GuardIAS

GUARDING EUROPEAN WATERS
FROM INVASIVE ALIEN SPECIES

Deliverable

D2.5 Workflow for detecting and prioritising emerging alien species based on GRIIS checklists and species occurrence cubes

Planned delivery date: 31/05/2026 (M17)
Actual submission date: 31/05/2026
Deliverable leader: EV INBO
Authors: Damiano Oldoni (EV INBO), Diederik Strubbe (EV INBO) & Tim Adriaens (EV INBO)
Internal Reviewers: Agnese Marchini (UNIPV), Stelios Katsanevakis (UAEGEAN)
Type of deliverable: Data
Work Package: WP2 - Early detection and monitoring
Work package Leader: Mark Costello (NORD)
Version: 1.1
Pages: 42

GA No. 101181413
Start date of the project: January 1st, 2025
www.guardias.eu

Funded by the European Union

GuardIAS receives funding from the European Union's Horizon Europe Research and Innovation Programme (ID No 101181413). Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the EU nor REA can be held responsible for them.

DISCLAIMER

This document contains information on GuardIAS core activities. Any reference to content in this document should clearly indicate the authors, source, organisation, and publication date. The authors of this document have taken any available measure in order for its content to be accurate, consistent and lawful. However, neither the project consortium as a whole nor the individual partners that implicitly or explicitly participated in the creation and publication of this document hold any sort of responsibility that might occur as a result of using its content.

COPYRIGHT NOTICE



This work by Parties of the GuardIAS Consortium is licensed under a Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/). “

‘GuardIAS receives funding from the European Union’s Horizon Europe Research and Innovation Programme (ID No 101181413). Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the EU nor REA can be held responsible for them.’

TABLE OF CONTENTS

Executive Summary 7

Acronyms & Abbreviations 8

1 Introduction 9

2 Methods 10

2.1 Create species occurrence cubes 11

2.1.1 Spatial dimension 11

2.1.2 Temporal dimension 11

2.1.3 Taxonomic dimension 12

2.1.4 Data quality filters 12

2.2 Detect appearing and reappearing alien species 12

2.3 Perform emerging trends analysis 12

2.3.1 Generalized Additive Model 13

2.3.2 Decision rules 16

2.4 Rank species by total emerging score 17

2.5 Automation 18

3 Results 19

3.1 Mediterranean Sea 20

3.1.1 Appearing species 20

3.1.2 Reappearing species 21

3.1.3 Most emerging species 24

3.2 North Sea 27

3.2.1 Appearing species 27

3.2.2 Reappearing species 28

3.2.3 Most emerging species 29

3.3 Baltic Sea 30

3.3.1 Appearing species 30

3.3.2 Reappearing species 31

3.3.3 Most emerging species 31
3.4 Marine species on the List of Invasive Alien Species of Union Concern 32
3.4.1 *Rugulopteryx okamurae* - Mediterranean Sea 33
3.4.2 *Rugulopteryx okamurae* - Iberian Coastal 34
3.4.3 *Plotosus lineatus* - Mediterranean Sea 35
3.4.4 *Eriocheir sinensis* - Iberian Coastal 36
3.4.5 *Mulinia lateralis* - Celtic-Biscay Shelf 37
4 Further Developments 38
5 Concluding remarks 38
6 Bibliography 41

LIST OF TABLES AND FIGURES

TABLE 1. TABLE OF ABBREVIATIONS AND ACRONYMS USED IN THE DOCUMENT

TABLE 2. EMERGING STATUSES AND SCORES USED IN THE EMERGING TRENDS ANALYSIS.

TABLE 3. RULES USED TO DETERMINE THE EMERGING STATUS BASED ON THE SIGN OF THE UPPER CONFIDENCE LIMIT (UCL) AND LOWER CONFIDENCE LIMIT (LCL) OF THE FIRST (1) AND SECOND (2) DERIVATIVES OF THE GAM FIT. THE TABLE SHOWS THE COMBINATIONS OF SIGNS (+ FOR POSITIVE, - FOR NEGATIVE) FOR THE DERIVATIVES AND THE RESULTING EMERGING STATUS.

TABLE 4. LIST OF DATASETS CONTRIBUTING WITH MORE THAN 100 THOUSANDS RECORDS TO THE GBIF OCCURRENCE CUBE, DOI: 10.15468/DL.RFN2YM.

TABLE 5. DECISION RULES FOR DETERMINING EMERGING STATUS AND SCORE BASED ON FOUR CRITERIA APPLIED TO A RESPONSE VARIABLE.

TABLE 6. WEIGHTS USED TO CALCULATE THE TOTAL EMERGING SCORE BASED ON THE YEAR AND THE RESPONSE VARIABLE.

TABLE 7. APPEARING SPECIES IN THE MEDITERRANEAN SEA, LISTED BY THE YEAR OF FIRST OBSERVATION.

TABLE 8. REAPPEARING SPECIES IN THE MEDITERRANEAN SEA, LISTED BY THE YEAR OF REAPPEARANCE.

TABLE 9. MOST EMERGING SPECIES IN THE MEDITERRANEAN SEA.

TABLE 10. APPEARING SPECIES IN THE NORTH SEA, LISTED BY THE YEAR OF FIRST OBSERVATION.

TABLE 11. REAPPEARING SPECIES IN THE NORTH SEA, LISTED BY THE YEAR OF REAPPEARANCE.

TABLE 12. MOST EMERGING SPECIES IN THE NORTH SEA.

TABLE 13. APPEARING SPECIES IN THE BALTIC SEA, LISTED BY THE YEAR OF FIRST OBSERVATION.

TABLE 14. REAPPEARING SPECIES IN THE BALTIC SEA, LISTED BY THE YEAR OF REAPPEARANCE.

TABLE 15. MOST EMERGING SPECIES IN THE BALTIC SEA.

TABLE 16. MARINE SPECIES ON THE LIST OF INVASIVE ALIEN SPECIES OF UNION CONCERN.

TABLE 17. SUMMARY OF CORRECTIONS AND ANNOTATIONS FLAGGED BY MARCHINI AGNESE DURING EXPERT REVIEW OF THE WP2 D2.5 DELIVERABLE. ORIGINAL ENTRIES REFLECT WHAT THE WORKFLOW DETECTED BASED ON GBIF DATA; CORRECTIONS INDICATE DISCREPANCIES IDENTIFIED AGAINST THE LITERATURE.

FIGURE 1. WORKFLOW FOR DETECTING AND PRIORITISING EMERGING ALIEN SPECIES. THE CIRCULAR ARROWS INDICATE THE PROCESS IS REITERATED OVER TIME. THE APPEARING AND REAPPEARING SPECIES AND THE RANKED EMERGING SPECIES ARE SHOWN ONLINE IN A DASHBOARD.

FIGURE 2. BLACK DOTS: THE NUMBER OF GRID CELLS (10KMX10KM) OCCUPIED BY RUGULOPTERYX OKAMURAE IN THE MEDITERRANEAN SEA PER YEAR. COLORED DOTS: EMERGING STATUS AS CALCULATED BASED ON THE FIRST AND SECOND DERIVATIVES OF THE UPPER AND LOWER CONFIDENCE LEVELS OF THE GAM FIT, SHOWN AS THE GREY BOUNDS.

FIGURE 3. BLACK DOTS: THE NUMBER OF OCCURRENCES OF RUGULOPTERYX OKAMURAE IN THE MEDITERRANEAN SEA PER YEAR. COLORED DOTS: EMERGING STATUS AS CALCULATED BASED ON THE FIRST AND SECOND DERIVATIVES OF THE UPPER AND LOWER CONFIDENCE LEVELS OF THE GAM FIT, SHOWN AS THE GREY BOUNDS.

FIGURE 4. BLACK DOTS: THE NUMBER OF GRID CELLS (10KMX10KM) OCCUPIED BY RUGULOPTERYX OKAMURAE IN THE IBERIAN COASTAL PER YEAR. COLORED DOTS: EMERGING STATUS AS CALCULATED BASED ON THE FIRST AND SECOND DERIVATIVES OF THE UPPER AND LOWER CONFIDENCE LEVELS OF THE GAM FIT, SHOWN AS THE GREY BOUNDS.

FIGURE 5. BLACK DOTS: THE NUMBER OF OCCURRENCES OF RUGULOPTERYX OKAMURAE IN THE IBERIAN COASTAL PER YEAR. COLORED DOTS: EMERGING STATUS AS CALCULATED BASED ON THE FIRST AND SECOND DERIVATIVES OF THE UPPER AND LOWER CONFIDENCE LEVELS OF THE GAM FIT, SHOWN AS THE GREY BOUNDS.

FIGURE 6. BLACK DOTS: THE NUMBER OF GRID CELLS (10KMX10KM) OCCUPIED BY PLOTOSUS LINEATUS IN THE MEDITERRANEAN SEA PER YEAR. COLORED DOTS: EMERGING STATUS AS CALCULATED BASED ON THE FIRST AND SECOND DERIVATIVES OF THE UPPER AND LOWER CONFIDENCE LEVELS OF THE GAM FIT, SHOWN AS THE GREY BOUNDS.

FIGURE 7. BLACK DOTS: THE NUMBER OF OCCURRENCES OF PLOTOSUS LINEATUS IN THE MEDITERRANEAN SEA PER YEAR. COLORED DOTS: EMERGING STATUS AS CALCULATED BASED ON THE FIRST AND SECOND DERIVATIVES OF THE UPPER AND LOWER CONFIDENCE LEVELS OF THE GAM FIT, SHOWN AS THE GREY BOUNDS.

FIGURE 8. THE NUMBER OF GRID CELLS (10KMX10KM) OCCUPIED BY ERIOCHEIR SINENSIS IN THE IBERIAN COASTAL PER YEAR.

FIGURE 9. BLACK DOTS: THE NUMBER OF OCCURRENCES OF ERIOCHEIR SINENSIS IN THE IBERIAN COASTAL PER YEAR.

FIGURE 10. THE NUMBER OF GRID CELLS (10KMX10KM) OCCUPIED BY MULINIA LATERALIS IN THE CELTIC-BISCAY SHELF PER YEAR.

FIGURE 11. BLACK DOTS: THE NUMBER OF OCCURRENCES OF MULINIA LATERALIS IN THE CELTIC-BISCAY SHELF PER YEAR.

Executive Summary

This report details a reproducible, data-driven workflow for detecting and prioritizing emerging invasive alien species (IAS) in European Large Marine Ecosystems (LMEs). The workflow integrates openly available data from biodiversity repositories such as the Global Biodiversity Information Facility (GBIF) and the Catalogue of the European Alien Species Information Network (EASIN), aggregated into occurrence cubes (i.e. a three-dimensional data structure spanning taxonomic, temporal, and geographic dimensions). The fully automated, self-updating pipeline flags two types of early-warning signals: *appearing and reappearing species* (i.e. alien species newly detected, or redetected after a period of absence, in a given sea region), and *emerging species* (i.e. established alien species showing a rapid and statistically significant increase in their distribution and/or abundance). Together, these outputs generate prioritized watchlists to support early warning and rapid response.

Results for the Mediterranean, North, and Baltic Seas provide illustrative prioritized lists of species based on an emerging score that weights recent trends and occupancy increases. Notably, multiple species reached the maximum emerging score during the 2022–2024 assessment period. While current data representations for marine environments are less comprehensive than for terrestrial ones, the framework is designed to automatically benefit from ongoing data mobilization efforts. Future developments will focus on improved accounting for variation in survey efforts, expanding spatial and taxonomic scope and increasing the spatial and taxonomic granularity to better support evidence-based risk assessment and management decisions across Europe.

