Ecological Informatics 82 (2024) 102713

Contents lists available at ScienceDirect

Ecological Informatics

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

A data platform for real-time monitoring and analysis of the brown marmorated stink bug in Northern Italy

Chiara Forresi [a], Enrico Gallinucci [a,*], Matteo Golfarelli [a], Lara Maistrello [b,c], Michele Preti [d], Giacomo Vaccari [e]

[a] *Department of Computer Science and Engineering, University of Bologna, Via dell'Universita 50, Cesena 47522, Italy*
[b] *Department of Life Sciences, University of Modena and Reggio Emilia, Via Amendola 2, Reggio Emilia 42122, Italy*
[c] *NBFC, National Biodiversity Future Center, Piazza Marina 61, Palermo 90133, Italy*
[d] *ASTRA Innovazione e Sviluppo srl, Via Tebano 45, Faenza 48018, Italy*
[e] *Consorzio Fitosanitario Provinciale di Modena, Via Santi 14, Modena 41123, Italy*

ARTICLE INFO

Keywords:
Halyomorpha halys
Integrated pest management
Data platform
Sustainable agriculture
Precision agriculture
Decision support system

ABSTRACT

The brown marmorated stink bug (*Halyomorpha halys*) is one of the main insect pest species causing economic damage to several agricultural commodities worldwide and one of the worst threats to tree fruit crops in northern Italy, especially in the Emilia-Romagna region. Previous efforts in implementing *H. halys* surveillance at the regional level were mainly focused on studying the *H. halys* phenology, but they were not designed to provide a public service. In this paper, we propose a data-driven approach to support the application of Integrated Pest Management strategies against *H. halys*. The proposal is based on the experience of a three-year project in which a network of monitoring traps has been deployed throughout the whole Emilia-Romagna region and a data platform has been implemented to enable the real-time tracking of *H. halys* occurrence and distribution, integrating these information with multiple data sources, and analytical capabilities through a public website. Besides the real-time pest surveillance, the data platform allowed us to increase our understanding about *H. halys* seasonal invasion dynamics and the main factors contributing to its spread. The results will help individual growers in protecting their crops and the whole region in promoting more efficient usage of insecticides and more sustainable and healthy agricultural productions.

1. Introduction

Recent years have seen an increasing interest from the research community towards the development and advancement of precision agriculture techniques (Gebbers and Adamchuk, 2010; Kamilaris et al., 2017). Several goals are pursued, including the reduction of resource consumption, the increase in sustainability of agricultural practices, and the improvement of counter-actions against insect pests that jeopardize the agricultural yields (Gallinucci et al., 2020). Indeed, land management issues resulting from intensive agriculture have promoted the need to rethink agricultural systems to make them socially and environmentally sustainable (Kloppenburg, 2004; Mcintyre et al., 2009). While there is considerable margin to pursue agricultural sustainability by mitigating the high levels of waste within the food supply chain, the debate on how to make technologies work for sustainability fits into this perspective (Bandh, 2021). In addition, recent research on the social context of data linkage has focused specifically on the use of data resources and technologies to challenge harmful models of intensive agriculture, including a reflection on how the incentive to use land intensively, which is the assumption of most digital technologies and data collection systems, needs to be linked to a more critical perspective that takes into account the ecological damage caused by agriculture (Leonelli and Williamson, 2023a; Leonelli and Williamson, 2023b).

One of the main threats to the sustainability of agricultural practices comes from invasive alien species, which adversely affect biodiversity and associated ecosystem services (Blackburn et al., 2019; Clarke and McGeoch, 2023). Specifically, invasive insects are the most responsible for the costs to the global economy and a major cause of damage to agricultural commodities worldwide (Bradshaw et al., 2016; Paini et al., 2016; Renault et al., 2022). Preventing and mitigating the negative

* Corresponding author.
E-mail addresses: chiara.forresi@unibo.it (C. Forresi), enrico.gallinucci@unibo.it (E. Gallinucci), matteo.golfarelli@unibo.it (M. Golfarelli), lara.maistrello@unimore.it (L. Maistrello), michele.preti@astrainnovazione.it (M. Preti), giacomo.vaccari@regione.emilia-romagna.it (G. Vaccari).

https://doi.org/10.1016/j.ecoinf.2024.102713
Received 9 February 2024; Received in revised form 16 May 2024; Accepted 28 June 2024
Available online 6 July 2024

impacts of invasive species is a top priority worldwide (European Parliament and Council, 2014; United States Department of Interior, 2016), as they are also seen as a threat to the success of adaptation to climate change (Invasive Species Advisory Council, 2023). Indeed, the importance of sampling and monitoring is crucial to mitigating the impact of invasive insect species (Morey and Venette, 2023).

Halyomorpha halys Stål 1855 (Heteroptera: Pentatomidae) (Taxonomic Serial No. 915660, ITIS.gov) is an invasive insect that currently represents one of the greatest threats to tree fruit crops in northern Italy. Also known as the brown marmorated stink bug, *H. halys* is a highly polyphagous agricultural insect pest that causes important economic losses to several horticultural, vegetable, and arable crops, as well as ornamental plants (Lee et al., 2013; Leskey et al., 2012). In addition, *H. halys* is considered an urban nuisance since in autumn it invades buildings to overwinter, causing problems not only to the growers but also to the citizens (Hancock et al., 2019; McPherson, 2018). Due to the high invasive capacity facilitated by trade and other human activities (Maistrello et al., 2018), this species is currently found on every continent in the northern hemisphere, North Africa, and South America. In Europe, it was recorded for the first time in 2004 in Zurich, Switzerland (Haye et al., 2014). In Italy, *H. halys* was first detected in 2012 (Lara and Paride, 2014) in Emilia-Romagna, which has been the earliest Italian region suffering economic damages caused by *H. halys* in tree fruit crops (Maistrello et al., 2017). In northern Italy, *H. halys* in 2019 caused about €740 million losses in fruit production, of which about €270 million in the Emilia-Romagna region; together with the direct damages caused to the agricultural commodities, consequent indirect losses have been recorded in several horticultural supply chains, where the lack of fruits (which were not harvested due to *H. halys* damages) in 2019 caused an estimated loss of over half a million working days (Cimice asiatica, 2022). Ten years after its first detection (Maistrello et al., 2016), *H. halys* can still be considered a phytosanitary emergency ubiquitous in all Emilia-Romagna provinces, due to its peculiar characteristics: great invasive capacity, easy adaptability, extreme polyphagia, high mobility, and occurrence both inside and outside agricultural areas.

In this paper, we support the application of Integrated Pest Management (IPM) strategies against *H. halys* in the Emilia-Romagna region by proposing a data-driven approach based on the experience of a three-year project. In particular, the goal of our research work is to provide growers and pest control advisors with an effective and reliable tool to monitor in real-time the *H. halys* seasonal invasion dynamics, its abundance and phenology across the region, and to improve the understanding of the main factors contributing to its spread. This goal is achieved by means of the following novel contributions.

1. The creation of an open source *H. halys* monitoring network covering the different provinces of the Emilia-Romagna region, standardizing the sampling techniques and the data collection in order to create and publish an area-wide reliable historical dataset. Previous efforts in implementing *H. halys* surveillance at regional level that involved the citizen science approach (Malek et al., 2018; Malek et al., 2019) were mainly focused on studying the *H. halys* phenology, but they were not designed to provide a public service. In the present network of traps covering different agroecosystems in the various provinces, one major goal was to deliver standardized information to the stakeholders; indeed, before this project, no standardization in terms of monitoring methods was considered among researchers, pest control advisors, and growers at the regional scale.
2. The implementation and deployment of a data platform for the real-time monitoring and analysis of *H. halys* occurrence and distribution. The data platform integrates data from the trap network with multiple environmental data sources to uncover correlated factors that describe the *H. halys* spread. These data sources include information on the weather parameters and the distribution of water resources, crops, buildings, and spontaneous vegetation close to the monitored sites. The integrated data are available to growers and pest control advisors via a web application to provide both tactical and strategic support. The web application and all datasets (including the code to process and analyze them) are publicly available (see point 4). To the best of our knowledge, this is the first proposal for a public data platform to support *H. halys* management practices.
3. The execution of a series of analytical evaluations to infer useful knowledge on the main environmental parameters driving the *H. halys* occurrence and distribution, and to demonstrate the value of a data-driven approach to agricultural problems. These results will not only help individual growers in protecting their crops, but will also contribute to the growth of the region as a whole by encouraging more efficient use of insecticides and thus more sustainable and healthier agricultural production.
4. The publication of the following open resources.

- The real-time monitoring web application: `https://big.csr.unibo.it/projects/stink-bug/monitoring/`.
- Open data and data processing code[1]: `https://github.com/big-unibo/stink-bug`.

The knowledge gained from our proposal will be crucial for the implementation of cost-effective management techniques, including the optimization of insecticide use and the biological control approach (mainly based on egg parasitoids, such as the exotic *Trissolcus japonicus*) through controlled releases of natural enemies where and when needed (Zapponi et al., 2021). Indeed, recent research work has demonstrated that the combination of on-field data with environmental information can prove decisive in estimating and predicting insect spread and in implementing effective control strategies in line with IPM guidelines (Bono Rosselló et al., 2023).

The paper outline is as follows. Section 2 summarizes the related literature, Section 3 describes our approach in full details, and Section 4 presents the obtained results; conclusions are drawn in Section 5. Finally, Appendix A digs deeper into the data and storage organization of the data platform.

2. Related literature

Integrated Pest Management programs rely on the mutual integration of several control methods and complementary practices to effectively protect agricultural crops from insect pests and plant diseases (Elliott et al., 1995; Flint and Van den Bosch, 2012). For instance, to keep an insect pest population at a level that is under the economic damage threshold, the use of insecticides is combined with other control tools, such as the application of good agronomical practices and the implementation of biological control dynamics. In order to minimize and rationalize the use of plant protection products, and consequently reduce the adverse side effects that they may cause, a key factor is to know exactly where and when to intervene. Field monitoring is therefore the basis to support any decision, allowing the prediction of pest outbreaks and consequently to act timely, choosing and effectively applying the appropriate intervention methods before any crop damage is caused (Dent and Binks, 2020). While reliable monitoring tools allow robust forecasting, the success of IPM strategies depends on the precision and effectiveness of the field monitoring programs. It is crucial to select a proper sampling technique to detect the occurrence of any insect pest species, to estimate their population size, and to measure changes in their abundance. Among the possible monitoring method options, the use of pheromone-baited traps is one of the most adopted worldwide (Prasad and Prabhakar, 2012). Insect pheromones are volatile organic compounds, usually very species-specific since they are chemical signals commonly used in the intra-specific communication between insects; their use as baits for monitoring traps guarantees a high selectivity of the

[1] Persistent identifier: `https://zenodo.org/records/10817206`

insect captures (Bell and Cardé, 2013).

Effective monitoring of the invasive pest *H. halys* is nowadays possible thanks to the notable and significant advances in the knowledge of its chemical ecology. *H. halys* aggregation pheromone has been identified and synthesized, a synergism with the pheromone of another Asian stink bug species has been documented, and several trap designs have been developed to provide the market with low-cost *H. halys* monitoring traps and lure solutions (Weber et al., 2017). The *H. halys* male-produced aggregation pheromone is a 3.5:1 mixture of two stereoisomers, (3*S*,6*S*,7*R*,10*S*)-10,11-epoxy-1-bisabolen-3-ol and (3*R*,6*S*,7*R*,10*S*)-10,11-epoxy-1-bisabolen-3-ol (known as murgantiol). This blend is synergized by methyl (2*E*,4*E*,6*Z*)-2,4,6-decatrienoate (known as MDT), which is the aggregation pheromone of the brown-winged green stink bug, *Plautia stali* Scott, providing a reliable attractant for *H. halys*, both juvenile forms and adults. Different trap designs have been tested to retain *H. halys* individuals (Rice et al., 2018) and the pyramid traps proved to be one of the most effective monitoring tools available in the market (Acebes-Doria et al., 2018). Several studies have been carried out to acquire further knowledge about *H. halys* monitoring systems, including the trap plume reach and the trapping area (Kirkpatrick et al., 2019), the effects of border habitat on the *H. halys* captures (Bergh et al., 2021), and the optimization of the pheromone lures (Leskey et al., 2021).

The monitoring of *H. halys* has been extensively exploited in several countries where this invasive pest has been detected, including North America and Europe (Morrison et al., 2017; Nielsen et al., 2013; Nielsen and Hamilton, 2009), as well as in the area of its origin (Lee et al., 2013). Extensive surveillance of *H. halys* at the regional level has already been investigated also in Italy, involving the citizen science approach (Maistrello et al., 2016; Maistrello et al., 2018; Malek et al., 2019), as it has been explored in other countries (Vétek et al., 2021). However, monitoring networks based on citizen science were not designed to provide a public service, but were mainly focused on studying the *H. halys* phenology and updating the spread of invasive populations in countries of introduction.

In the Veneto region (Italy), the public service “Veneto Agricoltura” provides an online weekly bulletin to inform farmers and pest management advisors about the status of *H. halys* in the area (Agenzia Veneta per l’innovazione nel Settore primario, 2024). This bulletin is similar to the one provided in our research project, but in this case there is no freely accessible map with individual *H. halys* capture data to visualize and download, as proposed in our network. In the U.S.A., the occurrence of *H. halys* is reported on a state-by-state basis (StopBMSB.org, 2024), which shows on a map the risk level of each state and also provides useful information such as the state-by-state occurrence of the main natural control agent of *H. halys*, *Trissolcus japonicus*. In this case, the website is rich in fact sheets and material to increase growers’ awareness and knowledge of this pest, but reports on pest distribution are at a national level. In our research project, we present a capillary distribution of the pest occurrence and infestation level, showing trap data at the farm level, with information useful for stakeholders in a given territory and province within the whole Emilia-Romagna region.

At the local scale in Emilia-Romagna (mainly in the province of Modena where this exotic species was initially recorded), *H. halys* occurrence has been constantly recorded over an 8-year monitoring period (2015–2022). Nevertheless, up to date, a network of traps covering different agroecosystems in the various provinces was lacking, as no standardization in terms of monitoring methods was considered among researchers, pest management advisors, and growers at a regional scale.

Additionally to the standardized monitoring method, one of our main contributions is the data-driven approach supported by a data platform to carry out analytical activities on *H. halys* spread, integrated with information from a wide range of data sources. In this paper, we fully discuss topics including the architecture of the data platform and the management and integration of data. While the interest in approaches that support precision agriculture activities is rapidly increasing, these topics are rarely the focus of related work. While many approaches are dedicated to sensors (Ojha et al., 2015) and remote sensing (Almstedt et al., 2023; Vibhute and Bodhe, 2012), the majority of related papers focus on the application of machine learning techniques to ad hoc agricultural datasets (Kamilaris et al., 2017). An open and integrated *cyber-physical infrastructure* was presented in (Chen et al., 2015), where heterogeneous monitoring sensors were integrated with different precision agriculture applications through a middleware built on open standards. However, the focus was mainly on the problem of data collection and no extensive support for analytical features was provided. Another work in this direction is (Sawant et al., 2016); here, the PRIDE (Progressive Rural Integrated Digital Enterprise) business model is presented to promote the advantages of providing timely and integrated data for precision agriculture applications, but the architecture and data model are not discussed in details.

More recently, a hybrid architecture for tactical and strategic precision agriculture called Mo.Re.Farming has been proposed in (Gallinucci et al., 2020) (and generalized in (Francia et al., 2021)). The goal of Mo.Re.Farming is to provide a Decision Support System for agricultural technicians and to enable analyses related to the use of water and chemical resources, in terms of optimization and environmental impact; also, it serves as a hub for agricultural data, collected and integrated from both public and private sources. The work presented in this paper builds on the results presented in (Gallinucci et al., 2020), extends it to further data sources, and applies it to the study of *H. halys* pest species to support the IPM strategies in the Emilia-Romagna region.

