A harmonised dataset of adult Culex pipiens mosquitoes from 17 administrative Italian regions, 2008–2022

Longo Eleonora[1,+,*], Blaha Margo[1,+], Accorsi Annalisa[2], Antognini Elisa[3], Arnoldi Daniele[4], Buonanno Marialuisa[5], Cafiero Maria Assunta[6], Calzolari Mattia[7], Corona Cristiano[2], De Liberato Claudio[8], de Martinis Claudio[5], Defilippo Francesco[7], Di Pasquale Maria Liliana[9], Foxi Cipriano[10], Gavaudan Stefano[3], Gobbo Federica[11], La Russa Francesco[9], Listorti Valeria[2], Montarsi Fabrizio[11], Mosca Andrea[12], Nardella La Porta Maria Concetta[6], Perna Mirko[12], Rizzoli Annapaola[4], Romiti Federico[8], Satta Giuseppe[10], Tessarolo Carlotta[2], Virgillito Chiara[13], Rosà Roberto[1], Caputo Beniamino[13,#], Da Re Daniele[4,1,#]

[1]Center Agriculture Food Environment, University of Trento, San Michele all'Adige (TN), Italy.

[2]Istituto Zooprofilattico Sperimentale del Piemonte, Liguria e Valle d'Aosta, Turin, Italy.

[3]Istituto Zooprofilattico dell'Umbria e delle Marche "Togo Rosati", Perugia, Italy.

[4]Fondazione Edmund Mach, San Michele all'Adige (TN), Italy.

[5]Istituto Zooprofilattico Sperimentale del Mezzogiorno, Portici (NA), Italy.

[6]Istituto Zooprofilattico Sperimentale della Puglia e della Basilicata, Foggia, Italy.

[7]Istituto Zooprofilattico Sperimentale della Lombardia e dell'Emilia-Romagna "Bruno Ubertini", Brescia, Italy.

[8]Istituto Zooprofilattico Sperimentale delle Regioni Lazio e Toscana, Rome, Italy.

[9]Istituto Zooprofilattico Sperimentale della Sicilia "A. Mirri", Palermo, Italy.

[10]Istituto Zooprofilattico Sperimentale della Sardegna, Sassari, Italy.

[11]Istituto Zooprofilattico Sperimentale delle Venezie, Legnaro, Italy.

[12]Istituto per le Piante da Legno e l'Ambiente S.p.A., Turin, Italy.

[13]Sapienza University of Rome, Department of Public Health and Infectious Diseases, Rome, Italy.

[+]These authors contributed equally.

[#]These authors contributed equally.

* Corresponding author.

Eleonora Longo, Center of Agriculture, Food and Environment (C3A), University of Trento, Via Edmund Mach, 1 - 38098 San Michele all'Adige (TN), Italy, eleonora.longo@unitn.it

Abstract

Culex pipiens s.l. is the primary vector of West Nile Virus in Italy. Despite the availability of abundant long-term surveillance data on adult *Cx. pipiens* observations in Italy, historical records are often fragmented across regional repositories with heterogeneous formats and metadata. This work introduces a comprehensive, harmonised dataset of adult *Cx. pipiens* observations collected across 17 Italian administrative regions from 2008 to 2022. The dataset, compiled from official surveillance programmes conducted by the Experimental Zooprophylactic Institutes network and collaborating institutions, encompasses 58,888 raw observational records of adult *Cx. pipiens*, corresponding to 37,070 unique trap deployments. To ensure cross-regional comparability, these overlapping local observations were harmonised and spatio-temporally aggregated onto a 9 × 9 km reference grid compatible with ERA5-Land climate reanalysis, providing weekly median abundances. The data were stratified by trap type and mosquito sex, resulting in 50,759 analysis-ready records. Following rigorous quality-control procedures for taxonomic reliability and spatio-temporal consistency, this dataset constitutes an essential resource for spatio-temporal modelling of vector phenology, epidemiological forecasting, and advancing integrated One Health preparedness.

Background & Summary

Culex pipiens s.l. is among the most widespread and abundant mosquito species across the Western Palearctic region[1,2] and the principal vector sustaining West Nile Virus (WNV) and Usutu virus (USUV) transmission in Europe[3,4]. Over the past two decades, WNV transmission has expanded markedly in both geographical range and intensity, with recurrent large-scale outbreaks across south-eastern and central Europe culminating in the unprecedent 2018 epidemic, which resulted in over 2,000 confirmed human cases in a single transmission season[5,6].

Italy has been among the most affected countries in this epidemiological scenario, experiencing recurrent seasonal WNV circulation with marked interannual variability and periodic outbreaks[5] since the re-emergence of the virus in 2008[5,7]. The population dynamics of this species are closely linked to environmental drivers, such as temperature and water availability, which determine larval development rates, survival, and adult activity[8,9]. In the temperate climatic zones that characterise Italy, *Cx. pipiens* displays a pronounced seasonal pattern, with densities typically reaching their peak during the warm summer months. To survive during the winter, some adult females enter a state of reproductive diapause, seeking refuge in sheltered microhabitats before resuming host-seeking and oviposition activities in the spring[10]. Routinely collected entomological surveillance data provide essential evidence on vector abundance and seasonal dynamics that can support risk assessment and targeted intervention planning.

Italy's WNV risk management relies on integrated surveillance on humans, animals, and vectors, coordinated within the National Arbovirus Prevention, Surveillance, and Response Plan[11] (PNA 2020–2025) and associated operational procedures. Within this framework, adult mosquito monitoring provides an actionable "upstream" signal supporting risk assessment, targeted vector-control decisions, and prevention of human infection through activation of blood and substance-of-human-origin safety measures. Entomological surveillance activities, targeting mainly *Cx. pipiens*, are implemented through a network of regional and national public-health and research institutions, including the Experimental Zooprophylactic Institutes (EZIs) and collaborating institutions, with reporting pathways aligned to national coordination.

Extensive long-term surveillance data on adult *Cx. pipiens* observations have been accumulated across Italy. However, they frequently differ markedly in structure, encoding and completeness (e.g. tabular layout, column naming conventions, and date formats), and are accompanied by uneven metadata (e.g., trap type, sampling duration, georeferencing conventions, and taxonomic reporting). Furthermore, sampling effort varies substantially among regions, with marked differences in spatial coverage, temporal extent, and sampling intensity. This fragmentation constrains reproducibility and cross-regional comparability, and it complicates the use of these data in statistical and mechanistic models that estimate or forecast mosquito abundance and seasonality.

Here we present a harmonised dataset of adult *Cx. pipiens* entomological observations from Italy covering 2008–2022, compiled from multiple surveillance programmes. The primary field observation is the number of *Cx. pipiens* specimens collected by a single trap in a single night of deployment. As traps were typically active for one night per session, the reported count represents a single trap-night observation. These trap-level observations were then spatially aggregated onto a 9 x 9 km reference grid compatible with ERA5-Land reanalysis[12] and temporally aggregated by sampling week. Specifically, for each unique combination of grid cell, sampling week, trap type, and mosquito sex, the dataset reports the median of all trap-night counts, yielding a weekly median abundance per grid cell. The use of the median as the aggregation function mitigates the impact of local outliers, while stratification by trap type prevents the conflation of data from devices with different capture efficiencies. Residual differences in sampling effort, such as the number of traps per cell, temporal sampling intensity and unequal surveillance coverage among regions, are not corrected by the aggregation itself. However, effort metadata are provided for each record to support

downstream analytical adjustments. This work aligns with broader community efforts to standardise and disseminate vector surveillance data through the VectAbundance initiative[13], which leverages harmonised time series to enhance ecological modelling and public-health preparedness.

To ensure such interoperability, our harmonisation focused on: (i) consistent representation of dates and locations; (ii) standardisation of sampling-effort fields used to derive trap-night counts; (iii) controlled vocabularies for trap typology and key descriptors; and (iv) transparent quality-control steps designed to preserve traceability to original records while producing an analysis-ready table.

Methods

In the context of the Italian PNA monitoring, we contacted EZIs representatives to ask for historical *Cx. pipiens* abundance data.

We compiled a harmonised, quality-controlled dataset of adult *Cx. pipiens* trap records collected between 2008 and 2022 across 17 Italian administrative regions (Figure 1 – Sampled regions). Trap records were spatially referenced and aggregated onto a 9 × 9 km national grid defined by the ERA5-Land climate reanalysis[12], ensuring compatibility with environmental covariates and geographical correspondence with VectAbundance released *Aedes albopictus* eggs database[13]. Trap records were also temporally aggregated by year and sampling week. To ensure comparability across regions, which varied in surveillance intensity and operational protocols, we implemented consistent data-cleaning procedures and standardised variable definitions. The complete list of variables included in the final dataset[14] is provided in Table 2.

Due to confidentiality agreements with the contributing institutions, the authors cannot publicly share the raw observational data. Researchers interested in accessing these raw *Cx. pipiens* trap records observations must reach out to the original providing institutes directly. The necessary contact details can be found under the “contact_person” and “contact_person_email” variables within the dataset. For each institution, provision of the raw records and formal approval for the publication of the manuscript and of the derived, harmonised dataset were documented through a structured consent process completed by the institutional referent and further formalised for a subset of institutions through a bilateral institutional agreement.

Sampling methods

Study Area

The study area encompasses the Italian peninsula and its two major islands, specifically comprising 17 administrative regions that actively contributed entomological data to the surveillance network: Piedmont, Liguria, Lombardy, Autonomous Province of Trento, Veneto, Friuli-Venezia Giulia, Emilia-Romagna, Tuscany, Umbria, Marche, Lazio, Campania, Calabria, Apulia, Basilicata, Sardinia, and Sicily.

This extensive spatial coverage incorporates a wide heterogeneity of landscapes and climatic regimes. Topographically, the sampled regions range from sea level along the coastal plains to higher elevations within the Alpine and Apennine mountain ranges, generating pronounced thermal and precipitation gradients[15]. Climatically, the northern plains and inland valleys are predominantly characterised by a temperate climate with year-round precipitation. In contrast, the central and southern territories, alongside Sardinia and Sicily, exhibit a Mediterranean climate defined by hot, dry summers and mild, wet winters. Land cover across these regions comprises diverse environments, including highly urbanised settlements, intensive agricultural areas sustained by extensive irrigation networks, mixed temperate forests, and Mediterranean scrub[15].

