University of Southampton

Faculty of Environmental and Life Sciences

School of Ocean and Earth Science

Ocean to Plate:
Tracing the Geographic Origin of Fish Products using Biochemical Forensic Techniques

by

Juliet Sarah Elizabeth Wilson

ORCID ID: https://orcid.org/0000-0003-4703-3128

Thesis for the degree of Doctor of Philosophy

March 2023

University of Southampton

Abstract

Faculty of Environmental and Life Sciences

School of Ocean and Earth Science

Doctor of Philosophy

Ocean to Plate:
Tracing the Geographic Origin of Fish Products using Biochemical Forensic Techniques

by

Juliet Sarah Elizabeth Wilson

Global consumption of marine food products is steadily increasing, with fisheries and aquaculture production now at a record high. Consequently, traceability of fish and seafood throughout supply chains is becoming ever more important for consumers, producers and regulators, as trade networks increase in complexity and fish stocks face increased exploitation pressure. Marine food products are known to be at particularly high risk of incorrect claims of geographic origin due to the spatially based management of fisheries combined with the highly globalised seafood market. Verifying the claimed catch location of fish and seafood products is challenging, so currently there are no widely accepted forensic tests for spatial origin, despite the vital importance of traceability for maintaining sustainable fisheries as well as protecting fishers, retailers and consumers.

This thesis investigates the potential for identifying the geographic origin of four commercially important fish species using three contrasting biochemical techniques. The number of samples analysed allowed a comprehensive investigation into the accuracy with which individuals could be traced to their catch region, on the spatial scale of ICES subareas. Firstly, I applied stable isotope analysis of carbon, nitrogen and sulfur to discriminate among Atlantic cod (*Gadus morhua*) from different regions across the Northeast Atlantic, and then extended this technique to two further whitefish species – haddock (*Melanogrammus aeglefinus*) and European hake (*Merluccius merluccius*). Atlantic cod is extremely important commercially as well as having social and cultural significance. In the UK, cod had the highest value of all demersal species landed in 2020. Haddock and hake are also very important commercial species in the north Atlantic, since hake is one of the top three most landed demersal species in the EU and haddock is a key species fished by the UK fleet. Both are among the most highly consumed species in the EU and UK. All three species demonstrated varying stable isotope ratios within their white muscle tissue depending on the spatial origin. Cod could be traced to their true origin with an overall mean assignment accuracy of 79%, with accuracies of 90% or greater for cod from certain discrete regions. Haddock and hake were assigned to their correct known origin with slightly lower overall mean accuracies of 70% and 72% respectively, although Mediterranean hake as well as haddock from the Norwegian Sea and Rockall were very distinct and had assignment success rates of 92-100%.

Subsequently, I investigated the use of stable isotope analysis for tracing the origin of the critically endangered European glass eel (*Anguilla anguilla*) and complemented this with another technique, fatty acid analysis, to assess the potential of these techniques to be used as a forensic tool for detecting illegal glass eel fishing and trade. Neither technique was able to reliably distinguish among glass eels from different European rivers at this stage of their migration, although there were indications that eels from the River Oria (Spain) and River Severn (UK) may be sufficiently distinct to be discriminated under certain circumstances.

Finally, trace elemental profiling by x-ray fluorescence (XRF) was explored as a possible technique for tracing the geographic origin of Atlantic cod, opportunistically using the same cod samples as were analysed for stable isotope composition earlier in the thesis. Two different XRF instruments were used – the Itrax core scanner and the Vanta analyser. The findings demonstrated that the multi-elemental composition of cod muscle tissue, estimated by both XRF instruments, varied with spatial origin. Cod from all sampled regions could be assigned to known origin with at least 74% accuracy, with success rates of >95% for certain regions. Classification to origin using stable isotope and trace element data combined gave an overall mean assignment accuracy of 91%, which is higher than the accuracy obtained using either of the two methods independently.

My research addresses the lack of established techniques available to trace the spatial origin of fish and seafood products, by demonstrating that biochemical techniques have the potential to determine provenance on a scale suitable for fisheries management. However, the findings indicate that any technique used independently is unlikely to provide a stand-alone test to confirm origin for all fishery areas with accuracy suitable for legislative action, but combining approaches may allow the required level of accuracy to be reached.

Table of Contents

Table of Tables

Table of Tables

Table of Figures

Research Thesis: Declaration of Authorship

Print name: Juliet Sarah Elizabeth Wilson

Title of thesis: Ocean to Plate: Tracing the Geographic Origin of Fish Products using Biochemical Forensic Techniques

I declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. None of this work has been published before submission

Signature: ..... Date: 05/03/2023

Acknowledgements

This research was funded by the Natural Environmental Research Council (grant number NE/R009783/1) as part of the SPITFIRE DTP. Additional funding was provided by a NERC 'Grant-in-Kind' for stable isotope analysis at the Life Sciences Mass Spectrometry Facility (grant number EK312-13/18).

First of all, I would like to express enormous thanks to my amazing supervisor, Clive Trueman, whose expertise and guidance was invaluable. I will always be grateful for his constant support and understanding through all the challenges I faced during the time of my PhD. His encouragement for me to apply for this PhD in the first place during a chance meeting is what got me the opportunity, and his belief in me throughout the PhD gave me the confidence to persevere.

I would also like to thank all those from numerous organisations who provided me with fish samples, including Marine Scotland, Faroe Marine Institute, Technical University of Denmark, Marine Institute (Ireland), Ifremer, Marine and Freshwater Research Institute (Iceland) and Institute of Marine Research (Norway). The effort and time spent collecting these samples for me is hugely appreciated and without them my PhD research would not have been possible.

A huge thanks to Rona McGill for all the stable isotope analysis and her expert advice. I would also like to thank Millie Bompard and all those at BOSCORF for the XRF analysis, particularly for analysing all my samples for me when I wasn't able to access the laboratory during Covid lockdowns. I am also very grateful for Millie's hard work in data processing, and her technical advice and help with data interpretation. In addition, thanks to Sarge Bray for his expertise and instruction in fatty acid analysis, as well as Bastian Hambach and Megan Wilding for their assistance with stable isotope analysis at NOCS.

I am very fortunate to have been part of such a friendly research group and office, and I am very grateful for all these friends who have made my PhD experience much more fun. I am also thankful for the support and advice they have given at numerous points along the way. A particular mention to Abbie for always being there for a chat and to help me out.

My family have been such a wonderful support and encouragement, for which I will always be thankful. It was due to their inspiration to follow my dreams of being a marine biologist that has allowed me to reach where I am today. Their last-minute proof reading was also greatly appreciated!

I want to thank Phil with all my heart for always being there for me, for his genuine interest in my research and for his kindness in caring for me (as well as for his expert help with debugging R code on many occasions!).

Finally and most importantly, I thank and praise God for helping me with my PhD and for giving me the strength and endurance to get through challenging times. I look forward to all the future adventures He has planned for me.

"I can do all things through Christ who strengthens me" *Philippians 4:13*

I dedicate my PhD thesis to Grandad who would have loved to read it.

Chapter 1 Thesis Introduction

1.1 The importance of seafood origin traceability

Over 3 billion people globally rely on seafood as their primary source of protein, since the ocean is the largest source of protein in the world (FAO, 2016). Seafood is particularly important for its nutritional value in the least developed countries, providing essential amino acids, fatty acids (including omega-3), vitamins and minerals. It also provides a valuable source of employment and income, since globally over 58 million people are employed in fisheries and aquaculture, and many developing countries rely on seafood as a source of foreign income (FAO, 2022c). When indirect and induced effects are included, the total economic impact of global marine capture fisheries is estimated to be between US$225 and US$240 billion a year, which is almost three times larger than the value of landings at first sale (Dyck and Sumaila, 2010). Seafood is also the most traded food commodity in the world with global trade worth over £180 billion a year, making it greater than wheat, rice, sugar and coffee (Asche, 2014).

Wild-caught seafood has a lower environmental impact than other sources of animal protein since capture fisheries are associated with relatively low greenhouse gas emissions (Parker et al., 2018, Gephart et al., 2021). Food production is responsible for 25% of global anthropogenic greenhouse gas emissions, so is a major contributor to climate change (Poore and Nemecek, 2018). Of these total global emissions, marine fisheries were found to contribute only 4% (179 million tonnes of CO_2-eq) whereas agriculture and livestock production accounted for 5 billion tonnes of CO_2-eq (Parker et al., 2018). Greenhouse gas emissions associated with wild seafood are reported to be roughly equivalent to poultry production, but around six times lower than that of beef and five times lower than that of lamb (Oceana, 2021). Farmed aquatic foods also have a relatively low environmental impact with similar levels of greenhouse gas emissions as chicken, though other adverse effects of aquaculture such as nitrogen and phosphorus discharge are greater issues than for wild-caught fish (Gephart et al., 2021). Therefore, marine food products from both aquaculture and capture fisheries are of significant importance in providing a global source of protein while also reducing the negative environmental impacts in a period of human-induced climate change.

With a rapidly growing world population there is an ever-increasing demand for seafood. Global consumption of aquatic foods has increased at an average rate of 3% per year from 1961 to 2019, and total fisheries and aquaculture production reached a record high of 214 million tonnes in 2020 (FAO, 2022c). However, wild fish stocks have suffered severe declines in the last few decades and now around 35% of the world's fish stocks are

overexploited (FAO, 2022c). Therefore, developing ways of sustainably managing fisheries to ensure future food security for people all around the world is a priority. Since fisheries are managed spatially, the correct capture origin of fish and seafood is essential information for sustainable management.

Food fraud is a major global issue, costing the UK food and drinks industry £11.2 billion a year (Gee et al., 2014). Fish and seafood are one of the foods most at risk of fraud (Interpol/Europol, 2017), which has serious implications for the sustainability of fisheries. Seafood fraud is a criminal activity committed when a product is placed on the market deliberately to deceive the customer, usually for the purpose of financial gain (National Food Crime Unit, 2016). Mislabelling by a retailer can also occur unknowingly, either through accidental substitution or from fraud further back in the supply chain. There are a number of types of seafood fraud, including species substitution (usually when a low-value species is used in place of a more expensive one), the undeclared use of additives such as water-binding agents to increase the weight of a product, mislabelling of the catch method and the fraudulent use of a brand name (FAO, 2018). However, the geographic origin of fish can also be a source of fraud by obscuring or claiming false origin for fish caught illegally, for example from a protected area or from areas where the fisher has no quota.

Illegal, unreported and unregulated (IUU) fishing is a major issue worldwide, contributing up to 20% of the global seafood catch and worth between US$10 billion and US$23.5 billion a year (Agnew et al., 2009). The total economic impact due to the diversion of seafood from the legitimate trade system to illicit trade is estimated to be even higher at US$26 to US$50 billion (Sumaila et al., 2020). The inaccurate reporting of illegal catches results in a significant part of the annual catch missing from assessments and so leads to overestimation of sustainable catches (Sumaila et al., 2006). Therefore, in addition to being illegal activities, IUU fishing and seafood fraud greatly hinder sustainable management efforts and threaten fish stocks. The ability to reliably identify the geographic origin of fish and seafood would deter IUU fishing by enabling detection and prosecution.

Furthermore, mislabelling and fraud presents food safety concerns. Seafood can expose consumers to a wide range of toxins and contaminants, including pathogenic microorganisms, heavy metals such as mercury, antibiotics from aquaculture and other chemical pollutants. These may be present locally and therefore present a greater risk in products from some regions than others. Production areas for bivalve molluscs are classified according to the levels of *E. coli* detected in the shellfish flesh. This determines the treatment required before bivalves can be marketed for human consumption and in some cases harvesting from the area may be prohibited if contamination levels are consistently above safe levels (Food Standards Agency, 2022). If the geographic origin of shellfish is not correctly identified, the appropriate pre-treatment may not be adhered to or illegally harvested products may be allowed to enter the supply chain. Therefore, it is

essential to identify contaminated or unsafe products before they can cause harm to consumers.

Food scandals such as the European horsemeat and avian flu crises have attracted much attention in the media (Quinn, 2013, Abbott and Pearson, 2004) and have led to a high demand from the public to know the origin of their food. Increased awareness of consumers to the issues of sustainability has also led to a greater demand for responsibly sourced products, which creates an incentive for more producers to sustainably harvest fish and seafood. Accurate traceability information allows consumers to make informed decisions on the products they purchase. Since sustainably sourced products often sell for a higher price, traceability also increases the economic gain for honest producers and retailers.

Therefore, traceability is vital for the sustainability of fisheries by enabling effective management and allowing regulations to be enforced, as well as detecting and deterring fraud. Reliable traceability also improves customer confidence in products, protects honest producers from the economic impacts of fraud and reduces health risks for consumers.

1.2 Prevalence of seafood fraud and mislabelling

Fish and seafood are particularly vulnerable to incorrect claims of geographic origin due to the spatially based management of fisheries combined with the globalisation of the seafood market and ever-increasing complexity in the trade network. This leads to a risk of both deliberate fraud for financial gain and accidental mislabelling of a product. Seafood products are usually transported from the producer through several processors and distributors before reaching the retailers and finally the consumer, often passing through several different countries (Leal et al., 2015). Management and regulation such as restricting access to fishing areas and imposing quotas, as well as being the basis for sustainable fisheries, create an incentive for fraud. As restrictions on legal fishing become greater and the market demand for a product increases, there is increased motivation to take part in illegal fishing. Therefore, when fish stocks decline any restrictions introduced can also exacerbate illegal fishing (Sumaila et al., 2006). Consumers may pay a higher price for sustainably caught fish, so this is also an incentive for fraud or mislabelling. In 2015 an INTERPOL–Europol investigation found that internationally-traded fish was in the third highest risk category for food fraud, and in 2013 the European Commission classified fish in the second-highest category for fraud (FAO, 2018).

Genetic tests are well established and can be used to reliably confirm the species in seafood products. This has revealed widespread species substitution and mislabelling across some sectors of the seafood trade network (Pardo et al., 2016, Naaum et al., 2016, Warner et al., 2016). In a major study by Oceana, reviewing over 200 studies from 55

countries worldwide, DNA testing found that 20% of fish species are mislabelled on average (Warner et al., 2016), but with much higher mislabelling rates in some species. For example, 98% of bluefin tuna samples tested in Brussels were found to be mislabelled, as well as 82% of swordfish, perch and grouper tested in Italy. In some cases the actual species identity was revealed to be a species of conservation concern – 55% of "shark" samples in Brazil were in fact found to contain largetooth sawfish, which is listed as critically endangered by the IUCN and for which trade is prohibited in Brazil (Warner et al., 2016). Pardo et al. (2016) also conducted a review of 51 peer-reviewed papers investigating seafood species mislabelling using genetic methods, including 4500 samples collected globally, and found on average 30% of samples tested were mislabelled with regard to species. In the UK, high rates of species mislabelling have also been found. A study by Miller and Mariani (2010) discovered that 25% of the cod and haddock products sampled in 2010 were a different species than labelled, and 82.4% of smoked fish samples were mislabelled.

These high levels of mislabelling, however, are likely not representative of the industry as a whole, since many studies deliberately target species known or suspected to have high levels of mislabelling (e.g. bluefin tuna, shark, or species such as snapper that are difficult to identify), as well as sections of the retail chain that are most likely to show mislabelling (such as restaurants and small independent retailers). Therefore, caution should be used before implying a similar level of mislabelling across the whole industry. For example, a study by Helyar et al. (2014) found much lower mislabelling rates for cod and haddock when samples from supermarkets were tested (94% of products correctly labelled), which suggests that in some parts of the retail chain mislabelling may be less prevalent (Pardo et al., 2016). A meta-analysis was carried out by Luque and Donlan (2019) using Bayesian methods, which aimed to improve on previous study-level mean estimates of mislabelling rates using all available data, although it was not possible to account for bias due to sample selection. This meta-analysis resulted in an estimate of 8% for the global mislabelling rate at product level, although the authors acknowledge that some products have more frequent mislabelling and should be a priority for research and improvement.

Following the European horsemeat scandal in 2013, which revealed the vulnerability of the food industry to international-level fraud, a number of initiatives and networks have been created to tackle food fraud. These include Europol, the Food Fraud Network of the European Commission, and the Food Fraud Database of the United States Pharmacopeial Convention (USP) (FAO, 2018). In the UK, a review into the integrity and assurance of food supply networks was carried out by Elliott (2014), which set out recommendations for a national food crime prevention framework. European Union (EU) legislation was also introduced in 2014 that requires all commercially landed fish to be labelled with their catch location (FAO area), amongst other details including the scientific species name and catch

method (European Union, 2013). Following the increased use of DNA testing and better labelling regulations, Warner et al. (2016) reported that the average species mislabelling rate declined in Europe from around 23% in 2011 to 8% in 2015. However, the meta-analysis by Luque and Donlan (2019) did not reveal a statistically significant decline in mislabelling in the EU during the period 2006-2015, so the effect of introducing new labelling and traceability regulations still remains uncertain. Nevertheless, the best available way of preventing fish and seafood mislabelling is through effective traceability and reliable methods of authenticity testing.

1.3 Current spatial traceability of marine food products

Genetic tests are now widely available and are routinely used to confirm the species contained in products, but it is much more difficult to identify the capture region of marine food products (FAO, 2018, El Sheikha and Montet, 2016, Gopi et al., 2019a, Cusa et al., 2021) because a tracer is required that shows variations between multiple different regions on a suitable spatial scale. Although EU legislation requires all commercially landed fish to be labelled with the catch location (European Union, 2013), currently there are no widely accepted analytical methods to verify the claimed geographic origin of traded seafood retrospectively. For this reason, there are also no estimates for the extent of geographic origin mislabelling within retail networks, as there are for species mislabelling. However, one small scale study identified possible spatial mislabelling of cod fillets sold in Swedish shops, where over half of 42 samples that claimed to have been caught in the North Sea were suspected to have actually originated from other areas, mainly the eastern Baltic (Nielsen et al., 2012a). This is important since eastern Baltic cod are considered to have lower quality flesh than cod from the North Sea and other North Atlantic stocks, and furthermore, fish from the eastern Baltic contain higher levels of organic pollutants (Nielsen et al., 2012a). The Food Standards Agency stated in their 2016 report that based on intelligence the “misrepresentation of fish origin is currently a greater concern than fish substitution” (National Food Crime Unit, 2016).

Current traceability systems rely mainly on document-based records at each stage in the supply chain. Vessel monitoring systems (VMS) that provide data at regular intervals about vessels’ location, course and speed are now used in many countries. This has improved traceability and reduced the options for fraud at first landing, but VMS are often only compulsory for larger vessels and cannot provide traceability throughout the whole retail chain. Chain of custody data using paper or electronic records also cannot prevent accidental or deliberate mis-association of products to specific vessels. Blockchain technology is a recent development that may improve confidence in supply chain documentation, but currently these techniques still rely on human input and physical tagging methods to correctly associate the product with blockchain transactions.

Therefore, there is potential for the data initially entered into the blockchain to be falsified or inaccurate (Blaha and Katafono, 2020).

The Marine Stewardship Council (MSC) and other eco-labels certify that seafood products can be traced to a sustainable source (Marine Stewardship Council, 2018). However, this also relies on following the chain of custody and, although this greatly improves transparency, there is still potential for accidental substitutions or counterfeit labels. Marko et al. (2011) reported that a small proportion of MSC-certified Patagonian toothfish products were found to be from an unsustainable stock rather than the certified one, demonstrating that it is possible for uncertified fish to be introduced into the MSC supply chain through deliberate or inadvertent mislabelling.

The development of a new forensic test for the geographic origin of fish and seafood products would allow the claimed origin to be verified at all points through the supply chain, reducing the prevalence of spatial mislabelling and providing a deterrent for fraud.

1.4 Use of stable isotope analysis in traceability of terrestrial food products

The major structural nutrient elements - carbon, nitrogen, sulfur, oxygen and hydrogen - all have multiple, naturally occurring stable isotopes (atoms with varying numbers of neutrons). The relative abundance of stable isotopes (the stable isotope ratio) contains information about geographic origin (Carter and Chesson, 2017, Meier-Augenstein, 2017, Trueman and St. John Glew, 2019). The ratios of these stable isotopes vary spatially in a manner that can be predicted based on the climatic conditions, precipitation, soil properties, volcanic activity, distance from the sea and the geology of the area, as well as a range of other factors (Camin et al., 2017). For carbon, the stable isotope ratio in organic tissues depends on the isotopic composition of inorganic carbon taken into the plant and the photosynthetic pathway of the plants at the base of the food web (Carter and Chesson, 2017). As a result, the stable isotope composition of the primary producers varies across geographic locations, which is subsequently transferred up the food chain into an animal's tissues through the diet, and therefore into the food product. This means that the isotopic composition of food products can be used to differentiate between those from different locations, acting as a natural marker for the area an animal was feeding (Carter and Chesson, 2017, MacKenzie et al., 2011, Trueman et al., 2012a, Trueman and St. John Glew, 2019).

For terrestrial food products, the stable isotope ratios of a number of elements including carbon, nitrogen, hydrogen, oxygen, sulfur and strontium are utilised for origin traceability (Carter and Chesson, 2017). Often multiple stable isotopes are required in combination for

tracing the geographic origin successfully (Camin et al., 2017). The use of stable isotope analysis has been applied extensively to verify the origin of a wide variety of terrestrial food products including wine, cheese, fruit juice, cocoa beans, olive oil and honey (West et al., 2007, Valenti et al., 2017, Rummel et al., 2010, Perini et al., 2016, Camin et al., 2010, Bontempo et al., 2017). In Europe there are an increasing number of geographic origins for food products that are officially protected, such as those awarded Protected Designation of Origin (PDO). These fetch significantly higher prices which leads to a greater incentive for fraud, so it is important to have a method of validating the claimed origin to protect consumers and honest producers.

Many studies have also investigated the use of stable isotopes for tracing the origin of meat products, such as beef, lamb and chicken. For example, Horacek and Min (2010) successfully traced beef to its country of origin using stable isotopes. They were able to reliably discriminate between beef from USA, Mexico, Australia and New Zealand since meat from each country had a statistically different carbon and hydrogen isotopic signature, generally with small variability. Bong et al. (2010) could also differentiate between Korean and Australian beef with a 100% success rate, and Korean from USA beef with a 90% success rate using carbon and oxygen isotopes. One study also used cattle tail hair to infer geographic origin, which gave an overall correct classification rate of 82.6% for four beef production regions in China based on carbon, nitrogen and hydrogen stable isotopes (Liu et al., 2013). However, this technique is only able to distinguish between products if they are from regions with distinct isotopic signatures. If two locations are very similar isotopically, it is difficult to trace the origin. Similarly, if products are from locations that are very close geographically, it may not be possible to reliably differentiate between them because the locations are likely to have very similar isotopic compositions.

Currently there are officially acknowledged methods in use to verify the claimed origin of wine, cheese, fruit juice and oil (Camin et al., 2017). Official isotopic databanks have been set up, such as the European wine databank that was established in 1991 or the databank for Grana Padano cheese, which contain reference isotopic data from a wide range of geographic regions and have been built up over many years (Camin et al., 2017). New data are gathered each year and are added to the databank to ensure accuracy. Suspect samples are compared to these reference datasets and if the sample does not fall within the limits of the reference values plus uncertainty, the sample is declared as not authentic. These databanks can assign origin with a 95% confidence level, so can be used for commercial products and in legal cases of suspected food fraud (Camin et al., 2017).

1.5 Techniques for origin traceability of marine food products

Developing a forensic test for the geographic origin of marine food products, particularly wild caught, is much more challenging than for terrestrial food products. The vast size and relative inaccessibility of the oceans means that collecting reference samples from a wide variety of regions is a difficult and costly task, whereas on land, farmers own the specific field where their produce is grown so it is relatively simple to derive a chemical fingerprint for that area. Furthermore, there is an economic incentive for farmers on land (or in aquaculture) to invest in sampling to characterise their own produce and protect against subsequent fraud, but this incentive does not exist for wild-caught seafood products where many fishers can operate in the same area (Martino et al., 2022b).

A forensic tracer for geographic origin is required that shows variations among regions on a suitable spatial scale for the application, and the tracer should be incorporated into the edible parts of the fish (muscle tissues) so that it can be analysed in processed or filleted food products. Ideally, the tracer should also be quickly and relatively inexpensively measured if the technique is to be widely used. There are several possible tracers that have potential application in the spatial traceability of fish and seafood products, which are discussed in detail below.

However, each of these techniques alone may not give sufficient accuracy for all species or regions for use in traceability applications. Use of two techniques in combination may result in a higher level of accuracy than can be achieved using any method independently (Ortea and Gallardo, 2015, Li et al., 2016, Gong et al., 2018). Cusa et al. (2021) investigated the possibility of combining genetic and stable isotope techniques using a Bayesian assignment method. This demonstrated that the assignment ability improved using a combined approach for certain species, such as sole and Albacore tuna, and the chance of misassignment to incorrect regions decreased. Cusa et al. (2021) also proposes a decision tree to assist with deciding which technique to use when, and suggests a combined genetic and stable isotope approach if the assignment accuracy is less than 90% using one method alone.

The assignment accuracy of regional tracers is usually tested by taking a subset of samples from the reference dataset and assigning these to origin using classification rules that were created from the reference data. The estimate of assignment accuracy achieved using this non-independent approach will therefore reflect the best possible case. Using test samples that were collected independently of the reference samples is likely to result in reduced performance. Very few traceability studies have used datasets that are genuinely

independent of each other to estimate accuracy (Cusa et al., 2021), but this is important for demonstrating the true effectiveness of a tracer in real world traceability settings.

1.5.1 Stable isotope ratios

The oceans show predictable regional variations in stable isotope ratios, just as in terrestrial environments (Somes et al., 2010, Trueman et al., 2012a, Magozzi et al., 2017) (Figure 1.1). In marine environments carbon and nitrogen stable isotopes are most commonly used for tracing origin of soft tissues, since hydrogen and oxygen isotope ratios are relatively constant spatially across the ocean (Trueman et al., 2012b). However, hydrogen and oxygen isotopes can be used in calcified structures because the differences become exaggerated (Martino et al., 2022b). Stable isotope ratios are measured as the difference compared to a standard – for example Vienna Pee Dee Belemnite for carbon, atmospheric N_2 for nitrogen, and Canyon Diablo Troilite for sulfur, expressed as $\delta^{13}C$, $\delta^{15}N$ and $\delta^{34}S$ values respectively.

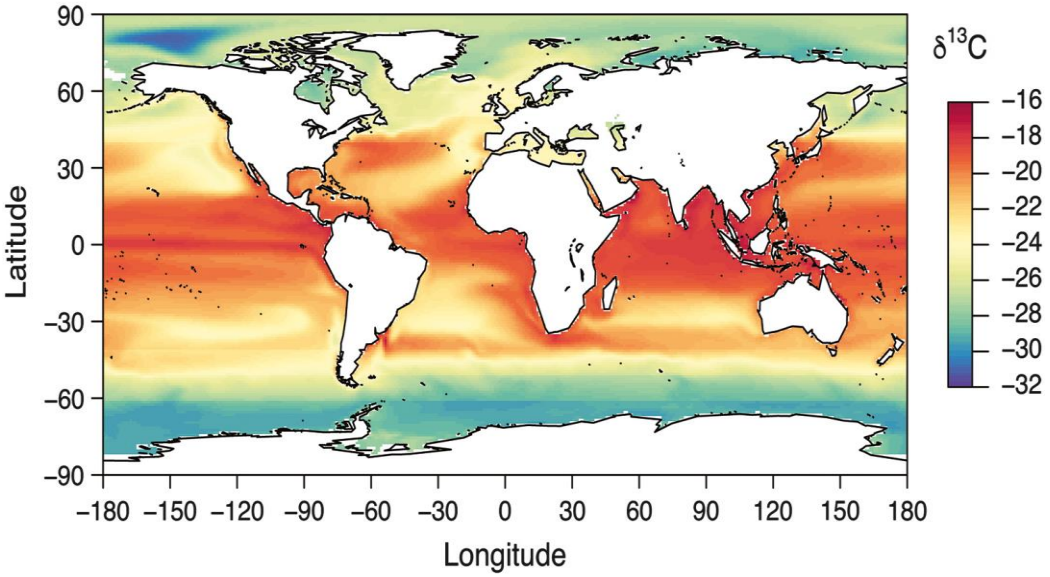


Figure 1.1 Modelled regional variation in carbon isotopic composition of phytoplankton across the global oceans (annually averaged surface water distribution in per mil) from Magozzi et al. (2017).

The isotopic composition of phytoplankton at the base of the food web varies spatially due to both variations in the isotopic composition of dissolved inorganic nutrients (e.g. CO_2, NO_2, SO_4), as well as isotopic fractionation associated with nutrient fixation (Barnes et al., 2009, Somes et al., 2010). Spatial variation in the isotopic composition of primary producers can be predicted a priori based on differences in phytoplankton growth rate and species, dissolved CO_2 concentration, nutrient source and water column depth, among

other factors (Somes et al., 2010, Magozzi et al., 2017). In organic tissues of animals from marine environments, carbon and nitrogen stable isotopes are the most useful for tracing geographic origin, since hydrogen and oxygen isotope ratios are relatively constant spatially across the oceans (Trueman et al., 2012a). Sulfur stable isotopes provide effective tracers for anoxic conditions commonly seen in nearshore and intertidal settings (Fry, 2006), but spatial variations in tissue $\delta^{34}S$ values have been less well studied in open marine environments. St. John Glew et al. (2019) have recently demonstrated that sulfur isotope ratios do show systematic spatial variability across UK shelf seas.

The stable isotope composition of an animal's tissues reflects the diet integrated over several weeks or months. Larger organisms or organisms with low metabolic rates have longer tissue turnover times and correspondingly slow isotopic assimilation rates (Jennings and Van Der Molen, 2015). This can mean that the results are difficult to interpret in large animals that migrate over long distances, since the isotope signatures from a range of locations become averaged. The location where an animal assimilates food may also not necessarily be the location where it spends most time in very mobile animals or those feeding on mobile prey (Trueman et al., 2017). Farmed animals can also pose a problem using stable isotope techniques for traceability, since the aquaculture feed is often imported from another region. Thus, the stable isotope composition of farmed animals will reflect the location where the feed was produced rather than where the farm is located. In such cases, stable isotope analyses may be useful in tracking the origin of feed products, and may discriminate among farm sites using different feed sources.

Isotopic ratios also vary temporally, so this needs to be taken into consideration when using the isotopic composition to trace the origin of products. Areas such as the North Sea are known to have relatively constant spatial differences in stable isotope ratios over decadal to centennial time periods (MacKenzie et al., 2014), whereas other sea areas with more fluctuating environmental conditions may show greater temporal variations. Test samples should be acquired during a similar timeframe as the reference dataset if they are to be reliably compared, due to the possible variation over time within the same region.

Finally, across food webs the heavier isotopes of nitrogen and to a lesser extent of carbon are preferentially accumulated in consumer tissues, so $\delta^{15}N$ and $\delta^{13}C$ values increase predictably up the food chain (typically 3-3.4‰ for nitrogen and ~1‰ for carbon with each trophic level (Minagawa and Wada, 1984, Peterson and Fry, 1987)), making $\delta^{15}N$ values a good indicator of trophic level (Jennings and Van Der Molen, 2015, Zanden and Rasmussen, 2001). The increase in the heavier isotope between the diet and consumer is often termed trophic fractionation, and it occurs due to the different rates the stable isotopes react during metabolism as well as variations in the proportions of amino acids that are synthesised *de novo* (Jennings and Van Der Molen, 2015). Consequently, species from the same geographic origin but of different trophic levels or with isotopically distinct

diets will have differing stable isotope ratios. Marine food webs are commonly structured by size, so $\delta^{15}N$ values typically increase systematically with body size in marine food webs (Figure 1.2). In order to infer location from the isotopic compositions of fish of different species or body sizes a correction is necessary to account for trophic effects, but this can be difficult to determine and introduces additional uncertainty.

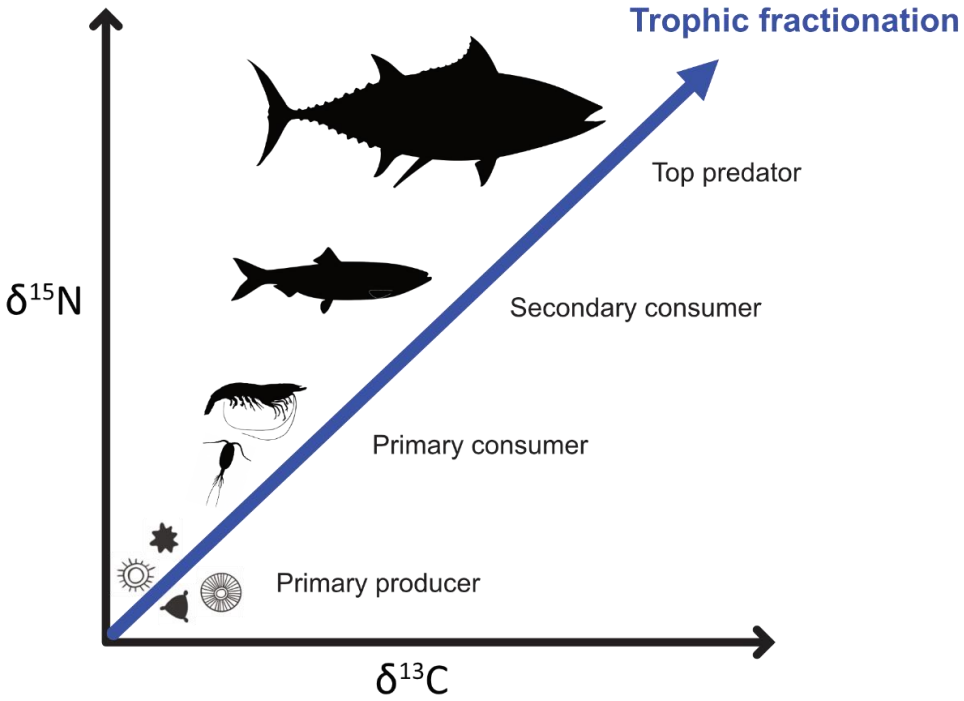


Figure 1.2 Schematic illustrating trophic enrichment of $\delta^{13}C$ and $\delta^{15}N$ values of marine organisms up the food chain.

1.5.1.1 Discrete and continuous assignment

There are two methods of assigning geographic origin using stable isotope analysis. The first is discrete assignment, where the sampling area is divided into a number of regions with defined boundaries (Figure 1.3). Reference samples are collected and measured for each region using known-origin samples, to which the test sample is compared. The most likely origin is the region that the test sample's isotope composition matches most closely. However, this method assumes the test sample can only have come from one of the sampled locations, and so it will give an incorrect result if the sample has in fact originated from elsewhere. Consequently for the method to be reliable, reference samples must be collected from all the possible regions of provenance.

