_{90} and MAF_{m}, respectively, which were used EFSA (2009).
[2] Degradation is defined as a chemical break-down of the chemical into metabolites

For soil applications (e.g., seed treatment or granules) concentration in plant parts (including pollen and nectar) may nevertheless remain more or less stable, due to a continuous uptake from soil and translocation to the upper parts of the plant. In such case, dissipation should not be accounted for in the risk assessment.

The assumption of a monotonic decline of residues post-application in the potentially sampled material is also not applicable for samples of multiple arthropods, where residues dynamics are influenced by more complex population dynamics. Arthropods, especially flying ones, move around in the environment, which has at least two consequences for the residue dynamics in samples of multiple arthropods in a certain area:

Decline of residues may be driven by exclusion of contaminated individuals due to death or emigration and/or by inclusion of uncontaminated ones due to birth or immigration. Furthermore, some processes can be partially reversible: some contaminated individuals may initially abandon the study area only to come back late;

Arthropods may pick up residues from other contaminated matrices (i.e., plants, soil) as they move on or through it. Indeed, it is not uncommon to observe a maximum level of residues in arthropods several days after the application of a substance.

This latter point is also relevant for samples of multiple earthworms. Residue dynamics in earthworms share in principle many aspects with arthropods, but movement into/out of the study area is generally more limited due to their slower migration.

In addition, depending upon the properties of the substance in question, residues may accumulate along a food chain and may therefore alter the decline dynamics depending on the position of a species within the food web.

Taken together, these unique factors mean that sampling of invertebrates - especially arthropods – in studies aiming to refine residue decline should consider more elements, which will be further discussed in other areas of this Annex.

2. Study types

In principle, laboratory, semi-field, and field studies provide relevant information for understanding residue dynamics. The study type is a trade-off between the ability to control the conditions of the experiment on one side and their environmental realism on the other.

Laboratory studies by design exclude the influence of some variables (e.g., wash-off due to rainfall, food web interaction, arthropod immigration/emigration, etc.) which allows more accurate description of some specific processes. For example, they might give a more accurate picture of the degradation dynamics by limiting many other dissipation processes. For the same reasons, however, studies in controlled conditions are considerably poorer than field studies in capturing the diversity of environmental processes that can influence overall residues decline in real agro-ecosystems. This is particularly true for residues in arthropods, but also for pollen and nectar. Furthermore, laboratory studies in invertebrates are generally limited to single species, whereas field studies investigate the whole arthropod and earthworm community, which together represents the potential food of insectivorous/vermivorous birds and mammals. Hence, a single species test is less representative compared to data obtained from the whole fauna, which is often the case in field studies. If single species studies are available, then clear justification of why the data are applicable to the diet of the model/focal species should be provided by the applicant.

On the other side of the spectrum, field studies provide a more comprehensive picture of the residue decline due to all relevant processes occurring at the same time. These kinds of studies are generally not able to distinguish the relative influence of individual processes, which makes extrapolation of the results from one specific set of conditions to another much more difficult. Nevertheless, this problem can be partially addressed by performing multiple field experiments, thereby capturing as much relevant environmental variability as possible.

In summary, field experiments are generally considered more appropriate for refining the residue decline in specific items in a risk assessment context. However, a suitable level of replication must be ensured to capture the relevant environmental variability.

For residue decline in pollen and nectar, the sampling can, in principle, be performed by collecting material directly from the treated plants, or indirectly by sampling bees foraging on the plants. Nevertheless, the second option is not acceptable, as any dynamics in the resulting pattern will be influenced by the foraging behaviour of the individual bees. This can in principle be accounted for in the landscape factors (LF_p and LF_n) but should not influence the quantification of the 'pure' dissipation dynamics from the contaminated matrices.

Sometimes studies performed for the quantification of residues for human health risk assessment can also inform on the residue decline for food items for birds and wild mammals, mostly plant materials. Since the focus should be on the rate of dissipation and degradation under practical use conditions, data from confined residue studies covering all routes of loss are more relevant than plant metabolism studies which are focussed on metabolic processes, rather than decline. Indeed, the confined residue studies may report residue values for several sampling points. It should be highlighted that these studies are usually not targeted at deriving a dissipation half-life (hereafter DT_{50}) or at describing residue dynamics on a time scale of few days after the initial exposure peaks. Still, studies with sampling starting directly after the initial application often exist and allow kinetic analyses. Care must be taken that the concentrations at individual data points refer to the same plant item (e.g., whole plant or green plant parts).

In general, due to the aforementioned limitation, it may be advantageous to conduct targeted plant dissipation studies if refinement of the dissipation kinetics is intended.

Terminology used for residue decline studies

Site or location: one geographically defined area where a specific experiment is carried out within a country/region/state. Sites/locations are characterised by homogenous landscape composition, topography, agronomic, and geo-climatic conditions. No clear boundaries can be set for site identification/separation. However, two sites should be considered independent if those have sufficient geographical distance to allow some difference in the general environmental conditions. European Commission (2019) mentions that different sites must be at least 20 km far from one another. EFSA (2014), for soil $DegT_{50}$, mentions that spatial variation in daily rainfall may be considerable on a scale of 100 km^2. For the present guidance, as rule of thumb, ≈ 100 km is considered a sufficient distance, but in case of very diverse landscape and agro-climatic conditions, smaller geographical distances may still be appropriate. If the distance between sites is 20 – 100 km, it must be demonstrated that the conditions are sufficiently different for factors influencing residues decline (e.g., weather patterns).

Trial: one independent residue decline experiment. It is characterised by a unique site, timing, target, sampling strategy, and application (in terms of rate and pattern). A trial may consist of several fields and/or plots, including control and one or more treated field/plots following a specific experimental protocol.

Field: a spatially well-defined unit on which the test plants are grown.

Plot: a spatially-characterised sub-unit used for a trial. It may be an entire field or a portion of a field (i.e., several plots may be identified in a single field). When plots in one trial are managed with the same experimental treatment (e.g., same crop, same applications, and same sampling strategy), they can be considered to be spatial (non-independent) replicates of a single trial. This is also applicable when the plots are defined within a single field. No general characteristics of plot dimension can be given. However, for trials aiming at measuring residues dissipation on arthropods a minimum of 1 ha is required for each plot EFSA (2009).

Replicate: generic term which indicates repeated measurements of the same distribution, aiming at quantifying the variability and the central tendency (i.e., mean, median) of that distribution. This can refer to spatial variability, temporal variability, or analytical/measurement variability.

3. Recommendations for performing and assessing residue decline studies

In order to perform residue decline trials that are suitable for the risk assessment, several aspects should be considered. These same aspects should be accounted for by the risk assessors when evaluating the intrinsic reliability and the relevance of these studies for a specific risk assessment.

Type of test: As mentioned in the previous Section, targeted field studies are preferred for refinement of the dissipation time. Other types of studies may be considered on a case-by-case basis if it can be shown that other aspects of the study were sufficiently worst-case to cover the uncertainty involved in the use of application apparatus other than what is foreseen based on the Good Agricultural Practice (GAP) and in deviating from environmental field conditions.

Test item: For a trial carried out in the context of a (re)authorisation at the EU level for an active substance, the test item should in principle be the representative formulation. If the context is the authorisation of a specific plant protection product, then preferably this should be the test item. If this is not the case, it should be assessed whether the tested formulation is sufficiently similar to the representative one (i.e., same type of formulation and similar co-formulants). Particular attention should be paid to elements which may alter the environmental fate of the active substance (e.g., encapsulation, solid vs. liquid form, etc.).