Ultimately, we register few efforts to design data platforms to support pest monitoring and analysis activities - not necessarily on *H. halys*. In (Liu et al., 2022), a big data platform to monitor generic diseases and pests in the Guizhou province of China is presented; it is based on just 10 traps, it is deployed in crops of hot peppers, and very few high-level details of the platform are disclosed. Similarly, little information is provided about the data platform envisioned in (Lina and Xiuming, 2020), whose main focus is an image classification algorithm for pest detection. (Wang et al., 2020) presents only a dataset for pest detection through deep learning techniques, while (Vanegas et al., 2018) proposes a method based on unmanned aerial vehicles to detect pests, but does not disclose details about the underlying platform. Finally, (Lagos-Ortiz et al., 2020) presents AgriEnt as a knowledge-based web platform for managing pests: its main focus is the ontological model that aids decision-making for insect diagnosis and treatment, but it is not coupled with a monitoring network and is envisioned as a web service to provide support to farmers based on pre-elaborated information.

3. Materials and methods

The study was conducted over three consecutive years (2020, 2021, and 2022) in the plain of the Emilia-Romagna region (Italy). Fig. 1 shows the duration of the monitoring sessions: in 2020, it began on June 29 and finished on October 18; in 2021, it began on February 26 and finished on October 24; in 2022, it began on March 21 and finished on October 23.

In the following, we present all the details about the monitoring network and the data platform in Sections 3.1 and 3.2, respectively.

3.1. The monitoring network

A total of 145, 165, and 101 farms specialized in tree fruit crops production were selected for the survey in 2020, 2021, and 2022, respectively; 43 of these farms were the same over the three-year survey. Most of the farms were located in the central-eastern part of the Emilia-Romagna region. This area is characterized by intensive agriculture and has a high level of specialization in fruit production: pome fruits (pears and apples), stone fruits (peaches and nectarines, apricots, plums, and cherries), kiwifruit and grapevine are the main crops, followed by other

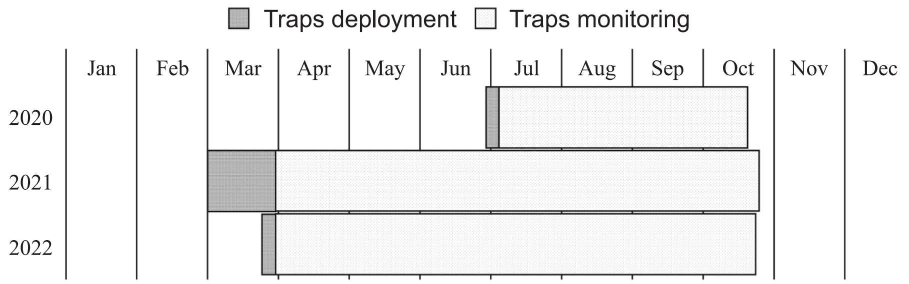


Fig. 1. Duration of monitoring sessions over the three-year project.

crops such as persimmon, walnut, and olive. All these crops can be considered host crops for the target pest *H. halys*, with a variable level of preference according to the plant's phenological stage, the cultivar, and the agroecosystem landscape. The geographical distribution of the traps over the three years is shown in Fig. 2.

3.1.1. Trap setup and monitoring system

In each farm, a black pyramid trap (Dead-Inn Trap, AgBio Inc., Westminster, CO, USA) was installed to monitor *H. halys* (an example is shown in Fig. 3). Traps were baited with a commercial aggregation pheromone dual lure (Pherocon BMSB DUAL Lure, Trécé Inc., Adair, OK, USA) containing a mixture of murgantiol and MDT and specifically attractive for *H. halys*. Lures were replaced at about 90-day intervals, as indicated by the manufacturer (12-week lure longevity). In each location, the monitoring trap was positioned near an orchard, keeping a distance of 5–10 m from the orchard edges and, when possible, near untreated vegetation (e.g., hedges, gardens, uncultivated areas) and buildings.

H. halys specimens captured in the collection jars of the monitoring traps were weekly removed and counted, dividing them into adults (both males and females), second and third age nymphs (small juvenile instars, or simply small instars), and fourth and fifth age nymphs (large juvenile instars, or simply large instars). In each monitoring season, the operators carried out the following tasks.

1. Trap installation: executed at the beginning of each season, it consisted in physically placing the monitoring trap and collecting information about the environmental elements in its surroundings.
2. Trap monitoring: performed on a weekly basis, it consisted in manually counting and removing the *H. halys* individuals captured by the trap. Each trap was consistently monitored on the same weekday (e.g., Monday) throughout the season. Information about the trap functionality, the number of *H. halys* individuals captured, divided into adults, small instars, and large instars, and the occurrence of other non-target species captures were weekly collected from each monitoring site.

3.1.2. On-field data crowdsourcing with CASE

The acquisition of data concerning the installation and monitoring of *H. halys* traps have been aided by CASE (Collaborative Agro SEnsing), a dynamic-questionnaire application for on-field data crowdsourcing in the agricultural domain. CASE has been developed to facilitate and standardize the communications between on-field operators with first-hand visuals of a given field/orchard and the technicians needing a 360-degree view of all fields/orchards for real-time monitoring, operational oversight, coordination, and analytical purposes. Details about CASE's databases are provided in Appendix A.1. In the context of this project, CASE has been used to collect data from the operators in charge of trap installation and monitoring activities, with ad-hoc questionnaires being devised for the *H. halys* use case. As shown in Fig. 4, the Android CASE application provides an easy-to-use interface for operators to register and automatically validate the data, thus minimizing the risk of incorrect data registration. CASE enables the registration and geopositioning of traps at installation time by automatically retrieving the latitude and longitude values from the smartphone's GPS. On a weekly basis, on-field operators receive the questionnaires to report trap monitoring information in the form of *tasks* to be fulfilled; CASE provides map-based visualizations to help operators locate the traps that are yet to be monitored in the current week and sends reminders on the yet-unfulfilled tasks.

3.2. The data platform

Fig. 5 gives an overview of the data platform in charge of collecting the data from the aforementioned sources, integrating them, and exposing them to end users through several fruition options. It is implemented on a two-rack cluster of 18 Ubuntu machines with a minimum configuration of i7 8-core CPU @3.2GHz, 32GB RAM, and 6 TB hard disk drives, each machine running the Cloudera Distribution for Apache Hadoop (CDH) 6.2.0.[2]

Data are organized into three tiers, in accordance with multi-tier architectures (Ravat and Zhao, 2019; Zburivsky and Partner, 2021) to logically separate the subsequent processing activities. The Raw tier hosts the *data lake* (Stein and Morrison, 2014), i.e., a storage repository that holds the data in its raw, native format. The Harbor tier provides an integrated and comprehensive view of the available data at the finest level of detail. The Access tier provides a higher-level view of data that is ready to be consumed for analytical purposes. Details about the technological stack and the conceptual/logical representations of the data stored in the different tiers are provided in Appendix A.

In the following, we disclose all the details of the main data platform components, starting with the presentation of all data sources (Section 3.2.1) and continuing with the integration and enrichment processes (Section 3.2.2), the validation and loading processes (Section 3.2.3), and concluding with the analytical fruition (Section 3.2.4).

3.2.1. Data sources

A summary of the many data sources is shown in Table 1, which reports the frequency of updates and space occupation for each source.

The **monitoring network** (managed by CASE, introduced in Section 3.1.2) is the main source, providing the number of *H. halys* captures on a weekly basis. For each deployed trap, we collected the coordinates of deployment and a list of environmental elements visually identified in the surroundings by the technician that deployed the trap (e.g., tree fruit crops, herbaceous crops, buildings). Then, for each deployed trap and week within the monitoring session, we collected the number of *H. halys* captured (divided into small instars, large instars, and adults), together with an indication of the correct functioning of the trap.

Satellite images of the Emilia-Romagna region were collected from the Sentinel-2 satellites of the European Space Agency (ESA). Each image is composed of one or more granules, each covering a squared 100 km^2 area with a 10 m resolution and containing several bands,

[2] The high amount of resources is due to the cluster being used to support several research projects, including the one presented in this paper.

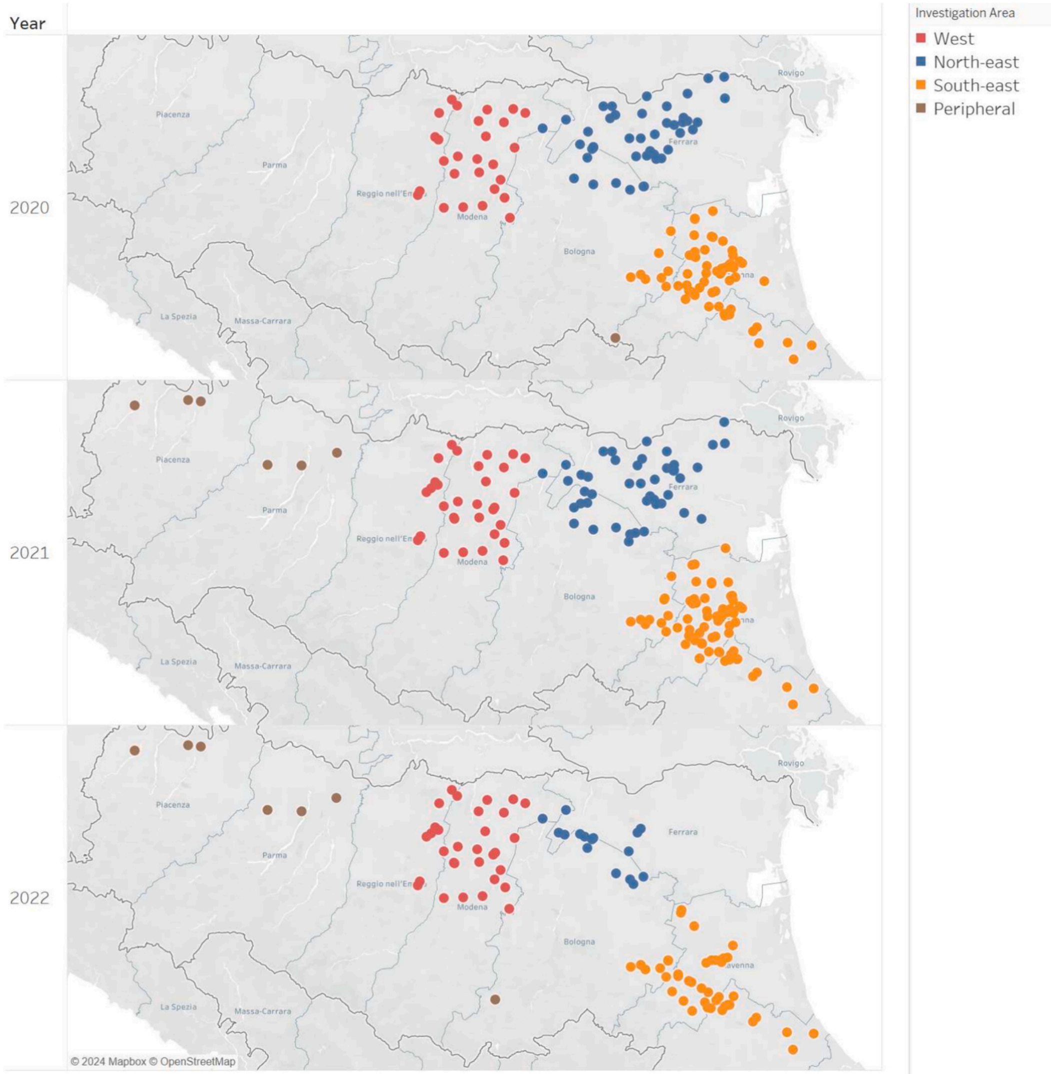


Fig. 2. Geographical distribution of *H. halys* monitoring traps network.

including visible (red, green, blue) and non-visible ones (e.g., infrared). Depending on the satellite's trajectory in space, up to 3 images are available each week. Satellite images are available as open data[3] and enable the computation of vegetation indexes, which are useful to infer the amount of vegetation in the trap's proximity.

Weather data were collected from the Regional Agency for Prevention, Environment, and Water (ARPAE), which publishes them as open data.[4] ARPAE creates a grid of squared 25 km^2 non-overlapping cells and provides daily weather data that are representative for each cell, including air temperature in °C (minimum, average, and maximum), wind speed in m/s (average and maximum), percentage of relative humidity (minimum, average, and maximum), and total precipitations in kg/m^2. The data is aggregated to get weekly totals and averages by the pre-processing tasks that load the data to the Raw tier.

Environment data were collected from the regional consortium for the Emilia-Romagna's irrigation channel (Consorzio Canale Emiliano Romagnolo, CER), which provides yearly snapshots of the main environmental elements in the region. For each year, we obtained a classification of all water basins and channels in the Emilia-Romagna region and a detailed report on all crops being cultivated in each plot of land of the Emilia-Romagna region.

Remarkably, all data are georeferenced, meaning that each record is associated with precise spatial identifiers. *H. halys* captures are identified by the location of each monitoring trap, obtained through the CASE application; satellite images and weather data are described by the coordinates of the granule/cell that they cover; environment elements are associated with geometrical data types (e.g., polygons, lines) describing the shape of the element and locating them on the map.

3.2.2. Integration processes and enrichment of *H. halys* captures

A fundamental aspect of the data platform is the integration of *H. halys* captures with the other data sources, which enables the study of the factors aiding/hindering the presence of *H. halys* individuals. This is achieved by relying on the spatial features of both *H. halys* traps and other sources to find overlaps from the spatial perspective and to extract relevant *enrichments*, i.e., additional information that can be correlated with *H. halys* captures.

The **integration with environment data** allows the identification of the environmental elements in the close surroundings of a trap.[5]

[3] ESA's satellite images available at https://scihub.copernicus.eu/dhus.

[4] ARPAE's weather data available at https://dati.arpae.it/dataset/erg5-interpolazione-su-griglia-di-dati-meteo.

[5] The "close surroundings of a trap" is meant as the circular area with a 200 m radius centered on the trap deployment location.

Fig. 3. Dead-Inn pyramid trap baited with aggregation pheromones used in the *H. halys* monitoring network.

Indeed, *H. halys* individuals may be attracted by specific crops (Zobel et al., 2016) or the presence of water basins that are related to the occurrence of wild vegetation (Moore, 2018).

The **integration with satellite images** allows the calculation of the *spontaneous vegetation percentage* (SVP) in the close surroundings of a trap. Indeed, spontaneous vegetation is a known attraction factor (Venugopal et al., 2015a; Venugopal et al., 2015b). This datum is also asked to the operator that deploys the trap; from here on, we indicate the latter with "SVP (manual)", whereas "SVP (auto)" refers to the one calculated from satellite images. Manually determining the correct SVP value by sight is not trivial, due to the operator's subjectivity and the often inconvenient observation point at the ground level. Satellite images offer a systematic and objective way to calculate it. Several techniques are known in the literature to extract vegetation statistics from satellite images, the most known being the Normalized Difference Vegetation Index[6] (NDVI) (Deering, 1978). In short, the index assigns to each pixel p a numeric value $ndvi(p) \in [-1, 1]$, where negative values indicate an absence of vegetation (Myneni et al., 1995). Based on (Dash and Curran, 2004), we consider $\tau = 0.7$ the minimum value of $ndvi(p)$ for p to be considered representative of spontaneous vegetation. It must be noted that high NDVI values do not necessarily imply the presence of spontaneous vegetation, as they could be measured even for certain crops (e.g., a pear orchard); however, these cases can be avoided by overlapping satellite images with the environment registers and excluding pixels falling within an area of a cultivated field. Ultimately, SVP (auto) for a trap is calculated as $\frac{|p \in P : ndvi(p) \geq \tau|}{|P|}$, where P is the set of pixels in the close surroundings of a trap not falling within any crop in the environment registers.