The second method is continuous assignment, using statistical models of isotopic spatial variability termed "isoscapes" that are created by interpolating between measurements from tissue samples of a baseline organism (Figure 1.3). By comparing the isotope values

of a test sample to those in each grid cell of the isoscape, the probability of occurrence for the test individual can be calculated to give the grid cells of most likely origin. By using an isoscape, individuals can be assigned to anywhere on the continuous surface, in contrast to discrete assignment where individuals can only be assigned to defined regions. This means that the geographic origin can be determined more precisely than the discrete regions, but it is more time consuming to create an isoscape. For example, Trueman et al. (2017) created an isoscape for the North Sea using jellyfish tissues that allowed scallops to be traced to their geographic origin to an area of 20% of the North Sea with 50% accuracy or to >40% of the North Sea with >90% accuracy. This demonstrates that although individuals can be assigned to a very small area, this results in the accuracy being lower. So for an assignment to have a high accuracy, test samples must be assigned to a larger geographic region. In the same way as with discrete assignment, the test sample is assumed to have originated from within the isoscape area.

Isoscapes have also been used for discrete origin assignments using a different approach. Slesser and Trueman (2021) used a continuous isoscape to generate reference predictions for selected discrete regions, and then assigned scampi to region of origin based on these predicted isotope values. A recent study by Martino et al. (2022b) also applied this technique for spatial assignment, whereby global ocean isoscapes were used to derive predicted stable isotope values for discrete regions. These predictions were used to assign fish, shellfish, and cephalopods to their most likely origin regions.

An isoscape can be used to trace the origin for any species, provided that calibration is carried out between the organism used to create the isoscape and the species and tissue to be assigned. However, this calibration is different for each animal due to trophic fractionation, so in practice it is difficult to determine accurately. The same species and tissues that were used to construct the isoscape are more likely to be assigned correctly, so this is the best option if possible. Individual variation in isotope values affects the accuracy of both of these methods for tracing origin, since larger variation between individuals can lead to a greater proportion of incorrect assignments. This variation in the isotope composition is created by small differences in movement patterns and diets between individuals of the same species combined with the gradual overturning of tissues over time.

All chapters in this thesis are based on discrete assignment, since the aim is for the methods developed to have practical use in fisheries management. Where well governed spatial management units exist, such as ICES subareas, discrete assignment is particularly well suited for traceability in a fisheries context.

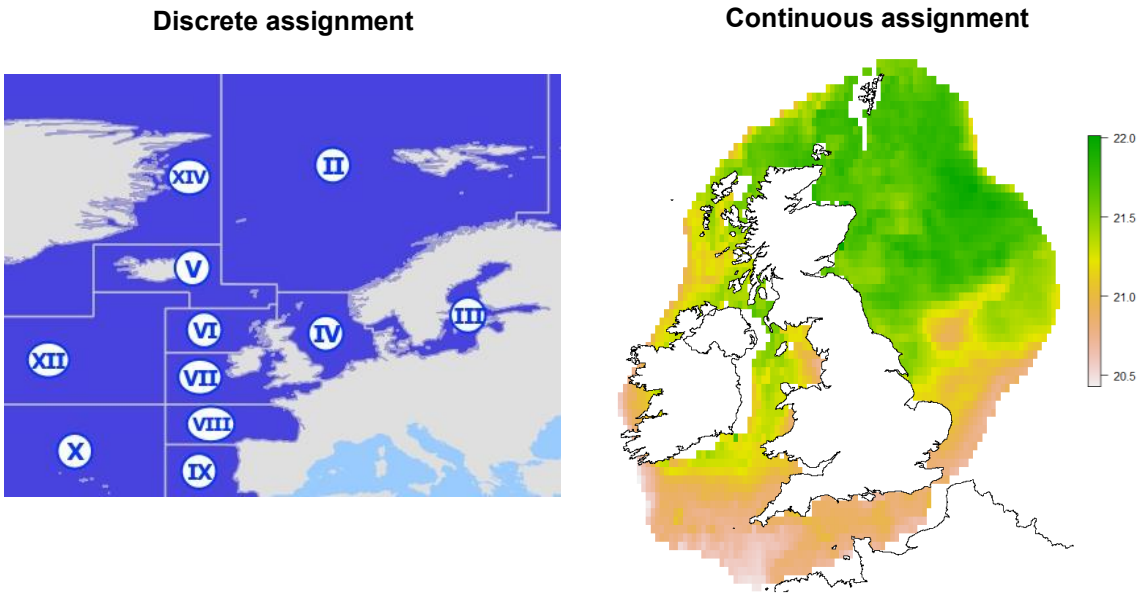


Figure 1.3 Two different methods of spatial assignment - discrete (left) and continuous (right). Discrete assignment divides the sampling area into regions with defined boundaries (image from European Commission (2022)), whereas continuous assignment uses isoscapes to assign individuals to anywhere on a continuous surface (figure created using data from St. John Glew et al. (2019)).

1.5.1.2 Stable isotope spatial traceability studies

Although stable isotope analysis is used routinely to test claims of geographic origin for terrestrial food products, it has not been investigated for the spatial traceability of marine food products to anywhere near the same extent, and those studies that do exist have largely investigated farmed products. For example, stable isotope analysis has been shown to differentiate between shrimps *Penaeus monodon* farmed in Indonesia, the Netherlands, Thailand and India, since the $\delta^{13}C$ and $\delta^{15}N$ values differed significantly (Kim et al., 2015a). However, stable isotope compositions of *Litopenaeus vannamei* farmed in Korea and Ecuador overlapped and could not be distinguished, which indicates that the feed used in those locations may have a very similar composition. Another study assigned farmed Mediterranean mussels to origin by combining stable isotope analysis with trace element analysis (del Rio-Lavín et al., 2022b). Using the stable isotope composition of the muscle tissue along gave an overall classification success of 80.5% and using trace elements in the shell the accuracy was 87.8%. However, a combined approach using both techniques increased the overall assignment accuracy to 97%.

Very few studies have investigated the use of stable isotopes for tracing the geographic origin of wild-caught seafood (Cusa et al., 2021), but among those that have been carried out to date, species include sea bass, tuna, plaice and hake, as well as commercially important invertebrates such as shrimp, scampi, jumbo squid and sea cucumbers. Example assignment studies are detailed in Table 1.1. Carbon and nitrogen stable isotopes were most commonly used to draw inferences about origin. The majority of studies used soft tissues, but a number also focused on traceability using hard tissues such as fish scales and otoliths or the shells of bivalves.

The use of hard tissues for isotopic analysis has the advantage that tissue is laid down incrementally and hard structures are metabolically inert, which means the isotopic tracers become permanently recorded. Hard structures thus give retrospective information on the spatial locations of organisms at different times during the life of the individual and can be used to trace the movements of migratory species, whereas soft tissues are metabolically active and have continuous turnover so only contain information on the last feeding location of the individual. However, the advantage of using soft tissues for traceability of fish and seafood is that the edible part of the product is being tested directly. In many cases hard structures are removed during filleting or processing to create the final product, leaving only the edible parts present. Therefore hard structures such as otoliths and scales often cannot be used for traceability of fish, although bivalve shells are more frequently retained. Since the purpose of the research contained in this thesis is to develop tests for the spatial origin of fish products at any point through the supply chain, all chapters focus on using soft tissues so that the findings can be applicable to a wider range of fish products all the way from landing to point of sale to the consumer.

Table 1.1 Examples of published studies that have measured the bulk stable isotope composition of tissues to determine the geographic origin of marine species.

Taxonomic group / species	Stable isotopes used	Tissue analysed	Citation of study
Hake	$\delta^{13}C$, $\delta^{15}N$	Muscle	Carrera and Gallardo (2017)
Jumbo squid	$\delta^{13}C$, $\delta^{15}N$	Muscle	Gong et al. (2018)
Manila clams	$\delta^{13}C$, $\delta^{15}N$, $\delta^{34}S$, $\delta^{18}O$, $\delta^{2}H$	Adductor muscle	Won et al. (2021)
	$\delta^{13}C$, $\delta^{15}N$	Adductor muscle	Go et al. (2022)
Mussels	$\delta^{13}C$, $\delta^{15}N$	Soft tissue	del Rio-Lavín et al. (2022b)

Plaice	$\delta^{13}C$, $\delta^{15}N$, $\delta^{34}S$, $\delta^{2}H$	Muscle	Pustjens et al. (2018)
Scallops	$\delta^{13}C$, $\delta^{15}N$	Adductor muscle	Trueman et al. (2017)
	$\delta^{13}C$, $\delta^{15}N$	Adductor muscle	Zhang et al. (2019)
Scampi	$\delta^{13}C$, $\delta^{15}N$, $\delta^{34}S$	Abdominal muscle	Slesser and Trueman (2021)
Sea bass	$\delta^{13}C$, $\delta^{15}N$	Muscle	Farabegoli et al. (2018)
	$\delta^{13}C$, $\delta^{15}N$	Muscle	Varrà et al. (2019)
	$\delta^{13}C$, $\delta^{15}N$	Scales	Cambiè et al. (2016)
Sea cucumbers	$\delta^{13}C$, $\delta^{15}N$	Body wall	Zhang et al. (2017)
	$\delta^{13}C$, $\delta^{15}N$, $\delta^{18}O$, $\delta^{2}H$	Body wall	Kang et al. (2020)
Shrimp	$\delta^{13}C$, $\delta^{15}N$	Abdominal muscle	Ortea and Gallardo (2015)
	$\delta^{13}C$, $\delta^{15}N$	Abdominal muscle	Gopi et al. (2019b)
	$\delta^{13}C$, $\delta^{15}N$	Abdominal muscle	de Aquino Ferreira et al. (2021)
Tuna	$\delta^{13}C$, $\delta^{15}N$	Muscle	Chouvelon et al. (2017)
	$\delta^{13}C$, $\delta^{15}N$	Muscle	Rampazzo et al. (2020)
Turbot	$\delta^{13}C$, $\delta^{15}N$	Muscle	Busetto et al. (2008)

Kim et al. (2015b) found differences in the isotopic ratios of three commercial fish species – Mackerel, Yellow Croaker and Pollock – depending on their country of origin. They suggested that $\delta^{13}C$ values could be used to determine geographic origin, but did not attempt to assign fish to their origin. Tanaka et al. (2010) was more successful and distinguished between anchovies from inshore and offshore habitats around Japan using $\delta^{13}C$ values. Yellowfin tuna have also been found to have variable $\delta^{15}N$ values in their muscle tissue across the Pacific and Indian Oceans, showing a north-south trend that reflects the baseline $\delta^{15}N$ values (Lorrain et al., 2015). This indicates that stable isotopes could be used for tracing the origin of tuna within ocean basins on a larger spatial scale.

A number of studies took the method further and attempted to assign individual fish to their geographic origin using stable isotopes analysis. Varrà et al. (2019) combined stable

isotope data (carbon and nitrogen) with rare earth element abundances in sea bass muscle tissue and found that fish could be assigned to the western, central and eastern Mediterranean Sea using these two techniques combined with 90% accuracy. Another study also investigated the provenance of sea bass using carbon and nitrogen stable isotopes present in the scales, and were able to successfully trace individuals to origin on a small spatial scale along the coast of Wales (Cambiè et al., 2016). Around 75% of the sea bass could be correctly assigned to their origin region based on the stable isotope composition. Carrera and Gallardo (2017) classified all commercial hake species to geographic origin in six clusters – European, North African, South African, North American, South American, and Australian – and these could all be clearly distinguished using $\delta^{13}C$ and $\delta^{15}N$ values.

Trueman et al. (2017) created spatial models of isotope variation (isoscapes) for carbon and nitrogen for the North Sea using jellyfish tissue (*Cyanea capillata*). By comparing the isotope values from the tissues of an individual to those in the isoscape, the probability of occurrence for the individual can be estimated to give a most likely origin. This assigns individuals to a continuous surface rather than to discrete regions. Using this technique, Trueman et al. (2017) were able to assign scallops to their geographic origin with >90% accuracy to regions making up >40% of the total area of the North Sea and with a mean error on the order of 100km, which is similar to using tags. Herring were also assigned to their most likely feeding areas, suggesting that larger fish were found in the central northern North Sea and smaller fish in the southern North Sea, which is consistent with acoustic and catch data.

Stable isotopes have also been used to successfully determine the harvesting region of marine invertebrates. For example, Zhang et al. (2019) were able to correctly classify three scallop species to their origin locations around China using carbon and nitrogen stable isotopes with an average accuracy of 92%. Sea cucumbers from various locations around China could also be differentiated based on their carbon, nitrogen, oxygen and hydrogen stable isotope composition (Kang et al., 2020). Another important seafood species for Asian and European countries, the jumbo squid (*Dosidicus gigas*), has been successfully traced to their geographic origin within the eastern Pacific Ocean using stable isotopes and fatty acid analysis (Gong et al., 2018). Slight overlap occurred between two of the geographic regions using stable isotope analysis alone, but this was improved by combining it with the fatty acid analysis and gave a high classification success rate. A study by de Aquino Ferreira et al. (2021) applied the approach to small-scale local shrimp fisheries in Brazil and aimed to contribute to the management of these coastal fisheries. Carbon and nitrogen stable isotopes were utilised to successfully distinguish among penaeid shrimps from different fishing locations, although those from two of the sampled

locations were isotopically similar. It was found that the $\delta^{13}C$ values showed more variability than $\delta^{15}N$ values and so were the most useful for differentiating among stocks.

Almost all studies inferring geographic origin using stable isotopes have focused on carbon and nitrogen. However, a sulfur isoscape has recently been created for the North Sea that indicates there is variability in sulfur isotope values across geographic regions (St. John Glew et al., 2019). One of the only studies to date that has used sulfur stable isotopes for traceability of marine organisms is that by Slesser and Trueman (2021). This study found that the ability to distinguish among catch locations of scampi (*Nephrops norvegicus*) was increased with sulfur isotopes in addition to carbon and nitrogen, achieving 60% accuracy of assignment to six possible fishery origins. Won et al. (2021) also used sulfur stable isotopes, together with a range of other isotopes, to classify Manila clams to three countries of origin.

A recent study by Martino et al. (2022b) made novel advances in developing a universal isoscape on a global scale that was successfully applied to identify the origin of a diverse range of species and taxa, based on oxygen stable isotope compositions in carbonate biominerals (otoliths, statoliths and shells). Assigning test samples of multiple different species to this oxygen isoscape achieved a classification accuracy of up to 90%, so is a promising technique for tracing the geographic origin of seafood on a large spatial scale. Reis-Santos et al. (2022) also highlights the potential for biogeochemical markers, including isotope ratios and trace elements, in calcified structures rather than muscle tissue as a valuable tool for seafood provenance.

It has also been suggested that a compound-specific isotope approach could be an improvement to using bulk stable isotope analyses. Zhao et al. (2018) investigated the use of amino acid carbon stable isotope fingerprinting for differentiating among sea cucumbers (*Apostichopus japonicus*) from eight different geographical areas within China. This study achieved a 100% correct classification rate to region of origin as well as correctly identifying the production method, which suggests that compound-specific isotope analysis of amino acids may be a promising method for tracing the provenance of seafood. Go et al. (2022) also tested this approach by analysing the carbon isotope values in fatty acids found in the adductor muscle of Manila clams, achieving a 100% correct assignment to origin among three countries in Asia.

1.5.2 Genetics

DNA barcoding based on the mitochondrial cytochrome c oxidase subunit 1 has proven to be a reliable method for identifying the species of fish, however this method is less useful for determining the geographic origin (FAO, 2018). Nielsen et al. (2001) demonstrated that Atlantic cod could be assigned correctly to region of origin using DNA microsatellites,

discriminating cod from the North Sea, Baltic Sea and Northeast Arctic. Cod were assigned to known origin with 97-100% accuracy. A study by Nielsen et al. (2012a) also assigned Atlantic cod to known origin among the North Sea, western Baltic and eastern Baltic, based on this method developed by Nielsen et al. (2001). Eastern Baltic cod were assigned with 91% accuracy and western Baltic cod with 75% accuracy. However these studies only distinguished among three relatively broad regions, and the authors noted that the genetic differences among fish populations using this technique are not very large and so it has a low discriminatory power (Nielsen et al., 2012a).

DNA analysis using advanced genetic techniques, such as next-generation sequencing, has recently been investigated for identifying fish provenance as part of the international FishPopTrace project and has shown better potential for differentiating among fish populations (European Commission, 2018). This project used SNP (single-nucleotide polymorphism) analysis as well as otolith analysis to trace cod, hake, herring and sole back to their population of origin. They identified a number of SNPs that varied in populations from different geographic regions. Their results demonstrated that this technique could discriminate between cod from Canada, North Sea, Baltic Sea and Northeast Arctic populations, and between hake from the Mediterranean and the Atlantic, with similar results for herring and sole. At large and medium geographic scales 100% correct assignment to population of origin was achieved. SNPs have also been investigated for tracing the geographic origin of other species including albacore tuna (Montes et al., 2017), anchovies (Montes et al., 2017), lobsters (Villacorta-Rath et al., 2016) and bivalves (Milan et al., 2019, del Rio-Lavín et al., 2022a). A study by del Rio-Lavín et al. (2022a) on the traceability of the Mediterranean mussel, *Mytilus galloprovincialis*, demonstrated that Atlantic, Mediterranean and south-eastern Pacific individuals could be differentiated using a suite of 10-25 SNPs and assigned them to population of origin with 90-100% accuracy.

However, for species or populations where there are smaller genetic differences between individuals from different origins, genetic methods including SNP approaches may be less successful. For example, jumbo squid have been shown to have low genetic diversity so DNA analysis would be unsuitable (Sandoval-Castellanos et al., 2010). Furthermore, on smaller spatial scales where regions to be differentiated are closer together, such as locations around the UK, the genetic differences may not be sufficient to allow discrimination. Low genetic differentiation is especially likely where individual fish migrate or larvae drift across wide regions (Cusa et al., 2021).

Next-generation sequencing was also used in a recent paper with a different approach. Pacific cod were genetically assigned to geographic origins along an isolation-by-distance gradient using RAD (restriction-site-associated DNA) sequencing (Drinan et al., 2018). This enabled cod to be assigned to locations along the gradient that had not been sampled, and in some cases gave an assignment success of >85%. The average distance between

capture and assignment location was 220km. However, this method is only applicable for species showing an isolation-by-distance pattern rather than populations with discrete boundaries.

Another approach has been to analyse the DNA of bacterial communities found on fish, which are assumed to be specific to the geographic area. Pangasius fish from five farms in different regions of Vietnam could be distinguished by carrying out DNA fingerprinting of the microbial communities from the fish skin, gills and intestines (Le Nguyen et al., 2008). Furthermore, distinct bacterial DNA fingerprints were found in the skin mucus of seabass from three farms in Portugal, distinguishing between locations only 500 metres apart (Pimentel et al., 2017).

1.5.3 Trace elements

Multi-elemental analysis of soft tissues has been used to successfully ascertain the geographic origin of several species including sea cucumbers, anchovies, Asian seabass, tuna, octopus, cuttlefish and shrimp. Several studies have also used hard tissues, such as the shells of bivalves, to trace marine species to origin. Example assignment studies using a range of tissue types are detailed in Table 1.2.

Table 1.2 Examples of published studies that have analysed the trace element composition of soft or hard tissues to determine the geographic origin of marine species.

Taxonomic group / species	Number of trace elements measured	Tissue analysed	Citation of study
Anchovy	52	Muscle	Varrà et al. (2021b)
Asian seabass	31	Muscle	Gopi et al. (2019c)
Cockles	5	Shell	Ricardo et al. (2017)
Corbicula clams	14	Shell	Iguchi et al. (2014)
Cuttlefish	52	Mantle and lateral fins	Varrà et al. (2021c)
Mussels	14	Shell	Forleo et al. (2021)
	16	Shell	del Rio-Lavín et al. (2022b)
Octopus	26	Arms	Martino et al. (2022a)

Sea cucumbers	15	Body wall	Liu et al. (2012)
Shrimp	12	Tails without shell	Smith and Watts (2009)
	20	Whole / tails with shell / tails without shell	Li et al. (2014)
	23	Tails with shell	Li et al. (2017)
	31	Abdominal muscle	Gopi et al. (2019b)
Tuna	7	Muscle	Chouvelon et al. (2017)

Li et al. (2014) classified shrimp, *Litopenaeus vannamei,* from three regions in the USA to their geographic origin using 20 elements with 100% success rate. This method was also used by the U.S. Customs and Border Protection Laboratory to confirm the geographic origin of imported farm-raised shrimp (Smith and Watts, 2009). Carter et al. (2015) combined multi-element and stable isotope analysis to distinguish prawns farmed in Australia and those imported from neighbouring Asian countries. Using either $\delta^{13}C$ and $\delta^{2}H$ values or zinc and arsenic concentrations, 100% of prawns were correctly classified to 'Australian' or 'imported' origin. One of the Australian prawn samples appeared more similar to the imported samples, which was due to the feed being sourced from Thailand. Ortea and Gallardo (2015) discovered that using stable isotopes and multi-elemental analyses in combination gave the best results for both wild and farmed shrimps from a range of locations. Using stable isotope analysis alone achieved correct classification of 70% of the samples, whereas the addition of multi-element analysis increased the correct classification to 100%. This emphasises the potential for combining both techniques for assigning geographic origins, as has been highlighted previously in other investigations (Li et al., 2016).

Trace element analysis has also been applied to bivalves, mainly using the elemental composition of the shells. For example, the country of origin of *Corbicula* clams was identified using a range of trace elements contained in the shells, with 89.8% of Japanese clams and 92.2% of Russian clams being correctly classified (Iguchi et al., 2014). Forleo et al. (2021) was able to trace the harvesting locations of *Mytilus galloprovincialis* from the Adriatic coast using the abundance of 14 elements in the mussel shells with a correct assignment rate of over 90%. Another invertebrate, the common cuttlefish, has also been classified to origin using the abundance of a range of elements present in the mantle and lateral fins, obtaining a 100% correct assignment to basin of origin among the Adriatic Sea, Mediterranean Sea and Northeastern Atlantic Ocean (Varrà et al., 2021c).

Varrà et al. (2021b) applied multi-elemental analysis to the provenance of processed fish products. This study measured the concentration of 52 elements in anchovies from three different fishing areas, both before and after processing (salt-ripening). Their results demonstrated that anchovies could be traced to origin using only five elements with over 96% accuracy even after processing.

Trace element analysis is now also being used in the UK seafood industry to prevent fraud, since Loch Duart, in partnership with Oritain, became the first fish farmer in the northern hemisphere to use this technique for traceability of their salmon in 2018 (Loch Duart, 2018). Loch Duart/Oritain claim to be able to trace products to the individual Scottish farm of origin using multi-elemental fingerprinting (Loch Duart, 2022), although raw data are commercially sensitive and such claims cannot be independently verified. Loch Duart is a premium brand, so this technique verifies the provenance of their products and ensures customers are buying genuine Loch Duart salmon.

Traditionally, elemental concentrations have been measured using inductively coupled plasma mass-spectrometry (ICP-MS), inductively coupled plasma atomic emission spectrometry (ICP-AES), or inductively coupled plasma optical emission spectrometry (ICP-OES), all of which involve time-consuming sample preparation (Varrà et al., 2021a). However, more recently x-ray fluorescence (XRF) has been applied for analysing the elemental abundance in a few marine species, which has a much higher sample throughput. Gadd et al. (2018) was the first study to explore the possibility of using the Itrax XRF core scanner to analyse soft biological tissues, rather than the typical application of geological cores, and developed a suitable method to implement this. Following this, Gopi et al. (2019b) investigated the use of the Itrax scanner for origin traceability and was able to assign farmed and wild-caught tiger prawns to geographic origin with correct classification rates of 98% and 100%, and found that there were seven essential elements (Mg, Cu, Ca, K, Fe, Se and Zn) for obtaining a high accuracy of assignment. Asian seabass (*Lates calcarifer*) could not be traced to origin with such a high level of accuracy, but achieved 72% correct classification using this technique (Gopi et al., 2019c). Martino et al. (2022a) also used the Itrax to determine provenance of wild-caught octopus from southeast Asia and Australia. This study combined the isotopic composition of the statoliths with multi-elemental analysis of the soft tissue, and resulted in 94.7% of octopus samples being correctly assigned to region of origin using the two techniques together.

However, a major limitation of multi-element analysis is that element concentrations are affected by factors such as size, age, sex, diet and sampling procedures (Li et al., 2016, Rainbow, 2018), and the mechanisms underlying these variations are not well understood. Therefore, the spatial variation in trace element concentrations cannot be predicted a priori as it can be for stable isotope compositions.

1.5.4 Fatty acids

The use of fatty acid compositions for discerning geographic origins has also been investigated, with mixed success. Fonseca et al. (2022) conducted a study on small-scale fisheries, and found that the bivalve *Scrobicularia plana* and European seabass could be classified to three locations within an estuary that were less than 30km apart using fatty acid profiling. They achieved a high classification accuracy of 80-100% using this technique. Go et al. (2022) also discovered that for Manila clams collected at different sites around Korea, the use of fatty acids in combination with stable isotope analysis improved assignment accuracies compared with using stable isotopes alone (success rates of 98% and 77% respectively). However, in other studies, fatty acids were not able to differentiate between either wild turbot from Denmark and the Netherlands (Busetto et al., 2008) or between Murray cod from different farms (Turchini et al., 2008).

Some studies have found that fatty acids can complement other approaches to increase the distinguishing ability. Gong et al. (2018) demonstrated that stable isotopes alone were not able to distinguish among giant squid from all three locations, whereas the addition of fatty acids resulted in 100% correct classification rate with cross-validation. Similarly, Zhang et al. (2017) found that sea cucumbers from different regions in China showed overlap in their isotopic compositions, so could not be used to distinguish those from all origins, but combining stable isotope and fatty acid markers allowed complete separation of the seven groups using discriminant analysis.

1.5.5 Other techniques

Additional techniques that have been explored for spatial traceability of marine animals include metabolomics and proteomics. Rochfort et al. (2013) found that blue mussels (*Mytilus galloprovincialis*) collected from two different sites in Australia showed significant variation in their aqueous and lipid metabolomic profiles, suggesting that metabolomic approaches could be useful for determining provenance of mussels. A study by Aru et al. (2020) supported these findings, demonstrating that the metabolomic composition of blue mussels from Denmark and Italy were significantly different, although clams (*Ruditapes philippinarum*) from the same two locations showed overlapping metabolomic profiles. Metabolomic approaches have also been successfully applied to determine the geographic origin of both wild and farmed prawns (Chatterjee et al., 2019).

Proteomics has been investigated as a potential tool for distinguishing among populations of European hake (Gonzalez et al., 2010). The results demonstrated that individuals from the Mediterranean Sea, Cantabrian Sea, and Atlantic Ocean could be correctly classified to capture region with 100% success rate using liver and brain protein biomarkers.

1.5.6 Strengths and weaknesses of different techniques

Each of the natural tracers discussed have different benefits and drawbacks for use in spatial traceability of marine organisms. Stable isotope ratios have an important advantage in being predictable a priori over time and space, so maps of the isotopic spatial variability can be created (isoscapes). This allows predictions to be made on the expected isotopic differences among individuals from different regions (Cusa et al., 2021), whereas the mechanisms underlying variations in trace element and fatty acid compositions of tissues are not well understood and so these cannot be reliably predicted. The elemental or fatty acid compositions of muscle tissues in fish may be influenced by a wide range of factors, such as size, age, sex, diet and sampling procedures (Li et al., 2016), but the extent of these impacts are unknown and unpredictable. However, a disadvantage of stable isotope analysis is that an expert user is required to operate the mass spectrometer and the process of preparing and analysing samples is time consuming. This also applies to fatty acid analysis, which has extensive sample preparation and also relies on an experienced user. Until recently, trace element analysis by techniques such as inductively coupled plasma mass-spectrometry (ICP-MS) has also been a time-consuming process, but the introduction of XRF instruments means that analysis of samples can now be completed very rapidly by relatively inexperienced operators. Genetic techniques are also usually relatively laborious, but novel technologies such as nanopore sequencing (Oxford Nanopore Technologies, 2022) may allow more rapid processing of samples for spatial traceability in the future.

Another crucial advantage of XRF technology is the ability to analyse samples in situ with portable devices, which would enable fish products to be tested at any location along the supply chain and provide results almost instantly. The emerging genetic nanopore sequencing technology also offers portability and real-time results, so has significant benefits in this regard. Conversely, stable isotope analysis and fatty acid analysis must be undertaken in a laboratory using instruments that are too heavy and sensitive to be transported.

Finally, the number of dimensions encompassing the tracer may affect how successfully individuals from different regions can be distinguished. Stable isotopes in the marine environment are limited to two or three dimensions ($\delta^{13}C$, $\delta^{15}N$ and sometimes $\delta^{34}S$), since stable isotopes of other elements are relatively constant spatially (Trueman et al., 2012a), whereas trace element analysis, fatty acid analysis and genetics (SNPs) have a much larger number of dimensions. For example, some studies have measured over 30 trace elements present within samples for spatial traceability purposes (Gopi et al., 2019b, Gopi et al., 2019c, Varrà et al., 2021c, Varrà et al., 2021b).

For the research in this thesis, I have chosen to focus on stable isotopes, trace elements using XRF and fatty acids as natural tracers for origin. For stable isotope analysis this was due to the predictability of sample isotopic compositions, whereas the portability and rapid analysis time of XRF technology made it a very promising technique that warranted investigation. Fatty acid analysis was also explored as an alternative technique due to the particularly challenging task of establishing origin for one important study species.

1.6 Equipment for sample analysis

1.6.1 Isotope ratio mass spectrometry

Isotope ratio mass spectrometry (IRMS) allows the accurate measurement of small differences in the abundances of naturally occurring isotopes, such as $^{13}C/^{12}C$, $^{15}N/^{14}N$ or $^{34}S/^{32}S$. The mass spectrometer is coupled with an elemental analyser, which first combusts the solid samples into gases (CO_2, N_2 and SO_2) that are isotopically representative of the original sample. These gases are then introduced into the mass spectrometer where they are ionised by an ion source, forming positive ions by electron impact. The ions are repelled and accelerated away from the ion source with a high voltage and are focused into a beam, before passing through a magnetic field. The ions are deflected by the magnetic field by greater or lesser amounts depending on their mass, resulting in separate ion beams. The ions of different masses are detected by Faraday cups, where the electric charge of the ions results in an electric current that is converted into a digital signal (Figure 1.4).

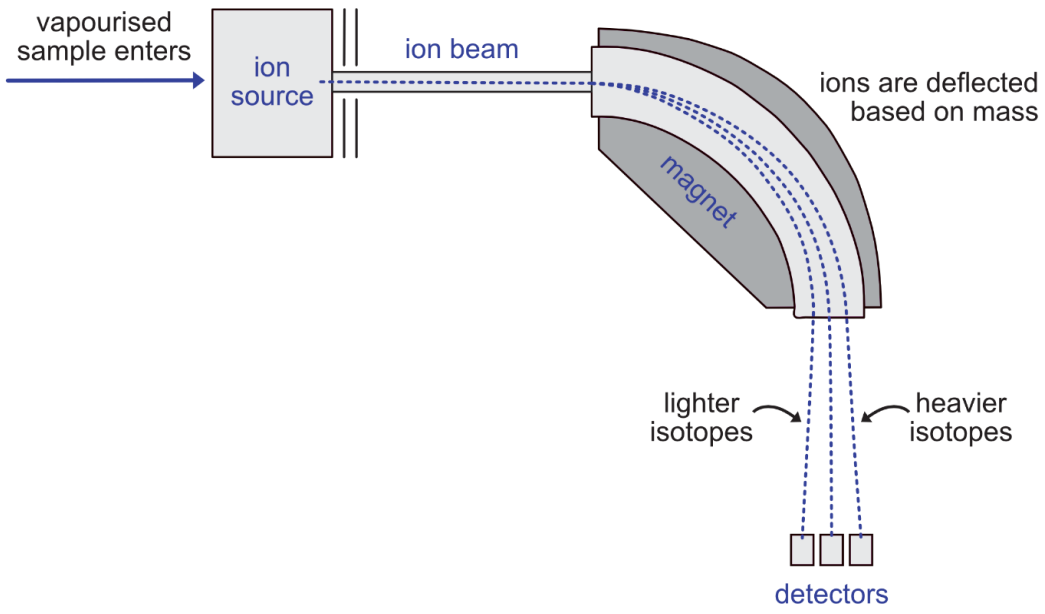


Figure 1.4 Schematic illustrating isotope ratio mass spectrometry (IRMS), which measures the stable isotope composition of samples.

1.6.2 Mass spectrometry for fatty acid analysis

Gas chromatography–mass spectrometry (GC-MS) is used to quantify fatty acids, and comprises a gas chromatograph combined with a mass spectrometer. The sample is introduced to the gas chromatograph (GC) and the molecules are retained by the GC column. Separation of the volatile compounds within the sample is then performed, whereby the sample is heated and the molecules come off the column (elute) at different times depending on their boiling points. Eluted molecules travel along the column and into the mass spectrometer, where a beam of electrons is directed at the molecules and breaks them into fragments. The fragments have different masses depending on where they break. The fragmentation also means that the resulting fragments have a charge. The fragments are repelled by a charged plate and focused into a beam, which passes through a quadrapole. Each pole of the quadrapole is charged differently and so, depending on the voltages, only one fragment size reaches the detector (Figure 1.5). The quadrapole scans through different voltages, producing a spectrum at each time interval. The detector measures the amounts of each fragment present, and the masses of the fragments detected indicate which fatty acids are in the sample.