Irrigation and rainfall: If relevant, the irrigation regime should be representative for the use under assessment. As in any field test, it is almost impossible to control for rain events. Nevertheless, rainfall can in some cases play a substantial role in determining the temporal pattern of residues, mainly via wash-off from the monitored item. The effect of the rainfall should be carefully evaluated when assessing the dissipation dynamics. In extreme cases, e.g., in case of storm/strong rainfall events occurring before 50% of the test item has dissipated, the outcome of the trial may be considered too biased to describe a typical dissipation kinetic. The way this process is modelled in FOCUS may help in such an assessment. According to Van den Berg et al. (2016), a rough classification of wash-off, based upon experimental results Leistra (2005) suggests that an average of 50% pesticide mass wash-off with 10 mm rainfall - and no rain during the first 6 h after application - may be expected to occur.

Application method: In principle, the trial should be carried out using the same application method specified in the GAP. However, this is not necessarily a strict requirement. The potential influence of the application method on the dissipation kinetics should carefully be assessed on a case-by-case basis. In case of spray application, special attention should be paid to the adjustment of the volume of the spray liquid towards the dimension of the crop, especially in orchards (tree size, tree spacing, row spacing, etc.).

Application timing - BBCH: The application pattern used in the trial should mimic the GAP at least in terms of the BBCH plant development stage at the time of the application(s), especially if the study aims to refine the decline on parts of the treated crop. Deviations are allowed when it could be established with reasonable certainty that the trial is carried out under worst-case conditions (e.g., slower growth rate than the expected one, in order to minimise dilution). It should be considered that the goal is to determine dissipation kinetics, which are not species-specific, but representative for the entire group of items (see 'analysed item (matrix)' below).

For trials focusing on arthropods, even more flexibility should be allowed, as these are often not carried out with the crop mentioned in the GAP (see 'Application target' – crop) and hence a BBCH match is less meaningful. This mismatch between crops for arthropod residues studies is purposeful and correct, as some crops would present poor arthropod communities in terms of both diversity and abundance, and hence trials in these crops would not be considered sufficiently informative/robust. This will depend upon the geographic area and crop in question and the organisms for which the risks are to be assessed. The BBCH stage is also obviously not meaningful when the trial focuses on seeds. If the GAP specifies the time of the year when application should be performed together with the BBCH, any deviation from such timing should be assessed in consideration of possible effects on the residue decline (e.g., higher temperatures, day length, etc.).

Application target – crop: Ideally, any residue decline trial should be carried out on the crop for which a refinement is sought. However, there are many instances for which this statement does not apply. First, the goal is to determine dissipation kinetics which are not crop-specific, but representative for the entire group of items (see 'analysed item' below). Furthermore, if the target of the experiment is the residue decline in invertebrates, other considerations must be made. For some crops and growth stages (e.g., early growth stages of vegetables), the abundance of arthropods is intrinsically low, due to the habitat structure they offer. Collecting enough arthropod material in such conditions can be very difficult, therefore surrogate crops and/or growth stages can be used. Arthropod residues decline trials are often performed in orchards, as these can host considerably more individuals and species. While this is, in general, a sensible choice, it is still important to consider other aspects: an orchard plot which has been intensively farmed for several decades, surrounded by other high production commercial orchards, may nevertheless contain a less-diverse and/or smaller arthropod community. Conversely, a small orchard out of production, surrounded by a diverse woodland habitat may hold an enormous number and type of arthropods but exchange with the surrounding source habitat may lead to an unrepresentatively fast dilution of individuals with a residue loading. In addition to the number and types of arthropod present, the likely exposure should also be considered, with some plant growth-stages/structures being likely to decrease direct exposure (spraying) and also build-up of residues due to interception of either the arthropods directly or the other plants/plant parts where the arthropods might most likely reside.

Application pattern: When the study is used to determine only a DT_{50} value, the number and the frequency of applications reported in the GAP should not necessarily be mirrored in the experimental design for a decline study, unless there is concern that application at different times over the season may result in different half-lives (e.g., because of different growth rates of the plant).

However, for birds and mammals, when the study is used to determine either an empirical MAF_{acute}[1] or a MAF_{repro}[1] × fTWA from the area under the residue vs. time curve (hereafter AUC), then it is important that the number and frequency of applications used in the experiment exactly mirror the ones in the GAP. It should be noted that, as discussed below, this is an option only for residue decline in invertebrates.

However, sampling of arthropods can be quite resource intensive. In consideration of this, if the GAP reports a high (>5) number of applications, it is suggested that any build-up due to multiple applications could be addressed with an experiment with a lower number of applications. However, an exact number cannot be defined *a priori* because this would depend on the specific GAP in question and take into consideration available information on the residue dynamics in arthropods. For example, if no build-up of residues is observed after 5 applications, it is unlikely that this will be observed after further applications. Similarly, if the build-up is very regular (i.e., the maximum value after each application increases always by a certain factor/amount compared to the previous one), then it should be possible to 'predict' the maximum value after more applications than the ones tested experimentally.

Application rate: The application rate (in terms of amount of test item applied per hectare) is, in principle, not relevant for assessing the dissipation kinetics. This is a general principle which

is applied in determining the environmental fate of substances in different compartments. Nevertheless, other types of considerations should be made. For example, the application rate should be high enough to guarantee a suitable number of time-points when the residues of the substances are still quantifiable (i.e., above the LOQ). On the other hand, application rates should not be significantly higher than the GAP if this causes a damage to the items being assessed and thus compromises the ability to collect enough materials at later time-points (e.g., if the arthropod community is decimated by the application of an insecticide). If the dosing in the study significantly deviates from the proposal in the GAP, it should be justified why this was the case.

Other treatments: For the same reasons as described above (see 'Application rate'), it should be ascertained whether other treatments might cause/have caused abnormalities in the target items. For instance, if the target is arthropods, no insecticides other than the test item should be used during the six-month period before the experimental dates. This is to ensure the maximum abundance of arthropods. Similarly, to prevent major habitat changes, no herbicides should be used during the trial unless it is necessary to assure a proper application of the test substance for the survival of the crop itself. Fungicides which are known to be toxic towards insects or plants should be treated in the same manner as insecticides or herbicides, accordingly.

Analysed item (matrix): In principle, the analysed items should mirror those considered for the risk assessment. However, a certain degree of extrapolation is possible. Extrapolation is generally considered appropriate within homogenous clusters of items, which are broadly defined as:

- Dicot plants (green parts and roots[3])
- Monocot plants (green parts and roots[3])
- Pollen
- Nectar
- Fruits
- Seeds (both weed seeds and cereals seeds)
- Foliar-dwelling arthropods
- Ground arthropods
- Flying insects
- Earthworms

No experience has yet been gained for the residue decline on flower buds. Once sufficient data is available, it can then be considered if it belongs to one of the above groups.

Extrapolation across groups is generally considered inappropriate, but common sense and expert judgement should be used in both directions: for example, extrapolation between dicot weeds and grass-like weeds may be possible for trials performed at late growth stage. On the contrary, extrapolating from maize to grass-like weeds is not recommended because of the extremely fast-growing nature of maize. Residues in any other matrix (e.g., soil, water, etc.) cannot be extrapolated to any foodstuff.

Plot size and number: for trials focusing on plant materials, the size of the plots should ensure that there is an area where the application is homogeneous enough, i.e., sufficiently far from the edge. Quantification of this area cannot be established *a priori* as it depends on the crop characteristics (e.g., size and height of the plants, space between rows) and on the application method. For trials focusing on arthropods, the plot area should also ensure that, when sampling in the inner part, emigration and immigration from untreated areas is minimised. An area of at

[3] Note that roots are not a relevant matrix for spray applications. However, both roots and green parts of plants are relevant for seed treatment and other types of soil applications.

least 1ha has historically been recommended as a size for the plots of arthropods residue decline studies (EFSA, 2009). This recommendation is maintained in this document, noting that deviations may be possible if appropriate justification is provided. Such justification should account for: (1) The dispersal/movement potential of the sample species; (2) The distance between the sampling spot and the edge of the plot; (3) typology of habitat surrounding the plot. In order to capture the intra-site variability, at least three replicate plots should be available and, ideally, enough material should be collected for each to allow separate analysis. While this minimum number of plots was already a requirement for arthropods studies in the previous Birds and Mammals Guidance Document (EFSA, 2009), this is now extended to all items. However, in order not to exclude useful information, this requirement should be interpreted with flexibility for existing studies on plant material (including seeds).

