UiA University of Agder

Analyzing the reliability of citizen science information using Pacific oysters as a model system

AMNA FAWAD

SUPERVISOR
Ane Timenes Laugen

University of Agder, 2023
Faculty for engineering and science
Department of Natural Sciences

University of Agder
Faculty of Engineering and Science
Department of Natural Science
Gimlemoen 25
4604 Kristiansand
http://www.uia.no

Acknowledgements:

I would like to express my gratitude to my supervisor Ane Timenes Laugen for all the assistance, guidance and writing expertise in this project. As well as Lars Korslund, Molly Raemon and Johanna Bjånes Marcussen for their respective contributions to field surveys and data analysis. Additionally, I also would like to thank Mari Tveit Birkeland for Peer-reviewing my paper and being an emotional support throughout this thesis. Lastly, I would like to thank all my friends and family that have been a part of this citizen science journey unwillingly.

Sammendrag:

Det biologiske mangfolds krisen har ført til at mange arter har blitt utryddet de siste 20 årene; En av årsakene til dette er økende nærvær av fremmede arter. Mange av disse artene er uidentifiserte, noe som gjør det vanskelig å skaffe kunnskap om hver enkelt av disse artene. En mulig løsning på dette kan være bruk av data fra folkeforskning, som har et uutnyttet potensial og kan svare på mange vitenskapelige spørsmål knyttet til flere forskningsfelt dersom de brukes riktig. Eksempler på slike åpent tilgengelig databaser er det norske rapportsystemet for arter, Artsobservasjoner og den tilsvarende svenske Arsportalen. Det er imidlertid en stor bekymring rundt gyldigheten og kvaliteten til denne typen data, da alle kan legge inn slike data i folkeforsknings prosjekter.

For å undersøke påliteligheten til folkeforsknings data, analyserte denne studien påliteligheten på tre måter ved hjelp av folkeforskningsrapporter og profesjonelle studier av den invasive kystarten stillehavsøsteres. Dette ble gjort ved først å sjekke kvaliteten på databasene ved hjelp av visuelt kart for å oppdage observasjoner som ble ansett som falske oppføringer. For det andre ble påliteligheten til databasene undersøkt ved å undersøke tilstedeværelse av stillehavsøsters i folkeforskningsdata, disse ble verifisert gjennom feltundersøkelse av to steder i det norske og svenske kystområdet. Videre ble det generert tilfeldige steder som en sammenligning med verifiserte folkeforskningsdata og geografisk skjevhet i utvalget. Etter feltundersøkelsen ble det laget en generalisert lineær modell for å analysere påliteligheten til de verifiserte stedene. Til slutt, for å undersøke om folkeforsknings databaser er raskere med å spore spredningen av arten enn profesjonelle studier, ble det laget en generalisert lineær regresjon som sammenliknet antall observasjoner i folkeforskning og profesjonelle studier i tid og rom. Basert på resultater fra denne studien viser det seg at folkeforsknings data er pålitelige under parameterne til denne studien, og at vitenskap samlet av folket er først til å spore utvidelsesområdet i visse regioner. Ytterligere studier bør prøve å iverksette bedre måter å tolke folkeforsknings data på og dermed forbedre den generelle kvaliteten på forskning utført av folket. Dette åpner muligheter for bredere forskningsspørsmål som kan gjennomføres uten omfattende bruk av penger og ressurser. I tillegg eliminerer det bekymringer om tidsmessige og romlige begrensninger som ofte begrenser offentlig finansiert forskning og overvåking.

Abstract

The biodiversity crisis has led to many species becoming extinct in the last 20 years; One of the drivers of the decline is non-native species. Many of these species are unidentified and that makes it difficult to keep track of every one of them. A possible solution to this can be the use of citizen science data, which has untapped potential and can answer many scientific questions related to several research fields if used correctly. Examples of such all-accessible databases are the Norwegian biodiversity information center and Swedish biodiversity information center´s databases. There is, however, a large concern around the validity and quality of this type of data as anyone can input such data into citizen science databases.

To investigate the reliability of citizen science data, this study analyzed the reliability in three ways. The first step was to check the quality of the databases using visual maps to detect observations that were considered false entries. Secondly, the validity of the databases was investigated by verifying presence of Pacific oysters in the citizen science databases, these observations were verified through field survey of two locations in the Norwegian and Swedish coastal area. Furthermore, random points were generated as a comparison to verified citizen data observations and checking for geographical bias in Pacific oyster observations conducted by the citizen scientists. After the field survey, a generalized linear model was made to analyze the reliability of the verified points. Lastly, to investigate if citizen science is faster in tracking the spread of the species, generalized linear regression was conducted and visualized. Based on results from this study, citizen science data is shown to be reliable under the parameters of this study and that citizen science is first to track the expansion range in certain regions. Further studies should try to implement better ways to interpret citizen science data thereby improving overall quality of citizen science research. This opens avenues for broader research questions that can be conducted without extensive use of money and resources. In addition, it eliminates concerns about the temporal and spatial constraints that often is imposed on publicly funded research and monitoring.

Contents

1 Introduction

1.1 Biodiversity crisis, anthropogenic disturbances, and the need for data

Biodiversity, encompassing all life forms, ecosystems, and ecological processes, is regarded as the basis of human survival and economic well-being (Millennium ecosystem assessment 2005; Pecl et al. 2017). While over 1 to 6 billion species are estimated to be found on earth, only a small percentage of those have been identified (almost 1.5 million), revealing a significant knowledge gap (Larsen et al. 2017). Scientists have estimated that there have been 5 mass extinctions on Earth, so it may be alarming to know that we are currently living under one right now (Elewa and Abdelhady 2020). Importantly, prior extinctions have occurred almost every million years (Singh 2002; Elewa and Abdelhady 2020). In contrast, anthropogenic disturbances and widespread diseases have led to mass extinction within a span of just under 200 years (Singh 2002; Elewa and Abdelhady 2020).

Global environmental changes caused by humans have resulted in the introduction of alien species, also known as non-native or invasive species (Pejchar and Mooney 2009). In addition to competing for resources with native species, these non-native species may disrupt food webs and alter ecosystem processes, thus threatening ecosystems, habitats, and species (Pejchar and Mooney 2009; David et al. 2017). In extreme cases, invasive species can even drive native species to extinction (Pejchar and Mooney 2009). Accordingly, following global warming, the establishment rate of invasive species has increased worldwide in the past century and is predicted to also continue in the future (Huang et al. 2011; Seebens et al. 2021). Moreover, more than half of the habitable surface on earth has already been altered by human interactions, significantly impacting the habitats that most species depend on and risking complex ecological interactions being disrupted and altered, which can endanger several species dependent on these interactions (Singh 2002; Koh et al. 2004). Without significant changes to how humans influence the environment, it is estimated that 20% of all species will be extinct within 30 years and 50% will be lost within 50 years (Singh 2002; Koh et al. 2004).

The large discrepancy between identified species and estimated number of species makes it difficult to know the exact number of invasive species in the world at any given moment. As a result of this uncertainty, it is significantly more difficult to keep track of each species than if they had already been identified. New methods such as eDNA (Darling and Blum 2007), remote sensing (Ustin et al. 2002), visual and acoustic survey equipment in combination with machine learning (Juanes 2018) and AI (Ashqar and Abu-Naser 2019) are being developed continuously and becoming important tools for mapping and monitoring the amount and degree of invasion. Although there have been many methodological developments, some research questions can only be answered by data collected in the field. However, it is impossible for researchers and managers of natural resources with limited resources to cover practically every inch of the land and sea within their jurisdiction (Pyšek et al. 2020). Therefore, one additional piece in the toolkit is the use of Citizen science.

1.2 Citizen science data

Before addressing the implications of citizen science, we need to define what citizen science is. Hackley et al. (2021) compared different definitions from different sources and concluded there are three main components that need to be present for the information to be considered citizen science. Firstly, it is the practice of the public, volunteers, professionals joining together to gather information. This is done by either collecting, categorizing , transcribing, or analyzing data by people with different expertise levels and therefore are not necessarily experts in the area which the data is being collected on (Bonney et al. 2014). In citizen science, the size of a collection project can vary from small and local information gathering from the public to millions of individuals reporting on a global scale (Bonney et al. 2014). Additionally, in just a short amount of time, the use of citizen science data has grown immensely in recent years (Bonney et al. 2014). This growth comes from the efficient integration of the internet, making it possible for citizen projects to gain increased visibility, functionality, and accessibility than ever before (Bonney et al. 2014). Secondly, the information that is gathered must be scientific in nature, either helping in answering a scientific question or conducting research (Haklay et al. 2021). Some examples includes documenting impacts of climate change, informing land management, observing and surveying various plants, animals, and water quality in the area (Bonney et al. 2014; Burgess et al. 2017). Citizen science can also help in formulating scientific questions and trying out new methods (Bonney et al. 2014). Lastly, the third component was the need to address or answer a politically relevant issue with citizen science, however this was not present in all definitions (Haklay et al. 2021). Some examples of citizen science include databases like the European Alien Species Information Network (EASIN) and the Global Biodiversity Information Facility (GBIF). This information can help tackle scientific questions involving the preservation of marine and terrestrial species (Aceves-Bueno et al. 2017).

There are several benefits of using citizen science as a method of data collection. The first is the immense scale at which data can be collected as volunteers can conduct data collection over large temporal and spatial scales (Burgess et al. 2017). However, it is seen as more complementary way to gather data than the localized, hypothesis-driven data, which has always been the traditional way for data gathering (Dickinson et al. 2010). Secondly, in comparison to the traditional way which demands resources and money, citizen science data is not reliant on any of those components, as its voluntary to input data (Aceves-Bueno et al. 2017). Thirdly, citizen science can also help in preventing project redundancy in a topic of interest that has already been covered (Bonney et al. 2014). This benefits in the long run as this leads to reduced design, testing and implementation expenses, compared to projects starting from "scratch" (Bonney et al. 2014). Additionally, it is also a known tool for gaining public interest and encouraging public engagement which can lead to higher collaboration rates between the scientists and the volunteers (Burgess et al. 2017). Finally, citizen science data can be used to inform and guide decision-making processes, including but not limited to policymaking, environmental management, and conservation efforts, thus leading to more informed and evidence-based policies and practices for future use (Conrad and Hilchey 2011). Decision makers and non-governmental organizations have already shown increased use of citizen science data to monitor and manage natural resources, track endangered species, and conserve protected areas (Conrad and Hilchey 2011; Aceves-Bueno et al. 2017).

While there is increasing interest in the use of citizen science in research, management and educating purposes, there is still uncertainty around the validity of citizen science data. The main concerns around the data collection are the variable level of expertise of the participants in any project involving citizen science (Dickinson et al. 2010). For instance, field assistants conducting research alongside professional scientists, there is a concern around sampling bias in such data, because the assistants vary in ability, expertise, experience, and type of training prior to conducting a task (Dickinson et al. 2010). Therefore, in the process of collecting data, uncertainty can arise (Dickinson et al. 2010). Another source of concern is possible variation in sampling effort among the participants; if the project managers allow any amount of sampling effort, there is a high likelihood of biased sampling. For instance, in projects involving species identification, there are issues with participants who are over-reporting the rare species and under-reporting the common species. In addition, there is also a concern about the observers only reporting the species that they find “interesting” (Dickinson et al. 2010). Lastly there is also the issue with spatial heterogeneity, which in the long run leads to biased estimates (Dickinson et al. 2010). This happens if the sampling sites chosen by the observers are not representative of the true geographic distribution in the population but rather due to the spatial sampling bias (Dickinson et al. 2010).

1.3 Invasive species in Scandinavia

The Scandinavian countries (Norway, Sweden, Denmark) hold large variety of biogeographic regions and different types of marine and coastal ecosystems (Albretsen et al. 2012; Reusch et al. 2018). These ecosystems support various marine and terrestrial life, including wildlife species of commercial importance (Barbier et al. 2011). Moreover, they provide many ecosystem services, including nutrient cycling, bio sequestration, and coastal protection (Barbier et al. 2011). However, there is a growing concern for the introduction of new species due to the drastic environmental changes and anthropogenic disturbances in the area (Pyšek et al. 2010). The European network on alien invasive species (NOBANIS) has recorded that these three countries have observed and identified around 2500 alien species each (Bevanger 2021).

Bevanger et al. (2021) noted that the introduction of invasive alien species in Scandinavia is largely attributed to the surrounding water, particularly through the transport of ballast water and sediments. However, the authors pointed out that there are several other pathways through which invasive species can enter the region. In fact, the study identified 19 potential pathways, including fisheries, transportation, aquaculture, and reintroduction. Therefore, it is essential to consider and address all these pathways to effectively manage the threat of invasive alien species in Scandinavia. Once a species has entered another area and established itself, it can influence its surroundings in both negative and positive ways, depending on the species (Gribben et al. 2013). Some species can even modify the physical and biological aspects of a habitat, often characterized as an ecosystem engineer (Wright and Jones 2006). Additionally, invasive alien species is said to be the main drivers of environmental change and is a threat to ecosystem services (Bevanger 2021). Furthermore, it can be a cause of financial loss in the aquaculture and agricultural sector due to some alien species leading to biodiversity loss in the area (Bevanger 2021). There are also alien species that affect human health as well (Bevanger 2021). Some notable examples of invasive species are red king crab, pacific oyster, and brook trout (Bevanger 2021).

1.4 Pacific oysters in Scandinavia

An important species in the Scandinavian waters is the Pacific oyster (*Magallana gigas*). It is characterized by its resilience to various environmental conditions, high growth, and reproductive rates (Laugen et al. 2015). In Scandinavian waters, Denmark introduced wild populations to the coast through abandoned aquaculture, as the assumption was that the oyster would not survive in the present environmental conditions. The establishment of wild populations were therefore accidental and probably happened sometime after the late 1990s (Laugen et al. 2015). The same could be said for Sweden, as abandoned aquaculture trails introduced oysters to the Swedish coast between 1973 and 1976 (Laugen et al. 2015). Wild populations were observed almost 30 years after in 2007 (Laugen et al. 2015). In Norway, the observations of oysters were recorded as early as 2002 and increased from the Swedish coast in Østfold and to southwestern-Norway in the following years (Laugen et al. 2015).

