Funded by
the European Union

Guideline for apicultural citizen science to use the honey bee colony for biomonitoring of the environment

Edition 2. June 2024

Funded by
the European Union

Guideline for apicultural citizen science to use the honey bee colony for biomonitoring of the environment

Authors

Jozef van der Steen[1], Flemming Vejsnæs[11], Hans Baveco[7], Robert Brodschneider[2], Valters Brusbardis[12], Bas Buddendorf[7], Norman L. Carreck[4], Leonidas Charistos[3], Giovanni Formato[8], Dirk de Graaf[13], Kristina Gratzer[2], Alison Gray[5], Fani Hatjina[3], Konstantinos Kasiotis[9], Evangelia Tzanetou[9], Ole Kilpinen[11], Ellen Danneels, Maria Murcia[10], Marco Pietropaoli[8], M. Alice Pinto[6], Andreia Quaresma[6, 14], Ivo Roessink[7], Amadeo Fernandez-Alba[10] M.J. Martínez Bueno[10]

Corresponding author

Jozef van der Steen,
Kerkstraat 96
5061 EL Oisterwijk
Netherlands
Email: alveusab@outlook.com

Affiliations

1. AlveusAB Consultancy, Oisterwijk, NL
2. Department of Biology, University of Graz, Graz, AT
3. Hellenic Agricultural Organization 'DEMETER' Nea Moudania, GR
4. Carreck Consultancy Ltd, Shipley, West Sussex, UK
5. University of Strathclyde, Glasgow, Scotland, UK
6. Centro de Investigação de Montanha (CIMO), Associate Laboratory for Sustainability and Technology in Mountains Regions (SUSTEC), Instituto Politécnico de Bragança, PT
7. Team Environmental Risk Assessment, Wageningen Environmental Research, Wageningen, NL
8. Istituto Zooprofilattico Sperimentale delle Regioni Lazio e Toscana "M.Aleandri", Rome, IT
9. Benaki Phytopathological Institute, Athens, GR
10. University of Almeria, Almeria, ES
11. Danmarks Biavlerforening, Sorø, DK
12. Latvian Beekeepers Association, Jelgava, LV
13. Gent University, Gent, BE
14. Laboratório Associado para a Sustentabilidade e Tecnologia em Regiões de Montanha (SusTEC)

Funded by
the European Union

This guideline reflects the rationale and organisation of the INSIGNIA-EU study "Preparatory action for monitoring of environmental pollution using honey bees". The study focuses on the scientific development of a substantiated Apicultural Citizen Scientist protocol, including data analysis, non-invasive sampling for the honey bee colony, in-hive passive sampler development, molecular detection of pollen origin, pesticide risk exposure modelling for honey bee colonies, and modelling of pollen availability and diversity for bees.

This study was granted by the European Commission, Directorate General Environment: No 09.200200/2021/864096/SER/ENV.D.2.

Content

Accounting

This is a general guideline for the use of honey bee colonies for biomonitoring and how to organise an apicultural citizen science study. The study objective determines the toolbox for the beekeepers. This guideline has been designed and tested for its practical and scientific merits in the INSIGNIA study of 2018-2021 and subsequently in the INSIGNIA-EU study of 2022-2024. This has resulted in a best practice guide for apicultural citizen science studies for biomonitoring with honey bee colonies.

Reading guide

This guideline consists of:

1. The *rationale* with the background of the use of the honey bee colony as a monitoring tool.
2. *Study set-up.* This chapter describes the outline of the study, linking relationships, research integrity, data integrity, and communication processes.
3. *Protocols* for the beekeeper in the role of Apicultural Citizen Scientist. In this Second Edition, June 2024, of the Guideline, the protocols for biomonitoring for pesticides, microplastics, air pollutants, heavy metals and biomonitoring for pollen diversity, with their specific toolboxes, are presented. The format of the protocols for different subjects is identical, to facilitate combining protocols for specific monitoring subjects in a study plan.
4. *Methods* applied in the laboratories for pesticide residue analysis, microplastic analysis and identification, Polycyclic aromatic hydrocarbons, Volatile organic compounds,m heavy metals and pollen ITS2 metabarcoding, statistical methods, and the description of the exposure risk and pollen availability models.

Rationale

Content

Honey bees for biomonitoring

The honey bee

The honey bee (*Apis mellifera* L.) is an insect that lives in a colony and feeds itself with nectar for carbohydrates, pollen for proteins, minerals and fatty acids, and water for minerals and hydration. The larvae are fed with jelly produced by the nurse bees. In addition, honey bees collect resins from plant buds to cover the inside of their nest with propolis, which has antibiotic properties. The honey bee lives in a colony with one queen many thousands of female workers, and in the summer, some hundred male drones. The forage distance is usually a 1 to 1.5 kilometre (km) radius from the colony but a three (3) km radius is considered to cover the majority of its foraging area. Optimal foraging locations further away can be scouted and explored as well. This extra forage distance depends on the honey bee colony's characteristic to focus on the most profitable food sources in terms of energy. The moment a flight uses more energy than is brought back as food, it will stop. If energy profitability and foraging time are equal the energy profitability will determine where to forage. For nectar with a high sugar concentration the bee will fly larger distances than for nectar with a low sugar concentration. Honey bees are typical mass foragers with a high preference for mass flowering crops, preferably in large amounts. They show a high degree of flower constancy. This means that during a foraging flight, most commonly, flowers of a single species are visited. The colony communicates using dances that show the way to foraging areas discovered by the scout cohort, which is constantly looking for new nectar, pollen, water, and propolis sources. The foragers of a colony do not divide themselves homogeneously over the foraging area but may focus on part of the area. Other colonies may forage in another part of the area with some overlapping. The more diverse the landscape and the fewer mass flowering crops, the more the foragers will collect their food from less profitable sources until there is no profitable source available.

The honey bee colony as a biomonitoring tool

The honey bee colony is an excellent biomonitoring tool for airborne pollutants and systemic pesticides because of seven typical honey bee features, and the fact that honey bees are man-kept and managed insects.

1) The foraging bees collect nectar and pollen from flowers and simultaneously unintentionally pick up contaminant particles that have been deposited in the flowers, as well as when they do in collecting surface water;

2) There is an intense in-hive exchange of food and materials that bees carry on their body and in their honey stomach by auto- and allogrooming (physical contact) and by trophallaxis (direct bee-to-bee) food exchange;

3) The collected food and contaminants are accumulated in a central location, the hive;

4) The foraging area of a colony ranges from about 7 to 28 km^2;

5) The number of food-collecting worker bees (forager cohort) is about 30 to 40% of the total number of bees in the colony, which ranges from 30,000 to 50,000 bees;

6) The constant physical chemical kinetic conditions in the brood area (Temperature 34.5 ^oC to 36^oC and Relative Humidity (RH) of 90% to 95%, make the results of the in-hive passive samplers directly comparable and independent from local weather conditions and ambient temperature.

7) To maintain the constant in-hive brood nest conditions there is an in-hive passive airflow due to temperature difference between the brood nest and the edges, and there is a forced airflow. This forced airflow is caused by bees fanning out air with relatively high CO_2 and low O_2 concentrations and letting fresh air in passively.

Global sampling - Beekeepers Citizen scientist

Honey bee colonies are managed all over the world except for the poles. Therefore, they can easily be used for biomonitoring in any country. In terms of biomonitoring, the honey bees do the sampling, and subsequently, the beekeeper subsamples the colony to collect the accumulated information; in other words, the honey bees do the major part of the work for free, and it takes the beekeeper a short time to do the subsampling by exchanging in-hive samplers. Sentinel honey bee colonies can be placed anywhere where nectar, pollen, and water are available and accessible.

The worldwide network of honey bee colonies provides the ultimate means to use these colonies for bio-sampling, both for large-scale overview monitoring and also for detailed small-scale monitoring. Biomonitoring means structured repeated bio-sampling. Focusing on large-scale studies, the land use and colony density determine the "region" and not the national borders. Sampling points should be spread over the region, depending on the study objective.

Environmental pollution: Source-Path-Receptor (SPR) concept

The SPR concept describes the source of environmental pollution, its path through the environment, and its reception. In the INSIGNIA-EU study subjects for biomonitoring are denoted as target compounds. The target compounds are:

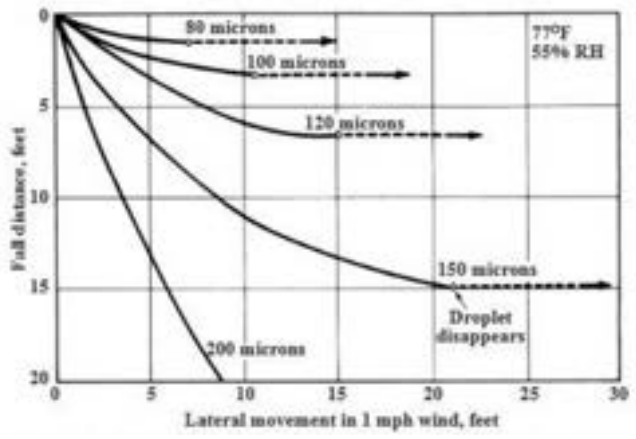


Figure 1. shows the impact of droplet size, humidity, temperature and wind speed on the lateral movements of spraying. (1 mph = weak wind = 1 Beaufort). Hofman and Solseq, 2004

Pesticides: synthetic chemicals produced to kill pests. The majority of pesticides are applied in agriculture. The term "pesticides" includes insecticides, fungicides, acaricides, herbicides, and Insect Growth Regulators (IGR). The pesticide SPR follows two routes. One is the direct route by direct overspraying of the honey bee during the pesticide application (Figure 1). This route should be limited due to advice to growers not to spray crops when bees are foraging. The second and dominant route is the indirect SPR. Pesticides are usually applied by spray, and part of the pesticide will land on the leaves and perform its protective function (Figure 1). The part will end on the soil where it becomes bound to (fatty) soil particles, but another part, aerosols will be taken away by the wind and may be deposited up to many kilometres from the application site. Part of the soil-bound pesticide particles may be relocated by soil erosion and soil dusting and be deposited far away from the application site. Pesticides in the air will be deposited and collected unintentionally by the honey bee foragers, regardless of the flowers they have

visited. Some synthetic pesticides, notably neonicotinoids are directly applied in the soil via coated seeds. These hydrophilic pesticides will partly enter the plant to protect it from biting insects, but may also end up in the pollen and nectar and be collected there. Another part of chemicals applied in this way may enter the surface water and be collected there.

Biocides are synthetic chemicals produced to control living organisms. For instance, anti-mosquito chemicals are biocides. The SPR route is often via the water and direct contact of the beekeeper with the beekeeper's equipment.

Microplastics are tiny plastic particles, smaller than 5 millimetres that result from both commercial product development and the breakdown of larger plastic items. They can be harmful to the environment. Microplastics do not break down chemically but instead break to form ever smaller particles. An important SPR route is synthetic fibres in clothes. These microplastics will, for example, enter the ambient air by the wearing of clothes, and their washing and drying. Part of the microplastics load will enter the hive via the airflow. Part of the microplastics will also be airborne deposited on flying bees or collected by them from flowers by the foragers.

Metals occur naturally in the soil as trace elements. Biota requires metals in trace quantities for vital processes. Increased quantities are the result of human activity such as fossil-fuelled transport, combustion, mining, metallurgical processes, agricultural activity, and leaching from metal structures. Increased metal concentrations may disturb biological processes, for example by interacting with the sulfhydryl group of enzymes and so inhibiting enzyme catalytical action. The SPR route is via contact of airborne deposition of soil particles and propolis, the sticky resin mass produced by flower buds and collected by bees.

Volatile Organic Compounds (VOC) and Polycyclic Aromatic Compounds (PAC) are serious air pollutants. VOC emission is often anthropogenic as the result of traffic and chemical industry, PAH emission is also anthropogenic as the result of combustion of organic matter, for example in biomass incinerators. The SPR route of these gaseous compounds is via the forced airflow inside the hive.

Pollen is not a contaminant, but is part of our biomonitoring program as a proxy for plant biodiversity and to provide information about where the bees are foraging. Pollen is collected by the forager pollen-collecting cohort. Pollen is stored in cells inside the hive, and the majority is eaten within days by the nurse bees to produce jelly for the brood. Pollen can be collected with a commercially available pollen trap.

The link between the honey bee colony and biomonitoring is the comprehensive interaction between the colony and the environment. Bees depend entirely for their food and in-hive resistance on the nectar, pollen, water, and propolis available in an approximately three to six-kilometre radius (28 to 113 km^2) around the colony. A smaller radius of about 1 to 1.5 km (3 to 7 km^2) is preferred and longer foraging distances have also been recorded (Garbuzov et al., 2015; Beekman & Ratnieks, 2000). Along with this collection, and due to the bees' behaviour on the flowers, contaminants, both solid in small and ultra-small particles and as a liquid, are inadvertently collected and brought into the hive. In the foraging season of the colony, about a quarter of the colony forages on average ten times a day and visits multiple flowers (Steen, et al, 2012; Winston, 1991). This is the power of the big numbers that the honey bee colony offers with its large forager cohort and the accumulation of what is collected in a central spot, the hive.

As a result, the colony reflects both the nutrient availability as well as the environmental pollution status of the foraging area (Chauzat et al., 2006; Conti & Botrè, 2001; Bromenshenk et al., 1985; Steen, et al., 2016). The pollutants collected are not restricted to pollutants in nectar, pollen, water, and propolis taken from the soil by the plants themselves, but also include pollutants deposited on the flowers, leaves, buds, and surface water by airborne processes, soil dusting, dust deposition and by human activities such as the application of plant protection products.

Bee-monitoring matrices

The basics of the use of the honey bee colony in biomonitoring is the non-intended influx and accumulation of pollutants in the colony by the worker bees and the subsequent sampling of the colony by the citizen scientist beekeeper with passive in-hive samplers and hive products.

A matrix is a substance that carries e.g. xenobiotics or plant pathogens and is not altered by these xenobiotics or micro-organisms. For example, nectar is a matrix for systemic pesticides and pesticide molecules deposited directly in the flower during spraying or as a result of drift. Pollen is a matrix for particles and molecules that are physically or chemically bound to its surface as a result of direct spraying or airborne deposition. Beeswax is a matrix for lipophilic molecules that enter the colony via nectar, pollen, or directly via contaminated bees. Propolis is a matrix for metals, both natural and as a result of airborne deposition to the sticky surface. The honey bee itself is a matrix for pollen grains, plant pathogens, and non-biological particulate matter stuck to the branched hairs and food-collecting parts of the bee. The pollen forager is focused on combing as much pollen as possible into the corbiculae, and exhibits frequent auto-grooming during foraging, whereas the nectar forager exhibits non-frequent auto-grooming.

Bee matrices' efficacy for monitoring of pesticides

In a review by Johnson et al. (2010), it is clearly shown that pollen is by far the most important vehicle for pesticides to enter the colony. Pooling together all data in their paper, the median amount of pesticides in wax, pollen, bees, and honey was 36, 61, 2, and 0.1 ppb respectively. In the studies where wax and pollen were analysed simultaneously (from the same colony), the median amount of pesticides in wax was 31 and in pollen 108 ppb. This difference in efficacy can be explained by the routes described. Honey is a poor matrix because of the quick dilution of nectar in the sprayed flowers, the in-hive dilution with non-contaminated nectars, and the temporal in-hive mixing of the nectars during honey processing (Schatz & Wallner, 2009). Also, bees themselves are not the most efficient matrices for biomonitoring because of autogrooming during the collection flight, in-hive auto- and allogrooming, detoxification in the alimentary tract of the bees, and the in-hive trophallaxis.

The practice and nature of sampling by the different matrices differ significantly. Every in-hive bee sample is by definition a mixture of all age cohorts of bees, which all have different levels of exposure (Steen et al., 2012). Forager bees are directly and the in-hive bees are indirectly exposed by trophallaxis and physical contact. Also, hive entrance samples comprise bees that have visited different locations (Visscher & Seeley, 1982; Waddington et al., 1994). Wax is a rather efficient matrix. However, it has the disadvantage that it is selective in its binding capacity and it is, because of strong binding and the long in-hive residence time of months to years, an archive from which it is impossible to distinguish between recent and older pesticide influxes.

Figure 2. Impression of collected pollen.

Pollen is the most efficient bee matrix (Figure 2). Working with trapped pollen, there is a tension between the duration of trapping and the impact this has on the colony; trapping for too long may have a negative impact on the colony, whereas short activation of the pollen trap will result in snapshot data as the flowering period of most flowers varies from only days to some weeks. Because of non-invasive sampling considerations and of aiming to affect the biomonitoring tool "honey bee colony" as little as possible, activation of the pollen trap should be limited to a maximum of four days, if weather conditions are such that no pollen is collected in a single day of collection. On the other hand, beebread is the accumulation of collected pollen. As most beebread is consumed in 2 weeks, it mostly holds recent information. However, beebread consumption depends strongly on the colony's condition, particularly the brood status (Crailsheim et al., 1992; Roessink & van der Steen, 2021). As beebread storage is randomly distributed in the colony over the edges of the brood nest, it is very time-consuming to collect this material, and it also has a disturbing impact on the colony. Furthermore, it is much more invasive and time-consuming to collect compared to pollen trapping.

Non-biological in-hive passive samplers and colony samples

To meet the requirements for non-invasive sampling of the honey bee colony, we used APIStrips for non-polar pesticides, APITraps for microplastics, silicone bands for the air pollutants volatile organic compounds (VOCs) and Polyaromatic Hydrocarbons (PAHs). For the colony samples, we collected fresh honey directly from the frame for polar pesticides and propolis for metal analysis.

APIStrip

Figure 3. APIStrip being collected from hive.

The APIStrip is an in-hive sampler made of a polystyrene strip covered with Tenax (a porous absorbent polymer; Figure 3). It is placed between two combs in the centre of the colony. The Tenax is very firmly bound to the plastic strip and binds pesticide molecules that circulate inside the colony on the bees' exterior, as loose particles and in gas form. In environmental research, these types of non-biological matrices are referred to as "passive samplers". The fabrication, and preparation for analysis of the APIStrips are described in Murcia-Morales et al. (2020). The mode of application is presented in this guideline in the Picture manual. The APIStrip is used to detect non-polar and semi-polar pesticides.

APITrap

Figure 4. Impression of the APITrap, note the hinge and swivels to close the trap when in use.

The APItrap is a sticky board in a gauze box, which is inserted in a frame (Figure 4). It is the in-hive passive sampler for microplastics, that circulate in the colony. The in-hive air is equal to the out-hive air. The gauze cage measures 30 x 20 cm inside. This size fits almost all frame sizes applied in the EU. The APITraps are placed next to the brood nest.

Silicone bands

Figure 5. Impression of silicone wrist band on frames.

Silicone wristband, composed exclusively of polydimethylsiloxane, (PMDS) binds air pollutants (Manzano et al., 2019, Samon et al., 2022).The silicone wristbands are applied as in-hive passive samplers to detect Volatile Organic Compounds (VOC) and Poly Aromatic Hydrocarbons (PAH) in the circulating in-hive air. The silicone wristbands are placed on top of the brood frames (Figure 5).

Colony matrices

Honey and propolis

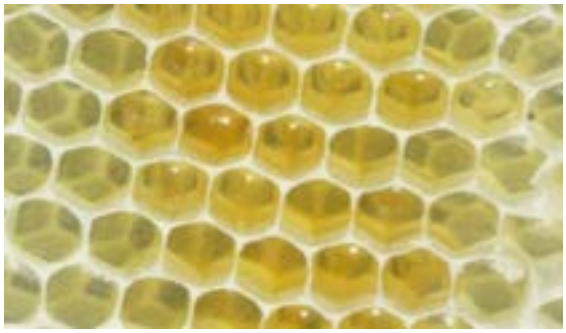


Figure 6. Honey comb.

Polar pesticides dissolve in honey and are not bound to the beeswax. To detect these pesticides we sample fresh honey from fresh combs (Figure 6).

Propolis was collected with silicone grids placed on top of the frames in the upper box (Figure 7). A bit of draft was induced by leaving a small opening between the grip and the hive cover to stimulate the bees to cover the grid with propolis.

Figure 7. Propolis grids after incubation on the hive.

Beekeeper citizen science

Citizen science is research, conducted partly or completely by citizens and non-professional scientists, often in collaboration with and under the direction of professional scientists and scientific institutions (WUR, Wikipedia). Apicultural citizen science is citizen science focused on beekeepers, who, in this project's case, sample their colonies or make regular recordings of hive parameters such as the number of bees, colony weight, foraging periods, etc. The main conditions for successful citizen science in general must be clear, unambiguously interpretable instructions about sampling, storage, shipping, etc. For apicultural citizen science, colony conditions, and beekeeping practices are additions to the general main conditions. Commitment of the beekeeper and support for the beekeeper are also essential. Citizen science is needed for large scaled research. In science and therefore also in citizen science, all parts of the chain from sample to result are equally important.

In-consortium communication

The in-consortium communication has been organised through monthly online meetings, which have proved extremely useful for keeping the study on track, strengthening the network, and for planning and coordination (Figure 8). The INSIGNIA cloud was used to host a blog, along with relevant documentation, publications, and administrative documents.

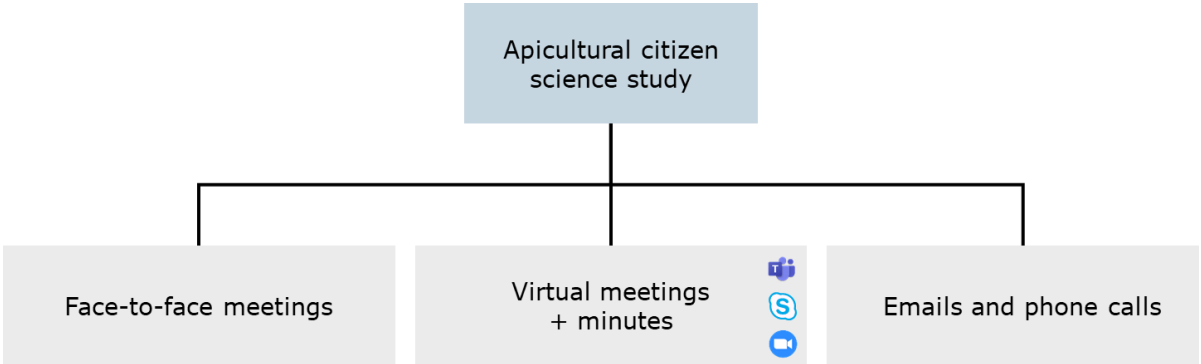


Figure 8. In-consortium communication

Beekeeper Instruction - tutorials and other means of instruction

From the National Coordinator (NatCo) to the Citizen Scientist beekeepers (CS)

It is the responsibility of the Apiculturalist Coordinator (AC) to ensure that the Citizen Scientist beekeepers (CS) are fully trained and understand exactly what they have to do. In the INSIGNIA study, instruction was achieved through an integrated approach via a written CS Picture Manual, instructional tutorial videos, face-to-face meetings, virtual training sessions, emails, and telephone calls (Figure 9). The CS were recruited for their high level of general beekeeping experience and competence, but INSIGNIA required them to learn additional skills, especially with the care taken in collecting, labelling, and storing samples.