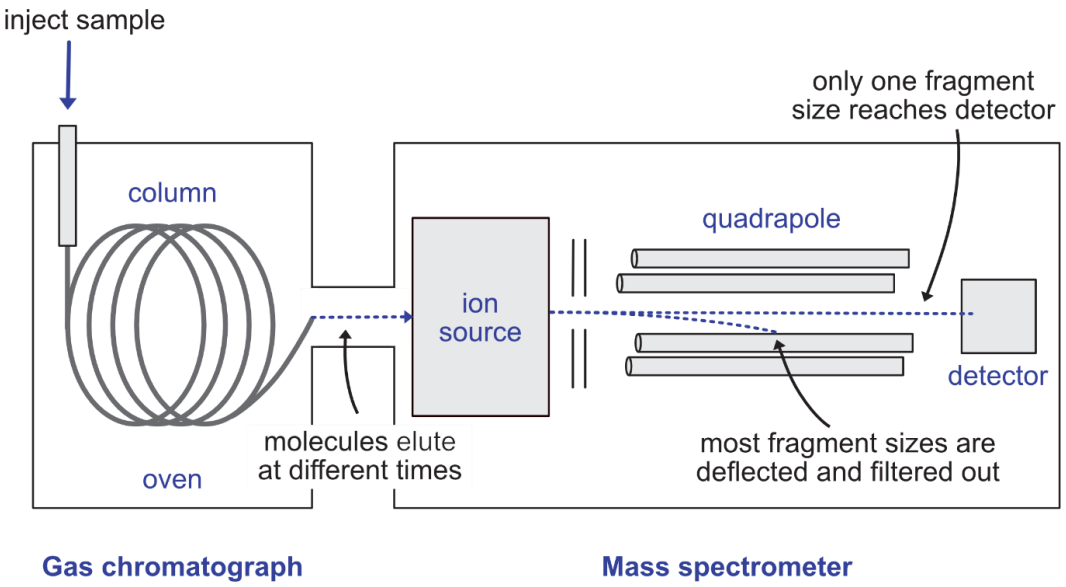


Figure 1.5 Schematic illustrating gas chromatography–mass spectrometry (GC-MS), which analyses the fatty acids present in a sample.

1.6.3 X-ray fluorescence spectrometry

Two different x-ray fluorescence (XRF) instruments are used in this research – the Itrax core scanner and the Vanta analyser. Both estimate the abundance of elements present in the sample in a similar way. An x-ray beam is produced by the XRF instrument that irradiates the sample surface. This ionises the atoms, causing inner-shell electrons to be ejected. Higher energy outer-shell electrons then move inwards to replace them, thereby releasing energy in the form of fluorescent x-rays which are detected by the XRF instrument (Figure 1.6) (Brouwer, 2006). The radiation emitted when an atom is ionised is unique to each element, since all elements have a different electron configuration. Therefore, a sample emits radiation that is characteristic of the atoms present, and the peak area for each element reflects the abundance within the sample (Brouwer, 2006).

The sensitivity of the XRF instruments to different elements varies, with light elements more poorly detected because they have lower fluorescence energies that are more susceptible to scatter and absorption effects, so may be unable to reach the detector at low concentrations (Potts, 1992). However, the Vanta analyser can detect certain light elements more reliably than the Itrax scanner, since the Vanta makes contact with the sample surface unlike the Itrax. The Vanta also internally calibrates measurements and so provides data on the absolute elemental concentrations in parts per million (Olympus, 2017), whereas the Itrax acquires data as relative elemental counts (Gadd et al., 2018).

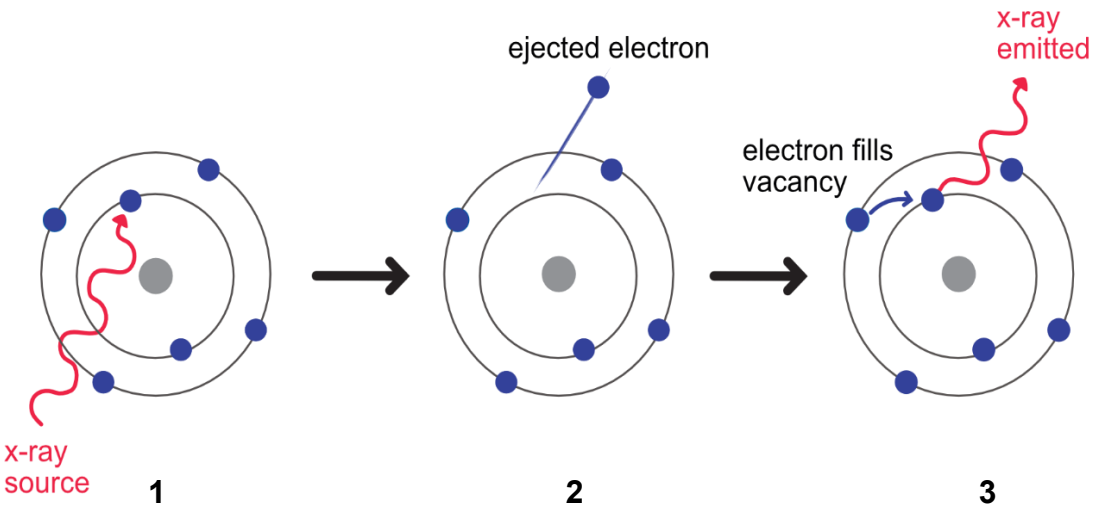


Figure 1.6 Schematic illustrating x-ray fluorescence - 1) an atom is irradiated with an x-ray beam, 2) electrons are ejected and 3) resultant fluorescent x-rays are produced.

1.7 Methods of assigning individuals to origin

Several assignment techniques are used in the research presented here to classify fish samples to their predicted origin based on the biochemical composition of the muscle tissue. The way that each technique works is explained in detail below.

1.7.1 Multivariate normal probability technique

This method uses the data collected to fit a multivariate normal probability distribution for each region sampled. A statistical comparison is then conducted between the test sample and these multivariate normal probability distributions, using the mean and covariance matrix from each region to calculate the likelihood of each being the origin of the sample. The region with the highest probability is deemed the most likely origin for that sample, as illustrated in Figure 1.7. This technique relies on the data for each region having a normal distribution.

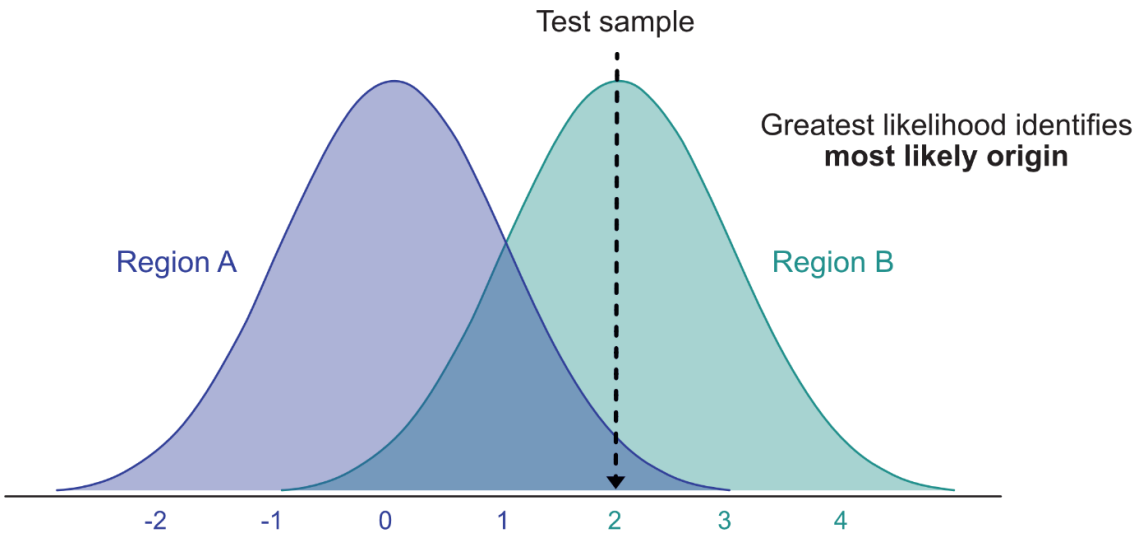


Figure 1.7 Schematic illustrating the multivariate normal probability technique of assigning individuals to origin. The test sample is compared to the multivariate normal probability distributions for each region and that with the greatest likelihood denotes the most likely origin.

To test the accuracy of this multivariate technique, a stratified jack-knifing approach is applied. This involves dividing the samples into two groups at random – the training subset that contains 75% of the samples and the test subset contains the remaining 25%. Each individual from the test subset is then assigned to its most likely origin using the multivariate normal probability distributions of the reference dataset, as explained previously. Since the test samples are of known origin, it can then be revealed whether each had been assigned back to the correct location. This is repeated 1000 times with

different subsets of training and test data selected at random, which allows a mean assignment accuracy to be calculated for each region.

1.7.2 Linear discriminant analysis

Linear discriminant analysis is a classification method that aims to maximise the distance among class means while minimising the variation within each class. The covariance matrix is calculated across all classes, and therefore all the classes have equal covariances. This technique also assumes the data are normally distributed, because the multivariate normal probability distribution is used to calculate the mean and covariance matrix. A linear discriminant function is used to calculate the probability of a test sample coming from each region, and again, the most likely origin is that with the greatest probability.

1.7.3 Random forest classification

Random forest classification is a supervised machine learning algorithm that uses an ensemble of decision trees operating together to classify samples. Each tree within the ensemble is created using a random selection from the training data with replacement (bagging), so that all trees are different. Each tree also uses a different random subset of the variables to classify with. This generates even greater variation among trees and ensures that they are uncorrelated. The decision trees classify samples by creating nodes based on what criteria is necessary to split the training data into their correct groups. A test sample is then run through each individual tree to give a prediction of which group the sample falls into. The group with the most votes from all trees in the random forest (the mode) becomes the algorithm's prediction (Figure 1.8). An ensemble of uncorrelated trees operating together will outperform any of the individual trees it contains.

Random forest classification estimates the error internally. Each tree within the ensemble is created using a different subset of the data at random, where a proportion of the samples are left out and are not used to construct the tree. These remaining samples are run through the tree to give a class prediction. This is carried out on all trees in the random forest and is then used to calculate an out-of-bag (OOB) error rate.

Random forest is a non-parametric technique that does not require data to be normally distributed. However, it can be biased by unbalanced sample sizes across groups as it is more likely to assign to classes containing a greater number of samples.

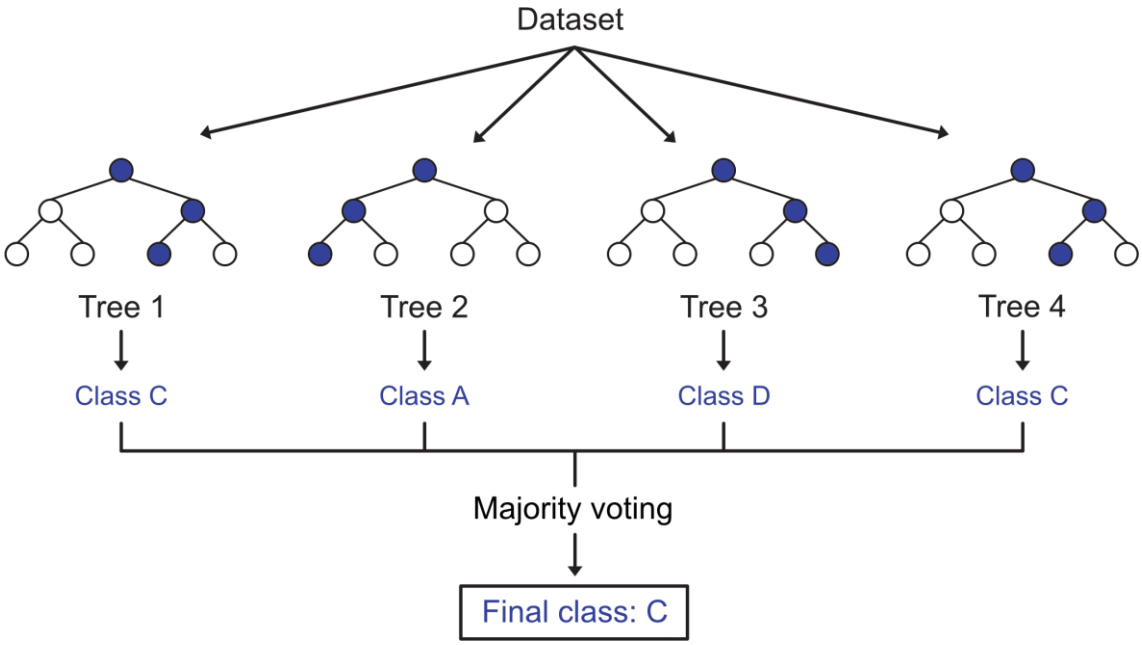


Figure 1.8 Schematic illustrating random forest classification, where a large number of decision trees predict the origin of the test sample and a majority vote takes place.

1.8 Thesis objectives and structure

The research in this thesis aims to address the lack of established techniques for tracing the spatial origin of fish and seafood products, despite its importance in the sustainable exploitation of fisheries as well as protecting fishers, retailers and consumers. The overall objective was to develop tests for the geographic origin of fish products, tailored to specific species, at a level that could have practical use in fisheries management.

Three contrasting techniques are investigated – stable isotope analysis, fatty acid analysis and trace element analysis using X-ray fluorescence. These techniques are applied to three species of commercially important whitefish, comprising Atlantic cod (Chapters 2 and 5), haddock (Chapter 3) and European hake (Chapter 3), as well as the critically endangered European eel (Chapter 4).

The specific objectives of the research were to:

- Create a reference dataset of the stable isotope ratios (carbon, nitrogen and sulfur) within Atlantic cod (*Gadus morhua*), haddock (*Melanogrammus aeglefinus*) and European hake (*Merluccius merluccius*) muscle tissue from a wide variety of regions across the Northeast Atlantic (all species) and the Mediterranean (hake).
- Investigate whether there are variations in the stable isotope compositions of cod, haddock and hake muscle tissue with respect to geographic origin.

- Determine the accuracies with which individual cod, haddock and hake can be assigned back to their known region of capture using stable isotope analysis.
- Measure the stable isotope and fatty acid composition of European glass eels (*Anguilla anguilla*) from four different rivers of origin.
- Investigate whether there are variations in the stable isotope and fatty acid compositions of glass eels among those of different origins.
- Determine the accuracy with which glass eels can be assigned to known origin firstly using stable isotopes and fatty acids independently and then using both approaches combined.
- Develop a successful method for analysing fish muscle samples using XRF instruments to estimate multi-elemental abundance.
- Investigate whether there are variations in the elemental composition of Atlantic cod muscle tissue, as estimated by two XRF instruments, with respect to geographic origin.
- Determine the accuracies with which individual cod can be assigned back to their known region of capture using the two different XRF instruments.
- Investigate the assignment accuracy of cod to region of origin using a combination of both stable isotope analysis and trace element analysis by XRF to determine whether the use of two approaches together increases the distinguishing power.
- Compare the effectiveness of stable isotope analysis and multi-elemental analysis using XRF to trace the spatial origin of Atlantic cod and assess whether these methods show potential for verifying the origin of traded fish products in real world scenarios.

Chapter 2: Tracing the geographic origin of Atlantic cod products using stable isotope analysis

This chapter presents a reference dataset (377 samples) of carbon, nitrogen and sulfur stable isotope ratios measured in Atlantic cod (*Gadus morhua*) muscle tissue collected from ten different fishery regions across the Northeast Atlantic (FAO Area 27). The regional variation in stable isotope composition of the cod samples is then investigated and the accuracy of assignment to known origin is estimated. The assignment accuracy using this technique is compared with that achieved using genetic techniques for cod from the same regions. The isotopic composition of samples collected here are also compared with those from previous datasets, and the accuracy of assignment using these independent samples is assessed.

Chapter 3: Tracing the geographic origin of haddock and European hake products using stable isotope analysis

Here a dataset is presented of the carbon, nitrogen and sulfur stable isotope compositions of 459 haddock (*Melanogrammus aeglefinus*) and 332 European hake (*Merluccius merluccius*) muscle tissue samples collected from a range of fishery regions across the Northeast Atlantic and, for hake, from one region of the Mediterranean Sea. The variation in stable isotope composition with region of origin is investigated and then the accuracy with which individuals can be assigned to known origin is assessed.

Chapter 4: Tracing the catch location of critically endangered European glass eels using stable isotope and fatty acid analyses

Here the findings of a study on the spatial traceability of the critically endangered European eel (*Anguilla anguilla*) are presented. The carbon, nitrogen and sulfur stable isotope ratios of 200 glass eel bodies are measured, and the fatty acid composition of a subset of these samples are also analysed. These data are used to investigate the variation among eels from four different European rivers of origin and the potential for inferring geographic origin of glass eels using stable isotopes and fatty acids are explored.

Chapter 5: Tracing the geographic origin of Atlantic cod products using two X-Ray Fluorescence instruments to estimate multi-elemental composition

The original plan for this chapter was to investigate the use of REIMS (rapid evaporative ionization mass spectrometry), which analyses the metabolites present in samples, for determining the origin of Atlantic cod. However, the analysis would have been conducted at an external organisation and due to the Covid-19 lockdowns, access to this location was not possible.

An alternative opportunity emerged to investigate the possibility of using x-ray fluorescence (XRF) technology to look for regional differentiation among the Atlantic cod samples, in particular the use of the Itrax core scanner and the Vanta handheld analyser. The Vanta has not to my knowledge been used in any studies prior to this for the spatial traceability of marine organisms or even for analysing fish samples in general. This chapter therefore is partly focused on method development, although not optimised, and then the application of this method to explore its potential use for regional differentiation of Atlantic cod and whether further research is justified. The multi-elemental concentrations within the cod muscle tissue samples, as estimated by both the Itrax scanner and Vanta analyser, are presented and the accuracy of assignment to origin using these two instruments are compared.

Chapter 6: Final discussion and summary

This chapter summarises the findings of the research presented in this thesis, and discusses the benefits and limitations of the techniques studied for the spatial traceability of fish products. Future work required to enable the widespread use of these methods in commercial fisheries settings at any point through the supply chain is also suggested.

Supplementary material

The full datasets and all R code scripts associated with this thesis are available in the GitHub repository found at:

https://github.com/julietsewilson/Thesis

Chapter 2 Tracing the geographic origin of Atlantic cod products using stable isotope analysis

2.1 Abstract

Increasing demand for fish and seafood means that traceability of marine products is becoming ever more important for consumers, producers and regulators. Highly complex and globalised supply networks create challenges for verifying the claimed catch region. Tracing the spatial origin of traded fish products retrospectively is crucial for the sustainable harvesting of wild fish and to remove the incentive for fraud or illegal fishing, but currently this is challenging and there are no widely accepted forensic tests for the geographic origin of seafood. Here I investigate whether carbon, nitrogen and sulfur stable isotope ratios in muscle tissue can be used to trace the catch region of an iconic whitefish species with extremely high commercial importance, Atlantic cod (*Gadus morhua*). I measured the isotopic composition of the muscle tissue from 377 cod from ten catch regions across the Northeast Atlantic and Northeast Arctic. Individual cod could be traced back to their true origin with an average assignment accuracy of 70-79% using three different assignment methods, and at least 90% accuracy for certain regions, although sulfur isotopes were found to contribute relatively low informational value. Assignment success rates comparable to those using genetic techniques were achieved when the same origin regions were selected. Stable isotope techniques are unlikely to provide stand-alone tests for proof of origin with accuracy suitable for legislative action, but could be used in conjunction with genetic techniques to result in a higher level of accuracy than could be achieved using either method independently.

2.2 Introduction

Traceability of marine products throughout retail chains is increasingly important for consumers, producers and regulators (Leal et al., 2015, Carter and Chesson, 2017, El Sheikha and Xu, 2017, Lewis and Boyle, 2017). Food fraud is known to be a serious global issue (Pardo et al., 2016, Warner et al., 2016, Interpol/Europol, 2018, FAO, 2018, Spink et al., 2019), whether substituting the species in a product, the undeclared use of additives, fraudulent use of a brand name or concealing the true geographic origin (FAO, 2018). Fish and seafood are particularly vulnerable to incorrect claims of geographic origin due to the spatially based management of global fisheries together with a highly globalised seafood

market and complex trade network (Leal et al., 2015), leading to a risk of both deliberate fraud for financial gain and accidental mislabelling of a product.

Genetic testing has revealed widespread species substitution and mislabelling across some sectors of the seafood trade network (Pardo et al., 2016, Naaum et al., 2016, Warner et al., 2016), including for whitefish (Miller and Mariani, 2010, Helyar et al., 2014). Pardo et al. (2016) conducted a review of 51 peer-reviewed papers investigating seafood species mislabelling using genetic methods, including 4500 samples collected globally, and found on average 30% of samples tested were mislabelled with regard to species. However, in Europe, the incidence of species mislabelling may have reduced since the widespread use of DNA tests (Warner et al., 2016). A recent meta-analysis using Bayesian methods suggests the true mislabelling rate is c.8% on average (Luque and Donlan, 2019).

Verifying the catch location is much more challenging than testing species' identity (El Sheikha and Montet, 2016, Gopi et al., 2019a, Cusa et al., 2021), because a tracer is required that shows variations between multiple different regions on a suitable spatial scale. EU legislation requires all commercially landed fish to be labelled with the catch location (European Union, 2013), but there are no widely accepted forensic tests for spatial origin of seafood products, and no equivalent estimates for the extent of geographic origin mislabelling within retail networks. Current traceability relies mainly on document-based records at each stage in the supply chain. Vessel monitoring systems are also now used in many countries and have improved traceability at first landing, but these systems are often not used on smaller vessels, and vessel-based monitoring with paper (or electronic) chain of custody data cannot address accidental or deliberate mis-association of products to specific vessels. Blockchain technology may strengthen confidence in supply chain documentation but at present blockchain techniques still rely on human input and physical tagging methods to associate the product with blockchain transactions, therefore the data initially entered into the blockchain can be falsified or inaccurate (Blaha and Katafono, 2020).

The stable isotope composition of structural elements in tissues (e.g. O, H, C, N, S, Sr) vary spatially across marine and terrestrial regions and so can provide indirect information on geographic origin. This technique is already widely used for verifying the origin of many terrestrial food products such as wine (West et al., 2007), cheese (Valenti et al., 2017), fruit juice (Rummel et al., 2010), olive oil (Camin et al., 2010) and honey (Bontempo et al., 2017), and is used as evidence in cases of disputed claims of Protected Designation of Origin (PDO) labels on products (Camin et al., 2017). For these products, the stable isotope ratios of a number of elements including carbon, nitrogen, hydrogen, oxygen, sulfur and strontium are utilised for origin traceability (Carter and Chesson, 2017). Natural chemical tags are considerably more challenging to apply for spatial traceability work in wild capture marine food forensic situations as unlike agriculture or mariculture settings,

wild fish are free to move across geographical regions and chemical gradients, and feed on a wide and incompletely known range of resources.

In the ocean, carbon and nitrogen stable isotopes are most often used as tracers for location, although oxygen isotopes can also be used where there are large temperature gradients (Martino et al., 2022b). Sulfur isotope ratios have more recently been shown to vary across open marine environments (St. John Glew et al., 2019), but have been rarely used for marine spatial traceability, with the exception of studies involving scampi (*Nephrops*) (Slesser and Trueman, 2021) and Manila clams (Won et al., 2021). The isotopic composition of phytoplankton varies spatially due to both variations in the isotopic composition of dissolved inorganic nutrients (e.g. CO_2, NO_2, SO_4), as well as isotopic fractionation associated with nutrient fixation. To some extent, spatial variation in the isotopic composition of primary producers can be predicted a priori based on factors such as sea surface temperature, phytoplankton growth rate and species, nutrient source, dissolved CO_2 concentration and water column depth (Magozzi et al., 2017, Somes et al., 2010). The spatially varying isotopic composition of primary producers is subsequently transferred through the food chain and becomes incorporated into the tissues of consumers, acting as a natural marker for animals feeding in different locations. Therefore the isotopic composition of food products can be used to differentiate among those from different locations and to infer origin (Trueman and St. John Glew, 2019). However, consumer tissues are systematically enriched in the heavier isotope of nitrogen and, to a lesser extent, of carbon with each trophic level due to preferential loss of lighter isotopes in excreted metabolites (the isotopic effect of which typically averages around 3-3.4‰ for nitrogen (Minagawa and Wada, 1984) and ~1‰ for carbon (Peterson and Fry, 1987)). Consequently, species from the same geographic origin but of different trophic levels will form tissues with contrasting stable isotope compositions.

Despite being used extensively to identify origin and movement in humans as well as terrestrial animals and food products, the increased practical difficulty of obtaining reference samples in the relatively inaccessible ocean means that stable isotope ratios have not been widely used to trace the origin of wild-caught marine food products. Most studies using stable isotope compositions for seafood origin have focused on farmed animals, such as the shrimps *Penaeus monodon* and *Litopenaeus vannamei* (Carter et al., 2015), often combined with other methods such as trace element analysis. For farmed animals, possible locations are discrete, fixed and relatively easily sampled. In addition, the stable isotope ratios largely depend on the composition of the aquaculture feed, and so this technique can discriminate among farm sites using different feed sources. However, a reference dataset is required comprising samples from each of the farms in question and new samples will be necessary if a different feed is used. Of the relatively few isotope-based studies of spatial origin of wild-caught seafood that have been carried out to date,

species include sea bass (Farabegoli et al., 2018, Varrà et al., 2019), turbot (Busetto et al., 2008), tuna (Chouvelon et al., 2017, Rampazzo et al., 2020) and hake (Carrera and Gallardo, 2017), as well as commercially important invertebrates such as shrimp (Ortea and Gallardo, 2015, Gopi et al., 2019b), scampi (*Nephrops*) (Slesser and Trueman, 2021), scallops (Trueman et al., 2017, Zhang et al., 2019), jumbo squid (Gong et al., 2018), Manila clams (Won et al., 2021) and sea cucumbers (Zhang et al., 2017, Kang et al., 2020). These studies suggest that stable isotope analysis is a promising method for determining origin in some wild-caught species. For example, Trueman et al. (2017) created an isoscape for the UK shelf seas using jellyfish, and used this to assign scallops to their geographic origin with 75% accuracy to an area 30% of the North Sea. In another study, Zhang *et al.* (2019) were able to correctly classify three scallop species to their origin locations around China using carbon and nitrogen stable isotopes with an average accuracy of 92% (Zhang et al., 2019). However, stable isotope analysis alone may not give sufficient accuracy for many traceability applications, and combining isotope markers with another technique, such as genetic, fatty acid or trace element markers, may give better results. For example, Ortea and Gallardo (2015) traced seven shrimp species to multiple geographic origins and found that 71-75.6% were correctly classified using stable isotopes alone, but this increased to 93.5-100% correct when used together with elemental analysis (Ortea and Gallardo, 2015). Gong et al. (2018) also correctly assigned 100% of jumbo squid to three origin locations using both stable isotopes and fatty acid analysis, whereas with stable isotopes alone there was a slight overlap between two origin locations (Gong et al., 2018).

Few studies have focused on tracing the catch locations of open marine whitefish such as cod, haddock and hake using stable isotope analysis. One of the few examples is a study by Carrera and Gallardo (2017) which differentiated among a range of hake species using stable isotopes. In another study, Cusa et al. (2021) predicted the ability to discriminate among Atlantic cod from MSC certified fishery areas by extracting isotopic compositions for each region from global biogeochemical models, and found that cod from the Barents Sea, North Sea and Icelandic waters could be assigned to origin with predicted accuracies of 90%, 89% and 50% respectively. However, this has not been investigated using measured isotope data from cod sampled within these regions. More research is required to better understand the ability of stable isotope ratios to discriminate among seafood products from different geographic origins, using a larger number of samples and from a wider range of locations on a spatial scale relevant to ICES fisheries areas. Furthermore, publishing bias favouring positive results likely skews apparent success rates in all forensic traceability tools.

Genetic techniques have already been successfully applied to determine the origin of whitefish from several populations on a relatively large spatial scale. Nielsen et al. (2012b)

used single nucleotide polymorphisms (SNPs) to discriminate between Atlantic cod among North Sea, Baltic Sea and Northeast Arctic populations, and traced individuals to a priori-defined populations of origin with 98-100% correct classification. Genetic markers clearly offer a powerful tool for assessing geographic origin among populations that are truly reproductively isolated, but this potential is reduced for species with mixed populations, for fisheries operating at the margins of discrete populations, or fisheries targeting common feeding grounds used by multiple reproductively isolated populations.

Combining genetic markers where the signal is based on reproductive isolation with chemical markers where the signal is based on spatial differences in the environment at the point of capture is a potentially powerful way to embrace the strengths and reduce weaknesses of both approaches. Cusa et al. (2021) assessed the potential for combining genetic and isotopic markers in seafood traceability applications, identifying a lack of studies combining both approaches on common samples or using common sampling approaches.

Atlantic cod is an extremely important commercial fish species, found on both the east and west sides of the Atlantic Ocean (ICES, 2005). It has been fished for almost a thousand years, starting in the Viking period and peaking in the 1960's with global landings of around 4 million tonnes a year (Kurlansky, 1998). Historically cod has had a vital influence on the economy and politics of several countries, through events such as the 'Cod Wars' of Iceland and the cod rush at the highly productive Grand Banks cod fishery of Newfoundland (Kurlansky, 1998). Despite drastic declines in its abundance, it remains one of the most commercially important species in the north Atlantic to the present day, with global production totalling 1.1 million tonnes in 2020 (FAO, 2022a).

Landings of cod in the EU were worth EUR 97 million in 2020 (European Commission, 2022), and in the UK cod had the highest value of all demersal species landed (£46.9 million) (MMO, 2021), so cod fisheries make a very significant contribution to employment and the economy in Europe. Atlantic cod is also one of the most highly consumed fish species in many European countries (Seafish, 2020, European Commission, 2022). However, cod has particular social importance in the UK with the longstanding tradition of fish and chips, usually made from cod, being deeply entrenched in British culture and identity. Fish and chips are also probably the most renowned British dish across the globe. UK fish and chip shops serve over 380 million meals a year, representing an annual spend of £1.2 billion (National Federation of Fish Friers, 2023), and fish and chips were responsible for 20% by volume of the total seafood sold in the British food service in 2021 (Seafish, 2022), demonstrating the importance of cod in UK retail.

A large majority of cod consumed in both the EU and UK is supplied by imports, with a value of over EUR 2 billion and £425 million in 2021 respectively (European Commission,

2022, Seafish, 2022). These imported cod are mainly caught in the north Norwegian, Barents Sea and Iceland fisheries (Seafish, 2019) which are certified as sustainable by the MSC (Marine Stewardship Council, 2020b). However, several other regional fisheries also supply cod into the market and some are more at risk of overexploitation. The eastern Baltic Sea has a zero recommended catch limit for cod in 2023 since the spawning stock biomass has decreased and the stock status is poor (ICES, 2022e). Other cod stocks, such as the West of Scotland, Celtic Sea and western Baltic stocks, are also severely depleted with high fishing pressure well above sustainable levels (ICES, 2022a, ICES, 2022b, ICES, 2022d).

The dramatic collapse in the 1990's of the Grand Banks fishery (Kurlansky, 1998), once one of the world's largest fisheries, demonstrates the vulnerability of even very abundant fish stocks to overexploitation. The collapse devastated the economy of Newfoundland, with complete closure of the fishery and the loss of traditional livelihoods for 30,000 people (Berry, 2020). Recovery has been extremely slow and even now, thirty years after the fishery closure, the Grand Banks cod stock is still at a critically low level (DFO, 2022). North Sea cod, another of the most important cod fisheries, also came close to devastating collapse due to high fishing pressure when the stock declined by 84% between the early 1970s and 2006 (Marine Stewardship Council, 2019). The introduction of management plans and a Cod Recovery Zone initially led to the recovery of North Sea cod to the point where it was certified as sustainable by the MSC in 2017 (Marine Stewardship Council, 2019), but the stock has since deteriorated again to below safe biological limits (ICES, 2022c). Consequently, ICES recommended a significant reduction in the total allowable catch (ICES, 2021b) and the MSC certification has been suspended (Marine Stewardship Council, 2019). These examples highlight the vital importance of effective fisheries management, which includes establishing or verifying the catch location of cod in the retail market, for the future sustainability of the cod fisheries.

The development of a reliable and effective test for the geographic origin of traded whitefish is therefore important to maintain sustainable fish stocks under increasing demand, to maximise consumer confidence in products traded as sustainable and to allow the recovery of overfished stocks. Any tool proposed to verify claimed catch location or establish catch location at any point through the supply chain must meet accuracy and precision standards. For terrestrial food products, 95% confidence limits are used for authenticity testing of commercial samples (Camin et al., 2017).

Discrete assignment traces a test sample to one of a number of distinct regions and known origin reference samples must be obtained from each of these regions. Test samples are compared to the reference samples from all the regions to determine which is the most likely origin. Samples can only be assigned to one of the discrete regions contained in the reference dataset, so any test samples originating from outside of these regions will be

incorrectly assigned. It is important to consider the method for testing accuracy of assignment, since falsely high estimates of accuracy and precision can result if the reference samples do not cover the full range of the fish, the regions are decided a priori based on expected differences, or the reference dataset is not independent of the test samples used to assess accuracy.

The aim of this study is to investigate whether the stable isotope composition of Atlantic cod muscle can provide geographic discrimination at a level that has practical use in fisheries management, to compare levels of spatial differentiation expressed in tissue isotope compositions with previously assessed genetic markers and assess the contribution that stable isotope markers can make as a forensic tool for verifying the spatial origin of traded fish products. We measured the carbon, nitrogen and sulfur stable isotope ratios in 377 Atlantic cod muscle tissue samples of known geographic origin to construct a reference dataset of fish from a wide range of catch locations across the Northeast Atlantic. We then used this to determine the accuracy with which individuals can be traced back to their true origin region.

2.3 Methodology

2.3.1 Sampling

Atlantic cod (*Gadus morhua*) was selected as a study species for this investigation due to the extensive research already published on its ecology, genetics and biochemical composition, in addition to the high commercial importance of the species. The use of genetic techniques for tracing cod from different populations of origin has already been explored by Nielsen et al. (2012b), who successfully discriminated among Atlantic cod from the North Sea, Baltic Sea and Northeast Arctic populations using single nucleotide polymorphisms (SNPs). Cusa et al. (2021) concluded that the direct comparison of genetic and stable isotope techniques is prevented by a lack of samples from the same populations and areas for both analyses. Therefore, investigation into the spatial traceability of cod to the same regions using alternative tracers, such as stable isotopes and trace elements, would allow comparison with genetic methods for the same model species.

Samples of Atlantic cod were collected between February and December 2018 from nine regions within the Northeast Atlantic Ocean, including those broadly comparable to regions sampled by Nielsen et al. (2012b), with multiple stations sampled at each region (Figure 2.1). The samples were obtained from research fisheries surveys or fish tagging studies carried out by the relevant research organisation in the area of interest to ensure authenticity of the catch locations. The sample suite was extended with additional samples provided by Young's Seafood Ltd. from the Barents Sea caught in 2017, however no exact

locations for these samples are known (only that they were caught in the Central Bank region). The research organisations and sources who supplied the samples are listed in Table 2.1.

The regions sampled were selected based on the main commercial fishery areas supplying the trade network as well as the availability of samples. The regions also correspond to the ICES subareas within FAO major fishing area 27, where stations for each region are contained within one subarea, with the exception of the region referred to as the Norwegian Sea which consists of stations falling within both ICES subareas 1 and 2 (Barents and Norwegian Sea subareas) close to the boundary between the two (Figure 2.1).