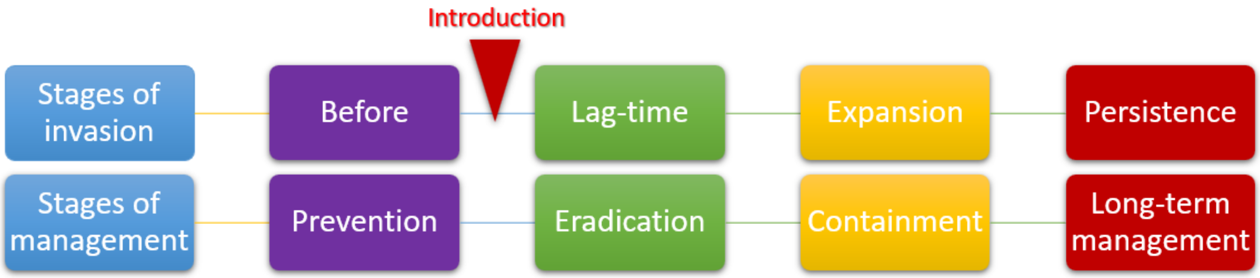


Figure 1. The different stages of invasion and suggested management solutions for non-native species, according to the theory of invasion stages as described by Geburzi and McCarthy (2018).

The spread of the Pacific oyster can therefore be divided into 5 invasion stages (Figure 2), this is based on the study of Geburzi and McCarthy et al (2018). These stages can change and extend over time as the invasion progresses in an area. The first stage is the transportation of the species, where the species in question leaves its native range to another location through several possible pathways (Bevanger 2021). This is classified as the before stage in the different invasion stages. Between the first and second stage is the introduction of the species where the species must be able to establish itself before it can be defined as an invasive species. In the second stage of the invasion, further establishment of the species is dependent on various factors that facilitate the spread of the species, such as the availability of resources in the area and environmental factors. The third stage, also characterized as the expansion stage, starts when the population has had the initial time to reproduce and has now become fully established in its introduction site, it will start to expand its range. The final stage of invasion is called the persistence stage and the species is now a part of the flora or fauna of the area in question. The severity of the impact that follows depends on the type of species and what interactions it has with its biotic and abiotic factors (Geburzi and McCarthy 2018).

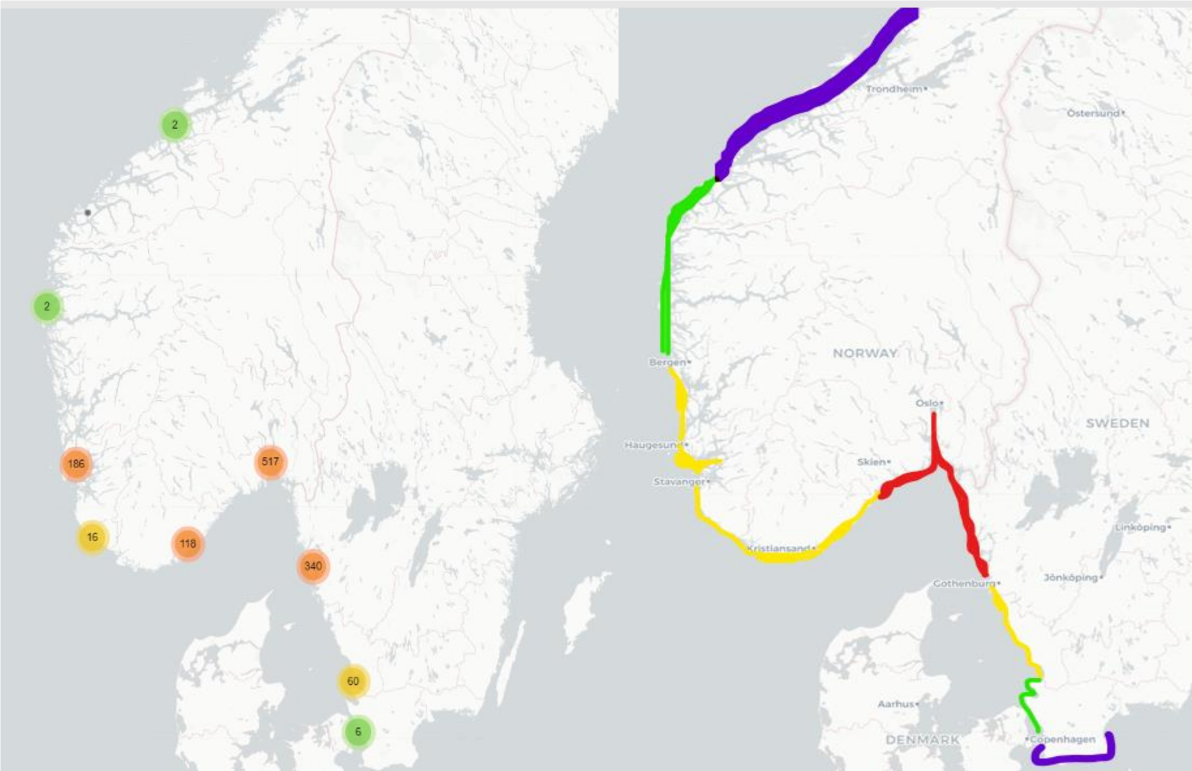


Figure 2. To investigate potential invasion zones in the Swedish and Norwegian coastal area, following visualizations were made. A(left): Citizen science observations of Pacific oysters from the Norwegian (https://www.artsobservasjoner.no/) and Swedish (https://www.artportalen.se/) species databases. The circle number represents the number of observations that were reported in

that area. The colors represent low density in green, intermediate density in yellow and high density in orange. B(right) : Approximate geographic location of the invasion stages of the Pacific oyster in the Norwegian and Swedish coast, as described in Figure 1 with the before stage in purple, lag-time stage in green, expansion stage in yellow and the persistence stage in red.

It is highly important to investigate the range expansion pattern of the Pacific oyster due to its high invasion rate and its ability to change the form and function of its surrounding environment (Wrange et al. 2010). The following species is defined as being an ecosystem engineer, which have the ability to cause major ecosystem changes (Wrange et al. 2010). In particular, the pacific oyster can modify the bottom substrates by forming reefs, and therefore leading to significant reduction in water currents in shallow areas (Wrange et al. 2010). This can cause fouling and changes in fauna (Wrange et al. 2010). Therefore, it is important to find and identify patterns in how fast the pacific oyster is dispersing. One method that can be handy in tracking the expansion is citizen science. Citizen science data can help in identifying when and where the first observations were found. Two sources of citizen science information related to the Pacific oyster in Scandinavia are databases such as the Norwegian Biodiversity Information Center and SLU Swedish Species Information Centre. According to these databases, Scandinavian citizen science data includes in total of 730 entries in Norway and 426 entries of observations in Sweden (Figure *3*).

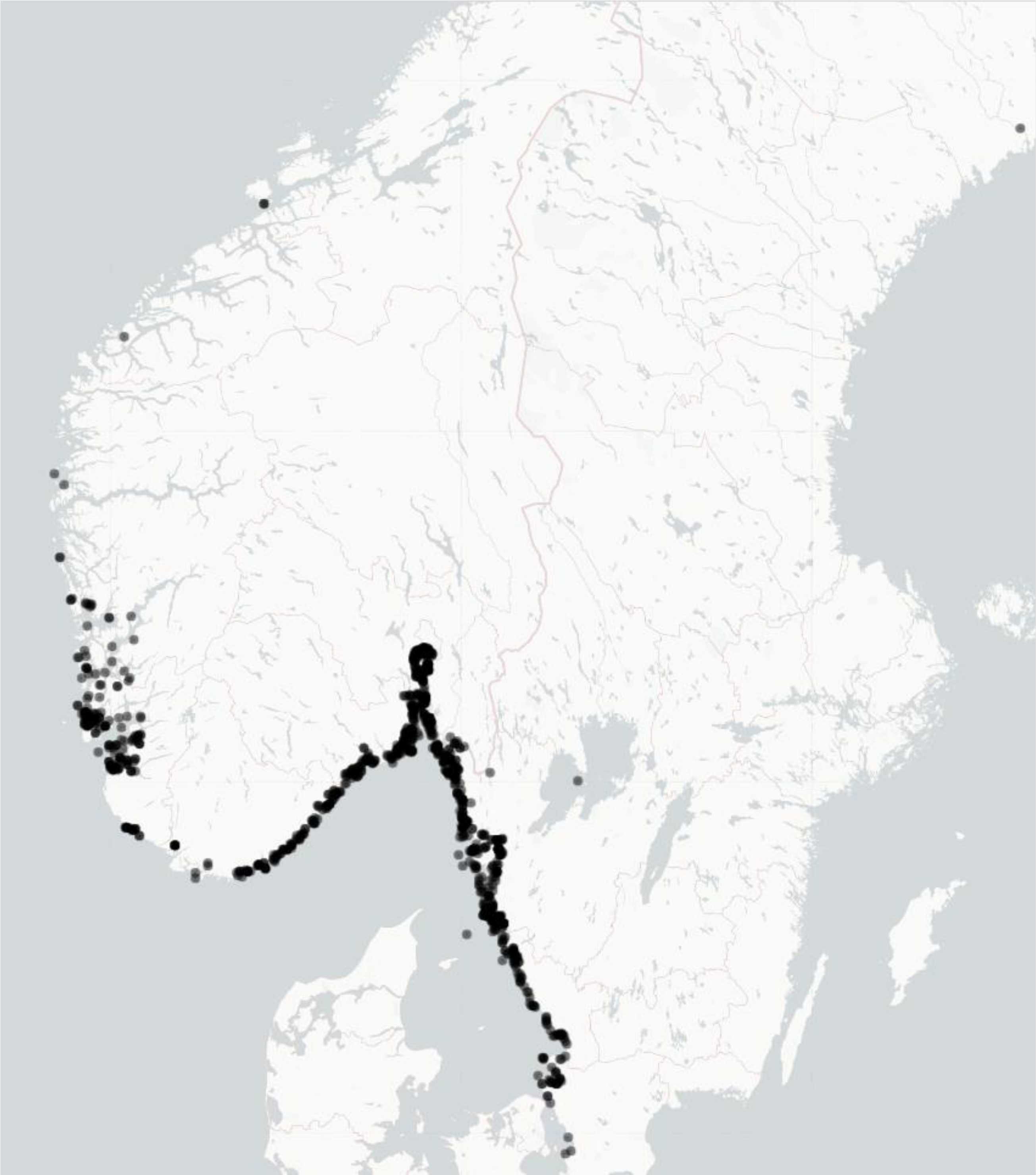


Figure 3. Map showing all Citizen science reports from the two species databases of Pacific oysters in Norway and Sweden from 2001 to the start of January 2023, One point represents a single person reporting the presence of Pacific oysters in one area (Data derived from the webpages: https://www.artsobservasjoner.no/ and https://www.artportalen.se/).

1.5 Aims

The overall aim of this study was to test if citizen science databases are useful for tracking the range expansion of invasive species, using the Pacific oyster as a model system. The aim was further broken down into three specific aims each corresponding to a common issue regarding the quality, validity, and use of species observation databases used by citizen scientists in tracking expansion.

The first aim was to evaluate the quality of the data gathered by citizen scientists (Wiggins et al. 2011) and published the available species observation databases. In general, data cleaning is essential in any raw data, but since this study is trying to evaluate if citizen science can be used to track expansion of an invasive species. It is essential that the reported observations are placed correctly. Otherwise, the false entries can lead to reaching the wrong conclusion about the current expansion rate of the species in question. Moreover, observations being reported in places that are impossible for the species to inhabit logically can give false ranges of the species and its ecology. If applied into other areas of research or management without checking the validity of the data, this can lead to fatal consequences.

Secondly, using Pacific oyster observations and field work, this study aimed to validate the citizen-science-collected species observation data. This was done by investigating if the database was factually correct in what kind of information was input into the database and if there was a geographic bias towards which places people found the Pacific oyster. One factor that was considered when evaluating observations from species databases was if the participants were only reporting on the most easily accessible sites, which would lead to a geographic bias in the data. In such cases, a species range would be skewed towards places that are available to all but wouldn't be representative of the species expansion in total. While there is a validation process in place for many open species observation databases that aim at safeguarding the databases against false or erroneous entries, this study aimed at checking if all observations are input correctly regardless of their verification status.

The third aim was to investigate if the Citizen scientists' databases were detecting and reporting the invasion before professional conducted surveys. If citizen science data can lead to more coverage of the expansion as there are more people conducting and inputting presence of oyster than scientists and managers are able to accomplish, such data can be a useful tool for management and risk-assessment of non-native species.

This project used citizen science data from species observation databases, available literature, and unpublished professional survey data on Pacific oyster densities and locations along the Norwegian and Swedish coasts of the Skagerrak, Kattegat and Øresund (Mortensen et al. 2022). Additionally, randomly generated geographic points were created from raster data containing water depth and wave exposure layers. The geographical range of the study was limited to the range of where the citizen science archive reported observations in the Swedish and Norwegian coastal borders; and where and when the "professional" surveys conducted by researchers were performed.

2 Methods

2.1 Cleaning the citizen science data

In this study, citizen science data was collected from two databases containing species information collected by the public, one database was selected from the Norwegian databases and the other from the Swedish ones. The following databases were chosen: 1) The Norwegian Biodiversity Information Center (https://www.artsobservasjoner.no/) and 2) The SLU Swedish Species Information Centre (https://www.artportalen.se/) (Artsdatabanken 2023; SLU Artdatabanken 2023). All reported observations gathered by the public and some from professional actors are displayed on these websites. It is, however, unclear to which extent this is done for professional observations. All observations of Pacific oysters were exported to an excel file. Additionally, a supplementary database called The Analysis portal for biodiversity data (https://www.analysisportal.se/) was used for the Swedish database to convert coordinates from Swereff coordinate system into standardized latitude and longitude coordinates. The joint database had a total of 1256 reports of Pacific oysters in Norway and Sweden. To get an indication of data quality in the database, observations thought to be false entries were taken out of the dataset manually. The identification of erroneous entries was conducted by inspecting the entered species observation on a map. The criteria for exclusion from the database were 1) occurrences in fresh water, and 2) occurrences on land and sea more than 10 km from the shore. Small deviations from the shoreline were ignored as GPS accuracy varies between devises used when registering observations.

2.2 Validating citizen science observations:

2.2.1 Available and generated data

To verify citizen science observations two different areas, Bohuslän (Sweden) and Agder (Norway) were chosen for field surveys in June 2022 and November 2022, respectively. The locations for the field survey were based on both density of citizen science observations (Figure 2A) as well as approximate invasion zones based on the theoretical invasion curve described by Geburzi and McCarthy et. al. (2018) (Figure 2B). Bohuslän was selected due to being a well-established area for the Pacific oysters and belonging to the last stage of invasion, while Agder was chosen as its in-transition stage between expansion to established stage. A total of 135 sites were surveyed in these two areas, there were 63 randomly generated points and 71 citizen science points that were checked during this period (Figure 4, Figure 5).