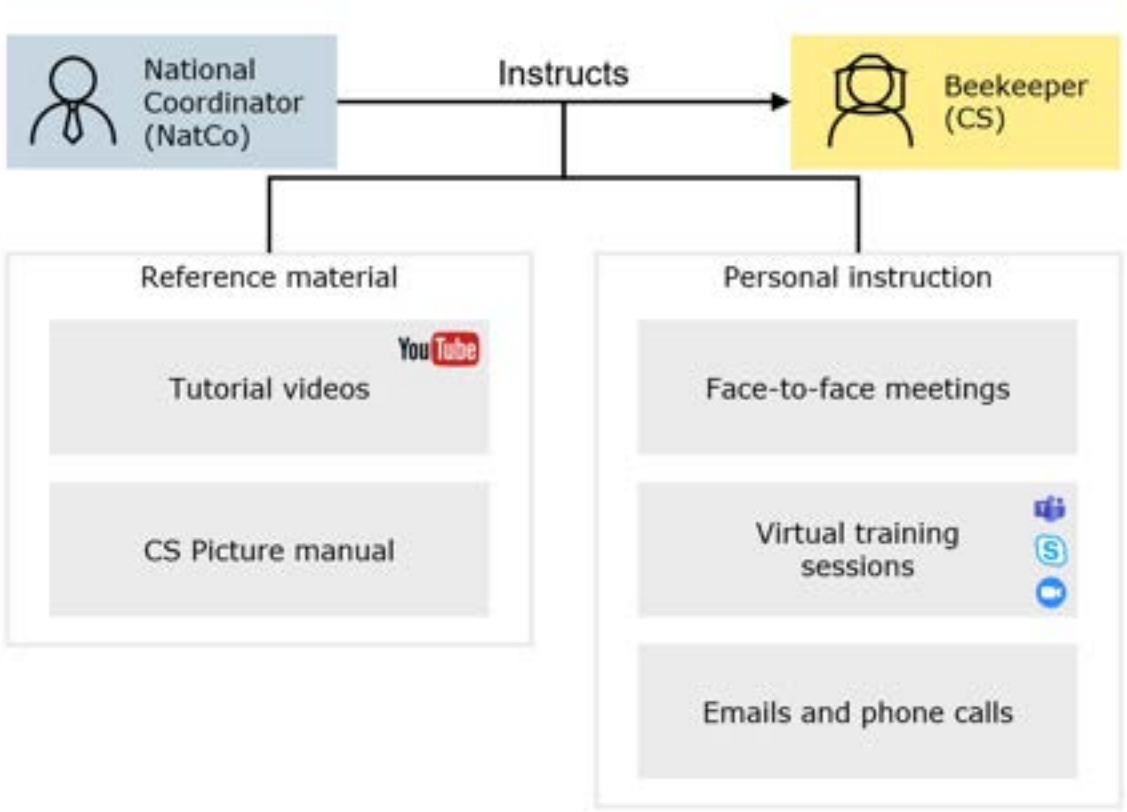


Figure 9. Communication from NatCo to CS.

NatCo cafés and face to face communication

The NatCo café sessions, organised once a month and intended to keep the participating beekeepers updated, were complementary events to the regular contact of the beekeepers with the NatCo. The subject matter of these meetings included objectives of the study, updates on results, and/or whatever was relevant to citizen science. These proved to be useful for helping to ensure the ongoing motivation and commitment of the NatCos and to allow the information/communication flow from NatCos to the participating apicultural citizen scientists. The individual NatCo - coordinator meetings were virtual meetings held and acceptance of the invitation was free. All NatCos accepted the invitations.

Communication with stakeholders through social media and journals

In a beekeeping citizen science study such as this, the stakeholders include other beekeepers, the general public, and decision-makers. In INSIGNIA, a strategy for dissemination to these stakeholders was established from the outset (Figure 10). The website https://www.insignia-bee.eu was set up as one of the first study deliverables, to become the core of the strategy. As well as information about the study and the members of the INSIGNIA consortium, the website had an active blog and links to social media including Instagram, Facebook, and YouTube. The regular blog entries and social media posts were written by many members of the INSIGNIA team and were intended to inform both the participating CSs and other stakeholders such as other beekeepers and the general public.

Lectures, presentations, and talks about the study have been given by various team members at a range of meetings, from international platforms such as the Apimondia International Apicultural Congress in Canada, the Apimondia Bee Health Symposium in Italy, and the COLOSS Asia Conferences in Thailand and Japan, through national and regional beekeepers meetings to local meetings of individual beekeepers' associations. Several articles arising from the study have been written and then translated into many local languages for publication in many local beekeeping journals.

A regular series of "Notes for stakeholders" have also been produced on a range of topics, and were distributed via a mailing list comprising beekeeping associations, environmental journalists, editors of beekeeping magazines, relevant government ministries, and relevant non-governmental organisations.

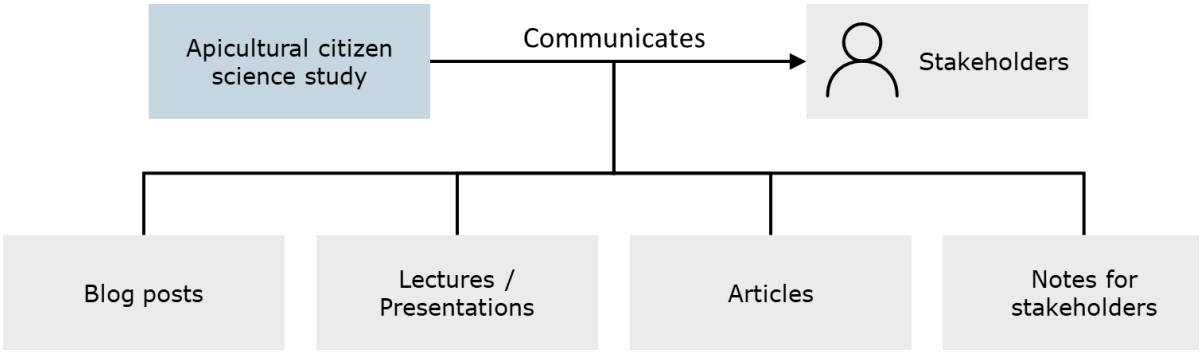


Figure 10. Consortium outward communication

Study set-up

Tasks - responsibilities - workflows of the Study Director, National Coordinator, and Data Curator

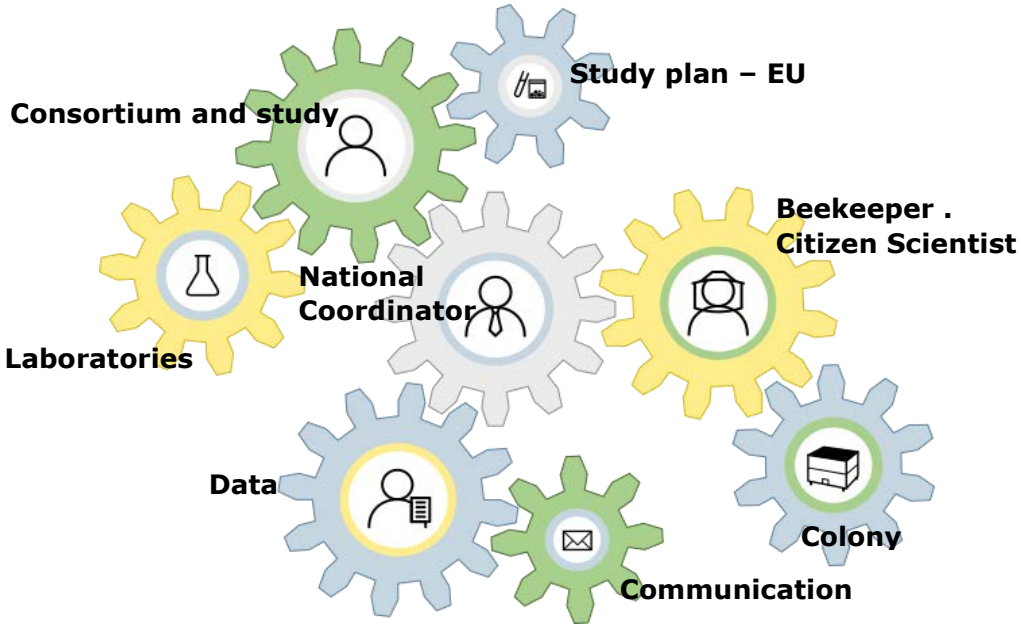


Figure 11. Schematic overview of the INSIGNIA action.

Introduction

The Study Director (SD) has the responsibility for the study in all its aspects. The SD coordinates and participates in writing the study plan with the National Coordinator (NatCo), the laboratory specialists, statisticians, and modellers involved, according to scientific standards. He/she is the pivot between the sponsor, the National Coordinators, the analytical laboratories, the statisticians and modellers and the stakeholders. This function comprises responsibilities for study coordination, study progress, study quality, coordination of CS instruction, data integrity, data storage, reporting to the sponsor about study content and finances, organising the communication routes, both internal and external, and the final study report.

The National Coordinator (NatCo) is responsible for the practical implementation and the course of the citizen science study in a particular country. He/she is the pivot between the SD and the Apicultural Citizen Scientist (CS). The function responsibilities comprise the recruitment of the apiculturalist CS, the content and assembly of the toolbox, the sample quality in terms of labelling, storage, transfer from CS to NatCo and subsequently from the NatCo to the laboratories, the sample flow recording and layman reporting to the participating CS.

Data integrity and communication, both internal and external, are key processes in the apicultural citizen science study. Although the SD is responsible for these processes and the NatCo for the practical implementation, these processes have such a comprehensive range that specific tasks should be delegated to control the process. The Data Curator's

tasks and processes are described in a separate paragraph "Data integrity and communication".

The study plan process is worked out in the paragraph "Study Plan". The role of the SD the consortium and NatCo is depicted in a flow chart and process descriptions are shown in the paragraphs "Study Director" and "Apiculturalist Coordinator".

List of abbreviations

SD: Study Director
NatCo: National Coordinator
DC: Data Curator
CS: Apiculturalist Citizen Scientist

Study Plan

Every research study starts with a question or hypothesis, formulated in a study proposal. A key process to turn a study proposal into a study plan is the question articulation, in other words, what exactly is the question or hypothesis? Next, this defined research question must be translated into a study plan.

In practice, a sponsor formulates a research question and an institute or consortium will, after verification of the definite study question, make the study plan.

In apicultural citizen science, the study plan process is an interplay between an SD, an AC, and the scientific specialists (e.g. chemists, geneticists, molecular biologists, statisticians, modellers, and environmentalists) involved in the study. The AC is the beekeeping specialist. In this interplay, the practical possibility, scientific standards, and financial and manpower possibilities are weighed, to come to a balanced study plan with timelines, responsibilities, and processes. This study plan is submitted to the sponsor who posed the research question.

In an apicultural citizen science study plan, the key process is the collaboration and interactions between the SD and the AC, a strict sampling or recording timeline, a well-defined protocol for the beekeepers, the actual citizen scientists, and an adequate toolbox for the beekeepers to do the job. The study plan comprises the layout of the study, the study course, the responsibilities on all levels - from the beekeeper, via the AC through to the SD - and the processes like the instructions, communication, recordings and reporting, research integrity, and overall data integrity.

*Research integrity**

Research integrity means conducting research in a way that allows others to have trust and confidence in the methods used and the findings that result from this. Researchers are expected to abide by the standards of research integrity:

- Honesty in all aspects of research including
    - Presentation of research goals, intentions, and findings;
    - Reporting on research methods and procedures;
    - Gathering data;
    - Using and acknowledging the work of other researchers;
    - Conveying valid interpretations and making justifiable claims based on research findings.

- Scrupulous care, thoroughness, and excellence in research practice
  - In performing research and using appropriate methods;
  - In adhering to an agreed protocol where appropriate;
  - In drawing interpretations and conclusions from the research;
  - In communicating the results.
- Transparency and open communications
  - In declaring conflicts of interest;
  - In reporting of research data collection methods, including the re-use of data collected for other purposes;
  - In the analysis and interpretation of data;
  - In making research findings widely available, including sharing negative results as appropriate;
  - In presenting the work to other researchers and the general public.
- Care and respect for
  - All participants in and subjects of research, including humans, animals, the environment, and cultural objects;
  - The stewardship of research and scholarship for future generations.

** The definition of research integrity presented is quoted from the University of Bath, UK (Definition of research integrity (bath.ac.UK))*

Study Director - tasks and responsibilities

The overall key task of the SD is to monitor the progress and the quality of the study in all aspects, to ensure data integrity, and to take action when needed. He/she is also responsible for the coordination of the practical and scientific output and paper-writing process.

The SD function towards the sponsor comprises:

- Reporting according to a reporting plan / deliverable list, of
  - the study progress reports
  - the final report
  - the financial reports
  - other agreed deliverables
- Reporting in writing any deviations and amendments to the study plan, including the rationale of any amendment and the implications for the study plan and outcome;
- Monitoring the study progress and quality.

The SD function towards the AC includes:

- Frequent contact about all relevant study matters and study updates;
- Organizing regular (virtual) meetings and making available the minutes on the study cloud or other relevant accessible storage tools;
- The internal communication in consultation with the Communication Curator;
- The coordination of the instruction of the beekeepers with all relevant tools;
- The direct contact with the beekeepers via the AC;
- All data integrity matters and updates concerning recordings and reporting in consultation with the Data Curator;
- The organization of plenary meetings.

The SD function towards the analytical laboratories, statisticians, and modellers comprise:

- Frequent contact about the analytical process;
- Clarifying and resolving through discussion any issues that arise;
- The organisation and checks of the uniformity of the sample identifications;
- Providing the statisticians and modellers with the relevant data files, being entries of the data warehouse, in collaboration with the Data Curator;
- The organisation of the plenary meetings.

The SD function towards the stakeholders comprises:

- Organising the communication with the stakeholders, in consultation with the Communication Curator, via e.g. notes and bee journal articles;
- Being the addressee for the stakeholders.

National Coordinator - tasks and responsibilities

The NatCo is responsible for the practical implementation of the study plan and is the pivot in the Apicultural citizen science study. In practice, the NatCo organises the local part of the study, that is the actual sampling in a particular area/region / country and all requirements for adequate sampling. The NatCo is a co-operator with the Data Curator. At the end of the study, NatCo will communicate the results with his/ her beekeepers in their country.

The NatCo function towards the beekeeper comprises:

- Recruitment of the beekeepers;
- Assigning a CS number to each beekeeper;
- Providing the beekeepers with all information on the "why and how" of the study and its specific goals;
- Instruction about the study protocol with all appropriate tools;
- Providing the beekeepers with the toolbox to conduct the sampling and recordings;
- Being the contact person for the beekeepers;
- Informing the beekeepers about the study's progress and results.

Apiculturalist Coordinator and Apiculturalist Citizen Scientist (NatCi and CS)

Recruitment

The beekeepers who will participate in the study as citizen scientists (CS) are selected and invited by the NatCo.

The basic conditions for a CS to participate are:

- Internet access and an email address;
- Having a minimum of 2 non-migrating honey bee colonies (well overwintered and of average strength as generally accepted in the particular region);
- Having time within the pre-determined sampling rounds to do the appointed pre-, post-, and actual sampling activities;
- Having an awareness of the environmental surroundings;
- Being reliable in sticking to the agreements and appointments;
- Being a skilled beekeeper;
- Having an affinity to science, as a pre-condition for involvement as a CS.

Information and contact

The quality and reliability of the data/results concerning pesticides, pollen, or anything else sampled depends on the quality of the sampling. This constitutes the basis of any monitoring study. Therefore, the CS must be well informed about the "why and how" of the study set-up, its specific goals, the study progress, and study results. It is up to the NatCo to inform the CS about these and to place the role of the CS in perspective as a key participant in the citizen scientist study. The basic condition for the relationship between the AC and CS to be successful is open and clear communication. All communication channels available in the study are at the disposal of the NatCo, including NatCo cafe meetings, WhatsApp groups, Signal groups, Facebook groups, or whatever is applicable to keep in touch with the CS. These will assist if questions arise from the CS, and will help to ensure the smooth running of the study and the continued and essential engagement of the CS.

Instruction of the CS

The importance of correct sampling, sample storage, and sample transport is explained in the previous subparagraph. This can only be achieved by crystal clear instructions.

These instructions are int the next paragraph "Picture manual"

Toolbox

Including sample tools, pre-printed labels, and return-transport boxes/ envelopes.

Consortium members prepare the toolbox and send it to the NatCo. The NatCos will distribute this among their beekeepers

A toolbox for bee-monitoring by the honey bee colony, designed for the CS contains:

- Tubes, bags, and pencils. The pencil is mandatory because all ink-based pen writing can be erased by alcohol, water, or other solvents;
- Non-biological passive samplers/matrices;
- Supporting material for storage;
- Pre-printed sample labels. This key aspect is elaborated in a sub-paragraph "Instruction for sample labelling";
- Envelopes/boxes for sending back to the AC by the CS. This key aspect is elaborated in a sub-paragraph "Samples from the CS to the AC".
- Whatever else is needed for proper sampling

Data integrity and communication

Both data integrity and external communication are responsibilities of the SD and the practical execution can be assigned or delegated to a participant; the Data Curator (DC).

The process of "data integrity" comprises that:

- SD guards data safety and integrity of the data warehouse;
- SD assigns data management tasks to the DC;
- NatCo checks labels and samples after return from the beekeeper;
- NatCo records samples and codes in his/her AC journal;
- NatCo sends the samples with the original label to the labs;
- DC uploads the results of the lab analyses;
- DC compiles data sets for statistical analysis and modelling.

The process of external communication comprises that:

- The communications section of the consortium: makes notes for external contacts;
- organises the writing of articles for bee journals based on the input by the SD and, via the SD, the laboratories, input by AC, and input of the beekeepers via the AC;
- takes care of the website and website upload;
- takes care of / feeds the social media;
- makes, or has made, instruction video and other visual instruction material;
- organises that participants contribute to the study blog, notes, and other outputs.

The Citizen science section of the consortium makes the format of the picture manuals, to be translated/adapted by the NatCo.

Communication

Automated communication

The texts for automated communication with CSs will be translated into the required languages, depending on the participants in the study, by the NatCo. The online survey setup allows the addition of other languages as needed. The NatCo makes sure that every participant in their country receives an online survey link for each sampling round. The NatCo is responsible for the process of data provision by the participants in consultation with the DC. Within the study platform, there is a document containing all online survey questions /content that has to be translated into the languages by the NatCo.

The online survey questionnaire

During the sample collection periods, the NatCo stays in contact with the DC doing the online survey with the citizen scientists by the citizen science section of the consortium. This includes the process of reminding the participating citizen scientists before each sampling date. The NatCo has the end responsibility for the study process in his/her country.

The online survey questionnaire serves two purposes. Firstly it confirms which dates the CSs have collected samples, and secondly, it is a way to collect additional data about the conditions in the apiaries over the sampling season.

There are various systems available for making questionnaires. In the INSIGNIA project, the online LimeSurvey survey software was used, but there are other systems such as Google Formula and Survey Monkey. It is the responsibility of the DC to set up the questionnaire and send out emails to all CSs for each sampling round. Only if individual CSs fail to answer the questionnaire will the AC be involved, and they will then make direct contact with the relevant CS to ensure that the questionnaire will be answered.

The online survey tool LimeSurvey was a good choice for the INSIGNIA project since it can issue individual invitations with a "token". This means that during the whole process, it is possible to keep track of all the responses from the beekeepers, issue reminders, and notify the AC to contact non-respondents. LimeSurvey also has a very good mapping facility, for determining the GPS position of each participating apiary. Information to be collected includes:

- Which samples have been collected in the sampling rounds, and the exact dates for the start and end of the subsequent sampling rounds?
- Information about the size of the colonies based on frame size and number of occupied bee lanes. This is important, as larger colonies will be expected to collect more food, nectar, and pollen, and consequently also pollutants.

- Finally, the questionnaire allows the CS to report (chemical) varroa treatments and other relevant information.

The DC collects the answers and adds information to the data warehouse. The DC also ensures that sampling dates from the questionnaire and the physical samples are consistent.

Picture manual / Instruction manual

This chapter show the INSIGNIA-EU 2023 picture manual in English. The translated picture manual in 22 other EU languages are available separately.

Funded by
the European Union

INSIGNIA - EU

PREPARATORY ACTION FOR MONITORING OF ENVIRONMENTAL POLLUTION USING HONEY BEES
N° 09.200200/2021/864096/SER/ENV.D.2

DELIVERABLE 3.4: Final version: Apicultural Citizen Scientist Picture Manual (available in 23 languages) (M24)

Lead: 2 - UoGraz

December, 2023

INSIGNIA-EU Picture Manual for the Citizen Scientist

INSTRUCTIONS FOR USE – Full study 2023

Funded by the European Union

Funded by
the European Union

INSIGNIA-EU

Picture manual

Version 02, March 2023

Communication tools

National Coordinator	xxINSERTxx
Email	xxINSERTxx
Phone number (contactable during sampling periods -Thursday to Sunday 9 am – 5 pm)	xxINSERTxx
WhatsApp group (optional)	xxINSERTxx
Website	https://www.insignia-bee.eu
YouTube channel	INSIGNIA-EU
Facebook	@insigniabeeEU
Instagram	@insigniabee_eu
Author picture manual	Kristina Gratzer

Thanks for your valuable contribution! We wish you interest and fulfilment with your citizen science study.

Support/Video Manual

You will find accompanying instruction videos and constantly updated blog entries on our INSIGNIA YouTube channel and on our website, respectively.

During the active sampling periods, the national coordinator is available under the above-mentioned phone number. If you have any questions or other requests outside the sampling periods, please write an email to the national coordinator. Further, you are also invited to join our national WhatsApp group to exchange your experiences with the other citizen scientists of your country.