The **integration with weather data** allows the calculation of the *degree days* (Haye et al., 2014) and the *cumulative degree days* registered at a given trap site during each week of a monitoring session. Degree days are a way to measure the development time needed by an insect species to reach a given status. According to the minimum and maximum insect development temperature thresholds and the actual air temperatures recorded in a given area, it is possible to predict the development period of an insect (from one instar to the following instar, from juvenile forms to the adult stage). In particular, degree days indicate the number of degrees cumulated per day by which the average daily air temperature is higher than a given threshold θ; based on (Haye et al., 2014), we consider $\theta = 12.2\ ^{\circ}\mathrm{C}$. Given a trap t and a day d with an average daily temperature of $avgTemp(d, t)$, they are calculated as $dd(t, d) = max(0, avgTemp(d, t) - \theta)$. Degree days are calculated at the week level by summing the values in the days of the monitored week (i.e., from the day of the previous monitoring — or, if not present, from the day of installation — to the one of the current monitoring). Then, the cumulative degree days $CDD(t, d)$ are the cumulative sum of degree days from the beginning of the season. The temperature values considered by each trap are those associated with the squared cell of 25 km^2 in the weather dataset that contains the trap.

The results of these integration and enrichment activities are materialized in the Harbor tier of the data platform, which provides an integrated and comprehensive view of the available data at the finest level of detail. The relational schema of the integrated data is discussed in Appendix A.2.

3.2.3. Validation and loading of enriched H. halys captures

Data about *H. halys* captures were collected weekly through the CASE application. This activity has been carried out by technicians who physically visited the field to directly check the monitoring traps and then use the app to register data. While this ensured a correct count of the captures (and the CASE app aided technicians in validating the data), some factors that caused non-uniformity in the data have to be considered. On the one hand, despite weekly reminders from the CASE app, technicians were not always able to visit all traps on a weekly basis. On the other hand, malfunctioning of traps may have prevented them from working correctly and returning trustworthy data on a weekly basis; in this case, the possible malfunctioning of a trap has been signaled through CASE.

For these reasons, a validation process is needed to ensure that analytical activities are carried out on clean and consistent data. In particular, we validate *H. halys* captures as follows.

- Traps with at least 10 consecutive weeks of non-entered data or with at least 5 registered malfunctioning were reported as "invalid" and not used in the data analysis.
- Traps with no more than 3 consecutive weeks of non-entered data were tolerated, reported as "noisy" and used in the data analysis.
- In any case, missed monitoring tasks would generate inconsistencies (as the registered number of captures is the result of two or more weeks of trap activity). In these cases, the captures registered with n weeks of non-entered data are divided by n and distributed in equal numbers among the current week and the preceding ones with non-entered data. For instance, let c be the number of *H. halys* adults captured at week w_i after n weeks of non-entered data; then, we normalize the data by registering c/n *H. halys* adults captured in week $w_j\ \forall j \in [i - ni]$. Each capture added with this method is reported as "normalized".

[6] The rationale in NDVI index computation is to compare the reflectance values on different spectral bands, as the chlorophyll in plants' leaves reflects light in different ways under different conditions.

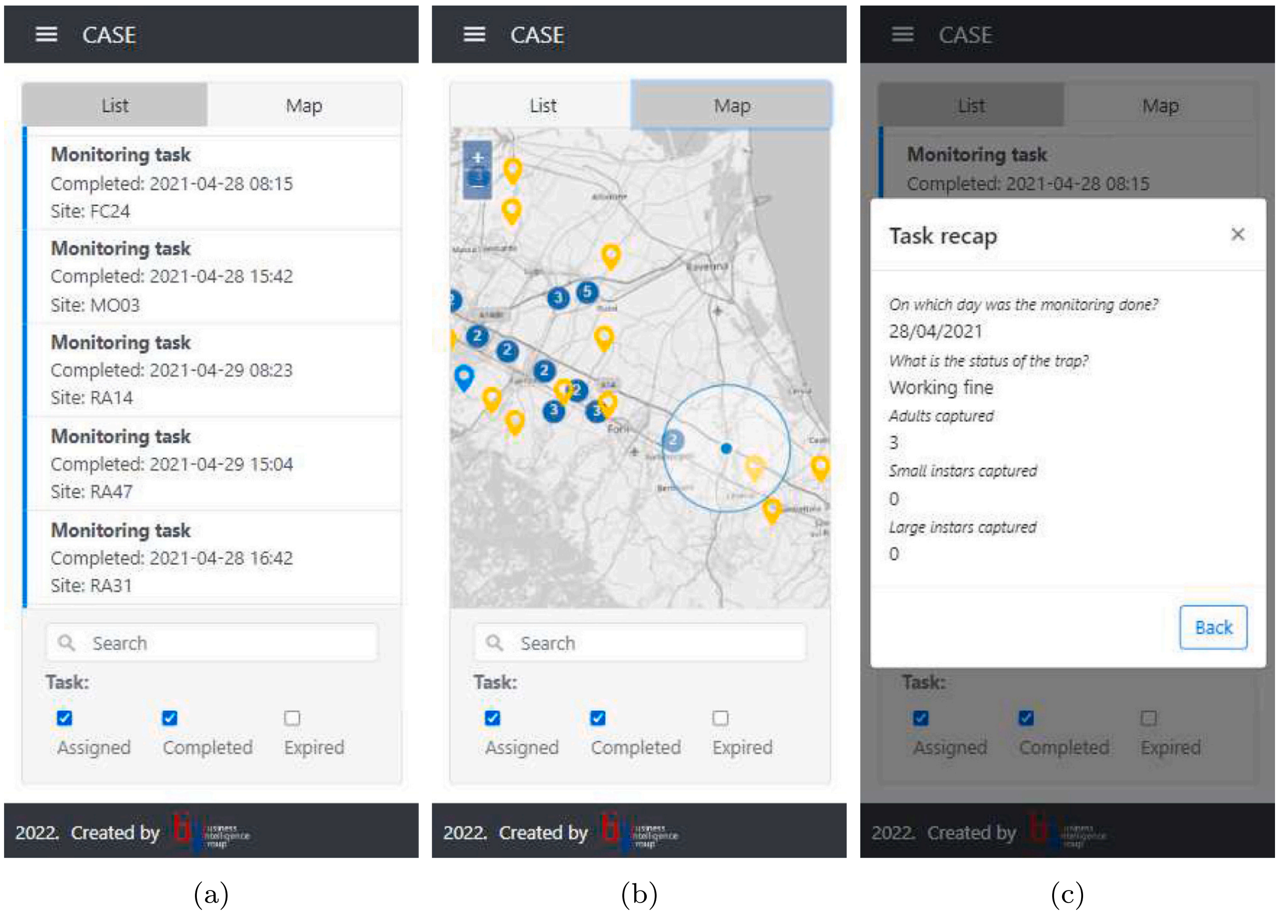


Fig. 4. Snapshots from the CASE application showing the list of monitoring tasks (a), the map of traps with associated monitoring tasks (b), and the summary of data sent in a task (c).

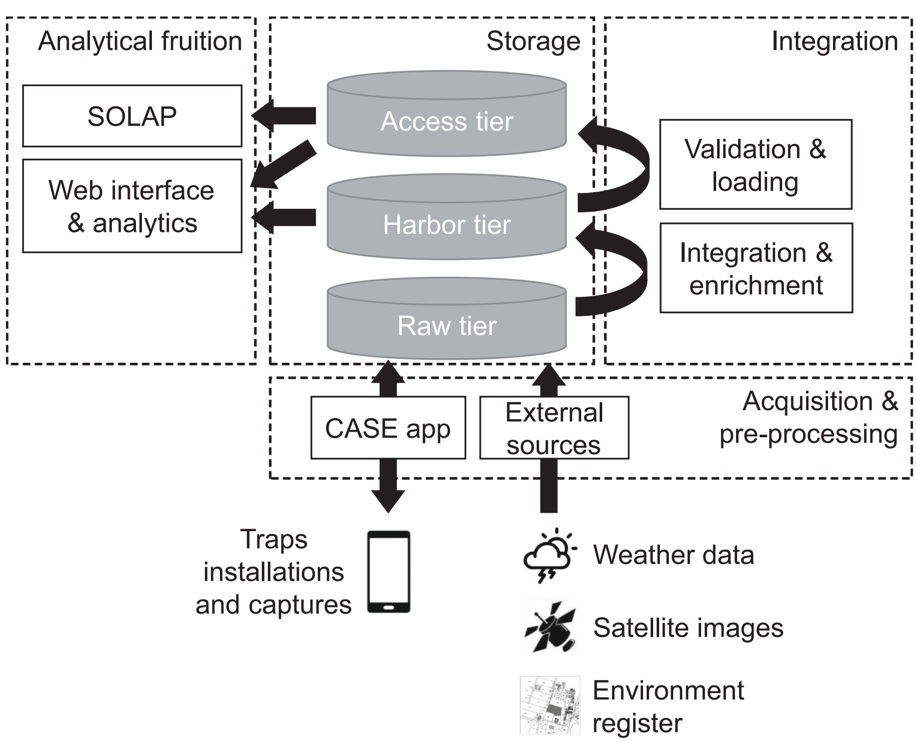


Fig. 5. Overview of the data platform.

Table 1
Summary of data sources.

Source	Provider	Frequency	Granularity	Yearly size
Monitoring network	CASE	Weekly	Monitoring task	5.2 MB
Satellite images	ESA	2–3 days	100 km^2 granule	370 GB
Weather	ARPAE	Daily	25 km^2 cell	6.5 GB
Environment	CER	Yearly	Environmental element	4.3 GB

The validated data is then loaded to the Access tier through an incremental ETL (extract, transform, and load) procedure acquiring the validated data from the Harbor tier.

3.2.4. Analytical fruition

The provisioning of analytical capabilities often builds on multidimensional representations of the data, which offer efficient support to query answering (Golfarelli and Rizzi, 2009). To this end, a multidimensional spatial cube called `Captures` is created in the Access tier. The schema of the cube is shown in Fig. 6 using the conceptual DFM notation (Golfarelli and Rizzi, 2009). Full details about the interpretation of the conceptual model are disclosed in Appendix A.3.

Domain experts have private access to the spatial cube, enabling SOLAP (Spatial OnLine Analytical Processing)[7] (Rivest, 2001) queries. Basic connectivity with common OLAP tools (e.g., Tableau, Power BI) is provided, enabling a self-service OLAP experience. Sample SOLAP queries on the cube using Tableau are shown in Fig. 7. The data from the `Captures` cube have been used also to perform statistical analyses with R software (v. 4.0.3, R Core Team 2020) (R. C. Team, 2013). *H. halys* captures have been analyzed using generalized linear mixed-effects models (glmer) with Poisson distribution. The models tested the effects of several factors surrounding the traps (`SVP` and several `Environmental elements`, such as the number of buildings, presence of tree fruit crops, herbaceous crops, river banks and channels, gardens and groves, hedges and borders) on the cumulated *H. halys* captures per trap over the season, considering these factors individually as predictors together with the `Year` and the trap location (`Province`). As expected, the `Year` and the trap location (`Province`) were found to have a significant effect on the trap captures ($p < 0.001$) and these factors were also set as random effects in the models. Akaike's information criteria (AIC) and residuals were used to select the fitted models. A multiple comparison post hoc test was performed on the fitted models (`glht` function from the `multcomp` package) and Tukey's test ($p < 0.05$) was adopted to discriminate significant differences.

A public website has also been developed and made openly available (link in Section 1). It presents a guided analytics experience, providing real-time open access to the Access and Harbor tiers and enabling farmers and technicians (and any other interested party) to continuously monitor the trend of *H. halys* captures throughout the regional territory and obtain useful insights at multiple analytical levels. Screenshots of the dashboards are shown in Fig. 8. The main component of the dashboard is the map (Fig. 8a), showing the distribution of traps on the territory, where the size of circles indicates the number of traps and the color indicates the number of *H. halys* captures; the top-right selection buttons allow to switch between monitoring seasons and to add layers from external sources (i.e., satellite images, weather data, and environmental elements). Below the map, a timeline allows time-traveling to previous weeks, gauges provide some statistics on the coverage, and a weekly bulletin is provided by domain experts to summarize the latest events in the monitoring network. The dashboards provide interactive reports at different levels of detail, e.g., plotting the trends of captures at the provincial level or showing all the data communicated by an on-field operator through CASE on a given date for a given trap.

[7] The term SOLAP refers to geo-business intelligence technologies allowing online analysis of a massive volume of multidimensional data.

4. Results and discussion

The monitoring website has been advertised among growers and pest control advisors across the region throughout conferences, events in the territory (FreshPlaza.it, 2021; Preti et al., 2021a; Vaccari et al., 2022), and the Emilia-Romagna IPM bulletins (Bollettini interprovinciali di produzione integrata e biologica 2022). Statistics by Google Analytics indicate that, in 2021 and 2022, the website has been visited by more than 2000 distinct users, mostly (79%) from Italy and the Emilia-Romagna region (60% of Italian users), with an average engagement time of 3 min. These statistics are encouraging as they indicate widespread interest in the initiative.

An important result obtained in this project is the publication of weekly bulletins (see Fig. 8a), i.e., short documents provided by domain experts to highlight the most relevant biological events, issued at the beginning of each week and detected during the previous week by the monitoring network (e.g.: first mating, start of ovipositions, emergence of a certain developmental stage, risk of *H. halys* migration in neighboring crops). The bulletins complement the information released through the website to identify the areas with a major presence of *H. halys*. Indeed, knowing when peak populations are expected in each area and which cropping systems are vulnerable to attack can be useful in conducting timely IPM decisions.

4.1. Analysis of H. halys captures

It is known that under the typical temperate Mediterranean climate conditions, *H. halys* develops two generations per year, and that in spring adults leave their overwintering sites, resume feeding, and begin reproduction (Costi et al., 2017). In this section, we present the monitored yearly trends of *H. halys* captures, providing exhaustive details about the generations' life cycles, verifying the consistency with the literature, and evaluating the consistency in the different investigation areas.

Analysis and results. Fig. 9 compares the trends of average weekly *H. halys* captures in the trap network, grouped by three life stages (total adults, both males and females; small juvenile instars, including second and third age nymphs; large juvenile instars, including fourth and fifth age nymphs) over the three years of monitoring. With the exception of 2020, the monitoring seasons began before the average daily temperatures stably exceeded 12 °C, a necessary condition for the noticeable adult emergence beginning (Rot et al., 2022).

In Tables 2 to 4, the CDD reflecting the main biological events highlighted in Fig. 9 are presented (in the tables, the term "comeback" refers to the beginning of a new increase in captures after the capture peak and the subsequent decrease of the captures).

Fig. 10 shows the pairwise comparison of average weekly captures between the three main `Investigation Areas` (West, North-east, and South-east) in the Emilia-Romagna region (see Fig. 2); the linear correlation values are then shown in Table 5.