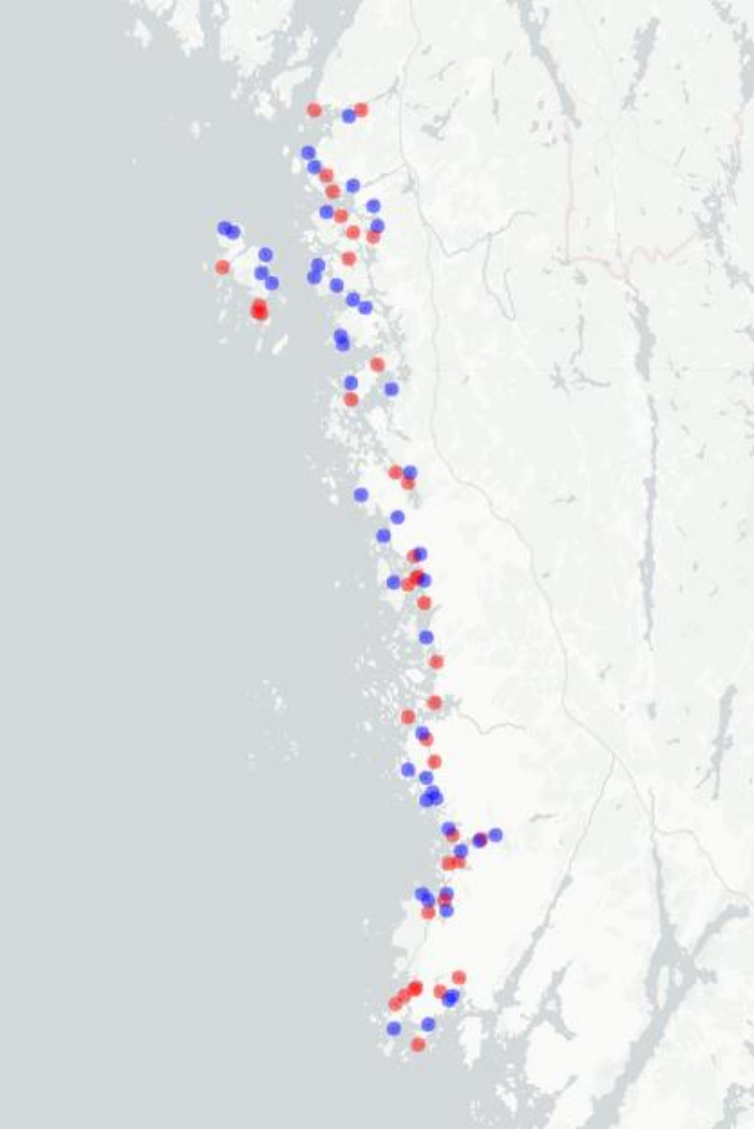


Figure 4. Picture showing the investigated sites for fieldwork in Bohuslän (ranging from Strömstad to Lysekil). The colors represent the randomly generated points in red while blue points represent the verified reports from citizen scientists.

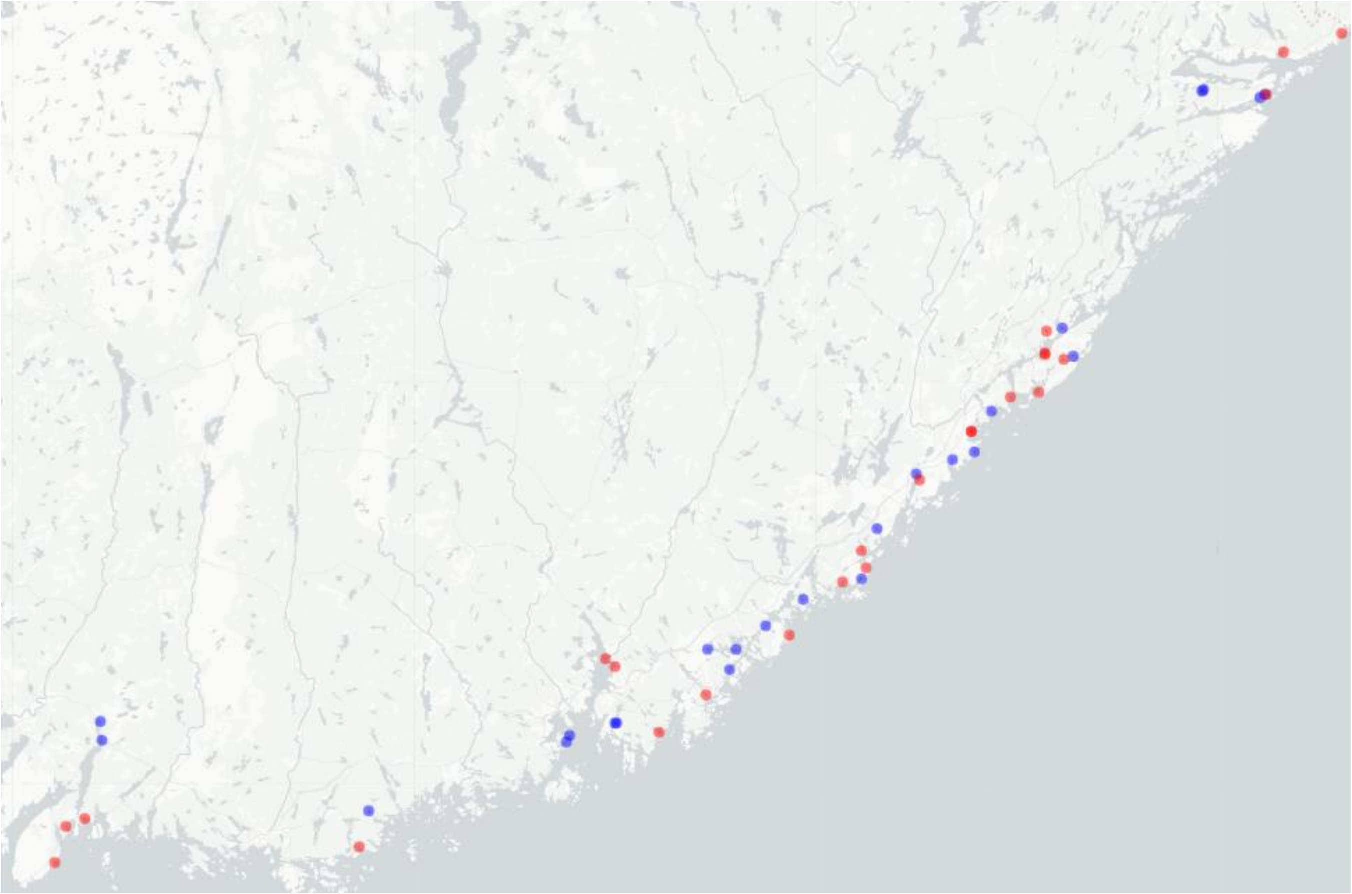


Figure 5. Picture showing the investigated sites for fieldwork.in southern Norway (ranging from Risør to Lindesnes).The colors represent the randomly generated points in red while blue points represent the verified reports from citizen scientists.

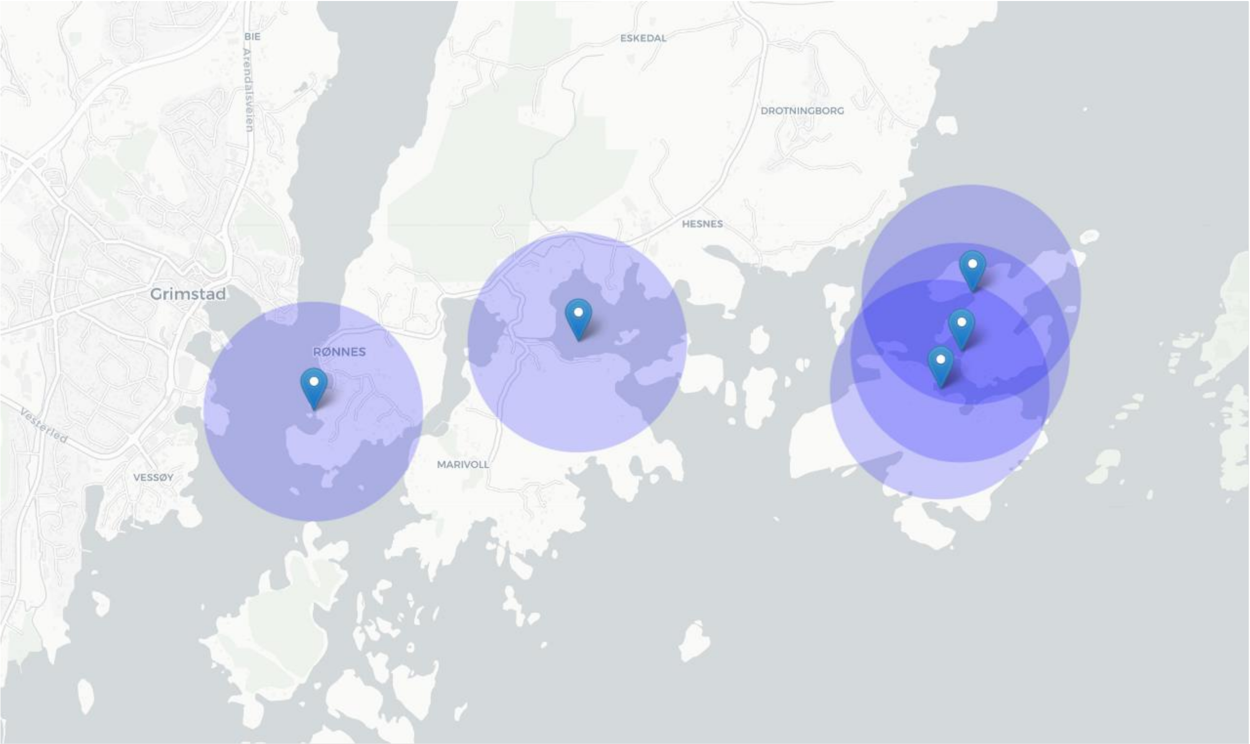


Figure 6. The following map shows blue markers as individual reports collected by citizen scientists, while the buffer around the markers is added as a layer to decide if there was an "hotspot" occurrence in the area. A hotspot was characterized as three or more markers overlapping with each other. With a hotspot identified, the point in the middle of the two points would be chosen to assess Pacific oyster presence in the area. This map shows a visual of how the point selection for citizen science data was conducted during fieldwork.

Within each area, the goal was to visit both areas where Pacific oysters had been observed as well as locations areas without citizen science observations. To ensure that as large an area as possible could be covered in the time available a buffer of 500 m radius to every citizen science observation on the map. If three or more points overlayed, the center point was considered a "Hotspot of points" (Figure 6) and was chosen for visitation. In addition to the citizen science observations chosen for visitation, randomly generated geographic locations were created from raster data of depth and exposure where the Pacific oyster is commonly found (Reamon, Marcussen, Strand and Laugen, unpublished data). The water depth was limited to 10 meters and above, and wave exposure was limited to the range from 'ultra sheltered' to 'moderately exposed' according to the EUNIS classification of exposure (Davies et al. 2004). The randomly generated points were also selected 500 meters away from the citizen science points.

Furthermore, the points were also created to prevent bias against areas that are inaccessible by humans.

Based on the assumption that citizen science observations clusters in highly populated or easily accessible areas, a descriptive factor called "accessibility" was established to characterize potential geographical bias in the citizen science data. This factor was used to determine if a point, either randomly generated or citizen science observation, was easy to access. The accessibility factor had three levels, where level one was the easiest to access and level three was an area which takes active action to access (Appendix B, Figures B.1-B.3). The visited sites were assigned to one of the three accessibility levels based on three criteria, described in detail in Appendix B. Since all but four points were checked by the same people, there is minimal sampling bias in how every area was categorized. For locations which could be put in two or more categories for the different criteria, the overall accessibility of the area and additional criteria such as accessibility by both boat and car, public or private area, or significant hindrances to get to the place, were taken into consideration.

2.2.2 Field procedures

Although the species databases provide the opportunity for citizen scientists to enter information into several available variables, only presence of the Pacific oyster was selected as verification variable as it would be too time-consuming to verify other measurements. For example, the number of oysters reported in the databases varies from one individual oyster to one thousand oysters.

During fieldwork by boat (Bohuslän), a randomly generated point was considered by choosing a number between 0 and 9 and then from its identity number at the end which was given to every point in the database to identify each point, those points were chosen. If there were the same numbers at the end, then the two last numbers would be considered in going to a point for verification. This was done every time a citizen science point was verified, so that sampling was conducted in the same areas as the reporting's from citizen science. For fieldwork conducted by car (Agder), the points were pre-planned using a map and the only criteria was practicality to access these points using a car. Citizen science points, however, were further selected using a pre-planned area that was selected to check for the day, and under this area, any point could be selected. Therefore, some sampling bias was expected by car, as only points that are reachable by car and by foot were checked. For instance, points under a steep cliff or an island had to be exempt from point selection. Randomly generated points that were in the same geographic areas as the citizen science reporting's were chosen. Finally, for points (randomly generated or citizen science) located in between two land areas, the nearest land area in a straight line was checked.

When arriving at a point by boat, one person was in control of the boat while the other searched for oysters using an aquascope. The search for oysters was limited to five minutes, under the assumption that nobody would actively search for oysters. If an oyster was found before the five minutes were up, presence of oysters was registered. However, if no oysters were found during the five minutes, the person checking for oysters would continue searching for the oysters wading along the coast for five more minutes (Figure 7). If nothing was found when wading or from land, absence of Pacific oysters was registered. When the points were sampled using a car, only the wading part was conducted. In addition to the presence and absences of the Pacific oysters, sampling method (boat or car), description of the substrate, accessibility, and comments and topography were noted.