Acronyms & Abbreviations

Table 1. Table of abbreviations and acronyms used in the document

Acronym / Abbreviation	Signification
EASIN	European Alien Species Information Network
EEA	European Environment Agency
GAM	Generalized Additive Model
GBIF	Global Biodiversity Information Facility
GRIIS	Global Register of Introduced and Invasive Species
GuardIAS	Guarding European Waters from Invasive Alien Species
IAS	Invasive Alien Species
LME	Large Marine Ecosystem
MIT	Massachusetts Institute of Technology
OBIS	Ocean Biodiversity Information System
TrIAS	Tracking Invasive Alien Species

1 Introduction

Invasive alien species (IAS) are among the principal drivers of biodiversity loss worldwide, and their early detection is a prerequisite for effective management and policy response. Meeting this challenge requires rapid, repeatable, and scalable monitoring workflows that draw on openly available biodiversity data. Large-scale data infrastructures such as the Global Biodiversity Information Facility (GBIF), the European Alien Species Information Network (EASIN), and the Ocean Biodiversity Information System (OBIS) now make such workflows increasingly achievable, integrating records from structured surveys, monitoring campaigns, and citizen science contributions alike (Groom et al. 2015; Adriaens et al. 2015).

This report and deliverable builds directly on the Tracking Invasive Alien Species (TrIAS) project (Vanderhoeven et al. 2017; Groom et al. 2018), which pioneered a data-driven, open-science framework for generating IAS indicators in Belgium. Central to TrIAS was the concept of the *species occurrence cube* (Oldoni et al. 2020), a three-dimensional aggregation of occurrence data across taxonomic, temporal, and geographic dimensions that accounts for the spatial uncertainty inherent in openly published records. Occurrence cubes have since been produced for Belgium and other European countries (Oldoni et al. 2024a, 2024b) and applied to generate indicators for both national and EU-level reporting. This occurrence cube-based workflow was applied to Belgium and its three regions; see the "Occurrence-based indicators" pipelines at https://trias-project.github.io/indicators (GitHub repository: trias-project/indicators), with results also published on Zenodo (Oldoni et al. 2024c, 2024d).

The workflow described here extends that experience in two important directions: it broadens the geographic scope to a European marine context, and for the first time fully automates all analytical steps using GitHub Actions, with outputs disseminated via a bespoke dashboard hosted on GitHub Pages, ensuring rapid and open access to the findings.

The workflow is built on occurrence data openly published on GBIF. Since OBIS (the dedicated global repository for marine biodiversity data) has been a contributing member of GBIF since 2001 and the two infrastructures operate a joint data-sharing strategy, the vast majority of OBIS-mediated marine records are accessible through GBIF and are therefore included in this workflow. This means the analysis effectively draws on the combined scope of both platforms for marine occurrence data.

Applying this workflow to marine IAS responds to a well-documented data and policy gap. While terrestrial IAS monitoring has benefited from decades of systematic recording and standardized data flows, marine environments remain underrepresented in open biodiversity databases. While closing this data gap ultimately depends on greater mobilization of marine occurrence data to open platforms, the occurrence cube framework is designed to make maximal use of all records currently available, drawing on GBIF regardless of whether they originate from citizen science platforms, professional monitoring datasets, or dedicated IAS surveillance programs.

From the occurrence cube, the workflow derives two categories of indicators. First, *appearing and reappearing species*, i.e. alien species newly detected or redetected in the study area, are identified through comparison with a baseline checklist, matching taxa from the EASIN Catalogue with the GBIF Backbone Taxonomy (GBIF Secretariat 2023) via the easin-gbif-taxa-matcher tool. Second, *emerging species* are defined as IAS showing a statistically significant increase in measured occupancy or observation counts in recent years, identified through Generalized Additive Modelling (GAM) and simple decision rules applied to the annual time series extracted from the cube. A ranking procedure then synthesizes these signals into a prioritized list of species to inform risk assessment and management decisions. As with any workflow relying on opportunistic occurrence data, trends derived from these records should be interpreted with appropriate caution, as changes in recording effort over time can influence the detected signals. Statistical corrections for this effect are possible and are planned for future versions of the workflow.

All data and code are openly available under an MIT License through the guardias-eu GitHub organization, encompassing tools for taxonomic matching (easin-gbif-taxa-matcher), cube construction (build-eu-cube), trend analysis (emtrends), and results dissemination via a dashboard (indicators). By adhering to open science principles, the workflow is freely adaptable to other countries, regions, or taxonomic scopes, and can be updated seamlessly as new checklist information or occurrence datasets are published on GBIF.

2 Methods

The proposed workflow to detect and prioritise emerging alien species can be divided into 4 steps:

1) Create species occurrence cubes
2) Detect appearing and reappearing IAS for each Large Marine Ecosystem (LME)
3) Perform emerging trends analysis on IAS for each LME
4) Calculate an emerging score at species level based on the emerging trends analysis for each LME and rank the IAS by this score.

This workflow does not take into account management feasibility aspects, but is conceived as a pure scientific, data driven tool to help detect (re)appearing IAS and rank those spreading or increasing in abundance the fastest.

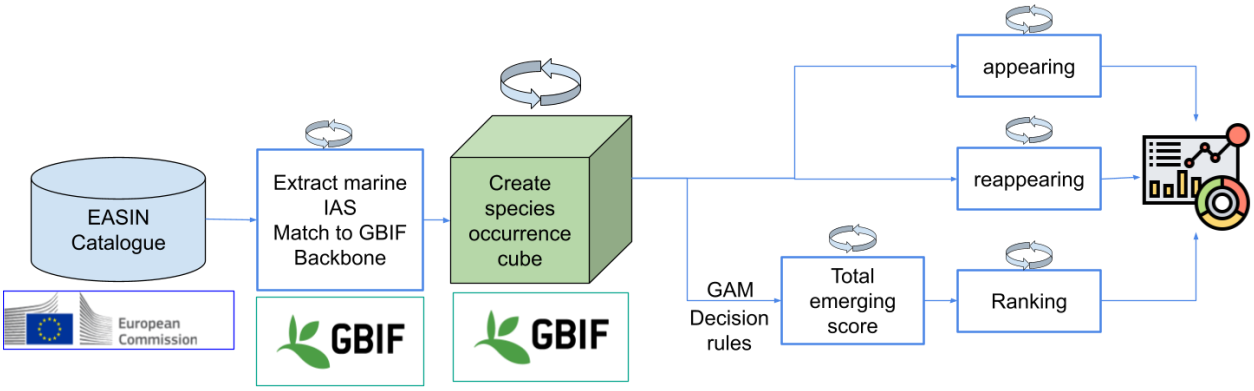


Figure 1. Workflow for detecting and prioritising emerging alien species. The circular arrows indicate the process is reiterated over time. The appearing and reappearing species and the ranked emerging species are shown online in a dashboard.

2.1 Create species occurrence cubes

The 'cube' approach was integrated in March 2025 within the GBIF platform as a data product directly accessible (see **species occurrence cube**). These species occurrence cubes allow to aggregate occurrence data along spatial, temporal and taxonomic dimensions.

Details on each of the three dimensions are provided in Sections 2.1.1–2.1.3. The code and the data are openly available in the GitHub repository build-eu-cube.

The generation of cubes relies on the GBIF SQL download API. The creation of a GBIF species occurrence cube is triggered using the R package rgbif (Chamberlain et al., 2026, Chamberlain, 2017), in particular the function occ_download_sql(). More details about how to use rgbif as an interface to the SQL download API can be found in the package vignette "GBIF SQL Downloads". The SQL query passed to GBIF and other details can be found in build-eu-cube/src/build_run_query.R.

2.1.1 Spatial dimension

We considered a bounding box around the European LMEs (saved in build-eu-cube/data/output/lme_eu_borders_bbox.txt).

Spatial granularity: 10 km x 10 km, corresponding to the EEA reference grid at 10 km resolution.

2.1.2 Temporal dimension

We limited the time range of our analysis to IAS occurrences from 1950 onwards, aggregating them at the year level.

2.1.3 Taxonomic dimension

We selected taxa living in marine and oligohaline environments from the EASIN Catalogue via the R package reasin (Oldoni, 2026). The match of the entire EASIN Catalogue to the GBIF Backbone is performed in the GitHub repository easin-gbif-taxa-matcher. At this moment we include only taxa with an **exact match** to **accepted species** (taxonKey = speciesKey) and taxa with an **exact match** to **synonyms** whose **accepted taxa** are **species** (acceptedKey = speciesKey). The occurrences are aggregated at species level (speciesKey).

2.1.4 Data quality filters

The advantage of using the GBIF SQL download API is its versatility. For example, quality controls can be added to the SQL query to exclude occurrences with spatial, temporal or identification-verification issues. See src/build_run_query.R#L61-L80 for details.

2.2 Detect appearing and reappearing alien species

The first step is the creation and subsequent download of the most recent species occurrence cube, see emtrends/src/calculate_indicators.R#L19-L40. A lookup table has been created to know which grid cells intersect each of the LMEs and it is saved as a csv file in emtrends/data/output/grid_cells_with_lme_info.csv. The script to generate it is stored in assign_grid_cells_to_lme.R.

For each LME, a list of **appearing** species is created by finding which species appear for the first time in grid cells intersecting the given LME in **2022** or later. Data from 2022 onwards are considered, reflecting the focus of this analysis on recent emerging trends (see Section 2.4 for more details on the choice of this time window). The list is stored as a csv file in emtrends/data/output/appearing_species.csv.

Similarly, a list of **reappearing** species is created by finding which species reappear in 2022 or later after being undetected for at least **5 years**. The list is stored as a csv file in emtrends/data/output/reappearing_species.csv.

The lists are created in the R script emtrends/src/calculate_indicators.R#L70-L313.

2.3 Perform emerging trends analysis

Emerging trends analysis is performed at LME level for species not defined as appearing or reappearing (as newly recorded species lack the observation history needed to fit a GAM or apply decision rules) and it uses functions from trias (Oldoni et al. 2026) R package. This analysis returns an **emerging status** and an **emerging score** for each year (Table 2) and is shown in the plots as dots color.

Table 2. Emerging statuses and scores used in the emerging trends analysis.

Status	Score	Color
Not emerging	0	black
Unclear	1	gray
Potentially emerging	2	orange
Emerging	3	red

The emerging trend analysis is implemented in the R script emtrends/src/calculate_indicators.R#L315-L508.

2.3.1 Generalized Additive Model

The emerging trends analysis is performed for each species applying a **Generalized Additive Model** (**GAM**) to two complementary response variables:

1. The **number of occurrences** per year
2. The **number of occupied grid cells** (measured occupancy) per year

The GAM fit is performed using the function apply_gam() from the trias (Oldoni et al. 2026) R package. The GAM is fitted separately for each response variable, producing a smooth curve that describes how the variable changes over time. Rather than assuming a fixed linear trend, the GAM flexibly captures the shape of the time series, e.g. whether a species is accelerating, stabilising, or declining in recent years. More details about the fitting in the Details section of the function documentation. Then, to assess whether an observed trend is statistically meaningful, the first and second derivatives of the fitted curve are computed. A positive first derivative means the number of occurrences/grid cells over the years is increasing, while a negative first derivative means the number of occurrences/grid cells over the years is decreasing. The second derivative is related to the rate of such increase or decrease: a positive second derivative means that the increase/decrease is speeding up, while a negative second derivative means that the increase/decrease is speeding down. The sign of the first and second derivative of the upper and lower confidence interval is then used to define the emerging status of a species at a specific year (Table 3).