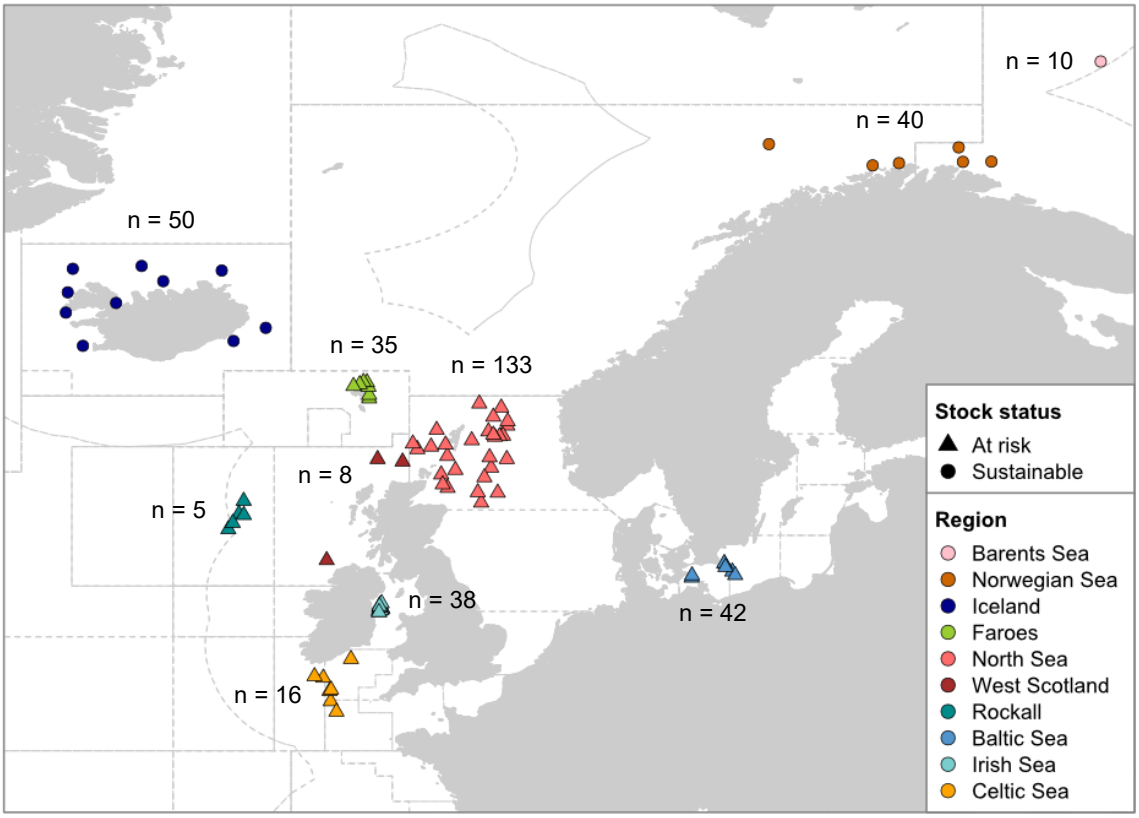


Figure 2.1 Locations of all stations where samples of Atlantic cod were obtained, coloured according to the geographic regions. The total number of samples (n) collected from each region are shown and the symbols indicate whether the stock is sustainable or at risk (as determined from the ICES 2022 advice for each stock (ICES, 2022h), based on the spawning stock biomass relative to the limit where reproduction of the stock is impaired (B_{lim})). ICES subarea boundaries are indicated by the dotted grey lines.

We aimed to sample at least 50 individuals from each of the regions, to estimate the natural variability of wild populations. However, this was not always possible. The total numbers of samples collected from each region are listed in Table 2.1. Sampling from multiple stations ensured that the samples were not all derived from one shoal, which may have travelled as

a group and consequently have very similar isotope compositions. Ideally, a maximum of five individuals were sampled from each station, although this was not always possible due to sampling constraints. The body sizes of the fish sampled were not measured, so that they would be representative of typical samples likely to be encountered in the context of traceability. However, only individuals of commercial size were used (>35cm).

A small sample of white muscle tissue (approximately $3cm^3$) was removed from each individual and placed in a zip-lock bag. The catch location, species, date collected and station number (haul) were recorded for each. The muscle tissue samples were placed immediately into a -20°C freezer for transport and were kept frozen until processing in the laboratory. The only exception were samples from the Faroe Islands, which were instead preserved in 70% ethanol due to the transportation time back to the UK.

Table 2.1 Total number of individual Atlantic cod collected from each geographic region, the number of stations sampled per region, and the source from which samples were obtained.

Catch region	Number of cod sampled	Number of stations	Year samples collected	Source of samples and organisation name
Barents Sea	10	Unknown	2017	Young's Seafood Ltd.
Norwegian Sea	40	6	2018	Annual fisheries survey - Institute of Marine Research
Iceland	50	10	2018	Annual fisheries survey - Marine and Freshwater Research Institute
Faroe Islands	35	7	2018	Annual fisheries survey - Faroe Marine Institute
North Sea	133	28	2018	Annual fisheries survey - Marine Scotland
West Scotland	8	3	2018	Annual fisheries survey - Marine Scotland
Rockall	5	5	2018	Annual fisheries survey - Marine Scotland
Baltic Sea	42	7	2018	Fish tagging survey - Technical University of Denmark
Irish Sea	38	16	2018	Fish tagging survey - Marine Institute
Celtic Sea	16	7	2018	Annual fisheries survey - Ifremer
Total	**377**	**89**		

2.3.2 Stable isotope analysis

The frozen whitefish muscle samples were freeze-dried at -55°C for 24 hours, and then homogenised into a powder with scissors. Care was taken not to include any skin with the muscle tissue. Between 2.5mg and 2.8mg of each powdered sample was weighed into tin capsules for analysis. A total of 377 samples were analysed for bulk carbon, nitrogen and sulfur stable isotope ratios. Most samples were analysed at the Life Sciences Mass Spectrometry Facility (LSMSF) in East Kilbride, United Kingdom, but a minority of samples were measured at the University of Southampton based at the National Oceanography Centre. Samples were analysed at both laboratories using an Elementar vario PYRO cube elemental analyser coupled with an Isoprime visION isotope ratio mass spectrometer.

Stable isotope compositions are expressed as $\delta^{13}C$, $\delta^{15}N$ and $\delta^{34}S$ values in per mille (‰) relative to the international standards (Vienna-PeeDee Belemnite, Air and Vienna-Canyon Diablo Troilite). Accuracy and precision were monitored through laboratory internal standards (LSMSF: MSAG, M2 and SAAG2; University of Southampton: sulfanilamide and a protein standard), as well as two common standards at run at both laboratories to ensure consistent measurements (fish muscle standard and an in-house glutamic acid standard). Blanks were also run in each batch.

Lipids are depleted in $\delta^{13}C$ compared to proteins and carbohydrates, therefore $\delta^{13}C$ values in samples with a high lipid content (C:N ratios >3.4) as defined by Skinner et al. (2016), were corrected arithmetically using the methodology described by Kiljunen et al. (2006). For samples where the C:N ratio was 3.4 or lower, no correction was necessary due to the low lipid content. A total of 2 cod samples were corrected.

Samples from the Barents Sea were collected a year prior to all other samples, so the $\delta^{13}C$ values for these were corrected for the Suess effect by applying a -0.022‰ per year adjustment (Young et al., 2013). The Suess effect causes decreasing $\delta^{13}C$ values over time as atmospheric CO_2 levels rise.

2.3.3 Statistical analysis

In this study, samples were assigned to one of a suite of discrete possible populations based on comparing the isotopic composition of each sample to the distribution of isotopic values measured within a reference population of known origin. Increasingly, continuous surface reference isotope models (isoscapes) are available, allowing probabilistic assignment of a sample to possible locations within a region without limiting the possible regions a priori (Wunder, 2010, Trueman and St. John Glew, 2019, Ma et al., 2020). However, isoscapes are not available across the entire sampled region, and genetic approaches cannot employ such continuous surface assignment methods. We therefore

employ discrete assignment methods here and divide the sampling area into pre-defined geographic regions reflecting geopolitical and hydrographic basins. Reference samples of known origin were collected and measured for each region, to which test samples can be compared. The most likely origin is the region that the test sample's isotope composition matches most closely. This method is well suited to traceability in a fisheries context, where the ocean is divided into distinct management units such as FAO subareas.

All analyses were performed using R version 3.6.2 (R Core Team, 2019). Multiple methods of assignment to origin were compared to determine the most effective. First, a statistical comparison between isotopic compositions in each 'unknown' sample and the reference population was conducted based on multivariate normal probability distributions. The carbon, nitrogen and sulfur stable isotope data for each location were used to fit a multivariate normal probability distribution for that location. A subset of test samples were then compared to these probability distributions. The 'dmvnorm' function within the 'mvtnorm' package (Genz et al., 2020) in R statistical software was used to calculate the probability density function of the multivariate normal distribution and therefore give the likelihoods of the samples having originated from each location. The location with the greatest likelihood is deemed the most likely origin for that sample.

To estimate the accuracy of assignment, a stratified jack-knifing approach was applied. A random subset of 25% of the samples from each location was extracted from the dataset and used as a test subset to be assigned, and the remaining 75% of the samples were used as the reference datasets. Each individual from the test subset was assigned to its most likely origin using the multivariate normal probability distributions of the reference dataset. Since the test samples are of known origin, it could then be revealed whether each had been assigned back to the correct location. This was repeated 1000 times with different random subsets of test and reference data and a mean was calculated for each region. The results of these repeated simulations were plotted on a boxplot to display the percentages assigned to each location.

A similar technique was also applied using only carbon and nitrogen stable isotope data, to allow the importance of sulfur isotopes for spatial traceability to be assessed. The bivariate normal probability distributions for each location were fitted using the carbon and nitrogen isotope data, and then a subset of test samples was compared to these. Again, the 'dmvnorm' function within the R package 'mvtnorm' (Genz et al., 2020) was used for this, calculating the probability density function of the bivariate normal distribution and therefore the likelihoods of the samples having originated from each location. The location with the greatest likelihood was deemed the most likely origin for each sample. A similar jack-knifing approach was employed, where 1000 repeat simulations were carried out using different random subsets of test and training data.

The multivariate probability density function used to calculate the likelihoods for multiple variables is defined as:

$$p(x \mid \mu, \Sigma) = \frac{1}{\sqrt{(2\pi)^d |\Sigma|}} \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1} (x - \mu)\right)$$

Where x is a random vector of size d, μ is the mean vector, Σ is the symmetric, positive definite covariance matrix of size d x d, $|\Sigma|$ is its determinant and T is its transpose.

As a further test of assignment accuracy using this multivariate normal probability method, leave-one-out cross validation was conducted by removing each sample in turn and assigning them using all the remaining samples. The most likely origin calculated for each sample was compared to their true geographic origin, and so the number of correctly assigned samples could be determined.

Two other tests were carried out to determine how well individuals could be assigned to their origin – linear discriminant analysis and random forest classification – and these results were compared to the multivariate normal probability method. Linear discriminant analysis was performed using the 'MASS' package (Venables and Ripley, 2002) in R statistical software. A leave-one-out approach was applied, where each sample in the dataset was taken as the test sample and assigned to its most likely origin in turn, with all the remaining data used as the reference subset to train the linear discriminant model.

Random forest classification is based on an ensemble of decision trees, and it is a non-parametric technique that does not require data to be normally distributed. Classification by random forest was conducted using the 'randomForest' package (Liaw and Wiener, 2002) in R statistical software. A leave-one-out approach was applied, where each sample in the dataset was taken as the test sample and assigned to its most likely origin in turn, using all the remaining data as the training subset to create the random forest model. The sample sizes were balanced using the 'sampsize' parameter within the randomForest package, which draws a specified number of random samples from each group to create the trees, since unbalanced sample sizes may result in the random forest being biased towards groups with a greater number of samples (Bader-El-Den et al., 2018). The sample size specified using 'sampsize' was set as one less than the number of samples in the smallest class in the training dataset (after removing one as the test sample), to allow for random variation among trees, and the same number of samples was drawn from all other regions. The classification was fine-tuned by selecting the values for 'ntree' (number of trees in the random forest) and 'mtry' (the number of variables used at each node split) to give the highest accuracy of assignment. The assignment accuracy was determined as the total number of correctly assigned test samples in each region, and this was also reported as a percentage of the total samples. Random forest classification also estimates the error internally. Each tree within the ensemble is created using a different subset of the data at

random, where a proportion of the samples are left out and are not used to construct the tree. This approach is used to give an out-of-bag error rate.

Each of these methods above suffers from non-independence of training and test samples, which is a common limitation in many assignment studies, and probably results in overestimated assignment accuracy. Assigning independently collected samples from populations of known origin would be a more robust test of assignment accuracy. In an attempt to address this issue, a dataset of carbon and nitrogen stable isotope ratios measured in cod was gathered by acquiring independent data from a range of sources. Published data from a previous study by Jennings and Cogan (2015) contributed isotope ratios measured in cod during a range of years spanning 2002 to 2010 from the North Sea, Irish Sea and Celtic Sea. Another dataset was provided by Ifremer from the EVHOE 2014 survey, which includes isotope ratios of cod caught in 2014 and 2015 in the Celtic Sea. These data were combined with results obtained from cod caught in the Barents Sea by the Institute of Marine Research, Norway and Icelandic cod provided by Young's seafood. The $\delta^{13}C$ data were corrected for the Suess effect by applying a -0.022‰ per year adjustment (Young et al., 2013) before being assigned to their most likely location using the bivariate assignment method described previously with the known origin cod in the current study as a reference dataset. This allowed the success to be investigated of tracing individual cod provenance when the test samples were independent and collected during different years than the reference samples.

2.3.3.1 Comparison with genetic techniques

Cod samples from the same three regions as in Nielsen et al. (2012b) - Barents Sea, North Sea and Baltic Sea - were selected from the full dataset. The same methodology for multivariate assignment as described previously was applied using only these regions, where the data were divided into test (25%) and training (75%) subsets and the test samples were assigned to the most likely origin to determine the accuracy of assignment. This was repeated 1000 times with different randomly selected test and training subsets.

As a further comparison to the assignment using genetic techniques in Nielsen et al. (2012b), where the likelihood of the test sample originating from the true population of origin was at least six times higher than for the second most likely population of origin, we discarded any assignments using stable isotope analysis from the same three regions where the probability for the most likely origin was less than six times that of the second most likely origin. The assignment accuracy with only these 'strong' assignments was then determined and compared to that obtained using genetic techniques.

2.3.3.2 North Sea isoscape

Cod samples were collected from 28 different stations within the North Sea, so the spatial variation among stations could also be investigated. Carbon, nitrogen and sulfur isoscapes were created for the sampled North Sea area using interpolation by kriging between the sampled stations, with a grid size of 0.1 degrees. Between two and seven samples were measured per station, so the mean value for each station was calculated. The associated spatial variances were also calculated for each of the three isoscapes. These analyses were performed using the 'gstat' package (Pebesma, 2004, Gräler et al., 2016) in R version 3.6.2 (R Core Team, 2019).

2.4 Results

Measurement error associated with the stable isotope analysis of the fish muscle samples is 0.1-0.4‰ for $\delta^{13}C$, 0.1-0.2‰ for $\delta^{15}N$ and 0.6-0.7‰ for $\delta^{34}S$, determined as the standard deviation of replicate measurements of two internal standards - a fish muscle standard and a glutamic acid standard - measured together with the samples. The internal standards used were consistent across analytical laboratories, allowing comparison between the measurements from the two sites. Means of repeated analyses of fish muscle and glutamic acid standards from both laboratories are within one standard deviation of each other (Table 2.2), demonstrating that there are no significant differences in the measurements between laboratories.

Table 2.2 Comparison of the mean values and uncertainties (standard deviations) in the stable isotope measurements of two internal standards measured at both laboratories where samples were analysed.

Standard	Laboratory	$\delta^{13}C$ (‰)		$\delta^{15}N$ (‰)		$\delta^{34}S$ (‰)	
		Mean	SD	Mean	SD	Mean	SD
Fish muscle	NOCS	-19.3	0.11	11.4	0.17	19.3	0.64
	LSMSF	-19.3	0.07	11.3	0.11	18.5	0.74
Glutamic acid	NOCS	-13.1	0.17	-3.8	0.06	NA	NA
	LSMSF	-13.6	0.43	-3.9	0.15	NA	NA

The C:N ratios indicate that the different storage method for the Faroes samples (70% ethanol prior to freeze drying) had very little effect on the stable isotope values, since the C:N ratios are very similar to the other samples and suggests that no extraction of lipids occurred.

The stable isotope compositions of Atlantic cod muscle samples differ significantly among the ten regions (ANOVA - $\delta^{13}C$: f=187.3, d.f.=10, $p<2e^{-16}$; $\delta^{15}N$: f=119.2, d.f.=10, $p<2e^{-16}$; $\delta^{34}S$: f=15.62, d.f.=10, $p<2e^{-16}$). The $\delta^{13}C$, $\delta^{15}N$ and $\delta^{34}S$ values measured in the cod samples from each region are shown in Figure 2.2, and the means and standard deviations for each location are listed in Table 2.3. Overall, $\delta^{13}C$ values range from -15.6‰ to -22.4‰. The lowest mean $\delta^{13}C$ value of -21.1‰ was found in the Baltic Sea, but the Norwegian Sea samples also have lower $\delta^{13}C$ values than those in almost all other regions. The greatest variation in $\delta^{13}C$ values occurs in the Iceland samples, with a range of 4.5‰. The Irish Sea also has a large range due to one outlier with a much lower $\delta^{13}C$ value than all others from the region. The Irish Sea has the highest mean $\delta^{13}C$ value of -16.7‰ as well as the highest mean $\delta^{15}N$ of 16.9‰. The Celtic and Barents Sea samples also have relatively high $\delta^{15}N$ values, whereas $\delta^{15}N$ values for the other locations are more similar – mean values between 12.7‰ and 14.3‰ – apart from cod from Rockall with a much lower mean $\delta^{15}N$ value of 10.5‰. The mean values of $\delta^{34}S$ show less variation across regions than for carbon and nitrogen isotopes, ranging between 18.1‰ and 19.7‰, so values for many regions overlap. However, the $\delta^{34}S$ values generally have greater variation among individuals within regions than for $\delta^{13}C$ or $\delta^{15}N$, with the largest ranges of 5.1‰ and 4.9‰ for samples from Iceland and the Baltic Sea respectively. The Norwegian Sea is the most distinctive in terms of sulfur isotope values, since the samples in this region have the highest mean $\delta^{34}S$ value and the range is relatively small compared to other locations, although the West Scotland samples also have relatively high $\delta^{34}S$ values. Figure 2.3 shows frequency distributions for the carbon, nitrogen and sulfur stable isotope data for cod from each region, and demonstrates that the majority of regions show a normal distribution.

Table 2.3 Mean and standard deviation of carbon (lipid corrected), nitrogen and sulfur stable isotope ratios from Atlantic cod caught in each of the sampled geographic regions.

Region	$\delta^{13}C$ (‰)		$\delta^{15}N$ (‰)		$\delta^{34}S$ (‰)	
	Mean	SD	Mean	SD	Mean	SD
Barents Sea	-19.5	0.48	15.3	0.64	18.2	0.64
Norwegian Sea	-20.2	0.30	14.3	0.42	19.7	0.50
Iceland	-19.0	1.05	13.8	0.87	18.7	0.92
Faroe Islands	-17.4	0.37	13.1	0.64	18.5	0.41
North Sea	-18.0	0.42	13.8	0.66	19.0	0.50
West Scotland	-18.3	0.41	13.5	0.73	19.4	0.55
Rockall	-18.8	0.43	10.5	0.63	18.7	0.63
Baltic Sea	-21.1	0.36	12.7	0.49	18.1	1.10
Irish Sea	-16.7	0.76	16.9	0.83	18.2	0.83
Celtic Sea	-17.4	0.51	15.5	0.77	18.8	0.70

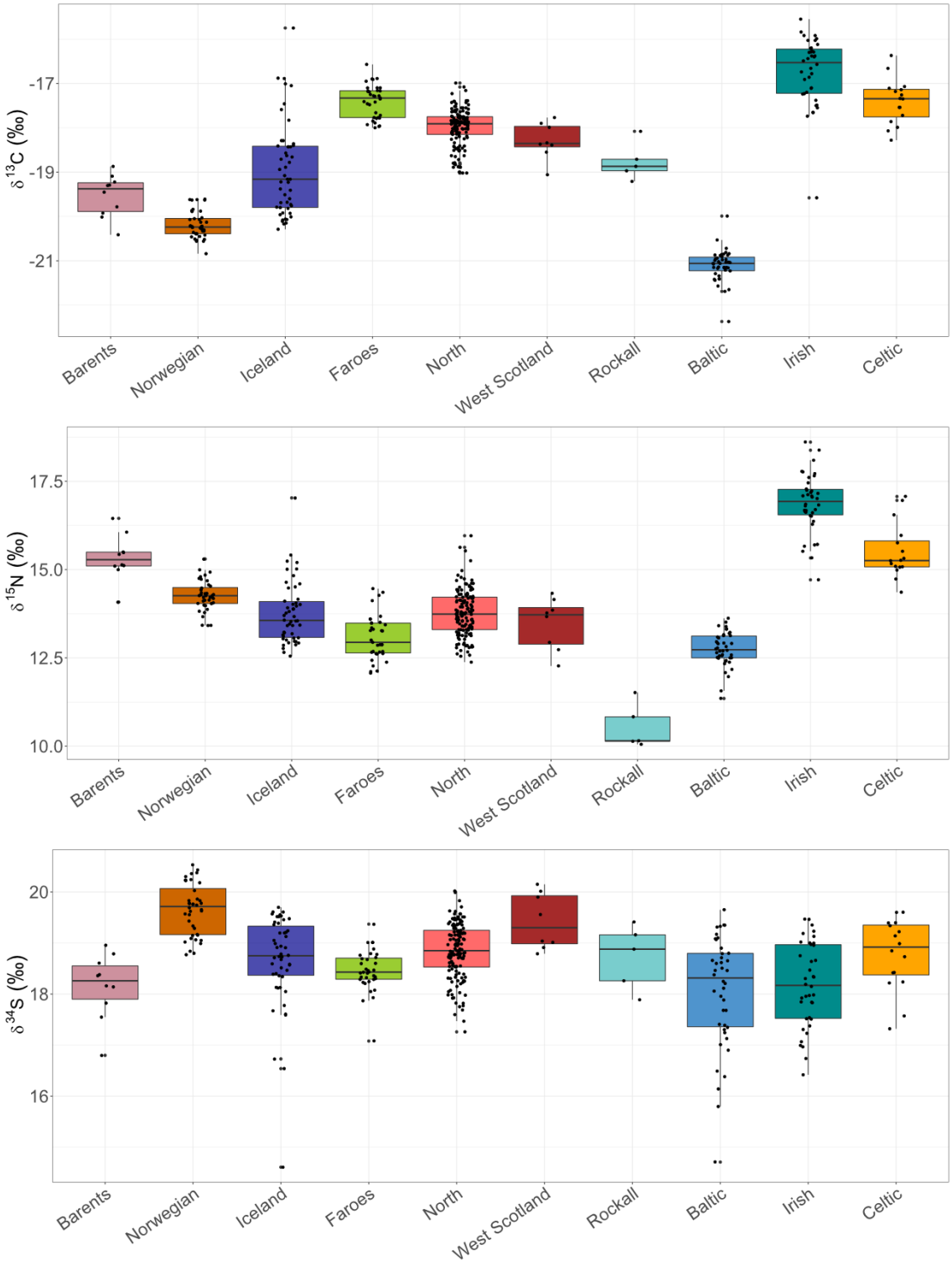


Figure 2.2 Lipid-corrected $\delta^{13}C$, $\delta^{15}N$ and $\delta^{34}S$ values measured in the Atlantic cod muscle tissue samples from each geographic region.

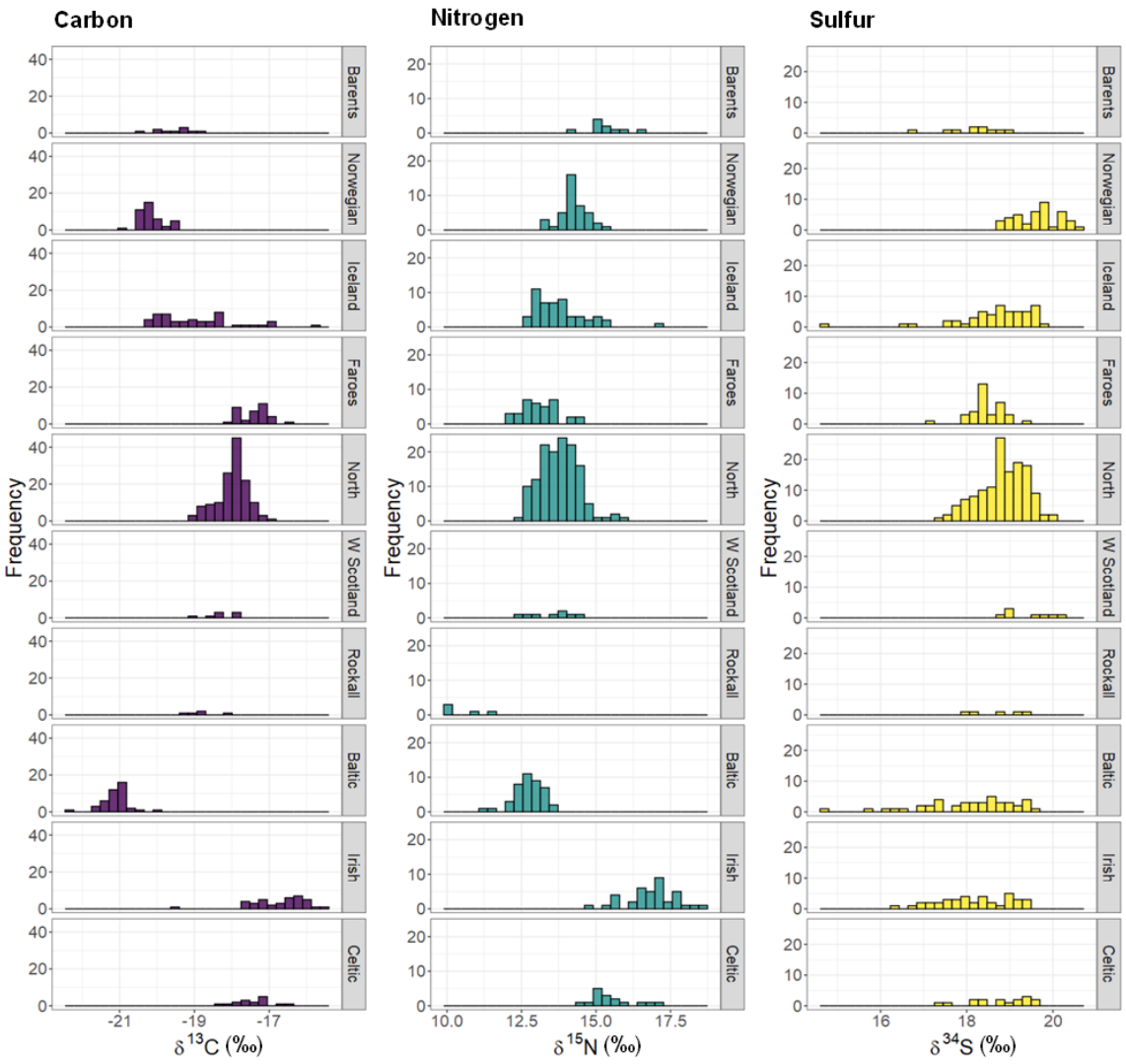


Figure 2.3 Frequency distributions of $\delta^{13}C$, $\delta^{15}N$ and $\delta^{34}S$ values from Atlantic cod caught within each of the ten sampled regions.

2.4.1 Clustering by region

The results of the stable isotope analysis reveal clustering of the cod samples by region, as shown in Figure 2.4. Overall, samples are separated by $\delta^{13}C$ values into two main groups. The first consists of the Baltic, Norwegian and Barents Seas with low $\delta^{13}C$ values due to the low salinity and the cold temperatures respectively, and then another group with warmer, more southerly regions around the Faroes and UK shelf seas where the $\delta^{13}C$ values are higher. Samples from Iceland show the largest spread of $\delta^{13}C$ values, overlapping both of these groups.

The Rockall samples are very distinct, with hardly any overlap with the other regions due to the lower $\delta^{15}N$ values, however this is based on a small sample size. The Baltic samples also form a distinct cluster, which do not overlap with samples from any other locations.

The Norwegian and Barents Sea samples show good separation from the remaining regions, but some samples from the two regions are isotopically similar. Other regions are less distinct in their isotopic composition, including the North Sea, Faroes, West Scotland, Celtic Sea and Iceland, which although form clear clusters, show a larger area of overlap in their $\delta^{13}C$ and $\delta^{15}N$ values. However, cod caught in the Irish Sea generally had higher $\delta^{15}N$ values than the other locations and only overlapped with samples from the Celtic Sea, likely due to the close geographic proximity of the two regions.

The $\delta^{34}S$ values are very variable and overlap in samples from all regions, indicating that sulfur isotopes would have limited use in distinguishing between samples from these locations.

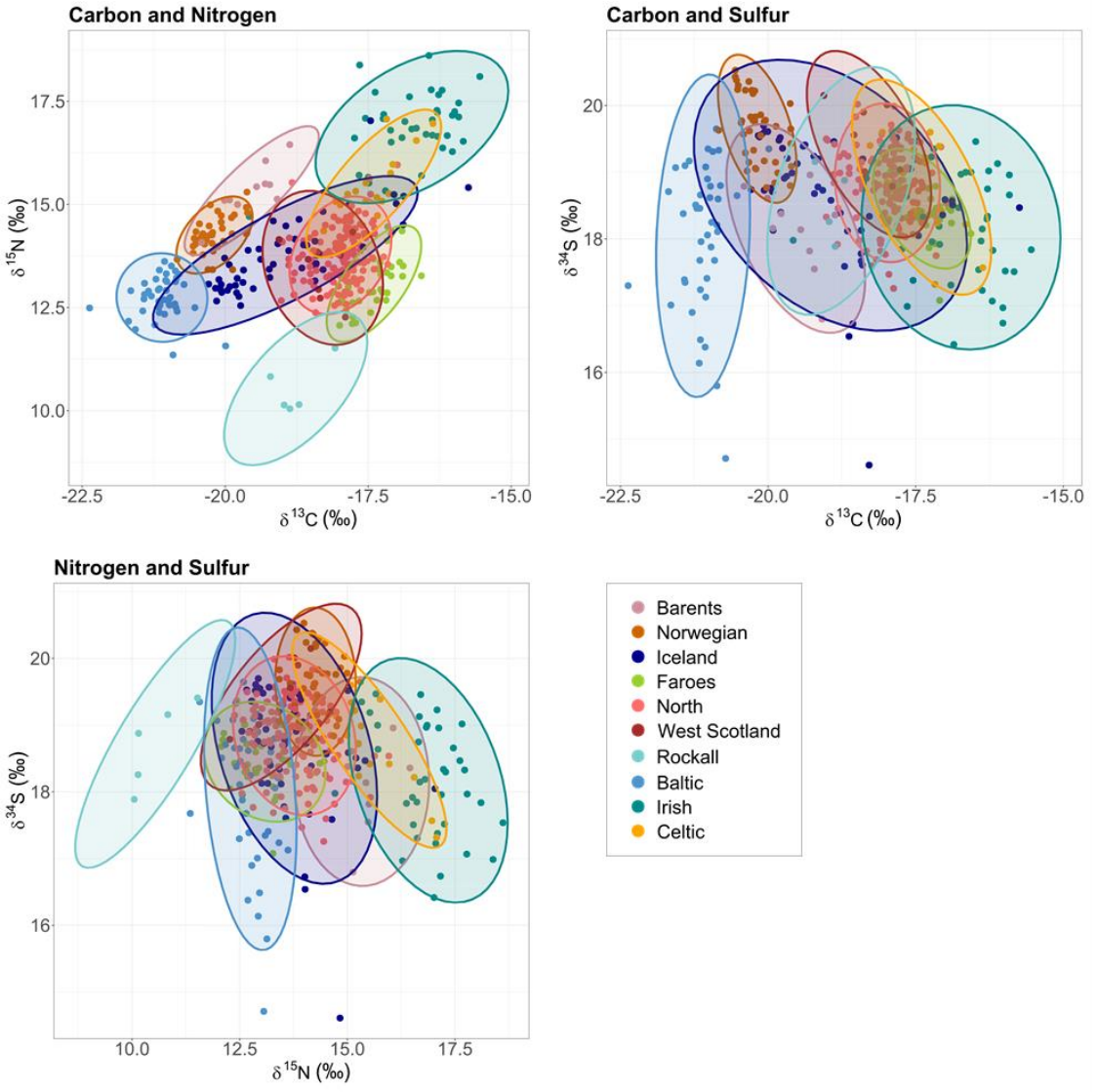


Figure 2.4 Carbon, nitrogen and sulfur stable isotope values for each individual cod sampled, coloured by region of origin. The 90% data ellipses are also shown for each geographic region.

2.4.2 Assignments to catch location

The average accuracy of assignment to known origin among all sampled regions was 70.8% using the multivariate normal probability technique with all three stable isotopes (carbon, nitrogen and sulfur), and the success rates for each location are shown in Table 2.4. Assignment accuracy varied with the repeated simulations depending on which samples were selected for the training and test datasets during the assignment process (Figure 2.5). Cod from the Norwegian and Baltic Seas were the most accurately assigned, with an average success rate of 95% and 94% respectively. The lowest assignment success was obtained for West Scotland and Rockall (44% and 40%), possibly due to the low sample numbers collected. The North Sea and Celtic Sea also had relatively low accuracies of assignment (52% and 62%), due to a number of cod samples being assigned to nearby regions of the UK shelf sea or Faroes. Most often the incorrect assignments were to the nearest region geographically, such as North Sea samples being assigned to the Faroes or West Scotland, or Irish Sea samples being assigned to the Celtic Sea, since the region in closest proximity is often the most isotopically similar.