Entomological Sampling Framework

The entomological surveillance of WNV in Italy operates under the PNA, established by the Ministry of Health in collaboration with regional authorities and the network of Italian EZIs. It's primary objective, unchanged throughout the study period, is the early detection of viral circulation in competent vectors, animal hosts and reservoirs in order to mitigate transmission risks to humans and equids.

The framework has been deeply shaped by Italy's increasingly complex epidemiological landscape. Since its first detection in Tuscany in 1998 during an equine outbreak[16], followed by a decade of epidemiological silence, WNV Lineage 1 (WNV-L1) re-emerged in 2008 in the Po River valley[7]. In 2011, WNV Lineage 2 (WNV-L2) was introduced and rapidly spread, progressively replacing WNV-L1 to become the dominant endemic strain in the period 2013-2017[17,18]. WNV-L2 was the primary cause of the severe 2018 epidemic, representing the largest WNV outbreak in European history[6]. More recently, the scenario has been characterised by the co-circulation of both lineages: after years of sporadic detection, a highly pathogenic WNV-L1 strain re-emerged in northern Italy during the 2021–2022 vector seasons, co-circulating with the endemic WNV-L2 and causing a significant surge in severe neurological cases[19]. In response, surveillance strategies and geographical stratification have continuously evolved (Figure 1). The country shifted from localised emergency responses to a structured, risk-based integrated surveillance system, which is directly reflected in the significant variations in sampling effort across the regions captured by the harmonised dataset[14] (Table 1).

Four phases of the surveillance system can be identified: an initial emergency response (2008), a standardisation and multi-trap phase (2009–2013), a consolidation phase focused on endemic areas (2014–2019), and the integrated surveillance phase under the PNA (2020–2025). While the underlying objective and the technical infrastructure remained stable across these phases, the spatial stratification, sampling intensity, and trap deployment strategy evolved substantially, as described below.

A common suite of mosquito trapping devices, targeting distinct physiological states and host-seeking behaviours, was used consistently throughout the study period, although their combination and intensity of deployment varied across phases and regions. Nocturnal host-seeking female mosquitoes were sampled with CDC miniature light traps (John W. Hock Company, USA). These traps utilise a suspended light source and a downdraft fan to draw insects into a collection net. Additionally, CDC-like traps were used. These traps consist of modified suction devices that operate via a similar fan mechanism but specifically lack a light source (Italian Mosquito Trap [IMT]; PeP, Cantù, Italy). Both types were often baited with carbon dioxide (dry ice or compressed gas) to mimic vertebrate respiration. In contrast, BG-Sentinel traps (Biogents AG, Germany), although primarily designed for diurnal anthropophilic species, were also deployed as they effectively collect *Cx. pipiens* during crepuscular activity periods. These traps employ a counter-flow principle, using a central fan to draw approaching mosquitoes downward into a collection net while dispersing synthetic attractants that reproduce human skin volatiles such as lactic acid and ammonia. Conversely, to capture mosquitoes at a later stage of their gonotrophic cycle, CDC Gravid traps (John W. Hock Company, USA) were utilised to selectively target *Culex* spp. females seeking oviposition sites. These devices feature a collection basin filled with an organic infusion (hay-macerated water) to provide olfactory cues, coupled with an overlying suction tube that captures gravid females as they approach the water surface. Finally, these passive trapping devices were complemented by active aspiration in animal shelters, such as stables and poultry houses. This targeted approach allowed for the collection of resting and blood-fed *Cx. pipiens* individuals that might otherwise evade standard traps, providing a more comprehensive overview of local host–vector interactions.

Emergency Surveillance and Baseline Data Collection (2008). Following the emergence of WNV in Italy, the Ministry of Health activated a National Surveillance Plan[20,21] focused on 15 identified risk areas selected

for their ecological features, such as wetlands and Important Bird Areas (IBAs). In particular, IBAs are internationally recognised sites of significance for bird conservation[22]. Spatially, each study area was defined by the cluster of standardised grid cells (20 km x 20 km), adopted from the National Bluetongue Surveillance Plan[23] and falling within a 20 km radius of specific central coordinates. The protocol mandated a comprehensive mosquito sampling strategy through combined larval and adult collections: larval surveys were conducted fortnightly with standard 0.5 L dippers, while host-seeking adults were monitored with CDC miniature light traps and CDC-like traps positioned at a height of 1.5–2.0 metres near animal shelters. These traps operated fortnightly from April to October for a single night (dusk to dawn) per session and were optionally baited with CO_2 to enhance sensitivity. Diurnal species were monitored with BG-Sentinel CO_2-baited traps, activated monthly from morning until dusk. Furthermore, to sample resting blood-fed females, aspiration catches were performed fortnightly during early morning hours inside animal shelters such as stables and poultry houses. All collected specimens were morphologically identified[2,24].

Standardisation and Multi-Trap Strategy (2009–2013). In 2009, the surveillance strategy underwent a significant refinement[25,26] to optimise the early detection of viral circulation, establishing a standardised protocol that remained in use until 2013[27-34]. The national territory was stratified based on epidemiological evidence into Areas with Viral Circulation (ACV), External Surveillance Areas (20 km buffer), and Risk Areas. To normalise sampling effort across provinces, the plan introduced "Geographical Reference Units" of approximately 1200–1600 km². From 2010 onwards, this standardisation became stricter, mandating surveillance in at least two sentinel farms per unit, selected based on historical WNV circulation or proximity to wetlands. Unlike the initial emergency phase, routine larval sampling was discontinued in established areas in favour of a comprehensive multi-trap protocol designed to target vectors with diverse ethological behaviours. Throughout this period, each sentinel site was equipped with three specific trap types operating simultaneously: CO_2-baited CDC light traps and CDC-like traps positioned at 1.5 m height, active for two consecutive nights per session (as opposed to the single dusk-to-dawn session used in 2008); CDC Gravid traps deployed in sheltered areas and baited with a fermented infusion of water, hay, and brewer's yeast (macerated for 24–48 hours); and BG-Sentinel traps operated continuously for 48 hours before sample collection. Sampling frequency evolved over the period: while initially conducted monthly (2009) to monitor seasonal dynamics and overwintering, the schedule was intensified between 2010 and 2013 to a fortnightly cadence during the vector season (typically March–October). Aspiration catches in animal shelters continued to complement passive trapping.

Consolidation of Endemic Areas (2014–2019). From 2014, the three-category zoning of 2009-2013 was simplified into two main categories: Endemic Areas (provinces with recurrent WNV circulation) and the Rest of the National Territory[35-39]. In Endemic Areas, the geographical reference unit remained the provincial territory subdivided into 1200–1600 km² zones. The entomological protocol prioritised the use of at least one CO_2-baited CDC or CDC-like trap per site. Collections were conducted fortnightly starting from April until two consecutive negative captures were recorded, typically ending in late October. In 2017, the surveillance scope expanded to include the Usutu virus (USUV) under a "One Health" Integrated National Surveillance Plan[40] (PNI), unifying veterinary and human surveillance efforts. Mosquitoes were collected alive or frozen, and identification was performed on cold-chain preserved specimens.

National Plan for Arbovirus Prevention, Surveillance, and Response (PNA 2020–2025). The current framework is regulated by the PNA[11], ratified by the State-Regions Conference on 15 January 2020 for the period 2020–2025 and subsequently extended through 2026, as the successor plan had not been yet ratified at the time of manuscript submission. Building on the 2014-2019 consolidation phase, the PNA further refines spatial stratification and formalises implementation strategies tailored to local epidemiological risk, determined by WNV circulation over the previous five years and eco-climatic conditions. Consequently, the national territory is stratified into three geographical zones.

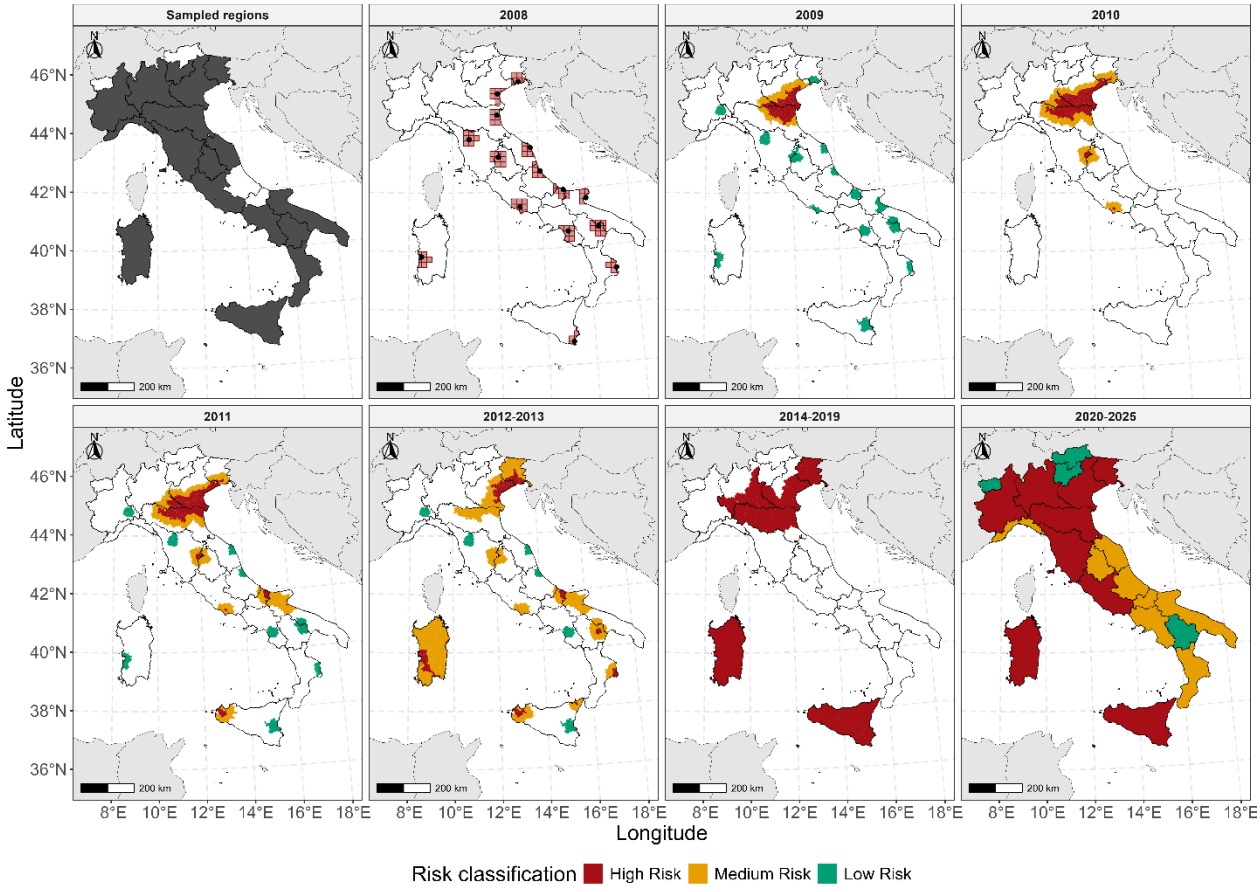
High Risk Areas (AR) include provinces where WNV is currently circulating or has circulated in at least one of the previous five years, showing recurrent infection episodes. Surveillance is comprehensive, encompassing monitoring of resident target bird species, entomological activities, passive surveillance of equids with neurological symptoms, testing of wild birds found dead, and human surveillance for neuroinvasive diseases.