Table 3. Rules used to determine the emerging status based on the sign of the upper confidence limit (ucl) and lower confidence limit (lcl) of the first (1) and second (2) derivatives of the GAM fit. The table shows the combinations of signs (+ for positive, - for negative) for the derivatives and the resulting emerging status.

ucl1	lcl1	ucl2	lcl2	Emerging status
+	+	+	+	emerging
+	+	+	-	emerging
+	+	-	-	potentially emerging

-	+	+	+	potentially emerging
+	-	+	-	unclear
+	-	-	-	not emerging
-	-	+	+	not emerging
-	-	+	-	not emerging
-	-	-	-	not emerging

GBIF occurrence data come from a wide variety of sources. The GBIF occurrence cube with download key 0010225-260507073636908 (DOI 10.15468/dl.rfn2ym, GBIF.org) used in this report includes occurrence data from 4995 datasets: see Table 4 for a list with the datasets contributing with more than 100 thousand records. The full list of datasets is available on 10.15468/dl.rfn2ym. It is important to note that the total number of available records increase over time, not necessarily because species are becoming more abundant, but also because observation and reporting effort continues to grow. If left uncorrected, this increase in recording effort can generate a spurious signal of population growth or range expansion. A standard approach to account for this bias is to include the total number of occurrences of a higher taxonomic group as a covariate in the GAM. The rationale is straightforward: if a focal species is increasing at the same rate as all species within its class, the observed trend is likely driven by increasing recording effort rather than by a genuine biological change. In contrast, if the focal species is increasing faster than the class as a whole, the trend is more likely to reflect a real increase in abundance or distribution. Implementation and evaluation of this correction are planned as part of future developments of the workflow. Because the current version does not yet include this adjustment, some species may be classified as emerging simply as a consequence of increased recording effort. Consequently, the current results should be interpreted cautiously and in a precautionary manner: species identified as emerging may not necessarily be undergoing a true expansion, whereas species not identified as emerging are less likely to represent false negatives. Accordingly, the emerging status generated by the workflow should be interpreted primarily as a prioritisation signal for expert review and further investigation. The workflow is designed to minimise the risk of overlooking potentially emerging species, accepting that some false positives may occur until recording-effort corrections are implemented.

Table 4. List of datasets contributing with more than 100 thousands records to the GBIF occurrence cube, DOI: 10.15468/dl.rfn2ym.

Dataset DOI	Dataset key	Dataset title	Dataset Citation	Number of records
10.15468/5nilie	8a863029-f435-446a-821e-275f4f641165	Observation.org, Nature data from around the World		252906
10.15468/3fl71g	724ec4af-c9a3-490b-b718-a1b202437c0f	Données piscicoles et astacicoles des cours d'eau de France métropolitaine bancarisées dans ASPE (OFB) - Données acquises selon le protocole de pêche complète à un ou plusieurs passages - ASPE - OFB	Inventaire National du Patrimoine Naturel (2021). Données piscicoles et astacicoles des cours d'eau de France métropolitaine bancarisées dans ASPE (OFB) - Données acquises selon le protocole de pêche complète à un ou plusieurs passages - ASPE - OFB. UAR PatriNat (OFB-MNHN-CNRS-IRD), Paris. Occurrence dataset https://doi.org/10.15468/3fl71g accessed via GBIF.org on 2025-11-18.	215478
10.15468/mtvg2m	9ba38fd3-e318-4623-bb86-2c5b5131ce9d	ICES Phytoplankton Community dataset	ICES Environmental Database (DOME), Phytoplankton community. Available online at http://dome.ices.dk. ICES, Copenhagen. Consulted on yyyy-mm-dd.	196498
10.15468/ab3s5x	50c9509d-22c7-4a22-a47d-8c48425ef4a7	iNaturalist Research-grade Observations		192444
10.15468/g6fwdc	60b641f5-ad90-48ed-9070-f716a277be8f	LajiGIS: Species surveys and monitoring at sea		178283
10.82144/5f44abaf	898b6c44-2891-45da-84e5-3d8bf6f5d38d	Grey seals GPS telemetry in Molene Archipelago	Vincent, C. 2019. Grey seals GPS telemetry in Molene Archipelago. Version 1.0.0. Dataset published in OBIS-SEAMAP. https://doi.org/10.82144/5f44abaf. accessed via GBIF.org on 2025-12-17.	138872
10.15468/9meff0	5f501df0-0c61-11dd-84ce-b8a03c50a862	SLU Aqua Institute of Freshwater Research National register of survey test-fishing - NORS	National Register of Survey test-fishing - NORS. yyyy. Swedish University of Agricultural Sciences, Department of Aquatic Resources. http://www.slu.se/sjoprovfiskedatabasen [yyyy-mm-dd].	138305
10.14284/431	47360f99-2f92-4e48-9898-1e4976d09d71	2013_ALBERTKANAAL - Acoustic telemetry data for European eel (Anguilla anguilla) and hatched Salmon (Salmo salar) in the Albert canal (Belgium)	Baeyens R, Vergeynst J, De Maerteleire N, Desmet P, Gelaude E, Milotic T, Mouton A, Pauwels I, Pieters S, Reyserhove L, Robberecht K, Verhelst P, Coeck J (2025). 2013_ALBERTKANAAL - Acoustic telemetry data for European eel (Anguilla anguilla) and hatched Salmon (Salmo salar) in the Albert canal (Belgium). Version 1.4.	134526

Dataset DOI	Dataset key	Dataset title	Dataset Citation	Number of records
			Research Institute for Nature and Forest (INBO). Occurrence dataset. https://doi.org/10.14284/431 accessed via GBIF.org on 2025-08-20. accessed via GBIF.org on 2025-09-30. accessed via GBIF.org on 2025-12-16.	
10.14284/428	0eab323a-364d-4a1b-983b-62b8098845b0	2012_LEOPOLDKANAAL - Acoustic telemetry data for European eel (Anguilla anguilla) in a polder area in Flanders (Belgium)	Verhelst P, Buysse D, Baeyens R, De Maerteleire N, Desmet P, Gelaude E, Jacobs Y, Milotic T, Mouton A, Reyserhove L, Robberechts K, Coeck J (2024). 2012_LEOPOLDKANAAL - Acoustic telemetry data for European eel (Anguilla anguilla) in a polder area in Flanders (Belgium). Version 1.5. Research Institute for Nature and Forest (INBO). Occurrence dataset. https://doi.org/10.14284/428 accessed via GBIF.org on 2025-08-20. accessed via GBIF.org on 2025-09-30. accessed via GBIF.org on 2025-12-16.	114037
10.15468/pdop0k	d03f444a-7192-4fa0-bf7a-63714585c611	England Non Native Species records 1965 to 2017	Environment Agency ([Insert download year]). England Non Native Species records 1965 to 2017. Occurrence dataset on the NBN Atlas	111648
10.15468/djr9yv	a24166fe-8c4b-4cf6-8653-646742dcf1b6	Danish benthic marine monitoring data from ODAM		106716
10.15468/psbyfu	2af0d55b-a0fc-475c-a32f-d68282b6cc67	ICES Zoobenthos Community dataset	ICES Environmental Database (DOME), Zoobenthos community. Available online at http://dome.ices.dk. ICES, Copenhagen. Consulted on yyyy-mm-dd.	104085
10.15468/q1g8w7	0b5f972e-0f84-4d07-9b84-96c09e2925da	Données piscicoles et astacicoles des cours d'eau de France métropolitaine bancarisées dans ASPE (OFB) - Données acquises selon le protocole de pêche partielle par points - ASPE - OFB	Inventaire National du Patrimoine Naturel (2021). Données piscicoles et astacicoles des cours d'eau de France métropolitaine bancarisées dans ASPE (OFB) - Données acquises selon le protocole de pêche partielle par points - ASPE - OFB. UAR PatriNat (OFB-MNHN-CNRS-IRD), Paris. Occurrence dataset https://doi.org/10.15468/q1g8w7 accessed via GBIF.org on 2025-11-17.	103969

2.3.2 Decision rules

It can happen that the GAM fit cannot statistically converge and so cannot be used. In this case, a set of deterministic **decision rules** are applied to define the emerging status at year level for the variables 'number of occurrences' and 'number of occupied grid cells'. The function apply_decision_rules() from the trias (Oldoni et al. 2026) R package is used.

Four criteria are applied point-wise for each year:

1. Does the response variable contain only one positive value at the evaluation year?
2. Is the response variable at the evaluation year above the median value?
3. Does the response variable contain only zeros in the five years before the evaluation year?
4. Is the value of the response variable at the evaluation year the maximum observed up to that year?

Based on the output (TRUE, FALSE) of these criteria, decision rules are defined as in Table 5.

Table 5. Decision rules for determining emerging status and score based on four criteria applied to a response variable.

Decision rules	Emerging status	Emerging score
1 and 3 are FALSE, 2 and 4 are TRUE	emerging	3
2 is TRUE, all others are FALSE	potentially emerging	2
1 is TRUE and 3 is FALSE or 1, 2 and 3 are FALSE	unclear	1
3 is TRUE	not emerging	0

2.4 Rank species by total emerging score

The emerging scores at year level are used to calculate a **total emerging score**. Thus, once emerging status scores have been assigned to each species for each evaluation year, by fitting a GAM on all available data from 1950 up to and including that year (2022, 2023 and 2024 respectively), these scores are combined into a single total emerging score per species. The combination of scores into a total emerging score applies differential weights to both years and response variables. For years, more recent evaluation years carry greater weight: scores from 2024 outweigh those from 2023, which in turn outweigh those from 2022, reflecting the assumption that recent trends are more informative for early warning purposes. For response variables, the number of occupied grid cells is weighted more heavily than the number of occurrences, as occupancy is considered a more robust indicator of genuine range expansion than raw observation counts, which are more sensitive to variation in recording effort. The resulting weights are summarised in Table 6.

Table 6. Weights used to calculate the total emerging score based on the year and the response variable.

Year	Variable	Weight
2022	Number of occurrences	1
2022	Number of occupied grid cells	1.5
2023	Number of occurrences	1.5
2023	Number of occupied grid cells	2
2024	Number of occurrences	2
2024	Number of occupied grid cells	2.5

These weights are used to calculate a weighted sum of the emerging scores shown in Table 2.

Based on these weights, the maximum emerging score is 31.5, obtained if the species is defined as emerging (emerging score: 3) for each year and variable.

In case there are species with the same total emerging score the ranking is further decided by two additional criteria:

1. Emerging scores calculated using GAM take precedence above the scores calculated using decision rules.
2. The higher the average minimal guaranteed growth of the number of the occupied grid cells/year, higher the rank.

Note that here, 'guaranteed' refers to the lower bound of the statistical confidence interval around the estimated growth rate, in other words, the minimum growth that can be confidently supported by the data. The ranking is performed in the R script emtrends/src/calculate_indicators.R#L511-L582.

Note that although the occurrence cube contains records from 2025 and 2026, these years are excluded from the GAM analysis as records for the most recent years may not yet have been fully published to GBIF, which would bias the trend estimates. These more recent records are however used in the separate step of identifying appearing and reappearing species, where only the presence of a record matters rather than a trend over time.