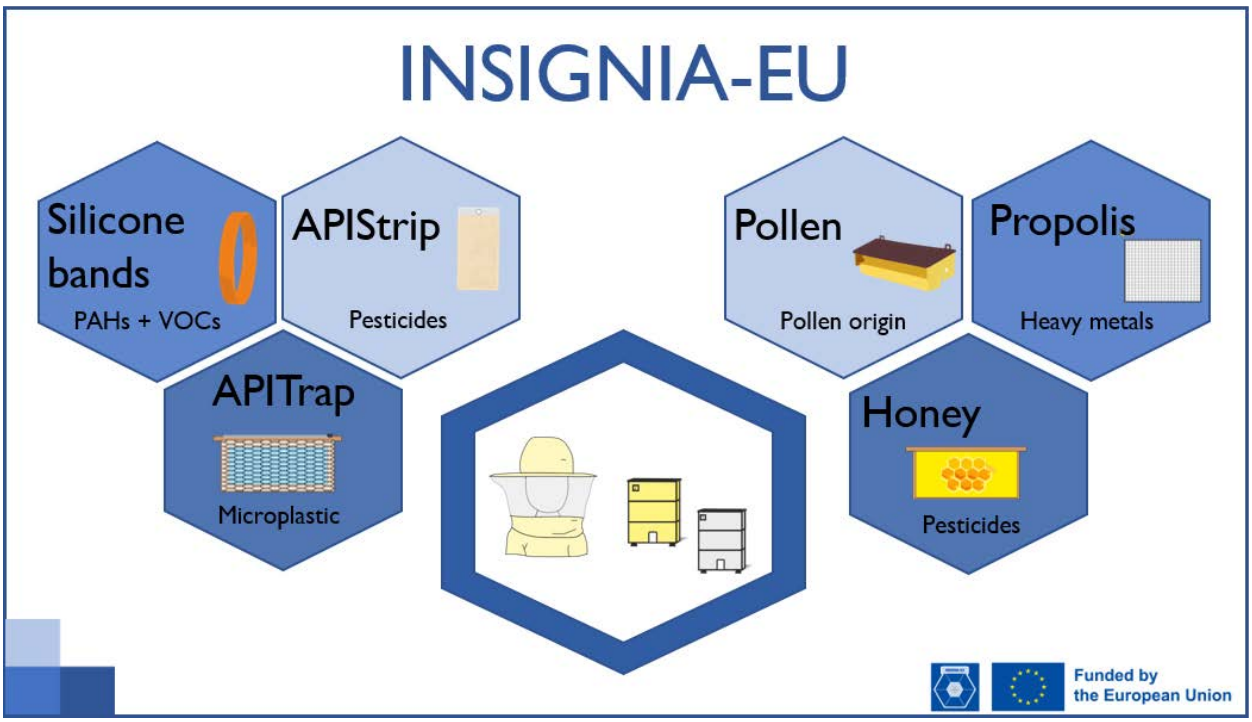
YouTube

Blog

The INSIGNIA-EU study 2023

Purpose of INSIGNIA-EU and study set-up

The INSIGNIA-EU study aims to provide information on environmental pollution on a pan-European scale that is covering all European countries. In 2023, 5-20 apiaries in each of the 27 EU countries will be sampled simultaneously. The outcome of the project will provide the first standardised EU-wide monitoring of environmental pollutants with honey bee colonies. This will be achieved by a network of beekeepers, sampling their own colonies as citizen scientists (YOU). This citizen science sample collection approach is complemented by leading research organizations implementing state of the art analytical, statistic and modelling methods. The results will broaden our understanding of anthropogenic pollution in the European Union, allowing identification of high and low pollution areas. This should facilitate actions to improve the environment.



Time table

Sampling round dates					
Sampling Round	Month	Sampling day			
		THU.	FRI.	SAT.	SUN.
START	April	20.04.	21.04.	22.04.	23.04.
SR01	May	04.05.	05.05.	06.05.	07.05.
SR02	May	18.05.	19.05.	20.05.	21.05.
SR03	June	01.06.	02.06.	03.06.	04.06.
SR04	June	15.06.	16.06.	17.06.	18.06.
SR05	June/July	29.06.	30.06.	01.07.	02.07.
SR06	July	13.07.	14.07.	15.07.	16.07.
SR07	July	27.07.	28.07.	29.07.	30.07.
SR08	August	10.08.	11.08.	12.08.	13.08.
SR09	August = end	24.08.	25.08.	26.08.	27.08.

(A) First steps

Choose 2 bee colonies for the INSIGNIA-EU study

The best choice will be well-overwintered, non-migrating colonies of average strength:

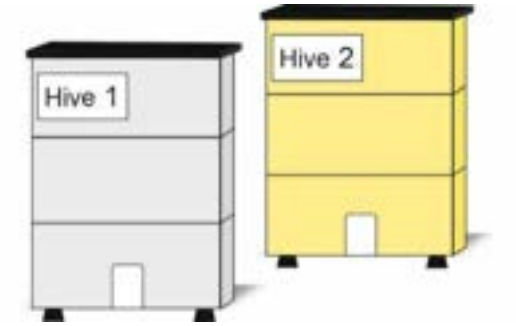
- Queenright
- All stages of brood present
- Adequate food storage
- No symptoms of diseases

The colonies must be situated at the same apiary and preferably near to your home.
If there are doubts, contact your national coordinator.

Test hives

Assign the two colonies the numbers 1 or 2 to clearly identify them during the whole sampling season.
You may use the stickers in your toolbox.

In case a bee colony has died, or has swarmed, you are able to exchange the colony with a different one from the same apiary.



Identification example of study colony 1:

Invitation

You will receive an **invitation email** from our project staff. Check if you can meet the requirements prior to confirming your study participation. Next, you will receive an invitation to complete an online questionnaire. Please follow the instructions to finish the data submission process.

Requirements for your participation

- ✓ Interested in science
- ✓ Internet access and an email address and ability to check emails regularly
- ✓ Telephone number
- ✓ Two honey bee colonies (see above)
- ✓ Two pollen traps
- ✓ One clean teaspoon per sampling round
- ✓ A fork for honey sampling
- ✓ A fridge (+4°C) or a freezer (- 18 °C) and enough space in it (~volume of 1 shoebox)
- ✓ Time within the predetermined sampling rounds (twice a month from April to August 2023)
- ✓ Consider bringing a small garbage pin or garbage bag with you to every sampling round

The tool box includes the equipment you will need for your activity as a citizen scientist within the INSIGNIA-EU project. It contains materials for 2 bee colonies and all of the 9 sampling rounds. Check the materials for completeness.

Materials	Amount	Checklist
Translated picture manual	1	☐
APIStrip sampling (9 times: SR01-SR09)		
APIStrips	18	☐
APIStrips "bench-control"	1	☐
Paper clips	18	☐
Sample bags	11	☐
Labels	10	☐
Big zip bag to collect all 9 sample bags	1	☐
Pollen sampling (9 times: SR01-SR09)		
Cups - plastic (green lid)	9	☐
Silica gel bags (6 g bags, 2 per plastic cup)	18	☐
Tea bags	9	☐
Sample bags	9	☐
Labels	9	☐
Big zip bag to collect all 9 sample bags	1	☐
Honey sampling (4 times: SR02, SR04, SR06, SR08)		
Tubes - plastic (blue lid)	4	☐
Sample bags	4	☐
Labels	4	☐
Big zip bag to collect all 4 sample bags	1	☐
Propolis sampling (4 times: SR02, SR04, SR06, SR08)		
Envelopes or bags	5	☐
(small) Propolis grids (4 per colony)	32	☐
Labels	4	☐
Silicone band sampling (4 times: SR02, SR04, SR06, SR08)		
(small) Honey jars - glass	4	☐
Silicone bands (1 per colony)	8	☐
Sample bags	5	☐
Labels	4	☐
Parafilm strips	4	☐
Big zip bag to collect all 4 sample bags	1	☐

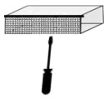
Tool Box

APITrap sampling (4 times: SR02, SR04, SR06, SR08)		
APITrap	1	☐
APITrap sheets - sticky (2 sticky APITrap sheets per APITrap)	8	☐
Envelopes C5 (1 for 2 sticky APITrap sheets)	5	☐
Labels	4	☐
Big zip bag to collect all 4 envelopes	1	☐
Stanley knife	1	☐
Additional material		
Sticker (hive allocation)	2	☐
Aluminium foil	1	☐
Gloves (package)	1	☐
Kitchen roll	1	☐
Pencil	1	☐
Cardboard transportation boxes	5	☐
Additional material from your home (not in the tool box)		
Clean teaspoon (pollen sampling)	1	☐
Clean fork (honey sampling)	1	☐
A wet and clean (paper) towel (honey sampling)	1	☐

Funded by the European Union

(B) Study start – installation of devices

The installation process of sampling matrices takes place between the **20th and 23th April 2023.** Please carefully read the following manual.



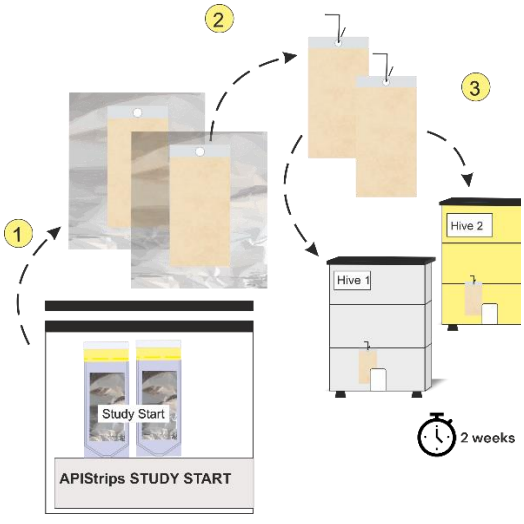
Install POLLEN TRAPS: Each test hive will be equipped with a pollen trap. Follow the specific instructions of the manufacturer for installing the trap. The pollen trap must be installed without additional gaps between the trap and the hive (bee tight). There are also plenty of YouTube Videos, if you need additional help. We will not provide any materials used for installation (screws, screwdriver, duct tape, …). After installation, make sure the trap is **not** activated.

From this point on use a pair of gloves from your tool box. You can use the same pair of gloves during the whole sampling procedure on the sampling day.

Place the APISTRIP with the label **BENCH CONTROL** at the place, you store your materials for beekeeping. Do not open it! This strip serves as a control for the lab. Leave it until the study ends.

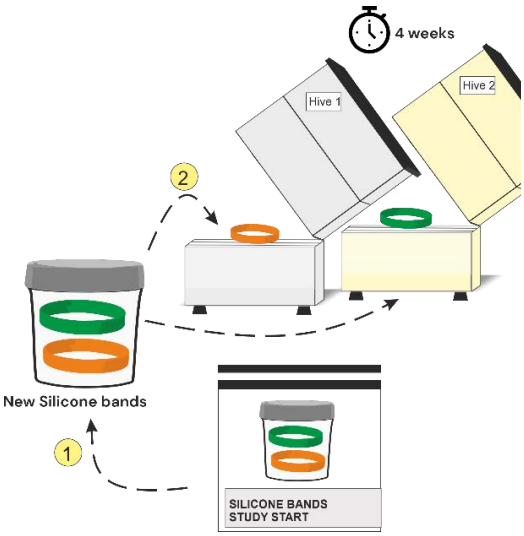
Put 1 APISTRIP in each hive:

1) For the installation process you will find 2 plastic envelopes containing the first 2 APIStrips (wrapped in aluminium foil) in your tool box (label "STUDY START").
2) Remember to wear gloves. Unpack one fresh APIStrip. You can touch the APIStrip on the not-coated area on top. Bend 1 paperclip into a "T" and attach it to the APIStrip. Use a thin twig to support the construction if necessary.
3) Insert the APIStrip in the middle of the broodnest (between frames) of **hive 1**.
4) Do not close the hive yet and insert the other matrices (silicone bands, propolis grids and the APITrap) first (see instructions below)! Afterwards, close hive 1 and repeat the steps for **hive 2.**
5) Leave the APIStrips in the hives for a total of **two weeks**. Store the plastic envelopes for further use, or even better, put them into the sample bag for SR09 (which is empty so far).



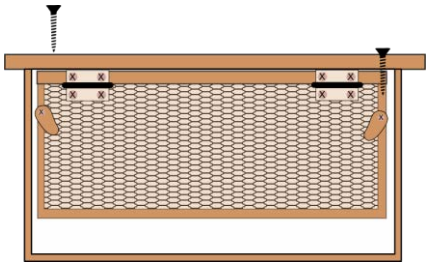
Put 1 SILICONE BAND in each hive:

1) Look for the SILICONE BANDS with the label STUDY START. You will find 1 small honey jar containing two silicone bands in different colours. Wear your gloves and open the jar.
2) Place 1 silicone band on the top of the frames in the middle of the brood nest of **hive 1** (e.g. above location APIStrip). If the room between boxes is not sufficient for the bands squeeze them.
3) Do not close the hive yet and insert the other matrices (the APITrap and the propolis grids) first (see instructions below)!
4) Afterwards, close **hive 1** and repeat the steps for **hive 2**. Leave the silicone bands in the hives for a total of **four weeks**. Store the jar for further use or even better, put them into the sample bag for SR08 (which is empty so far).

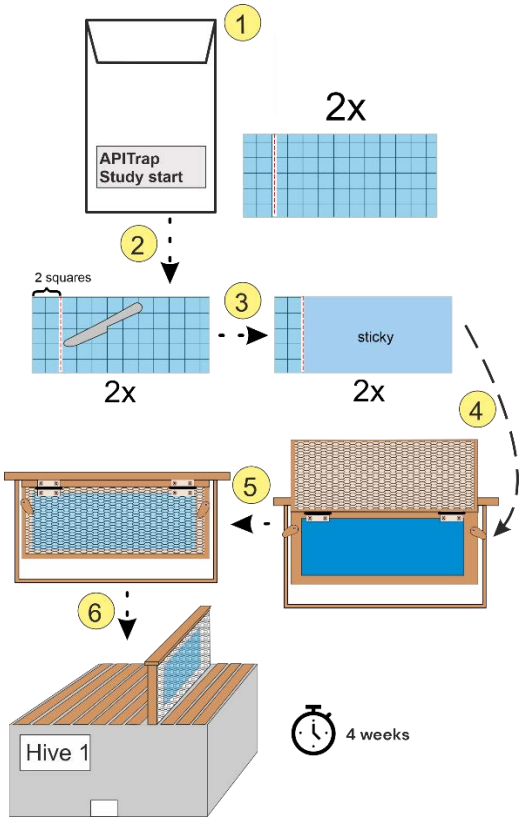


Put the APITRAP in hive 1:

The APITrap is basically a replacement for a frame next to the brood nest. Attach the APITrap to one of your empty frames by using a screwdriver and the enclosed screws. We recommend pre-drilling holes first before screwing in the screws.



1) Look for the envelope with the label "APITRAP STUDY START" and open it to find 2 APITrap sheets. Both sheets must be placed back-to-back in the trap to activate it.
2) Count two squares from the left (or right) and carefully cut the protecting backing paper with a (Stanley) knife from top to bottom. Repeat for the 2nd sheet. **Do not use too much pressure**!
3) Peel of the thin backing paper **only** from the **large area**. This results in a sticky zone and a "control" zone. Latter serves as a control in the lab during analysis. The sticky APITrap sheets are now ready for the APITrap.
4) Open the APITrap and carefully insert both sheets back-to-back. **Both sticky sides need to face outwards** (in the direction of the mesh).
5) Close and secure the APITrap. The APITrap is now **active**.
6) Take out the first frame next to the brood nest of **hive 1** in the lowest box and replace it with the activated APITrap. Leave it in for **four weeks** before replacing the sheets by fresh ones.



Note: If you prefer, you could cut the APITrap sheets at home and leave a dog-ear on one of the edges. If you do this, you only need to peel off the paper right before inserting the sheets.

Visit the INSIGNIA-EU YouTube channel, if you like to watch instruction videos.

Put 4 PROPOLIS GRIDS in each hive:

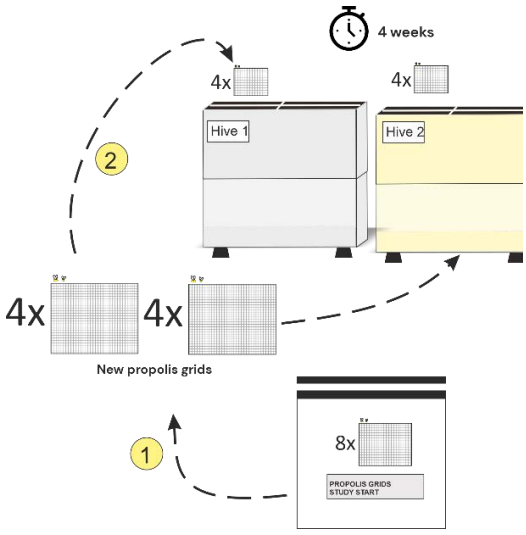
1) Look for an envelope with 8 PROPOLIS GRIDS with the label STUDY START. Use the same gloves and take out 4 propolis grids. The 4 grids equal one custom propolis grid, but we cut it to generate a better fit for the storage after sampling.

2) Start with **hive 1** and place the propolis grids on top of the frames of the upper box.

3) If you have successfully inserted the **APIStrips, the silicone bands, the APITrap and the propolis grids**, close **hive 1** and repeat the steps for **hive 2**.

4) Leave the propolis grids in the hives for a total of **four weeks**.

Note: Increased air ventilation above grids will result in higher propolis yields. Also, if you are working with 1-box colonies, make sure to cover ¾ of the top with propolis grids (cut them smaller if necessary) and use the remaining ¼ for the APIStrips and the bands.



The installation process is completed when you successfully set-up the following matrices:

Matrix	Per Hive	Per Apiary
Pollen trap	1	2
APIStrips in-hive	1	2
APIStrip "bench-control"	-	1
Silicone bands	1	2
Propolis grids (small)	4	8
APITrap	1	1

Important:

Not all matrices are sampled at every sampling round. Look carefully at the **sampling plan** on the next page and follow it for each sampling round.

If in doubt, contact your national coordinator.

Congratulations,

now you have time to rest and let the matrices and bees do their work…

The actual sampling starts in two weeks (04.05. – 07.05.2023).

(C) Sampling plan

Sampling Rounds	Sampling dates	Pollen	APIStrip	Honey	Silicone band	APITrap	Propolis
START	**20.04.-23.04.**	Install pollen traps	**Insert** 1 APIStrip/hive Total = 2		**Insert** 1 band/hive Total = 2	**Insert** 1 Trap in hive 1 Total = 1	**Insert** 4 small grids/hive Total = 8
SR01	**04.05.-07.05.**	Take sample	exchange				
SR02	**18.05.-21-05.**	Take sample	exchange	Take sample	exchange	exchange	exchange
SR03	**01.06.-04.06.**	Take sample	exchange				
SR04	**15.06.-18.06.**	Take sample	exchange	Take sample	exchange	exchange	exchange
SR05	**29.06.-02.07.**	Take sample	exchange				
SR06	**13.07.-16.07.**	Take sample	exchange	Take sample	exchange	exchange	exchange
SR07	**27.07.-30.07.**	Take sample	exchange				
SR08	**10.08.-13.08.**	Take sample	exchange	Take sample	Take sample	Take sample	Take sample
SR09	**24.08.-27.08.**	Take sample	Take sample				
Sample storage							

Funded by the European Union

Funded by the European Union

(C) Sampling by the citizen scientist

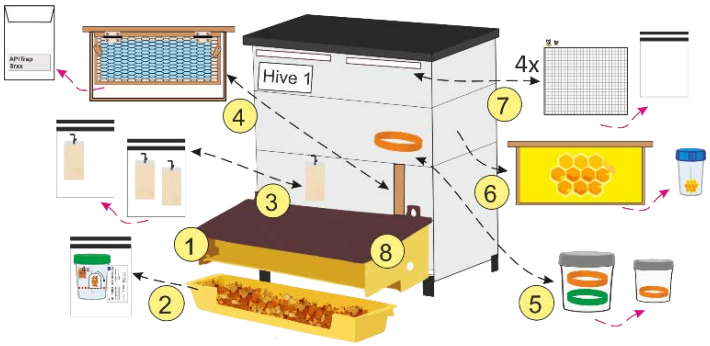
The first of the nine sampling rounds takes place between the **04th of May to the 07th of May 2023**. Please carefully read the following manual.

Prior to each sampling round

Invitation

Invitation email: prior to each sampling round you will receive an **invitation email**. Choose the one day during the specified 4-day sampling period with the best weather forecast, that fits best for you to collect the samples (table below). If you do not receive the email, please check your spam box or contact your national coordinator.

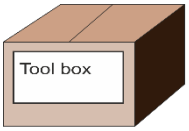
Sampling time window (example)				
Day	Thursday	Friday	Saturday	Sunday
Weather forecast & day job	working	free	free	free
Sampling		**24 hours before sampling** **(1)** close pollen trap (every SR)	**24 hours after closing pollen trap** **(2)** harvest pollen (every SR) **(3)** exchange APIStrips (every SR) **(4)** *exchange APITrap sheet **(5)** *exchange silicone bands **(6)** *take honey sample **(7)** *exchange propolis grids **(8)** open pollen trap if not enough pollen; leave trap closed repeat for hive 2 *only SR02, SR04, SR06, SR08	**IF NOT ENOUGH POLLEN YESTERDAY:** (2) harvest pollen (8) open pollen trap



Preparation

Preparation of sampling material

All matrices in your toolbox are packed separately. You will find e.g. the needed materials for the POLLEN sampling in one package, all material needed for the APISTRIPS in another package, etc. (do not forget to check for completeness: see *table TOOLBOX*).



Example Pollen:

Prepare for each pollen sampling by opening the big bag "**pollen"** and take out the sample bag for *SAMPLING ROUND 1*. You recognize it by taking a look on the code on the label attached onto the cup (xxxx-**SR**01-P). It contains: one plastic cup (green lid), a teabag, two silica gel bags and a label.

Do the same with the other matrices (APIStrips, silicone bands, propolis, APITrap sheets, honey).

Additional material needed:

- The picture manual
- A pencil
- Kitchen roll
- Aluminium foil
- Gloves
- Teaspoon, (Stanley) knife, fork, wet (and clean) towel

Correctly read and write the labels:

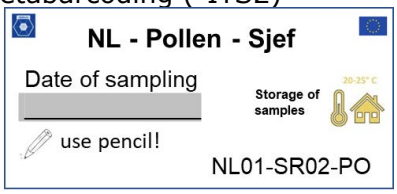
(1) The code on the bottom contains valuable information for you, and for us. It is the proof that the sample was taken, where, what and when it was taken from you.
What does it mean? - example on the right:
- Country code: NL = Netherlands
- Citizen scientist code: 01
- Sampling round 1: SR01
- Matrix: PO = pollen

(2) Use a pencil, otherwise the lettering could become blurred and unrecognizable.

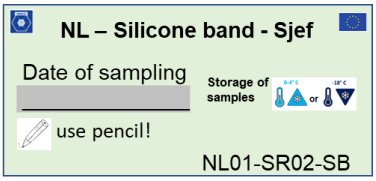
(3) Write the date of sampling on the labels in the format DD/MM".