Low Risk Areas (BR) are defined as territories with sporadic historical circulation or no prior detection yet possessing favourable eco-climatic characteristics for viral transmission. Surveillance activities generally mirror those of High-Risk Areas, although monitoring may focus on rural poultry farms as an alternative to resident wild birds.

Minimum Risk Areas (RM) comprise territories with no history of circulation and unfavourable conditions for transmission. Here, surveillance is primarily passive, focusing on neurological cases in equids, wild bird mortality, and human cases.

Entomological surveillance follows a standardised spatial protocol. The territory is divided into regular grid cells, excluding areas above 600 m a.s.l., with at least one trap per cell so that the area monitored by a single trap does not exceed 400 km² (typically a grid of 20×20 km). Sampling is conducted fortnightly from April to November, although regional authorities may adjust this timeframe based on local climatic and meteorological trends. Traps are active for a minimum of one night, operating from dusk until the following morning. When sampling extends over multiple nights, mosquitoes are collected and processed at the end of each capture interval.

Figure 1. Geographical coverage of the harmonised dataset and spatiotemporal evolution of West Nile Virus (WNV) risk stratification and entomological surveillance zones in Italy (2008–2025). The Sampled regions panel shows the geographical coverage of the harmonised dataset: regions contributing data are coloured in dark grey. The remaining panels illustrate the progression of surveillance strategies and risk zoning as defined by the evolving Ministerial communications and National Plan for Arbovirus Prevention, Surveillance, and Response (PNA). Colour codes represent varying risk levels; however, the specific epidemiological definitions associated with each colour differ across the regulatory periods as follows. 2008 (Emergency Phase): Surveillance focused on 15 specific sampling sites (red symbols) located in wetlands and Important Bird Areas (IBAs) to establish baseline data; 2009–2013 (Standardisation Phase): Stratification at the municipal level based on viral circulation evidence. Red: Areas with Viral Circulation (ACV); Orange: External Surveillance Areas (20 km buffer surrounding ACV); Green: Risk Areas; 2014–2019 (Consolidation Phase): Zoning simplified to the provincial level. Red: Endemic Areas (provinces with recurrent WNV circulation); uncoloured areas represent the rest of the national territory; 2020–2025 (Integrated Surveillance Phase): Stratification based on historical circulation (previous 5 years) and eco-climatic conditions. Red: High Risk Areas (AR – Aree ad Alto Rischio); Orange: Low Risk Areas (BR – Aree a Basso Rischio), characterised by sporadic circulation; Green: Minimum Risk Areas (RM – Aree a Rischio Minimo). Note: For the year 2013, the risk zoning was merged with that of 2012; unlike other years, it was not possible to retrieve the complete list of NUTS-3/NUTS-2 assigned to the various risk areas, although their distribution is considered similar to 2012. Also, the figure illustrates the full temporal evolution of the surveillance framework up to the current PNA (2020–2025), while the dataset presented in this study covers the period 2008–2022.



244

Table 1. Summary of spatial and temporal sampling effort and trapping devices deployed for entomological surveillance of *Cx. pipiens*, stratified by Institution and Region. Values represent the yearly mean number of sampled weeks, sampled cells, and specific trap types deployed, along with their respective standard deviations (± SD). Institutes: IZSPLV = Experimental Zooprophylactic Institute of Piedmont, Liguria and Valle d'Aosta; IPLA = Institute for Plants and Environment of Piedmont; IZSLER = Experimental Zooprophylactic Institute of Lombardy and Emilia-Romagna; FEM = Edmund Mach Foundation; IZSVE = Experimental Zooprophylactic Institute of the Venezie; IZSLT = Experimental Zooprophylactic Institute of Lazio and Tuscany; IZSUM = Experimental Zooprophylactic Institute of Umbria and Marche; IZSMPORTICI = Experimental Zooprophylactic Institute of Southern Italy (Portici); IZSPB = Experimental Zooprophylactic Institute of Apulia and Basilicata; IZSSARDEGNA = Experimental Zooprophylactic Institute of Sardinia; IZSSICILIA = Experimental Zooprophylactic Institute of Sicily; . Trap types: ASP = aspiration; BG = BG-Sentinel trap without CO_2; BG-CO2 = BG-Sentinel trap baited with CO_2; CDC = CDC miniature light trap without CO_2; CDC-CO2 = CDC miniature light trap baited with CO_2; CDC-LIKE = CDC-like trap without CO_2; CDC-LIKE-CO2 = CDC-like trap baited with CO_2; GRAVID = CDC Gravid trap. Note: trap types for 11 traps in the Liguria region were unknown and are excluded from the table.

Institution	Region	Sampling period	Yearly mean number of sampled week	Yearly mean number of sampled cells	Yearly mean number of traps deployed (± SD)							
					ASP	BG	BG-CO2	CDC	CDC-CO2	CDC-LIKE	CDC-LIKE-CO2	GRAVID
IZSPLV and IPLA	PIE	2014-2022	18.70 (± 1.41)	52.20 (± 11.20)	-	-	8.89 (± 1.69)	-	45.70 (± 12.00)	-	-	-
	LIG	2013-2022	20.00 (± 4.11)	10.60 (± 2.63)	-	8.90 (± 2.73)	-	0.30 (±0.48)	-	-	-	3.80 (± 2.10)
IZSLER	LOM	2014-2022	19.20 (± 2.05)	39.90 (± 3.92)	-	-	-	-	-	-	39.90 (± 3.92)	-
	EMR	2013-2022	20.40 (± 3.13)	86.10 (± 2.69)	-	-	-	-	-	-	78.70 (± 9.79)	8.00 (± 8.43)
FEM	TN	2008-2022	16.90 (± 9.29)	5.73 (± 2.76)	-	2.27 (± 1.49)	4.53 (± 3.48)	-	0.66 (± 1.63)	-	-	-
IZSVE	VEN	2010-2022	24.50 (± 2.44)	43.50 (± 11.40)	-	-	-	-	43.50 (± 11.40)	-	-	3.46 (± 3.91)
	FVG	2011-2022	22.20 (± 1.90)	15.60 (± 3.68)	-	-	-	-	15.60 (± 3.68)	-	-	0.50 (± 0.52)
IZSLT	TUS	2009-2022	18.90 (± 10.40)	12.70 (± 11.40)	-	-	-	-	12.70 (± 11.40)	-	-	-
	LAZ	2009-2022	15.10 (± 10.90)	9.36 (±9.85)	-	-	-	-	9.36 (± 9.85)	-	-	-
IZSUM	UMB	2020-2022	15.00 (± 6.04)	9.80 (± 2.77)	-	-	-	-	9.60 (± 2.79)	-	-	0.20 (± 0.44)
	MAR	2020-2022	19.40 (± 6.19)	29.60 (± 8.68)	-	-	-	-	29.60 (± 8.68)	-	-	1.20 (± 1.64)
IZSPB	APU	2020-2022	27.30 (± 8.08)	16.70 (± 7.23)	-	9.33 (± 2.31)	-	-	-	-	15.30 (± 7.77)	0.33 (±0.57)
	BAS	2021-2022	23.5 (± 4.95)	4.50 (± 0.70)	-	-	-	-	-	-	4.50 (± 0.70)	-
IZSMPORTICI	CAM	2022-2022	4.00	5.00	-	-	-	-	5.00	-	-	-
	CAL	2022-2022	1.00	1.00	-	-	-	-	1.00	-	-	-
IZSSARDEGNA	SAR	2015-2022	50.90 (± 1.73)	29.40 (± 2.83)	0.25 (± 0.46)	2.75 (± 1.67)	-	29.40 (± 2.83)	-	-	-	-
IZSSICILIA	SIC	2015-2022	34.20 (± 3.54)	19.50 (± 4.31)	-	10.80 (± 9.94)	-	7.75 (± 7.38)	-	8.87 (± 2.47)	-	3.88 (± 2.47)