2.5 Automation

The retrieval of all species from the EASIN Catalogue and the match of all species to the GBIF backbone is performed **daily** by the GitHub Action check_taxon_info.yaml. Changes in the EASIN Catalogue or in the match to the GBIF Backbone are detected by the action and result in a Pull Request and failing tests in test_output_match.R.

The creation of GBIF species occurrence cubes is performed **weekly** by the GitHub actions run-query-and-create-pr.yml and merge-prs-labelled-approved.yml. If the tests in test-area_wkt.R and test-list_downloads.R pass, the generation of cubes is fully automated and no human intervention is needed.

The creation of the appearing/reappearing species lists, the emerging trends analysis and the ranking are performed **weekly** by the GitHub action calculate-indicators-and-create-pr.yml. The action still requires a human review and changes are not directly applied to the main branch.

3 Results

The results can be explored by using the dashboard https://guardias-eu.github.io/indicators/. This dashboard is a Quarto dashboard hosted by GitHub Pages and based on the GitHub repository indicators.

After selecting a LME on the map, the dashboard homepage shows:

1 The **appearing species**: list of species occurring for the first time in 2022 or later.
2 The **reappearing species**: list of species occurring for the first time in 2022 or later after not being recorded for 5 years or more.
3 The **emerging species**: list with all other species ranked by the total emerging score described in 2, ordered in descending order.

For each species and LME, the plots of the number of occurrences and the number of occupied grid cells over time are also available. For the not (re)appearing species, the emerging score at year level is also shown as colored dots. If the emerging trends analysis was performed using GAM, the confidence intervals of the GAM fit are also shown in grey.

The plots can be accessed in two different ways:

1. On the Homepage by clicking on the species names shown in one of the three lists mentioned above.
2. On the Species Search page by typing the species name in the species search field and eventually choosing the LME in case the species is present in multiple LMEs. The species search field is provided with a dynamic autocompletion suggestion for easing the search among the many species.

The GitHub action update-data.yml runs **weekly**, **on demand** or every time a **change** occurs in the main branch: it retrieves the outputs from the emtrends GitHub repository and converts the ggplot2 plots to JSON files, enabling straightforward embedding in web pages. Also, every time changes occur in the main branch of the GitHub repository indicators the GitHub action publish.yml ensures that the dashboard is updated and rendered online via GitHub Pages.

In the tables below in section 3, the main outputs are shown as saved in release v1.1 of the GitHub repository emtrends, which is based on the GBIF species occurrence cube with download key 0010225-260507073636908 (DOI 10.15468/dl.rfn2ym, GBIF.org). The dashboard always

shows the most updated outputs, so the following tables and figures could slightly differ from the (dynamically updated) results in the dashboard.

In the next three sections the (re)appearing species and the most emerging species for the Mediterranean Sea, North Sea and Baltic Sea are shown. The final section provides an overview of the emerging status of marine species on the List of Invasive Alien Species of Union Concern.

Interpretation note: The rankings presented below identify species showing the strongest emerging signals in occurrence records available through GBIF. Because the current workflow does not explicitly account for changes in recording effort through time, species identified as emerging should be regarded as priority candidates for further assessment rather than definitive evidence of recent biological expansion.

3.1 Mediterranean Sea

3.1.1 Appearing species

Since 2022, a total of **35** species have been recorded for the first time.

Table 7. Appearing species in the Mediterranean Sea, listed by the year of first observation.

speciesKey	Species	First Year
2372396	*Oreochromis mossambicus*	2026
7577557	*Bispira polyomma*	2026
2292317	*Cuthona perca*	2025
2668224	*Sarcodiotheca gaudichaudii*	2025
4595317	*Saccostrea glomerata*	2025
5278964	*Gracilaria vermiculophylla*	2025
5855054	*Cephalothrix simula*	2025
5959034	*Spirobranchus kraussii*	2025
2257924	*Diadumene cincta*	2024
2501237	*Bolinopsis vitrea*	2024
4417705	*Sarscypridopsis lanzarotensis*	2024
6104442	*Thuridilla mazda*	2024
6454095	*Pleopis schmackeri*	2024
6496637	*Viriola bayani*	2024
6525922	*Pseudochama corbierei*	2024

speciesKey	Species	First Year
2372396	*Oreochromis mossambicus*	2026
7577557	*Bispira polyomma*	2026
2292317	*Cuthona perca*	2025
2320817	*Boccardia semibranchiata*	2023
4369982	*Cellana rota*	2023
4374562	*Isognomon legumen*	2023
5200463	*Corella eumyota*	2023
5200484	*Perophora viridis*	2023
5200684	*Microcosmus exasperatus*	2023
6532438	*Isognomon australicus*	2023
10933479	*Haliclona xena*	2023
2323543	*Diopatra marocensis*	2022
2323776	*Fauveliopsis glabra*	2022
2331256	*Distaplia bermudensis*	2022
2332248	*Symplegma rubra*	2022
2658889	*Grateloupia turuturu*	2022
2665285	*Galaxaura rugosa*	2022
4360387	*Crepipatella dilatata*	2022
5200487	*Perophora japonica*	2022
7668385	*Bugulina simplex*	2022
7815263	*Oscilla galilae*	2022
10755561	*Parthenina cossmanni*	2022
10822612	*Oscilla virginiae*	2022

3.1.2 Reappearing species

Since 2022, a total of **72** species have been recorded following a period of 5 years or longer during which they were not recorded.

Table 8. Reappearing species in the Mediterranean Sea, listed by the year of reappearance.

speciesKey	Species	Reappearance Year	Years Without Occurrences
2400787	*Liza carinata*	2026	9
4307290	*Thalamita poissonii*	2026	39
8365316	*Etrumeus golanii*	2026	5
2276151	*Ophiactis savignyi*	2025	18
2394612	*Morone saxatilis*	2025	6
2412987	*Spratelloides delicatulus*	2025	11
2659725	*Lophocladia trichoclados*	2025	8
4086523	*Verbena incompta*	2025	6
4360289	*Caloria indica*	2025	9
5175631	*Ligia oceanica*	2025	5
9471950	*Steromphala umbilicalis*	2025	7
9512405	*Trochus erithreus*	2025	28
10656713	*Deudorix livia*	2025	11
11916008	*Cryptocentrus steinhardti*	2025	7
12186722	*Acanthosepion pharaonis*	2025	23
2114214	*Pontella mediterranea*	2024	5
2301117	*Littorina littorea*	2024	5
2317572	*Eurythoe complanata*	2024	7
2331816	*Pyura tessellata*	2024	15
2379534	*Paracanthurus hepatus*	2024	9
2395422	*Petroscirtes ancylodon*	2024	12
4362606	*Cycloscala hyalina*	2024	15
4369932	*Smaragdia souverbiana*	2024	26
5166357	*Anomalocera patersonii*	2024	9
5193449	*Nucella lapillus*	2024	12
5209043	*Cryptocentrus caeruleopunctatus*	2024	9
8183323	*Anteaeolidiella lurana*	2024	5
9280149	*Steromphala adriatica*	2024	6
2220092	*Paracaprella pusilla*	2023	11

speciesKey	Species	Reappearance Year	Years Without Occurrences
2257920	*Diadumene leucolena*	2023	5
2266125	*Eudendrium carneum*	2023	9
2306710	*Chaetopleura angulata*	2023	39
2314184	*Lepidonotus tenuisetosus*	2023	5
2406485	*Chilomycterus reticulatus*	2023	11
2657651	*Antithamnionella spirographidis*	2023	9
4284320	*Lagocephalus lagocephalus*	2023	7
4333281	*Paracartia grani*	2023	5
5177561	*Ampithoe valida*	2023	6
5188827	*Xenostrobus securis*	2023	16
5194682	*Syllis jorgei*	2023	14
5199470	*Pileolaria militaris*	2023	5
5200503	*Polyclinum constellatum*	2023	5
5423960	*Emiliania huxleyi*	2023	6
6461590	*Ptilohyale littoralis*	2023	12
6461663	*Serejohyale spinidactylus*	2023	12
7563173	*Cryptorchestia cavimana*	2023	6
7992320	*Phorcus lineatus*	2023	5
8082159	*Palola valida*	2023	5
8116657	*Littorina saxatilis*	2023	5
9302212	*Phenacoccus peruvianus*	2023	6
9483153	*Triconia minuta*	2023	9
1048487	*Stator limbatus*	2022	5
2206602	*Paracerceis sculpta*	2022	5
2212511	*Mesanthura romulea*	2022	9
2218560	*Platorchestia platensis*	2022	6
2223416	*Penaeus semisulcatus*	2022	11
2260059	*Oulastrea crispata*	2022	5
2328734	*Novafabricia infratorquata*	2022	22

speciesKey	Species	Reappearance Year	Years Without Occurrences
2375011	*Priacanthus hamrur*	2022	5
2398885	*Stegastes variabilis*	2022	9
2509438	*Phascolion convestitum*	2022	44
2660969	*Laurencia chondrioides*	2022	20
3222430	*Vibrio vulnificus*	2022	18
4285694	*Oreochromis niloticus*	2022	11
4286150	*Tylosurus crocodilus*	2022	19
4360210	*Calma gobioophaga*	2022	5
5201040	*Hippocampus fuscus*	2022	10
5204778	*Holocentrus adscensionis*	2022	6
5717425	*Clorida albolitura*	2022	16
6467647	*Amphibalanus improvisus*	2022	5
6507885	*Pseudorhaphitoma iodolabiata*	2022	5
8082735	*Scrippsiella trochoidea*	2022	6

3.1.3 Most emerging species

A total of **63** species achieved the maximum total emerging score of 31.5. This score was derived from observations recorded during the 2022–2024 assessment period.