All labels (Sampling round 02)

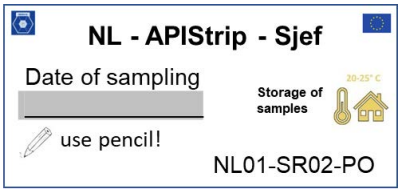
Pollen (-PO)
ITS2 metabarcoding (-ITS2)



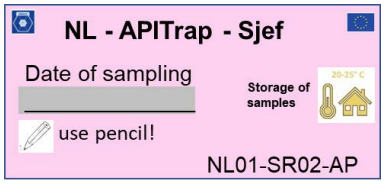
Silicone band (-SB)
Air pollution: VOC and PAH



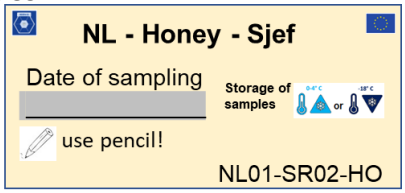
APIStrip (-AS)
Pesticides



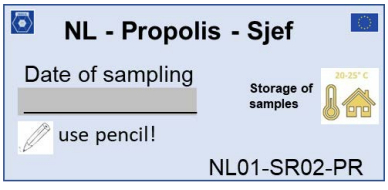
APITrap (-AP)
Microplastic



Honey (-HO)
Pesticides

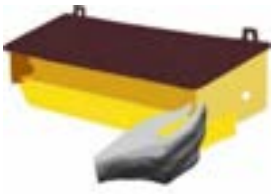


Propolis grid (-PR)
Heavy metals

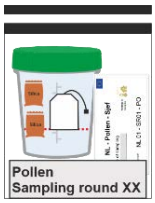


Pollen – sampling procedure

Activation of the pollen trap: activate the pollen trap 24 hours BEFORE the sampling.
Before activation, carefully clean the drawer with the kitchen roll paper to avoid cross-contamination between sampling rounds and remove any pollen grains or debris accumulated between samplings. Afterwards, line the collection tray of the pollen trap with kitchen roll to avoid cross contamination between sampling rounds. The trap needs to be closed for ~24 hours **(evening to evening)** to collect enough material - ~5 g - (if there is not enough material, let it closed up to 3 days within the THU-SUN sampling period).



Have you prepared the right sample bag and label?



Emptying the pollen trap/actual sampling:

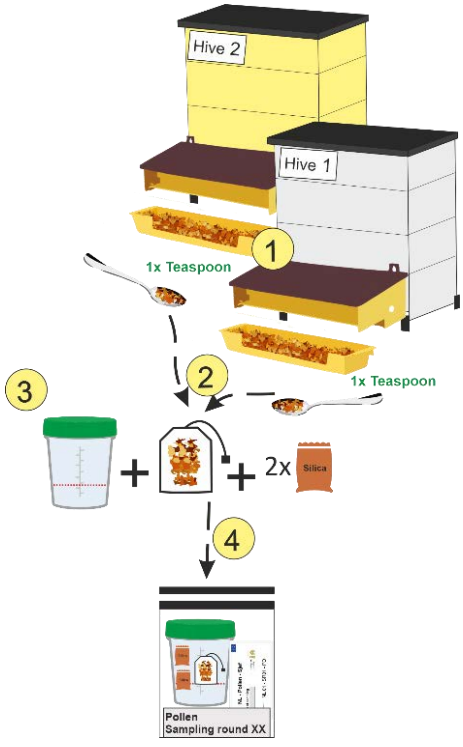
(1) 24 hours after the pollen trap was closed, empty the pollen traps of **hive 1** and **hive 2**. Put on your gloves and wear them for all the matrices. Further, stir the pollen in the trap with a spoon.

(2) Take one TEASPOON of pollen from **hive 1** and transfer it into the teabag in the plastic cup. Take one TEASPOON of pollen from **hive 2** and transfer it into the **same** teabag as before. We aim for a **well-mixed and representative** sample!

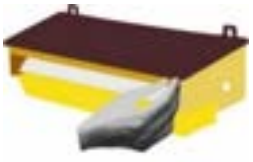
(3) Put the teabag together with 2x 6 g silica gel bags in the plastic cup and close the lid. In case of very wet circumstances one or two extra 6 g silica gels bags should be added.

(4) Use a pencil and write the date of sampling (today's date!!) on the label. Transfer the cup together with the right label into the sample bag and close it tight.

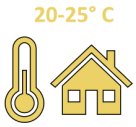
If there is remaining pollen: it is yours. Do not add more than a total of 2 teaspoons (~5 g), as the silica gel was calculated based on this amount.



De-activation of the pollen trap: after sample collection, open (=de-activate) the pollen trap to provide sufficient forage for your bees. Discard the kitchen roll and clean the drawer to avoid cross-contamination between sampling rounds.



Storage of pollen samples: Store the samples dark and at room temperature until shipment to the national coordinator.



Counting of occupied beelanes

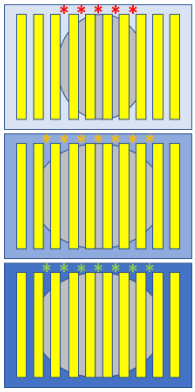
There are many different hive sizes in Europe. To put the generated data in context (=quantitative data), it is important to count the number of occupied beelanes and report it in each of the online questionnaires of each sampling round. The numbers help to understand what the results mean in the end and to compare an apiary in e.g. a northern country with one in southern Europe.

Therefore, count the occupied beelanes of each box and hive at every sampling round according to the following infographic (you are welcome to write the numbers in the table below):

Top view:

Top bar of frame
Beelane
Top bar of frame
Beelane
Top bar of frame
Beelane
Top bar of frame
Beelane

Side view:



bees

Third box occupied by 5 beelanes *

Second box occupied by 7 beelanes *

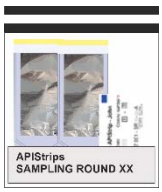
Bottom box (brood box) occupied by 7 beelanes *

Occupied beelanes				
Sampling round	Hive 1		Hive 2	
	Full box	Half box	Full box	Half box
SR01				
SR02				
SR03				
SR04				
SR05				
SR06				
SR07				
SR08				
SR09				

APIStrips – sampling procedure

Two weeks after the last APIStrips were placed in the hives (1 per hive = total of 2 APIStrips), they must be removed and replaced by new ones.

Have you prepared the right sample bag?
Have you prepared the right label?

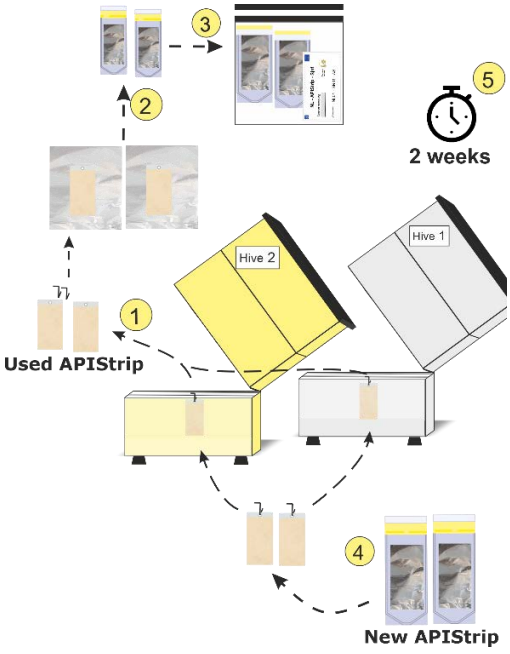


Further Preparations:

Take out the fresh strips (each wrapped in a plastic envelope and in aluminium foil) from the sample bag of the current sampling round and put them near you. Avoid touching the active portion of the APIStrip.
The bags are now empty. Write the sampling date on the labels (today's date!!). Use a pencil. **The empty bags are used for the APIStrips that stayed in the hives for two weeks and will be sampled today.**

Actual Sampling:

(1) Open **hive 1** and lift the APIStrip (avoid touching the active portion of the strip). Remove the wire/twig.
(2) Wrap the strip with aluminium foil and put it back into the plastic envelope from which you just removed the fresh APIStrip.
(3) Put the plastic envelope into the sample bag. Make sure that the appropriate label is inside the bag. The APIStrip from **hive 2** is kept in the same bag, but we recommend going through all the matrices first to avoid opening both hives at the same time.
(4) Hang in the prepared, fresh APIStrip into **hive 1** (avoid touching the active portion of the strip).



It is time to check the sample plan now! Do you need to exchange the other matrices as well?
If yes:
Do not close the hive yet and work through the other matrices (silicone bands, propolis grids, honey

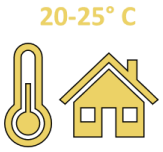
and the APITrap) first (see instructions below)! Afterwards, close **hive 1** and repeat the steps for **hive 2**.

If not: **Close hive 1** and repeat the process for **hive 2**.
In this case, the sampling is done.

(5) Leave the APIStrips in the hives for a total of **two weeks**.

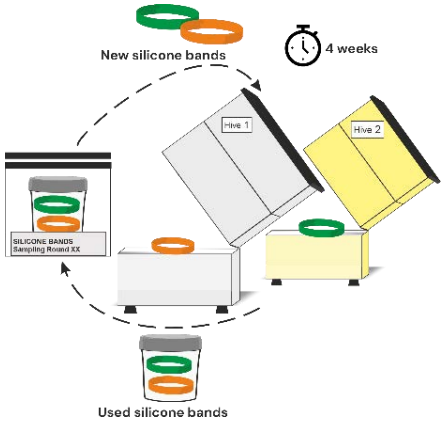
What happens if you have sampled the APIStrip but forgot to insert new ones?
In case of deviations from the sampling plan, please contact your national coordinator for further instructions. As this will result in a shorter or longer APIStrip exposure, you must indicate this in the online questionnaire.

Storage of APIStrip samples: Store the samples dark and at room temperature until shipment to the national coordinator.

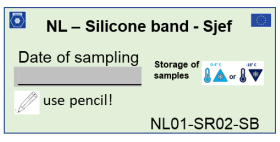


Silicone bands – sampling procedure

Four weeks after the last silicone bands were placed in the hives (1 per hive = total of 2 bands), they must be removed and replaced by new ones.

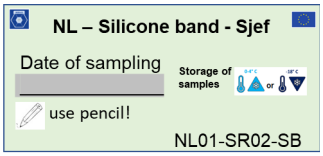


Have you prepared the right sample bag?
Have you prepared the right label?



Further Preparations:

Take the silicone bands from the honey jar in the sample bag and place them near you. It doesn't matter which colour is put in which hive. You will need the new bands soon.



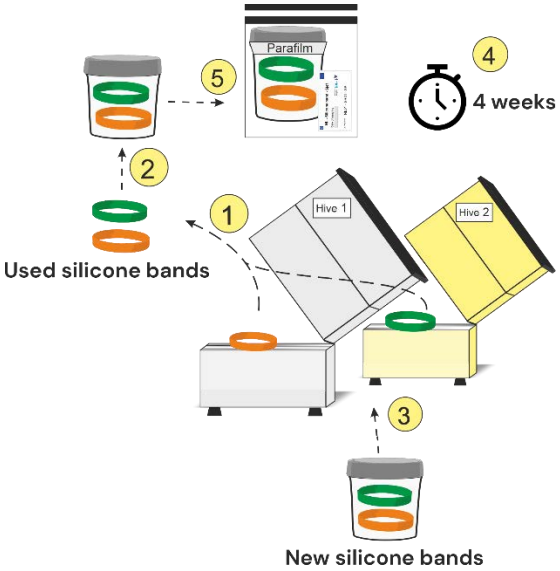
The honey jar and the bag are now empty. Write the date of sampling on the label (today's date!!). Use a pencil. The empty honey jar and bag will be used for the bands that have remained in the hives for **four weeks** and will be sampled today.

Actual Sampling:

(1) Are you still wearing your gloves? Start with **hive 1** and take out the used silicone band.

(2) Put the silicone band in the empty honey jar and close the jar lid. The silicone band from **hive 2** is kept in the same jar, but we recommend going through all the matrices first to avoid opening both hives at the same time.

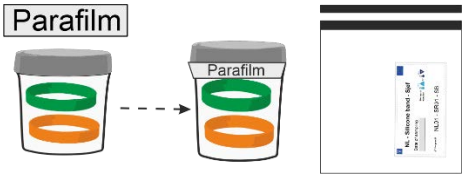
(3) Insert the prepared, fresh silicone band into **hive 1**.



Do not close the hive yet and work through the other matrices (the APITrap, honey, propolis grids) first (see instructions below)! Afterwards, close **hive 1** and repeat the steps for **hive 2.**

(4) Leave the silicone bands in the hives for a total of **four weeks**.

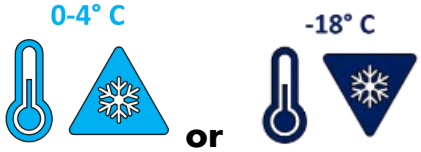
(5) **Make sure the lid of the jar is tightly closed and wrap a strip of parafilm around the lid for an additional airtight seal. Make sure that the appropriate label is inside the bag.**



What happens if you have sampled the silicone bands, but forgot to insert new ones?

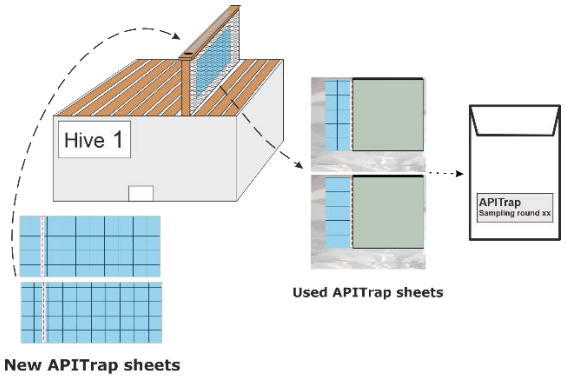
In case of deviations from the sampling plan, please contact your national coordinator for further instructions. As this will result in a shorter or longer silicone band exposure, you must indicate this in the online questionnaire.

Storage of silicone bands: Store the samples in the fridge (0-4° C) or freezer (-18° C) until shipment to the national coordinator.

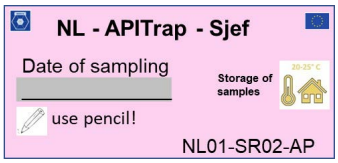


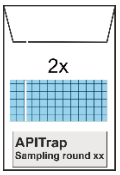
APITrap – sampling procedure

Four weeks after the last **APITrap sheets** were placed in the hive (1 APITrap in **hive 1**), the APITrap sheets must be removed and replaced by new ones.



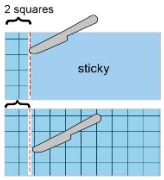
Have you prepared the right sample bag?
Have you prepared the right label?





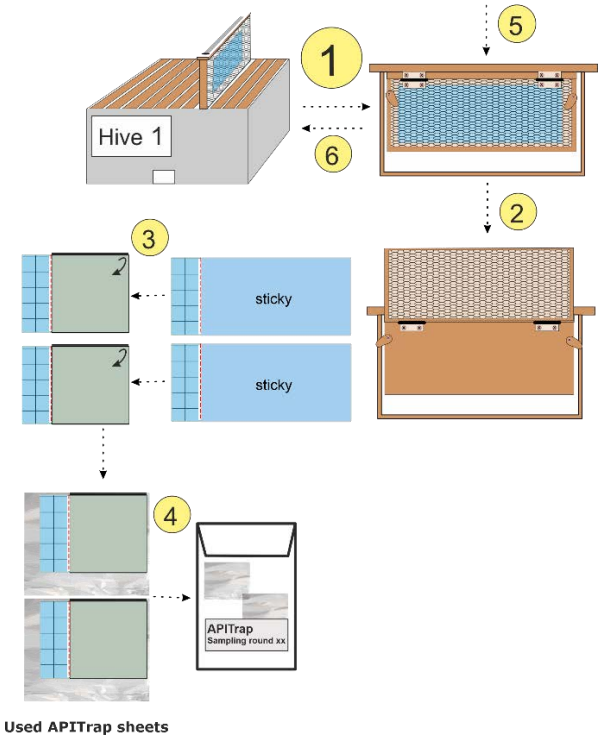
Further Preparations:

Take out the **in-active** APITrap sheets from the envelope and put them near you - you will need them soon.
The envelope is now empty. Write the sampling date on the label (today's date!!). Use a pencil. The empty envelope is used for the APITrap sheets that stayed for **four weeks** in the hive and will be sampled today.



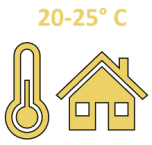
Actual Sampling:

(1) Start with **hive 1** and lift the APITrap.
(2) Open the APITrap
(3) Carefully remove both APITrap sheets one by one. Take the edge of the sticky area and bend it to the other edge of the sticky area (see illustration or video!!).
Do NOT cover the control area (non-sticky).
(4) Wrap the APITrap sheet in aluminium foil and put it in the envelope. Repeat for the second APITrap sheet and insert the label. Close the envelope.
(5) Insert two new APITrap sheets:
Count two squares from the left and carefully cut the protecting backing paper with a (Stanley) knife from top to bottom. Repeat for the 2nd sheet. Do not use too much pressure!
Peel of the thin backing paper only from the large area. This results in a sticky zone and a "control" zone.
(6) Insert both now **activated** APITrap sheets into the APITrap back-to-back (sticky side in direction of the mesh). Hang it back in the **hive 1** and work through the other matrices (honey, propolis grids).
(7) Leave the APITrap in the hive for **4 weeks.**



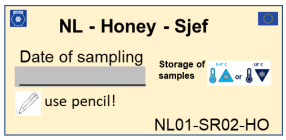
What happens if you have sampled the APITrap sheets, but forgot to insert new ones?
In case of deviations from the sampling plan, please contact your national coordinator for further instructions. As this will result in a shorter or longer APITrap exposure, you must indicate this in the online questionnaire.

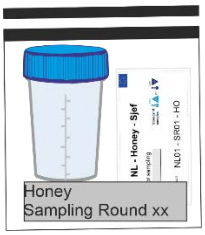
Storage of APITrap sheets: Store the samples dark and at room temperature until shipment to the national coordinator.



Honey – sampling procedure

Have you prepared the right sample bag?
Have you prepared the right label?





Further Preparations:
Take out the plastic cup (blue lid) from the sample bag and write the sampling date on the label (today's date!!). Use a pencil.

Actual Sampling:

(1) Start with **hive 1** and look for a freshly sealed honey comb. The target honey comb is located in one of the honey supers and consists of **fresh, light-coloured wax**.
Another option for the honey sample would be the APITrap frame, because we can be sure, that the wax in this frame is from 2023.
Lift the honey comb.

(2) Open the plastic cup with the blue lid and move the cup with light pressure along the honey comb (from top to bottom) to collect the sample. We aim for as little wax as possible. We need approximately 7.5 ml of honey per hive.
The honey sample from **hive 2** is kept in the same jar (mixed sample!), but we recommend going

through all the matrices first to avoid opening both hives at the same time. Thus, in the meantime close the lid, clean it with a wet towel and put it away.

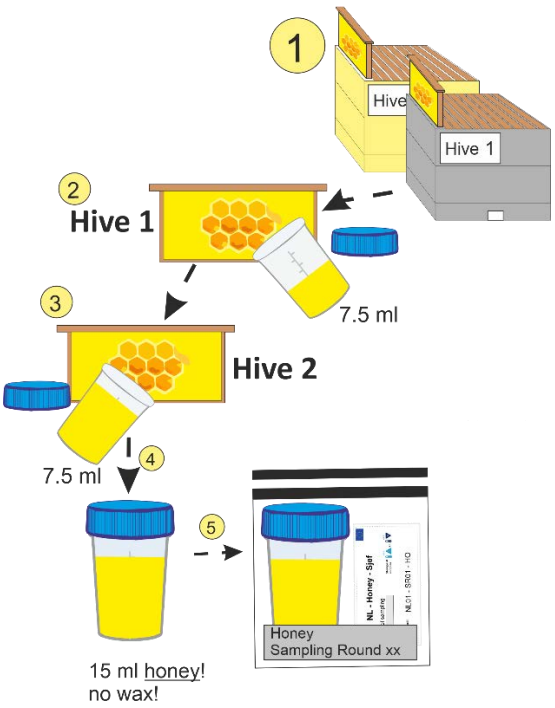
Do not close the hive yet and work through the other matrices (propolis grids) first (see instructions below)!

(3) Repeat the steps for **hive 2** and transfer the honey into the same plastic cup. We aim for **~15 ml honey**. Use a fork, or spoon to remove additional beeswax from the cup. Clean the cup with a wet towel.

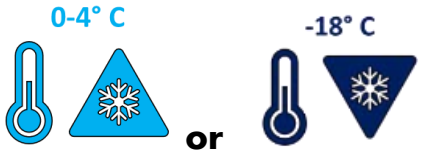
(4) Close the lid tightly.

(5) Transfer the full cup together with the label into the sample bag and seal it. If the cup is too sticky, we appreciate, if you clean it first with pure water and a paper towel before transferring it into the sample bag.

Note: Since we only need fresh honey for the analyses, stop the sampling at the end of the bee season in your country (if honey flow has stopped, or you start feeding).

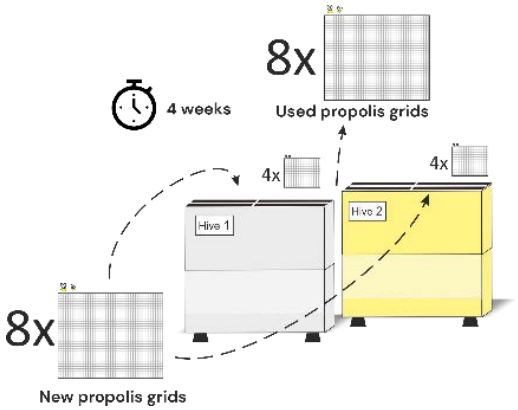


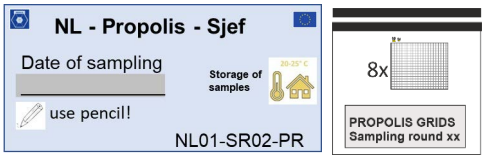
Storage of honey samples: Store the samples in the fridge (0-4° C) or freezer (-18° C) until shipment to the national coordinator.



Propolis grids – sampling procedure

Four weeks after the last **propolis grids** were placed in the hives (4 small grids per hive = total of 8 propolis grids), they must be removed and replaced by new ones. Are you still wearing your gloves?

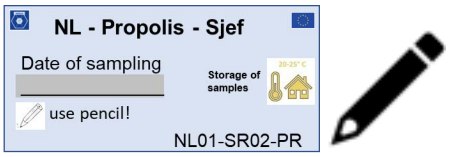




Have you prepared the right sample bag?
Have you prepared the right label?

Further Preparations:

Take out the propolis grids from the sample bag and put them near you. You will need the fresh grids soon.
The sample bag is now empty. Write the sampling date (today's date!!). Use a pencil. The empty bag is used for the grids that stayed in the hives for **four weeks** and will be sampled today.