Data Harmonisation and Quality Control

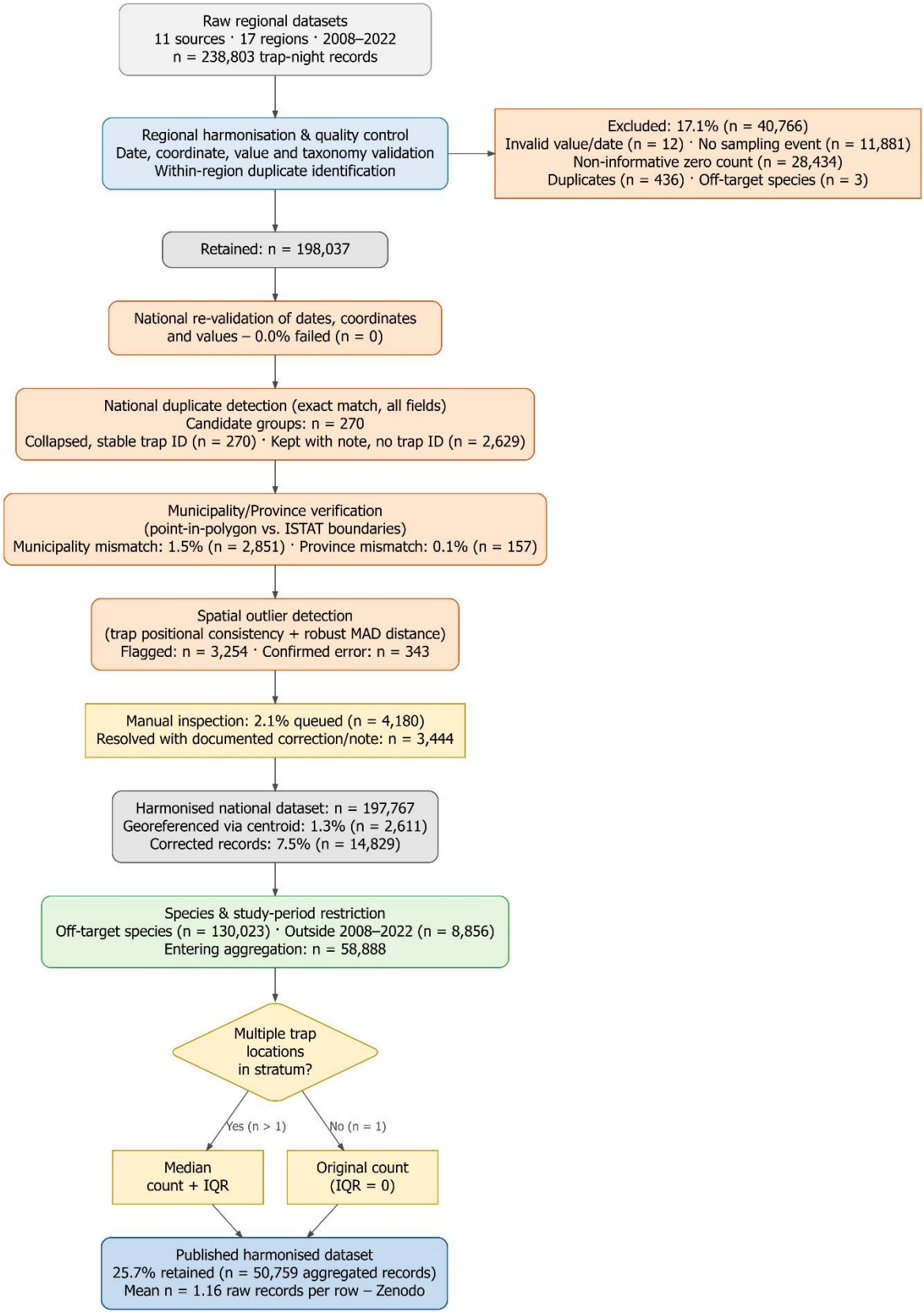
Given the heterogeneous nature of the raw data collected across 17 regions over 15 years, a rigorous harmonisation and quality-control pipeline was developed (Figure 2): region-specific pre-processing across the 11 source institutes, national-level merging, and spatio-temporal aggregation. All data manipulation, cleaning, and standardisation procedures were executed in the R statistical computing environment[41] (version 4.5.2), primarily utilising the *tidyverse* package[42] (version 2.0.0) for data wrangling and the *sf* (version 1.0-24) and *terra* (version 1.8-93) packages[43,44] for spatial operations.

At the regional stage, dataset structures were first unified by mapping regional column headers to a standardised schema. Categorical variables, such as trapping devices, were recoded into consistent operational formats (e.g., unifying various string representations into "CDC_CO2", “CDC_LIKE_CO2”, "BG", “BG_CO2”, "GRAVID" or “ASP”). To resolve taxonomic inconsistencies resulting from different regional data-entry protocols (e.g., use of abbreviations, synonyms, or typographical errors), a comprehensive lookup table mapped raw specie entries to their validated canonical taxonomic hierarchy, from Kingdom to Species. Following the protocol detailed in Da Re et al.[13], trap coordinates were first reprojected to WGS84 (EPSG:4326) using the R *sf* package[43] to ensure a unified coordinate reference system across all regions. A strict geospatial validation protocol was then applied. Municipality and Province names were systematically cleaned by correcting typographical errors, removing special characters, and standardising string cases to match official ISTAT registries[45]. Sampling events lacking precise trap coordinates were georeferenced by imputing the geometric centroid of the corresponding municipality, via a spatial join against reference spatial polygon dataset of Italian administrative boundaries[45]. For coastal regions, individual coordinate readings found by point-in-polygon intersection to fall in the sea were likewise replaced by the municipal centroid. Literal duplicate rows, identical across every field, were collapsed to a single observation at this regional stage. However, several regional sources were found to record abundance at the level of an individual laboratory sub-sample ("pool") rather than a complete field-collection event; as a result, a single trap visit could appear as multiple superficially duplicate rows sharing the same trap and date. Where this structure could be confirmed against the raw source, either by reproducing an independently reported capture total or by direct confirmation from the data owner, these records were summed to avoid under-counting true abundance. Where it could not be confirmed, records were left unaggregated and flagged for follow-up, rather than resolved by assumption. The regional outputs were then concatenated into a single dataframe. A second, national-level exact-duplicate check was applied, grouped on every field except the annotation field; this collapsed duplicate records only where a stable trap identifier confirmed them to be genuine re-entries of the same trap. For the subset of regions whose raw data provide no persistent trap identifier, records that were otherwise identical could not be safely distinguished from independent traps sharing a municipality-level coordinate. These were retained and flagged. At this stage, cross-regional consistency was also verified. A coordinate–administrative consistency check confirmed, for every non-

imputed coordinate, that it fell within the ISTAT polygon[45] of its declared municipality, and that the declared province matched administratively. Discordant records were resolved by visual inspection and geodesic distance to the candidate polygons, to distinguish genuine data errors from plausible boundary-adjacent trap locations. Genuine errors were corrected at source; boundary-adjacent locations were retained, with the discrepancy flagged. Independently, spatial outliers were flagged using two complementary criteria: excessive positional drift of a given trap from its own historical median location, and excessive distance of an individual record from its municipality's median position, relative to a robust, median-absolute-deviation-based threshold[46]. Every exclusion, correction, and unresolved flag for retained records was written to a record-level annotation field (note).

The merged database was then restricted to confirmed *Cx. pipiens* records within the 2008–2022 study period and assigned to a fixed spatial reference grid. This grid was built once from the ERA5-Land spatial template[12] (~9 km resolution) and published independently of the dataset[14], in two versions: a global grid retaining the unmodified ERA5-Land land–sea mask and cell numbering, and an Italy-specific grid in which coastal cells falling on the seaside of the ERA5-Land mask were gap-filled. Gap-filling used a 3×3 focal-mean window restricted to the immediate, one-ring (~10 km) neighbourhood of valid land cells[44], so that near-shore sampling locations resolved to a valid grid cell without discarding genuinely land-based records. Each cell was additionally attributed with its enclosing country and continent, using the Natural Earth administrative boundary dataset[47]. Occurrence records were assigned a grid-cell identifier by direct extraction from this pre-built Italy grid. Final diagnostic checks verified the uniqueness of the spatio-temporal aggregation key and the completeness of all fields.

Figure 2. Workflow of the data harmonisation and aggregation pipeline. The flowchart illustrates the stepwise processing of raw entomological *Cx. pipiens* surveillance data, from the raw regional inputs to the final published dataset. Raw regional datasets (n = 238,803 trap-night records, 11 institutional sources, 17 Italian regions, 2008–2022) were sequentially processed through regional harmonisation and quality control, national-level re-validation and duplicate detection, geospatial verification against ISTAT administrative boundaries, and spatial outlier screening, yielding a harmonised national dataset of 197,767 records. After restriction to confirmed *Cx. pipiens* records within the 2008–2022 study period, records entering spatio-temporal aggregation (n = 58,888) were summarised per stratum: strata with multiple trap locations (n > 1) were represented by the median mosquito count and interquartile range (IQR), while single-trap strata (n = 1) retained the original observed count (IQR = 0). The final dataset, deposited on Zenodo, comprises 50,759 harmonised records. Sample sizes and exclusion counts for each processing step are reported within the corresponding boxes.