Table 9. Most emerging species in the Mediterranean Sea.

speciesKey	Species	Emergence Score	Model	Growth
10906181	*Lamprohaminoea ovalis*	31.5	GAM	2.0434
4363957	*Ergalatax junionae*	31.5	GAM	1.7077
2225632	*Charybdis hellerii*	31.5	GAM	1.6777
5200315	*Branchiomma luctuosum*	31.5	GAM	1.6150
2208629	*Paranthura japonica*	31.5	GAM	1.5158
6525332	*Gafrarium savignyi*	31.5	GAM	1.4371
2115620	*Chthamalus stellatus*	31.5	GAM	1.4309

speciesKey	Species	Emergence Score	Model	Growth
2315466	*Pseudonereis anomala*	31.5	GAM	1.4216
5620016	*Opuntia elatior*	31.5	GAM	1.4202
8920970	*Sepioteuthis lessoniana*	31.5	GAM	1.3673
5188945	*Anadara transversa*	31.5	GAM	1.3417
1010034	*Schizoporella errata*	31.5	GAM	1.3404
9802456	*Perforatus perforatus*	31.5	GAM	1.3217
1005675	*Watersipora cucullata*	31.5	GAM	1.3152
2395735	*Parablennius incognitus*	31.5	GAM	1.3045
2395157	*Scartella cristata*	31.5	GAM	1.3014
5862470	*Amphibalanus eburneus*	31.5	GAM	1.2906
6480320	*Planocera ceratommata*	31.5	GAM	1.2892
5229142	*Vanellus spinosus*	31.5	GAM	1.2736
2330725	*Ecteinascidia turbinata*	31.5	GAM	1.2679
2395367	*Coryphoblennius galerita*	31.5	GAM	1.2554
5192721	*Truncatella subcylindrica*	31.5	GAM	1.2543
5729648	*Spondylus spinosus*	31.5	GAM	1.2468
2393905	*Sparisoma cretense*	31.5	GAM	1.2410
2390762	*Caranx crysos*	31.5	GAM	1.2402
8767949	*Microeuraphia depressa*	31.5	GAM	1.2358
2330558	*Phallusia nigra*	31.5	GAM	1.1890
2227756	*Atergatis roseus*	31.5	GAM	1.1771
7999925	*Atherinomorus forskalii*	31.5	GAM	1.1705
4372070	*Chama pacifica*	31.5	GAM	1.1637
2374391	*Pempheris rhomboidea*	31.5	GAM	1.1508
2379295	*Pomatoschistus marmoratus*	31.5	GAM	1.1501
2356802	*Sargocentron rubrum*	31.5	GAM	1.1499
5233152	*Beroe ovata*	31.5	GAM	1.1449
7791733	*Mya arenaria*	31.5	GAM	1.1400
4362895	*Cerithium scabridum*	31.5	GAM	1.1379

speciesKey	Species	Emergence Score	Model	Growth
2223441	*Penaeus kerathurus*	31.5	GAM	1.1333
8456531	*Saurida lessepsianus*	31.5	GAM	1.1246
5273309	*Ulva lactuca*	31.5	GAM	1.1225
9775198	*Himantura uarnak*	31.5	GAM	1.1216
5855429	*Ostrea stentina*	31.5	GAM	1.1176
5863622	*Portunus segnis*	31.5	GAM	1.1146
5210979	*Diplodus vulgaris*	31.5	GAM	1.1078
6519254	*Tayuva lilacina*	31.5	GAM	1.1038
4341995	*Synaptula reciprocans*	31.5	GAM	1.1031
2411944	*Atherina boyeri*	31.5	GAM	1.1008
8608139	*Tritia neritea*	31.5	GAM	1.1002
4372686	*Ruditapes philippinarum*	31.5	GAM	1.0881
2376059	*Zebrus zebrus*	31.5	GAM	1.0848
5724049	*Gibbula ardens*	31.5	GAM	1.0798
2293333	*Calliostoma zizyphinum*	31.5	GAM	1.0758
2393577	*Scarus ghobban*	31.5	GAM	1.0683
4360148	*Melibe viridis*	31.5	GAM	1.0609
2400863	*Chelon labrosus*	31.5	GAM	1.0571
4374596	*Brachidontes pharaonis*	31.5	GAM	1.0571
2397527	*Cheilodipterus novemstriatus*	31.5	GAM	1.0550
5201390	*Platycephalus indicus*	31.5	GAM	1.0515
2392508	*Sparus aurata*	31.5	GAM	1.0437
1009964	*Reptadeonella violacea*	31.5	GAM	1.0379
2388213	*Epinephelus coioides*	31.5	GAM	1.0308
5725688	*Bivetiella cancellata*	31.5	GAM	1.0234
5213449	*Stephanolepis diaspros*	31.5	GAM	1.0174
8337656	*Pseudo-nitzschia australis*	31.5	GAM	1.0136

3.2 North Sea

3.2.1 Appearing species

Since 2022, a total of **22** species have been recorded for the first time.

Table 10. Appearing species in the North Sea, listed by the year of first observation.

speciesKey	Species	First Year
1335213	*Megachile sculpturalis*	2025
4985433	*Victorella pavida*	2025
5278375	*Colaconema dasyae*	2025
2215522	*Nippoleucon hinumensis*	2024
2376462	*Tridentiger trigonocephalus*	2024
3457717	*Savoryella lignicola*	2024
4366020	*Bolinus brandaris*	2024
5177561	*Ampithoe valida*	2024
8140432	*Corbicula largillierti*	2024
4375643	*Grateloupia minima*	2023
6460808	*Aoroides longimerus*	2023
8187685	*Gobius cruentatus*	2023
8439213	*Chelicorophium sowinskyi*	2023
1007790	*Jellyella tuberculata*	2022
2114241	*Paracalanus indicus*	2022
2312404	*Hypereteone heteropoda*	2022
2397682	*Pampus argenteus*	2022
4346647	*Anisakis pegreffii*	2022
4359698	*Ercolania viridis*	2022
5178082	*Caprella scaura*	2022
6529629	*Anadara kagoshimensis*	2022
6531311	*Arcuatula senhousia*	2022

3.2.2 Reappearing species

Since 2022, a total of **43** species have been recorded following a period of 5 years or longer during which they were not recorded.

Table 11. Reappearing species in the North Sea, listed by the year of reappearance.

speciesKey	Species	Reappearance Year	Years Without Occurrences
5727801	*Cymbium olla*	2026	8
2114953	*Mesocalanus tenuicornis*	2025	6
3198332	*Microspongium globosum*	2025	8
3199159	*Ulonema rhizophorum*	2025	39
3199244	*Punctaria tenuissima*	2025	8
4363583	*Rapana venosa*	2025	5
2264670	*Haliclystus salpinx*	2024	9
2316069	*Perinereis cultrifera*	2024	6
2564481	*Haligena elaterophora*	2024	64
2665384	*Hildenbrandia crouaniorum*	2024	27
3193042	*Chaetoceros peruvianus*	2024	5
4287131	*Acipenser baerii*	2024	8
5166638	*Oithona davisae*	2024	8
5197206	*Clymenella torquata*	2024	8
6460818	*Aoroides semicurvatus*	2024	6
6467132	*Eurytemora carolleeae*	2024	5
7525773	*Rhizodomus tagatzi*	2024	5
2328302	*Hydroides elegans*	2023	6
2328383	*Hydroides ezoensis*	2023	5
2330832	*Aplidium proliferum*	2023	15
2480445	*Haliaeetus pelagicus*	2023	10
2565841	*Lulwoana uniseptata*	2023	7
2657432	*Antithamnion densum*	2023	6
3198195	*Myriactula rivulariae*	2023	22

speciesKey	Species	Reappearance Year	Years Without Occurrences
4408321	*Craspedacusta sowerbii*	2023	11
6180498	*Pachycordyle michaeli*	2023	13
2110811	*Mytilicola intestinalis*	2022	14
2115556	*Tortanus discaudatus*	2022	5
2288831	*Lyrodus pedicellatus*	2022	15
2309075	*Hesionides arenaria*	2022	26
2331954	*Styela plicata*	2022	6
2688081	*Mantoniella squamata*	2022	5
3198548	*Cladosiphon zosterae*	2022	17
3199341	*Acrothrix gracilis*	2022	6
4288621	*Cribrilaria innominata*	2022	5
4357756	*Haliotis pustulata*	2022	62
5273393	*Ulva californica*	2022	7
5273451	*Ulva gigantea*	2022	7
7531092	*Eutintinnus pectinis*	2022	10
7563173	*Cryptorchestia cavimana*	2022	6
7713874	*Pyropia yezoensis*	2022	17
8082735	*Scrippsiella trochoidea*	2022	6
12186722	*Acanthosepion pharaonis*	2022	40

3.2.3 Most emerging species

A total of **22** species achieved the maximum total emerging score of 31.5. This score was derived from observations recorded during the 2022–2024 assessment period.

Table 12. Most emerging species in the North Sea.

speciesKey	Species	Emergence Score	Model	Growth
2662823	*Gelidium vagum*	31.5	GAM	1.4505
4372686	*Ruditapes philippinarum*	31.5	GAM	1.4501
2227300	*Procambarus clarkii*	31.5	GAM	1.3720
5188971	*Rangia cuneata*	31.5	GAM	1.2799
2225772	*Hemigrapsus sanguineus*	31.5	GAM	1.2756
5279568	*Agardhiella subulata*	31.5	GAM	1.2574
2285681	*Mytilus trossulus*	31.5	GAM	1.2219
5272096	*Codium fragile*	31.5	GAM	1.1553
2287615	*Petricolaria pholadiformis*	31.5	GAM	1.1326
3196454	*Fucus distichus*	31.5	GAM	1.0987
2666539	*Caulacanthus okamurae*	31.5	GAM	1.0926
2224979	*Palaemon elegans*	31.5	GAM	1.0849
2287668	*Mercenaria mercenaria*	31.5	GAM	1.0646
5277182	*Ceramium tenuicorne*	31.5	GAM	1.0632
5277987	*Polysiphonia senticulosa*	31.5	GAM	1.0569
5166487	*Megabalanus tintinnabulum*	31.5	GAM	1.0568
9736038	*Necora puber*	31.5	GAM	1.0515
3196387	*Fucus evanescens*	31.5	GAM	1.0409
2301066	*Littorina obtusata*	31.5	GAM	1.0324
7791733	*Mya arenaria*	31.5	GAM	1.0265
7627628	*Leathesia marina*	31.5	GAM	1.0259
5178595	*Carcinus maenas*	31.5	GAM	1.0249

3.3 Baltic Sea

3.3.1 Appearing species

Since 2022, a total of **13** species have been recorded for the first time.

Table 13. Appearing species in the Baltic Sea, listed by the year of first observation.

speciesKey	Species	First Year
2288926	*Potamocorbula amurensis*	2025
2291865	*Corambe obscura*	2025
8190231	*Corbicula fluminea*	2025
2110813	*Mytilicola orientalis*	2024
2285980	*Pecten maximus*	2024
2376212	*Crystallogobius linearis*	2024
2578507	*Erysiphe corylacearum*	2024
2115556	*Tortanus discaudatus*	2023
2379243	*Proterorhinus nasalis*	2023
3193658	*Biddulphia sinensis*	2023
10933479	*Haliclona xena*	2023
2225646	*Callinectes sapidus*	2022
3197314	*Sargassum muticum*	2022

3.3.2 Reappearing species

Since 2022, a total of **5** species have been recorded following a period of 5 years or longer during which they were not recorded.

Table 14. Reappearing species in the Baltic Sea, listed by the year of reappearance.

speciesKey	Species	Reappearance Year	Years Without Occurrences
1450317	*Telmatogeton japonicus*	2025	5
2383720	*Centrolabrus exoletus*	2024	11
6467132	*Eurytemora carolleeae*	2024	5
2323931	*Aphelochaeta marioni*	2023	19
2266741	*Maeotias marginata*	2022	10

3.3.3 Most emerging species

A total of **6** species achieved the maximum total emerging score of 31.5. This score was derived from observations recorded during the 2022–2024 assessment period.

Table 15. Most emerging species in the Baltic Sea.

speciesKey	Species	Emergence Score	Model	Growth
2308528	*Tubificoides heterochaetus*	31.5	GAM	1.2828
2301066	*Littorina obtusata*	31.5	GAM	1.2373
3195275	*Apedinella radians*	31.5	GAM	1.1163
8368630	*Prorocentrum cordatum*	31.5	GAM	1.0785
2292354	*Tenellia adspersa*	31.5	GAM	1.0385
2315382	*Alitta succinea*	31.5	GAM	1.0319

3.4 Marine species on the List of Invasive Alien Species of Union Concern

This section provides illustrative examples of species included on the List of Invasive Alien Species of Union Concern. While not intended as a formal validation exercise, these examples provide an indication of the capacity of the workflow.