The results of multivariate assignment using only carbon and nitrogen isotopes, and not including sulfur isotopes, are shown in Figure 2.6. This resulted in an average assignment accuracy of 70.7% among all regions. This is almost exactly the same as the accuracy using all three isotopes, indicating that $\delta^{34}S$ contributes little to the spatial traceability of cod in the regions studied. For most locations the correct assignments were very similar when using $\delta^{13}C$ and $\delta^{15}N$ compared with using all three isotopes (Table 2.4), or even lower assignment success with sulfur included in some cases, such as the Celtic Sea and Rockall where more samples were assigned to West Scotland and the Barents Sea respectively (Figure 4). However, sulfur had more value for Icelandic cod, for which the assignment accuracy increased from 59% to 73% when $\delta^{34}S$ was used in addition to $\delta^{13}S$ and $\delta^{15}N$, as well as for the Norwegian Sea (13% increase with $\delta^{34}S$). The ability to distinguish between Iceland and West Scotland samples was particularly improved, since over 50% of Icelandic cod were assigned to West Scotland in some simulations using only carbon and nitrogen (Figure 2.6) compared with a mean of 1.7% over all repeat simulations with the addition of sulfur (Figure 2.5). In the same way, fewer West Scotland cod were assigned to Iceland when sulfur was included. Therefore, in limited cases sulfur isotopes gave an improvement in assignment accuracy, but for most regions sulfur did not provide significantly greater distinguishing power than using carbon and nitrogen isotopes alone.

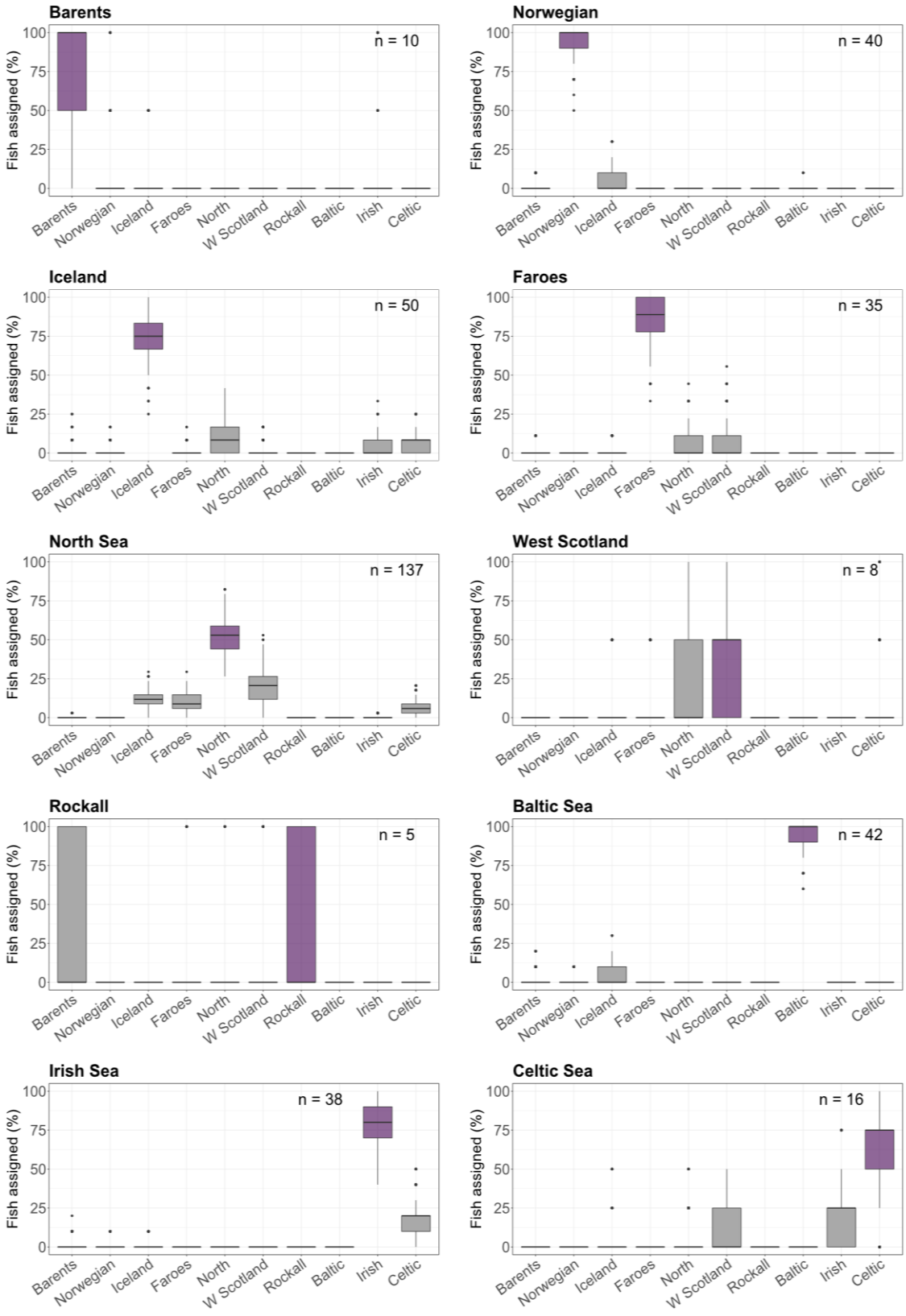


Figure 2.5 Assignment results using carbon, nitrogen and sulfur stable isotope data, showing the percentage of individuals from each known location assigned to all the possible regions over 1000 repeat simulations. The coloured boxes show the correct regions of origin.

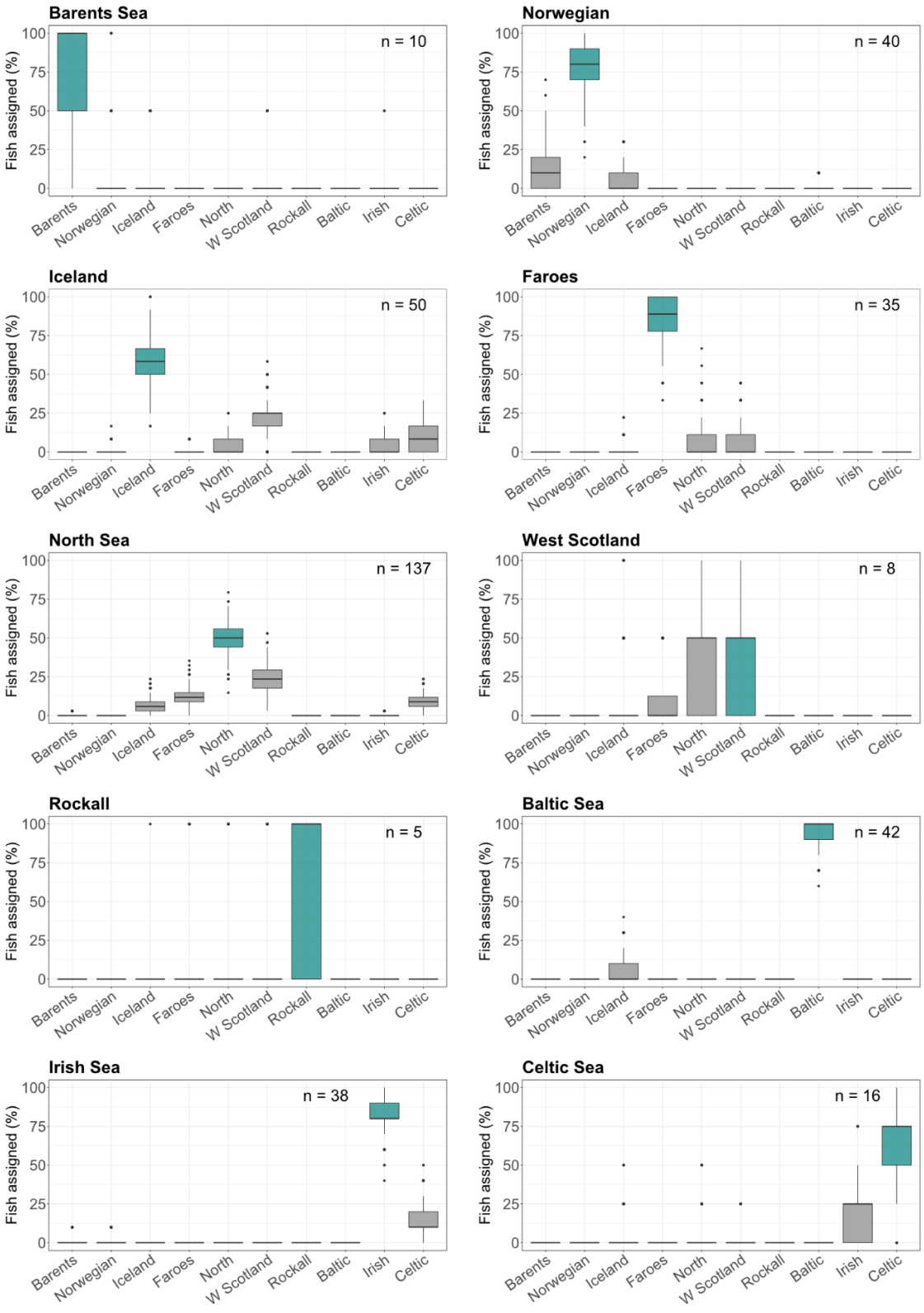


Figure 2.6 Assignment results using only carbon and nitrogen stable isotope data, showing the percentage of individuals from each known location assigned to all the possible regions over 1000 repeat simulations. The coloured boxes show the correct regions of origin.

Table 2.4 Mean percentage of individuals assigned to the correct origin region over 1000 repeat simulations using three isotopes ($\delta^{13}C$, $\delta^{15}N$ and $\delta^{34}S$) and two isotopes ($\delta^{13}C$ and $\delta^{15}N$).

Region	Mean correct assignments (%)	
	CNS	CN
Barents	83.0	80.0
Norwegian	95.1	82.0
Iceland	72.7	59.0
Faroes	85.1	86.8
North Sea	51.7	49.8
West Scotland	43.9	40.3
Rockall	40.4	61.0
Baltic	94.1	94.0
Irish	79.7	83.8
Celtic	61.9	70.4
Total	**70.8%**	**70.7%**

Leave-one-out cross validation using the multivariate normal probability method gave similar results to the stratified jack-knifing multivariate technique using all three stable isotopes, with an overall success rate of 72.2% (Table 2.5), since the two methods are very similar. However, cross validation performed better than the original multivariate technique for some regions and slightly worse for others. For the Barents Sea and West Scotland, the assignment accuracy increased to 90% and 50% respectively using cross validation, whereas the Iceland and North Sea assignment accuracies decreased to 70% and 47%. However, the results showed a maximum difference of 7% between the two methods, which gives confidence that the inferences are robust to the multivariate technique used, whether stratified jack-knifing or leave-one-out cross validation.

Table 2.5 Leave-one-out cross validation results using the multivariate normal probability method, showing the number of samples assigned to each of the geographic regions as well as the percentage of correct assignments for each region. True known origins are shown in the columns and the assigned most likely origins are shown in the rows.

Assigned origin	**True origin region – number assigned**									
	Barents	Norwegian	Iceland	Faroes	North Sea	West Scotland	Rockall	Baltic	Irish	Celtic
Barents	**9**	0	1	0	0	0	0	0	1	0
Norwegian	1	**39**	1	0	0	0	0	0	0	0
Iceland	0	1	**35**	1	15	1	0	1	0	1
Faroes	0	0	1	**30**	14	1	3	0	0	0
North Sea	0	0	4	2	**64**	2	0	0	0	0
West Scotland	0	0	2	2	33	**4**	0	0	0	1
Rockall	0	0	0	0	0	0	**2**	0	0	0
Baltic	0	0	0	0	0	0	0	**41**	0	0
Irish	0	0	2	0	1	0	0	0	**31**	4
Celtic	0	0	4	0	10	0	0	0	6	**10**
Percentage correct	**90%**	**98%**	**70%**	**86%**	**47%**	**50%**	**40%**	**98%**	**82%**	**63%**

The accuracy of assignment for each region of origin using the three techniques – multivariate normal probability distributions, linear discriminant analysis and random forest classification – are compared in Table 2.6. The assignments to all regions using linear discriminant analysis and random forest classification are shown in more detail in Tables 2.7 and 2.8 respectively. Random forest classification gave the highest overall accuracy of assignment to origin (mean 79% for all regions), whereas discriminant analysis and the multivariate normal probability technique both gave similar, slightly lower accuracies – 70% and 72% mean overall respectively. Linear discriminant analysis performed poorly for West Scotland (0% correct) and the Celtic Sea (42% correct), but gave the highest assignment accuracy of all three techniques for the Norwegian and North Seas (99% and 88% correct). The multivariate technique was the most successful for the Iceland assignments, achieving at least 6% greater accuracy than the other two methods, whereas it gave a low assignment accuracy for the Rockall samples (40% correct compared to 100% using random forest). Random forest classification was the most successful method for the Celtic Sea, giving either 13% or 33% greater accuracy than the other techniques for these regions, whereas for the Barents Sea, Faroes and West Scotland samples, the multivariate technique and random forest performed equally well (90%, 86% and 50% respectively). Therefore, each assignment technique performed better for some regions and worse for others, although random forest classification achieved the highest success rate overall as well as for the most individual regions. Figure 2.8 shows the assignment accuracy of cod to the ICES subareas within FAO region 27 using random forest classification, to demonstrate the potential to distinguish among these regions in a fisheries management context.

Linear discriminant analysis resulted in clear clustering of the stable isotope data by catch location, with relatively little overlap for many regions (Figure 2.7). Certain regions do show overlap with other clusters, in general with the regions that are closest geographically, such as the Barents and Norwegian Seas or West Scotland and the North Sea.

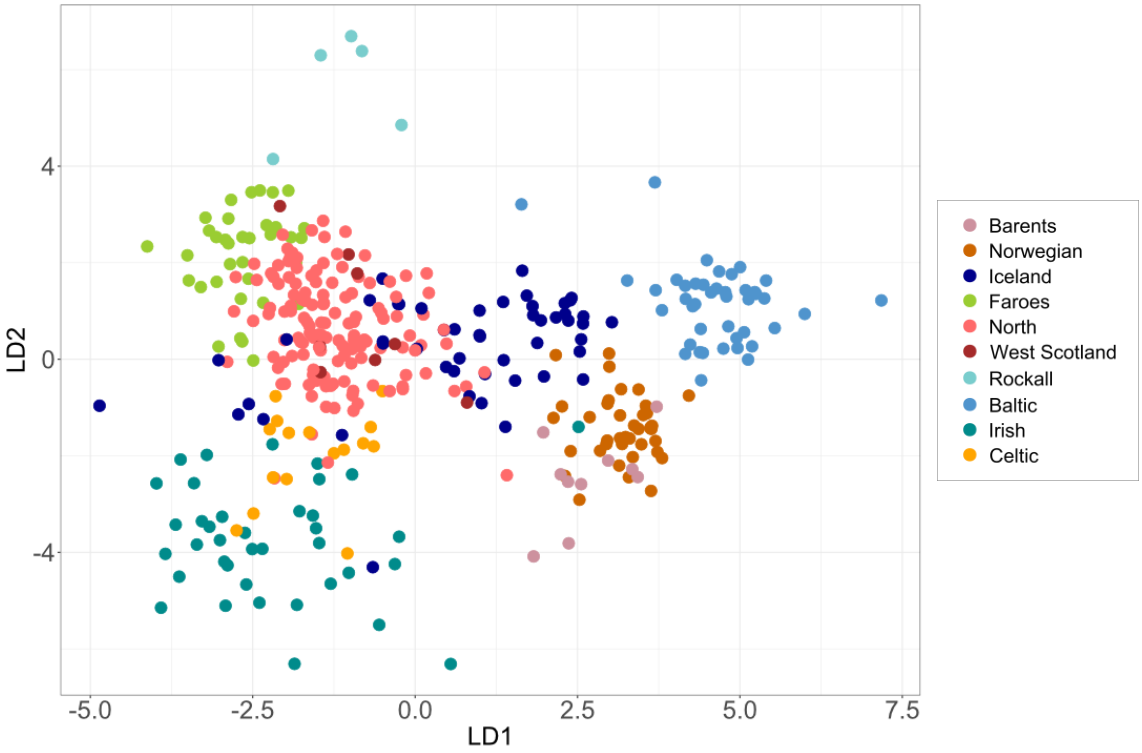


Figure 2.7 Linear discriminant analysis (LD1 and LD2) using the carbon, nitrogen and sulfur stable isotope compositions measured in cod muscle tissue from each of the sampled regions.

Table 2.6 Comparison of assignment accuracy for each geographic region using linear discriminant analysis, multivariate normal probability distributions and random forest classification. For linear discriminant analysis, the mean over 1000 repeat simulations is shown. For multivariate analysis and random forest, results are shown using a leave-one-out cross validation approach.

Location	Assignment accuracy (%)		
	Linear discriminant analysis	Multivariate analysis	Random forest
Barents Sea	78.5	90.0	90.0
Norwegian Sea	99.1	97.5	90.0
Iceland	60.9	70.0	64.0
Faroes	74.3	85.7	85.7
North Sea	87.9	46.7	51.8
West Scotland	0.0	50.0	50.0
Rockall	78.7	40.0	100.0
Baltic Sea	92.6	97.6	95.2
Irish Sea	84.3	81.6	84.2
Celtic Sea	41.8	62.5	75.0
Mean accuracy	**69.8%**	**72.2%**	**78.6%**

Table 2.7 Assignment results to each of the geographic regions using linear discriminant analysis, showing the mean percentage over 1000 repeat simulations. The correct assignments for each region are shown in bold.

Assigned origin	**True origin region – percentage assigned (%)**									
	Barents	Norwegian	Iceland	Faroes	North Sea	West Scotland	Rockall	Baltic	Irish	Celtic
Barents	**78.5**	0.0	0.2	0.0	0.8	0.0	0.0	0.0	2.6	0.0
Norwegian	21.5	**99.1**	6.3	0.0	0.0	0.3	0.0	5.0	0.0	0.0
Iceland	0.1	0.9	**60.9**	0.0	4.7	12.5	0.0	2.4	0.0	0.0
Faroes	0.0	0.0	0.1	**74.3**	4.4	13.7	21.3	0.0	0.0	0.0
North Sea	0.0	0.0	25.0	25.7	**87.9**	73.6	0.0	0.0	0.0	33.1
West Scotland	0.0	0.0	0.0	0.0	0.0	**0.0**	0.0	0.0	0.0	0.0
Rockall	0.0	0.0	0.0	0.0	0.0	0.0	**78.7**	0.0	0.0	0.0
Baltic	0.0	0.0	2.2	0.0	0.0	0.0	0.0	**92.6**	0.0	0.0
Irish	0.0	0.0	2.3	0.0	0.6	0.0	0.0	0.0	**84.3**	25.2
Celtic	0.0	0.0	2.9	0.0	1.5	0.0	0.0	0.0	13.1	**41.8**

Table 2.8 Leave-one-out cross validation results using random forest classification, showing the number of samples assigned to each of the geographic regions as well as the percentage of correct assignments for each region shown in bold.

Assigned origin	**True origin region – number assigned**									
	Barents	Norwegian	Iceland	Faroes	North Sea	West Scotland	Rockall	Baltic	Irish	Celtic
Barents	**9**	1	0	0	1	0	0	0	1	0
Norwegian	1	**36**	4	0	0	1	0	0	0	0
Iceland	0	2	**32**	0	10	0	0	1	0	0
Faroes	0	0	1	**30**	22	1	0	0	0	0
North Sea	0	0	3	2	**71**	2	0	0	0	0
West Scotland	0	0	3	3	25	**4**	0	0	0	1
Rockall	0	0	0	0	0	0	**5**	1	0	0
Baltic	0	1	0	0	0	0	0	**40**	0	0
Irish	0	0	1	0	0	0	0	0	**32**	3
Celtic	0	0	6	0	8	0	0	0	5	**12**
Percentage correct	**90%**	**90%**	**64%**	**86%**	**52%**	**50%**	**100%**	**95%**	**84%**	**75%**

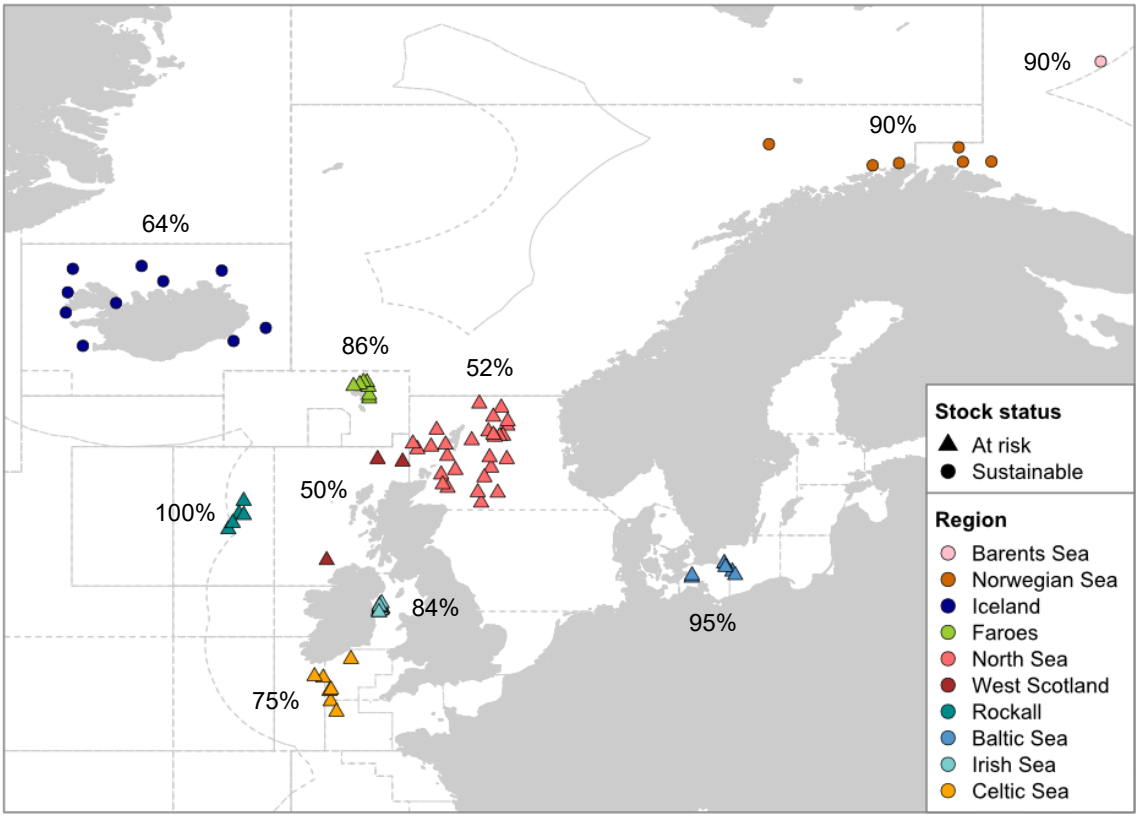


Figure 2.8 Locations of sampling stations within ICES sub area boundaries (grey lines), and assignment success rates in each region using random forest classification with a leave-one-out cross validation approach.

2.4.3 Assigning independent known origin samples to the reference dataset

Data from independent known origin samples from five locations (Barents Sea, Iceland, North Sea, Irish Sea and Celtic Sea), collected by Jennings and Cogan (2015), Institute of Marine Research (IMR, Norway), Young's Seafood Ltd. and Ifremer (Kopp, 2018), were corrected for the Suess effect and assigned to their most likely origin based on their carbon and nitrogen isotope composition, using the known origin samples analysed in this study as a reference dataset. A comparison between the $\delta^{13}C$ and $\delta^{15}N$ values of the samples from this study (2018) and these previous datasets is illustrated in Figure 2.9, and the results of the assignments are shown in Table 2.9. The isotopic composition of samples from all regions show good agreement between years following the Suess correction, and the data ellipses generally match well. The only region to show slight mismatch is the Barents Sea, since the Young's Seafood samples (reference data) had slightly higher $\delta^{13}C$ and $\delta^{15}N$ values than those collected by the Institute of Marine Research (independent samples), meaning that the stable isotope compositions of these independent samples are more similar to those of the Norwegian Sea cod in the reference dataset (shown in Figure 2.9).

The assignment successes were reduced for all regions compared to the estimates calculated using only the samples from the current study and assigning a subset of the same dataset to itself (non-independent samples). Since the stable isotope compositions of the independent Barents Sea samples were more similar to those of cod from the northern Norwegian Sea, only two of the independent samples were assigned to the Barents Sea, indicating that these samples more likely originated from north Norwegian coastal waters rather than the central Barents Sea (exact sampling locations are not known). In this case, the assignment accuracy would be 73% instead of 2% as it appears from Table 2.9, although this accuracy is still lower than expected based on the self-assignment of reference samples (90% success rate). Assignment accuracy for the other regions using independent samples ranged from 12% for the North Sea to 50% for Iceland, whereas using the reference samples the success rate was 52% for the North Sea and 64% for Iceland. Therefore, assignment accuracies for individual regions decreased by between 14% (Iceland) and 58% (Irish Sea) using independent samples. Incorrect assignments were most often to regions in close geographic proximity to the correct region, for example the North Sea samples were assigned to West Scotland or the Faroes, and Irish Sea samples were assigned to the Celtic Sea.

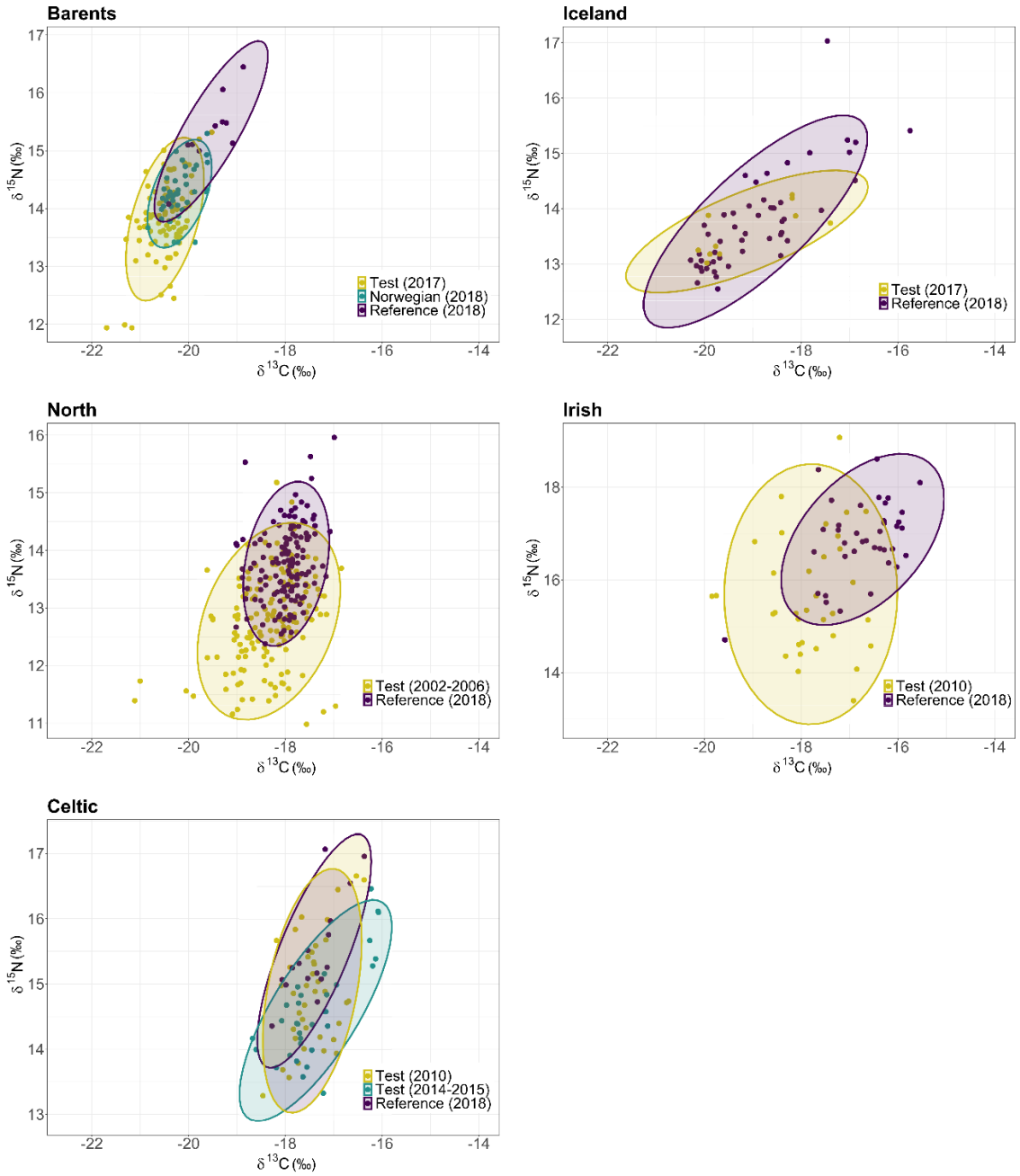


Figure 2.9 Carbon and nitrogen stable isotope values measured in cod samples collected in this study (reference samples) compared with those collected from previous years and studies (test samples), after applying Suess correction on $\delta^{13}C$ values. Previous data: Barents 2017 – collected by Institute of Marine Research (Norway); Iceland 2017 – provided by Young's Seafood Ltd.; North Sea 2002-2006 – from Jennings and Cogan (2015); Irish Sea 2010 – from Jennings and Cogan (2015); Celtic Sea 2010 – from Jennings and Cogan (2015), Celtic Sea 2014-2015 – collected by Ifremer (Kopp, 2018) from the EVHOE 2014 survey.

Table 2.9 Results of assigning cod samples measured in previous datasets to the sample data collected in our study as a reference dataset (with a correction for the Suess effect on $\delta^{13}C$ values), using only $\delta^{13}C$ and $\delta^{15}N$. The correct assignments for each region are highlighted in blue. Previous datasets: Barents Sea - collected by Institute of Marine Research (Norway); Iceland - provided by Young's Seafood Ltd.; North Sea – from Jennings and Cogan (2015); Irish Sea - from Jennings and Cogan (2015); Celtic Sea – from Jennings and Cogan (2015) and collected by Ifremer (Kopp, 2018) from the EVHOE 2014 survey.

True origin region	Assigned region	Cod assigned (count)	Cod assigned (%)
Barents	**Barents Sea**	**2**	**2.3**
	Norwegian Sea	63	73.3
	Iceland	3	3.5
	Faroes	0	0.0
	North Sea	0	0.0
	West Scotland	0	0.0
	Rockall	0	0.0
	Baltic Sea	18	20.9
	Irish Sea	0	0.0
	Celtic Sea	0	0.0
Iceland	Barents Sea	0	0.0
	Norwegian Sea	1	10.0
	Iceland	**5**	**50.0**
	Faroes	1	10.0
	North Sea	3	30.0
	West Scotland	0	0.0
	Rockall	0	0.0
	Baltic Sea	0	0.0
	Irish Sea	0	0.0
	Celtic Sea	0	0.0
North Sea	Barents Sea	0	0.0
	Norwegian Sea	0	0.0
	Iceland	18	11.7
	Faroes	31	20.1
	North Sea	**19**	**12.3**
	West Scotland	64	41.6
	Rockall	18	11.7
	Baltic Sea	2	1.3
	Irish Sea	0	0.0
	Celtic Sea	2	1.3

Irish Sea	Barents Sea	5	14.7
	Norwegian Sea	0	0.0
	Iceland	1	2.9
	Faroes	3	8.8
	North Sea	3	8.8
	West Scotland	3	8.8
	Rockall	0	0.0
	Baltic Sea	0	0.0
	Irish Sea	**9**	**26.5**
	Celtic Sea	10	29.4
Celtic Sea	Barents Sea	0	0.0
	Norwegian Sea	0	0.0
	Iceland	1	1.4
	Faroes	8	11.3
	North Sea	25	35.2
	West Scotland	3	4.2
	Rockall	0	0.0
	Baltic Sea	0	0.0
	Irish Sea	9	12.7
	Celtic Sea	**25**	**35.2**

2.4.4 Comparison to assignment using genetic techniques

Nielsen et al. (2012b) used single nucleotide polymorphism (SNPs) to assign Atlantic cod to three populations of origin in the north Atlantic. In the current study using stable isotope ratios, if the same cod populations as in Nielsen et al. (2012b) are selected from the dataset and all other sampled locations excluded (combining the Barents and Norwegian samples into one group – the Northeast Arctic), the assignment success rates are 99.6% for the Northeast Arctic, 99.2% for the North Sea and 94.4% for the Baltic Sea, so are very close to the high success rates achieved using genetic techniques for the same regions (98-100%). Using either the genetic or isotopic techniques, the Baltic Sea had the lowest assignment accuracy of the three regions. For the genetic assignments, the likelihood of the test sample originating from the true population of origin was at least six times higher than for the second most likely population of origin (Nielsen et al., 2012b). In our study using stable isotopes, the majority of assignments also had a probability at least six times greater for the true region of origin than the second most likely region – only between zero and eight samples out of a total of 82 test samples had a probability less than six times greater, varying with repeat simulations using different test subsets. Many samples had much higher likelihoods when assigned - the most likely origin location had a median

probability $6.0e^{54}$ times greater than the second most likely. If only 'strong' assignments are included, where the likelihood of the true location is at least six times greater than the second most likely origin, and others are discarded, the assignment accuracy increases to 99.9% for the Northeast Arctic, 99.3% for the North Sea and 95.7% for the Baltic Sea. Furthermore, using random forest classification, high assignment accuracies of 94%, 100% and 100% for the Northeast Arctic, North Sea and Baltic Sea were achieved respectively. Only three samples were misclassified in total, all from the Northeast Arctic. Therefore, it is demonstrated that isotopic techniques have the ability to assign cod to their region of origin with similarly high accuracies as those using genetic techniques.

2.4.5 Isotopic variation within discrete regions

Isotopic ratios of carbon, nitrogen and sulfur also vary within discrete regions, since there is a continuous variation across the oceans. Assessing and understanding the isotopic variance within regions has the potential to allow provenance tracing of fish on an even finer scale. In this study, we have collected samples from all around Iceland, except for the most southerly area, so the variation in the isotopic composition of cod muscle tissue within this region could be investigated. For a continuous approach a large number of samples are required, covering the entire area of interest. We collected enough samples from the northern North Sea to enable a continuous surface of isotopic variance (isoscape) to be created by interpolation between data points and therefore reveal the patterns of isotopic variation in the region.

2.4.5.1 Spatial variation around Iceland

Among the sampled regions, cod from Icelandic waters show the highest isotopic variance, with individuals grouping both with the Northeast Arctic/Baltic Sea and with the Hiberno-Britannic clusters (Figure 2.4).

Figure 2.10 shows the pattern of spatial variation in carbon, nitrogen and sulfur stable isotope values in cod from around Iceland. The average $\delta^{13}C$, $\delta^{15}N$ and $\delta^{34}S$ at each station differs by 2.1-3.3‰ between the highest and lowest station in all three isotopes. To the north and east of Iceland the carbon and nitrogen isotope values are generally lower, whereas the sulfur isotope values are higher.