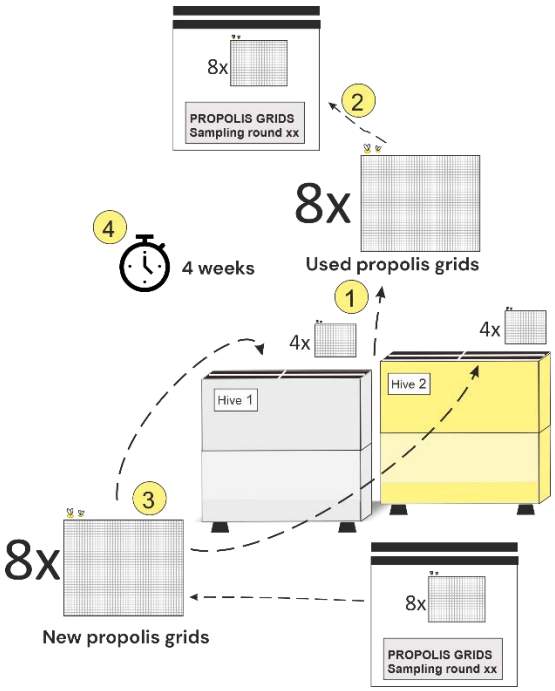


Actual Sampling:

(1) Start with **hive 1** and take out the four used propolis grids.
(2) Put **the 4 propolis grids from hive 1** into the corresponding sample bag. Close the bag with the right label in it. The propolis grids from **hive 2** are kept in the same bag, but we recommend going through all the matrices first to avoid opening both hives at the same time.
(3) Insert 4 fresh propolis grids to **hive 1**. Close the hive lid and repeat for **hive 2.**

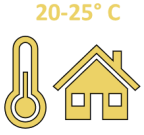
Note: Increased air ventilation above grids will result in higher propolis yields. You could use a carton, or small wood sticks.

(4) Leave the propolis grids in the hives for a total of **four weeks.**



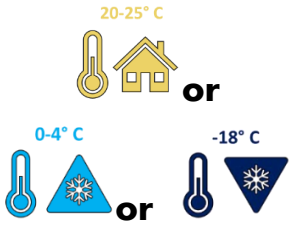
What happens if you sampled the propolis grids, but forgot to insert new ones?
In case of deviations from the sampling plan, please contact your national coordinator for further instructions. As this will result in a shorter or longer propolis grid exposure, you must indicate this in the online questionnaire.

Storage of propolis samples: Store the samples dark and at room temperature until shipment to the national coordinator.



(D) Storage, quality control, shipment procedure and data entry

Storage of samples: The samples will be either stored in the freezer or optionally in the fridge (honey and silicone bands), or at room temperature (APISTrips, APITrap sheets, pollen, propolis) until shipment days. Therefore, carefully read and follow the storage instructions at the end of each matrix protocol.



Quality control: From the moment, the samples were transferred to the sample bags, they must be permanently protected from light and be stored accurately (follow protocol) to avoid mould formation. Mould may destroy the valuable samples for e.g. the purpose of further DNA analysis.

Transport from the citizen scientist to the national coordinator: The process for transporting the samples from the beekeeper to the national coordinator will be decided on an individual basis and local conditions (postal service, by courier or by personal pick-up.). The national coordinator will give further instructions.
Transportation: 5 times during the study
Regardless of the method of transportation, the cost will be borne by the study and not by the beekeeper citizen scientist.

Data entry: Sample information is crucial. Therefore, it is necessary to answer the **online questionnaire** after each sampling round. Prior to each sampling round, you will receive a unique invitation email with a personalised link to your online questionnaire. The link allows the automatic recording of your answers/data entries from the present sampling round. If you have difficulties reaching the website or if you do not receive the **invitation email**, please check your spam box or contact the national coordinator. Note any derogations of the sampling scheme or irregularities, using the comment section of the questionnaire.

Successful data transmission: If you miss the data entry, you will receive a **reminder email**. If you have successfully submitted your data, within a few minutes you will receive a confirmation email. After data submission, the link will be de-activated and cannot be used again. If you made some mistakes during data entry, no problem - please immediately contact your national coordinator.

Additional information required: We would like to learn more about your individual way of beekeeping and strength of colonies, because both could possibly influence lab results. Therefore, we will ask special questions at the beginning and the end of the sampling season with the questionnaires.
Please prepare the following information for us:

- **Varroa treatment:** when and what substances?
- **Frame sizes**
- **Smoker material used**
- **Number of OCCUPIED beelanes** in your test hives (will be asked in **every** questionnaire).

(E) Checklist sampling

Actions	Checklist
Pollen sampling (9 times: SR01-SR09)	
24 hours before sampling: I closed 2 pollen traps	☐
24 hours after closing the traps: I sampled 1 teaspoon of pollen per pollen trap	☐
I put both teaspoons of pollen in the same tea bag	☐
I checked and filled out the label	☐
I have one sample bag with the sample including the label	☐
I de-activated the pollen trap	
I proper stored the pollen sample (dark; room temperature)	☐
APIStrip sampling (9 times: SR01-SR09)	
I sampled 1 APIStrip per hive	☐
I stored each APIStrip in aluminium foil and in separate plastic envelopes	☐
I put both plastic envelopes into the same sample bag	☐
I checked and filled out the label	☐
I have one sample bag with the sample including the label	☐
I inserted 1 fresh APIStrip in each hive	☐
APITrap sampling (4 times: SR02, SR04, SR06, SR08)	
I sampled 2 sticky APITrap sheets from hive 1, folded them accordingly and wrapped them in aluminium foil	☐
I checked and filled out the label	☐
I have one envelope with the sample including the label	☐
I proper stored the sample (dark; room temperature)	☐
I inserted 2 fresh APITrap sheets in the APITrap of hive 1 and activated them!	☐
Silicone band sampling (4 times: SR02, SR04, SR06, SR08)	
I sampled 1 silicone band per hive and stored both of them in the same jar	☐
I checked and filled out the label and sealed the jar with the parafilm strip	☐
I have one sample bag with the sample including the label	☐
I proper stored the sample (dark; fridge or freezer)	☐

I plan to transport them soon to my national coordinator	☐
I inserted 1 fresh silicone band in each hive	☐
Honey sampling (4 times: SR02, SR04, SR06, SR08)	
I sampled 7.5 ml of honey from each hive and I put them into the same vial (blue lid; total: 15 ml). I removed as much wax, as possible	☐
I checked and filled out the label	☐
I have one sample bag with the sample including the label	☐
I proper stored the sample (dark; fridge or freezer)	☐
I plan to transport them soon to my national coordinator	☐
Propolis grid sampling (4 times: SR02, SR04, SR06, SR08)	
I sampled 4 small propolis grids per hive and put all grids (total = 8) into one envelope/sample bag	☐
I checked and filled out the label	☐
I have one envelope/sample bag with the sample including the label	☐
I proper stored the sample (dark; room temperature)	☐
I plan to transport them soon to my national coordinator	☐
I inserted 4 fresh propolis grids in each hive	☐
OTHER Actions	
I counted the occupied beelanes	☐
I answered the online questionnaire	☐

INSIGNIA-EU

Thank you for being a part of the INSIGNIA-EU study 2023.

You make this study a success.

All questions are welcome. Please contact your national coordinator.

4 - Methods

APIStrip extraction procedure

The desorption procedure of the pesticides from the Tenax surface is carried out as follows: First, the APIStrips are cut into small pieces and placed inside 50 ml PTFE centrifuge tubes. Then, 10 ml of acetonitrile are added and the samples are automatically shaken at 1,250 rpm (Geno/Grinder 2010, SPEX) for 3.5 minutes. Subsequently the samples are centrifuged at 4,000 rpm for 5 minutes.

Procedural internal standards are used as surrogate standards to control the extraction performance: These standards include dichlorvos-D6, malathion-D10, carbendazim-D3, and triphenyl phosphate (TPP). The recovery of dichlorvos-D6, malathion-D10, and carbendazim-D3 is checked by liquid chromatography (LC) while the recovery of malathion-D10 and triphenyl phosphate is tested by gas chromatography (GC).

The injection vials varies according to the analytical technique used. For liquid chromatography (LC) vials, 1 ml of the extract is evaporated under a gentle nitrogen stream and reconstituted with 100 µl of acetonitrile. Subsequently, 400 µl of ultrapure water is added. For gas chromatography (GC), 500 µl of extract is evaporated and reconstituted with 50 µl of ethyl acetate. In both cases, the vial preparation entails a 10-fold concentration. Injection internal standards (dimethoate-d6 for LC, lindane-d6 for GC) are employed to check the variations in the injection volume.

Calibration curves are prepared as follows: a blank APIStrip extract (500 µl for GC-MS/MS, 1 ml for LC-MS/MS) is evaporated and reconstituted with an organic solvent (50 µλ ethyl acetate for GC-MS/MS, 100 µλ acetonitrile for LC-MS/MS) containing a mixture of pesticides at 0.5, 1, 5, 10, 50, 100 or 200 µg/l. For the LC vials, 400 µl of ultrapure water is added.

GC-QqQ-MS/MS analysis

The analyses by gas chromatography are performed using an Agilent Intuvo 9000 GC system equipped with an Agilent 7693 autosampler and an Agilent 7010 GC-MS/MS triple quadrupole. Data acquisition and processing is developed by Agilent MassHunter QQQ Acquisition and Quantitative Analysis software version 10.0. The samples are injected using a multimode injector inlet in splitless mode, through an Agilent ultra-inert inlet liner with a glass wool frit. The injection volume is 1 µl. The injector temperature is kept at 80 °C during the solvent evaporation stage (0.1 min) and then ramped up to 300 °C at 600 °C/min for 5 min and up to 250 ºC at 100 °C/min. Two planar columns (Agilent), HP-5MS UI 15 m long × 0.25 mm i.d. × 0.25 µm film thickness are used.

The oven temperature program is as follows: 60 °C for 0.5 min, up to 170 °C at 80 °C/min and finally up to 310 °C at 20 °C/min. The total run time is 12.4 min with 2.1 additional min for backflushing at 310 °C. The instrument works at a constant flow (1.28 ml/min column 1, 1.48 ml/min column 2). The system works in dynamic MRM, acquiring the transition in a ± 0.2 min window from the retention time of the analyte. Helium (99.999% purity) is used as the carrier and quenching gas, and nitrogen (99.999% purity) as the collision gas. The collision and quenching gas flows are 1.5 ml/min and 2.25 ml/min, respectively. Both the transfer line and the ion source -operated in electron ionisation- are

maintained at 280 °C. The quadrupole analyser temperature is fixed at 150 °C. The solvent delay is 2.6 min.

For the optimisation of the MS parametres, all pesticides are monitored in full-scan mode in the 50–550 *m/z* range. The first step is the selection of the precursor/s ion/s for each analyte and the retention time, injecting individual solutions for each pesticide at 1 mg kg^{-1} in full-scan mode. The ion with the highest intensity and *m/z* relationship is selected as the precursor ion. Precursor ion fragmentation is performed by collision-induced dissociation with nitrogen, from which the best fragment ions are chosen. Once the fragment ions are selected, the adequate CE for each transition is assayed in the 3–40 eV range.

LC-QqQ-MS/MS analysis

An Agilent UPLC 1290 Series (Agilent Technologies, Palo Alto, CA, USA) coupled to an Agilent Technologies 6490 Triple Quad LC/MS is used for this study. Data acquisition and processing is developed by Agilent MassHunter QQQ Acquisition and Quantitative Analysis software version 10.0, using dynamic MRM software features with a retention time window of 0.8 min. The injection volume is 5 µl, and the chromatographic separation is carried out with a Zorbax Eclipse Plus C8 column (Agilent), 2.1 mm × 100 mm × 1.8 µm. The system employs 0.1 % formic acid in milliQ water (mobile phase A) and 0.1 % formic acid and 5 % water in acetonitrile (mobile phase B) with the following gradient: 20 % of B for 2 min, a linear gradient up to 100 % of B in 13 min and finally an isocratic mode at 100 % of B for 2 min. Subsequently, an equilibration step coming back to 20% of B (2.5 min) is performed. The system is provided with a JetStream electrospray ion source, employing nitrogen as the nebulizer gas. This ion source is configured as follows: 120 °C for the drying gas temperature, 13 l/min for the drying gas flow rate, 45 psi for the nebulizer pressure, 375 °C for the sheath gas temperature and 10 l/min for the sheath gas flow rate. The MS uses nitrogen as the collision gas (99.999 % purity), 380 V for the fragmentor and 3,000 V for the capillary voltage, both in positive and negative mode.

For the optimisation of the MS parametres, all pesticides at 0.1 mg/l (acetonitrile: water, 1:1, v/v) are injected directly into the MS system in full-scan mode within the 50–800 *m/z* range. From this injection, the precursor ion is selected. Then, one more injection in product ion mode is needed to choose two fragment ions and the optimum collision energy (CE) for each transition. The most intense transition is selected as the quantifier transition (SRM1), while the second most intense is chosen as the qualifier transition (SRM2).

SWB extraction procedure

SWB Preprocess step

The matrix was the silicon wristband (SWB). Before their distribution to countries members of the Insignia EU Project and placing into the hives, all wristbands are solvent- clean up with four different organic solvents (methanol, hexane, acetone, ethyl acetate). For ergonomic and cost-effective reasons, the procedure is upscaled to forty silicone wristbands that can be cleaned simultaneously using the same organic solvent cleaning sequence eliminating matrix interferences. Therefore, they are sequentially extracted with a mixture of 750 mL of methanol, 750 mL of n-hexane, 750 mL of acetone and 750 mL of ethyl acetate. Each extraction occurs after a minimum of 30 min at 300 revolutions per minute (using a platform shaker) and 30 minutes at 30 °C ultrasound sonication. Finally the SWBs extracted with 750 mL of a mixture of ethyl acetate/acetone (1:1, *v:v*). The last extraction occurs a minimum of 15 min and 15 minutes sonification at 30 °C. The organic solvents are collected, distilled, and reused.

SWBs are left to dry overnight and then are positioned in a vacuum oven for 10-12 hours at elevated temperatures (200-220 °C). Afterward, the dried wristbands are placed in a glass laboratory desiccator prior to use as a sampler or blank samples.

During dispatch and before field deployment, the SWBs are placed in PTFE transport/storage bags. After the four weeks sampling, each sample is represented by two SWBs (green & pink) that is delivered back in the lab in the same glass container. The sample of two SWB is then cut into small pieces of 0.6 cm in a pool sample to be used for VOCs as well as for PAH determination following different sample preparation and analytical methods that they are described below.

Figure 4.1: Vacuum Oven for cleaning silicon bands

VOC Sample extraction & analysis

The extraction of the twenty target VOCs is carried through a pool sample of SWBs ~2.5 gr in a 20 mL headspace vial. Then 10 μL of ISs mixture (Benzene-D6, Toluene-D8, Chlorobenzene -D5, o-Xylene D10) is added and screwed immediately with the headspace tap. Internal standards showed good recoveries.

The analytical method is based on gas chromatography mass spectrometry (GC-MS) using external and internal standard quantification procedures. The method is validated with respect to linearity, precision, accuracy, limit of quantification and specificity. The stability of standard and sample solutions is also assessed. The standard and sample solutions are considered stable when stored at ambient and refrigerated for 4 days. Matrix effect is also assessed using procedural standards obtained from spiked blank pool SWB matrices and recovery correction is considered to be used when is needed (e.g recovery<95%).

Calibration curves are prepared as follows: A stock solution containing all 20 VOCs is diluted to polyethylene glycol to form a range of calibration solutions at 7 concentration levels. Then 10 µL of each std solution added in a 20 mL head space vial for HS analysis. From an examination of a low concentration calibration standard analysed as part of the linearity assessment, the limit of quantification (LOQ) is estimated. The limit of detection (LOD) is based on the signal to noise ratio of the chromatographic system.

VOC GC-QqQ-MS/MS analysis

The analysis of the VOC proceeded with head space analysis using a Chromtech Evolution 3 MS/MS triple quadrupole mass spectrometer built on an Agilent 5975 B inert XL EI/CI MSD system with CG-MS/MS eTrap system (cryocooling). The vial is heated at 100 °C for 16 min on the specific headspace heating tray of the gas chromatographer. Then, the vial is returned to the injection tray, and 1 mL of the headspace is withdrawn using a headspace syringe and injected into the GC-MS system equipped with an eTrap system (cryocooling), where trapped at -10 °C for 4.0 minutes. The trap is then rapidly heated to 250 °C and separation is performed using Agilent J&W DB-624 Ultra Inert GC capillary column with a length of 60 m, an internal diameter of 0.60 mm and a film thickness of 1.4 µm via programmed temperature (50 °C to 240 °C). All target VOCs are identified and measured separately, besides *p*-Xylene and *m*-Xylene that is not achievable to be chromatographically separated, even several modifications have been made. Therefore, they are given as *p,m*-Xylene.

For most VOC compounds, an MS/MS procedure is performed in selected reaction monitoring (SRM, also known as multiple reaction monitoring (MRM)) mode. For those compounds where two MS/MS transitions are not obtained, pseudo-selected reaction monitoring (pseudo-SRM) is used, selecting the precursor ion in the first quadrupole, applying zero collision energy (or 1 eV) and isolating the same one as the product ion in the third quadrupole. Data acquisition and processing is developed by Agilent MassHunter QQQ Acquisition and Quantitative Analysis software version 10.0 using dynamic MRM software features with a retention time window of 0.8 min.

Figure 4.2: Agilent GC-MS/MS with cryogenic e-trap.

PAH Sample extraction & analysis

The extraction of 35 target PAHs is carried through a pool sample of wristbands of ~2.6 gr in a borosilicate glass bottle. Then, 5 mL of ethyl acetate are added to the sample and 25µL of the internal standards mixture is added. Subsequently, an additional 25 mL of ethyl acetate is introduced into the glass bottle, and the samples are mixed in the platform

shaker for 30 minutes. Following the platform shaker agitation, the sample is transferred to an ultrasonic bath and subjected to 30 minutes of sonication. The extract is collected in a round-bottom flask. The same extraction process was repeated, and the final 30 ml extract were added to the flask. The combined 60 ml extract is evaporated to dryness, concentrating the extracted compounds. The final dried extract is reconstituted in 1mL of hexane: acetone (1:1, *v:v*), filtered (PTFE, syringe filter, 0.22 µm) and injected into the GC-MS/MS system.

The ISs used to monitor the variability of extraction procedure as well as the matrix effects and finally determine the concentration of the analytes in the sample more accurately are the following: Naphthalene-D8, Acenaphthene-D10, Phenanthrene-D10, Pyrene-D10. Matrix matched Calibration curves are prepared as follows: a blank SWB extract 1 mL is evaporated and reconstituted with an organic solvent of 1 mL containing a mixture of pesticides at 0.5, 1, 5, 10, 50, 100 or 200 µg/L. The LOQ was defined as the minimum calibration point that allowed the identification of each compound according to the criteria described in the SANTE Document No. 11312/2021.

PAH GC-QqQ-MS/MS analysis

Gas chromatograph coupled with Mass Spectrometer detector Triple quadrupole TQ-8040 NX Shimadzu Corporation (Kyoto, Japan) equipped with an AOC 600 Autosampler and a split/splitless injector is used for the determination of the PAHs. Aliquots of 1 µL of sample extract are injected into the gas chromatograph through a Shimadzu inlet for splitless with deactivated wool. The injector temperature is kept at 280 °C and sampling time is 1 min. The injector mode is splitless and split ratio is set -1. For the chromatographic separation of the compounds a MEGA-5 HT column (30 m ×0.25 mm I.D.×0.25 µm) is used. The column oven temperature program starts from 60 °C and is held for 1 min, increases to 180 °C at a rate 40 °C min^{-1} and is held for 6 min, then increases to 300 °C at a rate 25 °C min^{-1} and is held for 6 min, and finally increases to 320 °C and is held for 2 min. The total GC analysis time is 24.8 min. The helium 99.999% carrier gas flow rate is 1.69 mL min^{-1}.
Most of the target PAH are identified and measured separately, but three Benzo fluoranthenes are not chromatographically separated and determined as total Benzo[b,k,j] fluoranthenes.

The mass spectrometer was operated at EI-MS/MS mode and multiple reaction monitoring (MRM) data acquisition mode. For each target analytes, two MRM transitions is analysed, and the collision energy was optimized. Data obtained by GC/MS are processed using LabSolutions. The EI energy used was 70 eV. Solvent cut time was set at 1.5 min. The collision energy varied from 9–42 eV, depending on the precursor and product ions. The interface temperature is 300 °C and the ionization source temperature is set at 280 °C. Data are acquired without using CID Gas (Q3 Scan).
For the optimisation of the MS parameters
, all pesticides are monitored in full-scan mode in the 50–550 m/z range. The first step is the selection of the precursor/s ion/s for each analyte and the retention time, injecting individual solutions for each pesticide at 1 mg kg^{-1} in full-scan mode. The ion with the highest intensity and m/z relationship is selected as the precursor ion. Precursor ion fragmentation is performed by collision-induced dissociation with nitrogen, from which the best fragment ions are chosen. Once the fragment ions are selected, the adequate CE for each transition is assayed in the 3–40 eV range.

Honey extraction procedure

Polar pesticides Sample extraction & analysis

The extraction procedure of the eight polar pesticides from the honey samples is carried out as follows. First, 5 gr from the samples are weighted in 50 mL PTFE centrifuge tubes and spiked with 50µl from the isotopically labelled internal standard mix. Then, 7.5 mL of dd H_2O and 10 mL of MeOH containing 1% formic acid are added and the samples are automatically shaken for 15 minutes (Agytax shaking device). The extracts are completely freeze-out at -80 °C for 30 minutes. After that the samples are centrifuged at 4.500 rpm for 5 minutes and 1 mL from the supernatant is transferred into an Eppendorf containing 50mg C18 sorbent. The samples are centrifuged again at 13.000 rpm for 5 minutes and finally the extract is filtered into a plastic autosampler vial.
Procedural ISs are also used as surrogate standards to control the extraction performance: Glyphosate^{13}C2,^{15}N and AMPA^{13}C,^{15}N. The LOQ is defined in the context of this study as the lower fortification limit demonstrating acceptable precision and accuracy.