Table 16. Marine species on the List of Invasive Alien Species of Union Concern.

speciesKey	Species	LME	(re)appearing	Total emerging score
5824863	*Rugulopteryx okamurae*	Mediterranean Sea	-	13.5
5824863	*Rugulopteryx okamurae*	Iberian Coastal	-	18
2863941	*Nanozostera japonica*	-	-	-
7965247	*Plotosus lineatus*	Mediterranean Sea	-	0
2394604	*Morone americana*	-	-	-
5712056	*Fundulus heteroclitus*	-	-	-
2225776	*Eriocheir sinensis*	Iberian Coastal	Reappearing in 2024 after 57 years	-
5187508	*Asterias amurensis*	-	-	-
2286388	*Mulinia lateralis*	Celtic-Biscay Shelf	Appearing in 2024	-

3.4.1 Rugulopteryx okamurae - Mediterranean Sea

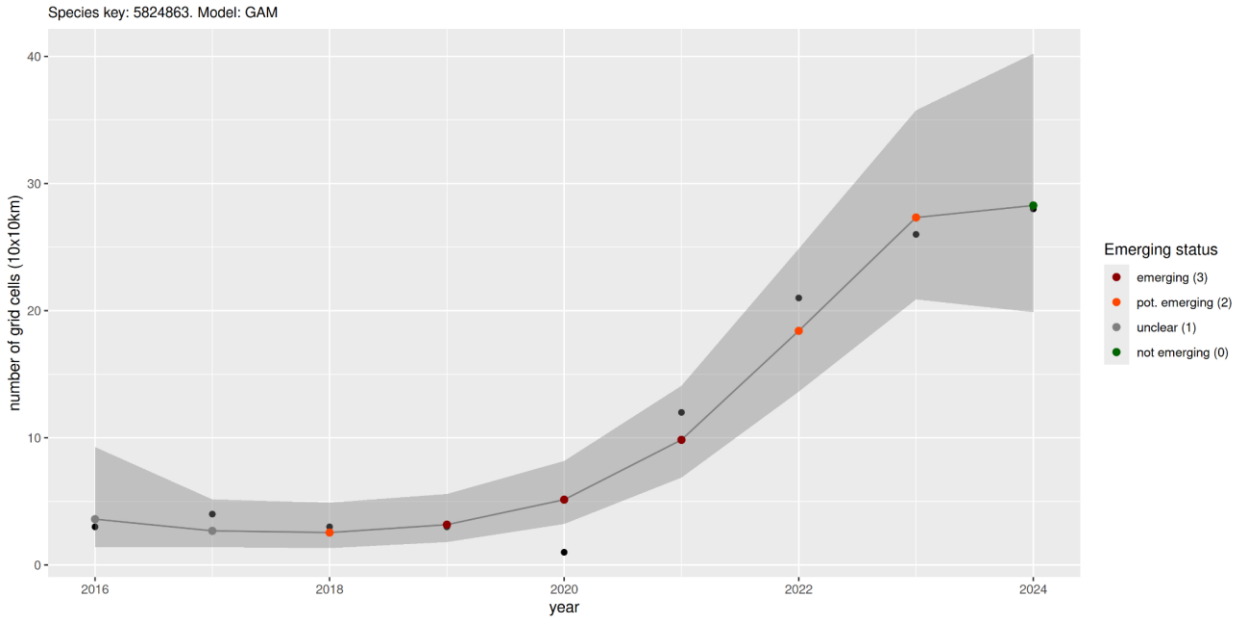


Figure 2. Black dots: the number of grid cells (10kmx10km) occupied by Rugulopteryx okamurae in the Mediterranean Sea per year. Colored dots: emerging status as calculated based on the first and second derivatives of the upper and lower confidence levels of the GAM fit, shown as the grey bounds.

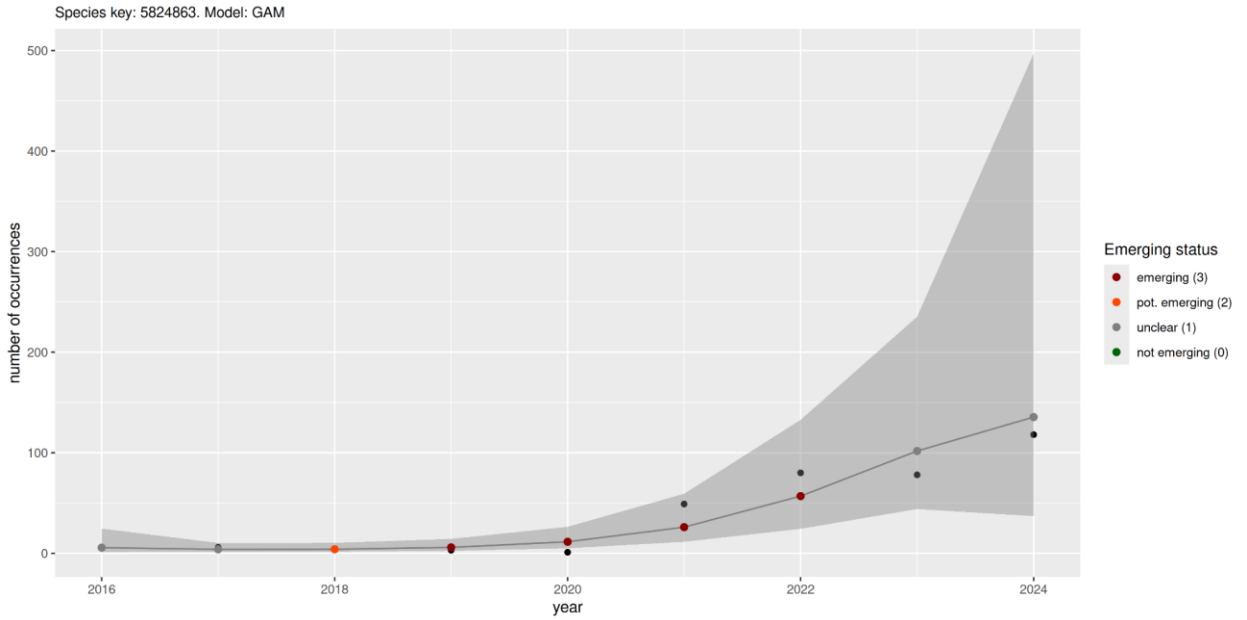


Figure 3. Black dots: the number of occurrences of Rugulopteryx okamurae in the Mediterranean Sea per year. Colored dots: emerging status as calculated based on the first and second derivatives of the upper and lower confidence levels of the GAM fit, shown as the grey bounds.

3.4.2 *Rugulopteryx okamurae* - Iberian Coastal

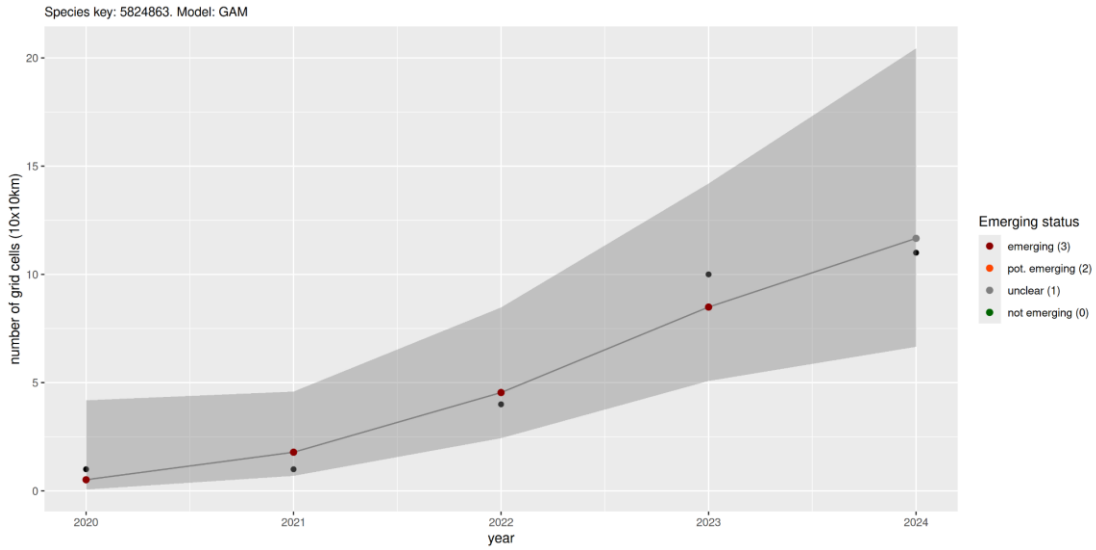


Figure 4. Black dots: the number of grid cells (10kmx10km) occupied by Rugulopteryx okamurae in the Iberian Coastal per year. Colored dots: emerging status as calculated based on the first and second derivatives of the upper and lower confidence levels of the GAM fit, shown as the grey bounds.

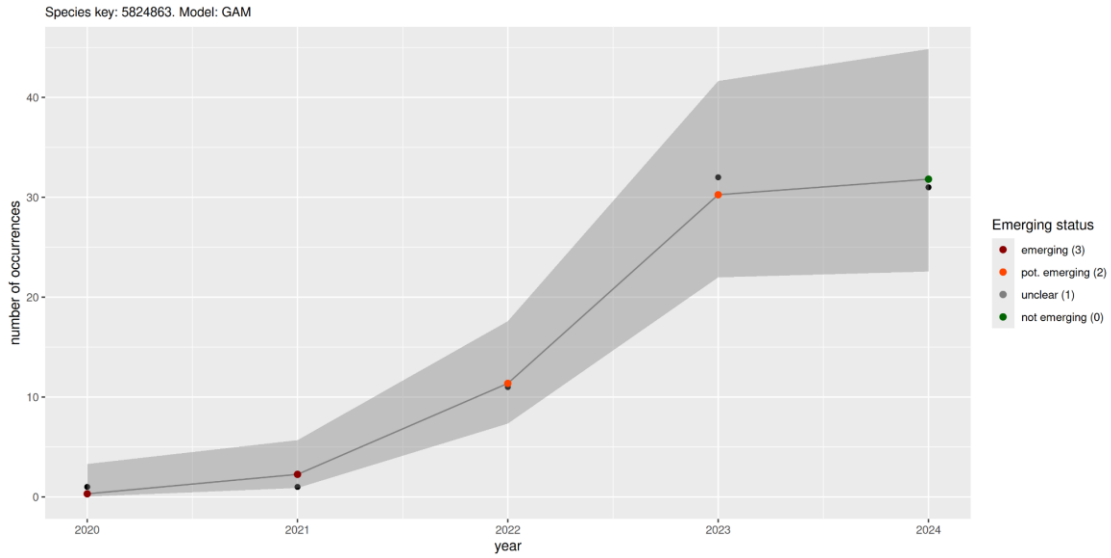


Figure 5. Black dots: the number of occurrences of Rugulopteryx okamurae in the Iberian Coastal per year. Colored dots: emerging status as calculated based on the first and second derivatives of the upper and lower confidence levels of the GAM fit, shown as the grey bounds.

3.4.3 Plotosus lineatus - Mediterranean Sea



Figure 6. Black dots: the number of grid cells (10kmx10km) occupied by Plotosus lineatus in the Mediterranean Sea per year. Colored dots: emerging status as calculated based on the first and second derivatives of the upper and lower confidence levels of the GAM fit, shown as the grey bounds.



Figure 7. Black dots: the number of occurrences of Plotosus lineatus in the Mediterranean Sea per year. Colored dots: emerging status as calculated based on the first and second derivatives of the upper and lower confidence levels of the GAM fit, shown as the grey bounds.