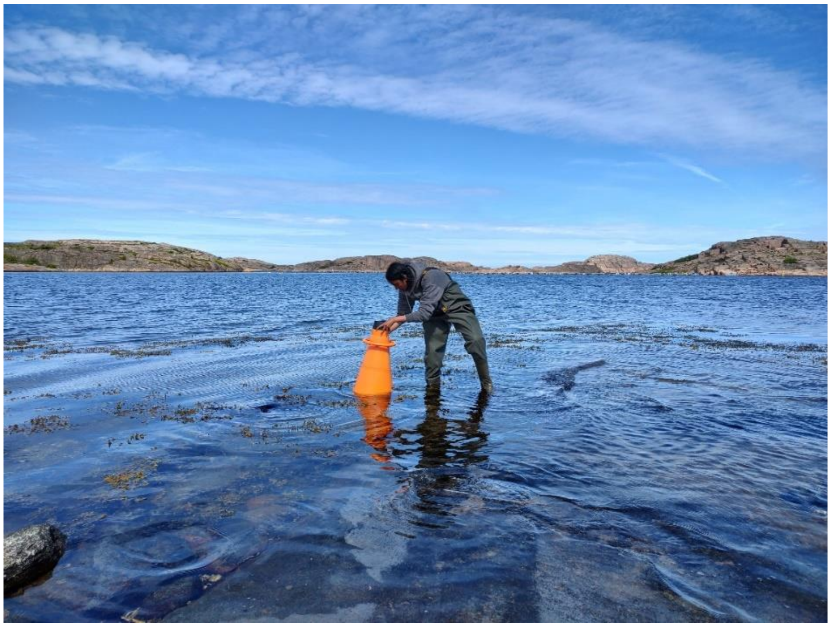


Figure 77. Picture showing the sampling method for wading/walking during fieldwork. Photo: Lars Korslund

2.2.3 Statistical analysis

All data wrangling, visualization, and analyses, were performed with the programming language R version 4.2.2 (R Development Core Team 2022). Using the **dplyr** package (Hadley Wickham 2021), the data was first transformed into a useable data frame. To determine if accessibility determines if citizen scientists observe and report observations in the Swedish and Norwegian species observation systems, a generalized linear model was fitted to the data. The presence of oysters was fitted as the response variable, which consisted of binomial data of 0s (absence) and 1s (presence). Both predictor variables were categorical and consisted of data type (citizen science observations or randomly generated geographical points that were visited for oyster presence during fieldwork), site accessibility (three levels), and two invasion stages (established area Bohusländ and Agder expansion area. The model was further simplified by removing site accessibility, as it was not a significant factor in the modul output.

2.3 Testing the usefulness of citizen science observations

To test if professional surveys or citizen science are reporting invasive species first, professional data comprised of surveys and studies conducted in the same region and in the same time frame as the observations in the citizen science databases was used (Table 1). If the spatial points were in the same place only one data point per year that had been observed first and data type was selected, this was done for both the professional surveys and the citizen science dataset. To determine if citizen science or professional surveys is first in determining the range expansion of Pacific oysters, a generalized linear model was fitted to the data. The number of observations was fitted as response variable. Three explanatory variables were fitted; year (2001 to 2022) was fitted as a continuous explanatory variable, the type of data (citizen science or professional survey) as categorical, and which county the individual observations belonged to as categorical. The model was fitted with Poisson distribution of errors as the number of observations is a count variable. Due to overdispersion the models used for hypothesis testing was fitted with quasipoission distribution of errors.

3 Results

3.1 Data cleaning

The dataset had a total of 1256 reports of Pacific oyster presences in both the Swedish and Norwegian coast. The first two observations overall were reported in Norway in 2001 by the same person. Both observations were in the same county of Rogaland but were recorded in different municipalities named Bokn and Tysvær. In contrast, the first observation in Sweden came some years after, the first one being reported in 2007 in the Västra Götaland county, specifically in the Lysekil municipality. The dataset had only 3 reports of Pacific oysters between 2001 and 2007. Further distribution of citizen science reporting's showed an increase in the number of reports and spatial range increased in the following years (Figure 8).

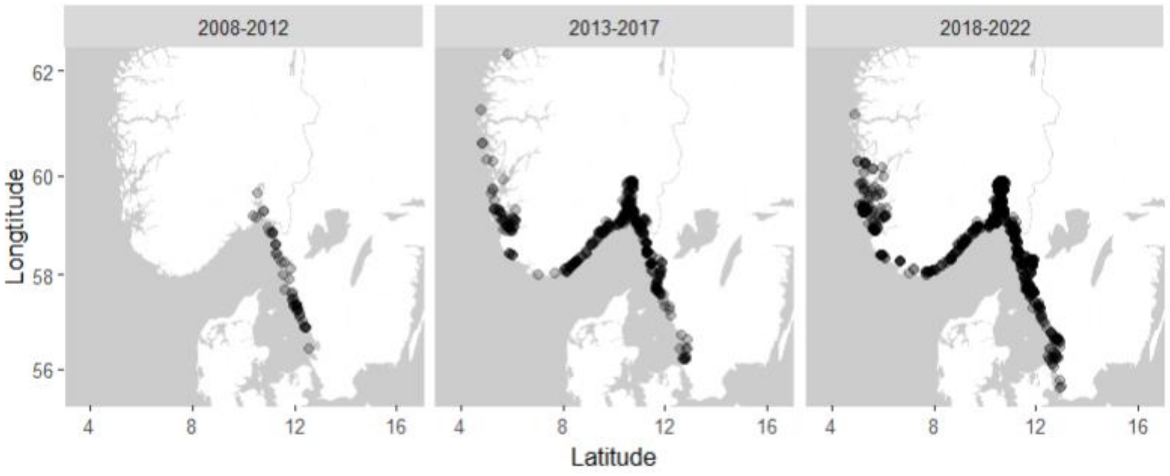
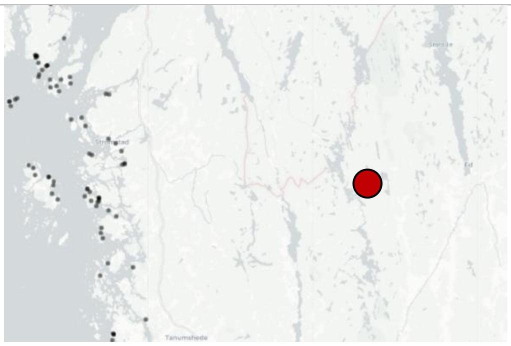
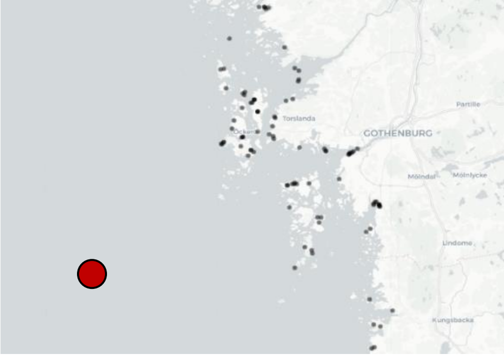
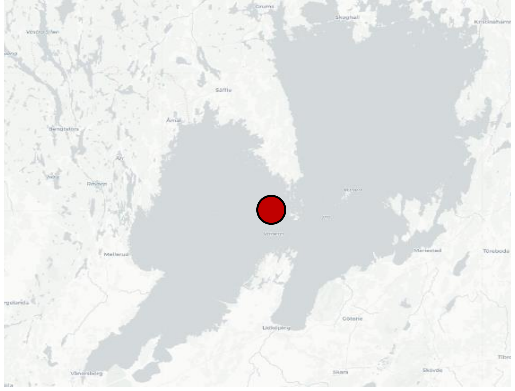


Figure 8. Pacific oyster observations from the Norwegian and Swedish species databases from 2008 to 2022. Data derived from https://www.artsobservasjoner.no/ and https://www.artportalen.se/

Investigating the citizen science observations for false or erroneous entries resulted in removal of 8 observations (Table 1). All points that were removed were in the Swedish region and most of these points were removed due to the points being too far out at sea to be considered a viable sighting of the Pacific oyster (Table 1).

Table 1. Datapoints manually removed from the dataset due to inaccurate or deliberately false entries in the database. Examples of removed points are highlighted in red. ID is the identity number for the observations from the citizen science databases.

ID:	Reason for removal:	Example:
90775919 90775926	Located on land	
90775921 87495648 81450715 90775927	Located too far out at sea	 Torslanda Öckerö GOTHENBURG Partille Mölndal Mölnlycke Lindome Kungsbacka
91666659	Lake (impossible given the ecology of the species)	 Grums Skoghall Västra Silen Säffle Åmål Bengtsfors Mellerud Vänern Mariestad Töreboda Götene Lidköping Skara Skövde Vänersborg

97003745	Umeå (impossible given that the invasion has yet to reach this far out)	Ume Örnsköldsvik

3.2 Validating citizen science data

The statistical analysis showed that the accessibility of a location did not explain any variation in the probability of detecting oyster in neither the citizen science data nor the randomly generated sampling points. The accessibility variable was thus removed from the final model. Both Geography and randomly generated points were shown to be significant in the simplified linear model (Table 2). The scatter plot (Figure 9) showed estimated effects of each explanatory variable (Types and Sampling method) on the response variable (Presence). The scatter plot visualized how the proportion of presence varies across the type of point it is and the sampling method. The results showed that it is higher probability to find presence of Pacific oysters using citizen science reports than random sampling. It is also higher probability to find presence of Pacific oysters in Bohusländ than Agder.

Table 2. Summary of the final generalized linear model estimating the probability of detecting Pacific oysters in two types of Scandinavian coastal locations (citizen science or randomly generated sampling points) and with two different geographical locations (Bohusländ in persistence stage and Agder in Expansion stage)

Coefficients	**Estimate**	**Std. Error**	**z-value**	**P-value**
Intercept	2.2467	0.4053	5.543	< 0.001
Types R	-0.8295	0.4468	-1.857	0.063
GeographyAgder	-1.3809	0.4417	-3.127	0.001

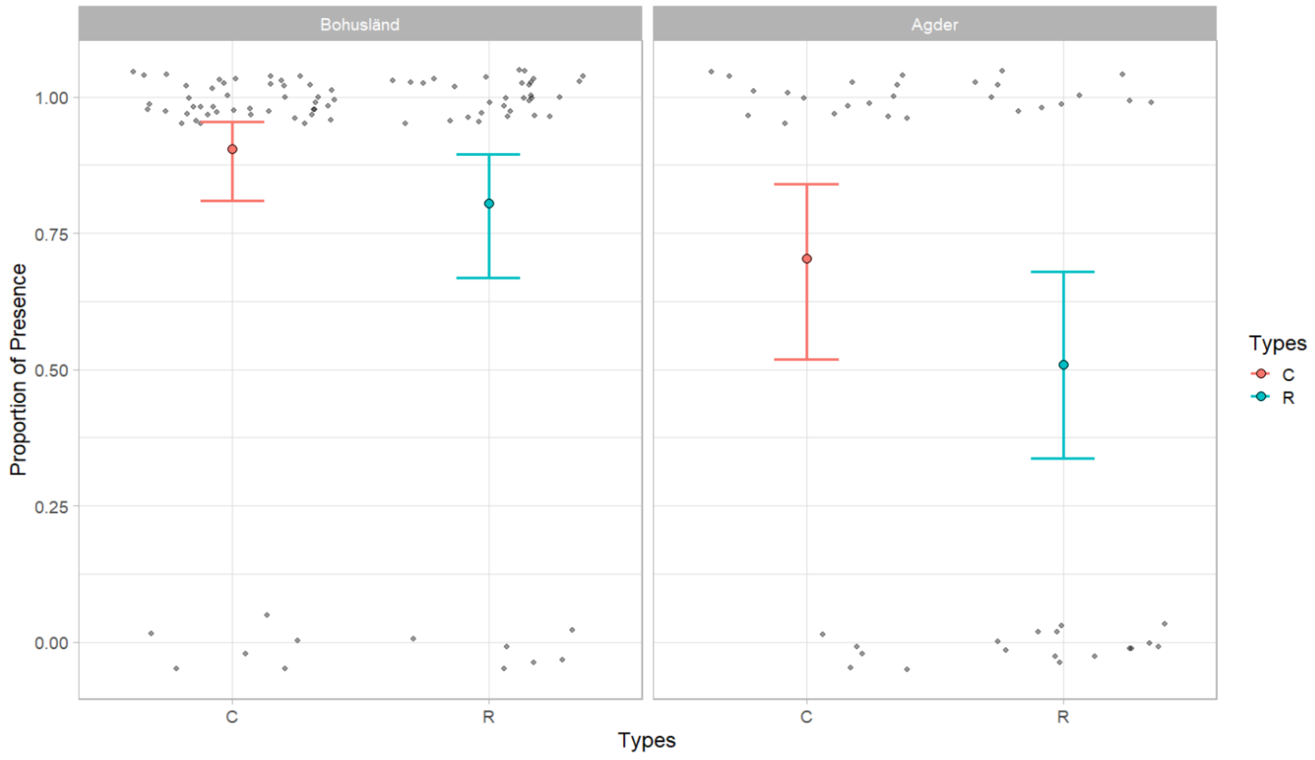


Figure 9. The effect of types of sampling location (citizen science observation, C, and randomly generated points, R) and geographic region on the probability of detecting presence of Pacific oysters in Scandinavian coastal waters. The grey points are raw data (presence or absence), and the error bars are the variation around the predictions from the model (Table.1).

3.3 Comparison of professional surveys with citizen science data

The results showed that there were several observations of Pacific oyster by citizen science recorded over time than professional survey points. The citizen science data also covers more of the geographic range, including the range limits, of the expanding Pacific oyster. Additionally, citizen science data covers larger areas than professional surveys more consistently over time. However, when professional surveys are conducted, they seem to be finding more Pacific oysters than citizen scientists (Figure 10).

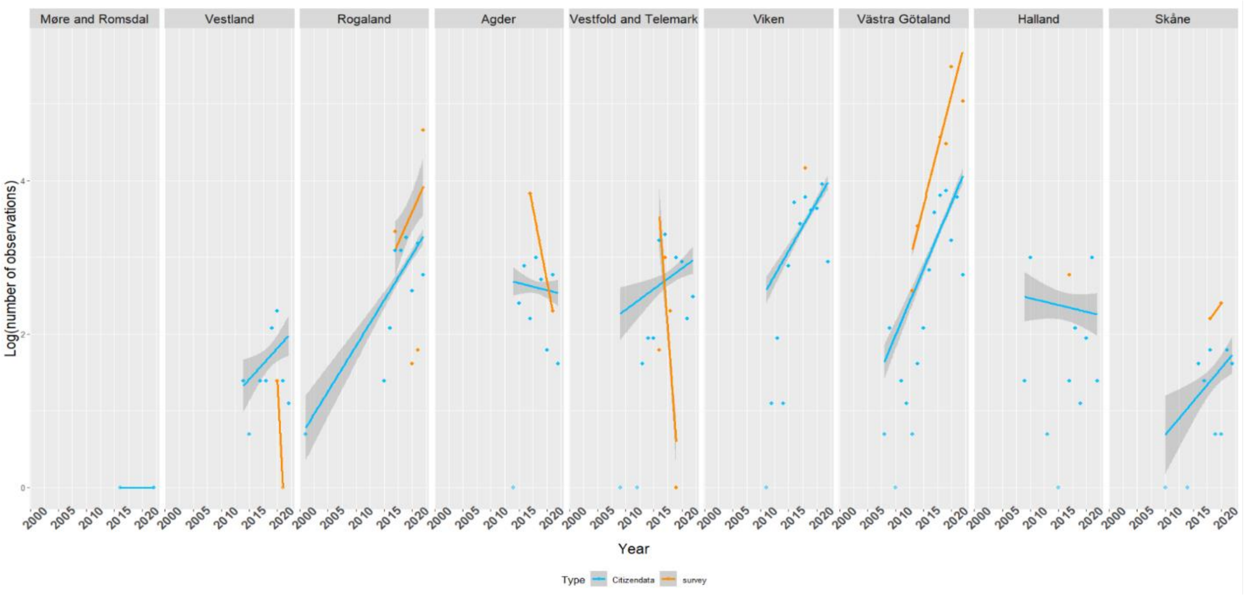


Figure 10. Distribution of Pacific oyster observations in the Norwegian and Swedish citizen science species databases and professionally conducted surveys in space (in geographical order from Møre and Romsdal in the north-west/left to Skåne in the south-east/right) and time (2000 – 2022). The data is log transformed.