Sampling locations: Sampling at the boundaries of the plot should be avoided. This is applicable for sampling of both plant material and invertebrates. Particularly for arthropods, an attempt should be made to collect organisms in the inner part of the plot, if enough organisms are available. For flying insects, samples should be taken at a greater distance from the boundary. In any case, the distance between the sampling location and the field/plot margin should always be reported.

Taxonomy of samples (arthropods only): In principle, information on the feeding habits of the model/focal species should be used for optimising the sampling and focussing on those arthropod classes (e.g., beetles, caterpillars, spiders, etc.) which constitute the largest part of the diet. Nevertheless, this is often extremely difficult in practice both because of lack of such detailed information on the model/focal species diet and because of limitations in the availability of a specific class of arthropods in the samples. Furthermore, according to the grouping presented above (see point 'analysed item (matrix)'), it is considered appropriate to extrapolate among classes of arthropods as long as the separation of the three main strata (i.e., ground, foliar and flying) is maintained. However, consistency in the relative taxonomic composition between successive sampling should be maximised. Smaller animals generally have a higher surface/weight ratio, so they might present a higher residue concentration than larger ones. Hence, an inconsistent sampling in terms of body size classes may produce considerable noise and hamper a proper description of the residue decline over time.

Sampling methods (arthropods only): In general, the sampling method should minimise any potential alteration of the residue values. Desiccation of the sample matrix should be avoided via quick handling times and storage of the samples as soon as possible in a deep freezer or on dry ice. Cross-contamination with other (non-arthropod) materials like soil particles or plant material should also be avoided, which is often the case with suction sampling techniques. A summary of potentially useful sampling methodologies is available in Appendix N of EFSA (2009). Further advancements have been made in the field since then, but those are not discussed here.

Sampling methods (pollen and nectar only): As discussed above, pollen and nectar should be sampled directly from the treated plant. Indirect sampling via collection of foraging bees should be avoided to reduce the influence of other processes not linked to the dissipation from the contaminated matrices.

Sampling frequency: The sampling scheme should be adjusted to the properties of the test substance, particularly in consideration of the expected dissipation time. As discussed in the Section outlining the parametric method, below, if the objective of the trial is the derivation of a DT_{50}, the experimental design should aim to provide at least five timepoints with quantifiable residue levels. However, if the objective is to determine an empirical AUC, a larger number of time points may be necessary, and it is important that the AUC is well described until at least *n* days after the last application (where *n* is the standard time-window for the calculation of the fTWA).

As a general rule of thumb, it is more important to accurately describe the dissipation shortly after the application, when most of the dissipation is likely to take place, particularly when the dissipation does not follow a single-first-order (SFO) kinetics. Hence, the interval between

sampling should be shorter at the beginning for the study and longer at the end. If it is expected that most of the substance will dissipate in the first few hours, then the interval between sampling should be calibrated accordingly. For such fast-dissipating substances, it might be necessary to perform more than one sampling on the day of the application, so that the shape of the decline is sufficiently captured. On the other hand, if the substance is expected to persist on the items long after the application, it is better to design a sampling schedule with longer intervals between samplings, which then extends further in time.

Some samples should also be obtained before the first application to adjust the sampling effort required for each method intended and to obtain reference matrix for the analytical laboratory.

Analytical phase: the analytical phase of the study should focus on the recommendations and the principles reported in the Guidance SANTE/2020/12830 (European Commission, 2021). Furthermore, for invertebrates, replicates (from each plot) should be analysed separately whenever possible.

4. Recommendations for reporting of the experimental phase

As with any other study included in a submission to a regulatory authority, residue decline trials should be performed according to GLP. The respective report should clearly describe the aim, all methods, protocol deviations, encountered problems and results of the experiment. The report should allow the evaluation of all the aspects described in the previous Section. Briefly, details should be reported including:

- Clear identification of the analysed item/matrix.
- Location (country/region/state/province, whichever is applicable) of the experimental site(s). If available, geographical coordinates of the plot locations should be reported as well.
- Basic figures (e.g., daily min, max, mean) for climatic variables such as temperature and rainfall. Location of the weather station and distance to the trial site(s) should also be reported.
- Identity of the test item. Particularly, information should be available on the type of formulation tested and the presence of relevant co-formulants.
- Treated crop in terms of species, variety (if known), and growth stage at the time(s) of application of the test item.
- For studies on specific plant species (i.e., not 'weeds'), the weight of a pre-defined number of plants (e.g., 20 plants) should be reported for several time points – at least for the first and the last sampling – in order to estimate the growth rate of the plants.
- Application technique, rate(s), timing and frequency.
- Sampling times, methodology, and size (number of samples per site and their weight). For arthropods, information on either the taxonomic or the class composition of the samples are also considered useful.
- Number, size, and relative location of the experimental plots.
- The distance between the sampling location and the field/plot margin.
- Information on the plot handling during the experiment. This includes any relevant agronomic practises such as irrigation, mowing, application of other plant protection products, etc.
- For studies focusing on arthropods, information on the landscape surrounding the experimental plots.
- The sample storage (including during transport) and stability.
- The analytical procedure including sample preparation, extraction, and purification.

- Information on the analytical validation requirements listed in SANTE/2020/12830 (European Commission, 2021). It is pivotal that LOQ and LOD are clearly reported.
- The results of the analysis should be reported to the maximum possible level of disaggregation (raw data).

5. Dissipation kinetics of a single trial

If a time-weighted average residue concentration is required for the risk assessment, it can be either determined parametrically after fitting a kinetic decline model or by considering the empirical AUC. The first method is always preferable, as the derivation of kinetic parameters allows calculation of fTWA × MAF_{repro}[1] and MAF_{acute}[1] (for birds and mammals) and of '*Conc*' (for bees) under any use pattern with the same active substance. Nevertheless, empirical quantification of AUCs provides useful information for a specific use, though it makes it very difficult to extrapolate to others unless a clear worst-case can be identified.

The use of an empirical AUC is not accepted for residue decline on plant material (including seeds) as, in general, the results of a well-conducted study can be reasonably fitted to a kinetic model. However, this is not always the case for invertebrates. Thus, for invertebrates, in the first instance an attempt should be made to fit the residue data to a kinetic model. However, if this attempt does not yield a satisfactory outcome, an empirical AUC can still be used as a basis for estimating a fTWA × MAF_{repro} and/or a MAF_{acute} for the risk assessment.

5.1. Parametric method: dissipation kinetics

The dissipation kinetic should be calculated considering comparable matrices over time (e.g., sampling of whole plants and samples of only leaves cannot be considered in the same dataset for deriving dissipation kinetics). Most of the recommendations are from FOCUS (2006). Further, more specific, recommendations can be retrieved therein. Worked example A, at the end of this Annex offers some practical guidance using a fictitious case with SFO kinetics. For further examples with other kinetics, please consult FOCUS (2006) and FOCUS (2014).

- At least five quantifiable time-points should be available for fitting the decline curve. In some exceptional cases four points may be enough (e.g., fast dissipation of the active substance or residues of metabolites with slow formation) but the points should never be fewer than four.
- If true replicates (i.e., samples from different plots) exist, they should be used in the fitting (averaging between replicates before fitting should be avoided). Analytical replicates, on the contrary, should always be averaged.
- It is recommended that a minimum of three true replicates are collected and used in the fitting. While this minimum number of plots was already a requirement for arthropods studies in the previous Guidance Document (EFSA, 2009), this is now extended to all items. In order to utilise all information, this requirement should be interpreted with flexibility for existing studies on plant material (including seeds). Generally, the reliability of kinetic estimation increases when more replicates are available.
- For values below the LOQ/LOD, the following procedure should be followed:
  - All values between the LOD and LOQ are set to the actual measured value. If the actual measured concentration has not been reported, use 0.5 x (LOQ + LOD).
  - All samples < LOD are set to ½ LOD.
  - The curve should be cut off after the pesticide has largely dissipated. All samples after the first non-detect (< LOD) should be omitted unless positive detections above LOQ are made later in the experiment. In that case, samples are included up to the first non-detect (<LOD) which is NOT followed by later positive samples above LOQ.