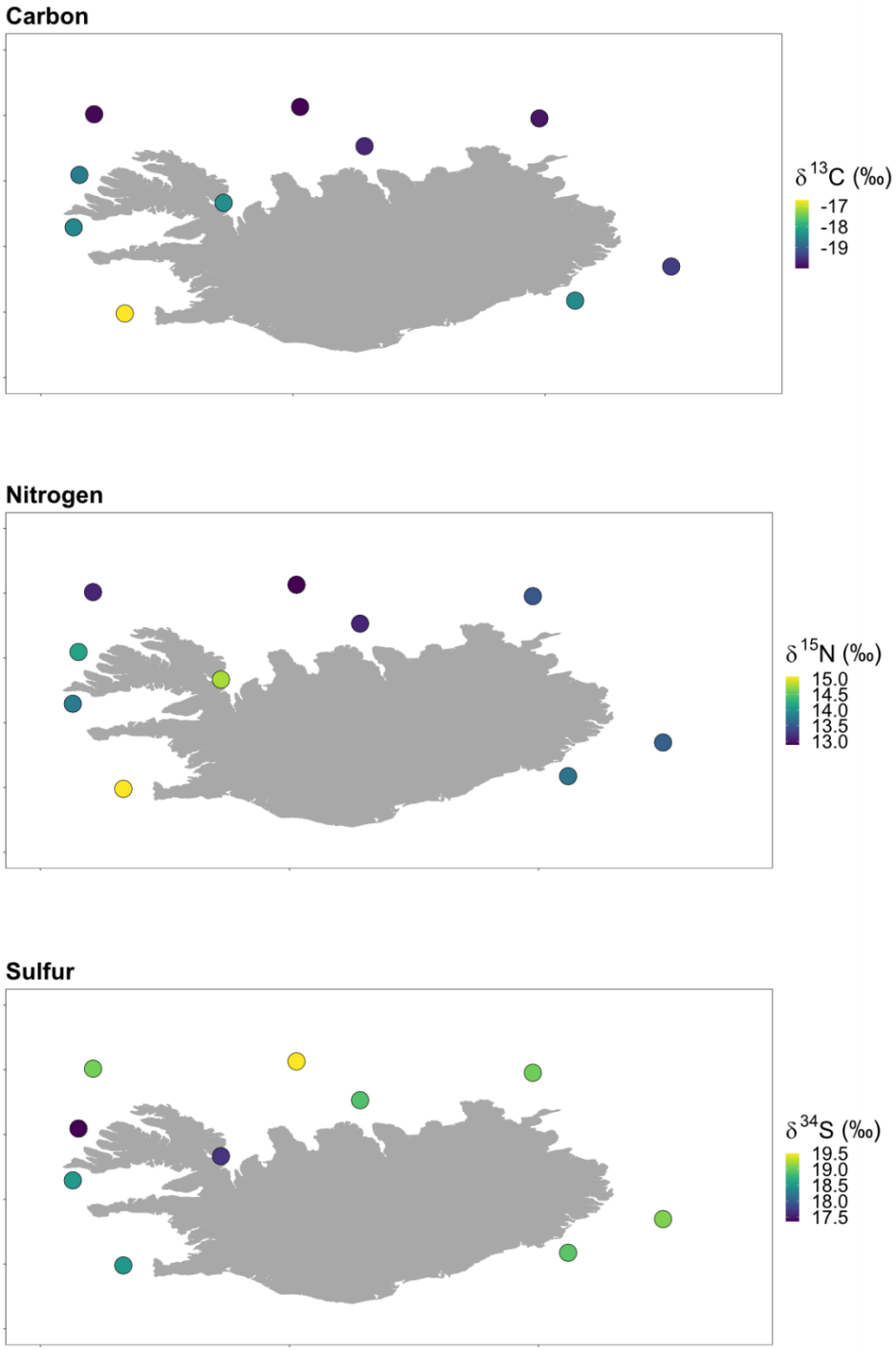


Figure 2.10 Spatial variation in $\delta^{13}C$, $\delta^{15}N$ and $\delta^{34}S$ values measured in cod from Icelandic waters, showing the mean value at each station.

2.4.5.2 North Sea isoscape

Cod samples were collected from 28 different stations within the North Sea, so the variation among stations could also be investigated. The mean station-level isotope values varied between -17.6‰ and -18.7‰ for $\delta^{13}C$, between 13.2‰ and 14.5‰ for $\delta^{15}N$ and between 18.3‰ and 19.5‰ for $\delta^{34}S$, meaning that there were variations of around 1.1‰, 1.3‰ and 1.2‰ respectively for $\delta^{13}C$, $\delta^{15}N$ and $\delta^{34}S$ across the sampled area. Carbon, nitrogen and sulfur isoscapes were created for this northern North Sea area using interpolation by kriging between the sampled stations (Figure 2.11). These reveal a pattern of lower $\delta^{13}C$ values in the northwest of the area and lower $\delta^{15}N$ values in the north, increasing southwards below the Shetland Islands. The highest $\delta^{34}S$ values are across the north and eastern parts of the isoscape, gradually decreasing to the southwest closer to mainland Scotland. The lowest $\delta^{34}S$ values are in the shallower water to the southwest of the Shetland Islands and east of mainland Scotland. The variances of all three stable isotopes are similar, although slightly lower for carbon. For all three isotopes, variances increase at the edges of the isoscapes, outside of the area with measured data points where the interpolation cannot be carried out reliably.

The sample variograms and associated variogram models for kriging interpolation are shown in Figure 2.12. A linear model fitted the variograms best for all three isotopes, although for sulfur an exponential model also fitted well. Without further reference data it could not be determined which model is more accurate, therefore in this case I used the linear model for consistency. The three variogram models all have very similar gradients, suggesting that carbon, nitrogen and sulfur isotope ratios in cod muscle vary over similar spatial scales within this region. However, the variance values are greater at all distances for nitrogen and sulfur than for carbon, meaning that at the same distance apart there are greater differences in cod $\delta^{15}N$ and $\delta^{34}S$ values than for $\delta^{13}C$. There is more scatter around the variogram model for nitrogen, indicating that the error in the model is greater and any spatial structure predicted in the isoscape is less reliable.

These cod isoscapes have the potential to assign Atlantic cod samples to origin at a higher resolution within the northern North Sea, if the samples are known to have originated from somewhere within the isoscape area.

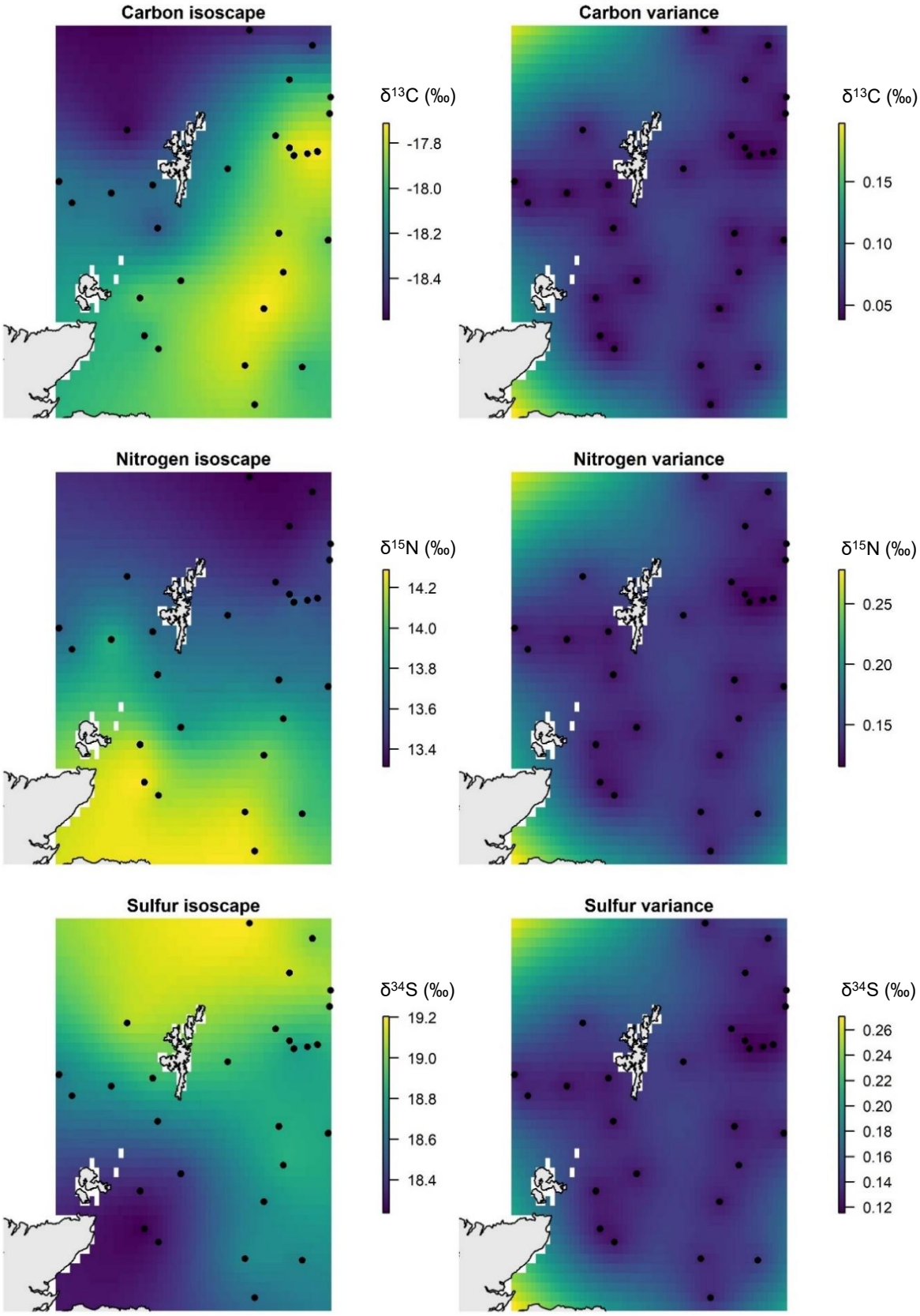


Figure 2.11 Carbon, nitrogen and sulfur isoscapes for the North Sea sampling area derived from cod muscle tissue (left), with the sampled stations indicated by black circles. The associated spatial variances for each isoscape are shown on the right.

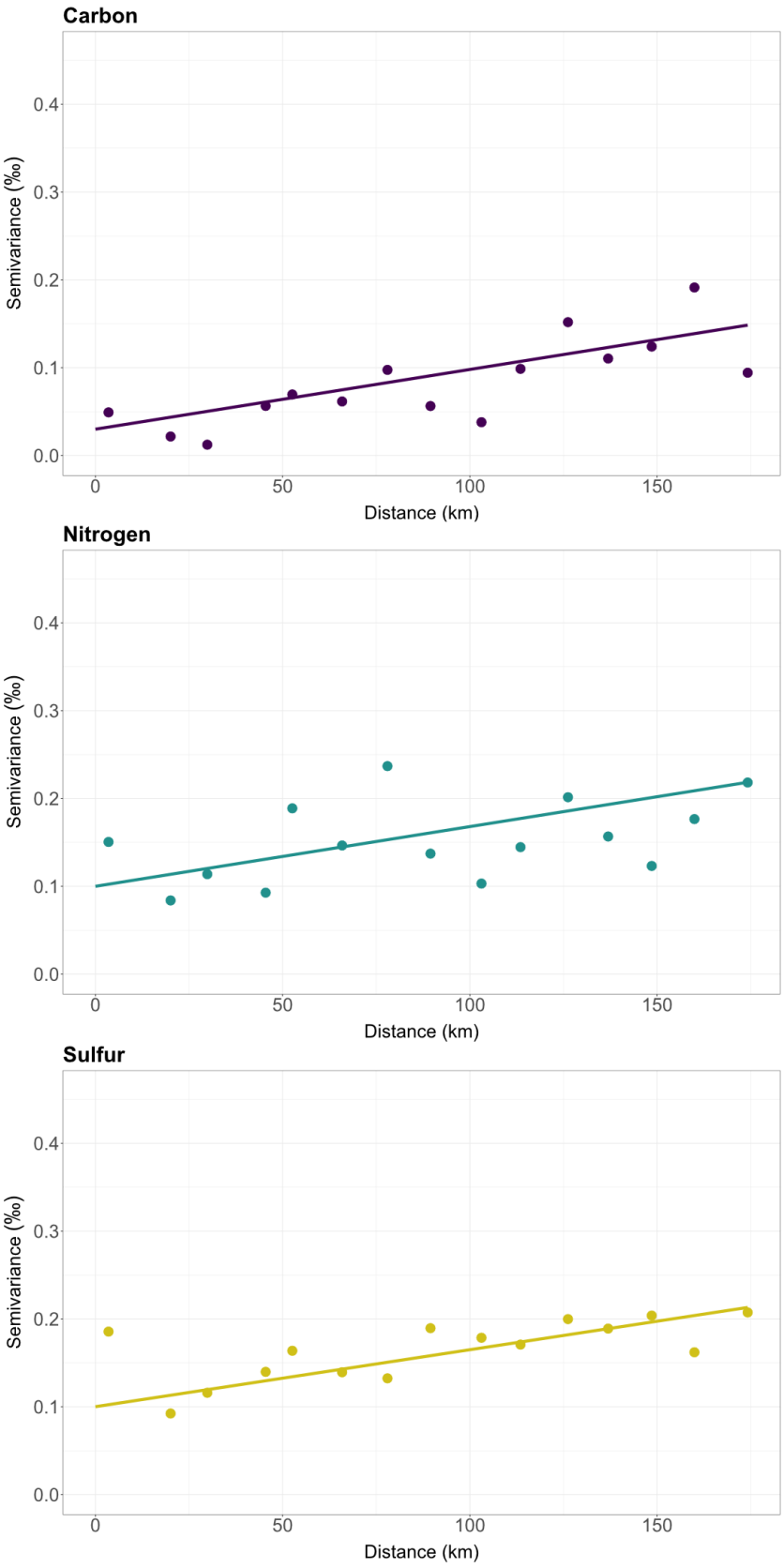


Figure 2.12 Variograms and the associated variogram models for carbon, nitrogen and sulfur stable isotope ratios using cod from the North Sea, illustrating the variance in isotope values with distance between the sampled stations.

2.5 Discussion

The results of this study demonstrate that Atlantic cod show varying stable isotope ratios within their muscle tissue depending on the spatial origin, and therefore demonstrate the potential use of stable isotopes as a forensic tool to indicate provenance of Atlantic cod products from the more isotopically distinct regions. However, cod from some regions were found to be isotopically similar and could not be reliably distinguished. This study contributes a large dataset of stable isotope ratios measured in Atlantic cod muscle tissue, covering a wide range of fishery areas, and is one of the only studies to have investigated the potential application of sulfur stable isotopes to the traceability of wild marine fish. The number of samples analysed allowed a comprehensive investigation into the accuracy with which cod could be traced to their region of origin, on the spatial scale of ICES subareas within FAO Major Fishing Area 27. The discrete regions cod were assigned to generally fall within the ICES subarea boundaries.

Cod from certain regions, such as the Barents Sea, Norwegian Sea, Baltic Sea and Rockall, could be traced to their catch area with high confidence (90-100%), whereas other regions were less isotopically distinct, such as Iceland, the North Sea and West Scotland, which resulted in lower assignment accuracies (64%, 52% and 50% respectively). The Baltic Sea samples were one of the most reliably discriminated, as is expected due to the different environmental conditions - the Baltic Sea is brackish because of freshwater runoff and has a shallow sill at the entrance to the Atlantic Ocean resulting in limited water exchange. Although the relatively low success for West Scotland may in part be due to the small sample number, the overlapping isotopic composition with the North Sea due to the close geographic proximity is also likely to have caused further difficulty in assigning the fish correctly. This is supported by the high proportion of West Scotland samples assigned incorrectly to the North Sea. Conversely, the North Sea also had a relatively low assignment success due to a number of misassignments to the closest neighbouring regions – the Faroes and West Scotland. Several Irish Sea and Celtic Sea cod were also incorrectly assigned to each other. The assignment success is increased if the Celtic and Irish Sea fish are combined into one group, which still has importance in terms of fisheries management since both regions are part of ICES subarea 27.7.

2.5.1 Comparison of assignment methods

Random forest classification performed the best overall for assignment accuracy to region of origin, achieving a significantly higher success rate than the other methods for several regions. However, it was necessary to balance the sample sizes for classification, because random forest can be biased by different sample sizes across groups, being more likely to assign to groups with a greater number of samples (Bader-El-Den et al., 2018), and

otherwise resulted in very low assignment success for locations with small sample numbers (e.g. West Scotland). The multivariate normal probability technique and linear discriminant analysis gave very similar assignment accuracies overall. Although these two techniques did not have such a high overall success rate as random forest classification, they had the advantage of performing better for certain discrete regions. In particular, linear discriminant analysis gave a significantly higher assignment accuracy for the North Sea than the other methods (88% compared with 52%). However, for other regions linear discriminant analysis resulted in much lower accuracies, for example it failed to assign any West Scotland samples correctly. The accuracies using the multivariate technique were more stable across regions, achieving moderate accuracies for all, although this method gave the highest success rate for Iceland. Therefore, random forest classification appears to be the best overall choice of assignment method for this dataset, and may improve further if more balanced sample sizes were obtained.

Cross validation gave very similar results to the multivariate technique, since the same multivariate assignment method was used but only removing one sample at a time rather than randomly selecting subsets of 25% from each region as test samples. In some cases this led to a higher assignment accuracy because fewer samples are removed at a time, leaving a larger reference subset to assign the test sample against. This is particularly beneficial for those with a small sample size, as reflected in the increased assignment accuracy for West Scotland (n = 8) using cross validation.

Sulfur stable isotope ratios have rarely been used in the marine environment for tracing the origin of animals, but it has been found to increase the ability to distinguish among catch locations of scampi (*Nephrops norvegicus*) (Slesser and Trueman, 2021). Anoxic conditions in seabed sediments result in anaerobic microbial respiration of sulfur, producing hydrogen sulfide with a relatively low $\delta^{34}S$ value which is subsequently incorporated into production (Yamanaka et al., 2003). Therefore, in benthic burrowing animals such as scampi, the tissues reflect spatial differences in $\delta^{34}S$ related to sediment oxidation conditions. St. John Glew et al. (2019) created a sulfur isoscape for the UK shelf seas using jellyfish tissues, which shows spatial variation in $\delta^{34}S$ also exists in pelagic ecosystems. Despite this, in the current study the inclusion of sulfur isotope ratios did not increase the proportion of cod correctly traced to origin for the majority of fishery regions. In fact, the addition of sulfur isotope measurements reduced the assignment accuracy in many cases. Of all the regions analysed, only Icelandic cod benefitted significantly from including sulfur isotopes, with an increase in assignment accuracy of 14%. A few other regions, such as the North Sea, Barents Sea and West Scotland were traced slightly more reliably using all three isotopes, but only small increases of 1.3-4% in the assignment success were observed. This is due to the large and overlapping ranges of $\delta^{34}S$ values in cod from almost all areas. Although Atlantic cod is a benthopelagic species and a

proportion of their diet is derived from benthic invertebrates (Mello and Rose, 2005), their similar $\delta^{34}S$ muscle tissue values across regions suggests that their sulfur isotopic composition is not significantly influenced by the benthic ecosystem or that $\delta^{34}S$ values of benthic prey are similar regionally, so sulfur isotopes may be of more importance in the spatial traceability of exclusively benthic or burrowing organisms. Sulfur stable isotopes may be useful for traceability of cod to certain regions, such as Iceland, but in the majority of cases it had relatively low informational value and could not be used to distinguish between individuals from different regions.

2.5.2 Assignment of independent known origin samples

Assigning other known origin cod from previous datasets to most likely origin using data from the current study as a reference dataset resulted in reduced assignment success in every region. This lower assignment accuracy compared to assigning non-independent samples is to be expected since the datasets were collected in different years (2002-2017 for the test data compared to 2018 for the reference data), were analysed in different laboratories, and the exact sampling locations within each region differed, which introduce additional variation. The Suess effect also leads to decreasing $\delta^{13}C$ values over time as anthropogenic CO_2 levels in the atmosphere increase (Young et al., 2013), although this was accounted for by applying a correction so that the datasets could be more reliably compared.

Temporal variation in the spatial distribution of $\delta^{13}C$ and $\delta^{15}N$ over seasonal and yearly time scales results from changes in the rate of hydrodynamic and biogeochemical processes with time (Magozzi et al., 2017), which is then transferred up the food web. The independent samples from the Barents Sea and Icelandic waters were only collected one year prior to those in the current study so the cod from these regions show similar isotopic compositions in both datasets (assuming the Barents Sea samples were actually collected in the north Norwegian Sea as discussed previously), whereas the Jennings and Cogan (2015) samples were collected 8-16 years earlier than in this study, so for these the temporal effect on assignment accuracy was more significant. In some areas, such as the North Sea, stable isotope ratios are known to remain relatively stable over decadal to centennial time periods (MacKenzie et al., 2014), whereas other sea areas with more fluctuating environmental conditions may show greater temporal variations. Therefore, certain areas may be more susceptible to temporal factors affecting the success in tracing origin using stable isotope ratios. Matching the timescale of reference sample collection to that of the unknown samples is likely to result in more reliable assignments.

The difference between the datasets could also partly be attributed to the differences in the exact locations of sampling within the same region. The samples in Jennings and Cogan (2015) from the North and Celtic seas were collected from similar locations to those here.

However, those from the Irish Sea cover a much wider area than in the current study, which may have led to the greater variation between individuals in the previous dataset (Figure 2.13). Creating a reference dataset using samples collected from across the whole of each target region better encompasses the true variance for the region and would likely result in a higher assignment accuracy.

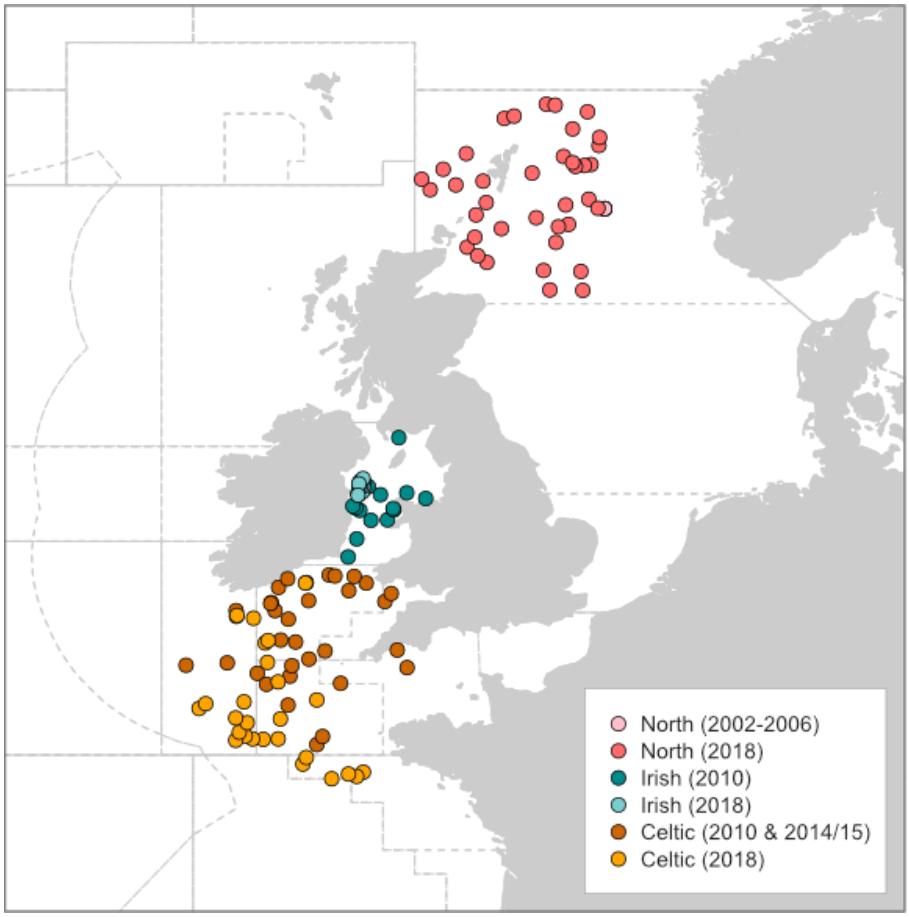


Figure 2.13 Locations where independent cod samples were collected previously by Jennings and Cogan (2015) compared to the locations where samples were collected for the current study in 2018 in the corresponding regions.

This lower assignment accuracy and precision when assigning previous datasets to the current data is important in demonstrating the true effectiveness of the methodology in a traceability context, because it tests samples that are truly independent from the reference data. When assigning a subset of the same dataset to itself, a higher success rate may be obtained than if used in real world scenarios in part due to the statistical artefact of having non-independent reference and test samples. Therefore, using independent assignment samples from other datasets is important when estimating accuracy.

2.5.3 Comparison of stable isotope analysis with genetic techniques

The use of DNA has also been investigated to trace the geographic origin of cod. Using advanced genetic techniques, such as next-generation sequencing, robust tools were developed to discriminate between cod from the Northeast Arctic, North Sea and Baltic Sea populations (Nielsen et al., 2012b). Between 98% and 100% of individuals were correctly assigned to area of origin using non-independent samples, with only one cod being incorrectly classified. This relied on identifying a number of SNPs (single-nucleotide polymorphisms) that varied in populations from different geographic origins. The use of genetics relies on populations being geographically and reproductively isolated over multigenerational timescales, so cod from certain regions may be clearly distinguished while the genetic differences between other regions may not be sufficient to allow discrimination. Nielsen et al. (2012b) did not discriminate between regions in such close proximity as in this study, such as the North Sea and Faroes or Irish and Celtic Seas, but these may have been more challenging to distinguish using genetics. With more potential regions that individuals can be assigned to, the proportion of correct assignments decreases, so it is to be expected that in our study the assignment accuracy did not reach that of Nielsen et al. (2012b). However, when assigning cod from the same three regions, both techniques gave comparable assignment success rates.

It is likely that combining the two approaches would reduce errors and give a higher confidence in tracing provenance than either technique used in isolation, as was found by Cusa et al. (2021), since genetics may give a more accurate result for some regions whereas for others stable isotopes may be more reliable.

2.5.4 Isotopic variation among and within discrete regions

Spatial variation in stable isotope ratios across the oceans can be predicted mechanistically based on oceanographic variables. The isotopic composition of phytoplankton at the base of the food chain varies spatially due to differences in the rate of photosynthesis and the nutrient source, among other factors (Magozzi et al., 2017, Somes et al., 2010). These spatial variations in isotope composition are then transferred up the food web and are assimilated into the tissues of the fish. Isotope-enabled global biogeochemical models allow a priori estimates of which fishery areas are likely to be isotopically distinct, and therefore where isotope markers may have predictive value (Cusa et al., 2021).

In addition to the spatial variation among discrete sea areas, variation in the isotopic composition of cod was also observed among stations within a single region. Around Iceland, cod collected from the north and east showed lower $\delta^{13}C$ and $\delta^{15}N$ values and higher $\delta^{34}S$ values compared to cod from the west of Iceland. It is possible that this is due to the presence of different currents flowing past and around Iceland. To the north and

east of Iceland, the cold waters of the East Greenland and East Icelandic currents flow southwards from the Arctic, but to the south and west the North Atlantic current and Irminger current bring warmer waters (Oskarsson et al., 2009). The different temperatures and other environmental variables associated with these water masses could have led to the variation in cod muscle isotope values from different regions around Iceland. No cod were analysed from the south of Iceland, but it could be expected that these would be similar to cod from western Iceland since they are exposed to the same water mass.

Variation among stations was also found in the North Sea, where there appears to be a transition to higher $\delta^{13}C$ and $\delta^{15}N$ values to the south of the Shetland Islands. The North Atlantic current flows eastwards from the Atlantic, following the edge of the continental shelf above the Shetlands, whereas within the North Sea is an anti-clockwise circulation around the basin (OSPAR Commission, 2000). Therefore, the differing isotope values may correspond to the water masses present in the region, and suggests that cod are not moving across this boundary during the time taken for the incorporation of stable isotopes into the muscle tissues. Sulfur isotope values also showed variation across the sampled North Sea area. Close to mainland Scotland in the south of the region, sulfur isotope ratios were lower than expected for fully marine water, indicating that lower $\delta^{34}S$ values from anoxic sediments may be incorporated into the cod through their diet in these areas.

Considering these differences in isotope signatures both around Iceland and in the North Sea, it may be possible to split these regions into two parts and assign individuals to a more precise area. Or if enough stations were sampled, test samples can be assigned more precisely using an isoscape, as in Trueman et al. (2017) where scallops could be traced to an area making up >40% of the North Sea with >90% accuracy using an isoscape created from jellyfish tissues. In most fisheries management or traceability contexts this may not be of additional value, since Iceland waters are all within ICES subarea 27.5 and the whole northern North Sea is contained in subarea 27.4.a, however it could be of importance for ensuring regulations are being adhered to within smaller geographic areas, such as in Marine Protected Areas.

2.5.5 Limitations and future research

There are limitations in the use of stable isotope ratios for provenance traceability related to the movement of individuals over time. If an individual migrates across multiple regions, the isotopic signature will be integrated across these boundaries and the results will be difficult to interpret. This is particularly a consideration for animals where the tissue turnover rate (time taken to assimilate the isotope values of the diet into tissue) is slow relative to the migratory movements of the animal. In this case, assignment to origin will be difficult because the individual moves across isotopic gradients faster than new tissue is formed. Likewise, the location an animal feeds may not necessarily be the location where it

is caught for very mobile animals, or animals could be feeding on mobile prey that have themselves crossed isotope gradients. Cod in the Northeast Atlantic are known to show site fidelity and have limited home ranges (Neat et al., 2014), but other more migratory species would be less successfully assigned to origin using stable isotopes.

Marine food webs are commonly structured by size, and since heavier isotopes increase with trophic levels, $\delta^{15}N$ and $\delta^{13}C$ values typically increase systematically with body size (Jennings et al., 2001, Jennings et al., 2008). In order to infer location from the isotopic compositions of fish of different species or body sizes a correction is necessary to account for trophic level effects, but this can be difficult to determine and introduces additional uncertainty (Jennings and Van Der Molen, 2015). In this study, body size was not recorded (although all fish were of commercial size) since the aim was to develop a tool that could be used in trade or retail situations, where the size before processing is unlikely to be known. Fish of a range of sizes were sampled from each region, so that the isotopic variation due to body size was taken into account. However, it is possible that errors could occur if an individual outside the size range of the reference dataset is assigned to origin.

In this study, discrete assignments were used to determine the most likely origin. Discrete assignments require a reference dataset of tissue isotope values from each of the potential origin areas. Therefore, discrete assignments assume that all possible origin regions have been sampled and are part of the reference dataset. Any test individuals from an origin not contained in the reference dataset will be incorrectly assigned. The accuracy of assignment also relies on the reference dataset sufficiently encompassing the true variance of the individuals from that region. To further test the use of stable isotopes for tracing the provenance of Atlantic cod, the regions sampled in this study could be expanded to cover all possible origin regions, and for some regions already sampled to increase the sample numbers to ensure the true variance has been sampled. Due to the temporal variation in stable isotope ratios, the reference dataset should also have been acquired during a similar timeframe as the test samples if they are to be reliably compared, since products from the same location could show different isotopic compositions if they were collected in different years or seasons. For terrestrial food products, new reference data are collected each year from the different geographic regions (Camin et al., 2017).

2.5.6 Conclusions

For terrestrial food products such as wines, 95% confidence limits are used in the authenticity testing of commercial samples (Camin et al., 2017). This indicates a similar accuracy of 95% would be necessary for any forensic tool used for origin traceability of traded marine food products in a commercial or legal setting. In the current study, the isotopic compositions of Atlantic cod muscle samples were not sufficiently distinct across all sampled regions to assign cod samples to the correct known origin with accuracy and

precision exceeding industry requirements of 95%. However, contrasts between certain discrete regions, such as the Baltic or Norwegian Sea compared with other regions, gave assignment accuracies around this level. This indicates that stable isotope techniques are unlikely to provide stand-alone tests for proof of origin with accuracy suitable for legislative action, but could be used in conjunction with genetic techniques to result in a higher level of accuracy than could be achieved using either method independently.

Chapter 3 Tracing the geographic origin of haddock and European hake products using stable isotope analysis

3.1 Abstract

Traceability of marine products throughout retail chains is increasingly important for the sustainable exploitation of fisheries, with rising demand for fish and seafood products around the world. Fraudulent practices and accidental mislabelling threaten sustainable management efforts by concealing the true catch location of products, but tracing the geographic origin of fish and seafood is currently challenging. In this study, I investigate whether carbon, nitrogen and sulfur stable isotope ratios in muscle tissue can be used to trace the catch region of two commercially important whitefish species – haddock (*Melanogrammus aeglefinus*) and European hake (*Merluccius merluccius*). I measured the isotopic composition of the white muscle tissue from 459 haddock from nine catch regions across the Northeast Atlantic, and 332 hake from five regions in the Northeast Atlantic and Mediterranean Sea. The stable isotope compositions of both haddock and hake varied according to catch region, and individuals of both species could be assigned to known origin with a similar mean classification success of 66-72%. However, haddock from the Norwegian Sea and Rockall as well as hake from the Mediterranean could be assigned to origin with high confidence (92-100% accuracy). Findings indicated that in general isotopic differences were conserved across whitefish species, so spatial differences from one species could be a reliable guide for other ecologically similar taxa. Therefore, stable isotope analysis shows good potential as a forensic tool to verify the spatial origin of haddock and hake products from certain regions, however it cannot provide sufficient accuracy for all fishery areas. Stable isotopes could be used in combination with genetic techniques to provide greater accuracy than either method used alone.

3.2 Introduction

Increasing demand for fish and seafood means that traceability of marine products is becoming ever more important for consumers, producers and regulators (Leal et al., 2015, Carter and Chesson, 2017, El Sheikha and Xu, 2017, Lewis and Boyle, 2017). Global consumption of aquatic foods has increased by an average of 3% per year since 1961, and fisheries and aquaculture production is now at a record high (FAO, 2022a). Highly complex and globalised supply networks (Leal et al., 2015) combined with the spatially based management of fisheries creates challenges for verifying the claimed catch region of

products. There is also incentive to conceal the geographic origin of illegally caught fish or those from a protected area (FAO, 2018). Therefore, fish and seafood are particularly vulnerable to incorrect claims of geographic origin, either by accidental mislabelling or through deliberate fraud for economic gain.