Raw regional datasets
11 sources · 17 regions · 2008–2022
n = 238,803 trap-night records
Regional harmonisation & quality control
Date, coordinate, value and taxonomy validation
Within-region duplicate identification
Excluded: 17.1% (n = 40,766)
Invalid value/date (n = 12) · No sampling event (n = 11,881)
Non-informative zero count (n = 28,434)
Duplicates (n = 436) · Off-target species (n = 3)
Retained: n = 198,037
National re-validation of dates, coordinates
and values – 0.0% failed (n = 0)
National duplicate detection (exact match, all fields)
Candidate groups: n = 270
Collapsed, stable trap ID (n = 270) · Kept with note, no trap ID (n = 2,629)
Municipality/Province verification
(point-in-polygon vs. ISTAT boundaries)
Municipality mismatch: 1.5% (n = 2,851) · Province mismatch: 0.1% (n = 157)
Spatial outlier detection
(trap positional consistency + robust MAD distance)
Flagged: n = 3,254 · Confirmed error: n = 343
Manual inspection: 2.1% queued (n = 4,180)
Resolved with documented correction/note: n = 3,444
Harmonised national dataset: n = 197,767
Georeferenced via centroid: 1.3% (n = 2,611)
Corrected records: 7.5% (n = 14,829)
Species & study-period restriction
Off-target species (n = 130,023) · Outside 2008–2022 (n = 8,856)
Entering aggregation: n = 58,888
Multiple trap
locations
in stratum?
Yes (n > 1)
No (n = 1)
Median
count + IQR
Original count
(IQR = 0)
Published harmonised dataset
25.7% retained (n = 50,759 aggregated records)
Mean n = 1.16 raw records per row – Zenodo

301

Data analysis

The harmonised trap-level records were spatially aggregated onto a ERA5-Land 9 × 9 km reference grid[12] using the *terra* package[44], and temporally by year and sampling week. Given the use of multiple capture devices (e.g., CDC light traps, Gravid traps) with varying capture efficiencies and to account for biological categories, the aggregation was stratified by trap type and mosquito sex. For each grid cell, year and sampling week, we calculated the median count of *Cx. pipiens* individuals independently for each trap category and sex, together with the associated 25th and 75th percentiles (interquartile range, IQR), in accordance with robust outlier-detection frameworks[46]. As a result, a single trap deployment may contribute up to multiple records in the final dataset when female, male and indefinite (sex is NA) counts are available. This approach prevents the conflation of data from different sampling methods and biological categories, while mitigating the impact of local outliers. However, the aggregation does not correct for residual differences in sampling effort, such as the number of traps per cell, deployment duration, or spatial coverage among regions. To support effort-aware analysis, each record includes the number of contributing raw records (n_records_aggregated), as a measure of sampling effort. When multiple records fall within a given grid cell for a particular combination of year, sampling week, trap type and mosquito sex (n_records_aggregated > 1), the reported value is the median of the corresponding trap-night counts. When only a single trap location contributes to the stratum (n_records_aggregated = 1), the reported value corresponds to the original observed count. The complete R code implementing the data-processing and aggregation pipeline, together with synthetic dummy datasets enabling full reproduction of the workflow, is available in a Zenodo repository[14].

Data Records

The harmonised dataset[14] describing *Cx. pipiens* adult trapping records in Italy integrate entomological surveillance data collected by the network of Italian EZIs and cooperating research centres. The dataset is publicly available on Zenodo under the DOI: 10.5281/zenodo.19494972. The data record is provided as a single tabular file in comma-separated values (.csv) format, comprising a total of 50,759 records.

Each row in the harmonised dataset represents a unique combination of grid cell (onto a 9×9 km grid), sampling week, year, trap type, and mosquito sex; consequently, a single trap deployment may generate two rows when both sexes are recorded separately. The core variables included in the dataset provide temporal metadata (year, week, date), spatial and administrative identifiers (centroid grid longitude and latitude, EPSG projection, Region, Province, Municipality), an indicator of spatial precision detailing whether single original traps coordinates before aggregation were derived from municipal centroids (centroid), trapping details (trap_type), full taxonomic and biological classification (from kingdom to species, including sex and life_stage), and the aggregated mosquito median abundance (value). Sex-specific recording practices varied among contributing institutes: some provided separate female and male counts, others reported combined unsexed totals, and others did not record sex at all. In selected cases, recording practices changed over time within the same institution (e.g., Marche and Umbria transitioned from combined to female-only counts in 2022). These categories were retained as provided, without reclassification, and are recorded in the sex variable as "F" (female), "M" (male), "F+M" (combined unsexed count) or NA (not recorded). Additionally, the dataset includes specific administrative fields (Institute, contact_person, contact_person_email) to facilitate formal requests for the underlying raw, disaggregated data. Finally, a dedicated column (note) is included to report any supplementary observations or remarks concerning methodological details of coordinate imputation, spatial corrections or data aggregation. A comprehensive description of all dataset variables, including their definitions and formats, is provided in Table 2.

Table 2. Description of the variables included in the harmonised *Cx. pipiens* dataset. The table outlines the core fields comprising the data record, providing the exact column names, descriptions, and data formats. Variables encompass temporal metadata, spatial and administrative identifiers, taxonomic classification, trap specifications, aggregated mosquito abundance, and institutional contact information for raw data requests.

Variable	Description	Format
cell_id	Unique identifier for the 9x9 km spatial reference grid cell	integre
year	Year of the sampling event	yyyy
week	ISO 8601 week number of the sampling event	integer
value	Median aggregated abundance of the sampled individuals per grid cell, date and trap category	numeric
value_q1	25th percentile(first quartile) of the median	numeric
value_q3	75th percentile /third quartile) of the median	numeric
Country	Country of the sampling site	string
Region	Administrative region of the sampling site	string
longitude	Longitude of the 9x9 km grid cell centroid	numeric
latitude	Latitude of the 9x9 km grid cell centroid	numeric
municipality_centroid	Indicator of wheter spatial disaggregated coordinates were derived from the municipality centroid	string
Institute	Institute responsible for data collection and ownership	string
contact_person	Name of the institutional contact person	string
contact_person_email	Email address of the institutional contact person	string
trap_type	Type of trap used for sampling	string
Canonical_name	Accepted taxonomic name of the sampled taxon	string
kingdom	Taxonomic Kingdom	string
phylum	Taxonomic Phylum	string
class	Taxonomic Class	string
order	Taxonomic Order	string
family	Taxonomic Family	string
genus	Taxonomic Genus	string
species	Taxonomic Species	string
life_stage	Developmental life stage of the sampled individuals	string
sex	Sex of the sampled individuals	string
EPSG	Coordinate Reference System (CRS) designated by its EPSG code	integer
n_records_aggregated	Number of unique sampling location aggregated in the cell	integer
note	Methodological notes regarding coordinate imputation, spatial corrections, or data aggregation	string

Data overview

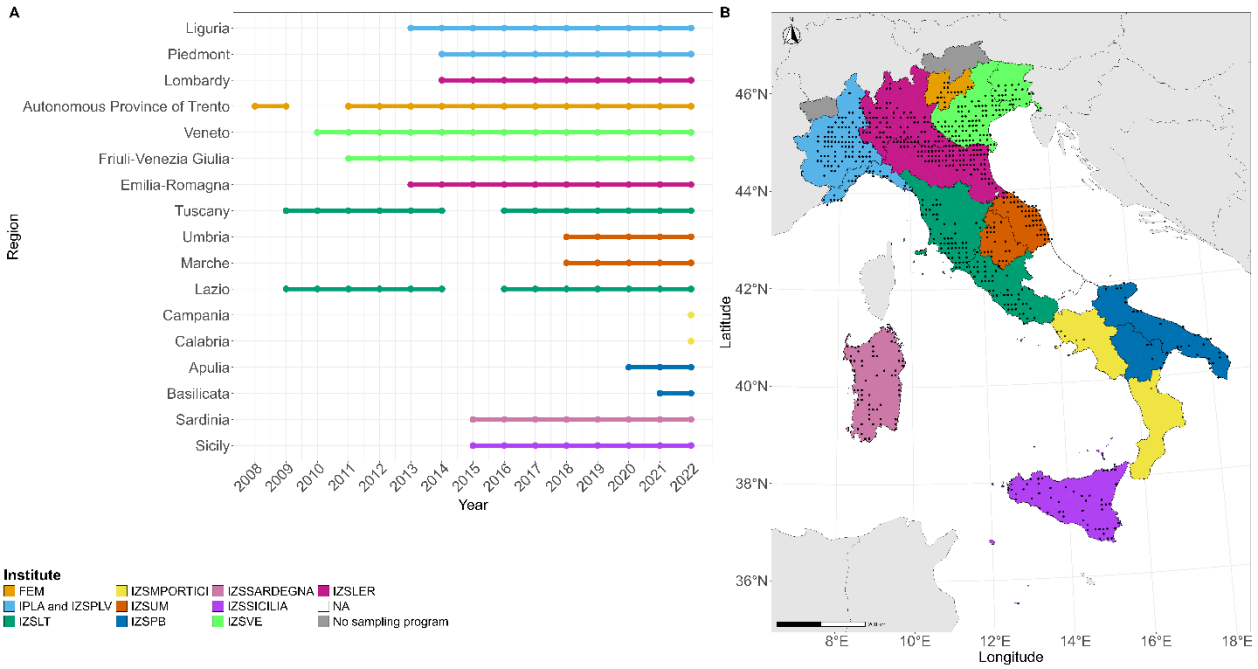
Filtering and quality control processing (Figure 2) began with 238,803 raw trap-night records collected across 11 institutions and 17 Italian regions (2008–2022). Initial regional stage excluded 17.1% of records (n = 40,766), mainly due to non-informative zeros (n = 28,434) and unlinked sampling events (n = 11,881), alongside minor exclusions for duplicates (n = 436), invalid entries (n = 12), and off-target species (non-mosquitoes; n = 3), leaving 198,037 records for national processing. At the national stage, duplicate detection collapsed 270 confirmed re-entries. Geospatial checks against ISTAT boundaries identified 2,851 municipality and 157 province mismatches, while spatial outlier detection flagged 3,254 records (343 confirmed errors). Targeted manual inspection resolved 3,444 of 4,180 queued entries, yielding a harmonised dataframe of 197,767 records (1.3% municipal-centroid imputed; 7.5% carrying documented corrections). Finally, restricting the scope to *Cx. pipiens* within the 2008–2022 period removed 130,023 off-target species (other mosquitoes) entries and 8,856 out-of-period records. This left a final pre-aggregation baseline of 58,888 individual trap-level records of adult *Cx. pipiens* ready for spatio-temporal summarisation. The aggregation yielded 50,759 records, each representing a unique combination of grid cell, sampling week, year, trap type, and mosquito sex (corresponding to 37,070 unique trap deployments; the difference reflects the sex-stratified structure, where a single deployment may contribute up to three records when female, male and indefinite – sex is NA- counts are available).

Overall, the temporal extent of *Cx. pipiens* sampling varies by administrative region and responsible institute, covering a 15-year period from 2008 to 2022 (Figure 3A). Across the four phases of the Italian surveillance system, the harmonised dataset documents a progressive expansion of data availability.