- If an outlier is rejected based on expert judgement, this must be clearly indicated in the report and, where possible, supported by statistical analysis.
- Initial values should, as a first step, be included in the optimisation (not constrained). If a constrained procedure is to be used, this should be well-justified.
- The first time point should preferably report residues sampled after the application, during the application dates. It is acknowledged that this may be practically unfeasible for night-active arthropods, thus this aspect should be interpreted with flexibility for that specific case.
- The kinetic model (SFO, FOMC, DFOP, HS, etc.) used to fit the data should be reported, together with the relevant parameter estimates (and related 95% uncertainty limits).
- Goodness-of-fit should be assessed using four indicators, all of which should be clearly reported. It should be noted these indicators should be evaluated as a whole and not in a hierarchical manner:
  - Visual fit-> plot of time ~ concentration should be provided. Ideally, the fitted line should pass through (or in the vicinity of) the measurement points.
  - Residual plot -> Plot of time ~ residuals against the y=0 line should be provided. Points should ideally be scattered around the zero line. Regular patterns are generally indicative that the used kinetic model is not appropriate. Underestimation of the last time points is indicative of an under-conservative kinetic. For further explanation and figures see FOCUS (2006).
  - Chi-square (χ^2) error % -> should be reported and should possibly be <15%. Note that Chi-square error should be calculated using the mean of true replicates.
  - *t*-test and confidence intervals of individual model parameters should be reported. *t*-test resulting in p-values >0.05 (or confidence intervals including zero) indicate large uncertainty in the estimation of the model parameters. In cases where p-values are between 0.05 to 0.1, the parameter may still be considered acceptable, however further discussion and justification based on the fit, as well as on the weight-of- evidence from other available data for the substance is then necessary.
- The software/package used to fit the data should be clearly reported. Preferably, the calculator tool accompanying the relevant guidance document should be used.
- The selection of the appropriate kinetic (and thus of the appropriate endpoint) should follow the recommendations of FOCUS (2006) about how to derive endpoints as modelling inputs (Section 7.1.2, particularly Figure 7.2). Briefly, SFO kinetics are always preferred if the fitting is acceptable (even if other kinetic models may give a slightly better fit). Particularly for plant materials, SFO are always preferred unless there is an indication that the fitted model underestimates the DT_{50} (it is not worst-case). The use of pseudo-DT_{50} (i.e., $DT_{90}/3.32$) obtained with the FOMC model may be appropriate if SFO kinetics cannot be used. Following the recommendation of FOCUS (2006) FOMC should not be used for sequential metabolites (i.e., fitted together with the parent).
- It is recommended that the expert evaluation of the kinetic modelling is presented in the (d)RR/DAR/RAR.

5.2. Empirical method: area under the curve (AUC)

For invertebrates, whenever a derivation of kinetic parameters via fitting a model is not possible, a more empirical method can still be used to refine the residue levels over time. This method consists of integration of the area under the residue vs. time curve (see Figure 1). It is important that the dataset covers:

- For the derivation of the MAF_{acute}: at least a few days after the last application

- For the derivation of the fTWA × MAF_{repro} (to be determined together): at least 21 days after the last application.

The reference residue value is always the highest residue value following the first application (and before the second). This corresponds to the 'standard' practice for quantifying the majority of default RUD values.

The MAF_{acute} would correspond to the ratio between the highest recorded residue value and the reference residue value.

The derivation of the fTWA × MAF_{repro} is a bit more laborious. As a first step, a moving-time-window approach is used to identify the 21-day window with the largest AUC. Next, this value is divided by the time window in days (21) to derive the average daily residue value in the worst-case 21-day window. Finally, the fTWA × MAF_{repro} corresponds to the ratio between the average daily residue value and the reference residue value.

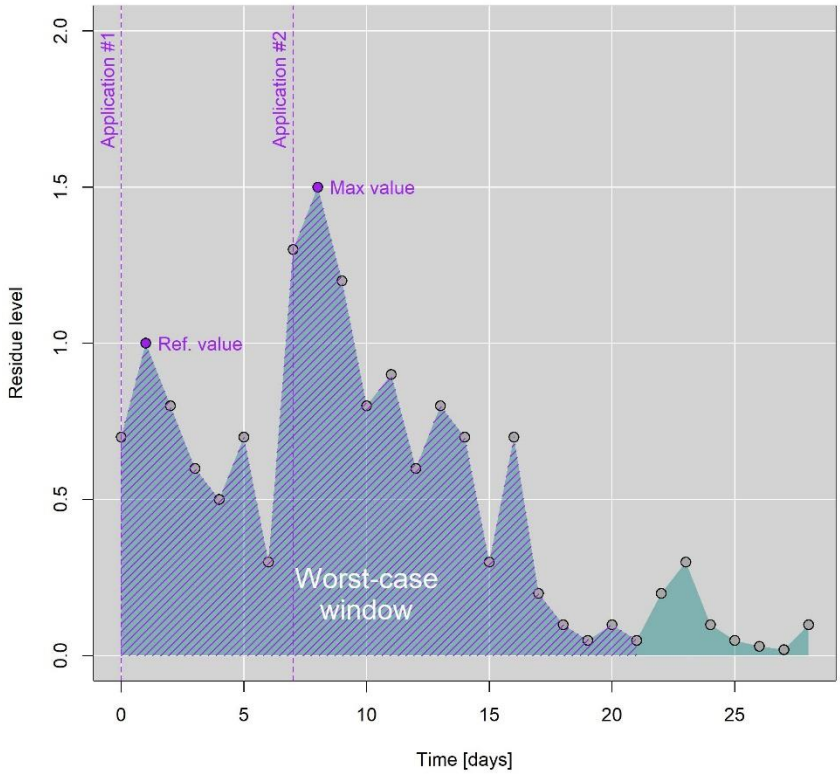


Figure 1: Illustrative example of the main elements used to derive MAF_{acute} and fTWA × MAF_{repro} from an empirical AUC.

It should be noted that this method is not equivalent to using the AUC directly in the risk assessment. The latter would mean ignoring the default RUD values considered in the Guidance Document and is thus not recommended. Hence, the AUC should only be used to derive multiplicative factors for the default RUDs (MAF_{acute} and/or fTWA × MAF_{repro}), unless these have also been refined by a suitable substance-specific dataset.

6. Dealing with multiple application trials

In case of residues measured after multiple applications, if the decline can be described parametrically, two alternatives are possible (Worked example B at the end of this Annex, relates to this issue). The first one is:

- consider each application (and the following points until the next application) as a stand-alone trial;
- calculate as many DT_{50}s as the number of the applications;
- calculate the geometric mean of the calculated DT_{50}s as a representative for the multiple application trial.

Otherwise:

- express all concentrations in terms of % of the one measured on the last application date (i.e., on the day of each application, the value will be 100%);
- calculate the time between each measurement time point and the date of the last application;
- use the newly derived values for the fitting exercise.[4]

7. Use of residue dissipation in the risk assessment

7.1. Minimum number of trials and types of data

The sites and conditions of the residue decline trials should be representative of the proposed usage. Data from a trial conducted in a northern Member State should generally be appropriate for a northern MS risk assessment. However, it may be possible to use data from region A to support uses in region B if it is obvious that the conditions in region A tend to be worse than in region B so that the risk will not be underestimated. The acceptability of this extrapolation should be considered on a case-by-case basis.