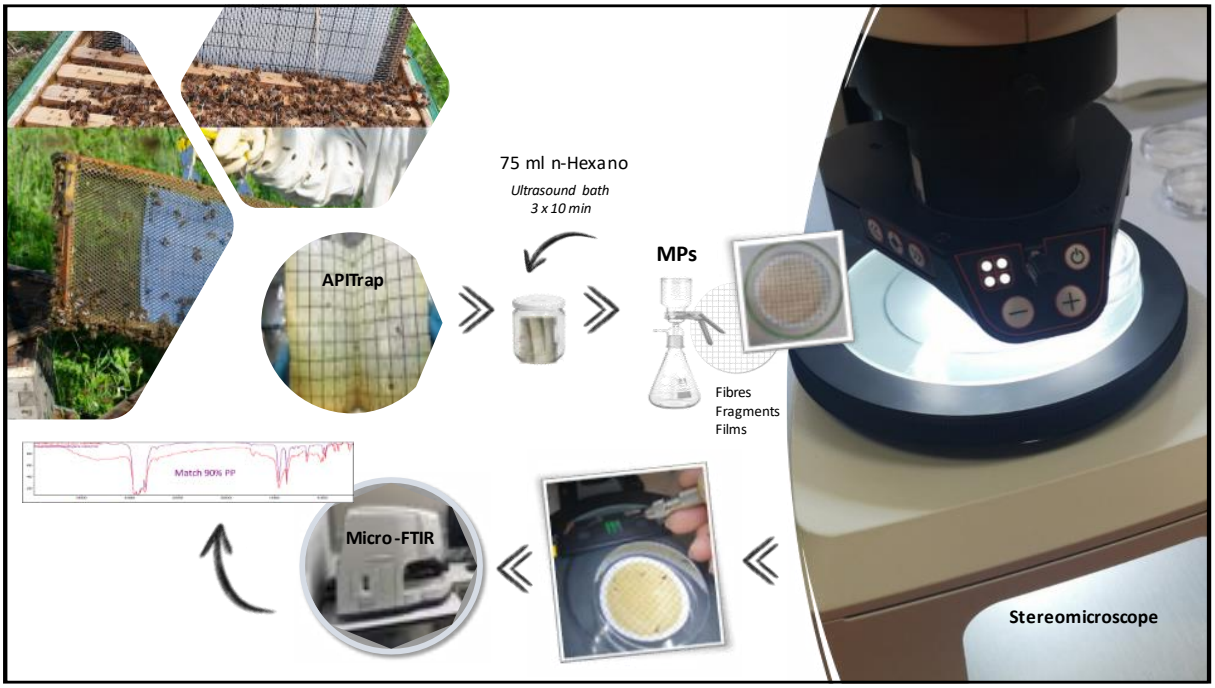
Polar Pesticides LC-MS/MS analysis

The analysis by liquid chromatography was performed in a SHIMADZU 8060NX LC system with triple quadrupole mass spectrometer on negative ionization mode. Mobile phase A was 95% water and 5% MeOH whereas mobile phase B was methanol, both mobile phases contained 1% acetic acid. For the separation, a Hypercarb column (2.1 mm in diameter, 100mm in length, with 5µm particle size) was used combined with a Hypercarb Guard column (2.1 mm in diameter, 10mm in length, with 5µm particle size). The column was thermostatted at 40 °C. The autosampler was thermostatted at 4°C and the injection volume was 10µL. The gradient program is described in detail at the Quick Method for the Analysis of Highly Polar Pesticides in Food Involving Extraction with Acidified Methanol and LC- or IC-MS/MS Measurement, Food of Plant Origin (QuPPe-PO-Method), Version 12.2, 2023.

Microplastics

Sample extraction & analysis

The desorption procedure of the MPs from the plastic polyethylene surface is carried out as follows: First, each APITrap is placed inside 250 ml pre-cleaned glass jars. Then, 75 ml of filtered n-hexane is added. The bottles are covered to avoid contamination, and placed in an ultrasonic bath for 10 minutes, applying three extraction cycles. The supernatant is then filtered under vacuum on a cellulose esters filter (S-Pak Filters, pore size 1.2 µm, diametre 47 mm, Merck Millipore, Milford, MA, USA). During the filtration process, each beaker is flushed with n-hexane to rinse adhering materials. All filters are placed in Petri dishes and left in the dark at room temperature in a drying chamber. A stereomicroscope is used to identify the number of MPs and classify them in terms of morphology, colour and type. After morphological identification, a micro-FTIR is used to characterize MPs and identify the polymer types.



Metals in propolis

Harvesting propolis

Nylon grids have been provided to each beekeeper. These grids were positioned on top of the hive (on the frames) while a small draft was created by lifting the cover of the hive slightly using a peg or piece of wood. In order to stop the draft, bees will cover de grid with propolis, sealing off the hive once more. Four grids per hive were installed to properly cover the whole top-side. As two hives were used per apiary, eight grids were collected once every two weeks and send to the laboratory in Wageningen for further processing.

Prior to collecting the propolis, grids were stored in the freezer to stiffen up the grids and material attached to it. Wearing gloves, the frozen grids were gently rolled and rubbed in order to detach the propolis, which was collected in a plastic container. The sample was properly mixed and collected in a 50 mL vial.

Processing propolis

Using a mortar and a pestle the propolis was mixed with liquid nitrogen and ground to a fine powder. Using a funnel, the powder was collected in a 50 mL vial. Care should be taken to properly clean the used materials between processing different samples.

Approximately 4 to 5 grams of propolis is transferred to a porcelain crucible after which this is covered with a lid and incinerated in an oven at 450 degrees. The oven should heat up in small steps in order to prevent the propolis from boiling over the rim of the crucible. After the incineration process is completed, the crucibles are transferred to a desiccator to cool down to room temperature.

The incinerated material is weighed and dissolved in 30 mL of 0.43M HNO_3 solution. After thorough mixing for approximately 1 hour a 15 mL aliquot is collected and used for further analysis.

Analysis of metals in propolis

The analysis of aluminum (Al), magnesium (Mg), iron (Fe), copper (Cu), zinc (Zn), chromium (Cr), cobalt (Co), nickel (Ni), and lead (Pb) of the propolis was performed using ICP-MS (Inductively Coupled Plasma Mass Spectrometry). Both calibration and internal standards were used in the analysis and the limit of detection and quantification varied between 0.03 and 2.5 and 0.1 and 7.5 ug/L in solution, respectively (Table 1). The results were expressed in mg/L or µg/L and had to be transformed to µg metal/g propolis using the initial weight of the incinerated sample and the volume of HNO_3 used to dissolve the metals.

Table 1. Limits of quantification (LOQ) and detection (LOD) for the analysed metals in µg/L.

	Mg	Al	Fe	Cu	Zn	Cr	Co	Ni	As	Se	Pb
LOD	2.5	3	0.6	1	1.5	0.03	0.037	0.05	0.03	0.03	0.03
LOQ	7.5	10	2	3.3	5	0.1	0.12	0.15	0.1	0.1	0.1

Pollen

Botanical identification of mixed-pollen samples by ITS2 metabarcoding

Molecular identification of the botanical origin of pollen in mixed bee-collected pollen pellets is performed by DNA metabarcoding using high-throughput sequencing (HTS) with the nuclear barcoding marker ITS2 (internal transcribed spacer 2 regions of nuclear ribosomal DNA). To that end, first, DNA is extracted from pollen using the Maxwell® RSC PureFood GMO and Authentication Kit automatic magnetic bead-based protocol implemented in the Maxwell® automatic extractor, with an in-house-made combination of zirconia beads of varying sizes to assure exine rupture of pollen grains of varying sizes. The DNA extracts are then analysed using the ITS2 metabarcoding protocol. Briefly, DNA extracts are PCR-amplified with the universal primers ITS-S2F (Chen et al., 2010) and ITS-S4R (White et al., 1990), as part of an oligo scaffold that incorporates the MiSeq-specific adapters and the unique indexes for sample multiplexing. PCRs are run in triplicate to account for PCR bias. Library preparation for HTS is performed using a dual-indexing approach. The mixed pollen samples are sequenced on the Illumina MiSeq platform using 2×250 cycles v2 chemistry for pools of 373 samples plus positive and negative controls. Analysis of sequence reads and taxa assignments are performed using a curated and updated ITS2 reference database and a bioinformatics pipeline, as described in Quaresma et al. (2024). In addition to ITS2 sequences retrieved from GenBank, the reference database also includes sequences representing bee plants collected by the INSIGNIA consortium. Prior to large-scale implementation of this metabarcoding protocol to samples collected by citizen scientists in 2019, 2020, 2022 and 2023, sample storage / transportation, DNA extraction (Quaresma et al 2021), PCR conditions and primers were assessed, and the protocols were optimised accordingly. The protocols are presented in chapter 4-Methods.

The ITS2 reference database

The ITS2 sequence reads generated in the Illumina MiSeq platform for each mixed-pollen sample are aligned with the sequences of the reference database, and taxonomic classification is performed using VSEARCH v2.15.2 (Rognes et al., 2016), embedded in the bioinformatics pipeline (modified from Sickel et al., 2015). The success and accuracy of taxonomic identification depends, to a large extent, on the quality and breadth of the reference database; ideally, accurately curated sequences of every single plant species visited by the colonies under study would be represented in the reference database. Additionally, the more geographically suited the reference database is, the more accurate can be the classification at the species level; ideally, the database would only include ITS2 sequences for the plant taxa occurring in the area visited by the colonies under study. To address these issues, in INSIGNIA, we constructed a new database for classifying the pollen samples collected by CS across Europe. To that end, an ITS2 reference database was developed with ITS2 sequences recently deposited in GenBank (Quaresma et al., 2024). Due to higher plant diversity, and possibly to lower sequencing efforts, Mediterranean species are probably underrepresented in GenBank. Furthermore, although a greater breadth of the ITS2 database is expected for northern Europe, not all species are represented in GenBank. Therefore, in an attempt to improve taxonomic classification of the bee-collected pollen samples, bee plants (list prepared by INSIGNIA) identified to the species level by experts were gathered from across Europe by ACs and shipped dried to the pollen analytical laboratory (IPB, Portugal) for Sanger-sequencing of ITS2. The newly developed sequences were added to the ITS2 reference database. Second, the GBIF (https://www.gbif.org/) and the EuroMed PlantBase

(https://www.emplantbase.org/home.html) databases were used for generating the lists of plant taxa occurring in the INSIGNIA countries (for further details, see Quaresma et al., 2024).

Metabarcoding procedure

All steps, from pollen homogenisation, DNA extraction, to sample preparation for PCR, should be carried out in a safety cabinet to prevent any contamination from airborne pollen. Additionally, before starting, all equipment and materials (e.g. safety cabinet, micropipettes, magnetic stirrer, centrifuge) should be carefully cleaned with 10% bleach and 70% ethanol and autoclaved (e.g. beakers, magnets, racks). The UV light of the safety cabinet (and of the room, if it exists) should be turned on for 1 hour. Negative controls should be included in every batch of DNA extraction and PCR plate, for monitoring potential cross contamination.

Pollen homogenisation

Pollen homogenisation is carried out by weighing 2 g of pollen pellets into a 50 ml beaker, followed by adding 4 ml of sterile ultrapure water. This solution is homogenised using a magnetic stirrer until the pollen pellets are totally dissolved. A volume of 200 µl of the homogenised solution (~50 mg of pollen) is placed in a 1.5 ml microcentrifuge tube and then centrifuged at maximum speed for 3 minutes. After centrifugation, the supernatant is discarded and 1 ml of absolute ethanol is added to the tube, which is then stored at -20°C until DNA extraction.

DNA extraction

DNA is extracted from the ~50 mg pollen sample using Maxwell® RSC PureFood GMO and Authentication Kit automatic magnetic bead-based protocol implemented in the Maxwell® automatic extractor, according to manufacturer's instructions, with minor modifications as follows:

a) The ethanol-stored pollen samples are centrifuged at maximum speed for 3 minutes, and the supernatant (ethanol) is discarded;
b) The pollen is transferred to a 2.0 ml screwcap tube containing a mix of zirconia beads of varying sizes (100 µm, 200 µm, 400 µm, 800 µm, 3 mm) to target all pollen grains;
c) A volume of 300 µl of CTAB lysis buffer is added to the 2.0 ml tube and the mixture is ground in a powerful tissue homogeniser (e.g., Precellys, Bertin, for 3 grinding cycles at 6200 rpm for 5 sec) to assure exine rupture of all pollen grains;
d) The next DNA extraction steps strictly follow manufacturer's instructions.

Library preparation

ITS2 library preparation follows a dual indexing strategy (Kozich et al., 2013) and starts with a two-stage PCR: amplicon and indexing. PCR products are sequenced in the Illumina MiSeq platform using a paired-end protocol to generate millions of overlapping reads.

Amplicon PCR

In stage-one PCR, the ITS2 nuclear region is amplified using the universal primers ITS-S2F (Chen et al., 2010) and ITS-S4R (White et al., 1990), which have been expanded to include the Illumina sequence adaptors for the stage-two PCR (Figure 12; Table 2).

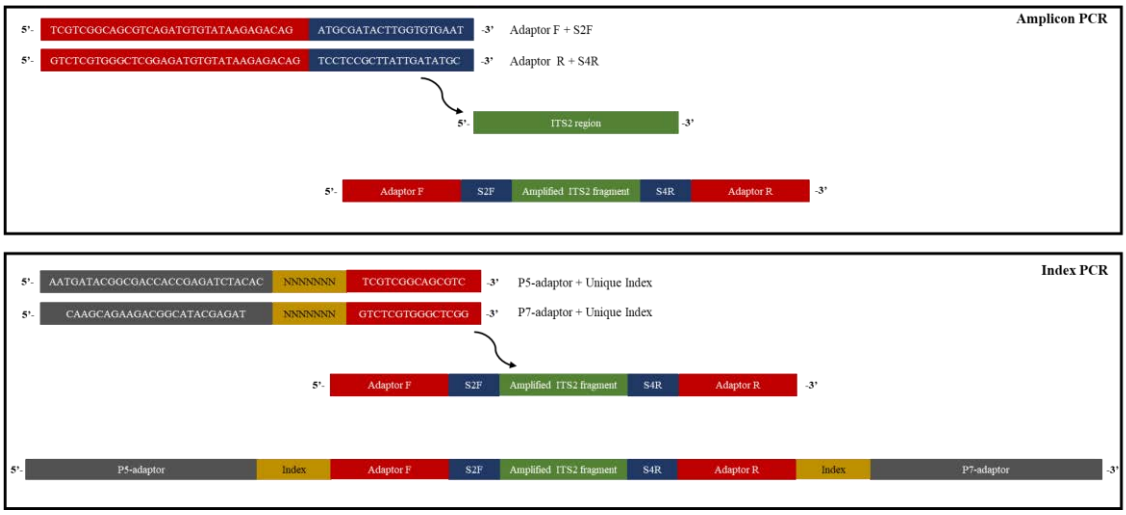


Figure 12. Structure of the Illumina library.

To overcome the problem of low base diversity, typical of amplicon libraries, and therefore improving sequence read base quality, a pool of seven different forward and reverse oligos are used in the PCR (Wu et al., 2015). The difference between each oligo is the addition of any nucleotide "N" between the adaptor and both primers (Table 2).

Table 2. Oligos with variable number of Ns (underlined). The 33-nt adapters are marked in red. The ITS2-specific forward (ITS-S2F) and reverse (ITS-S4R) primers are boldface. A 10 µM pool of the seven different oligos is used in stage-one PCR.

Adaptor-ITS-S2F (forward)	**Adaptor-ITS-S4R (reverse)**
5'TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG**ATGCGATACTTGGTGTGAAT** 3'	5'GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG**TCCTCCGCTTATTGATATGC** 3'
5'TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNATGCGATACTTGGTGTGAAT 3'	5'GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNTCCTCCGCTTATTGATATGC 3'
5'TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNATGCGATACTTGGTGTGAAT 3'	5'GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNTCCTCCGCTTATTGATATGC 3'
5'TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNATGCGATACTTGGTGTGAAT 3'	5'GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNTCCTCCGCTTATTGATATGC 3'
5'TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNATGCGATACTTGGTGTGAAT 3'	5'GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNNTCCTCCGCTTATTGATATGC 3'
5'TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNNATGCGATACTTGGTGTGAAT 3'	5'GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNNNTCCTCCGCTTATTGATATGC 3'
5'TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNNNATGCGATACTTGGTGTGAAT 3'	5'GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNNNNTCCTCCGCTTATTGATATGC 3'

PCR reactions are carried out in a T100™ Thermal Cycler (BioRad, California, USA) using a 10 µl total volume containing 1 µl of DNA, 5 µl of Q5® High-Fidelity 2X Master Mix (NEB, MA, USA), and 0.5 µl of the extended ITS-S2F and ITS-S4R oligo pools at 10 µM (Table 1). Each pollen sample is amplified in triplicate (e.g., three PCR plates for 96 samples, including controls). Thermal cycling conditions are 98°C for 3 min, 35 cycles of 98° C for 10 sec, 52° C for 30 sec and 72° C for 40 sec, and a final extension of 72°C for 2 min.

For checking amplification success, 2 µl per sample of PCR products are loaded on a 1% agarose gel and run at 95V. Before the indexing PCR, PCR products are diluted 1:1.

Indexing PCR

In the stage-two indexing PCR, unique 7-nt sequence indexes are incorporated into the amplicons. These unique sequences allow the pooling (multiplexing) of hundreds of samples in the sequencing library. The indexing PCR is prepared individually for each replicate using 2.8 µl of 1:1 dilution of the previously amplicon, 7 µl of KAPA HiFi HotStart ReadyMixPCR Kit (Kapa Biosystems, MA, USA), 1.4 µl of 1 µM of forward and reverse oligos, in a total volume of 14 µl. The oligo scaffold includes the MiSeq-specific sequence (forward: P5, reverse: P7), to attach to the sequencing flow cell, a 7-nt unique index, to allow multiplexing, and the 14-15 nt adapters to attach to the stage-one amplicon (Figure 12, Table 3). Thermal cycling conditions are 95°C for 3 min, followed by 8 cycles of 95°C for 30 sec, 55°C for 30 sec, 72°C for 30 sec, and a final extension of 72 °C for 5 min. Products are run on a 1% agarose gel, for quality checking of the indexing reaction.

Table 3. Oligos used in the indexing PCR. The sequences in yellow and blue are the MiSeq-specific adapters P5 and P7, respectively, read by the MiSeq flow cell. The 7-nt underlined sequences are examples of unique indexes, which allow sample pooling. The size of the pool depends on the number of unique indexes used. The remaining 14-15 nt adaptors will attach to the amplicon of stage-one PCR.

P5-index (forward)	P7-index (reverse)
5'AATGATACGGCGACCACCGAGATCTACACAC CAACT TCGTCGGCAGCGTC 3'	5'CAAGCAGAAGACGGCATACGAGATAGG TATT GTCTCGTGGGCTCGG 3'
5'AATGATACGGCGACCACCGAGATCTACACCG AATGC TCGTCGGCAGCGTC 3'	5'CAAGCAGAAGACGGCATACGAGATCCT ATCT GTCTCGTGGGCTCGG 3'
5'AATGATACGGCGACCACCGAGATCTACACGA TCTCG TCGTCGGCAGCGTC 3'	5'CAAGCAGAAGACGGCATACGAGATCCT GCGA GTCTCGTGGGCTCGG 3'

Library normalisation, pool size determination, and quantification

Indexed amplicons are normalised using SequalPrep™ Normalisation Kit 96 wells (Invitrogen, Carlsbad, CA, USA), to assure more equalized library sizes. After normalisation, all the samples from each PCR plate are combined into one pool. Each plate pool is quantified using the Qubit Fluorometric Quantification (ThermoFisher Scientific, Massachusetts, USA) and then equimolarly combined into one final pool. For further confirmation that the correct ITS2 fragment has been amplified, the size distribution of the amplicons is determined on a TapeStation 2200 using the HS D5000 kit (Agilent Technologies, CA, USA). Next, the final pool is quantified by a SYBR green quantitative PCR (qPCR) assay using the KAPA Library Quantification kit (Kapa Biosystems, MA, USA) and the QuantStudio™ 5 Real-Time PCR (ThermoFisher Scientific, Massachusetts, USA).

Library sequencing and bioinformatics

The sequencing library is diluted to 2 nM, spiked with 10% Illumina-generated PhiX control library and then sequenced on MiSeq (Illumina, CA, USA) using the MiSeq Reagent Kit v2 (500 cycles) with paired-end reads, according to manufacturer's instructions. Once the sequencing run is completed (after ~39 hours), the pools are de-multiplexed into the original samples, based on the unique 7-nt index, and the primers are trimmed using the BaseSpace™ platform. The sequence reads generated for each sample for the ITS2 fragment are imported into a bioinformatics pipeline (modified from Sickel et al., 2015) for semi-automated analysis. Briefly, the reads are merged, quality filtered and then classified to the family, genus or species level using a curated ITS2 reference database that includes

sequences representing bee plants collected by the INSIGNIA consortium (Quaresma et al., 2024). The entire work flow for ITS2 metabarcoding of mixed pollen samples is shown in Figure 13.

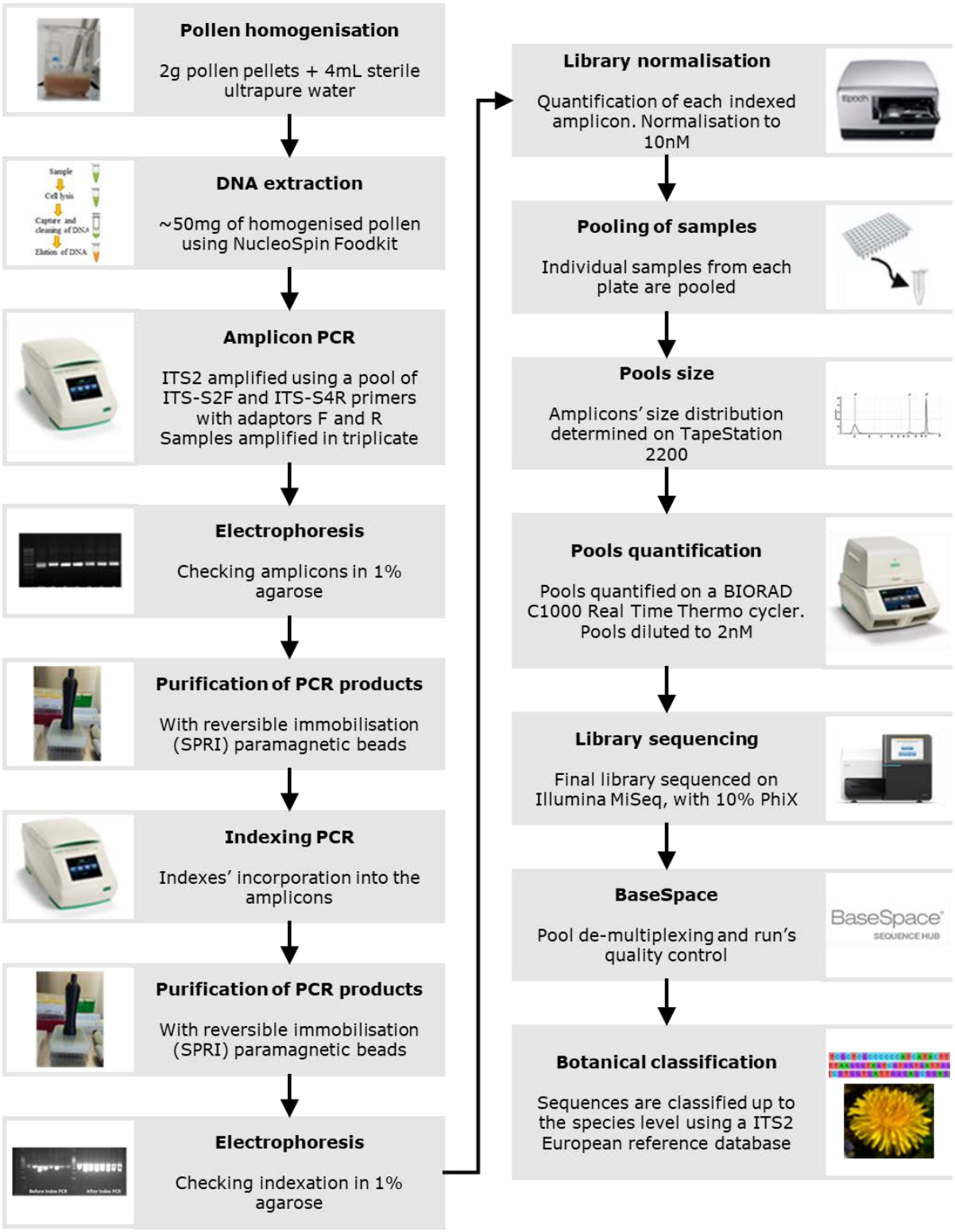


Figure 13. Work flow for ITS2 metabarcoding of mixed pollen samples.