4 Discussion

4.1 How much cleaning is necessary?

At first glance, most observations in the citizen science species databases looked reliable (Figure 3). However, by focusing on the individual inputs, there seemed to be some points that didn't quite make sense. Therefore, to achieve a more quality dataset before analysis, some entries were removed from the dataset. One possible explanation for the false entries shown in Table 1 can be wrong input of either latitude or longitude. False entry of longitude values could explain why certain points ended up on land or in the middle of the sea far away from the coast but within the latitude bounds when compared to the other points. This would generate east-west spatial error in the databases. For instance, it is more likely that the one observation in the lake Vänern is explained by false entry of longitude than that the observer was unaware of the habitat which Pacific oysters usually inhabit. There is, however, always a possibility that oysters (live or dead) have been transported to and dumped at this location. Erroneous longitude entries could be detrimental to the analysis of range expansion as it could lead to wrong decisions and conclusions around the impact of the non-native species in both management and conservation. For conservation purposes, false data entries will give bias population density estimates and indicate a larger range preferred environments than what is possible. This especially involves points that were in the deeper sea. According to the literature, these ranges are impossible considering the depth range and preferences of the species which are coastal subtidal and intertidal areas (Diederich 2006). While such false entries will create an impression of a larger range, the biological reality will be a much smaller range. Another point of interest that was taken out had ended up in the city of Umea in Sweden. This is both unexpected and unrealistic, as the invasion of Pacific oysters have yet to reach this part of Sweden. It is thus probably due to false entry of both latitude and longitude. If this single observation was to be taken seriously, the area for early-stage of invasion (Geburzi and McCarthy 2018; Figure 2B) would have been expanded considerably. It is therefore imperative that citizen science data set is investigated, cleaned, and preferably validated, before using it for further analysis of species distributions and applications in management. Lastly, it would be interesting to note that the obviously erroneous entries were only found in the Swedish dataset. One possible explanation other than that the people put in wrong data is the actual structure of the data input in this database. There can be some difficulties with how the data can be reported in with the database not specifying which

GPS system is being used. This is a common factor that can decrease the quality of a dataset, as cited by Balazes et al.(2021), the data quality decreases if there is problem in how a project is designed.

There is overall skepticism and mistrust around the use of citizen science in relation to gathering geographic information (Haklay 2012). The current analysis showed that for tracking the range expansion of Pacific oysters in the Norwegian and Swedish species observation databases are fairly reliable and with most of the observations in an acceptable range and quality. At the same time, a little healthy skepticism about some of the observations is necessary. Indeed, this study shows that the future use of citizen science observations of any species must be filtered through established knowledge about the species to be able to detect spatial anomalies before conducting or reaching conclusions. One solution could be to constrict the ranges that can be entered in the public database. However, this requires advanced knowledge about possible ranges of every species and will thus defy the purpose of public databases where the main aim is to get the help of interested citizens in the collection of data on species´ geographic ranges. Moreover, it is important to remember that geographical positional accuracy might not be relevant for other types of projects, therefore cleaning should be personalized to the data that is being analyzed (Balázs et al. 2021). Either way, using established knowledge about the species to justify cleaning is a good first step in ensuring that the quality of the data is optimized.

4.2 Validation of citizen science data

This study showed that sampled locations based on citizen science observations was a more reliable source of information than randomly generated points within the preferred exposure and depth ranges of the focal species (Figure 9). More specifically, it is more likely to find pacific oysters in places where people have reported the presence of Pacific oysters than in randomly generated points. This shows that reports from non-scientists are indeed generally reliable. Studies of other species such as green crab, Mangrove red snapper and a study tracking various invasive mosquito species found similar evidence of reliability where professionally collected data is in par with the volunteer’s work (Grason et al. 2018; Pernat et al. 2021; Langeneck et al. 2022). Bonney et al. (2014) emphasized that with appropriate protocols, training and oversight, volunteers could collect data with the same quality as the experts. Nonetheless, other studies do not agree with this statement, a review conducted by Wazny (2017) emphasized concerns around using citizen science. These included concerns around generalizability of the reporting's and the possibility of malicious participants in the study. One example of malicious participation is deliberate false entries. However, it is unlikely to cause a major problem in public databases as most of the citizen science reportings showed presence where the participants had reported the species in question. Nevertheless, if citizen science data are to be used for further research or management uses, a discussion around the implications of using the data should be mandatory. Furthermore, with the large sample size that is available through citizen science databases, large differences arising from chance is minimal (Tipton et al. 2017).

Geography was also an important explanatory variable for the variation in detection probability (Table *2*). The model showed that there was a lower probability of finding oysters in Agder region than in Bohuslän. This could be explained by the fact that the non-native species has not fully established in the Agder region yet. This would also support that the distribution of the oysters according to number of reported observations in an area is equal to the true distribution of an invasion stage. According to the non-simplified statistical model, there was no indication of a geographical bias in where people are finding the oysters as we did not find any effect of accessibility. In almost every coastal area where Pacific oysters are expected to be found, the presence of the species was reported by the public. The only areas without observations from the public were coastal wetlands, which were only visited through randomly generated points. Even with the random sampling, there were only two observations in that type of habitat. Moreover, public and private areas were surveyed using citizen science reports. Therefore, citizen science data shows the ability to transcend both formal and informal borders, which can lead to a more accurate scale of research than the traditional method (Wazny 2017).

Only one of the visited citizen science locations was misidentified as having presence of oysters. However, a dead Pacific oyster was found at the site in 2018, so there is a possibility that the individual died between the citizen science observation and the validation survey. This implies that the Pacific oyster was sufficiently familiar to most people, despite this species now being perceived as a well-known non-native species. Due to the higher success rate of simple tasks over those that are more complex, this is logical (Bonney et al. 2014). Most of the "false" entries in this dataset could therefore be assumed to have more to do with the wrong input of coordinates rather than the assumption that people are unaware of the species living spaces or its ecology, since no misidentifications were made. To further strengthen this argument, many of the observed absence sites had dead oysters, which logically implies previous presence of live oysters, unless the oysters or dead shells were transported there and dumped.

Since some of the citizen science observations were outside the currently known habitat preferences, there was an underlying assumption that somebody had misidentified oysters when reporting. The points in question were in areas where substantial freshwater input was expected. During the field surveys, however, several Pacific oysters were found in places where freshwater input was prevalent, showing the oysters range far beyond their officially known capabilities. Therefore, citizen science data can help in setting new expansion ranges for a species; unexpected observations are not necessarily wrong. Additionally, because citizen science observations of Pacific oysters are reliable, such data can be used for several other purposes such as identification of populations for other types of ecological research. This saves the researchers time and resources in finding study sites.

While citizen science data can give many benefits to researchers and people involved in conservations and management efforts (Conrad and Hilchey 2011; Aceves-Bueno et al. 2017) this availability of information can also work against the same goals it is trying to achieve. The goal of the species databases is to map and categorize biotopes and species found in their respective regions. If there is public information on the location of endangered or rare species, there is thus a risk of poaching or illegal wildlife trade of these species, as they can be located and taken. Another passive risk that comes with these public species databases is people wanting to visit sites with endangered species or biotopes and thereby causing unintended disturbances or destruction such as littering, trampling, or picking up species in an area. Lastly, there is also a privacy issue in these databases; both the Norwegian and Swedish databases have public information on who is entering information in the database, which can be problematic if the information is used for targeting or harassing the volunteers.

4.3 Professionally conducted surveys contra citizen science observations

This study showed that the citizen science data stretches over further distances and more years than professional survey data. It also has a larger overall sample size in terms of number of total observations of pacific oysters. Moreover, citizen science data is the first to gather information about the presence of the Pacific oyster in all counties. One possible explanation as to why citizen science has been further along in tracking the invasion is the fact that citizen scientists often are tracking the invasion individually in areas that they inhabit, thus having the advantage of local expertise in the area. This would also give advantage in areas that are low density as they might notice it first. This phenomenon was also observed in a study that conducted research on a predatory fish called *Lutjanus argentimaculatus* in the Mediterranean sea was only reported by the citizen scientists and was completely undetected in the professional fishery studies (Langeneck et al. 2022). However, whenever a professional survey is conducted, there are generally a larger number of observations compared to citizen science (Figure 10). This shows similarity with another study which compared mosquito identification with professional data and citizen science (Pernat et al. 2021). The study found higher species richness for the professional data than citizen science reporting's as they were actively monitoring for the specific species, rather than passive monitoring through citizen science (Pernat et al. 2021). Furthermore, there is a pattern of increase in counts per year for both surveys and citizen science. This could be explained by the fact that with increased coverage of Pacific oysters in traditional and alternative media, higher awareness among the public leads to higher amounts of people reporting to the citizen science databases. Higher awareness among natural resource managers may also lead to more allocation of resources for research and monitoring, thus also explaining the general increase in professional surveys over time.

The overall number of observations in the different counties seem to coincide with the four stages of invasion. There is a clear pattern in counts of observations decreasing from middle of the geographic range to the end on both sides. Vestfold and Telemark ,Viken, and Västra Götaland have the highest number of reports of Pacific oysters from citizen scientists and are also considered as belonging to the last stage of invasion, the so-called persistence stage (Geburzi and McCarthy 2018). These three counties also show an increase over time in citizen science observations. One of the counties, Västra Götaland, also has the highest number observations from currently known professional surveys. One explanation may be the large presence of marine research institutions that are heavily invested in tracking the expansion of these oysters (Mortensen et al. 2022) in the area. This leads to the first observation of professional surveys already being recorded in 2007 due to early investment in research efforts in the area (Faust et al. 2017). Another contributing factor may be that this county has the largest coastal area with environmental conditions relevant for the Pacific oyster in Sweden (Laugen et al. 2015). Viken only had one available professional survey, which could be due to bias in available surveys, or that there is not enough sampling effort from surveys. The survey in question is a rapport from Institute of marine research in Norway that surveyed the biomass and presence of Pacific oyster from the year 2017 to 2019 (Jelmert et al. 2020). This survey was also available in the citizen science database; therefore, it is important to consider the fact that even professionals input their sightings into the database. This naturally leads to more reports into the citizen science database.

Halland and Agder have a trend of decreasing the number of citizen science observations through the years, while the number of observations seems to be increasing for Rogaland in both survey and citizen science. This area represents the expansion stage of invasion. There are only a few surveys conducted in this area. The first two citizen science observations come from Rogaland in 2001, but no comparable data from professional surveys exist. A rapport from Norwegian Environment Agency notes that observations in the early 2000s time period could be explained by abandoned aquaculture facilities in the area (Bodvin et al. 2010).

Vestland and Skåne are both in the second stage of invasion. This is not surprising as the distribution of the oyster according to literature found presences of oyster up to 60°N in 2009 and further down south in Kattegat, Sweden (Wrange et al. 2010). They correlate well with the presences being found by the public during the same years (Figure 10). The northernmost citizen science observation is from Smøla in Møre and Romsdal. This area is in the first stage of the invasion. No professionally conducted surveys have to date been conducted in this county, which further indicates that citizen scientists are leading the charge in documenting the range expansion of Pacific oysters. As also pointed out by a study conducted on green crabs expansion with use of citizen science (Grason et al. 2018), citizen science is contributing to both early monitoring of a species and covering large spatial areas that are not necessarily available for government and academic monitoring alone. This provides support to the hypothesis that citizen science has observed Pacific oysters' range expansion faster than professional surveys that are generally limited to areas further south in Norway. The range expansion according to citizen science thus seems to be a good indicator of how the Pacific oyster is increasing rapidly and can be a valuable tool for various conservation and management efforts in coastal areas.

4.4 Study design (choices and limitations)

For this specific study, Pacific oyster was chosen as the model system for two main reasons. Firstly, the study design relies on revisiting sites where a species was last observed according to the citizen science data, and the Pacific oyster gives us the opportunity to revisit those sites, as they have a sessile lifestyle after settling, unless the oysters are removed manually. Secondly, Pacific oysters are also quite resilient to external factors and are relatively long-lived, which makes it possible to observe this species even after a significant period (Laugen et al. 2015; Takeuchi et al. 2016). The second step in the study design was to figure out how to validate the citizen science data. Many studies have tried to investigate if the data gathered from citizen science resources are reliable and qualitative analysis and field surveys are common ways for determining reliability and validity of data (Wiggins et al. 2011; Aceves-Bueno et al. 2017). A major reoccurring problem with citizen science data is that the data-sets are often not comparable, due to a lack of a control group (Alabri and Hunter 2010; Aceves-Bueno et al. 2017). This study avoided this problem by generating points in the same area as the citizen science points that were re-checked through the field survey.

Thirdly, the study design for field procedures was based upon imitating passive monitoring. The assumption was as follows; the people that are reporting in presences are not actively looking for oysters, rather visiting these sites randomly. Therefore, the time limit was set to 5 minutes. This seemed to be sufficient time to detect an oyster's presence in the area. However, there was an underlying concern about the points in areas that weren't limited to how far it was appropriate to check in our timeframe of 5 minutes around a point. Both because of the accuracy of the spatial points input in the dataset can vary geographically and the possibility for concluding a false absence . This became apparent when arriving at a point where it was concluded an absence in pacific oyster population in the five minutes that the area was surveyed, yet 50 meters to the other side of the island, live oysters were spotted from the boat. More sampling time would have resulted in covering a larger area, which could be more advantageous for areas with low density of the oysters. This would indicate that the time for checking each point should be increased if the experiment was to be repeated, to accommodate the extra time it takes to locate oysters in low-density areas and thus to avoid false negatives, but also be aware that the sampling time should imitate real observers and not turn the sampling into an active monitoring process. Fourthly, the choice to make randomly generated data assumed that not everyone is visiting the whole ecological range of the Pacific oysters, while there was no accessibility bias found in the model, the study did find one habitat that was not sampled by citizen science, this could indicate some bias towards certain habitats rather than accessibility areas.