A linear mixed model with the monitored substance as the random part of the model and fitted to all data on green parts (monocots+dicots) from the Lahr et al. (2018) database, revealed that the region of sampling has a significant influence on the DT_{50}. While the same result was not replicated for trials carried out on fruits and arthropods (influence borderline significant), it is proposed to nevertheless consider, as initial step, a differentiation between the regulatory zones.

More than one site should be used as between-site variations are likely to be greater than within one site. This puts more emphasis on spatial variability, although it is acknowledged that temporal variability may also play an important role (e.g., weather conditions for the same dates in different years may diverge greatly).

The number of sites studied should cover an appropriate range of situations to ensure that the data are representative of the proposed uses. In theory, the minimum number of trials should be calibrated in order to ensure a sufficient level of confidence in the derived endpoint (e.g., DT_{50}, AUC, etc.). If one would take the DT_{50} as the reference endpoint, the expected variance between trials could be used as a basis for determining the width of the confidence interval around the average in relation to the number of trials.

In reality, such variance will change for every item and substance. Therefore, using such an evaluation to establish an *a priori* number of minimum trials is not straightforward. In addition, there is a need to balance minimum requirements between statistical confidence and experimental feasibility.

For residues on vegetable materials, a minimum of ***trials carried out in four sites per item category and regulatory zone*** is recommended in order to establish a reliable refinement of the dissipation of the pesticide. This number is proposed considering that: (I) three to five trials have historically been considered the minimum for refining DT_{50} values; (II) the residue data requirement for MRL (minor crops) specifies that at least four independent trials should be available; (III) at least four soils should be tested for establishing a valid DT_{50} in soil. However,

[4] Note that this approach is likely to be successful with plant residues, while it might not be so for residues on invertebrates, where the dynamic is further complicated by other processes such as organisms' movement, different surface/body weight ratios, bioaccumulation, etc.

it should be considered that particular climatic conditions of certain areas may allow extrapolation, to some extent (e.g., Northern France could be representative of the central zone or the northern zone despite being in the southern zone).

For invertebrates, the Lahr et al. (2018) database reports a smaller variance than for plants, although this must be considered with care, due to a significantly more limited data availability. In consideration of this, and in view of the objective practical difficulties in carrying out these studies, a minimum of ***trials carried out in two sites per item category and regulatory zone*** is recommended.

For seeds, residue refinements for treated seeds can only be performed with treated seeds, and residues refinements for over-sprayed seeds should only be performed with over-sprayed seeds. **Since seed treatments are interzonal, trials should be performed in all zones unless it is clearly demonstrated that the seeds will or cannot be sown in one of the zones (e.g., due to the growing conditions, crop, etc.). Similar to vegetable materials, there should be at least four sites per item category and regulatory zone.** Residues may decline differently in seeds which are buried or partially buried in the soil. However, these seeds are also less available for birds and mammals, depending on the depth at which the seeds are drilled. Therefore:

- For seeds which are drilled at a **depth of ≤3cm**, birds/mammals may dig up buried seeds and therefore it must be considered whether residues trials with both protected (covered) and unprotected (surface) seeds are needed to adequately refine the residues level (i.e., the assessment needs to cover both seeds on the soil surface and those which are buried).
- For seeds drilled at a **depth >3cm**, it is assumed that the seeds are too deep to be readily dug up by birds and mammals, residue trials should be performed using unprotected (surface) seeds (i.e., the assessment needs to cover only seeds on the soil surface).

7.2. Combining trials

As a general rule, results from different trials on equivalent items and carried out in the same regulatory zone should be averaged before being used in the risk assessment.

Combining DT_{50}s

FOCUS (2006) highlights the use of geometric mean DT_{50}s, as it ensures consistency between the averaging of DT_{50}s and the averaging of the corresponding dissipation constants in the SFO kinetics.

When the dissipation in all trials is described by SFO kinetics, the averaging of the DT_{50} (or of the dissipation constant $k = \log(2)/DT_{50}$) is straightforward. However, if other kinetics were used, the appropriate figure to be included in the average is the following:

- FOMC: $DT_{90}/3.32$
- DFOP and HS: slower DT_{50}, unless HS: $Ct<0.1$ M0 at breakpoint or DFOP: $g>0.9$ → In this case use fastest DT_{50}. This is because if >90% of the residue decline is described with the faster DT_{50} of the DFOP or HS, then the faster DT_{50} could be used instead of slower DT_{50}.

However, the use of the **worst-case value** should be considered when:

- the dataset presents limited reliability (see Section on recommendations for performing and assessing residue decline studies)
- the dataset x is relatively small (4-6 studies) and one value is considerably higher than the others (e.g., $\max(\log(x))-\text{mean}(\log(x))>2*\text{sd}(\log(x))$)

Combining $fTWA \times MAF_{repro}$

For invertebrates, when a reliable DT_{50} cannot be obtained for some of the available trials, it is still possible to combine the information from the available studies into a single figure on the basis of their fTWA × MAF_{repro} for a specific item. The main condition is that the trials should be homogenous in terms of their application pattern (i.e., number and frequency of applications).

In this case, it is recommended that the arithmetic mean of the available fTWA × MAF_{repro} is used in the risk assessment. As when combining DT_{50}s, the use of worst-case values may still be considered under specific circumstances.

No combination of trials is recommended for refining the MAF_{acute}. In such case, a 90^{th} percentile can be used if this could be calculated with sufficient confidence (i.e., at least 20 trials with the same use pattern). Otherwise, the worst-case value should be used in the exposure assessment.

7.3. Splitting or merging datasets

When DT_{50} values are available over multiple items and geographical areas (e.g., regulatory EU zones), a consideration should be given as to whether the DT_{50} estimations are part of the same distribution. When this is the case, merging the datasets can simplify the risk assessment and provide more robust dissipation estimation[5].

This could be assessed visually (generally visualisations like boxplots helps in this case) or statistically by running the appropriate tests. When running such tests, alpha (α) levels higher than the standard 0.05 may be considered, if there is a concern that large dataset variations would lead to accepting the null hypothesis despite potentially influential differences in the mean values. If a higher alpha is used, this should be documented, and underpinning reasons should be explained.

- If only two groups are present, then a *t*-test is normally the most straightforward approach. When data from the two groups clearly deviate from a normal distribution[6], then an equivalent non-parametric test can be used (Wilcoxon or Mann-Whitney tests). If the difference is significant at the chosen alpha, it is likely that the considered factor (crop, geographical area, etc.) plays a role in the dissipation speed, and hence the two datasets should not be merged.
- If more than two groups are present (e.g., data for the three regulatory areas) an ANOVA test is the most straightforward approach. A common non-parametric equivalent for ANOVA is the Kruskall-Wallis test. Once again, if the difference between group means is found to be significant at α=0.05, it is likely that the considered factor plays a role in the dissipation speed, and hence the datasets should not be merged.
- Whenever more discriminatory variables are considered to be potentially influencing (e.g., crop AND regulatory areas) a two-way ANOVA (with or without interaction term) can be run. Non-parametric equivalent to two-way ANOVA exist, but they are not as straightforward as in the previous cases.

7.4. Dealing with toxic metabolites (birds and mammals only)

Whenever the toxicity of the metabolites is comparable to that of the parent (within a factor of 2-3), the use of the TWA factor (fTWA) based on the parent only should not be considered acceptable. In this case, the residue measurements should report for both the parent and the appropriate metabolite.

FOCUS (2006) specifies how to fit residue data for the parent and the metabolite from the same experiment. However, one should be aware that, in this case, the number of parameters to be

[5] This procedure should not be considered a 'shortcut' for reducing the number of trials per item group and regulatory area. The minimum number of trials per item group and zone should still be respected, before making any conclusion on the possibility of merging datasets.

[6] Often the number of values is too small for appreciating the shape of the distribution. In this case, it is suggested to stick to the parametric test (generally more powerful).

estimated is generally rather high. Therefore, the number of data points needed for a reliable fitting exercise increases accordingly.