Coverage grew from a single region during the 2008 emergency phase to 7 regions in 2009–2013, 13 in 2014–2019, and 17 in 2020–2022. Correspondingly, unique trap locations expanded from 20 to 204, 822, and 704 across the four phases, distributed over 3, 162, 571, and 539 9 × 9 km grid cells, respectively. Within this overall trajectory, the average monitoring duration per region was 8 (± 4.43) years. While some areas exhibit long-term monitoring spanning up to 14 years (e.g., Autonomous Province of Trento and Veneto region, starting in 2008 and 2010, respectively), others display gaps or discontinuous, shorter sampling windows (e.g., Campania and Calabria regions, where monitoring activities provided only 1 year of data). This reflects the specific surveillance mandates and project-based activities of the 11 contributing institutes. Notably, Valle d'Aosta region and the Autonomous Province of Bolzano are absent from the dataset, as systematic adult *Cx. pipiens* trapping was not implemented during the study period. Spatially, the dataset achieves widespread coverage across the Italian peninsula and islands, encompassing 728 unique 9×9 km grid cells (Figure 3B). The geographic distribution of these cells clearly delineates the territorial competence of the contributing institutes (e.g., IZSLER in Lombardy and Emilia-Romagna, IZSLT in Lazio and Tuscany, and IZSSICILIA and IZSSARDEGNA in the islands).

The intensity of the surveillance effort varied geographically: the highest yearly mean number of sampled weeks was recorded in the islands, specifically in Sardinia (50.90 ± 1.73 weeks) and Sicily (34.20 ± 3.54 weeks) (Table 1). In terms of spatial coverage, highly monitored regions such as Emilia-Romagna, Piedmont, and Veneto maintained the highest yearly mean of sampled cells, with 86.10 (± 2.69), 52.20 (± 11.20), and 43.50 (± 11.40) cells, respectively (Table 1). Furthermore, descriptive statistics of the aggregated records reveal the specific contribution of each trap type to the dataset (Table 1). CO_2-baited traps were the predominantly used collection method, with CDC-CO_2 and CDC-like CO_2 traps accounting for the vast majority of the aggregated records (n = 15,817 and n = 10,364, respectively). This effort was primarily driven by the intense monitoring in northern and central regions. These were followed by unbaited collection methods, including standard CDC (n = 3,905) and CDC-like traps (n = 8), primarily deployed in Sardinia and Sicily. The remaining sampling effort was covered by BG-Sentinel traps, including variants both with and without CO_2 (BG-CO_2, n = 1,267; BG, n = 1,582), Gravid traps (n = 1,343), and aspirators (ASP, n = 16).

Figure 3. Temporal and Spatial Coverage of *Cx. pipiens* Sampling by Italian Experimental Zooprophylactic Institutes (EZIs). (A) Temporal range of *Cx. pipiens* samples collected by Italian EZIs across regions (y-axis) and years (x-axis). Gaps represent years without available data. (B) Spatial distribution of *Cx. pipiens* sampling sites aggregated on a 9×9 km grid. Each point corresponds to the centroid of a sampled grid cell, coloured by the responsible EZI. Grey areas denote regions without a structured regional sampling program during the study period, while white areas (NA) indicate regions where data were not available or not included in this dataset. Institutes: IZSPLV = Experimental Zooprophylactic Institute of Piedmont, Liguria and Aosta Valley; IPLA = Institute for Plants and Environment of Piedmont; IZSLER = Experimental Zooprophylactic Institute of Lombardy and Emilia-Romagna; FEM = Edmund Mach Foundation; IZSVE = Experimental Zooprophylactic Institute of the Venezie; IZSLT = Experimental Zooprophylactic Institute of Lazio and Tuscany; IZSUM = Experimental Zooprophylactic Institute of Umbria and Marche; IZSMPORTICI = Experimental Zooprophylactic Institute of Southern Italy (Portici); IZSPB = Experimental Zooprophylactic Institute of Apulia and Basilicata; IZSSARDEGNA = Experimental Zooprophylactic Institute of Sardinia; IZSSICILIA = Experimental Zooprophylactic Institute of Sicily.



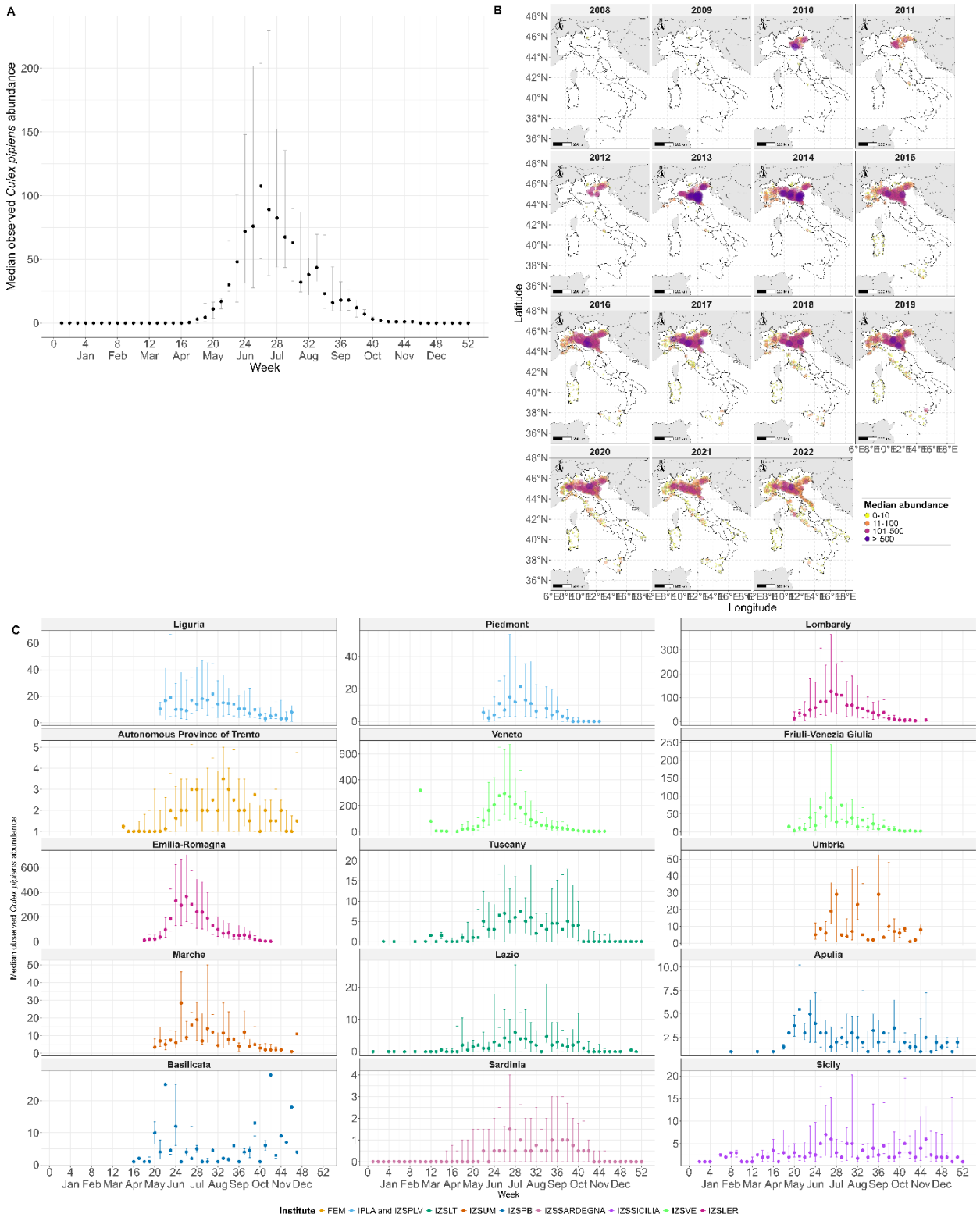
Beyond sampling effort, the contributing institutes also recorded sex-disaggregated abundance data, allowing for female-specific analysis in selected regions and years. Among regions with female-specific records, the highest median counts of female *Cx. pipiens* per trap per week were observed in Emilia-Romagna and Veneto, where values peaked at 240.00 (in 2014; Q1-Q3 = 100.00-473.00, n = 731) and 120.00 (in 2013; Q1-Q3 = 28.20-420.00, n = 426), respectively. High abundances were also recorded in Liguria, Lombardy and Friuli-Venezia Giulia, where values peaked at 31.50 (in 2013; Q1–Q3 = 13.00–48.60, n = 34), at 89.00 (in 2014; Q1-Q3 = 26.00-226.00, n = 250) and at 40.00 (in 2014; Q1-Q3 = 11.00-112.00, n = 111), respectively. Piedmont recorded more moderate female abundances compared to other northern regions, with a maximum median of 8.00 (in 2019; Q1-Q3 = 2.00-29.40, n = 406). Central regions such as Lazio and Tuscany consistently recorded lower female abundances, with medians typically ranging from 0.00 to 3.75 per trap per week across most years, reaching a maximum of 5.00 in Tuscany region (in 2022; Q1-Q3 = 0.00-18.80, n = 322), with only one missing year (2015). Southern regions such as Apulia and Basilicata were characterized by shorter and more recent time series (2020–2022 and 2021–2022, respectively). In Apulia, annual median counts remained at 2.00 throughout the study period (Q1–Q3 up to 1.00–5.00, n = 149 in 2021), whereas in Basilicata, medians ranged from 2.00 to 3.00, peaking in 2022 (Q1–Q3 = 1.00–6.00, n = 45). Similarly, Sardinia provided a consistent time series (2015–2022) in which the weekly median remained at 0.00 in every year, reflecting a high proportion of zero-catch samples; non-zero captures did occur regularly, however, reaching up to 171.50 individuals in a single week (2018, week 25), with the upper quartile peaking in 2018 (Q1–Q3 = 0.00–2.50, n = 476). Sicily also provided female-specific counts across its entire monitored period (2015–2022), with medians ranging from 2.00 to 4.00 across years, reaching the maximum in 2017 (Q1–Q3 = 2.00–9.50, n = 61) and 2020 (Q1–Q3 = 2.00–15.20, n = 104). The Autonomous Province of Trento provides the longest sex-specific time series in the dataset, spanning the entire study period (2008-2022) with only one missing year (2010), and female medians ranging from 1.00 to 3.00 per trap per week. For the remaining regions, some provided data as combined counts of both sexes (females+males) or no data on sex at all (NA). For instance, in Marche and Umbria, that recorded combined values (females+males) until 2021, but started recording only females from 2022, reporting a median abundance of 6.00 (Q1–Q3 = 2.00–16.00, n = 413) in Marche and 5.00 (Q1–Q3 = 2.00–19.00, n = 82) in Umbria.