Discussion. The cycles of *H. halys* in the three life stages are evident in Fig. 9 – except for the first cycles of adults in 2020 due to the delayed deployment of the monitoring trap network. The observed trends are consistent across the investigation areas, as Fig. 10 shows: despite the level of *H. halys* captures being different (with the western area registering more captures than the others), the linear regression indicates the existence of a good correlation (see the correlation values in Table 5) in

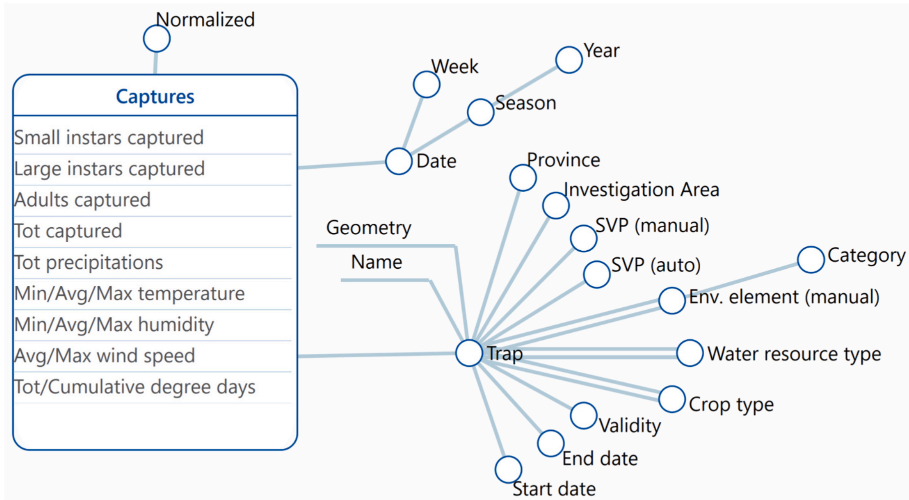
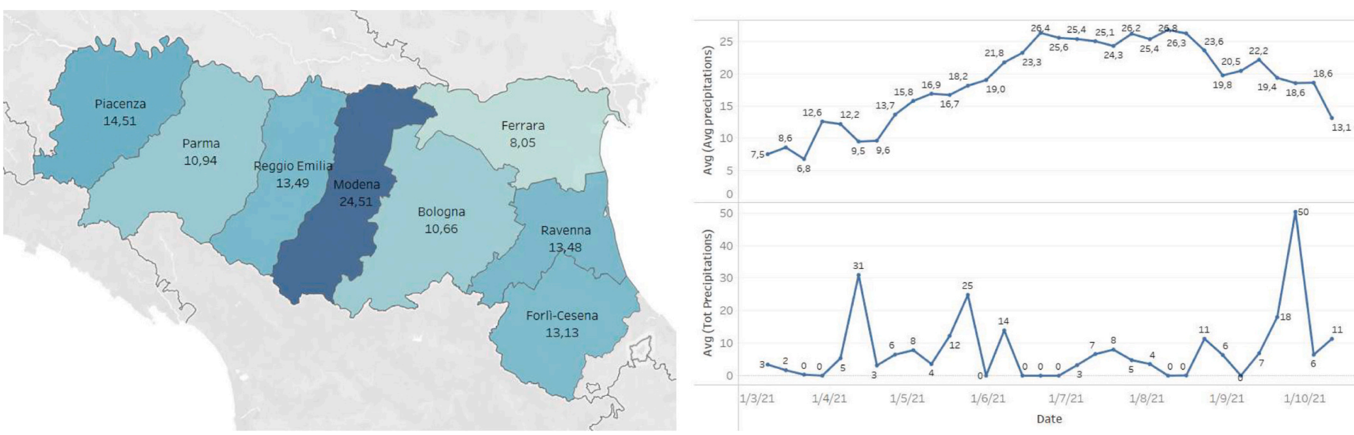


Fig. 6. Multidimensional schema of the `Captures` spatial cube.



(a)　　　　(b)

Fig. 7. Sample results of SOLAP queries, showing the average values of `Tot captured` by `Province` with filter `Year` = 2021 (a), and the average values of `Tot precipitations` and `Tot precipitations` by `Date` with filter `Year` = 2021 and `Province` = "Bologna" (b).

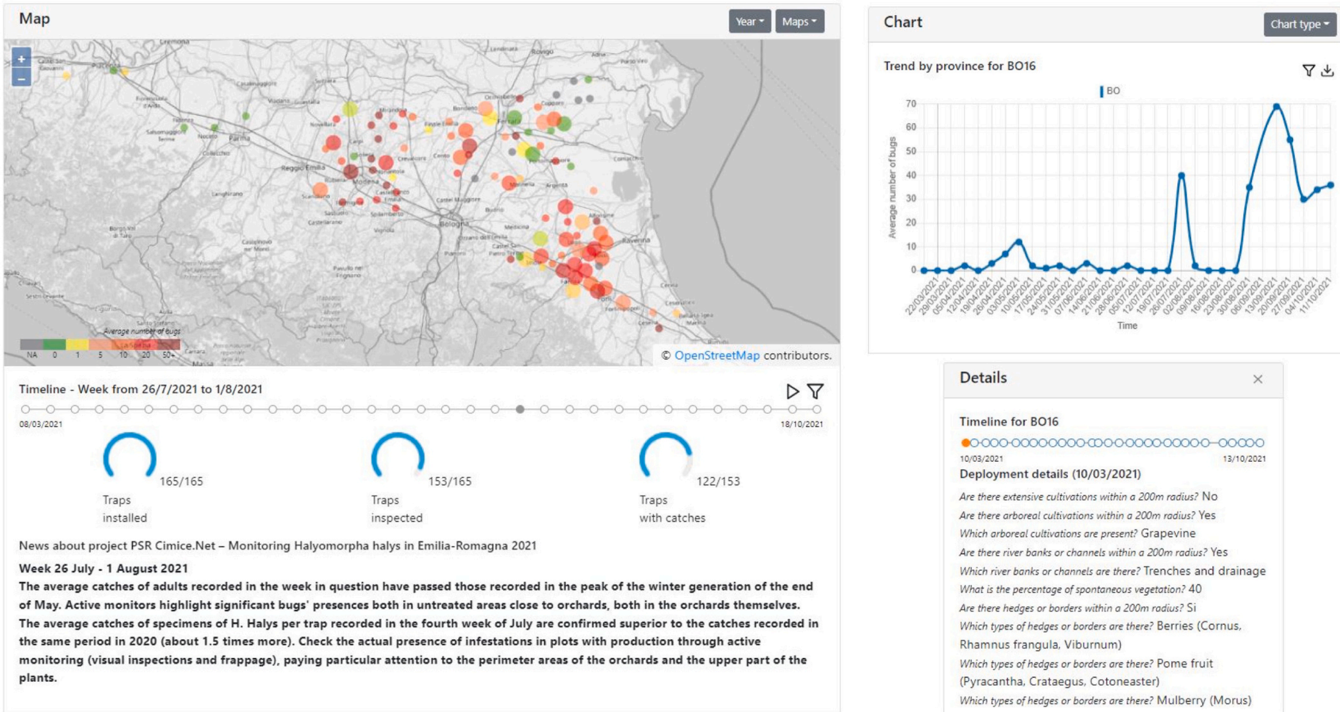
terms of average weekly captures. Therefore, regardless of the size of the *H. halys* population in the different sites, the trend of captures across the season was comparable for the different areas.

The observed trends are also consistent with another study on *H. halys* biology and phenology conducted on the same geographic area, but based on active monitoring techniques (Maistrello et al., 2017). In all monitoring years, adult captures became noticeable from the end of March and had a first peak between the mid and the end of May. From the beginning of July, adult captures started to increase again until the end of July-beginning of August and remained at the same levels until the beginning of September. In September, there was a very strong increase in adult captures, which reached a peak at the beginning of October. The captures of juvenile forms instead showed only two peaks that anticipate the second and third adult peaks by 5–6 weeks in the case of small juvenile instar and by about 3 weeks in the case of large juvenile instar.

In general, the capture peaks of the same *H. halys* life stage can vary between one and four weeks among different years (Tables 2 to 4). Neverthless, considering the CDD these differences among years are lower: the maximum variability can be observed in late summer when the second peak of large nymphs is reached with differences of 214 CDD between 2021 and 2022 (Table 4). Our results collected in Emilia-Romagna are consistent with Rot et al. (Rot et al., 2022), who investigated the biological parameters of *H. halys* in Nova Gorica (western Slovenia) calculating the CDD with the same temperature thresholds. In particular, in the Slovenian study, the first juvenile instar (first age nymphs) is reported to occur at 200 CDD, while the peak of oviposition of the overwintering generation at 200–400 CDD (Rot et al., 2022). We recorded the first capture of small instars (second age nymphs) between 271 and 278 CDD and the peak of the small instars (second and third age nymphs) at 521–541 CDD, totally in line with (Rot et al., 2022). The first juvenile instar (first age nymphs) of the second generation in Slovenia was recorded at ca. 917.8 CDD (between late July and early August), while we observed an increase of the small instars (second and third age nymphs) between 1002 and 1080 CDD. Finally, in Slovenia the second generation of adults occurs just before 1400 CDD; in our Italian study, we recorded the second increase of adult captures between 1375 and 1398 CDD.

4.2. Validation of the automatic SVP quantification

Spontaneous vegetation is usually one of the factors that attracts *H. halys* individuals (Venugopal et al., 2015a; Venugopal et al., 2015b); in this study, its quantification in the surroundings of a monitoring trap (i.e., the spontaneous vegetation percentage within a 200 m radius from the trap) is done both manually (via CASE, by the technician that installs



(a) (b)

Fig. 8. Screenshot of the website's dashboards.

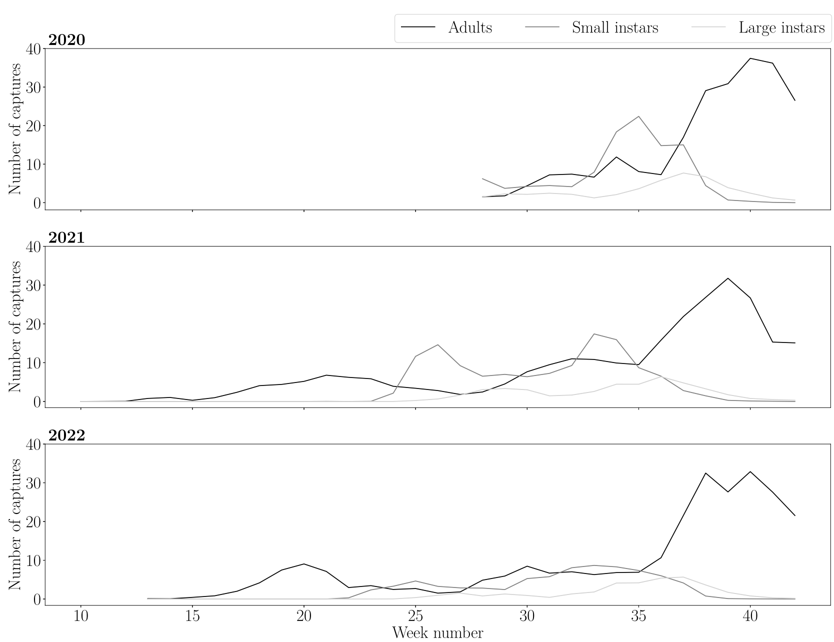


Fig. 9. Trend of average weekly *H. halys* captures by insect stage and monitoring season.

Table 2
Timing and average cumulative degree days (CDD) of the main events in the lifecycle of *H. halys* adult individuals.

Year	1st Peak		1st Comeback		2nd Peak		2nd Comeback		3rd Peak	
	Week	CDD	Week	CDD	Week	CDD	Week	CDD	Week	CDD
2020			Jul 13–19 (W29)	729	Aug 17–23 (W34)	1169	Sep 7–13 (W37)	1398	Sep 28-Oct 4 (W40)	1552
2021	May 24–30 (W21)	163	Jul 12–18 (W28)	722	Aug 9–15 (W32)	1099	Sep 6–12 (W36)	1382	Sep 27-Oct 3 (W39)	1547
2022	May 16–22 (W20)	141	Jul 4–10 (W27)	727	Jul 25–31 (W30)	1025	Aug 22–28 (W34)	1375	Sep 19–25 (W38)	1646

Table 3
Timing and average cumulative degree days (CDD) of the main events in the lifecycle of *H. halys* small instars (second and third age nymphs).

Year	1st Capture		1st Peak		1st Comeback		2nd Peak	
	Week	CDD	Week	CDD	Week	CDD	Week	CDD
2020					Aug 10–16 (W33)	1080	Aug 24–30 (W35)	1256
2021	Jun 7–13 (W23)	278	Jun 28-Jul 4 (W26)	551	Aug 2–8 (W31)	1002	Aug 16–22 (W33)	1199
2022	May 30-Jun 5 (W22)	271	Jun 20–26 (W25)	521	Jul 25–31 (W30)	1025	Aug 15–21 (W33)	1299

Table 4
Timing and average cumulative degree days (CDD) of the main events in the lifecycle of *H. halys* large instars (fourth and fifth age nymphs).

Year	1st Capture		1st Peak		1st Comeback		2nd Peak	
	Week	CDD	Week	CDD	Week	CDD	Week	CDD
2020			Jul 27-Aug 2 (W31)	901	Aug 17–23 (W34)	1169	Sep 7–13 (W37)	1398
2021	Jun 21–27 (W25)	455	Jul 19–25 (W29)	810	Aug 9–15 (W32)	1099	Sep 6–12 (W36)	1382
2022	Jun 13–19 (W24)	424	Jul 4–10 (W27)	727	Aug 8–14 (W32)	1205	Sep 12–18 (W37)	1596

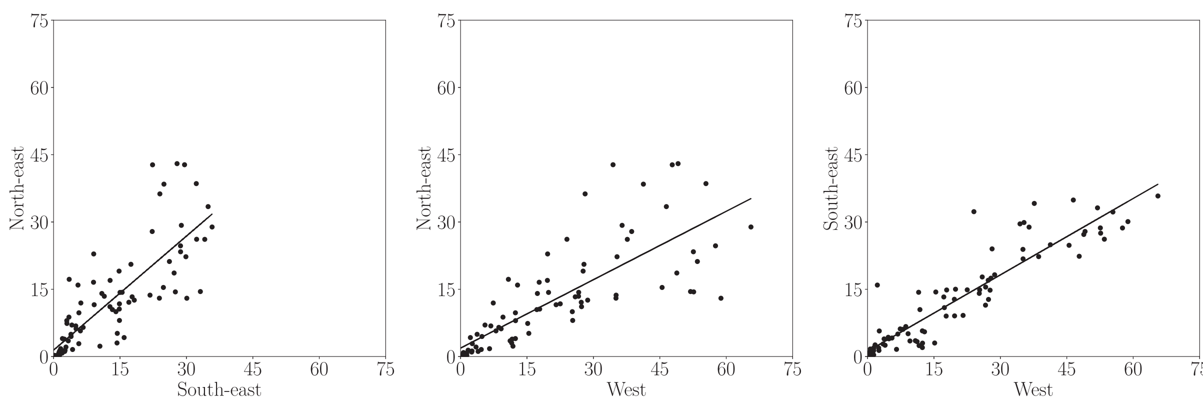


Fig. 10. Pairwise comparison of average weekly *H. halys* captures across the three main `Investigation Areas` monitored in the Emilia-Romagna region.

the trap and gives a visual estimate) and automatically (through the procedure presented in Section 3.2.2 that relies on satellite images and the NDVI index). In this section, we verify the correctness of the automatic quantification to recommend it as a better alternative to the manual one and to validate the subsequent analyses.

Analysis and results. The scatter plot in Fig. 11 indicates, for each monitored trap, the corresponding SVP values quantified employing the manual and automatic methods. The correlation is noticeably weak, with a Pearson correlation coefficient (Benesty et al., 2009) of 0.61.