When the toxicity of the parent and the metabolite is similar, an alternative approach could be to sum their concentrations at any time point and then fit the obtained data as if it was a single chemical. The derived DT_{50} would then be valid for the sum of parent+metabolite.

In other cases, the difference in the toxicity may be larger, but it may not be possible to conclude that the risk assessment for the parent 'covers' for the metabolite. This is for example the case when:

- the metabolite toxicity is significantly higher than the parent toxicity;
- the metabolite toxicity is significantly (more than a factor 2-3) lower than the parent, but a high formation factor and longer persistence makes the exposure to the metabolite non-negligible compared to that to the parent.

In such cases, a more elaborate approach can be taken by 'normalising' the exposure of the metabolite in consideration of the toxicity ratio with the parent. Such normalised exposure can then be summed to the one of the parent for a combined risk assessment using the endpoint from the parent.

Illustrative examples of both approaches are presented in the Worked example C at the end of this Annex.

References

EFSA, (European Food Safety Authority) 2009. Risk Assessment for Birds and Mammals. EFSA Journal, 2009;7(12):1438 https://doi.org/10.2903/j.efsa.2009.1438

EFSA, (European Food Safety Authority) 2014. EFSA Guidance Document for evaluating laboratory and field dissipation studies to obtain DegT50 values of active substances of plant protection products and transformation products of these active substances in soil. EFSA Journal, 2014;5(5):3662 https://doi.org/10.2903/j.efsa.2014.3662

European Commission, 2019. Technical Guidelines on data requirements for setting maximum residue levels, comparability of residue trials and extrapolation of residue data on products from plant and animal origin. European Commission document reference SANTE/2019/12752.

European Commission, 2021. Guidance Document on Pesticide Analytical Methods for Risk Assessment and Post-approval Control and Monitoring Purposes. SANTE/2020/12830, Rev.1.

FOCUS, (FOrum for the Co-ordination of pesticide fate models and their USe), 2006. Guidance document on estimating persistence and degradation kinetics from environmental fate studies on pesticides in EU Registration Report of the FOCUS Work Group on Degradation Kinetics. EC Document Reference SANCO/10058/2005-v. 2.0.

FOCUS, (FOrum for the Co-ordination of pesticide fate models and their USe), 2014. Assessing Potential for Movement of Active Substances and their Metabolites to Ground Water in the EU" Report of the FOCUS Ground Water Work Group. European Commission document reference Sanco/13144/2010 v. 3.

Lahr, Joost, Krämer, Wolfgang, Mazerolles, Vanessa, Poulsen, Véronique, Jölli, Daniela, Müller, Marc, McVey, Emily, Wassenberg, Jacoba, Derkx, Ria, Brouwer, Ans, Deneer, John, Beltman, Wim, Lammertsma, Dennis, Jansman, Hugh & Buij, Ralph 2018. Data collection for the estimation of ecological data (specific focal species, time spent in treated areas collecting food, composition of diet), residue level and residue decline on food items to be used in the risk assessment for birds and mammals. 2018;15(11):1513 https://doi.org/10.2903/sp.efsa.2018.EN-1513

Leistra, Minze, 2005. Estimating input data for computations on the volatilisation of pesticides from plant canopies and competing processes. ALTERRA, Wageningen. Available online: https://edepot.wur.nl/25119

Van den Berg, F, Leistra, Minze, Van der Linden, AMA, Boesten, JJTI & Tiktak, Aaldrik, 2016. PEARL model for pesticide behaviour and emissions in soil-plant systems: description of the processes in FOCUS PEARL. The Statutory Research Tasks Unit for Nature & the Environment MILIEU, W. N.

Appendix A – Worked example A - Data treatment for a single trial

The available fake dataset is supposed to contain residue data for a random pesticide applied once to cereals. Results are related to the whole plant. The fictious dataset was conceived to show an example of a good kinetic fitting.

The method of analysis was satisfactory reported, and the LOD and LOQ are reported below:

LOQ	2 µg/kg
LOD	0.66 µg/kg

The available dataset is reported in the following table.

DAT	Plot	Conc. [µg/kg]			
		Analysis 1	Analysis 2	Average analysis	Average per DAT
0	A	10.45	10.82	10.635	10.71
0	B	9.32	9.37	9.345	
0	C	12.4	11.9	12.15	
1	A	9.2	9.43	9.315	8.65
1	B	8.3	7.97	8.135	
1	C	8.56	8.43	8.495	
3	A	7.2	7.45	7.325	6.55
3	B	5.34	5.65	5.495	
3	C	6.78	6.9	6.84	
7	A	4.2	4.07	4.135	3.79
7	B	3.42	3.57	3.495	
7	C	3.8	3.65	3.725	
12	A	1.97	1.99	1.98	1.73
12	B	1.56	1.52	1.54	
12	C	1.7	1.63	1.665	
21	A	<LOD	<LOD	0.33	0.33
21	B	<LOD	<LOD	0.33	
21	C	<LOD	<LOD	0.33	
42	A	<LOD	<LOD	N/C	N/C
42	B	<LOD	<LOD	N/C	
42	C	<LOD	<LOD	N/C	

DAT= Days after treatment

Note that:

- Values at 12 DAT were below the LOQ. Nevertheless, as values were still available, these were used in the fitting exercise (if they were not available, default values halfway between LOD and LOQ would have been used).
- Values at 21 DAT were all below the LOD. Therefore, they were set to half of the LOD.
- Values at 42 DAT were all below the LOD. However, the curve should be cut at the first "non-detection" (happening in this case at DAT 21)
- Values used for the fitting are reported under the column "Average analysis".
  - Values in the rightmost column (average per DAT) will not be used in the dissipation fitting (single true replicates are considered independent)
  - Values in the column "Analysis 1" and "Analysis 2" represent analytical replicates of the same sample, and they are averaged before the fitting.

SFO model fit

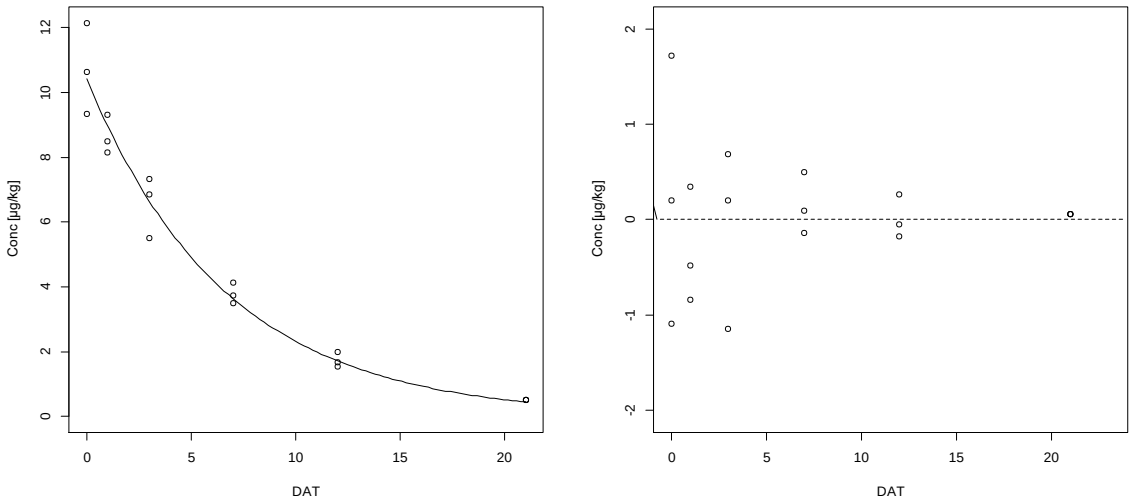


Figure A1: Dissipation curve fitting (left) and related residuals (right)

Model parameters (SFO)