3.4.4 *Eriocheir sinensis* - Iberian Coastal

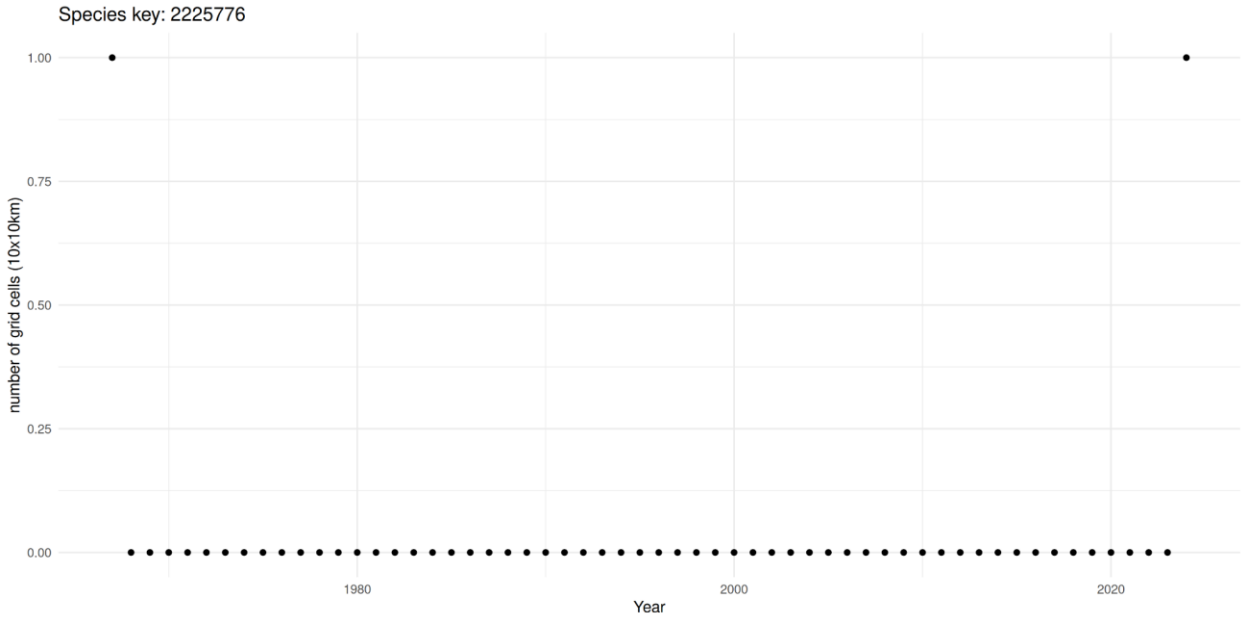


Figure 8. The number of grid cells (10kmx10km) occupied by Eriocheir sinensis in the Iberian Coastal per year.

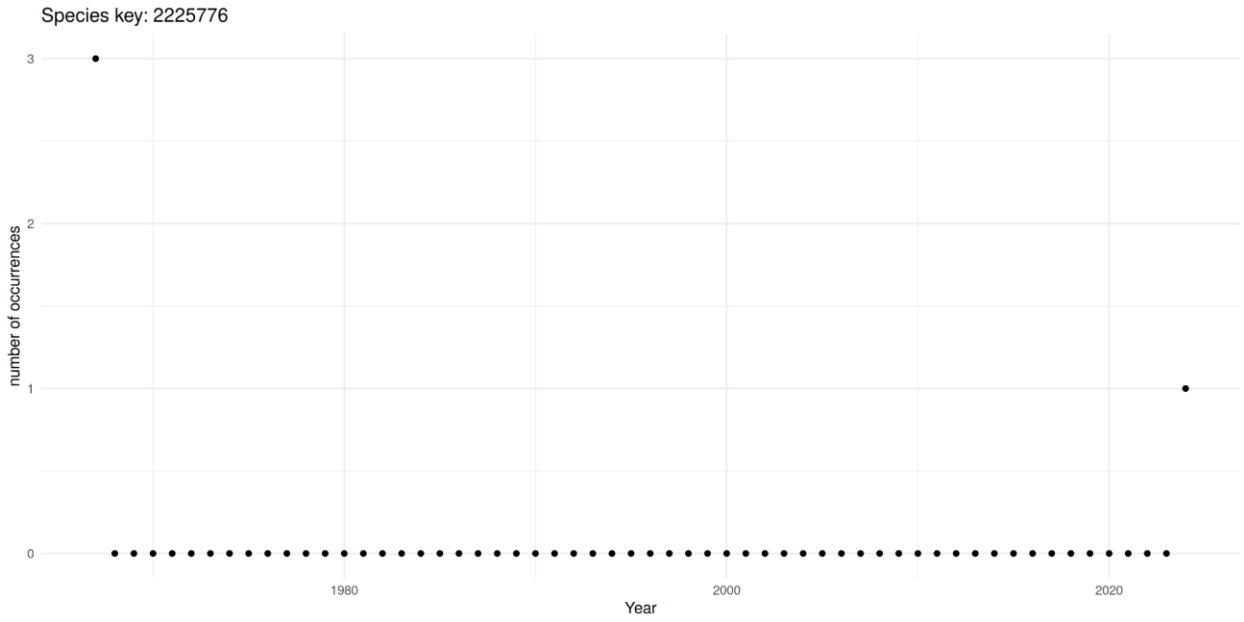


Figure 9. Black dots: the number of occurrences of Eriocheir sinensis in the Iberian Coastal per year.

3.4.5 Mulinia lateralis - Celtic-Biscay Shelf

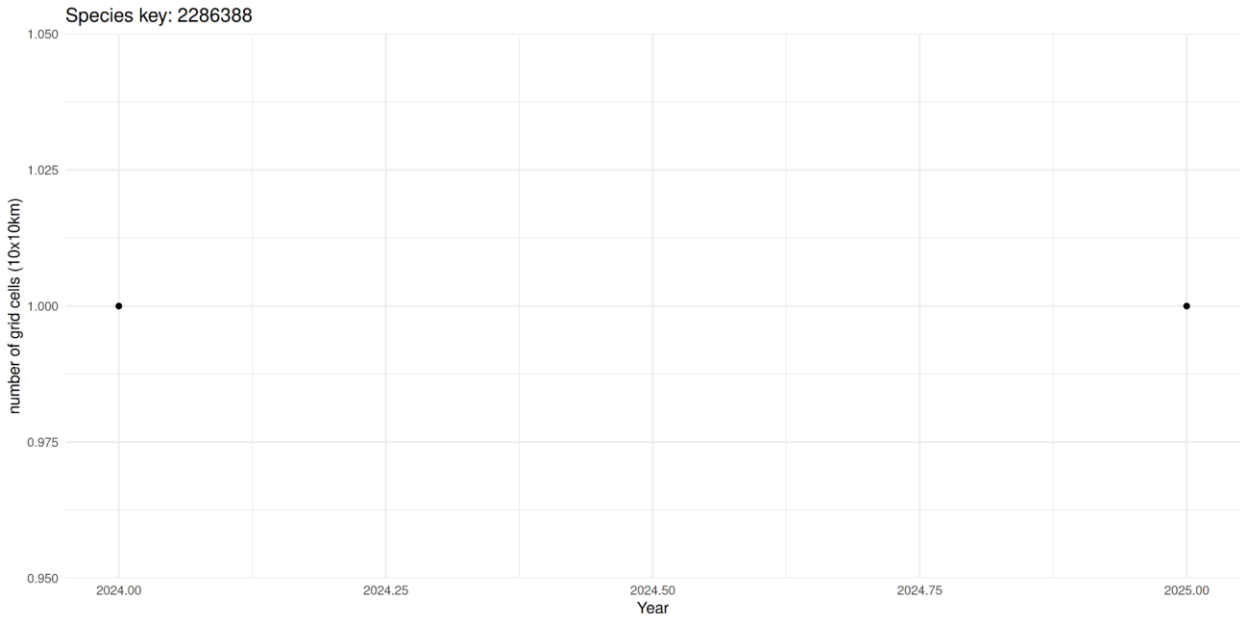


Figure 10. The number of grid cells (10kmx10km) occupied by Mulinia lateralis in the Celtic-Biscay Shelf per year.

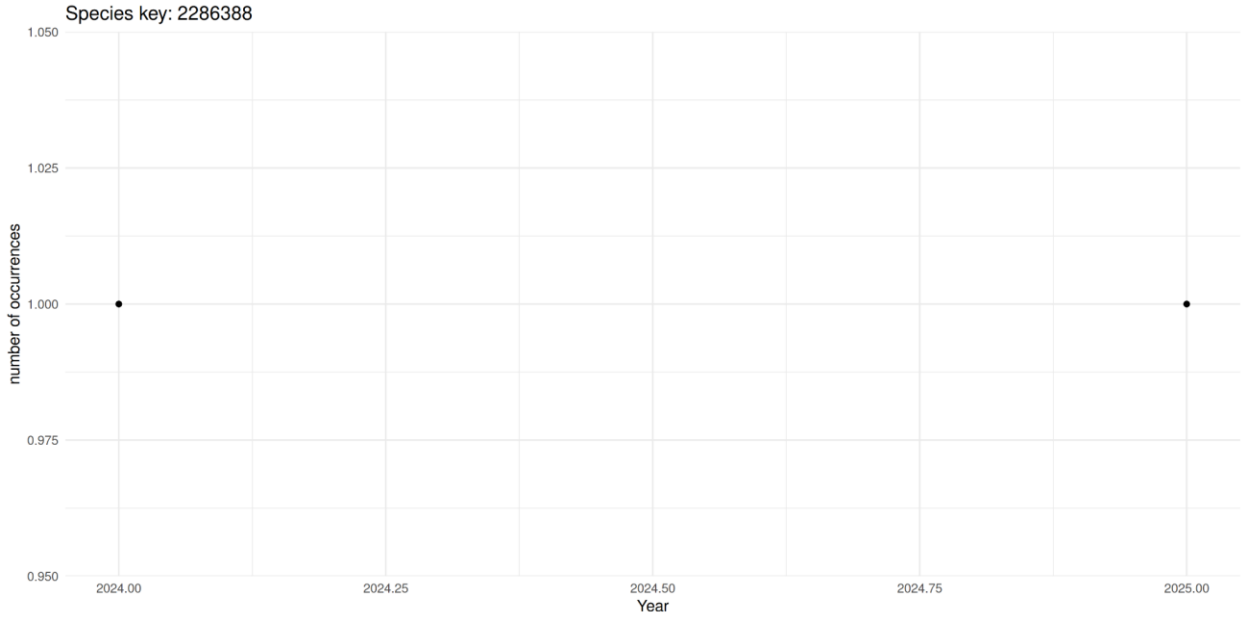


Figure 11. Black dots: the number of occurrences of Mulinia lateralis in the Celtic-Biscay Shelf per year.

4 Further Developments

The workflow proposed here will be further improved during the second half of the GuardIAS project.

Taxonomically, the OneSTOP list of introduced and invasive species in Europe, which will be produced by the OneSTOP project by the end of 2026, should be used (emtrends#15). This checklist will be an aggregation of national (GRIIS, Global Register for Introduced and Invasive Species) checklists. The European checklist will include both terrestrial, marine and freshwater species(emtrends#16).