Statistical methods

Several data analyses are considered here: for assessing pesticides found in APIStrips, PAHs/VOCs from analysis of silicone wrist bands (SWBs), microplastics (fibres and fragments/films) found in APITraps, and heavy metals found in propolis samples. The relationships between these various pollutants and pollen diversity are also considered.

The aims and approach

In most of the above, the focus has been on qualitative analysis, based on the number of compounds only, rather than quantitative analysis of actual concentrations detected of any one pesticide residue or PAH/VOC, for example. For heavy metals, quantities of eleven metals were reported for all samples analysed and quantitative analysis is appropriate.

The main aims are to assess the overall situation across the EU in terms of the presence and load of pollutants, differences between the numbers recorded over sampling rounds, differences between countries, differences between the participating apiaries within a country (geographical variation), and differences between apiaries as a whole according to the main land use of their surrounding areas (here land use was categorised either as mainly Agricultural, mainly Artificial/urban, or mainly Forest/natural landscape). Land use was of particular interest in the analysis, both for the EU as a whole and per country.

Data processing

The data are supplied as Excel spreadsheets. Each one includes details of country, apiary code, colony number (if relevant) and sampling round for each analysed sample, as well as the data from the analytical laboratory analysis of the sample, which was further processed by the Data Curator. Each row or record corresponds to a colony / apiary / sampling date combination within any given country. If this does not directly contain numbers of compounds detected in each record (for each apiary (and possibly colony) per sampling occasion for each participating beekeeper in each country), to obtain presence/absence data the dataset will need further processing. Any uninformative records not providing any quantitative information may be removed. Any blanks/not detected codes in the data cells are then set to 0 and any recorded quantities or concentrations are set to 1, to give absence / presence data. For pesticide data, for example, summing across the columns, one column per pesticide, then gives the number of pesticides detected for each row/sample, for use in further qualitative analysis. In the case of the heavy metals data, the recorded values were used as given, for quantitative analysis.

In the INSIGNIA-EU study all detections of compounds were treated as indicating presence of that compound, regardless of the recorded concentration, and any values <LOQ were also included and taken as LOQ/2 before recoding to 1 (presence), as advised by the consortium's analytical chemists.

Descriptive statistics and plots

Summary statistics (including, for example, number of observations, minimum, mean, median, maximum, lower and upper quartiles) should be used to form an informal impression of the data overall, and of any differences between sampling rounds or land use categories. This can be done at EU level and also per country if required. It may be of interest to do the same to compare apiaries within any one country.

Histograms or bar charts, and / or boxplots, are useful for visualising the distribution of the data. Bar charts are suitable if the maximum number of compounds recorded for any one sample is not large, otherwise histograms may be preferred. Histograms are more appropriate for quantitative data such as the heavy metals. Boxplots are useful for identification of any outliers (extreme or unusual observations) as well as easily assessing level and spread.

Parallel boxplots of the number of compounds are recommended for indication of unusual observations and comparisons of the level and spread of the data between sampling rounds or land use categories of the apiaries (or apiaries per country). Each individual boxplot will correspond to all the data recorded for one sampling round or land use category (or an apiary for a given country).

Two-way frequency tables (for qualitative data) and image plots (frequency plots) displaying colour-coded frequencies of the number of compounds may be useful to show differences between sampling rounds or land use categories, for example, for a given country or overall. Greater depth of colour in one cell of the plot indicates a greater frequency of occurrence of that number of compounds for that sampling round or land use category. The rows will correspond either to the sampling rounds or land uses respectively, and the columns will correspond to number of compounds, or vice versa. The same approach could be used for apiaries within any one country. The image plots make visual identification of any patterns straightforward.

It is useful to find the proportion of samples in which a compound was detected at all or was detected for different sampling rounds or land use categories (in order to compare them- see below).

Statistical tests

Tests will formally assess differences in recorded numbers of compounds between sampling rounds or land use categories, for the EU as a whole or per country, or to compare numbers recorded for apiaries in a given country. As the number of compounds found is likely to have a skewed distribution and is count data, sometimes having a limited range, a Kruskal Wallis test of difference in distribution in number of compounds is appropriate to compare different sampling rounds, land use categories, or different apiaries within one country. If this indicates significant differences, this is followed by multiple comparisons tests to identify the differences.
In the INSIGNIA-EU study, Dunn's test was used for multiple comparisons after a Kruskal Wallis test, or pairwise Wilcoxon rank sum tests corrected for multiple testing (here we used the Benjamini-Hochberg correction for multiple testing).
To test the significance of any differences in the proportions of times compounds were detected at all for different sampling rounds or land use categories, a Pearson's Chi-squared test can be used on the frequencies of detection and non-detection, or alternatively, if the observed frequencies of detection/non-detection are low, a Fisher's exact test is more appropriate.

Quantitative analysis and relationships among variables

For quantitative analysis, for the heavy metals, relationships between the quantities of the metals can be examined using correlations, tests of correlation and correlation plots. As these data are also skewed, due to the presence of outliers, Spearman's correlation is appropriate. Log transformation of the heavy metal data is also useful for plotting the data. Principal components analysis can be used to visualise the combinations of the heavy metals for the different samples, identify unusual patterns and examine any grouping in the data.

Examining correlations can also be used to relate measures of pollen diversity or the number of plant families, genera or species to pollutant load. Alternatively, statistical modelling (generalised linear models or generalised additive models) could be used to model pollen diversity in terms of pesticide numbers and apiary location, for example.

Modelling

The authors of the WUR models are well aware of the EFSA Scientific Opinion on good modelling practice. The Gradient Boosting Machine model used in INSIGNIA is a statistical model. It does not fall under the umbrella of the EFSA Scientific Opinion which focusses on mechanistic models. The pollen foraging model is a variant of the model published by Baveco et al. (2016), and is mechanistic. This model has been developed prior to the EFSA SO. The current version of the pollen foraging model is under development. Future evolutions of the model may lead to applicability in the risk assessment of PPPs, at which point the model, model description and documentation will be assessed against the EFSA SO. Suggestions made therein, as well as following the rationale of the TRACE framework by Schmolke et al. (2010) and updated by Grimm et al. (2014) will be taken into consideration.

Occurrence of pollen and pesticides in the INSIGNIA landscape

In order to visualise the occurrence of pollen and pesticides in the landscape, pollen and pesticides entering the hive that were sampled during the monitoring campaign were used in two different model approaches. These approaches comprised a 'mechanistic pollen foraging model' and a 'pollen and pollutant prediction model'. The mechanistic model (Figure 14) simulates bees foraging on pollen in a realistic landscape. It has its roots in energetics, which means that bees will forage in fields where the energetic gain from pollen is maximised. This could mean, for example, that bees forage on a pollen resource with a lower energy content if the net energy gain is higher compared to a pollen resource with a higher energy content but which may be further away from the hive. Ultimately, the types of pollen and pesticides (in the pollen) that are returned to the hive are emergent properties of the model. Detailed information on the landscape, land-use, crop-specific pesticide usage, application times and rates and residues around a hive location are needed, as well as data on plant phenology to infer pollen availability and assumptions on the linkage between landscape and pollen sources (i.e., which pollen originate from which types of landscape elements).

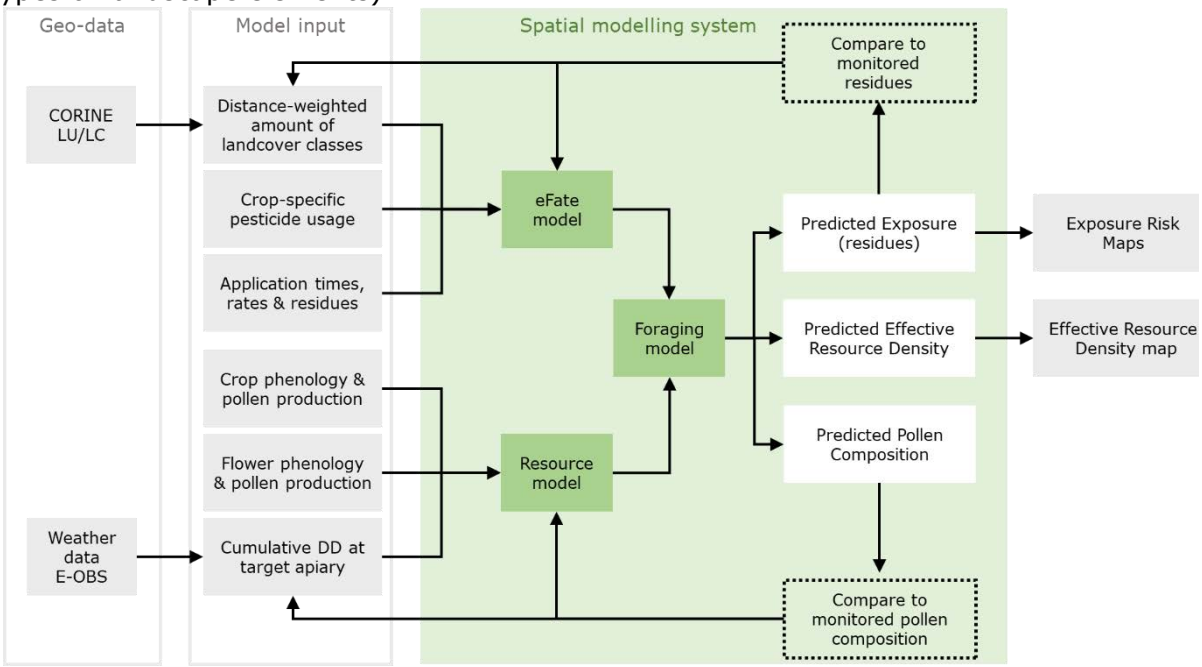


Figure 14. Overview of the modelling approach. The different modules deal with (1) the geodata sets used to define the landscape; (2) the definition of the (pollen) resource landscape; (3) the definition of the exposure landscape; (4) the pollen foraging modelling from which the foraging locations and the pesticide residues in the collected pollen can be obtained

The prediction models (Figure 15) are machine learning models (gradient boosting machine) that allow for predictions of the number of pollen and pesticides and quantities of metals that are likely to be present at a given location and at large spatial scales (e.g., at the scale of the EU 27). Although, in contrast to the mechanistic model, it is challenging to gain insights in the driving processes, the machine learning models have a much lower data requirement.

Moreover, they are free from assumptions on the links between the landscape, land-use, and available pollen or presence of pesticides or metals, and do not need information on pesticide usage, application times and rates, and residues that may be available for local areas, but are difficult to obtain at large spatial scales.

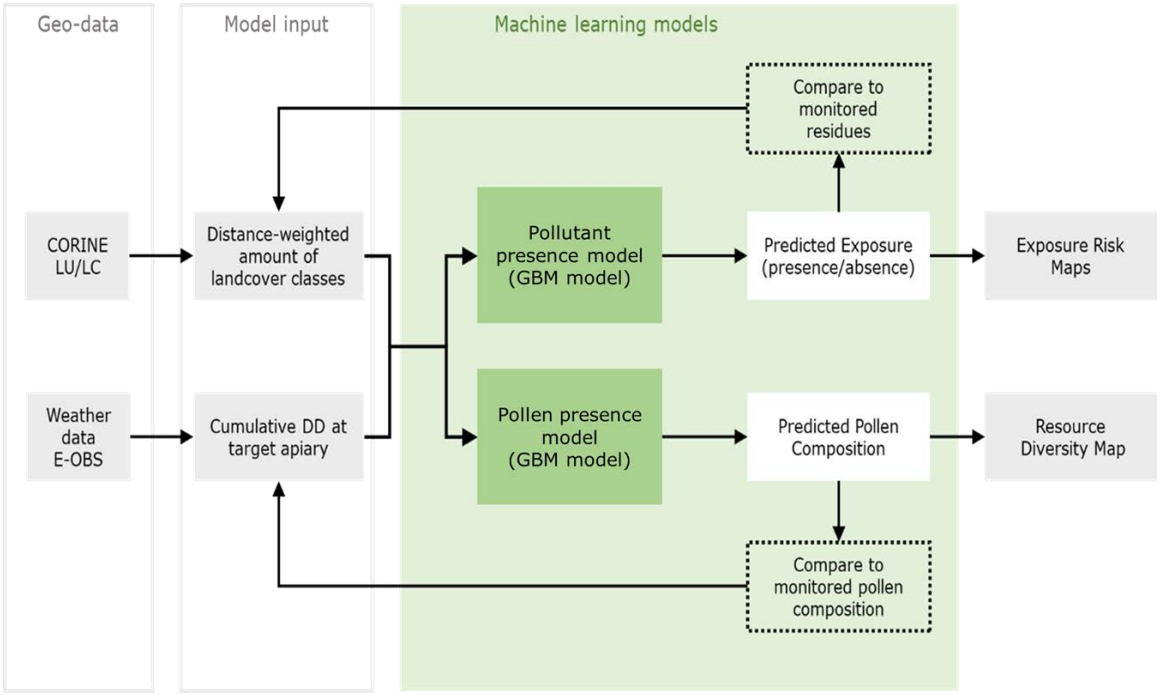


Figure 15. Overview of the machine-learning based modelling of the pollutants and pollen expected in pollen**.**

To apply either type of model, two main types of input geo-data need to be collected and prepared (see Figure 14 and Figure 15): monitoring data for pollen and pesticides collected in the INSIGNIA study; and two sources of spatial and temporal geodata, i.e. the 'CORINE Land Use and Land Cover' data and the 'pan-European weather' data set. They form the heart of the mechanistic and statistical models, the technical details of which are described in "Mechanistic pollen foraging model" and "Pollen and pollutant prediction model" respectively. Before describing these models, first the underlying geo-data sets are discussed.

Spatial and temporal geodata

CORINE Land Use and Land Cover data

The first step in preparing the necessary data for the mechanistic and classification models is the geo-data (see box 'Geo-data' in Figure 14 and Figure 15). After the first processing step, which is described in the section "Kernel model", the data can be used as input to the models (see box 'Input data' in Figure 14 and Figure 15). The data describe the landscape around the locations of hives used in this study and are used to form a link between the physical landscape and the presence of pollen and pesticides. The data set used is the CORINE Land Cover (CLC) dataset, which is based on satellite imagery. It was first developed in the 1980s with the aim to standardise data collection on land and land-use in Europe and to support environmental research and policy. The dataset contains information on the land use and land cover over the whole of Europe at a relatively coarse 100 m resolution. In other words, the whole of Europe is divided into 1 hectare grids, and each grid cell is placed into a category describing the land-use. In total 44 different land-use classes are recognised. They are sub-divided into five main categories: artificial surfaces (e.g., roads, built up areas), agricultural areas (e.g., pastures, arable fields), forests and semi-natural areas (e.g., rough pastures), wetlands (e.g., bogs, marshes), and waterbodies (e.g., rivers and lakes). The most recent CORINE raster version (coordinate reference system EPSG:3035 - ETRS89-extended / LAEA Europe – Projected; 2018) was downloaded from https://land.copernicus.eu/pan-european/corine-land-cover/clc2018?tab=download and used for this study. Obviously, there is no need to use the entire data set for the whole of Europe for each hive. Therefore, in order to extract the relevant information from the CORINE database (i.e., within the vicinity of a hive) Python scripts are used to obtain so-called clipped maps that contain information on the type of CORINE CLCs in the local area. Pre-processing steps of the models expect maps in ascii-grid format. In the INSIGNIA study, two spatial scales are used to extract CLC data. Firstly, bee foraging is assumed to be limited to the landscape within a 3.25 km radius of the hive. With this distance in mind, for pollen resources a hive location is defined by a clip of 65 by 65 cells with the hive coordinates in the centre. In other words, the hive is at the centre of a 'patch' with an area of 6.5 by 6.5km. Secondly, the foraging landscape that is used by bees can be influenced by pesticides coming from within this area but also from outside this area by, for example, drift deposition. Therefore, for the landscape that determines the potential sources of exposure to pesticides, a larger clip area is used, i.e., 513 by 513 cells.

Pan-European weather data

The second primary source of input geo-data required by both models is on temperature patterns (see Weather data E-OBS in the box 'Geo-data' in Figure 14 and Figure 15). The reason is that the phenology of plant species determines when that species starts flowering and thus when pollen is available. It is, however, impossible to obtain information on plant phenology for individual plant species from geo-data directly. The number of cumulative degree days (CDD10) above a threshold value (10 °C in this report) is generally considered to be an important species-specific requirement. Indeed, species-specific thresholds in CDD10 for first bloom and / or full bloom can be found in phenology calendars such as: https://www.oardc.ohio-state.edu/CDD10/. For maximum generality, for each pollen resource the flowering period is defined as a distribution over CDD10 values with a base temperature of 10 °C. To apply the models for specific location and time of the year, the CDD10 up to this date for this location will be needed. The CDD10 can be calculated from temperature time-series by summing the difference between the daily average temperature (geometric mean between minimum and maximum) and the threshold temperature.

See Figure 16 for an example of accumulated CDD10s for INSIGNIA apiaries in 2019. To avoid the accumulated CDD10 from decreasing after cold days (i.e., incorrectly reverse plant development), if the minimum or maximum temperature are below the threshold temperature, they get the same value as the threshold. This is the so-called Modified Average Method (for more information see http://cues.cfans.umn.edu/old/IPM-Tactics/IPM-tactics.html chapter 11: Using Degree-Days and Plant Phenology to Predict Pest Activity).

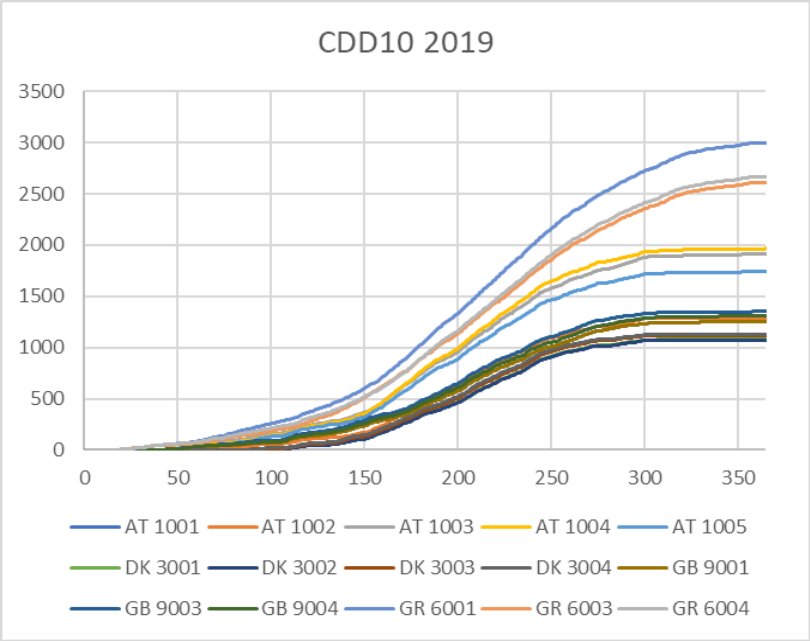


Figure 16. Cumulative degree days with base temperature 10 °C for the INSIGNIA apiaries in 2019

For any location in Europe, and for a large number of years (from 1950 onwards), time-series of daily temperatures were obtained from the daily gridded observational dataset for precipitation, temperature and sea level pressure in Europe (E-OBS), provided by the Copernicus Institute, on a 0.1 degree regular grid (https://www.ecad.eu/download/ensembles/download.php. For the calculations, version 21e was used. Note that for a probabilistic approach, one can download ensemble spread data (difference between the 5th and 95th percentile indicating the 90% uncertainty range). However, here a deterministic approach was followed, based on the median values (i.e., the median values of time-series of minimum and maximum temperatures of an ensemble of models). Additionally, the ensemble median values can be downloaded in smaller chunks from the website: http://surfobs.climate.copernicus.eu/dataaccess/access_eobs_chunks.php.). Finally, for recent periods (July 2020 – January 2021 at the time of writing: 4 March 2021) monthly datasets are available from: http://surfobs.climate.copernicus.eu/dataaccess/access_eobs_months.php). An example of the contents of such an E-OBS dataset, for 31 December 2019, is shown with the apiaries in Denmark, Austria, UK and Greece, in Figure 17. Using the R packages *ncdf4*, *raster*, *rgdal* and *ggplot2*, the timeseries for given hive coordinates and year can be extracted for use in the model (R-script available upon request to the authors). With the temperature data pre-processed, the required input geo-data are complete and can be used as the first set of input data (see Figure 14 and Figure 15).