Parameter	Mean estimate	95% Confidence interval	p-value (t-test)
M0 (initial value) [µg/kg]	10.43	9.78-11.1	<0.0001
k (dissipation constant) [day^{-1}]	0.15	0.13-0.18	<0.0001

Model outcome

DT_{50} [days]	4.6
DT_{90} [days]	15.3
χ^2 error %	2.84

Evaluation of the fitting:

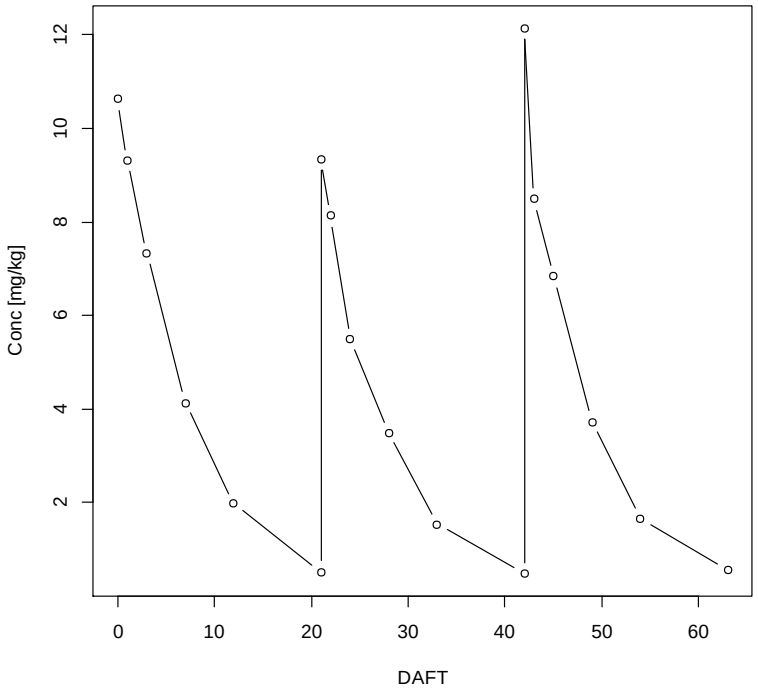
- Visual fit-> plot of time ~ conc shows that the fitted line passes in the vicinity of the measurement points for all dates.
- Residual plot -> Plot of time ~ residuals against the y=0 line shows that points are scattered around the zero line. No regular patterns are identified and no systematic underestimation of overestimation is present at any date.
- Chi-square error % -> well below 15%
- *t*-test for individual model parameters-> *t*-test resulted in *p*-values well below 0.05, indicating high confidence in the estimation of the model parameters.

Appendix B – Worked example B -Dealing with multiple applications

Within this Annex two different options are presented to derive a DT_{50} from a single residue trial with multiple applications. The available fake dataset is supposed to contain residue data for a random pesticide applied three times to pome fruits, with an interval of 21 days. Results are related to fruit.

DAFT	Conc. [mg/kg]
0	10.64
1	9.32
3	7.33
7	4.14
12	1.98
21	0.51
21	9.35
22	8.14
24	5.50
28	3.50
33	1.54
42	0.48
42	12.15
43	8.50
45	6.84
49	3.73
54	1.67
63	0.56

DAFT= Days after first treatment



Option 1: consider each application as a standalone trial

The first procedure presented in the Section "Dealing with multiple applications trials" is followed in the example below.

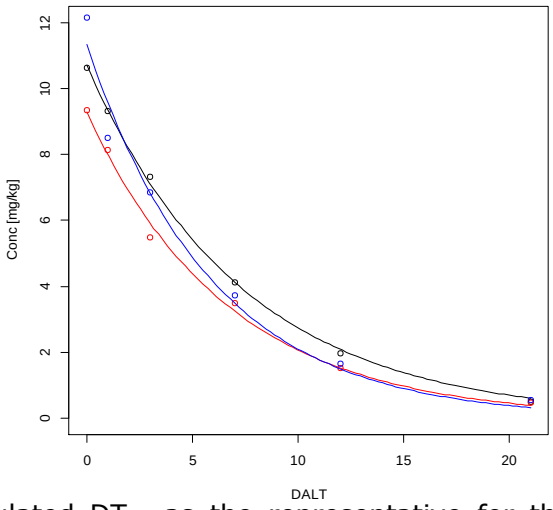
1. Consider each application (and the following points until the next application) as a standalone trial;

DALT	Conc. [mg/kg] Sub-trial 1	Sub-trial 2	Sub-trial 3
0	10.64	9.35	12.15
1	9.32	8.14	8.50
3	7.33	5.50	6.84
7	4.14	3.50	3.73
12	1.98	1.54	1.67
21	0.51	0.48	0.56

DALT= Days after last treatment

2. Calculate as many DT_{50} as the number of the applications;

Sub-trial	DT_{50} [days]
Sub-trial 1 (black line)	5.09
Sub-trial 2 (red line)	4.62
Sub-trial 3 (blue line)	4.11



3. Calculate the geomean of the calculated DT_{50} as the representative for the multiple application trial.

Sub-trial	DT_{50} [days]
Sub-trial 1 (black line)	5.09
Sub-trial 2 (red line)	4.62
Sub-trial 3 (blue line)	4.11
Geomean	**4.59**

Option 2: consider all applications in a unique fitting

The second procedure presented in the Section "Dealing with multiple applications trials" is followed in the example below.

1. Express all concentrations in terms of fraction of the one measured on the last application date (i.e. on the day of each application, the value will be 1);

DAFT	Conc. [mg/kg]	Last application on DAFT	Conc. as fraction of 0 DALT [adimensional]
0	10.64	0	1.00
1	9.32		0.88
3	7.33		0.69
7	4.14		0.39
12	1.98		0.19
21	0.51		0.05
21	9.35	21	1.00
22	8.14		0.87
24	5.50		0.59
28	3.50		0.37
33	1.54		0.16
42	0.48		0.05
42	12.15	42	1.00
43	8.50		0.70
45	6.84		0.56
49	3.73		0.31
54	1.67		0.14
63	0.56		0.05

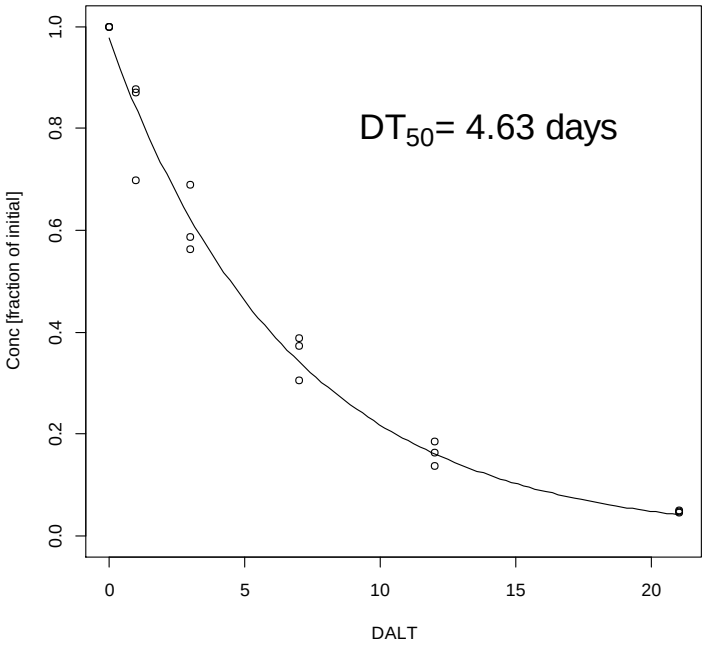
DAFT= Days after first treatment

DALT= Days after last treatment

2. Calculate the time between each measurement time point and the date of the last application;

DALT	Conc. as fraction of 0 DALT [adimensional]
0	1.00
1	0.88
3	0.69
7	0.39
12	0.19
21	0.05
0	1.00
1	0.87
3	0.59
7	0.37
12	0.16
21	0.05
0	1.00
1	0.70
3	0.56
7	0.31
12	0.14
21	0.05

3. Use the newly derived values for the fitting exercise.



Appendix C – Worked example C - Dealing with toxic metabolites (birds and mammals only)

In case the toxicity of the metabolite is comparable to the one of the parent, two approaches can be followed.