Also, taxa with different **taxonomic rank** than species must be included (emtrends#36). However, this means creating different occurrence cubes as it means grouping occurrences at different taxonomic levels, e.g. at genus level (GROUP BY genusKey) or at taxonomic level in case of subspecies, forms or varieties (GROUP BY taxonKey). The reason is that join queries or sub-queries are not allowed by the GBIF SQL Download API. More details can be found in section Supported SQL in the GBIF API SQL Downloads documentation.

Spatially, next to the LME we will include all the European countries. We will also try to increase the spatial granularity by using the 1km x 1km EEA reference grid (emtrends#13). This increase in resolution may be more relevant for freshwater IAS.

Methodologically, we will test the use of the number of occurrences and the number of occupied grid cells at class level as covariates of the GAM fit, in order to account for the research effort bias (emtrends#12). This means that a new species occurrence cube must be generated: this cube must have the same spatial and temporal resolution and constraints, but the occurrences must be aggregated at class level.

Future work should also include a systematic evaluation of workflow performance through expert validation and comparison with independent invasion assessments. Such evaluations would allow estimation of false-positive and false-negative rates and help refine decision thresholds used in the emerging-species classification framework.

5 Concluding remarks

This deliverable presents a reproducible, data-driven workflow for detecting and prioritizing emerging IAS in European Large Marine Ecosystems, building on openly available occurrence data from GBIF. The workflow combines occurrence cubes with statistical trend analysis using Generalized Additive Models (GAM) and a set of deterministic decision rules as a fallback, to assign emerging status scores to IAS across multiple years and regions. The primary output of this deliverable is the workflow itself; the species-level results presented here are illustrative of its

capabilities, and their reliability is inherently constrained by the quality and completeness of the underlying data.

Marine occurrence data are substantially less well represented in open biodiversity databases than terrestrial data. This means that species listed as newly appearing or emerging may in reality have been present and spreading for considerably longer, with earlier records remaining unpublished or scattered across grey literature and personal datasets. Coordinated efforts to mobilize such historical marine occurrence data, such as those by Katsanevakis et al. (2020) and Ragkousis et al. (2023), who compiled thousands of unpublished Mediterranean and Black Sea records through large collaborative initiatives, are paramount to improving the reliability of workflows like the one presented here. However, for these efforts to have their full impact, the compiled records should ultimately be published to GBIF or equivalent open biodiversity databases; data that remain in journal appendices or institutional repositories are only semi-open and cannot be automatically harvested by workflows such as this one. As more historical data are made openly available, the workflow will automatically benefit from them without any structural changes.

Because of these data limitations, the current results should be interpreted with caution. Expert review during the preparation of this report (Agnese Marchini, pers. comm, see Table 17) identified several species in the appearing species lists whose first GBIF-mediated record postdates known literature records by years or even decades, highlighting that while the workflow correctly reports the first year a species appears in the available data, this does not always correspond to the true first introduction record. Additionally, the current workflow version does not yet correct for research effort bias, meaning that some species may be flagged as emerging simply due to growing recording activity rather than genuine biological expansion. Planned improvements include incorporating class-level occurrence data as a covariate to correct for this bias, extending the taxonomic scope to include non-species ranks, broadening the spatial coverage to all European countries, and integrating the forthcoming OneSTOP European checklist of introduced and invasive species.

The workflow should be viewed as a scalable and transparent screening tool that supports horizon scanning and prioritisation rather than as a standalone assessment of invasion status. Combining automated detection with expert evaluation will maximise its usefulness for early warning and rapid response activities across European marine ecosystems.

Table 17. Summary of corrections and annotations flagged by Marchini Agnese during expert review of the WP2 D2.5 deliverable. Original entries reflect what the workflow detected based on GBIF data; corrections indicate discrepancies identified against the literature.

Table	Location / Section	Reviewer	Original text / entry	Correction / note
Table 6	Mediterranean Sea — Appearing species	Marchini Agnese (pers. comm.)	*Bispira polyomma* — *First Year: 2026*	First record in Italy actually 2014; GBIF record lags literature by 12 years

Table	Location / Section	Reviewer	Original text / entry	Correction / note
Table 6	Mediterranean Sea — Appearing species	Marchini Agnese (pers. comm.)	*Cuthona perca — First Year: 2025*	First record in Italy actually 1976–77; GBIF record lags literature by ~49 years
Table 6	Mediterranean Sea — Appearing species	Marchini Agnese (pers. comm.)	*Gracilaria vermiculophylla — First Year: 2025*	First record in Italy actually 2008; GBIF record lags literature by 17 years
Table 6	Mediterranean Sea — Appearing species	Marchini Agnese (pers. comm.)	*Diadumene cincta — First Year: 2024*	First record in Italy actually 1993; GBIF record lags literature by 31 years
Table 6	Mediterranean Sea — Appearing species	Marchini Agnese (pers. comm.)	*Bolinopsis vitrea — First Year: 2024*	Mediterranean native; reported as possible ballast-water introduction only in the Black Sea. Classification as appearing IAS questionable
Table 6	Mediterranean Sea — Appearing species	Marchini Agnese (pers. comm.)	*Sarscypridopsis lanzarotensis — First Year: 2024*	Freshwater species from springs and caves; should not appear in marine IAS list
Table 6	Mediterranean Sea — Appearing species	Marchini Agnese (pers. comm.)	*Cellana rota — First Year: 2023*	First record in Greece actually 1999; GBIF record lags literature by 24 years
Table 6	Mediterranean Sea — Appearing species	Marchini Agnese (pers. comm.)	*Oreochromis mossambicus — First Year: 2026*	Freshwater species (Nile tilapia); presence in marine appearing species list requires verification
Table 12	Baltic Sea — Appearing species	Marchini Agnese (pers. comm.)	*Mytilicola orientalis — First Year: 2024*	First record in the Baltic Sea actually 2018 per: Brenner et al. (2019), BioInvasions Record 8(3). GBIF record lags literature by 6 years

Note: 'First Year' in the original tables refers to the first year a species appeared in the GBIF occurrence cube, not necessarily the true first introduction record. Entries in italics in the 'Original text' column reflect the exact wording in the deliverable.

6 Bibliography

Adriaens T, Sutton-Croft M, Owen K, Brosens D, van Valkenburg J, Kilbey D, Groom Q, Ehmig C, Thürkow F, Van Hende P, Schneider K (2015. Trying to engage the crowd in recording invasive alien species in Europe: experiences from two smartphone applications in northwest Europe. Management of Biological Invasions 6(2): 215-25.

Brenner M, Schulze J, Fischer J, Wegner KM (2019) First record of the parasitic copepod (Mytilicola orientalis Mori, 1935) in blue mussels (Mytilus spp.) of the Baltic Sea. BioInvasions Records 8(3): 623–632, https://doi.org/10.3391/bir. 2019.8.3.19

Chamberlain S, Barve V, Mcglinn D, Oldoni D, Desmet P, Geffert L, Ram K (2026). _rgbif: Interface to the Global Biodiversity Information Facility API_. R package version 3.8.5, https://CRAN.R-project.org/package=rgbif

Chamberlain S, Boettiger C (2017). "R Python, and Ruby clients for GBIF species occurrence data." *PeerJ PrePrints*. https://doi.org/10.7287/peerj.preprints.3304v1

Desmet P, Reyserhove L, Oldoni D, Groom Q, Adriaens T, Vanderhoeven S, Pagad S (2025). Global Register of Introduced and Invasive Species - Belgium. Version 1.15. Invasive Species Specialist Group ISSG. Checklist dataset https://doi.org/10.15468/xoidmd accessed via GBIF.org on 2026-05-05.

GBIF Secretariat (2023). GBIF Backbone Taxonomy. Checklist dataset https://doi.org/10.15468/39omei accessed via GBIF.org on 2026-05-05.

GBIF.org (11 May 2026) GBIF Occurrence Download https://doi.org/10.15468/dl.rfn2ym

Groom Q, Adriaens T, Oldoni D, Reyserhove L, Strubbe D, Vanderhoeven S, Desmet P (2018). TrIAS, leveraging citizen science data to monitor invasive species in Belgium. Biodiversity Information Science and Standards. 2: e24749.

Groom, Q., Adriaens, T., Oldoni, D., Reyserhove, L., Strubbe, D., Vanderhoeven, S. and Desmet, P., 2018. TrIAS, leveraging citizen science data to monitor invasive species in Belgium. Biodiversity Information Science and Standards, 2, p.e24749.

Katsanevakis S, Poursanidis D, Hoffman R, Rizgalla J, Rothman SB-S, Levitt-Barmats Y, Hadjioannou L, et al. (2020) Unpublished Mediterranean records of marine alien and cryptogenic species. BioInvasions Records 9(2): 165–182, https://doi.org/10.3391/bir.2020.9.2.01

Oldoni, D. (2026). reasin: Interface to the European Alien Species Information Network API (Version 0.0.1.0) [Computer software]. https://github.com/guardias-eu/reasin

Oldoni, D., Delva, S., Devisscher, S., & Varewyck, M. (2026). trias: Process Data for the Project Tracking Invasive Alien Species (TrIAS) (Version 3.2.4) [Computer software]. https://github.com/trias-project/trias

Oldoni, D., Groom, Q., Adriaens, T., Davis, A. J. S., Reyserhove, L., Strubbe, D., Vanderhoeven, S., & Desmet, P. (2020). Occurrence cubes: a new paradigm for aggregating species occurrence data. BioRxiv. https://doi.org/10.1101/2020.03.23.983601

Oldoni, D., Groom, Q., Adriaens, T., Hillaert, J., Reyserhove, L., Strubbe, D., Vanderhoeven, S., & Desmet, P. (2024a). Occurrence cubes at species level for European countries (Version 20240124) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.10528506

Oldoni, D., Groom, Q., Adriaens, T., Davis, A. J. S., Reyserhove, L., Strubbe, D., Vanderhoeven, S., & Desmet, P. (2024b). Occurrence cubes for non-native taxa in Belgium and Europe (Version 20240118) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.10527772

Oldoni, D., Delva, S., Devisscher, S., Van Daele, T., Vanderhoeven, S., & Adriaens, T. (2024c). A list of newly (re)appearing alien species in Belgium in support of decision making (Version 20240703) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.12637525

Oldoni, D., Delva, S., Desmet, P., Devisscher, S., Van Daele, T., Vanderhoeven, S., & Adriaens, T. (2024d). Prioritized lists of alien species in Belgium and its regions (Version 20240704) [Data set]. Research Institute for Nature and Forest (INBO). https://doi.org/10.5281/zenodo.10897501

Ragkousis M, Zenetos A, Ben Souissi J, Hoffman R, Ghanem R, Taşkın E, Muresan M, et al. (2023) Unpublished Mediterranean and Black Sea records of marine alien, cryptogenic, and neonative species. BioInvasions Records 12(2): 339–369, https://doi.org/10.3391/bir.2023.12.2.01

Vanderhoeven S, Adriaens T, Desmet P, Strubbe D, Backeljau T, Barbier Y, Brosens D, Cigar J, Coupremanne M, De Troch R, Eggermont H, Heughebaert A, Hostens K, Huybrechts P, Jacquemart A, Lens L, Monty A, Paquet J, Prévot C, Robertson T, Termonia P, Van De Kerchove R, Van Hoey G, Van Schaeybroeck B, Vercayie D, Verleye T, Welby S, Groom Q (2017) Tracking Invasive Alien Species (TrIAS): Building a data-driven framework to inform policy. Research Ideas and Outcomes 3: e13414. https://doi.org/10.3897/rio.3.e13414