To assess the accuracy of the two methods, we conducted a manual validation process with the support of a domain expert on a sample of 15 traps, randomly selected with varying agreement levels (5 among those with the highest agreement, 5 among those with the lowest agreement, 5 with an average agreement). By relying on a high-resolution Google Maps satellite image, the domain expert used the QGIS[8] software tool to draw the polygons corresponding to the areas containing spontaneous vegetation. Then, a *ground truth* reference for the actual SVP of a trap has been obtained by dividing the sum of the areas of such polygons by the area corresponding to a 200 m radius. The process is exemplified in Fig. 12.

Discussion. The comparison of the quantified SVP values with the ground truth shows that the automatic SVP is only slightly less precise than the manual one, with an average error of 8% and 5%, respectively. Several factors can negatively impact the automatic quantification. On the one side, the environment registry is not always accurate; for instance, crops may be positioned over areas of spontaneous vegetation

Table 5
Correlation of average weekly *H. halys* captures across the three main `Investigation Areas` monitored in the Emilia-Romagna region.

Investigation Area	West	North-east	South-east
West	1.00	0.80	0.93
North-east	0.80	1.00	0.82
South-east	0.93	0.82	1.00

(causing an underestimation of the SVP by the algorithm) or not reported at all (causing an overestimation, as the algorithm mistakes them for spontaneous vegetation). On the other side, the low granularity of Sentinel-2 satellite images (i.e., 10 m^2 for the RGB bands, and 20 m^2 for the near-infrared band required to compute NDVI) fails to capture areas of spontaneous vegetation that are not significantly wide (e.g., a line of trees on the side of a river or a road).

The slight difference between the two methods suggests that the automatic SVP quantification method can be used as a valid and effective replacement for the manual one. Ultimately, this result (i) demonstrates how the power of the data platform can be harnessed to relieve technicians from a burdensome and easily error-prone manual activity, and (ii) provides the necessary validation to use the automatic calculation for the subsequent analytical activities.

4.3. Factors influencing H. halys presence

The integration and enrichment activities in the data platform enable the association of *H. halys* captures with several attributes (or *features*) that can motivate the presence of the target pest. The nature of these

[8] QGIS: https://www.qgis.org/en/site/

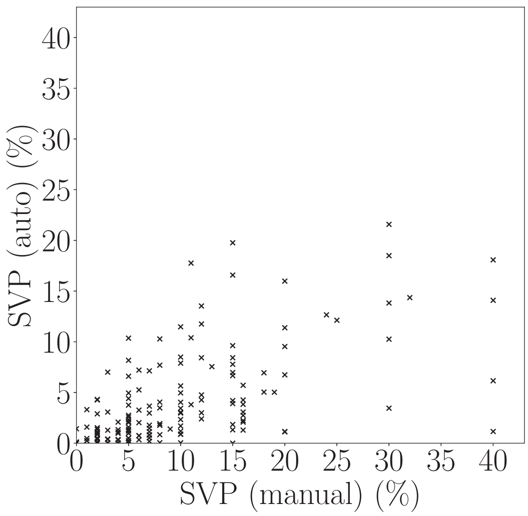


Fig. 11. Scatter plot comparing manual and automatic SVP.

factors is manifold, as some vary on a weekly basis (e.g., the temperature), while others remain the same throughout the season (e.g., the proximity to water resources). Thus, we analyze these features at both weekly and seasonal levels.

4.3.1. Factors influencing H. halys presence on a weekly level

Analysis and results. On a weekly level (i.e., considering the single events in the spatial cube), we calculate the maximal information coefficient (MIC) between all the measures available in the cube.[9] The correlation matrix is calculated with two methods. In the first one, for each trap in the monitoring network we calculate an individual correlation matrix with the trap's weekly data; then, the final matrix is determined with the median values in each cell. In the second method, we directly calculate the final correlation matrix with the average weekly values from all traps. The results are shown in Fig. 13 (first method on the left, second method on the right).

Fig. 14 shows the weekly captures compared with the total precipitation (a) and average wind speed (b) registered on the same week; the red line is the LOESS (locally estimated scatterplot smoothing) (Cleveland, 1979).

Discussion. First of all, we notice that the correlation values in the matrixes of Fig. 13 are consistent across the two methods. Besides trivial correlations (e.g., the ones between temperature measurements), the matrices show a clear correlation between the total (and adult) *H. halys* captures and the cumulative values of degree days (CDD). This is a confirmation of the results from recent research showing that temperature and photoperiod (which can vary with latitude and other geographical parameters) have a great influence on *H. halys* development, survival, voltinism, population density, size, and overwintering behavior (Costi et al., 2017; Haye et al., 2014; Rot et al., 2022). The absence of correlation with non-cumulative values is expected, as captures are primarily driven by the cycles of the *H. halys* individuals (which are well modeled by the cumulative values of the same measurements). Ultimately, the correlation is less evident with small and large juvenile instars as the majority of captures are of adult individuals (the latter composing the 59% of all captures).

The correlation matrices do not reveal a correlation between *H. halys* captures and meteorological factors such as wind speed and precipitation. Nevertheless, these factors have an influence on the target pest captures, because *H. halys* individuals are less likely to move when it is either windy or rainy. This is observable in the scatter plots in Fig. 14. Indeed, low values of `Tot precipitations` and `Average wind speed` are not determinant, but high values visibly account for fewer *H. halys* captures on the traps. This is also confirmed by the LOESS curve, which does not increase significantly when precipitations and wind are absent.

4.3.2. Factors influencing H. halys presence on a seasonal level

Analysis and results. The evaluation on a seasonal level is done by considering the total number of *H. halys* captures cumulated for every trap in a whole season and taking into account both the different trap locations and the three years of survey.

The average yearly captures are analyzed by every categorical feature that remains constant throughout the year; the most interesting comparisons are shown in Fig. 15. In addition, Fig. 16 focuses on the presence of buildings in the surroundings of the traps[10]; the number of buildings surrounding the traps was grouped into four levels: sites without buildings (0), sites with one or two buildings (1–2), sites with three or four buildings (3–4), and sites with five or more buildings (≥ 5).

Discussion. Interestingly, the features showing a distinctive difference in the number of trap captures are those related to the environmental elements in the surroundings of the traps (Fig. 15). In particular, we notice significantly higher cumulated *H. halys* captures with higher SVP ($\chi^2 = 2057.9$; $p < 0.001$) and in presence of gardens and groves ($\chi^2 = 172.9$; $p < 0.001$), hedges and borders ($\chi^2 = 211.5$; $p < 0.001$), and river banks and channels ($\chi^2 = 1004.1$; $p < 0.001$). The effect of these factors was expected since the above-mentioned environmental elements are known to be attractive for *H. halys* considering that wild host plants can grow among the spontaneous vegetation, such as in garden and groves, hedges and borders, and along the water sources. The complexity of the agroecosystem with one or more of such environmental elements can therefore serve as a refuge for this pest species, providing a hospitable habitat in which to feed, develop, and lay eggs.

On the contrary, the presence of herbaceous crops in the surrounding of the traps had no significant effect on the trap captures ($\chi^2 = 1.44$; $p = 0.23$), with a comparable number of cumulated *H. halys* captures (number of traps; mean value $\pm$ SE) in sites where herbaceous crops occurred close to the traps ($N = 256$; 410 ± 20) and in sites where there were no herbaceous crops in the surrounding of the traps ($N = 92$; 380 ± 30). This result can be easily understood because not all the herbaceous crops present in this study are attractive to *H. halys* and the ones that can be considered host plants are infested only in specific phenological stages, for instance when bearing fruits or seeds. Therefore, the restricted time period of pest occurrence in a given attractive herbaceous crop can not be appreciated taking into consideration the totality of the herbaceous crops and the entire vegetative season.

Further analyses should consider defined host crops (e.g., soybean) and specific phenological periods (e.g., after seed formation) to better evaluate the effect of the herbaceous crops on the *H. halys* trap captures. Such evaluations were not carried out in this study considering the limited number of sites ($N = 10$) with comparable conditions. Regarding the effect of tree fruit crops on the trap captures, no analysis was carried out since only a few sites ($N = 5$) had no orchards in the surrounding of the monitoring traps; in fact, the whole trap network was set up taking into consideration the tree fruit production of Emilia-Romagna region and the majority of the sites ($N = 343$) intentionally had tree fruit crops in the area around the monitoring trap.

As to the presence of buildings (Fig. 16), there are significant differences in terms of captures among the different levels of buildings ($\chi^2 = 471.2$; $p < 0.001$) when looking at the cumulated *H. halys*

[9] MIC (Reshef et al., 2011) is a statistical algorithm to determine both linear and non-linear correlation between two variables, returning values between 0 (no correlation) and 1 (maximum correlation).

[10] Agricultural buildings and the farmhouses in the surroundings of the traps were summarized together since *H. halys* can overwinter in both types of buildings.

Fig. 12. Left (a): a Google Maps image centered on a trap (red circle in the middle), with its 200 m radius; polygons with white borders are crops mapped in the CER dataset; polygons in green are indications of spontaneous vegetation by the domain expert. Middle (b): the same setup, but with a Sentinel-2 image in the background. Right (c): the visible 10 m^2 squares contain an NDVI value ≥ 0.7 (the brighter the shade of green, the higher the NDVI value). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

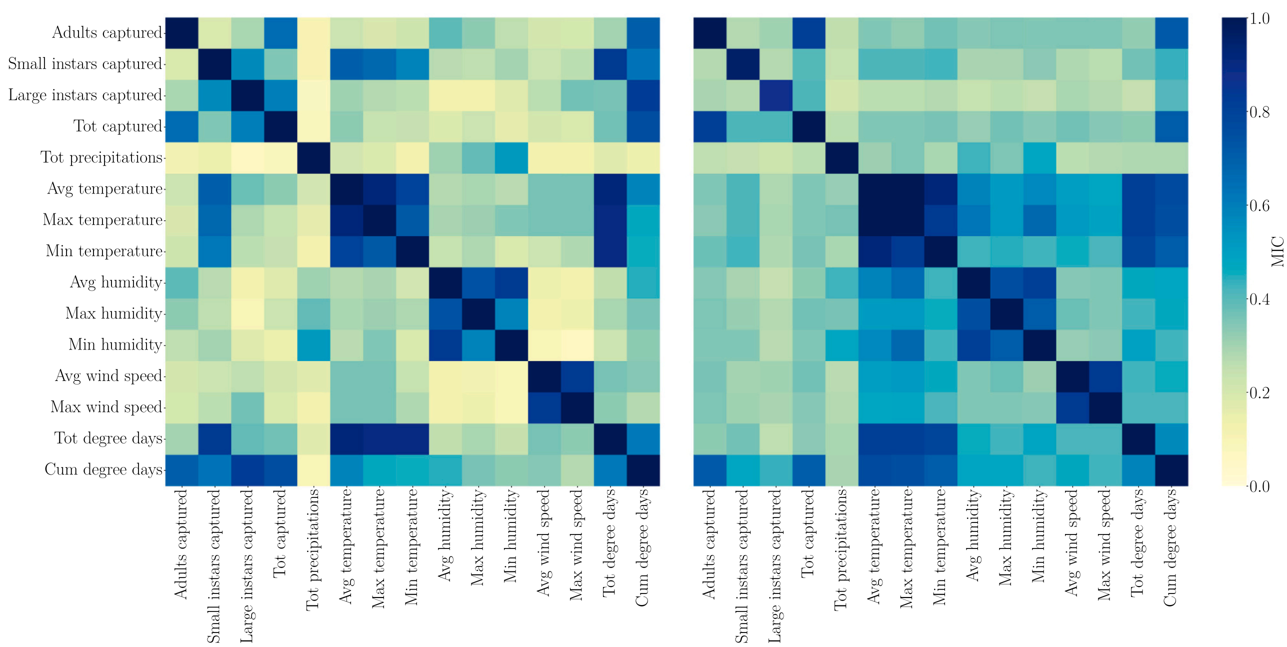
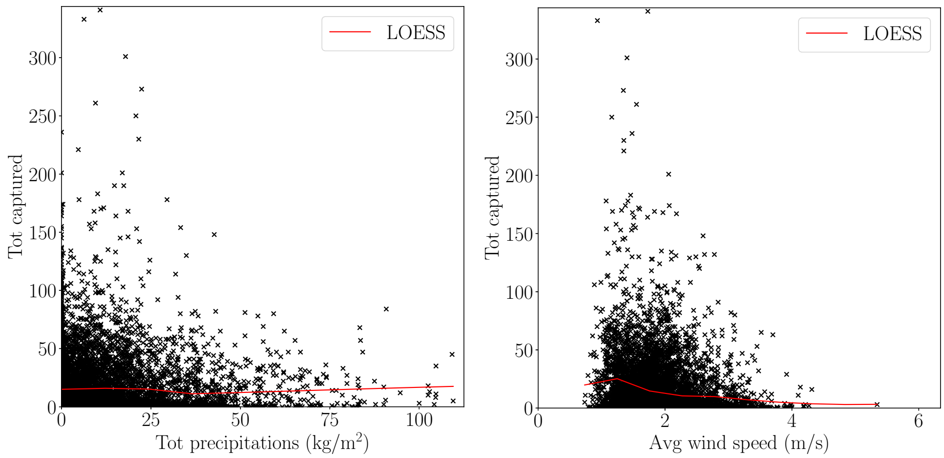


Fig. 13. Correlation matrices with MIC on a weekly level, obtained by taking median values from the matrices calculated for every single trap (left) or by simply considering average weekly values (right).

captures over the entire season (Fig. 16a). However, the trend of these captures is not clear since various factors not included in the analysis (such as the attractiveness of the surrounding host species in the different phenological stages and the agroecosystem complexity) may affect the number of captures during the season much more than the presence of buildings. Considering the cumulated *H. halys* captures only in the spring period (i.e., from the beginning of the monitoring until May 31 of each year), there is a significant effect of the presence of buildings ($\chi^2 = 102.4$; $p < 0.001$) and, as typically observed by the growers and pest control advisors, the *H. halys* captures are higher in sites with buildings compared to those without. This trend is very well explained by the behavior of the overwintering *H. halys* adults that during autumn aggregate in protected and dry shelters such as the buildings; however, in spring no difference in terms of captures was observed among sites with different numbers of buildings (Fig. 16b). Finally, taking into consideration only the cumulated *H. halys* during the autumn period (i.e., from September 1 until the end of the monitoring session of each year), the effect of buildings on the captures is also significant ($\chi^2 = 274.8$; $p < 0.001$), but again the trend of captures among different number of buildings is not clear. In this period, the summer adults look for shelters in which to overwinter, but the building factor alone cannot explain the *H. halys* captures in autumn because other factors not included in the analysis (e.g. crops not yet harvested in September and the temperature trend during September–October) affect the *H. halys* dispersal before overwintering.