Simpler approach

The simpler approach can be followed when it can be assumed that:

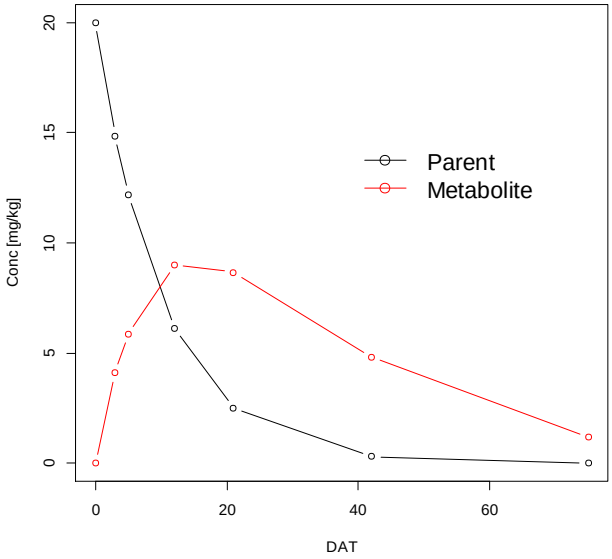
toxicity of the parent ≈ toxicity of the metabolite (within a factor 2-3)

In this case it is enough to sum the residue concentrations at any time point of the parent and the metabolite. Then fit the obtained data as if it was a single chemical.

Let's assume we have a dataset reporting measured residues of parent and one metabolite of equal toxicity. The measurements were performed on 6 dates, from DAT 0 to DAT 75.

DAT	Conc. [mg/kg]	
	Parent	Metabolite
0	20	0
3	14.86	4.11
5	12.19	5.86
12	6.1	9.02
21	2.5	8.66
42	0.31	4.81
75	0.01	1.18

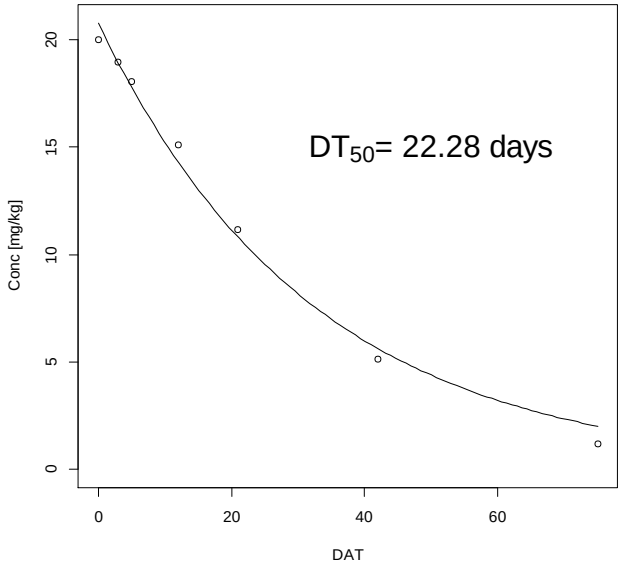
DAT= Days after treatment



Since the toxicity of the two compounds is practically equal, it makes sense to sum the concentration at each time point.

DAT	Conc. [mg/kg]		
	Parent	Metabolite	Sum
0	20	0	20
3	14.86	4.11	18.97
5	12.19	5.86	18.05
12	6.1	9.02	15.12
21	2.5	8.66	11.16
42	0.31	4.81	5.12
75	0.01	1.18	1.19

Once calculated the sum, this can be used for fitting a decline curve.



The resulting DT_{50} can be used to calculate refined MAF_{acute}, $MAF_{chronic}$ × fTWA. These can replace the default values.

It should be noted that the risk assessment with the refined values should be carried out by considering the application rate of the parent. As this approach is applicable when the toxicity of the parent and that of the metabolite are virtually equivalent, the choice of the endpoint for the risk assessment is not so determinant. Following a conservative approach, the lower toxicity value can be used.

In case multiple trials are available, the individual DT_{50}s can be averaged, as described in the main text of this Annex.

More complex approach (parent-normalised fTWA)

This approach should be followed when the toxicity of the metabolite and the one of the parent are comparable (less than a factor of 10) but cannot be assumed as equivalent. It is anticipated that the following approach will mainly be used when:

- the metabolite has a high formation fraction and is more persistent than the parent.
- the metabolite is formed in medium-high amount and is more toxic than the parent.

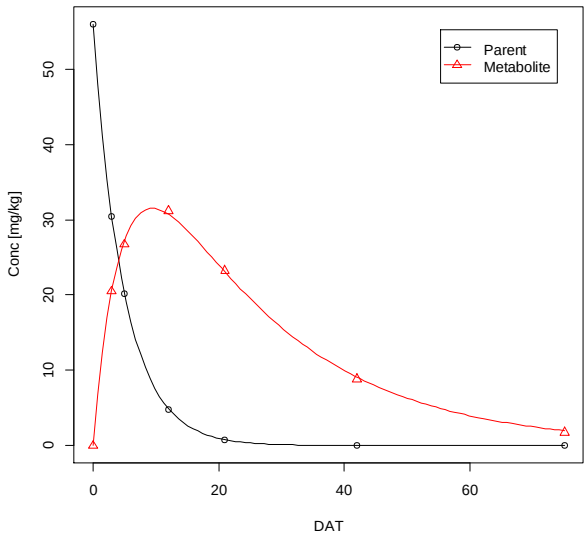
The example dataset includes measured residues of parent and one metabolite. The measurements were performed on 6 dates, from DAT 0 to DAT 75. The toxicity of the metabolite is known to be equivalent to ¼ of the one of the parent.

	Conc. [mg/kg]	
DAT	**Parent**	**Metabolite**
0	56	0
3	30.38	20.5
5	20.21	26.7
12	4.85	31.17
21	0.77	23.23

42	0.01	8.82
75	0	1.73

1. Estimate the kinetic parameters for the parent and the metabolite together.

Parameter	**Parent**	**Metabolite**
Kinetic model	SFO	SFO
M0 (concentration at *t*=0) [mg/kg]	55.89	-
DT_{50} [days]	3.42	15.02
Formation fraction [%]	-	87.25
χ^2%	0.44	1.22



2. Calculate the fTWA of the parent and of the metabolite between DAT 0 and DAT 21. Care should be taken as the fTWA of the metabolite cannot be derived analytically with the standard equation

$fTWA = (1-e^{-k*Time})/(k*Time)$

On the contrary, this should be quantified as area under the curve. The fTWA for the metabolite is calculated as TWA Concentration of the metabolite (same interval used for the parent) divided by the initial concentration of the parent.

fTWA Parent = 0.23

fTWA Metabolite = 0.41

3. Re-calculate the fTWA of the metabolite accounting for the toxicity ratio with the parent (in this case ratio= 0.25) and finally sum up this value with the fTWA for the parent.

Total fTWA = (0.41*0.25) + 0.23= 0.347

The principle of this calculation is to convert the combined exposure to parent and metabolite to a total exposure expressed in the same 'currency', i.e., the toxicity equivalent of the parent. As such, this fTWA should be used in the risk assessment together with the application rate and the endpoint for the parent.

If multiple trials are available, the averaging among trials should not be performed directly on the DT_{50}, as suggested for the parent alone, but can be performed directly on the fTWA. In this case, an arithmetic mean can be used for averaging.

This approach could, in principle, also be used when there are multiple applications, but in such cases the calculations of both MAF_{acute} and $MAF_{chronic}$ × fTWA are more complex, as this would entail adopting a moving time window approach to be applied to the normalised area under the curve.