Furthermore, for the field procedure, there were certain aspects that needed to be discussed, for example an aquascope allows the user to look underwater and is highly dependent on light. Therefore, almost all observations were conducted during daylight hours apart from one observation site that used a flashlight to sense the presence of Pacific oyster. In any case, it is also important that the boat has a steady pace when the other person is checking for oysters. If the boat is too fast, observing the oyster becomes more difficult as sediment resuspension makes it harder to see the water clearly. In addition, sometimes putting on a wading suit wasn't necessary as the oysters could be observed from land, either because of the shallowness of the water, or if oysters were firmly settled on a bridge (normally a wooden or cement bridge by the coast). Another concern was that after several times of testing sites on different days with different weather, it became clear that the sampling design is indeed dependent on clear weather. While it was not impossible to observe oysters in rainy or windy weather, it posed challenges in the field. Along with these factors, three things had to be checked when arriving by boat to check for oysters. The first being the intensity of tides, the second being the general direction of the wind and lastly the level of exposure in the area. One randomly generated point was unavailable due to the moderately high exposure area, with tide and wind direction going towards the bay, which would have made it impossible for the boat to get out. Tidal and wind direction are crucial to check before making the boat float along the area, this ensures that an area can be checked safely, and the boat is not thrown further away or towards land. Some points were skipped due to these factors.

In retrospect, certain variables could have been added to the field protocol. The first is noting which sites had been subjected to oyster removals. There were locals present during the field surveys who talked about annual cleaning in the area, either performed individually or as a community because they expressed the need to keep their recreational areas free of Pacific oysters to prevent injury. The second is noting the number of dead oysters in the area, which would have indicated a previous presence, but was still counted as an absence since invasive species expansion only looks at active invasion of the species. Therefore, this should be taken into consideration if the research was to be replicated. This study could also be improved by implementing raster layers to create an accessibility variable that was spatially generalized for all areas. That way the uncertainty of geographical bias or accessibility could have been checked in a more objective way than was possible in this study. For instance, as observed in field, some locations can be harder to access in real life than it looks on a map. In addition, the two countries have relevant data layers that are free and available for use.

4.5 Further work and recommendations

Further work should focus on improving data collection systems in citizen science databases where anyone can enter data, therefore decreasing the number of erroneous data entries. If similar studies are to be conducted for other model species, it is important to be aware of the biological characteristics of a species and therefore making a study design suited to the individual species rather than a generalized interpretation of this study design. This study has proved that citizen science data is indeed reliable for the current model species and probably also for species with similar habitat preferences. Further research should focus on implementing citizen science data as an extension to already established data. As previously mentioned, citizen science data can cover higher rates of temporal and spatial areas making it possible to investigate scientific questions in a larger picture and can also contribute to achieving a higher sample size when looking for ecological patterns in different species.

To improve the species databases evaluated in this study, GPS systems should be standardized. If there is too much deviation between the devices and sensors used to gather data, the information can be placed incorrectly or misreported (Balázs et al. 2021).While there were very few false entries for the chosen species, it would be premature to assume the same for all the others. Standardization would improve the overall accuracy of the data. Another thing would be to give volunteers certifications and badges to further promote them to continue contributing to the program, as told by Crall et al. (2011). Furthermore, to improve availability of use for this data there should be options to change the coordinate system when exporting this data, as different coordinate systems make it cumbersome to merge data sets.

There should also be more focus on mitigating unintended uses for citizen science data. It would be appropriate to prevent unintended consequences to the species and biotopes by making the public aware of the ethical implications of such data, but also protect sensitive information that could potentially harm endangered species in the long run. Bowser et.al (2014) suggested that certain reports and locations could be hidden. As mentioned earlier, there is also a privacy issue in these datasets, anyone can see where and who sampled in the database. To protect the samplers, there should be requirement of minimum personal data collected from the samplers (Bowser et al. 2014), rather than having a public name-list of samplers there should be individual sample ID codes in the form of numbers to protect the identities of the samplers.

4.6 Conclusion

In conclusion, there was an appropriate amount of evidence to suggest that citizen species databases such as the Norwegian Biodiversity Information Center and the SLU Swedish Species Information Centre are reliable enough to spatially evaluate invasion of a species like the Pacific oyster. Further research should, however, implement stricter rules for data validation through field campaigns and also be aware that each species has its own characteristics. As this study has shown general reliability of citizen science data, this data collection tool should be used to implement better ways to improve research areas with big knowledge gaps and reach larger geographic areas than what is possible from publicly funded research and monitoring.

5 References

Aceves-Bueno E, Adeleye AS, Feraud M, Huang Y, Tao M, Yang Y, Anderson SE. 2017. The accuracy of citizen science data: a quantitative review. *Bulletin of the Ecological Society of America* **98**: 278-290.

Alabri A, Hunter J. 2010. Enhancing the quality and trust of citizen science data. in *2010 IEEE sixth international conference on e-science*, pp. 81-88. IEEE.

Albretsen J, Aure J, Sætre R, Danielssen DS. 2012. Climatic variability in the Skagerrak and coastal waters of Norway. *ICES Journal of Marine Science* **69**: 758-763.

Artsdatabanken. 2023. Artsobservasjoner

Ashqar BA, Abu-Naser SS. 2019. Identifying images of invasive hydrangea using pre-trained deep convolutional neural networks. *International Journal of Academic Engineering Research (IJAER)* **3**: 28-36.

Balázs B, Mooney P, Nováková E, Bastin L, Arsanjani JJ. 2021. Data quality in citizen science. *The science of citizen science* **139**.

Barbier EB, Hacker SD, Kennedy C, Koch EW, Stier AC, Silliman BR. 2011. The value of estuarine and coastal ecosystem services. *Ecological monographs* **81**: 169-193.

Bevanger K. 2021. Invasive Alien Species in Scandinavia. *Invasive Alien Species: Observations and Issues from Around the World* **3**: 1-41.

Bodvin T, Norling P, Smit A, Jelmert A, Oug E. 2010. Mulige effekter av etablering av stillehavsøsters (Crassostrea gigas) i Norge. *Norwegian Directorate for Nature Management, DN-utredning*: 1-2010.

Bonney R, Shirk JL, Phillips TB, Wiggins A, Ballard HL, Miller-Rushing AJ, Parrish JK. 2014. Next steps for citizen science. *Science* **343**: 1436-1437.

Bowser A, Wiggins A, Shanley L, Preece J, Henderson S. 2014. Sharing data while protecting privacy in citizen science. *interactions* **21**: 70-73.

Burgess HK, DeBey L, Froehlich H, Schmidt N, Theobald EJ, Ettinger AK, HilleRisLambers J, Tewksbury J, Parrish JK. 2017. The science of citizen science: Exploring barriers to use as a primary research tool. *Biological Conservation* **208**: 113-120.

Conrad CC, Hilchey KG. 2011. A review of citizen science and community-based environmental monitoring: issues and opportunities. *Environmental monitoring and assessment* **176**: 273-291.

Darling JA, Blum MJ. 2007. DNA-based methods for monitoring invasive species: a review and prospectus. *Biological Invasions* **9**: 751-765.

David P, Thebault E, Anneville O, Duyck P-F, Chapuis E, Loeuille N. 2017. Impacts of invasive species on food webs: a review of empirical data. *Advances in ecological research* **56**: 1-60.

Davies CE, Moss D, Hill MO. 2004. EUNIS habitat classification revised 2004. *Report to: European environment agency-European topic centre on nature protection and biodiversity*: 127-143.

Dickinson JL, Zuckerberg B, Bonter DN. 2010. Citizen science as an ecological research tool: challenges and benefits. *Annual review of ecology, evolution, and systematics* **41**: 149-172.

Diederich S. 2006. High survival and growth rates of introduced Pacific oysters may cause restrictions on habitat use by native mussels in the Wadden Sea. *Journal of Experimental Marine Biology and Ecology* **328**: 211-227.

Elewa AM, Abdelhady AA. 2020. Past, present, and future mass extinctions. *Journal of African Earth Sciences* **162**: 103678.

Faust E, André C, Meurling S, Kochmann J, Christiansen H, Jensen LF, Charrier G, Laugen AT, Strand Å. 2017. Origin and route of establishment of the invasive Pacific oyster Crassostrea gigas in Scandinavia. *Marine Ecology Progress Series* **575**: 95-105.

Geburzi JC, McCarthy ML. 2018. How do they do it?–Understanding the success of marine invasive species. in *YOUMARES 8–Oceans Across Boundaries: Learning from each other: Proceedings of the 2017 conference for YOUng MARine RESearchers in Kiel, Germany,* pp. 109-124. Springer International Publishing.

Grason E, McDonald S, Adams J, Litle K, Apple J, Pleus A. 2018. Citizen science program detects range expansion of the globally invasive European green crab in Washington State.

Gribben PE, Byers JE, Wright JT, Glasby TM. 2013. Positive versus negative effects of an invasive ecosystem engineer on different components of a marine ecosystem. *Oikos* **122**: 816-824.

Hadley Wickham RF, Lionel Henry, Kirill Müller. 2021. dplyr: A Grammar of Data Manipulation (version 1.1.1).

Haklay M. 2012. Citizen science and volunteered geographic information: Overview and typology of participation. *Crowdsourcing geographic knowledge: Volunteered geographic information (VGI) in theory and practice*: 105-122.

Haklay MM, Dörler D, Heigl F, Manzoni M, Hecker S, Vohland K. 2021. What is citizen science? The challenges of definition. *The science of citizen science* **13**.

Huang D, Haack RA, Zhang R. 2011. Does global warming increase establishment rates of invasive alien species? A centurial time series analysis. *PloS one* **6**: e24733.

Jelmert A, Espeland SH, Ohldieck MJ, van Son TC, Naustvoll LJ. 2020. Kartlegging av Stillehavsøsters (Crassostrea gigas)-Bestandskartlegging Karmøy-Svenskegrensa 2017-2019. *Rapport fra havforskningen*.

Juanes F. 2018. Visual and acoustic sensors for early detection of biological invasions: Current uses and future potential. *Journal for nature conservation* **42**: 7-11.

Koh LP, Dunn RR, Sodhi NS, Colwell RK, Proctor HC, Smith VS. 2004. Species coextinctions and the biodiversity crisis. *science* **305**: 1632-1634.

Langeneck J, Minasidis V, Doumpas N, Giovos I, Kaminas A, Kleitou P, Tiralongo F, Crocetta F. 2022. Citizen Science Helps in Tracking the Range Expansions of Non-Indigenous and Neo-Native Species in Greece and Cyprus (Eastern Mediterranean Sea). *Journal of Marine Science and Engineering* **10**: 256.

Larsen BB, Miller EC, Rhodes MK, Wiens JJ. 2017. Inordinate fondness multiplied and redistributed: the number of species on earth and the new pie of life. *The Quarterly Review of Biology* **92**: 229-265.

Laugen AT, Hollander J, Obst M, Strand Å. 2015. The Pacific oyster (*Crassostrea gigas*) invasion in Scandinavian coastal waters: impact on local ecosystem services. *Biological invasions in aquatic and terrestrial systems: biogeography, ecological impacts, predictions, and management De Gruyter Open, Berlin*: 230-252.

Millennium ecosystem assessment M. 2005. *Ecosystems and human well-being*. Island press Washington, DC.

Mortensen S, Timenes Laugen A, Strand Å, Dolmer P, Naustvoll L-J, Jelmert A, Albretsen J, Broström G, Gustafsson M, Durkin A. 2022. Stillehavsøsters i Norden: Datainnsamling og bestandsvurderinger som grunnlag for forvaltning og høsting av nordiske bestander av stillehavsøsters, *Crassostrea gigas*. Nordisk Ministerråd.

Pecl GT, Araújo MB, Bell JD, Blanchard J, Bonebrake TC, Chen I-C, Clark TD, Colwell RK, Danielsen F, Evengård B. 2017. Biodiversity redistribution under climate change: Impacts on ecosystems and human well-being. *Science* **355**: eaai9214.

Pejchar L, Mooney HA. 2009. Invasive species, ecosystem services and human well-being. *Trends in ecology & evolution* **24**: 497-504.

Pernat N, Kampen H, Jeschke JM, Werner D. 2021. Citizen science versus professional data collection: Comparison of approaches to mosquito monitoring in Germany. *Journal of Applied Ecology* **58**: 214-223.

Pyšek P, Hulme PE, Simberloff D, Bacher S, Blackburn TM, Carlton JT, Dawson W, Essl F, Foxcroft LC, Genovesi P. 2020. Scientists' warning on invasive alien species. *Biological Reviews* **95**: 1511-1534.

Pyšek P, Jarošík V, Hulme PE, Kühn I, Wild J, Arianoutsou M, Bacher S, Chiron F, Didžiulis V, Essl F. 2010. Disentangling the role of environmental and human pressures on biological invasions across Europe. *Proceedings of the National Academy of Sciences* **107**: 12157-12162.

R Development Core Team. 2022. R: A language and environment for statistical computing. R Foundation for Statistical Computing

Reusch TB, Dierking J, Andersson HC, Bonsdorff E, Carstensen J, Casini M, Czajkowski M, Hasler B, Hinsby K, Hyytiäinen K. 2018. The Baltic Sea as a time machine for the future coastal ocean. *Science Advances* **4**: eaar8195.

Seebens H, Bacher S, Blackburn TM, Capinha C, Dawson W, Dullinger S, Genovesi P, Hulme PE, van Kleunen M, Kühn I. 2021. Projecting the continental accumulation of alien species through to 2050. *Global Change Biology* **27**: 970-982.

Singh J. 2002. The biodiversity crisis: a multifaceted review. *Current Science*: 638-647.

SLU Artdatabanken. 2023. Artportalen

Takeuchi T, Koyanagi R, Gyoja F, Kanda M, Hisata K, Fujie M, Goto H, Yamasaki S, Nagai K, Morino Y. 2016. Bivalve-specific gene expansion in the pearl oyster genome: implications of adaptation to a sessile lifestyle. *Zoological letters* **2**: 1-13.