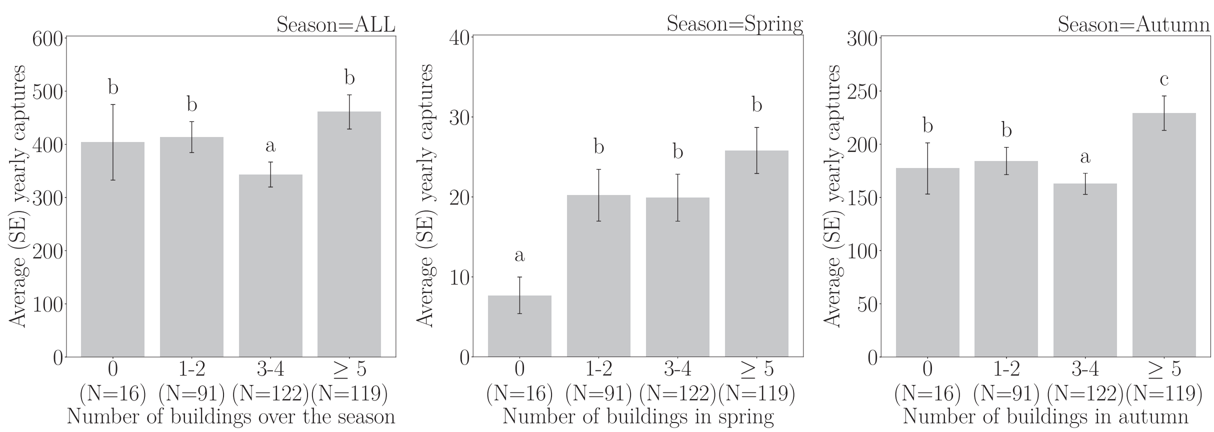
4.4. Degree-day phenological model

From the discussion in Section 4.3 it emerged that CDD are highly correlated with weekly *H. halys* captures. In this section, we demonstrate



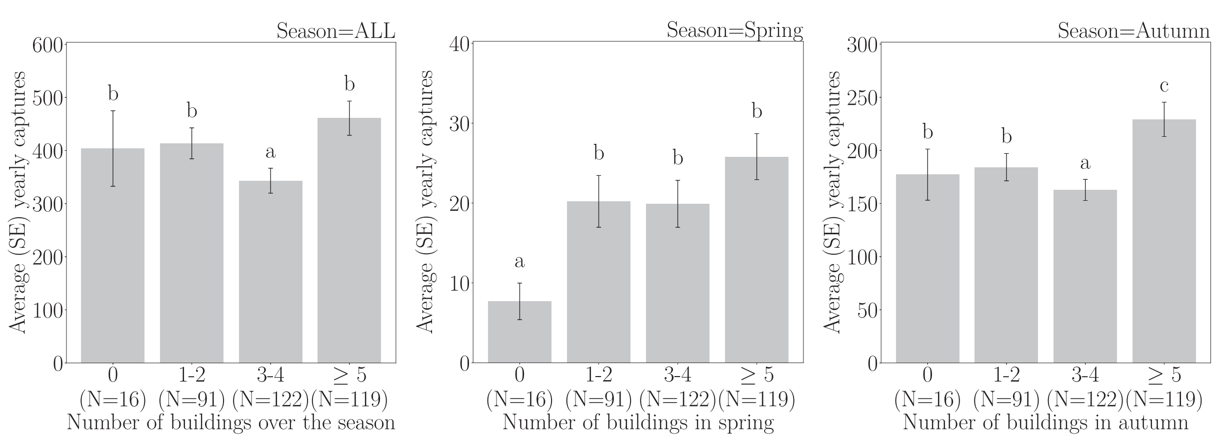
(a) (b)

Fig. 14. Scatter plot of `Tot captured` against either `Tot precipitations` (a) or `Avg wind speed` (b).



(a) (b) (c)

Fig. 15. Average yearly number of *H. halys* captures by range of `SVP (manual)` (a) and by closeness to at least one environmental element of `Category` "Gardens and groves" (b), "Hedges and borders" (c), and "River lands and channels" (d). Confidence intervals represent the standard error (SE).



(a) (b) (c)

Fig. 16. Average yearly number of *H. halys* captures by the number of buildings in the close surroundings of the traps, considering captures over the whole season (a), only in spring (b), or in autumn (c). Confidence intervals represent the standard error (SE).

that the quality and quantity of the data collected in our study enables the study and creation of a phenological model for *H. halys*.

Analysis and results. In the literature, phenological models based on CDD have been defined to describe and predict the spread of insect pests, including *H. halys* (Kamiyama et al., 2021; Nielsen et al., 2008; Nielsen et al., 2016). Existing models are not directly applicable to our study, as they are either based on features that we do not monitor ((Nielsen et al., 2008) studies the mortality of *H. halys* individuals in a closed and controlled environment) or they do not provide sufficient elements to ensure their reproducibility (Kamiyama et al., 2021; Nielsen et al., 2016).

For a demonstrative purpose, we adopt here the phenological model presented in (Damos et al., 2018), which is based on *Cydia pomonella*, the key Lepidopteran pest of pome fruit crops. (Damos et al., 2018) presents a three-parameter non-linear regression model:

$$F(x,\alpha,\beta,\gamma) = \frac{\alpha}{1+e^{-\left(\frac{x-\gamma}{\beta}\right)}} + \epsilon_i$$

In the equation above, F is the cumulative percentage of *H. halys* captures, x is the value of CDD, and ϵ_i is the standard error term that is assumed to have a normal distribution and zero variance. The behavior of this curve is affected by three key constant variables: α, β, and γ. The model is quite flexible since the curve can be twisted around to fit most conceivable variations of its basic shape depending on the parameter values (Damos et al., 2018); α and β designate the upper and lower asymptotes, respectively, and set the vertical limits of the curve. The parameter γ is the gradient that sets the length of time of the curve, which represents the CDD of 50% *H. halys* captures. Separate models are created for each cycle of each specimen. Values for α, β, and γ have been calculated from the trends in Section 4.1 by manually isolating the cycles of each specimen. For simplicity, the cycles of each specimen have been assumed as non-overlapping, even though they actually are (Costi et al., 2017). The obtained models are shown in Fig. 17, together with the data collected in 2020 and 2021.

Discussion. The model shows high accuracy, with an RMSE of 8.41, confirming the goodness of CDD as a measure highly correlated with *H. halys* captures. The curve of the third cycle for *H. halys* adults shows a different behavior, due to monitoring season being closed before all adults have found shelter for the winter. While the model hereby proposed is based on a different insect species and presents many simplifications, the results demonstrate the goodness of the data collected in this study and encouragingly call for further developments in this direction.

5. Conclusions

In this paper, we presented a data-driven approach to monitor and analyze the occurrence, distribution, and spread of *H. halys* in the Emilia-Romagna region, Italy. The field monitoring was carried out for three consecutive years and relied on a wide network of traps across the whole region, supported by a data platform for the collection and integration of data on *H. halys* trap captures with several external data sources, including weather data and information concerning environmental elements close to the traps. The data platform enabled analytical activities that led to a deeper understanding of *H. halys* population dynamic, considering both its occurrence, distribution, and phenology in the monitored territory. Most importantly, the study revealed how weekly *H. halys* captures are highly correlated with degree days and partially influenced by atmospheric events like rain and wind, whereas yearly captures are mainly driven by the amount of spontaneous vegetation in the close surroundings, such as by gardens and groves, hedges and borders, and the presence of water sources. Some factors occurring in the agroecosystem affect the captures only in specific periods of the season; for instance, in early spring, the presence of buildings acting as shelter for the overwintered population of *H. halys* adults significantly impacted on the trap captures compared to sites without buildings. Finally, other factors are very difficult to control and analyze, such as the

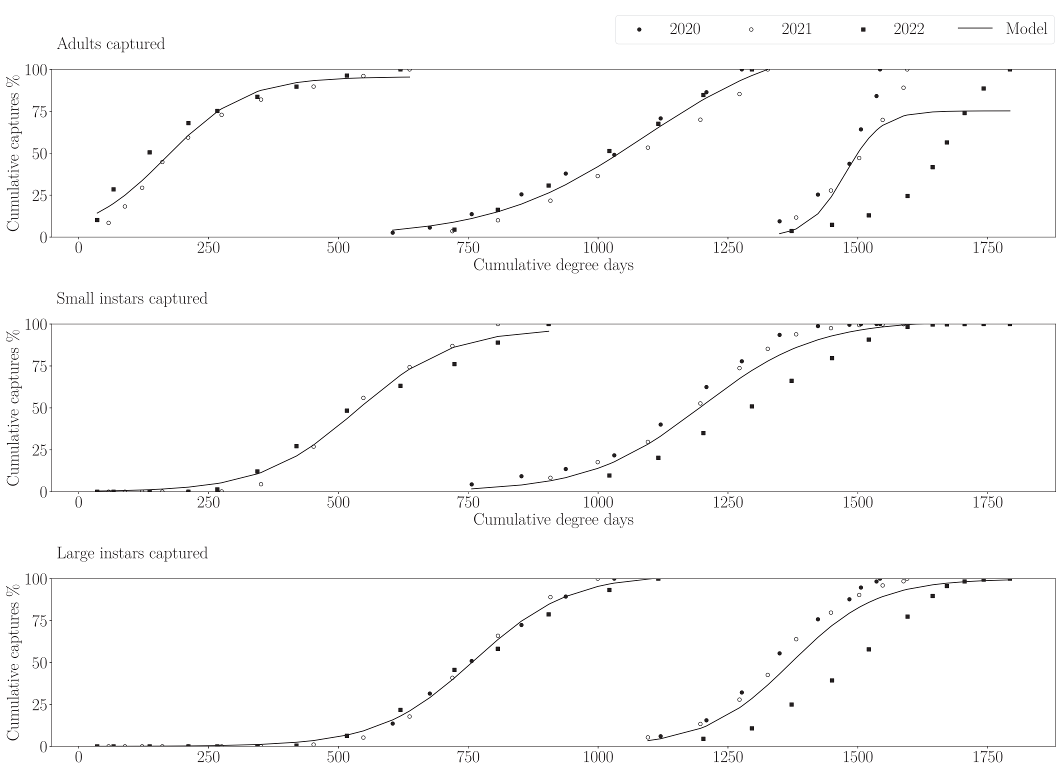


Fig. 17. Phenological models by specimen and cycle.

different attractiveness of the host plants over the season, according to the plant phenology stages and the ripening calendar of the different crops, cultivars, and wild vegetation, considering their occurrence and mutual interaction within each area infested by *H. halys*.

The knowledge on *H. halys* occurrence, distribution, and phenology is crucial to effectively apply the IPM strategies; in particular, highlighting the periods of the pest's high abundance can help focus and rationalize the insecticide-based control (Short et al., 2016). Our study reached these needs, providing the stakeholders (both growers and pest control advisors) with several useful information to set the alerts, to choose the appropriate control methods, and therefore to optimize the insecticide usage in the management of *H. halys* during the project. The weekly bulletins provided with this study, together with the level of *H. halys* captures in the monitored areas, were available for the stakeholders to better evaluate their control programs against this pest, acting as a supplementary decision support system. In fact, the information provided by the punctual monitoring at the orchard and farm scales, combined with the information derived from this area-wide monitoring at the territory scale, is crucial to implement effective IPM strategies, which differ locally case by case requiring as much information as possible to be reasonable. In addition, the CDD calculated on the field data collected during the period 2020–2022 could be exploited to set intervention thresholds based on the different biological events of *H. halys* during the vegetative season and according to the weather trend. Ultimately, the data collected in this study can be further exploited to develop, calibrate, and validate a phenological model of the *H. halys* development, in order to facilitate growers in predicting the lifecycle parameters of this pest and adopting a more efficient and sustainable usage of counteractive measures.

The results presented in this paper open several opportunities for further development. First of all, smart traps (similar to those used in (Gallinucci et al., 2020)) can be used to automatically count the trapped individuals and send the data to the data platform; this solution would alleviate the huge effort of manually checking all the traps on a weekly basis (Preti et al., 2021b). A further step is a novel automated monitoring system that combines drone imagery with artificial intelligence-based insect detection. This was first demonstrated by (Giannetti et al., 2024), who were able to develop a system capable of detecting and quantifying the presence of *H. halys* using high-altitude, high-resolution imagery; this makes the method potentially applicable to a range of crops and pests. Another important improvement that this study calls for, concerns the quality of the data in the regional environmental registry, which currently limits the effectiveness of automatic SVP detection techniques. This aspect is crucial to combine the field scouting activities with remote sensing data and implement data platforms exploitable for integrated analyses combining different information.

In conclusion, the analytical results of this project demonstrate the importance of adopting a data-driven approach and call for IPM strategies in agriculture, specifically for insect pests such as *H. halys* that require an area-wide monitoring and management approach.

CRediT authorship contribution statement

Chiara Forresi: Writing – review & editing, Writing – original draft, Software, Methodology, Formal analysis, Data curation. **Enrico Gallinucci:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Conceptualization. **Matteo Golfarelli:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Lara Maistrello:** Writing – review & editing. **Michele Preti:** Writing – review & editing, Writing – original draft, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Giacomo Vaccari:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Links to download data and code can be found within the paper

Acknowledgements

This work was supported by the regional project PSR 2014-2020 Op. 16.1.01 - Go Pei-Agri Focus Area 4B Pr. "Cimice.Net" during the field seasons 2020 and 2021. The project prosecution during the field season 2022 was supported by the Emilia-Romagna Producers Organizations, thanks to the coordination of Ri.Nova.

Appendix A. The data platform's multi-layer storage

Storage in the data platform is organized into three tiers, so as to logically separate the subsequent processing activities. In this section, we provide fine-grained details about the data stored in the different tiers.

A.1. The Raw tier

The Raw tier hosts the *data lake* (Stein and Morrison, 2014), i.e., a storage repository that holds a vast amount of raw data in its native format, including structured, semi-structured, and unstructured data. The data lake is used to store all the data coming from external sources in their raw format and to host the enrichment activities required before their integration. The Raw tier is composed of the Hadoop Distributed File System (HDFS) (used to store the raw and enriched files), the distributed database Apache HBase (used to store the data extracted from weather files), and the PostgreSQL RDBMS (used to store data collected from CASE).

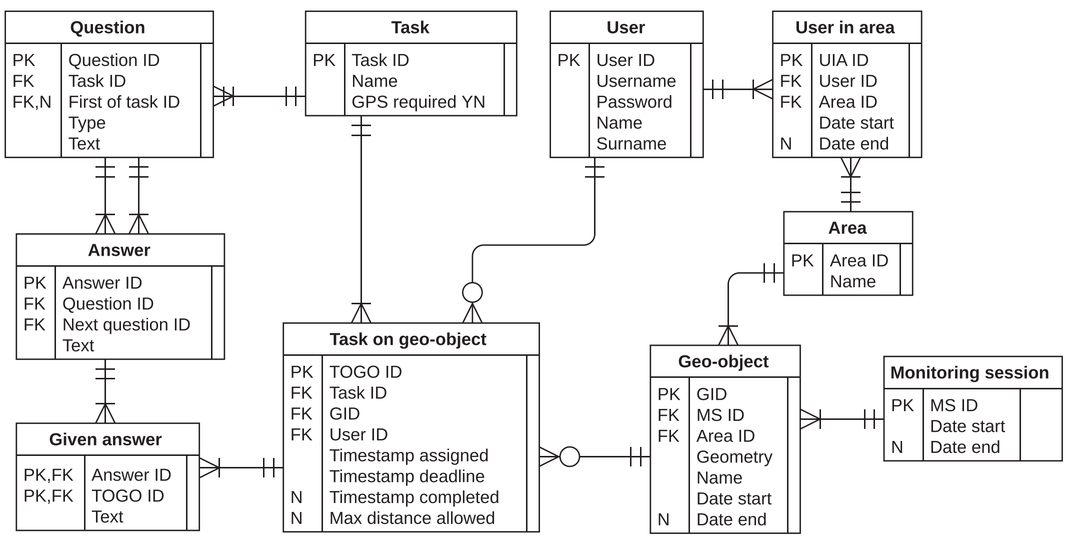


Fig. A.18. Relational schema of the data managed by CASE.