The analysis of aggregated weekly data reveals a distinct unimodal seasonal pattern for *Cx. pipiens* activity when evaluated nationwide, from the Alps to the Islands (Figure 4A). Crucially, these trends represent the median observed abundance of female mosquitoes, calculated by pooling female-specific records from all deployed trap types (e.g., CDC and CDC-like baited with CO_2, unbaited CDC and CDC-like, BG-Sentinel and BG-Sentinel baited with CO_2, and Gravid traps) into a single aggregated metric. Although this combined approach allows for a comprehensive overview of the national infestation pressure, it is predominantly driven by the CDC-CO2 collections (as detailed in Table 1). Furthermore, while this nationwide aggregation highlights a clear general trend, the data exhibit substantial interannual variability (as indicated by the wide interquartile ranges in Figure 3A), and the overall magnitude is heavily influenced by the massive volume of records originating from the Emilia-Romagna region. Female vector abundance remains negligible during winter and early spring, with detections beginning to rise consistently from mid-spring (around week 18, early May). The median observed abundance escalates rapidly, reaching a pronounced peak between late June (week 26) and late July (week 30), with maximum national median values reaching 83 mosquitoes in week 26. Following this maximum, the median abundance declines steadily throughout late summer and early autumn, approaching zero after week 42 (mid-October).

Spatially, the aggregated map reveals marked heterogeneity in female vector density across the Italian territory (Figure 4B). The spatial median observed abundance ranges from negligible values (<1 mosquito records) in areas with low surveillance yields to significant hotspots. The highest median densities are prominently concentrated in the Emilia-Romagna region, where local median values peak at 881 females *Cx. pipiens* mosquitoes.

Stratifying the seasonal analysis by geographic macro-region (Figure 4C) highlights latitudinal nuances in female vector phenology. For this specific evaluation, data from Campania and Calabria were excluded, as their monitoring activities provided only a single year of sampling data (2022), which is insufficient for robust multi-year seasonal profiling. While the peak median abundance window is largely synchronized across the country, the duration of the activity season varies notably among macro-regions. The islands exhibit the broadest activity window, characterised by an exceptionally early onset (week 8, mid-February) and sustained activity throughout the year (up to week 49, early December). In contrast, the northern and central regions display a more compressed seasonal profile. Specifically, vector activity onset is recorded around week 10 (early March) in the North and week 12 (late March) in the Centre, with the season concluding by week 47 (late November) and week 48 (early December), respectively.

Figure 4. Spatio-temporal dynamics of female *Cx. pipiens* median abundance in sampled areas by Italian Experimental Zooprophylactic Institutes (EZIs) (2008–2022). (A) Overall seasonal profile aggregated across all years and regions. Points represent the median number of observed female mosquitoes per week, and vertical bars show the Interquartile Range (IQR, 25th–75th percentiles). (B) Annual spatial distribution of female vector density aggregated on a 9×9 km grid. Each point represents the median of weekly median counts calculated for that grid cell, with size and color intensity proportional to the observed abundance. (C) Seasonal profile stratified by EZIs, highlighting latitudinal differences in phenology and activity windows. Points represent the median number of observed female mosquitoes per week, and vertical bars show the Interquartile Range (IQR, 25th–75th percentiles). Institutes: IZSPLV = Experimental Zooprophylactic Institute of Piedmont, Liguria and Aosta Valley; IPLA = Institute for Plants and Environment of Piedmont; IZSLER = Experimental Zooprophylactic Institute of Lombardy and Emilia-Romagna; FEM = Edmund Mach Foundation; IZSVE = Experimental Zooprophylactic Institute of the Venezie; IZSLT = Experimental Zooprophylactic Institute of Lazio and Tuscany; IZSUM = Experimental Zooprophylactic Institute of Umbria and Marche; IZSPB = Experimental Zooprophylactic Institute of Apulia and Basilicata; IZSSARDEGNA = Experimental Zooprophylactic Institute of Sardinia; IZSSICILIA = Experimental Zooprophylactic Institute of Sicily.



Technical Validation

Adult *Cx. pipiens* specimens were identified by trained medical/veterinary entomologists within the institutions responsible for routine arbovirus vector surveillance in Italy. These records originate from the official entomological activities that feed national and regional WNV preparedness and response workflows;

as such, they reflect operational surveillance outputs rather than opportunistic research sampling. Identification was performed using established morphological criteria and standard laboratory practices routinely adopted in Italian surveillance services, strictly following the evolving national protocols and guidelines issued throughout the study period[20,21,25-40]. To ensure high data quality, the identification process included internal supervision by senior staff and, where required, the escalation of ambiguous specimens for expert review.

Quality control procedures were applied throughout the sampling and identification workflow. Field collection followed standardized protocols to ensure specimen integrity. Samples intended for virological analysis were maintained under a strict cold chain—transported at +4°C from the field and subsequently preserved at -80°C upon laboratory arrival—to prevent viral degradation. Alternatively, specimens collected exclusively for morphological analysis were preserved in silica gel. Identification was performed using established morphological keys[2,24] and standard laboratory practices, under the coordination of the National Reference Centre for Exotic Diseases (CESME) and regional reference laboratories. This hierarchical supervision ensured consistency in taxonomic assignment across different regional units (as defined in the PNA[11]).

Beyond taxonomic identification, we evaluated the internal consistency and ecological plausibility of the time series during harmonisation and quality control. The resulting trap-night-standardised counts reproduce expected seasonal dynamics for *Cx. pipiens* in Italy: the unimodal abundance peak between late June and late July, with activity onset in mid-spring and cessation by mid-October in northern and central regions, closely matches the phenology reported in multi-year field studies from the Po plain[8] and northwestern Italy[9]. The latitudinal gradient in season length, with broader activity windows in Sardinia and Sicily compared to northern continental regions, is consistent with the known influence of Mediterranean climatic conditions on *Cx. pipiens* population dynamics, and the negligible winter abundance followed by spring emergence aligns with documented diapause dynamics in central Italian populations[10]. This geographically structured variation provides confidence in the ecological plausibility of the harmonised dataset.

We further screened the compiled data for logical inconsistencies (e.g., impossible dates, duplicated records, invalid coordinate formats), and for outliers or discontinuities potentially attributable to data-entry issues (e. g., non-numeric strings recorded in abundance columns, such as qualitative trap malfunction notes) or atypical operational conditions (e.g., records with implausible dates or misformatted coordinates). Records failing basic validity checks were corrected when traceable to source information or flagged/excluded according to the documented quality-control rules detailed in the Methods section. No statistical outlier detection was applied to abundance values; ecologically extreme counts were retained in the dataset.

Collectively, the provenience of the data as official surveillance outputs, the involvement of expert entomologists in identification, and the consistency of observed seasonal patterns provide confidence that the dataset is fit for downstream analyses of abundance, phenology, and spatio-temporal dynamics.

Usage Notes

This dataset provides harmonised adult *Cx. pipiens* observations from 17 Italian administrative regions (2008–2022). The core metric is the weekly median of standardised counts (adults per trap-night), spatially aggregated on a 9-km reference grid and accompanied by sampling metadata to support reuse in reproducible analyses. It is suited to spatio-temporal modelling of adult abundance and phenology (e.g., onset, peak timing, season length), comparative studies of interannual variability and geographic synchrony, and the evaluation of meteorological and environmental associations with vector dynamics. The long-time span and multi-site structure also make the dataset useful for developing, benchmarking, and validating forecasting and early-warning approaches, and for assessing surveillance design and performance (coverage, sampling

intensity, and the consequences of missingness). Finally, the data can support integrative One Health analyses that relate vector abundance patterns to indicators of WNV circulation in vectors and hosts and, where appropriate, to human outcomes.

For reuse, users should note that trap-night standardisation and spatial aggregation improve comparability across sampling intervals and reduce the impact of local outliers. However, remaining heterogeneity in trap type, placement, and operational protocols may still influence observed densities. As the dataset provides aggregated values stratified by trap type, analyses that combine data across different surveillance programmes should consider adjustment for trap typology and/or site- and programme-level effects, rather than interpreting raw differences in magnitude as purely ecological signals. As monitoring objectives and site-selection criteria may have differed among contributing institutions, users are encouraged to assess potential institution-level effects by stratifying analyses using the institute variable. Variability is also reflected in the choice of trapping devices, which have inherently different capture efficiencies for *Cx. pipiens*: CO_2-baited CDC and CDC-like traps (the reference standard for nocturnal host-seeking *Culex*) capture actively host-seeking females, unbaited variants typically yield lower catches; BG-Sentinel traps capture *Cx. pipiens* with lower and more variable efficiency than their *Aedes* target; Gravid traps selectively sample ovipositing females, and aspiration targets resting individuals. These systematic differences mean that trap-type effects should not be ignored in multi-trap analyses. Furthermore, as is common for entomological surveillance, the time series may be irregular, and underlying distributions can be overdispersed. Trap locations represent sentinel surveillance rather than probability-based sampling; therefore, inference should be framed as surveillance-informed patterns at monitored locations. Spatial prediction exercises should report uncertainty and the limits of extrapolation beyond the covariate and geographic space covered by the monitoring network. In addition, repeated observations from the same trap over time introduce spatial and temporal pseudoreplication, which should be accommodated via random effects or equivalent hierarchical modelling structures. Finally, the use of the median as the aggregation function, while robust to sampling artefacts, may attenuate genuine ecological extremes such as localised abundance hotspots within a grid cell. For analyses requiring the full range of trap-level variability, the raw observational data can be requested from the contributing institutions via the contact information provided in the dataset.