Widespread species substitution and mislabelling have been reported in some sectors of the seafood network through DNA testing (Pardo et al., 2016, Naaum et al., 2016, Warner et al., 2016). This includes whitefish, for example Miller and Mariani (2010) found that 25% of cod and haddock sampled were identified as a different species, and Helyar et al. (2014) discovered a significant number of mislabelled products from supermarket chains that claimed to contain cod or haddock. A 2016 review of 51 peer-reviewed papers using genetic methods to investigate seafood species substitution revealed that an average of 30% of samples were mislabelled with regard to species (Pardo et al., 2016). However, the occurrence of species mislabelling may have reduced in Europe in recent years with the widespread use of DNA testing (Warner et al., 2016). A more recent meta-analysis suggests the true species mislabelling rate is around 8% on average (Luque and Donlan, 2019).

However, verifying the catch location of fish products is much more challenging than identifying the species (El Sheikha and Montet, 2016, Gopi et al., 2019a, Cusa et al., 2021), because a tracer must show unique variations among multiple different regions on a suitable spatial scale. Currently there are no widely used forensic tests for the spatial origin of fish and seafood products, and therefore no equivalent estimates of the extent of geographic origin mislabelling exist. Traceability in the seafood supply chain at present relies primarily on document-based records, complemented by vessel monitoring systems on larger vessels. However, these vessel monitoring systems only improve traceability at first landing and are often not used on smaller fishing boats. The use of paper or electronic chain of custody data also cannot prevent accidental or deliberate mis-association of products to specific vessels. Blockchain technology may improve the confidence in supply chain documentation, but currently these techniques still rely on human input and therefore data initially entered into the blockchain can be falsified or inaccurate (Blaha and Katafono, 2020).

The stable isotope ratios of structural elements in tissues of living organisms (e.g. O, H, C, N, S, Sr) show spatial variation across marine and terrestrial environments and so can provide information on geographic origin. This technique is already routinely used for verifying the origin of many terrestrial food products, such as wine (West et al., 2007), cheese (Valenti et al., 2017), fruit juice (Rummel et al., 2010), olive oil (Camin et al., 2010) and honey (Bontempo et al., 2017), and is used to confirm claims of Protected Designation of Origin (PDO) labels on products (Camin et al., 2017). However, stable isotope tracers are significantly more challenging to apply to the spatial traceability of wild capture marine

food products, since wild fish are free to move across geographic regions and chemical gradients, and to feed on a wide range of incompletely known resources.

In the ocean, carbon and nitrogen stable isotopes are most frequently used as tracers for origin, since hydrogen and oxygen isotope ratios are relatively constant spatially across the ocean (Trueman et al., 2012b). However, oxygen isotopes can be used where there are large temperature gradients (Martino et al., 2022b). Sulfur isotopes have also been found to show spatial variations across open ocean marine environments (St. John Glew et al., 2019), although few studies so far have used sulfur isotopes for tracing capture location, with the exception of studies on scampi (Slesser and Trueman, 2021) and Manila clams (Won et al., 2021). The isotopic composition of carbon and, to a lesser extent, nitrogen in phytoplankton at the base of the food chain can be predicted mechanistically based on factors such as sea surface temperature, phytoplankton growth rate and species, nutrient source, dissolved CO_2 concentration and water column depth (Magozzi et al., 2017, Somes et al., 2010). The spatially varying isotopic composition of primary producers is subsequently transferred through the food chain and becomes incorporated into the tissues of consumers, acting as a natural marker for animals feeding in different locations. Therefore the isotopic composition of food products can be used to differentiate between those from different locations and to infer origin (Trueman and St. John Glew, 2019). However, tissue-diet isotopic spacing (where consumer tissues are systematically enriched in the heavier isotope of nitrogen and, to a lesser extent, of carbon through the food chain) means that species from the same geographic region but with different diets will form tissues with differing stable isotope compositions. This also applies to the same species where individuals of different body sizes occupy different trophic levels, through an ontogenetic shift in diet (Jennings et al., 2008).

Despite their widespread use in terrestrial food traceability, stable isotopes ratios have not been widely used to trace the origin of wild-caught marine food products due to the significant practical difficulty of obtaining reference samples in the vast and relatively inaccessible open ocean. Most studies to date have focused on farmed animals, such as shrimps (Carter et al., 2015). For farmed animals, possible origin locations are discrete, fixed and relatively easily sampled, and the exact composition of the aquaculture feed is known. Of the relatively few isotope-based studies on tracing the spatial origin of wild-caught seafood undertaken so far, species include sea bass (Farabegoli et al., 2018, Varrà et al., 2019), turbot (Busetto et al., 2008), tuna (Chouvelon et al., 2017, Rampazzo et al., 2020) and hake (Carrera and Gallardo, 2017), as well as commercially important invertebrates such as shrimp (Ortea and Gallardo, 2015, Gopi et al., 2019b), scampi (*Nephrops norvegicus*) (Slesser and Trueman, 2021), scallops (Trueman et al., 2017, Zhang et al., 2019), jumbo squid (Gong et al., 2018), Manila clams (Won et al., 2021) and sea cucumbers (Zhang et al., 2017, Kang et al., 2020). These studies suggest that stable

isotope analysis is a promising method for determining origin in some wild-caught species. For example, Zhang *et al.* (2019) were able to correctly classify three scallop species to their origin locations around China using carbon and nitrogen stable isotopes with an average accuracy of 92% (Zhang et al., 2019). Another study successfully assigned sea bass to their origin region along the coast of Wales using the stable isotope composition of the scales, and achieved an assignment accuracy of around 75% (Cambiè et al., 2016). One of the only studies to date that has used sulfur stable isotopes for traceability of marine organisms is that by Slesser and Trueman (2021) who found the ability to distinguish among catch locations of scampi was increased with sulfur isotopes in addition to carbon and nitrogen. However, stable isotope analysis alone may not give sufficient accuracy, and combining isotope markers with another technique, such as genetic, fatty acid or trace element markers, may give better results. For example, Ortea and Gallardo (2015) traced seven shrimp species to multiple geographic origins and found that 71-75.6% were correctly classified using stable isotopes alone, but this increased to 93.5-100% correct when used together with elemental analysis (Ortea and Gallardo, 2015). Furthermore, my research in Chapter 2 demonstrates that isotope-based tracers in Atlantic cod are not sufficiently distinct across all sampled regions to be used as a stand-alone test for origin, and that it would be necessary to combine stable isotopes with another approach to achieve a level of accuracy that meets industry requirements.

Traceability studies have been applied to a very small number of species. Of all the marine finfish species with a global capture production over 1 million tonnes live weight in 2020 (total 13 species) as reported by the FAO (FAO, 2022c), only three have had spatial traceability methods applied (Atlantic herring, cod and mackerel (Trueman et al., 2017, Nielsen et al., 2012b, Rampazzo et al., 2020)), accounting for only 5.6% of the total finfish production. This lack of traceability studies partly reflects the cost involved with creating reference datasets and the absence of economic incentive for producers, suppliers and retailers to conduct that work. In previous work (Chapter 2), I demonstrated the potential for stable isotope tracers to discriminate among cod fisheries in the Northeast Atlantic. Here I aim to extend that work to determine whether the same tracers have similar discriminatory power for two other commercially important whitefish species – haddock and European hake.

Advanced genetic techniques have been successfully applied to determine the origin of whitefish from several populations on a relatively large spatial scale. Single nucleotide polymorphisms (SNPs) were used to discriminate between European hake from the Atlantic and Mediterranean, and achieved a success rate of 98% when individuals were assigned to either ocean basin (Nielsen et al., 2012b). Nielsen et al. (2012b) also used SNPs to discriminate among Atlantic cod from the North Sea, Baltic Sea and Northeast Arctic populations and traced individuals to a priori-defined population of origin with 98-100%

correct classification. However, these populations on a large spatial scale were selected due to their expected genetic differences, and classification accuracy was determined based on individuals sampled within these populations (rather than spread across the total range of the species). The potential for accurate assignment using genetic markers is reduced for species with mixed populations, fisheries operating at the margins of discrete populations, or fisheries targeting common feeding grounds used by multiple reproductively isolated populations. Studies have shown that the division between the genetically isolated Atlantic and Mediterranean hake populations is very distinct, but hake populations within one ocean basin have low genetic differentiation (Milano et al., 2014, Pita et al., 2011), as do haddock populations (Berg et al., 2021). Proteomics has also been investigated by Gonzalez et al. (2010) as a potential tool for distinguishing among European hake populations. Their results demonstrated that individuals from the Mediterranean Sea, Cantabrian Sea, and Atlantic Ocean could be correctly classified to capture region with 100% success rate using liver and brain protein biomarkers.

Combining genetic techniques with stable isotope analysis may be a powerful way to benefit from the strengths and reduce the weaknesses of both approaches, since for genetic markers the signal is based on reproductive isolation whereas stable isotope tracers rely on spatial differences in the environment at the point of capture. The potential for combining genetic and isotopic markers for seafood traceability was assessed by Cusa et al. (2021). This identified a lack of studies combining both approaches on common samples or using common sampling approaches.

Haddock (*Melanogrammus aeglefinus*) is one of the most commercially important demersal fish species in the north Atlantic. It is a key species fished by the UK fleet who landed 28,500 tonnes into the UK in 2020, with a value of £39.6 million (MMO, 2021). The household consumption of haddock in the UK is the second highest of any fish or seafood species, and therefore the UK is a net importer of haddock, importing this species to a value of almost £160 million in 2020 (MMO, 2021). The Marine Stewardship Council has certified haddock fisheries in the Northeast Arctic, Iceland, Rockall and the North Sea as being sustainable (Marine Stewardship Council, 2022). However for certain stocks, including the Northeast Arctic, Faroes and Rockall fisheries, although the spawning stock biomass is healthy, the fishing mortality is higher than that consistent with achieving Maximum Sustainable Yield (ICES, 2021h, ICES, 2021f, ICES, 2021g). Therefore, although haddock stocks are sustainably fished at present, these fisheries may be at risk of overexploitation in the future.

European hake (*Merluccius merluccius*) is also an important commercial species in the North Atlantic and Mediterranean Sea. Global capture production totalled 107,619 tonnes live weight in 2020 (FAO, 2022b). Hake also had the second highest landings of all demersal fish in the EU in 2018 with a value of 490 million euros, of which European hake

accounted for around 66% (European Commission, 2022). Hake is among the top ten species consumed in the EU, particularly in Spain, where it has the highest consumption of all fish and seafood (European Commission, 2022). Hake is also a very important commercial species in the Mediterranean, where it is one of the top three commercial species in terms of value (FAO, 2020a). The main European hake fisheries are in the areas north and west of Scotland, Celtic Sea, Bay of Biscay and the Atlantic coast of the Iberian Peninsula, as well as in the Mediterranean (FAO, 2022b). European hake fisheries in the Atlantic are divided into two management units by ICES - the northern and southern stocks. The geographic boundary between these two stocks lies at Cap Breton Canyon, close to the border between the French and Spanish coasts (ICES, 2013). The northern stock is currently fished at sustainable levels and the spawning stock biomass is above the target level (ICES, 2022g). Therefore, there are fisheries in the North Sea and Cornwall that have been certified as sustainable by the Marine Stewardship Council (MSC) (Marine Stewardship Council, 2022). However, the Mediterranean hake stock is severely depleted due to overexploitation (FAO, 2020a) and has been listed as vulnerable to extinction on the IUCN Red List of Threatened Species since 2007 (IUCN, 2022). The minimum landing size of European hake differs between the Atlantic (27cm) and the Mediterranean (20cm) (GFCM, 2016, MMO, 2018), which may give an incentive to claim a false catch location for undersized Atlantic fish. Therefore, the ability to reliably determine the catch location of hake is of vital importance to ensure any management plans succeed and to secure the future sustainability of the species, especially where stocks are in contrasting states of health.

The spatial traceability of haddock has not been well studied, in fact no studies could be found in the literature that investigated the accuracy with which haddock could be assigned to origin. Since it is such an important traded fish species, studies into the potential for verifying the geographic origin of haddock using natural tracers would be very beneficial. Similarly, very few published studies on the spatial traceability of hake could be found. Nielsen et al. (2012b) distinguished between hake from the Atlantic and Mediterranean using SNPs (single-nucleotide polymorphisms) but no studies to date have been conducted on a smaller spatial scale, so more research is needed to determine whether European hake can be traced to fishery of origin using natural biochemical markers.

The development of a reliable and effective test for the geographic origin of traded whitefish is therefore important to maintain sustainable fish stocks under increasing demand, to allow the recovery of overfished stocks and to ensure consumer confidence in products marketed as sustainable. For terrestrial food products, 95% confidence limits are used for authenticity testing of commercial samples and in legal cases (Camin et al., 2017). Any tool developed to verify the claimed catch location of products throughout the supply chain must meet these accuracy and precision standards.

Discrete assignment traces a test sample to one of a number of distinct regions, and known origin reference samples must be obtained from each of these regions. Test samples are compared to the reference samples from all the regions to determine which is the most likely origin. Samples can only be assigned to one of the discrete regions contained in the reference dataset, so any test samples originating from outside of these regions will be incorrectly assigned. It is important to consider the method for testing accuracy of assignment, since falsely high estimates of accuracy can result if the reference samples do not cover all regions across the full distributional range of the fish, the regions are decided a priori based on expected differences, or the reference dataset is not independent of the test samples used to assess accuracy.

The aim of this study is to investigate whether the stable isotope composition of haddock and European hake muscle can provide geographic discrimination at a level that has practical use in fisheries management and to assess the potential for stable isotope tracers to be used as a forensic tool for verifying the spatial origin of traded fish products. To test this, we measured the carbon, nitrogen and sulfur stable isotope ratios in 459 haddock and 332 hake muscle tissue samples of known geographic origin to construct a reference dataset of fish from a wide range of catch locations across the Northeast Atlantic and one region of the Mediterranean Sea. We then used this to determine the accuracy with which individuals can be traced back to their true origin region.

3.3 Methodology

3.3.1 Sampling

Haddock (*Melanogrammus aeglefinus*) and European hake (*Merluccius merluccius*) were chosen for the study due to their commercial importance in the Northeast Atlantic. Samples of haddock and hake were collected from nine regions within the Northeast Atlantic Ocean and for hake also from one region in the Mediterranean Sea, with multiple stations sampled at each region, between February and December 2018 (Figures 3.1 and 3.2). To ensure authenticity of the catch locations, samples were obtained from research fisheries surveys carried out by the relevant research organisation in the target regions. The research organisations who supplied the samples are listed in Tables 3.1 and 3.2. The sampling regions were selected to cover the main commercial fishery areas, but were also based on the availability of samples. Regions sampled correspond to the ICES subareas within FAO major fishing area 27. Stations for each region are located within one subarea, except for the region referred to as the Norwegian Sea in which stations fall within both ICES subareas 1 and 2 (Barents and Norwegian Sea subareas) close to the boundary between the two (Figure 3.1).

Haddock

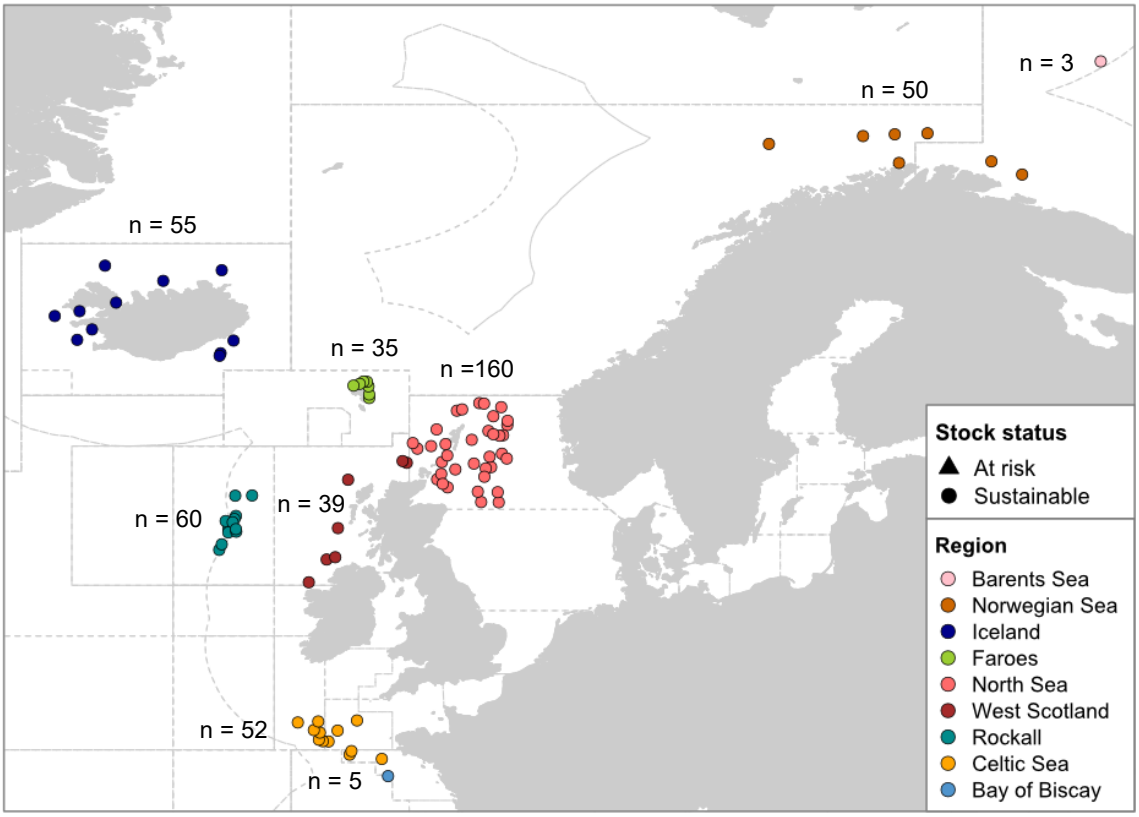


Figure 3.1 Locations of all stations where samples of haddock were obtained, coloured according to the geographic regions. The total number of samples (n) collected from each region are shown and the symbols indicate whether the stock is sustainable or at risk (as determined from the ICES 2022 advice for each stock (ICES, 2022h), based on the spawning stock biomass relative to the limit where reproduction of the stock is impaired (B_{lim})). ICES subarea boundaries are indicated by the dotted grey lines.

European hake

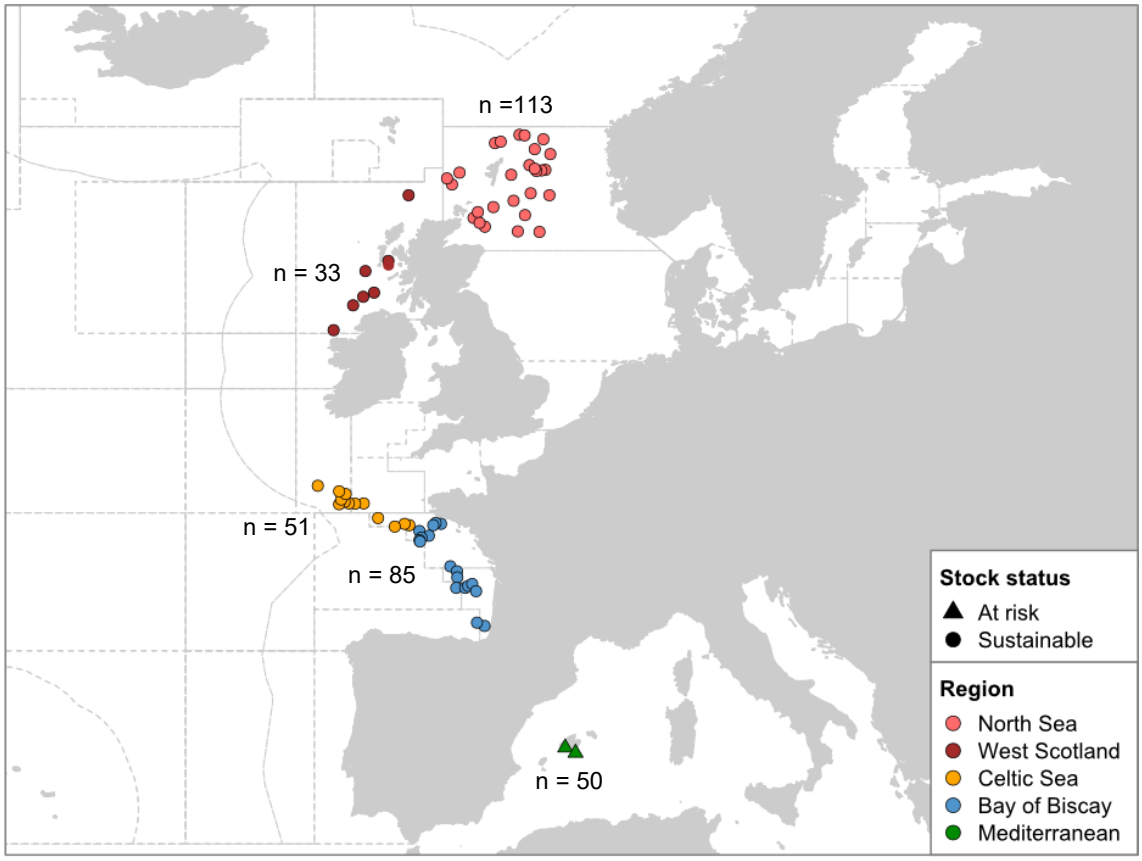


Figure 3.2 Locations of all stations where samples of hake were obtained, coloured according to the geographic regions. The total number of samples (n) collected from each region are shown and the symbols indicate whether the stock is sustainable or at risk (as determined from the ICES 2022 advice for each stock (ICES, 2022h), based on the spawning stock biomass relative to the limit where reproduction of the stock is impaired (B_{lim})). ICES subarea boundaries are indicated by the dotted grey lines.

The aim was to sample at least 50 individuals of each species from each of the target regions so that the samples would encompass the natural variability of wild populations, however this was not always possible. The total numbers of samples collected from each region are listed in Tables 3.1 and 3.2. To ensure fish were not all caught from the same shoal, which may have travelled as one group and therefore have very similar isotope compositions, multiple stations within each region were sampled. A maximum of five individuals were collected from each station if possible, although sampling constraints meant that in some cases a greater number were used from certain stations. In order to better represent samples encountered in real world traceability situations, the body sizes of fish sampled were not measured, however only individuals of commercial size were used (>30cm for haddock and >27cm for hake).

A piece of white muscle tissue of approximately 3cm^3 was taken from each individual and placed in a zip-lock bag. Details including the catch location, species, date collected and station number (haul) were recorded for each. Muscle tissue samples were immediately frozen at -20°C and were stored frozen until preparation for analysis, excluding samples from the Faroe Islands which were instead preserved in 70% ethanol for transport, due to the transportation time back to the UK, before being stored frozen in the laboratory.

Table 3.1 Total number of individual haddock collected from each geographic region, the number of stations sampled per region, and the source from which samples were obtained.

Catch region	Number of haddock analysed	Number of stations	Year samples collected	Source of samples and organisation name
Barents Sea	3	Unknown	2017	Young's Seafood Ltd.
Norwegian Sea	50	7	2018	Annual fisheries survey - Institute of Marine Research
Iceland	55	11	2018	Annual fisheries survey - Marine and Freshwater Research Institute
Faroe Islands	35	7	2018	Annual fisheries survey - Faroe Marine Institute
North Sea	160	35	2018	Annual fisheries survey - Marine Scotland
West Scotland	39	7	2018	Annual fisheries survey - Marine Scotland
Rockall	60	12	2018	Annual fisheries survey - Marine Scotland
Celtic Sea	52	12	2018	Annual fisheries survey - Ifremer
Bay of Biscay	5	1	2018	Annual fisheries survey - Ifremer
Total	**459**	**92**		

Table 3.2 Total number of individual hake collected from each geographic region, the number of stations sampled per region, and the source from which samples were obtained.

Catch region	Number of hake analysed	Number of stations	Year samples collected	Source of samples and organisation name
North Sea	113	27	2018	Annual fisheries survey - Marine Scotland
West Scotland	33	8	2018	Annual fisheries survey - Marine Scotland
Celtic Sea	51	13	2018	Annual fisheries survey - Ifremer
Bay of Biscay	85	18	2018	Annual fisheries survey - Ifremer
Mediterranean	50	2	2018	Local fish market in Palma de Mallorca - IMEDEA
Total	**332**	**68**		

3.3.2 Stable isotope analysis

The frozen whitefish muscle samples were prepared for stable isotope analysis by freeze-drying at -55°C for 24 hours, before being ground to a powder using scissors and weighed into tin capsules (2.5-2.8mg per sample). A total of 459 haddock and 332 hake samples were analysed for bulk carbon, nitrogen and sulfur stable isotope ratios at the Life Sciences Mass Spectrometry Facility (LSMSF) in East Kilbride, United Kingdom, using an Elementar vario PYRO cube elemental analyser coupled with an Isoprime visION isotope ratio mass spectrometer.

Stable isotope compositions are expressed as $\delta^{13}C$, $\delta^{15}N$ and $\delta^{34}S$ values in per mille (‰) relative to the international standards (Vienna-PeeDee Belemnite, air, and Vienna-Canyon Diablo Troilite respectively). Laboratory internal standards (MSAG, M2 and SAAG2), two other standards (fish muscle standard and an in-house glutamic acid standard) and blanks were run together with the samples to monitor accuracy and precision of measurements.

Lipid correction was carried out in the same way as described in Chapter 2, whereby samples with a high lipid content were arithmetically corrected. A total of 15 haddock and 47 hake samples had a correction applied.

3.3.3 Statistical analysis

The sampling area was divided into pre-defined geographic regions reflecting geopolitical and hydrographic basins. Samples were assigned to one of these discrete possible

populations by comparing the isotopic composition of each sample to the distribution of isotopic values measured within a reference population of known origin. The region that most closely matches the isotopic composition of the test sample is taken to be the most likely origin. This method is particularly well suited to traceability of fish and seafood products where the ocean is divided into distinct management units such as FAO subareas.

Two methods of assignment to origin were compared to determine which is most effective – a multivariate normal probability technique and random forest classification. These assignment methods were carried out as described in Chapter 2 (page 42), for haddock and hake separately. The multivariate technique was conducted first using all three isotopes and then using only carbon and nitrogen isotope data to allow the contribution of sulfur isotopes to spatial traceability to be assessed. Random forest classification was carried out using carbon, nitrogen and sulfur stable isotope data. All analyses were carried out using R version 3.6.2 (R Core Team, 2019).

Both of the assignment methods above suffer from non-independence of the training and test samples, which may result in overestimation of the assignment accuracy. This is a common limitation in many assignment studies. A more robust test of assignment accuracy would be to assign independently collected test samples of known origin. Therefore, in this study a dataset of carbon and nitrogen stable isotope ratios measured in haddock and hake was obtained from Ifremer from the EVHOE 2014 survey. Further haddock samples were acquired from Young's Seafood Ltd., collected in 2016-2017 from the Norwegian Sea and Iceland. Supplied $\delta^{13}C$ values were corrected for the Suess effect by applying a -0.022‰ per year adjustment (Young et al., 2013). Samples were then assigned to their most likely origin region using the bivariate normal probability assignment method described in Chapter 2 and the known origin samples from the current study as a reference dataset. This allowed us to investigate the success of tracing provenance using independent test samples that were collected during different years than the reference dataset.

3.3.3.1 North Sea isoscape

Haddock samples were collected from 35 different stations within the North Sea and hake from 27 different stations, to allow the spatial variation among sampling stations to be investigated. Carbon, nitrogen and sulfur isoscapes were created for the sampled North Sea area for haddock and hake separately, using interpolation by kriging between the sampled stations with a grid size of 0.1 degrees. Between one and five samples were measured at each station, and where more than one sample was collected at a station the mean value was used. The associated spatial variances were also calculated for each of the three isoscapes. These analyses were performed using the 'gstat' package (Pebesma, 2004, Gräler et al., 2016) in R version 3.6.2 (R Core Team, 2019).

3.4 Results

Measurement error associated with the stable isotope analysis of the fish muscle samples is 0.1-0.4‰ for $\delta^{13}C$, 0.1-0.2‰ for $\delta^{15}N$ and 0.6-0.7‰ for $\delta^{34}S$, determined as the standard deviation of replicate measurements of two internal standards - a fish muscle standard and a glutamic acid standard - measured together with the samples. The internal standards used were consistent across analytical laboratories, allowing comparison between the measurements from the two sites. Means of repeated analyses of fish muscle and glutamic acid standards from both laboratories are within one standard deviation of each other (Table 3.3), demonstrating that there are no significant differences in the measurements between laboratories.

Table 3.3 Comparison of the mean values and uncertainties (standard deviations) in the stable isotope measurements of two internal standards measured at both laboratories where samples were analysed.

Standard	Laboratory	$\delta^{13}C$ (‰)		$\delta^{15}N$ (‰)		$\delta^{34}S$ (‰)	
		Mean	SD	Mean	SD	Mean	SD
Fish muscle	NOCS	-19.3	0.11	11.4	0.17	19.3	0.64
	LSMSF	-19.3	0.07	11.3	0.11	18.5	0.74
Glutamic acid	NOCS	-13.1	0.17	-3.8	0.06	NA	NA
	LSMSF	-13.6	0.43	-3.9	0.15	NA	NA

The C:N ratios are very similar across all samples, which suggests that the storage of the Faroes haddock samples in 70% ethanol had very little effect on the stable isotope values, since no extraction of lipids occurred. The stable isotope compositions of haddock muscle samples differed significantly among the nine regions (ANOVA - $\delta^{13}C$: f=116, d.f.=8, $p<2e^{-16}$; $\delta^{15}N$: f=97.23, d.f.=8, $p<2e^{-16}$; $\delta^{34}S$: f=29.87, d.f.=8, $p<2e^{-16}$). The $\delta^{13}C$, $\delta^{15}N$ and $\delta^{34}S$ values measured in the haddock samples from each region are shown in Figure 3.3, and the means and standard deviations for each location are listed in Table 3.4. Overall, $\delta^{13}C$ values in haddock muscle range from -21.7‰ to -16.6‰. Haddock from the Norwegian Sea have the lowest $\delta^{13}C$ values of all the regions, with a mean of -20.2‰ (Figure 3.3). All other regions show a similar range of $\delta^{13}C$ values with a slight increase from south to north, with the exception of the Barents Sea haddock which also have fairly low $\delta^{13}C$ values (mean of -18.9‰). The nitrogen isotope values also show a general trend of increasing $\delta^{15}N$ values from south to north, apart from the Celtic Sea and Bay of Biscay samples which have relatively high $\delta^{15}N$ values. The Barents Sea haddock show the highest mean $\delta^{15}N$ value of 14.9‰, and the haddock from Rockall have the lowest values, with a mean of 10.6‰. Many regions show similar $\delta^{34}S$ values, ranging from 13.4‰ to 20.8‰ overall.

However, the Norwegian Sea haddock are the most distinct since this region has slightly higher sulfur isotope values than the other regions. The Icelandic and North Sea samples have the greatest range of $\delta^{34}S$ values (5.4‰ and 6.2‰ respectively).

Figure 3.4 shows frequency distributions for the carbon, nitrogen and sulfur stable isotope data for haddock from each region. The majority of regions show a normal distribution, with a few exceptions.

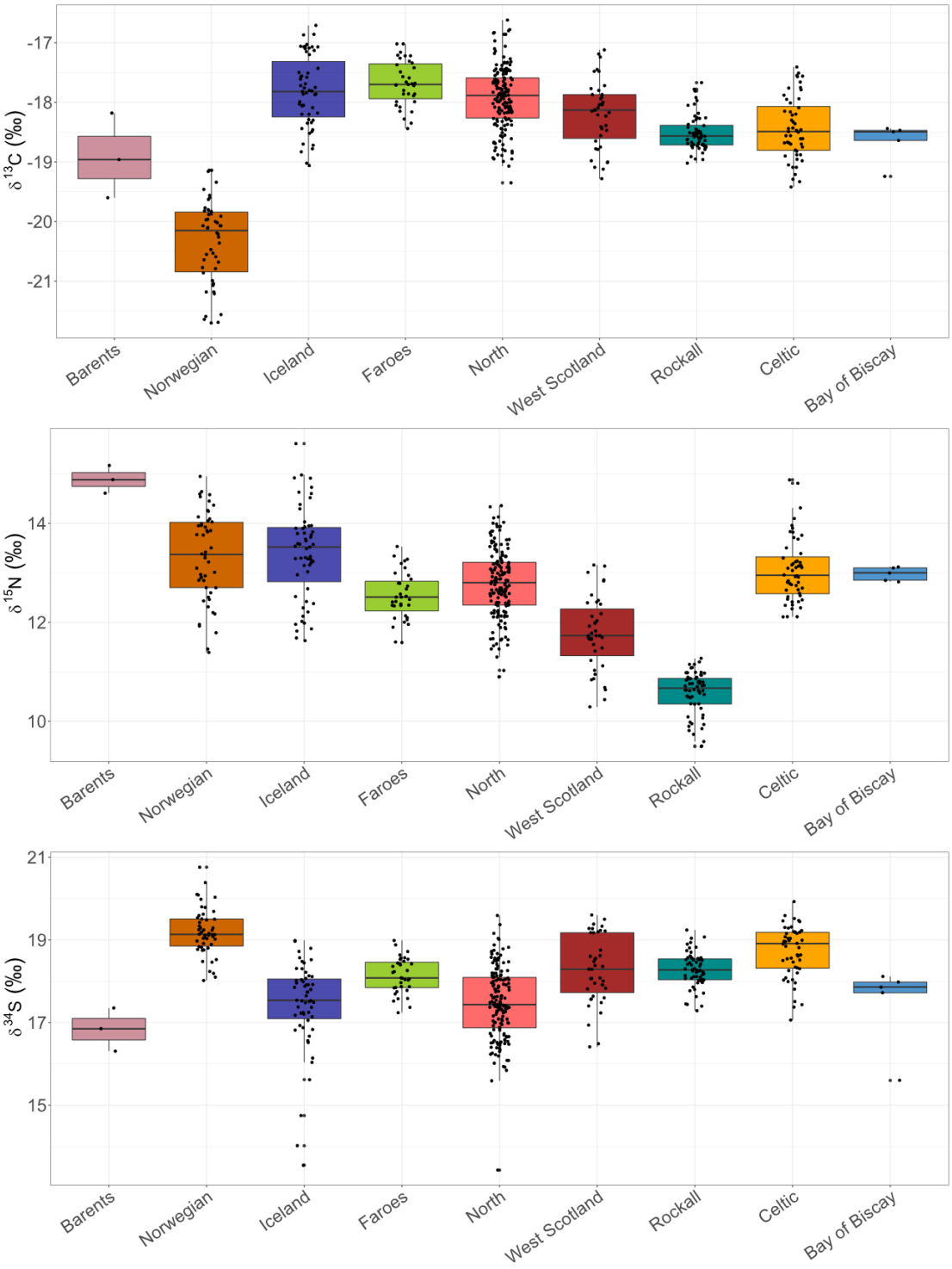


Figure 3.3 Lipid-corrected $\delta^{13}C$, $\delta^{15}N$ and $\delta^{34}S$ values measured in the haddock muscle tissue samples from each geographic region.