The relational schema of CASE's database is shown in Fig. A.18. CASE is based on a concept of `Task`, i.e., a questionnaire that must be fulfilled by a `User` (i.e., an on-field operator). Each task is composed by a set of `Questions`, categorized in different types (e.g., multiple answers, date picker) each composed by a set of predefined `Answers` (one of which could be an open answer, defined by the user when selecting the `Given answer`). The dynamic aspect of CASE's questionnaires lies in the association of each possible answer with a different one to be shown next. Thus, users fulfilling the same task may follow different paths of questions depending on the given answers. This enables a customized and efficient user experience, as detailed questions on a given topic may be asked only if such a topic has been mentioned by the user. Instances of tasks (i.e., `Task on geo-object`) are associated with `Geo-objects`, i.e., georeferenced elements (e.g., fields, plants, traps) that are the subject of the questionnaire. Thus, the fulfillment of a task is limited to users who are located within a given maximum distance. Users have visibility on the geo-objects in a given `Area` (e.g., a farm, a province), and geo-objects are associated with a period of activity named `Monitoring session`.

A.2. The Harbor tier

The Harbor tier provides an integrated and comprehensive view of the available data at the finest level of detail. Data integration takes place at this level and is mainly based on the spatial features of data: the relationships between heterogeneous data are found thanks to geopositioning even in the absence of direct references. The Harbor tier is implemented on the same PostgreSQL instance of the Raw tier, which enables spatial features through its spatial extension called PostGIS.

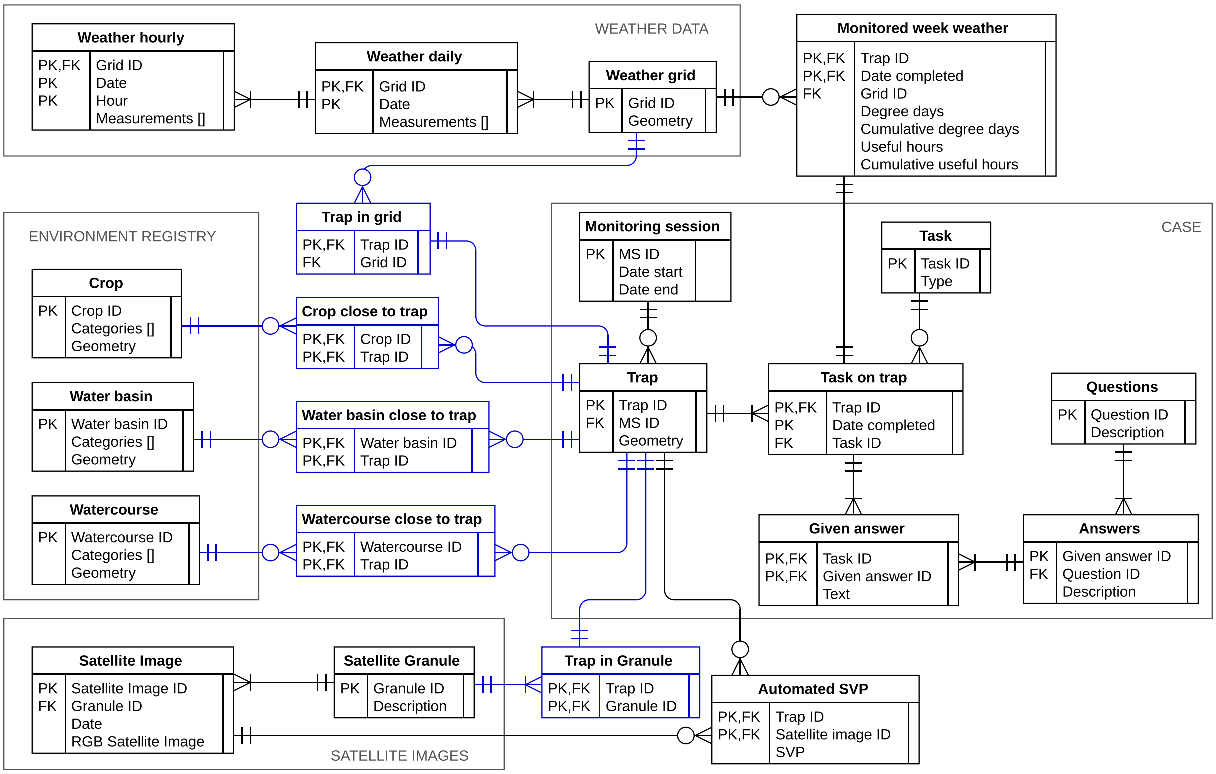


Fig. A.19. Relational schema of the integrated data in the Harbor tier.

The relational schema of the integrated data is shown in Fig. A.19. The main role is played by the *H. halys* traps, whose spatial coordinates enable the integration with the different data sources. As CASE is a multi-purpose application, only the data relevant to *H. halys* captures is extracted; also, with a slight abuse of notation, the `Measurements []` and `Categories []` columns in the weather and environment registry datasets are placeholders for the respective lists of measurements and categories, that are here omitted for better readability. Tables highlighted in blue materialize the spatial proximity joins between traps and static references in the other datasets (e.g., a trap is always localized within the same weather grid). Tables `Automated SVP` and `Monitored week weather` materialize measures computed as indicated above.

A.3. The Access tier

The Access tier provides a higher-level view of data that is ready to be consumed for analytical purposes. It mainly relies on a *spatial cube* to enable SOLAP activities. Multidimensional data are organized in cubes describing domain events called *facts*, which are characterized by numerical indicators called *measures* (e.g., the number of *H. halys* captures) and *dimensions* to be used for analyses (e.g., time, traps). Each dimension is described by a *hierarchy* of concepts that describes the dimension of analysis at different granularity levels (e.g., a trap is placed in a province, which in turn belongs to a macro-area). Data can be analyzed using SOLAP operators such as *spatial slice* and *spatial drill*, which allow for aggregating measure values along hierarchies with SQL operators. The Access tier is implemented on a different spatially-enabled instance of PostgreSQL.

The schema is obtained with a data-driven approach, i.e., by choosing the weekly *H. halys* captures as the fact of interest, then following the functional dependencies coded within the integrated schema, and finally validating the result during a meeting with domain experts.

The `Captures` cube (shown in Fig. 6) features three dimensions: `Date`, `Trap`, and `Normalized`. The latter features a single attribute with a boolean value (either "yes" or "no") depending on whether the indicated *H. halys* captures are the result of the normalization process. The `Date` dimension develops into a temporal hierarchy and contains all dates within each monitoring season; each `Year` is decomposed into three `Seasons` ("Spring" from the beginning of the monitoring to the 31st of May;"Summer" from the 31st of June to the 31st of August;"Autumn" from the 31st of September to the end of the monitoring). The `Trap` dimension develops into a hierarchy with all the information associated with each trap, either collected manually at installation time or obtained automatically in the integration phase. Most importantly:

- `SVP` refers to the spontaneous vegetation percentage; `SVP (manual)` is the one collected via CASE, while `SVP (auto)` is the average of the values calculated via satellite images throughout the season. Both `SVP` values are discretized into bins of 15% width.
- `Validity` indicates the trustworthiness of the trap; its value is either "valid", "noisy", or "invalid".
- `Crop type` and `Water resource type` are the elements within the close surroundings, obtained from the environmental registries.

- `Environmental element (manual)` are the elements indicated by the technician at installation time (e.g., fruit trees, crops, buildings), which can be aggregated by `Category` (i.e., “Agricultural buildings”, “Tree fruit crops”, “Herbaceous crops”, “Hedges and borders”, “Residential buildings”, “Gardens and groves”, and “River banks and channels”).
- `Investigation Area` is a grouping of traps based on their geographical concentration (see Fig. 2). The main values are “West” (including traps in the provinces of Modena and Reggio Emilia), “North-east” (including traps in the province of Ferrara and the north-eastern part of Bologna), and “South-east” (including traps in the city of Imola and the provinces of Ravenna and Forlì-Cesena); traps too distant from these groupings (i.e., those in the provinces of Piacenza and Parma, and those close to the Tuscany region) are classified as “Peripheral”.

The events in this cube are the weekly captures of each trap, enriched with further data and statistics obtained from the external sources. In particular:

- `Small instars captured`, `Large instars captured`, and `Adults captured` are the normalized values, computed as discussed above. `Tot captured` is the sum of the three measures.
- `Avg/Max/Min temperature/humidity/wind` speed and `Tot precipitations` are the weather information obtained from ARPAE.
- `Tot/Cumulative degree days` are the additional measures computed as discussed in Section 3.2.2.

References

Acebes-Doria, A.L., Morrison III, W.R., Short, B.D., Rice, K.B., Bush, H.G., Kuhar, T.P., Duthie, C., Leskey, T.C., 2018. Monitoring and biosurveillance tools for the brown marmorated stink bug, *Halyomorpha halys* (Stål)(Hemiptera: Pentatomidae). Insects 9 (3), 82.

Agenzia Veneta per l’’innovazione nel Settore primario, 2024. Bollettino cimice asiatica. https://www.venetoagricoltura.org/argomento/bollettino-cimice-asiatica/, [Online; accessed 14-May-2024].

Almstedt, L., Baltieri, D., Sorbelli, F.B., Cattozzi, D., Giannetti, D., Kargar, A., Maistrello, L., Navarra, A., Niederprüm, D., O’Flynn, B., Palazzetti, L., Patelli, N., Piccinini, L., Pinotti, C.M., Wolf, L., Zorbas, D., 2023. Technological innovations in agriculture for scouting *Halyomorpha halys* in orchards. In: In: 2023 19th International Conference on Distributed Computing in Smart Systems and the Internet of Things (DCOSS-IoT). IEEE. https://doi.org/10.1109/dcoss-iot58021.2023.00110.

Bandh, S.A., 2021. Sustainable agriculture: Technical progressions and transitions. https://doi.org/10.1007/978-3-030-83066-3.

Bell, W.J., Cardé, R.T., 2013. Chemical Ecology of Insects. Springer.

Benesty, J., Chen, J., Huang, Y., Cohen, I., 2009. Pearson correlation coefficient. In: Noise Reduction in Speech Processing. Springer, pp. 1–4.

Bergh, J.C., Morrison III, W.R., Stallrich, J.W., Short, B.D., Cullum, J.P., Leskey, T.C., 2021. Border habitat effects on captures of *Halyomorpha halys* (Hemiptera: Pentatomidae) in pheromone traps and fruit injury at harvest in apple and peach orchards in the mid-Atlantic, USA. Insects 12 (5), 419.

Blackburn, T.M., Bellard, C., Ricciardi, A., 2019. Alien versus native species as drivers of recent extinctions. Front. Ecol. Environ. 17 (4), 203–207. https://doi.org/10.1002/fee.2020.

Bollettini interprovinciali di produzione integrata e biologica 2022 (2022). URL https://agricoltura.regione.emilia-romagna.it/fitosanitario/temi/difesa-sostenibile-delle-produzioni/bollettini/interprovinciali-produzione-integrata-biologica-2022.

Bono Rosselló, N., Rossini, L., Speranza, S., Garone, E., 2023. Towards pest outbreak predictions: are models supported by field monitoring the new hope? Eco. Inform. 78, 102310 https://doi.org/10.1016/j.ecoinf.2023.102310.

Bradshaw, C.J.A., Leroy, B., Bellard, C., Roiz, D., Albert, C., Fournier, A., Barbet-Massin, M., Salles, J.-M., Simard, F., Courchamp, F., 2016. Massive yet grossly underestimated global costs of invasive insects. Nat. Commun. 7 https://doi.org/10.1038/ncomms12986.

Chen, N., Zhang, X., Wang, C., 2015. Integrated open geospatial web service enabled cyber-physical information infrastructure for precision agriculture monitoring. Comput. Electron. Agric. 111, 78–91.

Cimice asiatica, 2022. i danni del 2019 Superano I 740 Milioni. URL https://www.italiafruit.net/DettaglioNews/53075/in-diretta-da/cimice-asiatica-i-danni-del-2019-superano-i-740-milioni.

Clarke, D.A., McGeoch, M.A., 2023. Invasive alien insects represent a clear but variable threat to biodiversity. Curr. Res. Insect Sci. 4 https://doi.org/10.1016/j.cris.2023.100065.

Cleveland, W.S., 1979. Robust locally weighted regression and smoothing scatterplots. J. Am. Stat. Assoc. 74 (368), 829–836.

Costi, E., Haye, T., Maistrello, L., 2017. Biological parameters of the invasive brown marmorated stink bug, *Halyomorpha halys*, in southern Europe. J. Pest. Sci. 90 (4), 1059–1067.

Damos, P.T., Kouloussis, N.A., Koveos, D.S., 2018. A degree-day phenological model for *Cydia pomonella* and its validation in a mediterranean climate. Bull. Insectol. 71, 131–142.

Dash, J., Curran, P.J., 2004. The meris terrestrial chlorophyll index. Int. J. Remote Sens. 25 (23), 5403–5413. https://doi.org/10.1080/0143116042000274015.

Deering, D.W., 1978. Rangeland Reflectance Characteristics Measured by Aircraft and Spacecraft Sensors. Ph.D. thesis. Texas A&M Univ, College Station.

Dent, D., Binks, R.H., 2020. Insect Pest Management. Cabi.

Elliott, N., Farrell, J., Gutierrez, A., van Lenteren, J.C., Walton, M., Wratten, S., 1995. Integrated Pest Management. Springer Science & Business Media.

European Parliament and Council, 2014. Consolidated text: Regulation (eu) no 1143/2014 of the european parliament and of the council of 22 october 2014 on the prevention and management of the introduction and spread of invasive alien species, http://data.europa.eu/eli/reg/2014/1143/2019-12-14, [Online; accessed 11-March-2024].

Flint, M.L., Van den Bosch, R., 2012. Introduction to Integrated Pest Management. Springer Science & Business Media.

Francia, M., Gallinucci, E., Golfarelli, M., Leoni, A.G., Rizzi, S., Santolini, N., 2021. Making data platforms smarter with moses. Futur. Gener. Comput. Syst. 125, 299–313. https://doi.org/10.1016/j.future.2021.06.031.

FreshPlaza.it, 2021. Cimici asiatiche in rete oppure una rete contro le cimici? https://www.freshplaza.it/article/9316773/cimici-asiatiche-in-rete-oppure-una-rete-contro-le-cimici/, [Online; accessed 8-November-2022].

Gallinucci, E., Golfarelli, M., Rizzi, S., 2020. Mo.Re.Farming: a hybrid architecture for tactical and strategic precision agriculture. Data Knowl. Eng. 129, 101836 https://doi.org/10.1016/j.datak.2020.101836.

Gebbers, R., Adamchuk, V.I., 2010. Precision agriculture and food security. Science 327 (5967), 828–831.

Giannetti, D., Patelli, N., Palazzetti, L., Sorbelli, F.B., Pinotti, C.M., Maistrello, L., 2024. First use of unmanned aerial vehicles to monitor *Halyomorpha halys* and recognize it using artificial intelligence. Pest Manag. Sci. https://doi.org/10.1002/ps.8115.

Golfarelli, M., Rizzi, S., 2009. Data Warehouse Design: Modern Principles and Methodologies. McGraw-Hill, Inc.

Hancock, T.J., Lee, D.-H., Bergh, J.C., Morrison III, W.R., Leskey, T.C., 2019. Presence of the invasive brown marmorated stink bug *Halyomorpha halys* (Stål) (Hemiptera: Pentatomidae) on home exteriors during the autumn dispersal period: results generated by citizen scientists. Agric. For. Entomol. 21 (1), 99–108. https://doi.org/10.1111/afe.12312.

Haye, T., Abdallah, S., Gariepy, T., Wyniger, D., 2014. Phenology, life table analysis and temperature requirements of the invasive brown marmorated stink bug, *Halyomorpha halys*, in Europe. J. Pest. Sci. 87 (3), 407–418.

Invasive Species Advisory Council, 2023. Invasive species threaten the success of climate change adaptation efforts. https://www.cakex.org/documents/invasive-species-threaten-success-climate-change-adaptation-efforts, [Online; accessed 11-March-2024].