Data availability

This dataset[14] is permanently archived on Zenodo v1 release (10.5281/zenodo.19494972).

Code availability

The custom R code used for cleaning, harmonising, and aggregating the raw observational records into the final dataset[14] is available in the Zenodo repository (10.5281/zenodo.19494972). Please note that these scripts are specifically designed to operate on the original, disaggregated data shared by the contributing institutes. As detailed in the Materials and Methods section, these raw records cannot be shared publicly due to confidentiality agreements. Therefore, to allow readers to test the scripts and reproduce the data processing workflow, a synthetic, disaggregated “dummy” dataset for each EZI is provided in the repository in place of the official private raw observational records. Researchers interested in obtaining the raw data to run the scripts must contact the providing institutes directly; the relevant contact information is available in the 'contact_person' variable within the public dataset.

Fundings

This study is supported by the Italian Ministry of University and Research, Piano Nazionale di Ripresa e Resilienza, and EU within the Extended Partnership on Emerging Infectious Diseases project number PE00000007 (One Health Basic and Translational Actions Addressing Unmet Needs on Emerging Infectious Diseases, INF-ACT). This work was also supported by the Italian Ministry of University and Research (MUR) PRIN2020 “Tackling mosquitoes in Italy: from citizen to bench and back” (N. 2020XYBN88).

Bibliography

1. Petersen, L. R., Brault, A. C. & Nasci, R. S. West Nile virus: review of the literature. JAMA 310, 308–315 (2013). https://doi.org/10.1001/jama.2013.8042
2. Becker, N. et al. Mosquitoes and their control 2nd edn. (Springer, 2010). https://doi.org/10.1007/978-3-540-92874-4
3. Rizzoli, A. et al. The challenge of West Nile virus in Europe: knowledge gaps and research priorities. Euro Surveill. 20, 21135 (2015). https://doi.org/10.2807/1560-7917.ES2015.20.20.21135
4. Ashraf, U. et al. Usutu virus: an emerging flavivirus in Europe. Viruses 7, 219-238 (2015). https://doi.org/10.3390/v7010219
5. European Centre for Disease Prevention and Control & European Food Safety Authority. Surveillance of West Nile virus infections in humans and animals in Europe, monthly report – data submitted up to 3 December 2025. ECDC (2025). https://www.ecdc.europa.eu/en/infectious-disease-topics/west-nile-virus-infection/surveillance-and-disease-data/monthly-updates
6. Veo, C. et al. Evolutionary dynamics of the lineage 2 West Nile virus that caused the largest European epidemic: Italy 2011–2018. Viruses 11, 814 (2019). https://doi.org/10.3390/v11090814
7. Monaco, F. et al. Re-emergence of West Nile virus in Italy. Zoonoses Public Health 57, 476–486 (2010). https://doi.org/10.1111/j.1863-2378.2009.01245.x
8. Carrieri, M. et al. Weather factors influencing the population dynamics of Culex pipiens (Diptera: Culicidae) in the Po plain valley, Italy (1997-2011). Environ. Entomol. 43, 482–490 (2014). https://doi.org/10.1603/EN13173
9. Marini, G. et al. The role of climatic and density dependent factors in shaping mosquito population dynamics: the case of *Culex pipiens* in northwestern Italy. PLoS ONE 11, e0154018 (2016). https://doi.org/10.1371/journal.pone.0154018
10. Romiti, F. et al. Characterization of overwintering sites (hibernacula) of the West Nile vector *Culex pipiens* in central Italy. Parasit. Vectors 18, 74 (2025). https://doi.org/10.1186/s13071-025-06710-5
11. Ministry of Health. Piano nazionale di prevenzione, sorveglianza e risposta alle arbovirosi (PNA) 2020–2025. (Direzione Generale della Prevenzione Sanitaria, 2019).
12. Muñoz Sabater, J. ERA5-Land hourly data from 1950 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). https://doi.org/10.24381/cds.e2161bac (2019).
13. Da Re, D. et al. VectAbundance: A spatio-temporal database of Aedes mosquitoes observations. Sci. Data 11, 636 (2024). https://doi.org/10.1038/s41597-024-03482-y
14. Longo, E. et al. A harmonised dataset of adult *Culex pipiens* mosquitoes from 17 administrative Italian regions, 2008–2022. Zenodo (2026). https://doi.org/10.5281/zenodo.19494972
15. Cervellini, M. et al. A grid-based map for the biogeographical regions of Europe. Biodivers. Data J. 8, e53720 (2020). https://doi.org/10.3897/BDJ.8.e53720

16. Cantile, C., Di Guardo, G., Eleni, C. & Arispici, M. Clinical and neuropathological features of West Nile virus equine encephalomyelitis in Italy. Equine Vet. J. 32, 31–35 (2000). https://doi.org/10.2746/042516400777612080
17. Savini, G. et al. Evidence of West Nile virus lineage 2 circulation in northern Italy. Vet. Microbiol. 158, 267–273 (2012). https://doi.org/10.1016/j.vetmic.2012.02.018
18. Rizzo, C. et al. West Nile virus transmission: results from the integrated surveillance system in Italy, 2008 to 2015. Eurosurveillance 21, 30340 (2016). https://doi.org/10.2807/1560-7917.ES.2016.21.37.30340
19. Barzon, L. et al. Early start of seasonal transmission and co-circulation of West Nile virus lineage 2 and a newly introduced lineage 1 strain, northern Italy, June 2022. Eurosurveillance 27, 2200548 (2022). https://doi.org/10.2807/1560-7917.ES.2022.27.29.2200548
20. Ministry of Health. Decree of 12 February 2008. Gazz. Uff. della Repub. Ital. 36 (2008).
21. Experimental Zooprophylactic Institute of Abruzzo and Molise “G. Caporale”. West Nile disease in Italia nel 2008. (IZSAM, 2008).
22. BirdLife International. Important Bird Areas in Europe: priority sites for conservation. BirdLife International https://datazone.birdlife.org/publications/important-bird-areas-in-europe-priority-sites-for-conservation (2000).
23. Ministry of Health. Piano di sorveglianza entomologica (Nota n. 600.6/BT/1279). (Direzione Generale della Sanità Pubblica Veterinaria, 2001).
24. Severini, F., Toma, L., Di Luca, M. & Romi, R. Italian mosquitoes: general information and identification of adults (Diptera, Culicidae). Fragm. Entomol. 41, 213–372 (2009). https://doi.org/10.13133/2284-4880/92
25. Ministry of Health. Decree of 02 October 2009. Gazz. Uff. della Repub. Ital. 229 (2009).
26. Experimental Zooprophylactic Institute of Abruzzo and Molise “G. Caporale”. West Nile disease in Italia nel 2009. (IZSAM, 2009).
27. Ministry of Health. Decree of 29 April 2010. Gazz. Uff. della Repub. Ital. 99 (2010).
28. Ministry of Health. Decree of 10 September 2011. Gazz. Uff. della Repub. Ital. 209 (2011).
29. Ministry of Health. Decree of 10 September 2012. Gazz. Uff. della Repub. Ital. 211 (2012).
30. Ministry of Health. Decree of 10 September 2013. Gazz. Uff. della Repub. Ital. 211 (2013).
31. Experimental Zooprophylactic Institute of Abruzzo and Molise “G. Caporale”. West Nile disease in Italia nel 2010. (IZSAM, 2010).
32. Experimental Zooprophylactic Institute of Abruzzo and Molise “G. Caporale”. West Nile disease in Italia nel 2011. (IZSAM, 2011).
33. Experimental Zooprophylactic Institute of Abruzzo and Molise “G. Caporale”. West Nile disease in Italia nel 2012. (IZSAM, 2012).
34. Experimental Zooprophylactic Institute of Abruzzo and Molise “G. Caporale”. West Nile disease in Italia nel 2013. (IZSAM, 2013).
35. Ministry of Health. Decree of 18 July 2014. Gazz. Uff. della Repub. Ital. 165 (2014).
36. Experimental Zooprophylactic Institute of Abruzzo and Molise “G. Caporale”. West Nile disease in Italia nel 2014. (IZSAM, 2014).
37. Experimental Zooprophylactic Institute of Abruzzo and Molise “G. Caporale”. West Nile disease in Italia nel 2015. (IZSAM, 2015).
38. Experimental Zooprophylactic Institute of Abruzzo and Molise “G. Caporale”. West Nile disease in Italia nel 2016. (IZSAM, 2016).
39. Ministry of Health. Piano nazionale integrato di sorveglianza e risposta al virus della West Nile - 2016. (Direzione Generale della Prevenzione Sanitaria & Direzione Generale della Sanità Animale e dei Farmaci Veterinari, 2016).
40. Ministry of Health. Piano nazionale integrato di sorveglianza e risposta ai virus West Nile e Usutu - 2017. (Direzione Generale della Prevenzione Sanitaria, 2017).

41. R Core Team. R: A Language and Environment for Statistical Computing. (R Foundation for Statistical Computing, 2025). https://www.R-project.org/
42. Wickham, H. et al. Welcome to the Tidyverse. J. Open Source Softw. 4, 1686 (2019). https://doi.org/10.21105/joss.01686
43. Pebesma, E. Simple features for R: standardized support for spatial vector data. R J. 10, 439–446 (2018). https://doi.org/10.32614/RJ-2018-009
44. Hijmans, R. J. terra: Spatial data analysis (R package version 1.8-93) (2023). https://CRAN.R-project.org/package=terra
45. Istituto Nazionale di Statistica (ISTAT). Confini delle unità amministrative a fini statistici al 1° gennaio 2018. ISTAT (2018). https://www.istat.it/notizia/confini-delle-unita-amministrative-a-fini-statistici-al-1-gennaio-2018-2/
46. Leys, C., Ley, C., Klein, O., Bernard, P. & Licata, L. Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. J. Exp. Soc. Psychol. 49, 764–766 (2013). https://doi.org/10.1016/j.jesp.2013.03.013
47. Natural Earth. Admin 0 – Countries and Admin 0 – Continents (1:10m Cultural Vectors). (2024). https://www.naturalearthdata.com