Tipton E, Hallberg K, Hedges LV, Chan W. 2017. Implications of small samples for generalization: Adjustments and rules of thumb. *Evaluation review* **41**: 472-505.

Ustin SL, DiPietro D, Olmstead K, Underwood E, Scheer GJ. 2002. Hyperspectral remote sensing for invasive species detection and mapping. in *IEEE International Geoscience and Remote Sensing Symposium*, pp. 1658-1660. IEEE.

Wazny K. 2017. "Crowdsourcing" ten years in: A review. *Journal of global health* **7**.

Wiggins A, Newman G, Stevenson RD, Crowston K. 2011. Mechanisms for data quality and validation in citizen science. in *2011 IEEE seventh international conference on e-Science Workshops*, pp. 14-19. IEEE.

Wrange A-L, Valero J, Harkestad LS, Strand Ø, Lindegarth S, Christensen HT, Dolmer P, Kristensen PS, Mortensen S. 2010. Massive settlements of the Pacific oyster, *Crassostrea gigas*, in Scandinavia. *Biological Invasions* **12**: 1145-1152.

Wright JP, Jones CG. 2006. The concept of organisms as ecosystem engineers ten years on: progress, limitations, and challenges. *BioScience* **56**: 203-209.

Appendix A: Protocol for Pacific oyster presence, field survey 2022

There are certain criteria's for checking a citizen science point or random point:

1) If the point is on land and not very close on the coastal side according to the visual map, we do not check the point as it can become too vague and up to interpretation where the point should have been.
2) If the point is on sea in between two potential land-areas, we chose the site closest to the land-area.

When arriving to a point:

1) Assess whether it is possible to reach the point by boat and wading(this also involves if you can see the oysters by walking on land).
2) If it's only possible to sample by one , then only sample with one sampling method.
3) The protocol suggests you take 5 minutes searching from the boat, if nothing is found under those 5 minutes, you put on a wading suit and check 5 minutes from land.

Sampling protocol:

1) Write the number of the observation (ID)
2) Write the substrate description (S), example: boulder/hard substrate if the shell is attached to a rock, or clay/mud. Write how the substrate is with or without finding presence.
3) Write presence or absence of the pacific oyster by using 0 and 1, B1 or B0 (for sampling by boat), W1 or W0 (for sampling by wading). If one sampling method is unavailable in a certain area, do not write W or B and then explain in (C) for comments why it is unavailable. If you find it by boat and know the sampling is possible by wading write B1W1.
4) Write down the time it took to find something 5 min timer for both wading and by using the boat, example: B1(4,5 min)
5) Write (Grason et al.) = (The T stands for "Tilgjenglihet", which translates to accessibility in English) / The L stands for landscape), for availability in terms of sampling method, the T is divided into three groups of categories, one being easily available to sample and observe, while three being the most difficult place to get to and observe and sample the oysters. L in (Grason et al.) stands for landscape , this is for writing down why you chose

the number, for example a beach is an easy place to access , therefore (Beach 1). While a steep cliff going straight down would make the accessibility highly difficult putting it at a category 3, therefore (Cliff 3). The landscape description is also not just limited to geographical/biological structures, if a pacific oyster is found on a bridge or somebody’s backyard, write (House 1) or (Bridge 1).

Criteria for accessibility are as follows:

1) Look at the terrain, easy flat terrains are high access areas whereas rocky or cliff structures make it harder to access. Any significant hindrance to entering the site can decrease accessibility.
2) Building structures nearby increase accessibility.(Includes bridges, houses , cabins, roads, trails)
3) Passive or active sighting chance for the oysters in the area, do you think people have high level of accessibility to the area. For instance, any beach will be an area that is easy to arrive to while wetland areas are harder to access.

6) Write the time of the sampling with (Ti)
7) Use (C) for comments or unusual circumstances.

Appendix B: Criteria for determining accessibility

Terrain. The terrain could determine how difficult it was to have access to a certain area. The first thing was to observe if the terrain was mainly flat or if there were a certain degree of slopes or obstacles in the area. Big boulders or stones in the water and the coast was considered greater difficult than pebbles and gravel in the area. The level of mud in the surrounding area also greatly affected how difficult it was to walk in the area. As a result, the areas in question became less accessible. Concrete example of these terrain differences were flat terrain sites like beaches which were considered under category one as these were the easiest to access in terms of walking, wading, and swimming; these points were also easily accessible by boat as there were no obstacles in the water. Category two was anything in between very easy and impossible in terms of accessibility. Typically, this would be terrains like islets or any point with significant access difficulties. Category three were the hardest points to reach, and cliffs were categorized as that because you can only reach them by swimming in the cliff area or checking from a boat.

Man-made structures. If there were any man-made building structures such as houses, roads, and bridges, in the area, the assumption is that the point is easier to access.

General accessibility. If the participants were actively or passively coming to an area, for instance a trail to the observation indicated human activity and sign of people passing by. The proposed idea behind it was the question "Could people venture through a place like this voluntarily unless they were actively seeking the Pacific oyster? ” Beaches and bridges were put under category one, as most people actively coming to a beach or a bridge were most likely not trying to find an oyster, rather passively observing the oyster through swimming, taking out their boats or walking on the bridge/harbor (Figure B.1). Furthermore, as the oysters are sharp and hard to set foot on, they would be easier to notice on beaches. Wetland areas however were classified as category three since people rarely venture through these areas voluntarily if they are walking through forests or driving on a boat (Figure B.3).

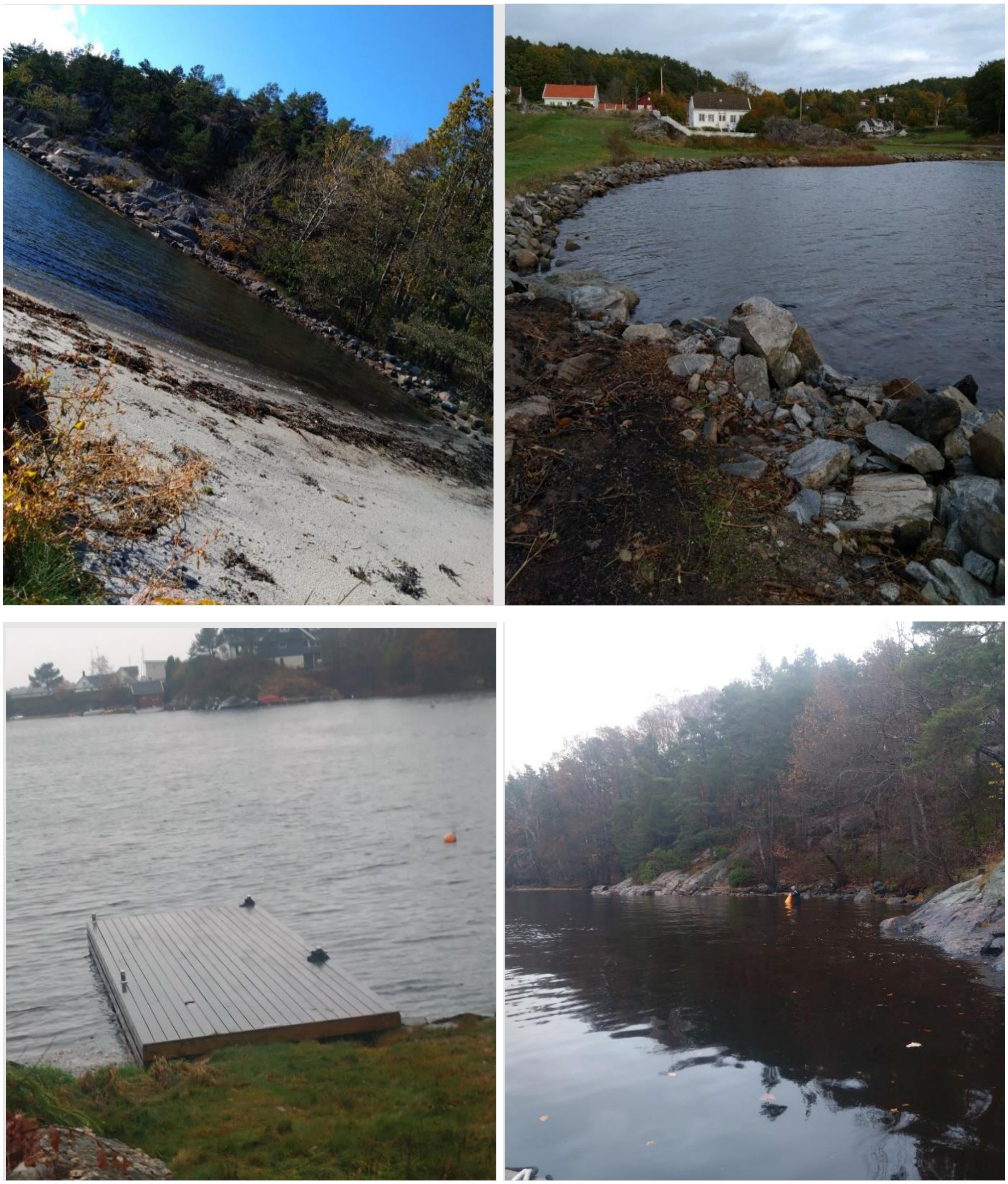


Figure B.1 Examples of high accessibility sites (Accessibility level 1).

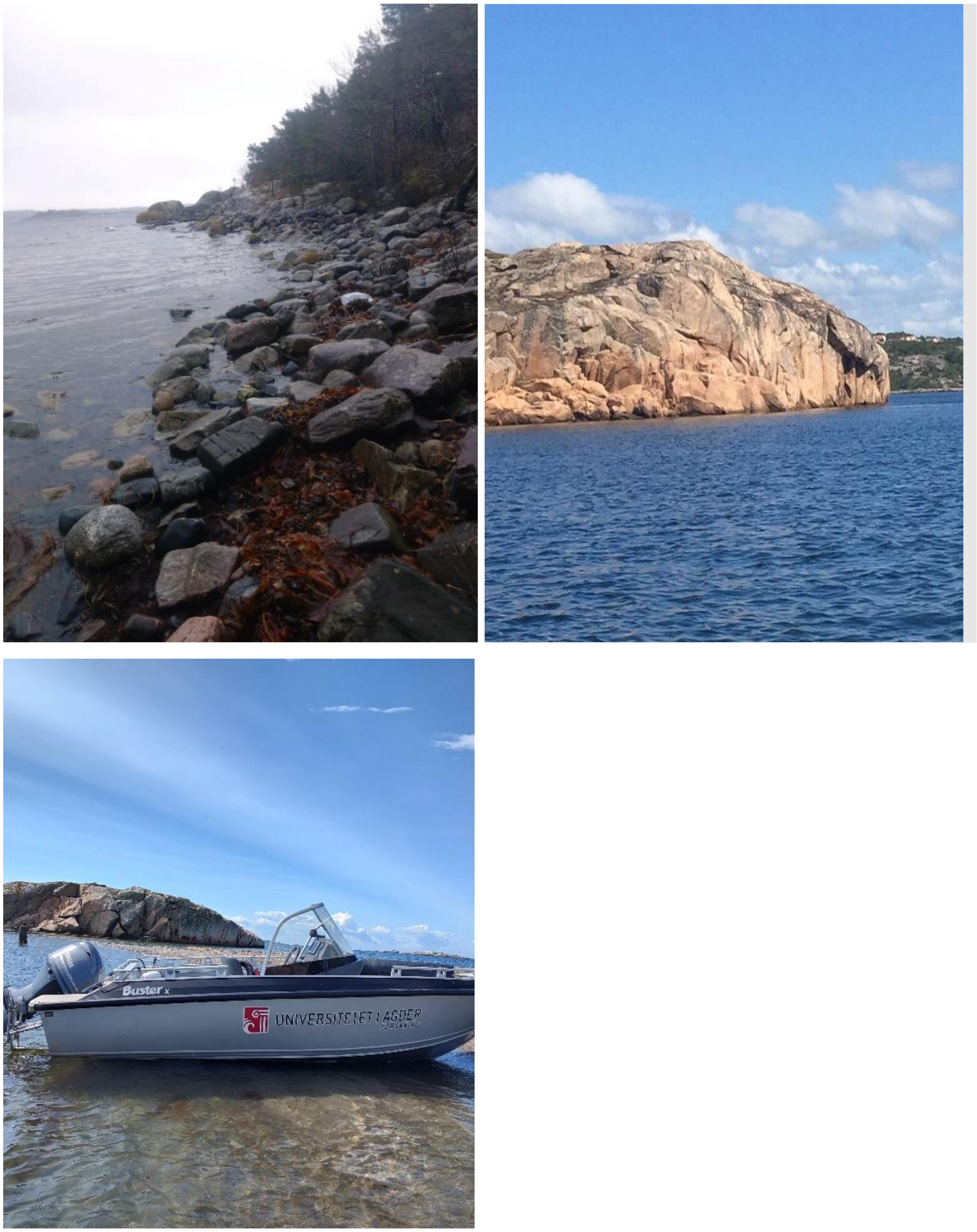


Figure B.2 Examples of intermediate accessibility sites (Accessibility level 2).