Kamilaris, A., Kartakoullis, A., Prenafeta-Boldú, F.X., 2017. A review on the practice of big data analysis in agriculture. Comput. Electron. Agric. 143, 23–37.

Kamiyama, M.T., Matsuura, K., Yoshimura, T., Yang, C.-C.S., 2021. Improving invasive species management using predictive phenology models: an example from brown marmorated stink bug (*Halyomorpha halys*) in Japan. Pest Manag. Sci. 77 (12), 5489–5497.

Kirkpatrick, D.M., Acebes-Doria, A.L., Rice, K.B., Short, B.D., Adams, C.G., Gut, L.J., Leskey, T.C., 2019. Estimating monitoring trap plume reach and trapping area for nymphal and adult *Halyomorpha halys* (Hemiptera: Pentatomidae) in crop and non-crop habitats. Environ. Entomol. 48 (5), 1104–1112.

Kloppenburg, J.R., 2004. First the Seed: The Political Economy of Plant Biotechnology, 1492–2000, Second edition. Cambridge University Press. https://doi.org/10.2307/3115976.

Lagos-Ortiz, K., Salas-Zárate, M.D.P., Paredes-Valverde, M.A., García-Díaz, J.A., Valencia-Garcia, R., 2020. Agrient: a knowledge-based web platform for managing insect pests of field crops. Appl. Sci. 10 (3), 1040.

Lara, M., Paride, D., 2014. *Halyomorpha halys* stål 1855, Trovata per la prima Volta Nelle Alpi Centrali Italiane (insecta: Heteroptera: Pentatomidae).

Lee, D.-H., Short, B.D., Joseph, S.V., Bergh, J.C., Leskey, T.C., 2013. Review of the biology, ecology, and management of *Halyomorpha halys* (Hemiptera: Pentatomidae) in China, Japan, and the Republic of Korea. Environ. Entomol. 42 (4), 627–641.

Leonelli, S., Williamson, H.F., 2023a. Artificial intelligence in plant and agricultural research, artificial intelligence for science. A Deep Learn. Revolut. 319–333. https://doi.org/10.1142/9789811265679_0018.

Leonelli, S., Williamson, H.F., 2023b. Introduction: Towards Responsible Plant Data Linkage. Springer International Publishing, Cham, pp. 1–24. https://doi.org/10.1007/978-3-031-13276-6_1.

Leskey, T.C., Hamilton, G.C., Nielsen, A.L., Polk, D.F., Rodriguez-Saona, C., Bergh, J.C., Herbert, D.A., Kuhar, T.P., Pfeiffer, D., Dively, G.P., et al., 2012. Pest status of the brown marmorated stink bug, *Halyomorpha halys* in the USA. Outlooks Pest Manag. 23 (5), 218–226.

Leskey, T.C., Andrews, H., Bády, A., Benvenuto, L., Bernardinelli, I., Blaauw, B., Bortolotti, P.P., Bosco, L., Di Bella, E., Hamilton, G., et al., 2021. Refining pheromone lures for the invasive *Halyomorpha halys* (Hemiptera: Pentatomidae) through collaborative trials in the United States and Europe. J. Econ. Entomol. 114 (4), 1666–1673.

Lina, Y., Xiuming, Y., 2020. Design of intelligent pest monitoring system based on image classification algorithm. In: In: 2020 3rd International Conference on Control and Robots (ICCR). IEEE, pp. 21–24.

Liu, K., Mu, Y., Chen, X., Ding, Z., Song, M., Xing, D., Li, M., 2022. Towards developing an epidemic monitoring and warning system for diseases and pests of hot peppers in Guizhou, China. Agronomy 12 (5), 1034.

Maistrello, L., Dioli, P., Bariselli, M., Mazzoli, G.L., Giacalone-Forini, I., 2016. Citizen science and early detection of invasive species: phenology of first occurrences of *Halyomorpha halys* in southern Europe. Biol. Invasions 18 (11), 3109–3116.

Maistrello, L., Vaccari, G., Caruso, S., Costi, E., Bortolini, S., Macavei, L.I., Foca, G., Ulrici, A., Bortolotti, P.P., Nannini, R., Casoli, L., Fornaciari, M., Mazzoli, G.L., Dioli, P., 2017. Monitoring of the invasive *Halyomorpha halys*, a new key pest of fruit orchards in northern Italy. J. Pest. Sci. 90, 1231–1244.

Maistrello, L., Dioli, P., Dutto, M., Volani, S., Pasquali, S., Gilioli, G., 2018. Tracking the spread of sneaking aliens by integrating crowdsourcing and spatial modeling: the Italian invasion of *Halyomorpha halys*. BioScience 68 (12), 979–989.

Malek, R., Tattoni, C., Ciolli, M., Corradini, S., Andreis, D., Ibrahim, A., Mazzoni, V., Eriksson, A., Anfora, G., 2018. Coupling traditional monitoring and citizen science to disentangle the invasion of *Halyomorpha halys*. ISPRS Int. J. Geo Inf. 7, 171.

Malek, R., Zapponi, L., Eriksson, A., Ciolli, M., Mazzoni, V., Anfora, G., Tattoni, C., 2019. Monitoring 2.0: update on the *Halyomorpha halys* invasion of Trentino. ISPRS Int. J. Geo Inf. 8, 564.

Mcintyre, B., Herren, H., Wakhungu, J., Watson, R., 2009. Agriculture at a Crossroads: The Global Report. Island Press. URL https://wedocs.unep.org/20.500.11822/8590.

McPherson, J.E., 2018. Invasive Stink Bugs and Related Species (Pentatomoidea): Biology, Higher Systematics, Semiochemistry, and Management. CRC Press.

Moore, L.C., 2018. Population Dynamics of the Brown Marmorated Stink Bug, *Halyomorpha halys* (Stal), in the Veneto Region of Italy and its Damage Potential to Kiwifruit and Cherry. Ph.D. thesis,. University of Georgia.

Morey, A., Venette, R., 2023. Mitigating invasive insect species: Eradication, long-term management, and the importance of sampling and monitoring. In: Fountain, M., Pope, T. (Eds.), Advances in Monitoring of Native and Invasive Insect Pests of Crops. Burleigh Dodds Science Publishing, pp. 325–364. https://doi.org/10.19103/as.2022.0113.15.

Morrison, W.R., Milonas, P., Kapantaidaki, D.E., Cesari, M., Di Bella, E., Guidetti, R., Haye, T., Maistrello, L., Moraglio, S.T., Piemontese, L., et al., 2017. Attraction of *Halyomorpha halys* (Hemiptera: Pentatomidae) haplotypes in North America and Europe to baited traps. Sci. Rep. 7 (1), 1–11.

Myneni, R.B., Hall, F.G., Sellers, P.J., Marshak, A.L., 1995. The interpretation of spectral vegetation indexes. IEEE Trans. Geosci. Remote Sens. 33 (2), 481–486.

Nielsen, A.L., Hamilton, G.C., 2009. Life history of the invasive species *Halyomorpha halys* (Hemiptera: Pentatomidae) in northeastern United States. Ann. Entomol. Soc. Am. 102 (4), 608–616.

Nielsen, A.L., Hamilton, G.C., Matadha, D., 2008. Developmental rate estimation and life table analysis for *Halyomorpha halys* (Hemiptera: Pentatomidae). Environ. Entomol. 37 (2), 348–355.

Nielsen, A.L., Holmstrom, K., Hamilton, G.C., Cambridge, J., Ingerson-Mahar, J., 2013. Use of black light traps to monitor the abundance, spread, and flight behavior of *Halyomorpha halys* (Hemiptera: Pentatomidae). J. Econ. Entomol. 106 (3), 1495–1502.

Nielsen, A.L., Chen, S., Fleischer, S.J., 2016. Coupling developmental physiology, photoperiod, and temperature to model phenology and dynamics of an invasive heteropteran, *Halyomorpha halys*. Front. Physiol. 7 https://doi.org/10.3389/fphys.2016.00165.

Ojha, T., Misra, S., Raghuwanshi, N.S., 2015. Wireless sensor networks for agriculture: the state-of-the-art in practice and future challenges. Comput. Electron. Agric. 118, 66–84.

Paini, D.R., Sheppard, A.W., Cook, D.C., De Barro, P.J., Worner, S.P., Thomas, M.B., 2016. Global threat to agriculture from invasive species. Proc. Natl. Acad. Sci. USA 113 (27), 7575–7579. https://doi.org/10.1073/pnas.1602205113.

Prasad, Y., Prabhakar, M., 2012. Pest monitoring and forecasting, integrated pest management: principles and practice. Oxfordshire, UK: CABI, pp. 41–57.

Preti, M., Vaccari, G., Landi, L., Rufolo, E., Fagioli, L., Manucci, F., Fabbri, G., Civolani, S., Profeta, M., Delvago, C., Colla, R., Golfarelli, M., Chiari, G., Tommasini, M.G., Butturini, A., Bariselli, M., Maistrello, L., 2021a. Monitorare la cimice asiatica con Cimice. Net, L'Informatore Agrario 32 (2021), 58–61.

Preti, M., Verheggen, F., Angeli, S., 2021b. Insect pest monitoring with camera-equipped traps: strengths and limitations. J. Pest. Sci. 94 (2), 203–217.

R. C. Team, 2013. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. R-project. org/, vienna, austria. http://www.

Ravat, F., Zhao, Y., 2019. Data lakes: Trends and perspectives. In: Proc. DEXA, Linz, Austria, pp. 304–313.

Renault, D., Angulo, E., Cuthbert, R.N., Haubrock, P.J., Capinha, C., Bang, A., Kramer, A. M., Courchamp, F., 2022. The magnitude, diversity, and distribution of the economic costs of invasive terrestrial invertebrates worldwide. Sci. Total Environ. 835 https://doi.org/10.1016/j.scitotenv.2022.155391.

Reshef, D.N., Reshef, Y.A., Finucane, H.K., Grossman, S.R., McVean, G., Turnbaugh, P.J., Lander, E.S., Mitzenmacher, M., Sabeti, P.C., 2011. Detecting novel associations in large data sets. Science 334 (6062), 1518–1524.

Rice, K.B., Morrison III, W.R., Short, B.D., Acebes-Doria, A., Bergh, J.C., Leskey, T.C., 2018. Improved trap designs and retention mechanisms for *Halyomorpha halys* (Hemiptera: Pentatomidae). J. Econ. Entomol. 111 (5), 2136–2142.

Rivest, S., 2001. Toward better support for spatial decision making: defining the characteristics of spatial on-line analytical processing (SOLAP). Geomatica 55 (4), 539–555.

Rot, M., Maistrello, L., Costi, E., Trdan, S., 2022. Biological parameters, phenology and temperature requirements of *Halyomorpha halys* (Hemiptera: Pentatomidae) in the sub-Mediterranean climate of Western Slovenia. Insects 13 (10), 956. https://doi.org/10.3390/insects13100956.

Sawant, M., Urkude, R., Jawale, S., 2016. Organized data and information for efficacious agriculture using PRIDE model. Int. Food Agribusiness Manag. Rev. 19 (A).

Short, B.D., Khrimian, A., Leskey, T.C., 2016. Pheromone-based decision support tools for management of *Halyomorpha halys* in apple orchards: development of a trap-based treatment threshold. J. Pest. Sci. 90, 1191–1204.

Stein, B., Morrison, A., 2014. The enterprise data lake: better integration and deeper analytics. PwC Technol. Forecast: Rethink. Integrat. 1, 1–9.

StopBMSB.org, 2024. Management of brown marmorated stink bug in us specialty crops. https://www.stopbmsb.org/where-is-bmsb/state-by-state/, [Online; accessed 14-May-2024].

United States Department of Interior, 2016. National Invasive Species Council, Nisc management plan 2016–2018. https://www.doi.gov/sites/doi.gov/files/uploads/2016-2018-nisc-management-plan.pdf, [Online; accessed 11-March-2024].

Vaccari, G., Preti, M., Bariselli, M., Fagioli, L., Civolani, S., Tommasini, M.G., Golfarelli, M., Gallinucci, E., Forresi, C., Maistrello, L., 2022. A territorial monitoring system for *Halyomorpha halys* in Emilia-Romagna region, PheroFruit22. In: Joint Meeting of the IOBC/WPRS Working Groups "Pheromones and other semiochemicals in integrated production" "Integrated Protection of Fruit Crops", 25-29 September 2022. Girona, Spain.

Vanegas, F., Bratanov, D., Powell, K., Weiss, J., Gonzalez, F., 2018. A novel methodology for improving plant pest surveillance in vineyards and crops using uav-based hyperspectral and spatial data. Sensors 18 (1), 260.

Venugopal, P.D., Martinson, H.M., Bergmann, E.J., Shrewsbury, P.M., Raupp, M.J., 2015a. Edge effects influence the abundance of the invasive *Halyomorpha halys* (Hemiptera: Pentatomidae) in woody plant nurseries. Environ. Entomol. 44 (3), 474–479.

Venugopal, P.D., Dively, G.P., Lamp, W.O., 2015b. Spatiotemporal dynamics of the invasive *Halyomorpha halys* (Hemiptera: Pentatomidae) in and between adjacent corn and soybean fields. J. Econ. Entomol. 108 (5), 2231–2241.

Vétek, G., Melifronidou-Pantelidou, A., Koukkoularidou, D., Martinou, A.F., 2021. Initiation of a monitoring programme for early detection of *Halyomorpha halys* in Cyprus by using pheromone-baited traps and involving citizen science. Manag. Biol. Invasions. 12 (2), 331–343.

Vibhute, A., Bodhe, S., 2012. Applications of image processing in agriculture: a survey. Int. J. Comput. Appl. 52 (2).

Wang, Q.-J., Zhang, S.-Y., Dong, S.-F., Zhang, G.-C., Yang, J., Li, R., Wang, H.-Q., 2020. Pest24: a large-scale very small object data set of agricultural pests for multi-target detection. Comput. Electron. Agric. 175, 105585 https://doi.org/10.1016/j.compag.2020.105585.

Weber, D.C., Morrison, W.R., Khrimian, A., Rice, K.B., Leskey, T.C., Rodriguez-Saona, C., Nielsen, A.L., Blaauw, B.R., 2017. Chemical ecology of *Halyomorpha halys*: discoveries and applications. J. Pest. Sci. 90 (4), 989–1008.

Zapponi, L., Tortorici, F., Anfora, G., Bardella, S., Bariselli, M., Benvenuto, L., Bernardinelli, I., Butturini, A., Caruso, S., Colla, R., Costi, E., Culatti, P., Di Bella, E., Falagiarda, M., Giovannini, L., Haye, T., Maistrello, L., Malossini, G., Marazzi, C., Marianelli, L., Mele, A., Michelon, L., Moraglio, S.T., Pozzebon, A., Preti, M., Salvetti, M., Scaccini, D., Schmidt, S., Szalatnay, D., Roversi, P.F., Tavella, L., Tommasini, M.G., Vaccari, G., Zandigiacomo, P., Sabbatini-Peverieri, G., 2021. Assessing the distribution of exotic egg parasitoids of *Halyomorpha halys* in europe with a large-scale monitoring program. Insects 12 (4), 316. https://doi.org/10.3390/insects12040316.

Zburivsky, D., Partner, L., 2021. Designing Cloud Data Platforms. Simon and Schuster.

Zobel, E., Hooks, C., Dively, G., 2016. Seasonal abundance, host suitability, and feeding injury of the brown marmorated stink bug, *Halyomorpha halys* (Heteroptera: Pentatomidae), in selected vegetables. J. Econ. Entomol. 109 (3), 1289–1302.