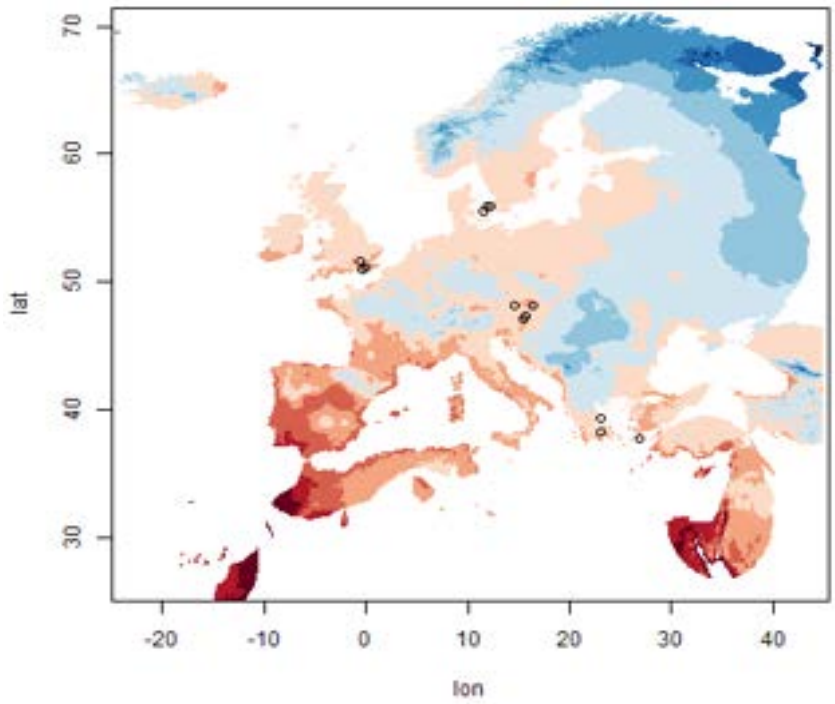


Figure 17. Location of the apiaries plotted on the map with temperatures on 31 December 2019. Indicative colours range from warm (dark red) to cold (dark blue).

Mechanistic pollen foraging model

Kernel model

For a given hive location, the processed CORINE CLC data represents the landscape around the hive. It is assumed that within this local landscape, areas closer to the hive will have a stronger influence on pollen and pesticides than areas that are further away. To account for this, a so-called kernel model is used to analyse the landscape around a hive location. The model calculates a distance-weighted impact for each grid cell, where grid cells closer to the hive "weigh heavier" than grid cells further away. The impact of the CLC that is at the hive location has an impact value of 1 and the further away a cell is, the smaller its influence will be. At this stage it is assumed that the influence of a land cover class at the hive location is representative of the experienced foraging landscape used by bees belonging to the hive in the centre. In other words, bees are assumed to use all of their surroundings at the same distance equally. Next, all calculated values are summed for each of the CORINE CLCs which gives a total "impact" of a certain land-use category at the location of the hive. As mentioned, the relevant landscape for pollen is defined as a 65x65 grid (cell size 100m) and for pesticides we take into account an eight times larger landscape on a 512x512 grid. Furthermore, the influence of the surrounding area is assumed to decrease exponentially. Four different spatial kernels were tested to "distribute" the impact of CLCs around the source cells such that impact is halved at distances of 69, 347, 693 and 3466 m.

Principal component analysis and hierarchical clustering of kernel data

Depending on their location, apiaries may "experience" a similar landscape. To investigate relationships between apiary sites in the INSIGNIA study and using the different kernel distances, a Principal Component Analysis (PCA) was performed followed by a clustering analysis. The clustering analysis gives insight into how apiaries are associated to each other in increasingly larger "experienced" landscapes. The analyses were done using the *FactoMineR* package in R (script available from authors upon request). In the PCA analysis,

the number of dimensions (i.e., the number of axes fitted to the data) was reduced so that a minimum of 85% of variance is explained. For all four kernel distances, this resulted in retaining the first seven PC-axes. The principal component scores from the PC-axes were then used as input for a hierarchical clustering analysis.

Plant phenology

Crop and flower phenology are input requirements into the mechanistic model (see Resources box in Figure 4.3). Plant phenology dictates when and which pollen resources are available to bees. In the foraging model, for a given location and date, a CDD10 is calculated (see Pan-European weather data for details). Subsequently, using the CDD10 value, plant phenology is implemented as a filter for scaling the basic resource availability. In the INSIGNIA study, phenology is defined with a Gaussian distribution around the mean optimal flowering CDD10 for a plant species, and in case of an optional second flowering period, with a second Gaussian distribution around the mean CDD10 for the second flowering period. To make the distribution more discrete, values smaller than 0.01 fraction flowering are set to zero. The function is normalised, dividing it by the maximum value (at the mean), resulting in the function for the fraction of blooming plants depending on the CDD10 (x) for a single flowering period:

Equation 1 $$\mathbf{F(x) = e^{-\frac{1(x-\mu)^2}{2\ \sigma^2}}}$$

with m representing the mean optimal flowering CDD10 and s the standard deviation in CDD10 days.

Pollen resources

Utilising the phenology information, the landscape in terms of land-cover classes is converted into the 'raw' resource landscape by setting the presence of the main (pollen) resources (flowering plant species) in each of the land-cover classes (see Resource model balloon in Figure 15). The following section provides a more detailed description of the resource model.

Presence is defined in terms of the amount of pollen available when the species is in full bloom, as pollen density in g per m^2. As spatial units are 1 ha cells, this density applies to the whole cell. For pollen-providing trees, it is more conveniently calculated as the number of individuals per ha divided by 10,000. The absolute amount of resources available in the 1 ha cell is then equal to the tree density * 10,000 (g ha^{-1}). For resources that cannot be assumed to be uniformly distributed within the 1 ha grid cell, a coverage fraction smaller than 1.0 should be set. This will be the case for composite land cover classes that represent a combination of landscape elements, e.g., a mosaic of small agricultural fields and semi-natural elements. The effective pollen density in such an area, for a flower species that occurs in e.g., the semi-natural elements, will then be 1 / coverage * density. Note that for foraging choices made by the bees, this pollen density may be the relevant unit, while for estimating the amount of resources provided by the landscape the absolute amount (for the whole grid cell) may play a role. To define the basic resource landscape, a table with pollen density per plant species against (relevant) land-cover classes is required and a similar table with coverage fractions. These tables thus define $N_{i,j}$ and $c_{i,j}$ as the pollen provider density (ind m^{-2}) and coverage fraction (-) of resource i in land-cover class j, respectively.

The actual effective resource availability R (in g pollen per m^2) provided by species i in a cell with landcover class j at time t is defined as:

Equation 2 $$\mathbf{R_{i,j,t} = F_{i,t} \cdot 1/c_{i,j} \cdot P_{i,j}}$$

The phenology of the species acts as a filter (*F*, between 0 and 1) scaling the pollen resource availability. The absolute amount of resource G in kg per ha offered by the resource over the whole grid cell is then given by:

Equation 3 $$\mathbf{G_{i,j,t} = F_{i,t} \cdot P_{i,j} \cdot \frac{10000}{1000}}$$

The pollen density while optimally flowering, $P_{i,j}$, is the product of the pollen provider density, $N_{i,j}$, and characteristics of the pollen provider that do not depend on habitat nor time, S_i (g pollen ind^{-1}):

Equation 4 $$\mathbf{P_{i,j} = N_{i,j} \cdot S_i}$$

This S_i is defined as the product:

Equation 5 $$\mathbf{S_i = flower(head)s/individual \cdot pollen/flower(head)s}$$

As input for the model, the abundance of the specific pollen provider, the coverage fraction of the pollen provider's habitat and the invariant pollen provider characteristics thus need to be defined (Table 4).

Table 4. Abundance N_{ij} [ind m^{-2}] of plant species i in each CLC j, fraction coverage of the pollen provider habitat within the grid cell c_{ij} and the invariant pollen-provider properties (amount of pollen per individual plant) S_i.

Species *i*	$N_{i,1}$	$c_{i,1}$	...	$N_{i,j}$	$c_{i,j}$	S_i	

Exposure landscape

The exposure to pesticides in the mechanistic model (see "eFate model" balloon in Figure 15) is based on 11 CORINE CLCs that refer to agricultural use (Table 5). For six of these 11, the actual crop grown is at least broadly defined. For the other classes, additional (national, regional) data would be needed to assign a crop to them. Even when a (dominant) crop would be known, additional data would be needed on the use of pesticides (e.g., active substances, application rates, number of applications) on this crop. Given the lack of such data, we do not attempt to define an environmental fate model. Rather, we take a simplified approach based on the kernel calculations described in the section "Kernel model" and using only agricultural CORINE CLCs. With additional information on the application rate of a pesticide in a given CLC and of the fraction of the applied substance that will be distributed outside the area of direct application and of the relationship between deposition and residue in pollen (see box "Exposure' in Figure 15), the summed impact can be converted into expected residues at the hive.

Table 5. Land-cover classes in the CORINE database referring to agricultural use

Land-cover class	*Agricultural use*
211	Non-irrigated arable land
212	Permanently irrigated land
213	Rice fields
221	Vineyards
222	Fruit trees and berry plantations
223	Olive groves
231	Pastures
241	Annual crops associated with permanent crops
242	Complex cultivation patterns
243	Land principally occupied by agriculture with significant areas of natural vegetation
244	Agro-forestry areas

Pollen foraging

With the outputs from the "Resource" and "eFate" models, the "foraging model" can be run (see Figure 15). The outputs from the foraging model are: predicted exposure (residues); predicted mixture risk; predicted pollen composition; and the predicted effective resource density (Figure 15). The outputs can be compared against the monitoring data, and the model can be iteratively refined. Once a best-performing model has been established, the outputs can be used to produce exposure and risk maps, as well as effective resource density maps (Figure 15). The following provides an overview of the foraging model.

For honey bee pollen foragers, we assume it is the amount of pollen protein that can be obtained from a pollen resource that determines its attractiveness (Liolios et al., 2016). This amount is determined by both the protein content of the pollen and the presence of the resource in the surrounding landscape. Gain of the pollen foraging is the amount of protein collected. To be able to deal with it in a framework balancing gains and costs of collecting the pollen protein, gains and costs need to be expressed in a common currency, for which we use energy (in J). Energy gain is thus the amount of pollen collected, multiplied by its protein content and the conversion factor of protein to energy (17 kJ/g). Energetic costs are the transport costs (unloaded flight from hive to foraging location and loaded flight back to the hive) and the foraging itself. For the latter we apply a single coefficient, lumping the outcome of the processes of movement from flower to flower and from plant to plant, and the actual extraction of the pollen from each flower. The rate with which pollen is collected (*f*, in mg s^{-1}) depends also on the resource density (g m^{-2}), following a functional response function that puts an upper limit on the rate of collection:

Equation 6 $$\mathbf{f = \frac{aR*1000}{1+ahR*1000}}$$

Here pollen density (*R*) is converted to mg per m^2, *a* represents the attack rate (m^2 s^{-1}), and *h* the handling time (s mg^{-1}). NB compared to the nectar foraging model of Baveco et al. (2016) here the species-specific differences in time required to handle a single flower (head) are ignored. The time required to collect a full load of pollen from the resource patch is then:

Equation 7 $$\mathbf{t_L = \frac{\gamma}{f}}$$

with γ representing the full capacity (mg) of the forager. Selection of resource patches to exploit is based on the net energetic efficiency (*NEE*) associated with each. Net energetic efficiency is defined as:

Equation 8 $$\mathbf{NEE} = \frac{\mathbf{EI-EE}}{\mathbf{EE}}$$

with *EI* representing energy intake (in J), the energetic gain, and *EE* energy expenditure (in J), the energetic costs. Each resource patch in the landscape that surrounds a hive will have a specific value of *NEE*. From all resource patches, the ones with the highest *NEE* will be selected. As a colony is often exploiting several pollen resources at the same time, more than 1 resource may be selected. The number of pollen foragers is divided over the resource patches, proportional to their *NEE*. Each hourly timestep of the model, an absolute amount of pollen is collected of a particular pollen providing species and originating from a specific location. Of the same species, pollen may be collected from different locations (grid cells). Similarly, at the same location (grid cell) pollen may be collected from different plant species. Note that in our terminology, "resource patch" refers to a species-grid cell combination.

In the resource patches, the pollen foraging in one hour lowers the available pollen density the next hour. With a single colony and grid cells of default 1 ha size, the impact of foraging on density may be small. However, when pollen densities are low, or only a fraction of the 1 ha cell is suitable habitat for the pollen provided, or an apiary with multiple colonies is considered, or a considerable background competition for pollen is assumed, foraging might well lead to severe resource depletion in the exploited patches, resulting in spatial shifts of foraging effort over the landscape.

Pollen and pollutant prediction model

Gradient boosting machine

The goal for predictive models here is to predict a number (pollen, pesticides) or a quantity (metals), for which Gradient Boosting Machine (GBM) models are ideally suited. GBM models were fitted using the *gbm* package (Ridgeway, 2024; Ridgeway and GBM Developers, 2024) in R (R Core Team, 2023). GBM models are based on the concepts of machine learning presented by e.g., Friedman et al. (2000); Friedman (2001, 2002). The basic approach of a GBM model is that, without going into the mathematical details described in (Ridgeway, 2024), it builds an ensemble of shallow trees, where each subsequent tree is trained on the residuals of the previous tree. In other words, the algorithm starts with a weak predictor and builds on this with (many) additional weak predictors which together make up a strong predictor ensemble (Figure 18).

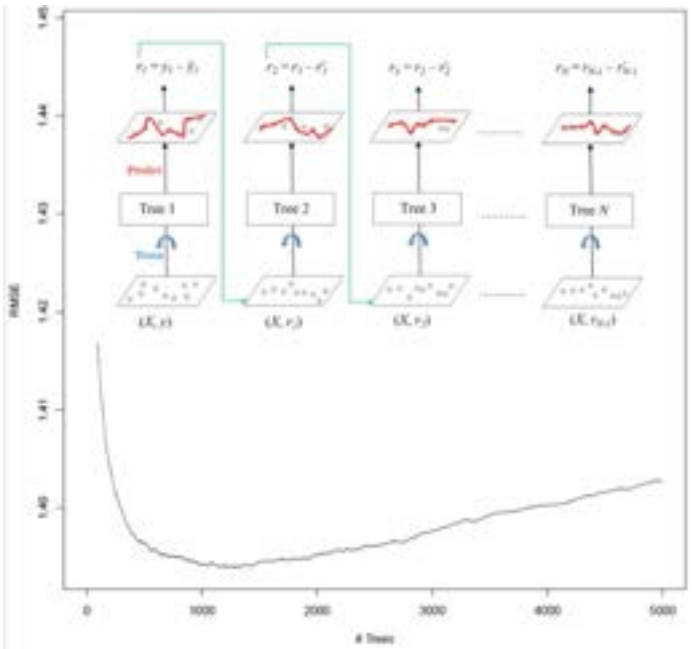


Figure 18. Graphical representation of a gradient boosting machine model. Each individual tree is a weak learner. From the second tree onwards, trees are trained on residuals of the preceding tree. Thus the model 'learns' from previous mistakes. With an increasing number of iterations the prediction error is reduced, up to the point where the model starts to overfit the data and the error tends to increase again. Here, the goal of fitting is to find the number of iterations (i.e., the number of trees) where the loss function (error) is minimised. Figure inset from: https://www.geeksforgeeks.org/ml-gradient-boosting/.

There are a number of parameters that can be optimised in the fitting procedure, the optimised values of which depend on the data set, the most important variables and their optimised values are: 1) the learning rate, the smaller the learning rate, the smaller the change in RMSE per iteration but the longer fitting will take; 2) the maximum number of splits in a tree, a lower number results in a more shallow tree, but this trades off against the number of trees needed; 3) the minimum number of observations in a terminal node, higher values prevent the model from learning highly specific relationships and puts emphasis on generic trends in the data; and finally 4) the bag fraction, the fraction of the training set observations randomly selected to propose the next tree in the expansion, this introduces randomness into the model fit.

A final predictive GBM model was fitted for each of the matrices, using the optimised parameters and all available data. To test the influence of each of the explanatory variables on the model results, predictions were made by iteratively permuting one predictor variable and logging the change in the RMSE. For important variables a large change in the RMSE is expected, whereas for variables that carry less information smaller changes are expected. Note importance of a variable does not indicate the relationship to the number of predicted plant families. These relationships can be strongly non-linear and depend on the interactions between all variables in the model.

Fitting models to INSIGNIA pollen and pollutant data

The number of pollen, pesticides and quantities of metals in a sample were calculated. Distance weighted CORINE CLC data and CDD10 data (see Figure 15) were collected and linked to individual observations, so that each observation is associated with the landscape around the respective hives, and cumulative temperature for pollen. The long distance kernel (see section "Kernel model") was deemed most suitable to describe the surrounding landscape of a hive. Based on the long-distance kernel, CORINE CLC impact was calculated

for each apiary (see section "Kernel model"). Note that, in contrast to the mechanistic model, no assumptions on the likely sources of pollen and / or pollutant are made with respect to the different landcover classes, and therefore here all 44 CORINE CLCs are used. CDD10 information was extracted for the observation's location and date (see section "Pan-European weather data") for pollen.

Country-scale prediction of pollen and pollutants to produce maps

The objective of the final model is to make predictions for five months of the year (i.e. May – September). The required data (see Figure 15) was collected based on a 10km spacing of points for which CORINE CLC kernel impacts were extracted using the long distance kernel.

For every 10 km grid cell the model predicts the pollen genera/families, pesticide numbers and metal quantities present.

References

Baveco, J. M., A. Focks, D. Belgers, J. J. M. van der Steen, J. J. T. I. Boesten, & I. Roessink. (2016). An energetics-based honey bee nectar-foraging model used to assess the potential for landscape-level pesticide exposure dilution. *PeerJ* 4: e2293.

Beekman, M., & Ratnieks, F. L. W. (2000). Long-range foraging by the honey-bee, *Apis mellifera* L. *Functional Ecology*, *14*(4), 490-496.

Bromenshenk, J. J., Carlson, S. R., Simpson, J. C., & Thomas, J. M. (1985). Pollution monitoring of Puget Sound with honey bees. *Science*, *227*(4687), 632-634.

Chauzat, M. P., Faucon, J. P., Martel, A. C., Lachaize, J., Cougoule, N., & Aubert, M. (2006). A survey of pesticide residues in pollen loads collected by honey bees in France. *Journal of Economic Entomology*, *99*(2), 253-262.

Conti, M. E., & Botrè, F. (2001). Honey bees and their products as potential bioindicators of heavy metal contamination. *Environmental Monitoring and Assessment*, *69*(3), 267-282.

Crailsheim, K., Schneider, L. H. W., Hrassnigg, N., Bühlmann, G., Brosch, U., Gmeinbauer, R., & Schöffmann, B. (1992). Pollen consumption and utilization in worker honey bees (*Apis mellifera carnica*): dependence on individual age and function. *Journal of Insect Physiology*, *38*(6), 409-419.

Friedman J, Hastie T and Tibshirani R, 2000. Special Invited Paper. Additive Logistic Regression: A Statistical View of Boosting. The Annals of Statistics, 28, 337-374.

Friedman JH, 2001. Greedy function approximation: A gradient boosting machine. The Annals of Statistics, 29, 1189-1232, 1144.

Friedman JH, 2002. Stochastic gradient boosting. Computational Statistics & Data Analysis, 38, 367-378.

Grimm, V., Augusiak, J., Focks, A., Frank, B. M., Gabsi, F., Johnston, A. S., ... & Railsback, S. F. (2014). Towards better modelling and decision support: Documenting model development, testing, and analysis using TRACE. *Ecological modelling*, 280, 129-139.

Hofman, V., & Solseng, E. (2004). Spray equipment and calibration. NDSI Extension service North Dakota State University. Sept 2004

Johnson, R. M., Ellis, M. D., Mullin, C. A., & Frazier, M. (2010). Pesticides and honey bee toxicity – USA. *Apidologie*, *41*(3), 312-331.

Liolios, V., C. Tananaki, M. Dimou, D. Kanelis, G. Goras, E. Karazafiris, & A. Thrasyvoulou. (2016). Ranking pollen from bee plants according to their protein contribution to honey bees. *Journal of Apicultural Research* 54:1-11.

Manzano, C.A., Dodder, N.G., Hoh, E. & Morales, R. (2019). Patterns of Personal Exposure to Urban Pollutants Using Personal Passive Samplers and GC × GC/ToF−MS. *Environmental Science & Technology*, *53*, 614-624.

Ridgeway G 2024. Generalized Boosted Models: A guide to the gbm package.

Ridgeway G and GBM Developers, 2024. gbm: Generalized Boosted Regression Models.

Schmolke, A., Thorbek, P., DeAngelis, D. L., & Grimm, V. (2010). Ecological models supporting environmental decision making: a strategy for the future. *Trends in Ecology & Evolution*, 25(8), 479-486.
Garbuzov, M., Schürch, R., & Ratnieks, F. L. (2015). Eating locally: dance decoding demonstrates that urban honey bees in Brighton, UK, forage mainly in the surrounding urban area. *Urban Ecosystems*, *18*(2), 411-418.

Murcia-Morales, M., Van der Steen, J. J., Vejsnæs, F., Díaz-Galiano, F. J., Flores, J. M., & Fernández-Alba, A. R. (2020). APIStrip, a new tool for environmental contaminant sampling through honey bee colonies. *Science of The Total Environment*, *729*, 138948.

Quaresma, A., Ankenbrand, M.J., Garcia, C.A.Y. et al. (2024). Semi-automated sequence curation for reliable reference datasets in ITS2 vascular plant DNA (meta-)barcoding. *Sci Data* 11, 129. https://doi.org/10.1038/s41597-024-02962-5.

Roessink, I., & van der Steen, J.J.M. (2021) Beebread consumption by honey bees is fast: results of a six-weeks field-study. *Journal of Apicultural Research.* 60(5): 659–664..

Samon, S.M., Hammel, S.C., Stapleton, H.M., Anderson, K.A. (2022). Silicone wristbands as personal passive sampling devices: Current knowledge, recommendations for use, and future directions. *Environment International*, *169*, 107339.

Schatz, F., & Wallner, K. (2009). *Pflanzenschutzmittelapplikation in blühenden raps (*Brassica napus*) und deren auswirkungen auf die rückstandssituation in honig, nektar und pollen der honigbiene (*Apis mellifera *L.)* Doctoral dissertation, Thesis (Diploma), Universität Hohenheim, Hohenheim, Germany.

Steen, J. J.M van der, Cornelissen, B., Donders, J., Blacquière, T., & van Dooremalen, C. (2012). How honey bees of successive age classes are distributed over a one storey, ten frames hive. *Journal of Apicultural Research*, *51*(2), 174-178.

Steen, J. J. M. van der, Cornelissen, B., Blacquière, T., Pijnenburg, J. E. M. L., & Severijnen, M. (2016). Think regionally, act locally: metals in honey bee workers in the Netherlands (surveillance study 2008). *Environmental Monitoring and Assessment, 188*(8), 463.

Visscher, P. K., & Seeley, T. D. (1982). Foraging strategy of honey bee colonies in a temperate deciduous forest. *Ecology*, *63*(6), 1790-1801.

Waddington, K. D., Herbert, T. J., Visscher, P. K., & Richter, M. R. (1994). Comparisons of forager distributions from matched honey bee colonies in suburban environments. *Behavioural Ecology and Sociobiology*, *35*(6), 423-429.

Winston, M. L. (1991). *The biology of the honey bee*. Harvard University Press.