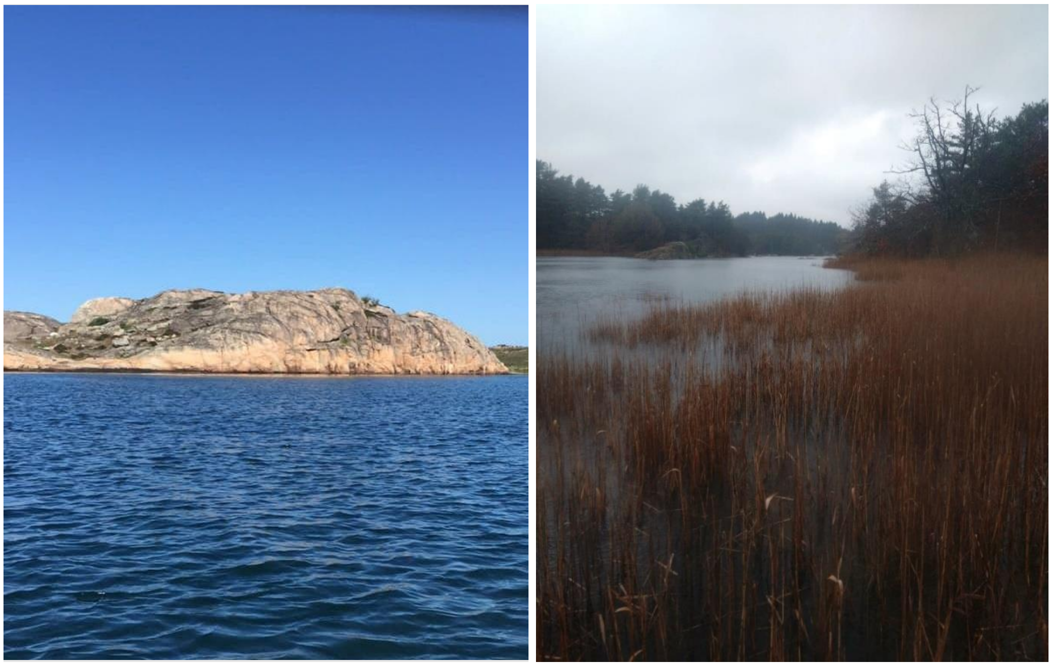
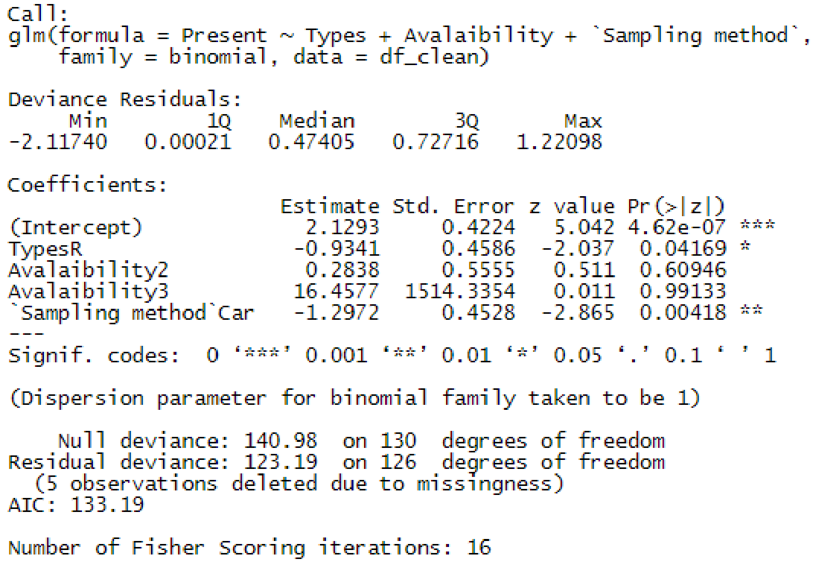


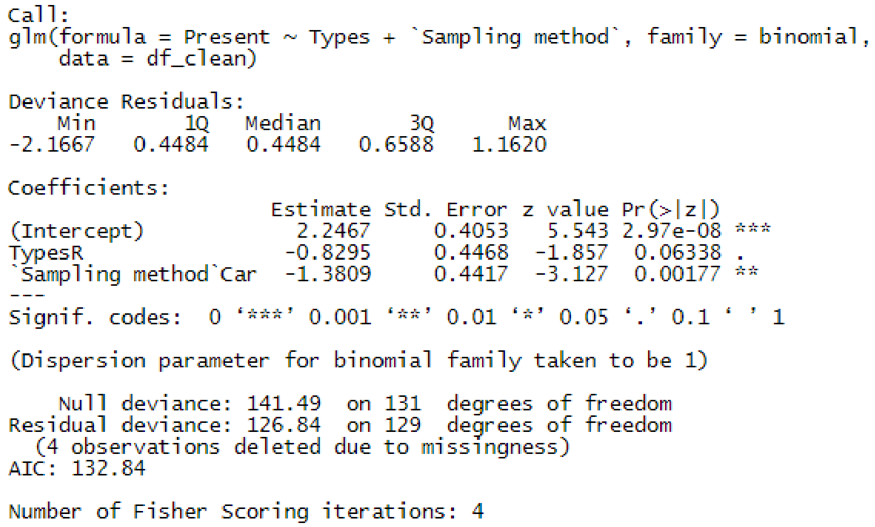
Figure B.3. Pictures showing examples of low accessibility sites (Accessibility level 3).

Appendix C: Generalized linear model output.



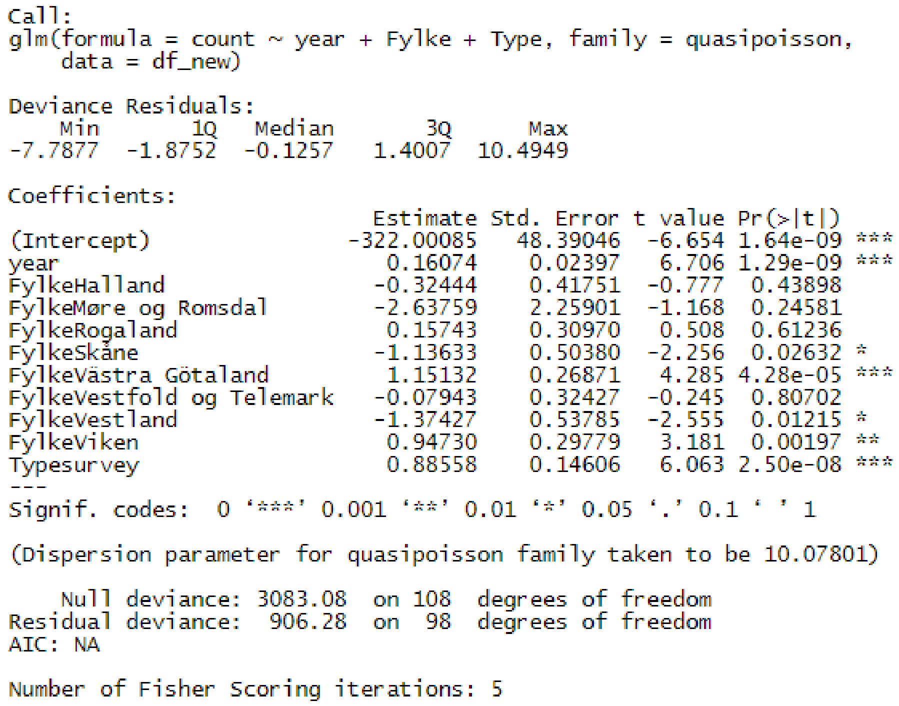
```
Call:
glm(formula = Present ~ Types + Avalaibility + `Sampling method`, 
    family = binomial, data = df_clean)

Deviance Residuals: 
     Min        1Q    Median        3Q       Max  
-2.11740   0.00021   0.47405   0.72716   1.22098  

Coefficients:
                     Estimate Std. Error z value Pr(>|z|)    
(Intercept)            2.1293     0.4224   5.042 4.62e-07 ***
TypesR                -0.9341     0.4586  -2.037  0.04169 *  
Avalaibility2          0.2838     0.5555   0.511  0.60946    
Avalaibility3         16.4577  1514.3354   0.011  0.99133    
`Sampling method`Car  -1.2972     0.4528  -2.865  0.00418 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 140.98  on 130  degrees of freedom
Residual deviance: 123.19  on 126  degrees of freedom
  (5 observations deleted due to missingness)
AIC: 133.19

Number of Fisher Scoring iterations: 16
```

Figure C.1 The generalized linear model output for presences with availability as a factor. Availability categories shown to be not significant in the model.



```
Call:
glm(formula = Present ~ Types + `Sampling method`, family = binomial, 
    data = df_clean)

Deviance Residuals: 
    Min       1Q   Median       3Q      Max  
-2.1667   0.4484   0.4484   0.6588   1.1620  

Coefficients:
                     Estimate Std. Error z value Pr(>|z|)    
(Intercept)            2.2467     0.4053   5.543 2.97e-08 ***
TypesR                -0.8295     0.4468  -1.857  0.06338 .  
`Sampling method`Car  -1.3809     0.4417  -3.127  0.00177 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 141.49  on 131  degrees of freedom
Residual deviance: 126.84  on 129  degrees of freedom
  (4 observations deleted due to missingness)
AIC: 132.84

Number of Fisher Scoring iterations: 4
```

Figure C.2 The simplified generalized linear model output for presences without availability as a factor.

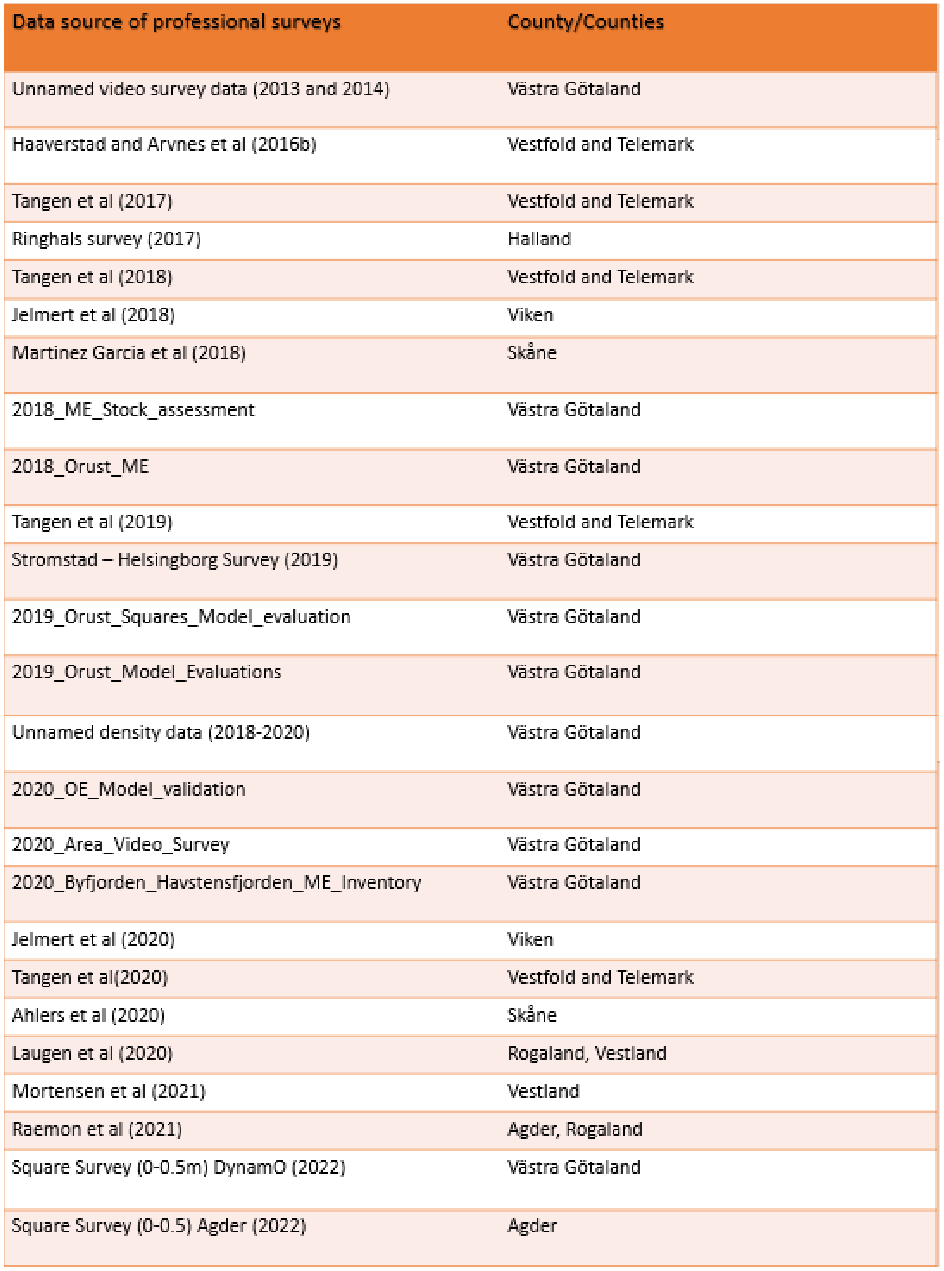


```
Call:
glm(formula = count ~ year + Fylke + Type, family = quasipoisson,
    data = df_new)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-7.7877  -1.8752  -0.1257   1.4007  10.4949

Coefficients:
                            Estimate Std. Error t value Pr(>|t|)
(Intercept)               -322.00085   48.39046  -6.654 1.64e-09 ***
year                         0.16074    0.02397   6.706 1.29e-09 ***
FylkeHalland                -0.32444    0.41751  -0.777  0.43898
FylkeMøre og Romsdal        -2.63759    2.25901  -1.168  0.24581
FylkeRogaland                0.15743    0.30970   0.508  0.61236
FylkeSkåne                  -1.13633    0.50380  -2.256  0.02632 *
FylkeVästra Götaland         1.15132    0.26871   4.285 4.28e-05 ***
FylkeVestfold og Telemark   -0.07943    0.32427  -0.245  0.80702
FylkeVestland               -1.37427    0.53785  -2.555  0.01215 *
FylkeViken                   0.94730    0.29779   3.181  0.00197 **
Typesurvey                   0.88558    0.14606   6.063 2.50e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for quasipoisson family taken to be 10.07801)

    Null deviance: 3083.08  on 108  degrees of freedom
Residual deviance:  906.28  on  98  degrees of freedom
AIC: NA

Number of Fisher Scoring iterations: 5
```

Figure C.3 The generalized linear model output for counts of observations divided by counties in Norway and Sweden.

Appendix D: Professional survey data details

Table D.1. Overview of professional surveys data used in this study.



Data source of professional surveys	County/Counties
Unnamed video survey data (2013 and 2014)	Västra Götaland
Haaverstad and Arvnes et al (2016b)	Vestfold and Telemark
Tangen et al (2017)	Vestfold and Telemark
Ringhals survey (2017)	Halland
Tangen et al (2018)	Vestfold and Telemark
Jelmert et al (2018)	Viken
Martinez Garcia et al (2018)	Skåne
2018_ME_Stock_assessment	Västra Götaland
2018_Orust_ME	Västra Götaland
Tangen et al (2019)	Vestfold and Telemark
Stromstad – Helsingborg Survey (2019)	Västra Götaland
2019_Orust_Squares_Model_evaluation	Västra Götaland
2019_Orust_Model_Evaluations	Västra Götaland
Unnamed density data (2018-2020)	Västra Götaland
2020_OE_Model_validation	Västra Götaland
2020_Area_Video_Survey	Västra Götaland
2020_Byfjorden_Havstensfjorden_ME_Inventory	Västra Götaland
Jelmert et al (2020)	Viken
Tangen et al(2020)	Vestfold and Telemark
Ahlers et al (2020)	Skåne
Laugen et al (2020)	Rogaland, Vestland
Mortensen et al (2021)	Vestland
Raemon et al (2021)	Agder, Rogaland
Square Survey (0-0.5m) DynamO (2022)	Västra Götaland
Square Survey (0-0.5) Agder (2022)	Agder