Table 3.4 Mean and standard deviation of carbon (lipid corrected), nitrogen and sulfur stable isotope ratios from haddock caught in each geographic region.

Region	$\delta^{13}C$ (‰)		$\delta^{15}N$ (‰)		$\delta^{34}S$ (‰)	
	Mean	SD	Mean	SD	Mean	SD
Barents Sea	-18.9	0.41	14.9	0.16	16.8	0.30
Norwegian Sea	-20.2	0.11	13.5	0.13	18.9	0.13
Iceland	-17.8	0.08	13.4	0.12	17.4	0.14
Faroes	-17.7	0.06	12.5	0.08	18.1	0.07
North Sea	-17.9	0.04	12.8	0.06	17.5	0.07
West Scotland	-18.2	0.09	11.8	0.12	18.3	0.13
Rockall	-18.5	0.04	10.6	0.06	18.3	0.05
Celtic Sea	-18.4	0.07	13.1	0.09	18.7	0.09
Bay of Biscay	-18.7	0.15	13.0	0.06	17.5	0.47

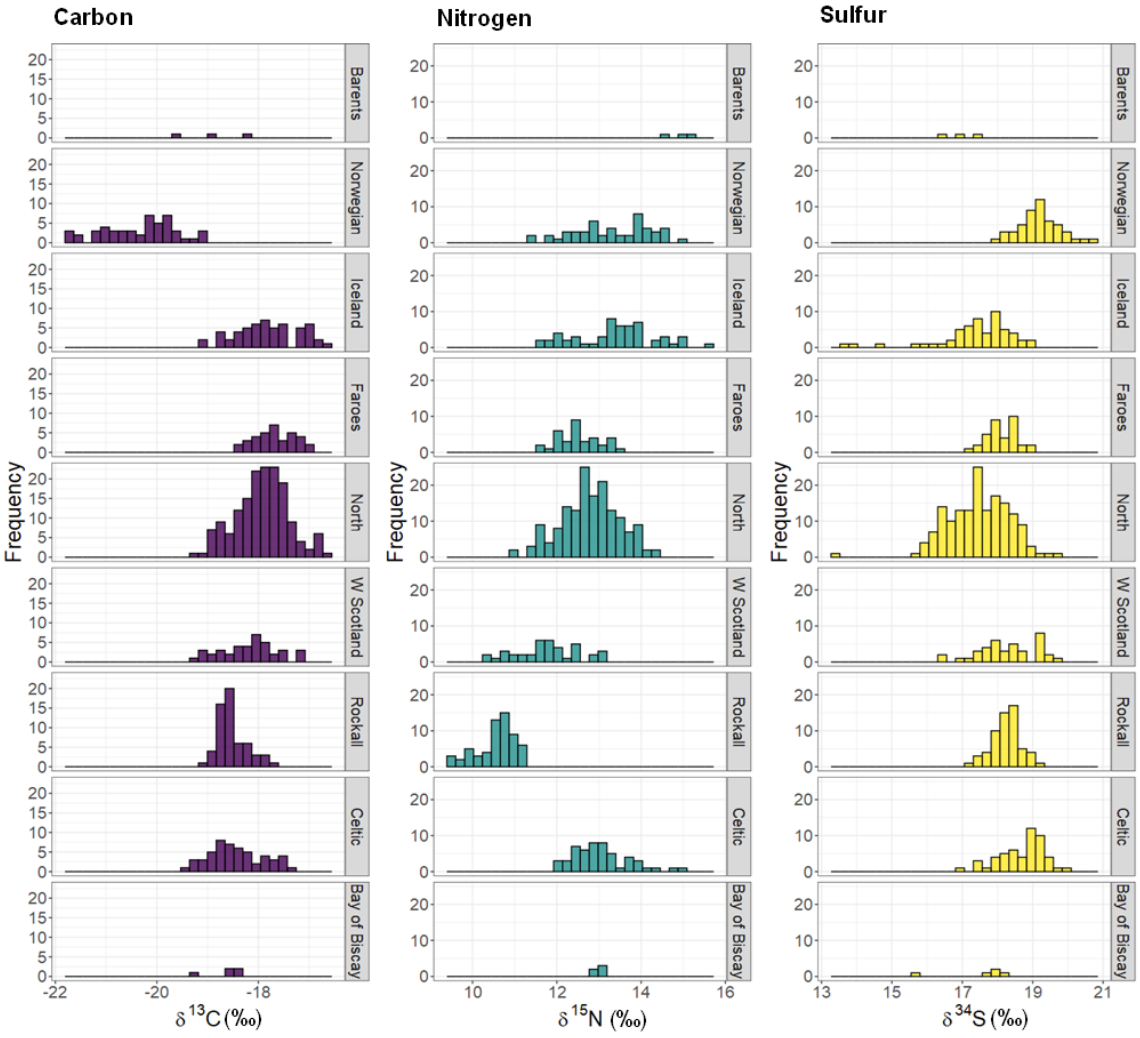


Figure 3.4 Frequency distributions of $\delta^{13}C$, $\delta^{15}N$ and $\delta^{34}S$ values from haddock caught within each of the nine sampled regions.

The stable isotope compositions of hake muscle samples also differed significantly among the five sampled regions (ANOVA - $\delta^{13}C$: f=21.16, d.f.=5, $p<2e^{-16}$; $\delta^{15}N$: f=354.7, d.f.=5, $p<2e^{-16}$; $\delta^{34}S$: f=94.91, d.f.=5, $p<2e^{-16}$). The $\delta^{13}C$, $\delta^{15}N$ and $\delta^{34}S$ values measured in the hake samples from each region are shown in Figure 3.5, and the means and standard deviations for each location are listed in Table 3.5. The $\delta^{13}C$ values are relatively similar across all regions, ranging from -20.5‰ to -17.2‰. The Mediterranean hake have slightly lower $\delta^{13}C$ values, but these still show overlap with all other regions. The Mediterranean samples are also the most distinct in their $\delta^{15}N$ values, since they are significantly lower than the other four regions, with a mean of 8.9‰ compared to a range of 11.1‰ to 15.5‰ at the remaining regions (Table 3.5). Similarly, the Mediterranean hake are differentiated from the Atlantic regions by their greater $\delta^{34}S$ values (mean of 19.9‰). The Atlantic regions all show overlapping $\delta^{34}S$ values with relatively large ranges, from a minimum of 15.2‰ to a maximum of 19.5‰. The hake from the Bay of Biscay have a slightly lower mean $\delta^{34}S$ value, but still overlap with the other regions (Figure 3.5)

Frequency distributions for the carbon, nitrogen and sulfur stable isotope data are shown in Figure 3.6, and reveal that the data for hake from most regions are normally distributed. In particular, all regions show a normal distribution in the $\delta^{15}N$ values, although for $\delta^{13}C$ and $\delta^{34}S$ there are exceptions.

Table 3.5 Mean and standard deviation of carbon (lipid corrected), nitrogen and sulfur stable isotope ratios from European hake caught in each geographic region.

Region	$\delta^{13}C$ (‰)		$\delta^{15}N$ (‰)		$\delta^{34}S$ (‰)	
	Mean	SD	Mean	SD	Mean	SD
North Sea	-19.1	0.04	13.5	0.06	18.1	0.06
West Scotland	-19.0	0.09	13.6	0.07	18.1	0.09
Celtic Sea	-18.7	0.10	12.9	0.15	18.5	0.08
Bay of Biscay	-18.8	0.04	13.4	0.06	17.6	0.08
Mediterranean	-19.5	0.05	8.9	0.10	19.9	0.04

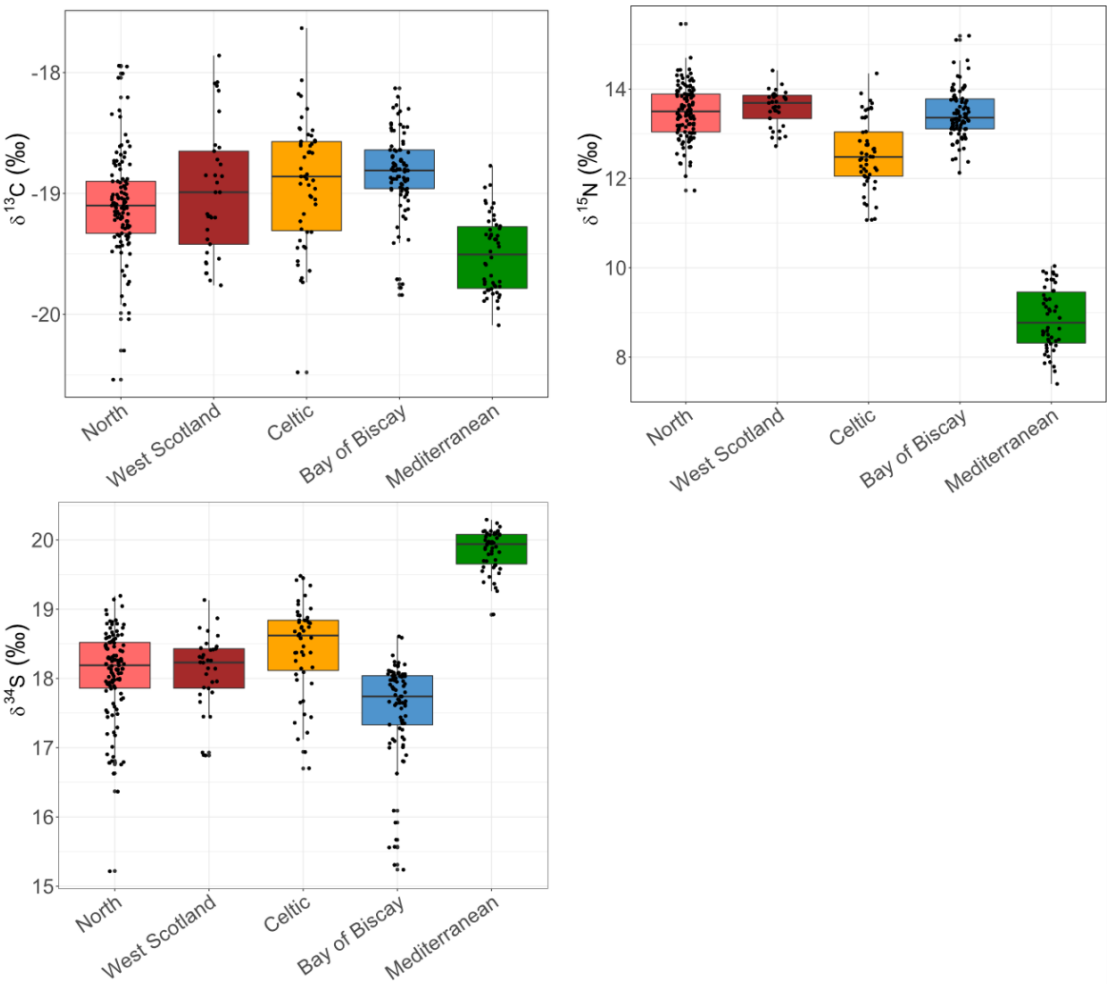


Figure 3.5 Lipid-corrected $\delta^{13}C$, $\delta^{15}N$ and $\delta^{34}S$ values measured in the hake muscle tissue samples from each geographic region.

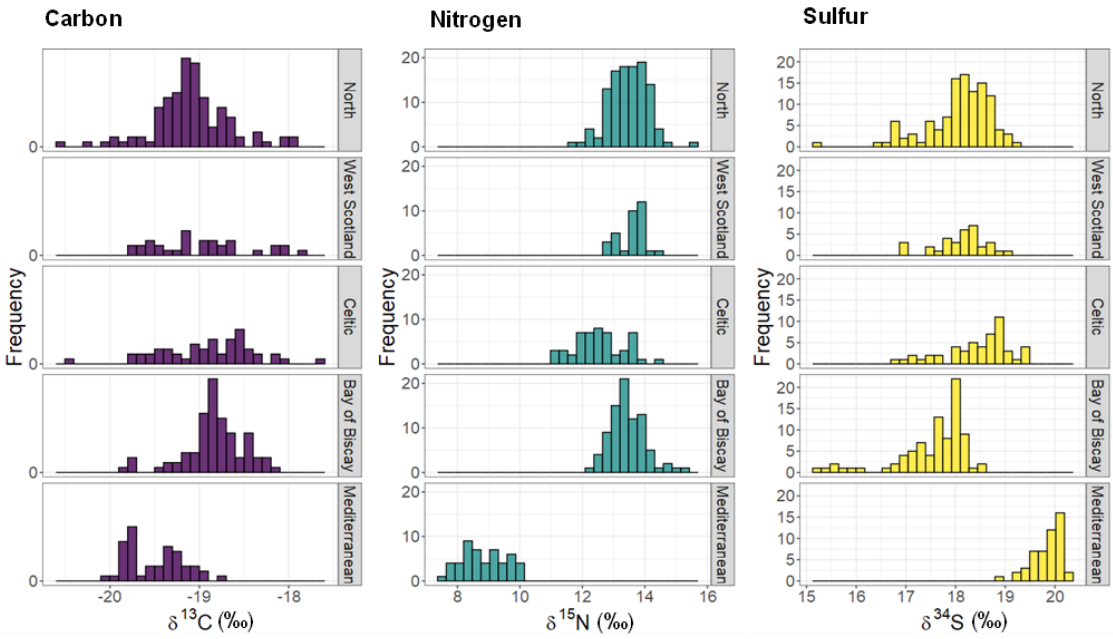


Figure 3.6 Frequency distributions of $\delta^{13}C$, $\delta^{15}N$ and $\delta^{34}S$ values from European hake caught within each of the five sampled regions.

3.4.1 Clustering by region

3.4.1.1 Haddock

The stable isotope compositions of the haddock muscle tissue reveal some clustering by region, although there is significant overlapping of regions (Figure 3.7). The Norwegian Sea samples form the most distinct cluster, which has little overlap with other regions due to its lower $\delta^{13}C$ values. The haddock from Rockall are also quite easily separated from the other regions, forming a clear cluster with lower $\delta^{15}N$ values than most other samples and only showing overlap with the West Scotland haddock. The Iceland, Faroes, North Sea, Celtic Sea and Bay of Biscay clusters all overlap each other and are difficult to distinguish, since they all appear to have relatively similar isotopic compositions. The $\delta^{34}S$ values are similar across all regions, indicating that sulfur isotopes may have limited use in assigning samples to origin.

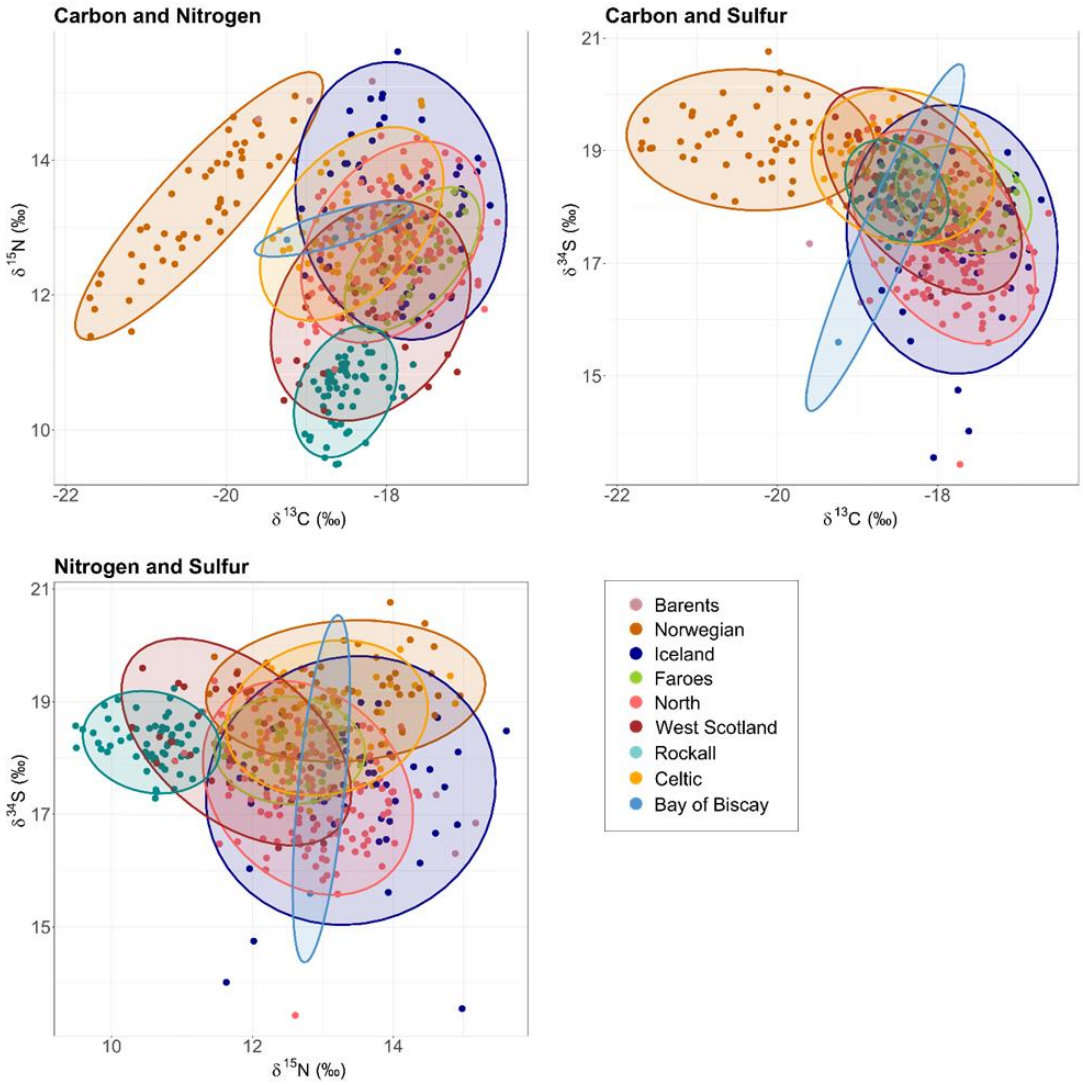


Figure 3.7 Carbon, nitrogen and sulfur stable isotope values for each individual haddock sampled, coloured by region of origin. The 90% data ellipses are also shown for each geographic region.

3.4.1.2 Hake

The European hake samples also show clear clustering by origin region (Figure 3.8). The Mediterranean hake are easily distinguished from the Atlantic hake with no overlap, due to the significantly lower $\delta^{15}N$ values and slightly higher $\delta^{34}S$ values. The Atlantic regions are less distinct, showing large overlaps among regions. However, hake from the Celtic Sea have slightly lower $\delta^{15}N$ values and therefore may be more easily differentiated from the other Atlantic regions. The $\delta^{13}C$ values are similar across all regions, both Mediterranean and Atlantic, so may not provide significant benefit in assigning samples to origin. Similarly, the $\delta^{34}S$ values are similar across all Atlantic regions.

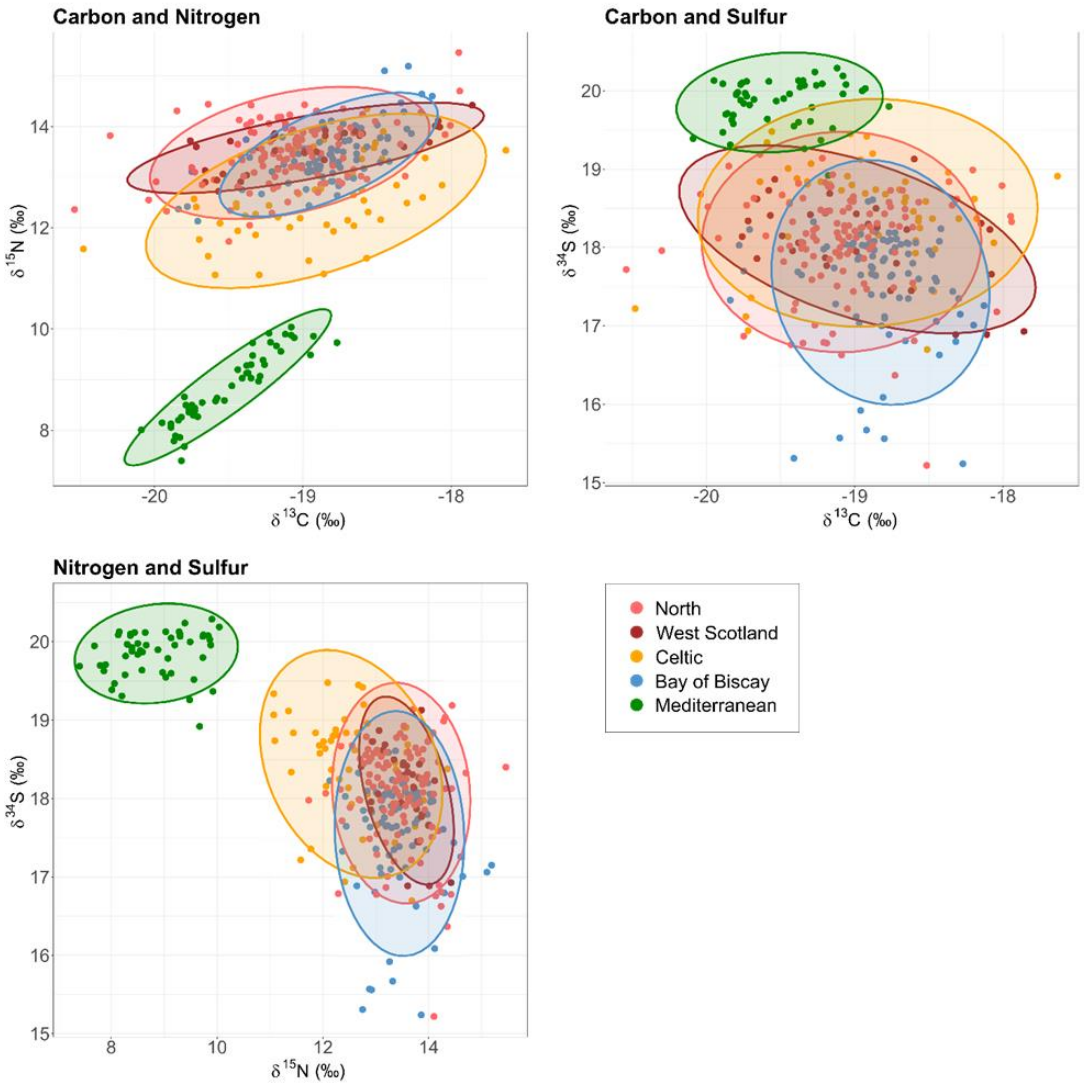


Figure 3.8 Carbon, nitrogen and sulfur stable isotope values for each individual hake sampled, coloured by region of origin. The 90% data ellipses are also shown for each geographic region.

3.4.2 Assignments to catch region

3.4.2.1 Haddock assignment using multivariate technique

The average accuracy of assignment to known origin among all sampled regions was 60.3% using the multivariate normal probability technique with all three stable isotopes (carbon, nitrogen and sulfur), and the success rates for each location are shown in Table 3.6. Assignment accuracy varied with the repeated simulations depending on which samples were selected for the training and test datasets during the assignment process (Figure 3.9). Haddock from the Norwegian Sea and Rockall were the most accurately assigned, with an average success rate of 98% and 92% respectively. Where the Rockall samples were incorrectly assigned, it was to West Scotland which is the next closest region. The Faroes and Celtic Sea samples could also be assigned with moderate accuracy (67-70%). However, other regions were more challenging to discriminate. The North Sea and Iceland haddock could only be assigned with mean accuracy of 38%, with samples being assigned to a range of incorrect regions (Figure 3.9). The Bay of Biscay also had a low assignment success rate of 21%, although this could be partly a result of the small sample size (n = 5). The incorrectly assigned Bay of Biscay samples were also classified to the nearest neighbouring region – the Celtic Sea.

The results of multivariate assignment when sulfur isotope values were excluded, and instead only carbon and nitrogen isotopes were used, are shown in Table 3.6 and illustrated in Figure 3.10. The overall correct assignment rate was very similar to using all three isotopes (58.8%). However, the assignment accuracy for the Bay of Biscay was greatly increased, from 21% to 60%. Other regions showed a much reduced assignment accuracy, for example the success rate for the North Sea decreased from 38% to 17% and the Celtic Sea assignment success reduced from 67% to 40%. Therefore, overall this indicates that the use of all three stable isotopes ($\delta^{13}C$, $\delta^{15}N$ and $\delta^{34}S$) results in the highest rate of correct assignments to origin for haddock, since sulfur does contribute additional power to distinguish among regions.

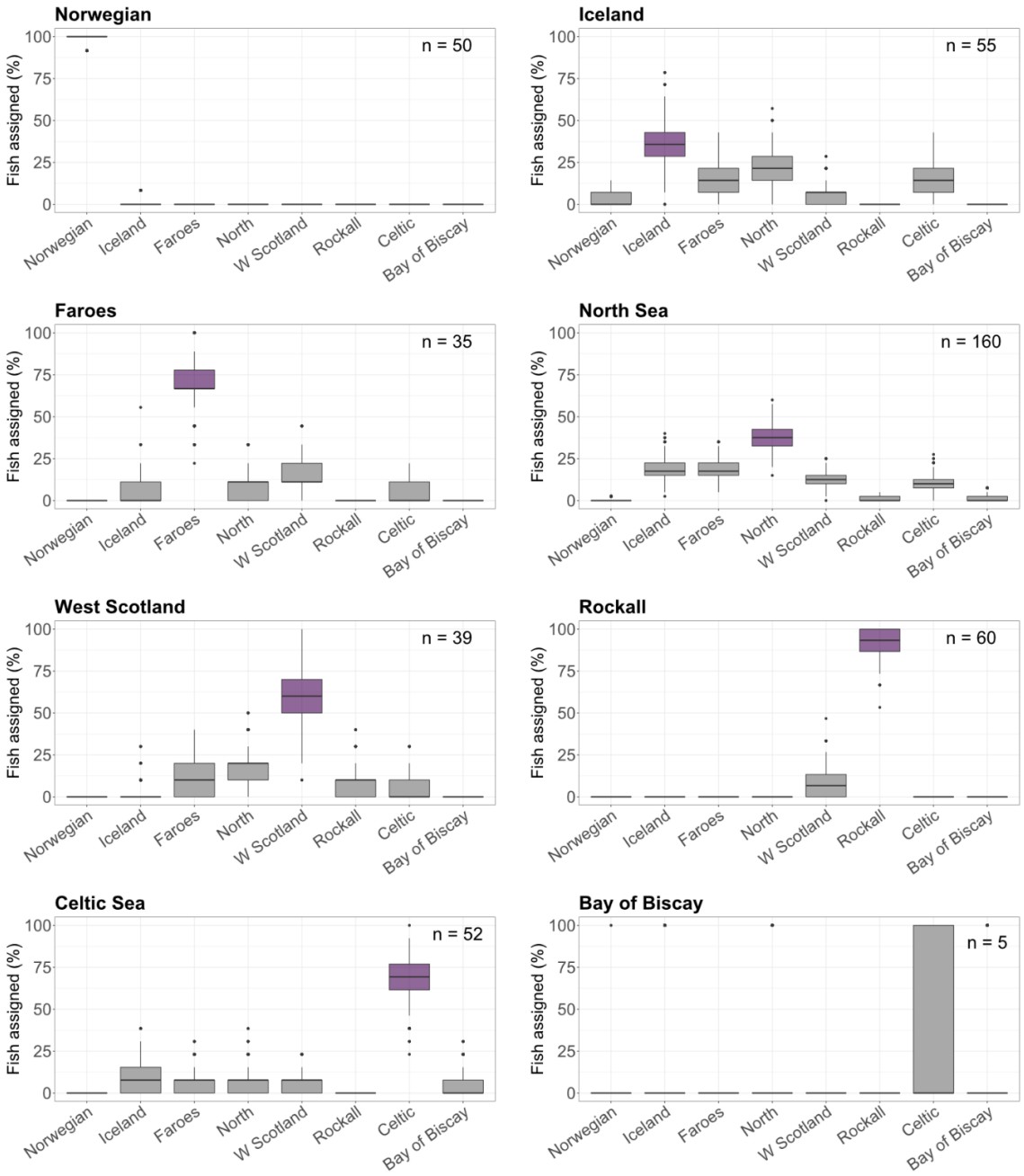


Figure 3.9 Assignment results for haddock using carbon, nitrogen and sulfur stable isotope data, showing the percentage of individuals from each known location assigned to all the possible regions over 1000 repeat simulations. The coloured boxes show the correct regions of origin.

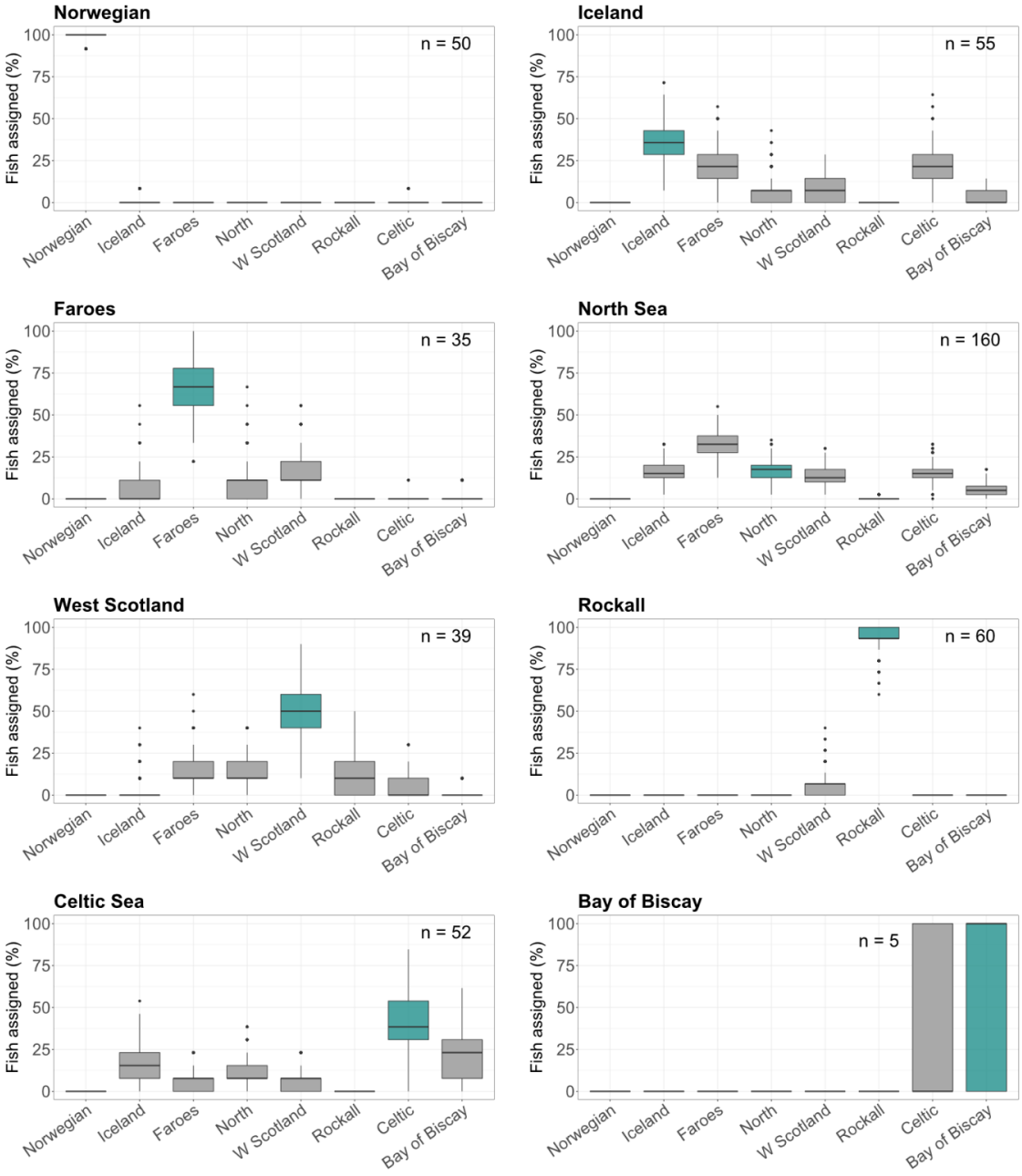


Figure 3.10 Assignment results for haddock using only carbon and nitrogen stable isotope data, showing the percentage of individuals from each known location assigned to all the possible regions over 1000 repeat simulations. The coloured boxes show the correct regions of origin.

Table 3.6 Mean percentage of individual haddock assigned to the correct origin region over 1000 repeat simulations using three isotopes ($\delta^{13}C$, $\delta^{15}N$ and $\delta^{34}S$) and two isotopes ($\delta^{13}C$ and $\delta^{15}N$).

Region	Mean correct assignments (%)	
	CNS	CN
Norwegian Sea	98.1	98.5
Iceland	38.4	36.8
Faroes	70.5	68.7
North Sea	37.6	17.2
West Scotland	57.9	54.6
Rockall	92.2	94.5
Celtic Sea	67.1	40.1
Bay of Biscay	20.7	60.4
Total	**60.3%**	**58.8%**

3.4.2.2 Hake assignment using multivariate technique

The mean assignment accuracy to known origin for all regions was 62% using the multivariate technique was all three stable isotopes (carbon, nitrogen and sulfur), and the success rates for each location are shown in Table 3.7. The assignment accuracies with repeated simulations where different subsets were selected for the training and test datasets are shown in Figure 3.11. Hake from the Mediterranean Sea were the most accurately assigned, with a very high success rate of 99%. The Celtic Sea samples also had a relatively high assignment accuracy, with 73% of individuals correctly assigned. However, hake from the North Sea, West Scotland and Bay of Biscay were more difficult to assign reliably, where the success rates varied from 27% to 58%, and those incorrectly assigned were classified to all other regions in the Atlantic (Figure 3.11).

Excluding sulfur isotope values from the multivariate assignment and using only carbon and nitrogen stable isotopes resulted in an overall success rate of 60.8%, which is very similar to that using all three isotopes (Table 3.7). The results of the repeated assignments with only $\delta^{13}C$ and $\delta^{15}N$ are illustrated in Figure 3.12. For several regions the assignment success was very similar, but for the Bay of Biscay hake the correct assignments increased from 52% to 62% when sulfur was excluded. However, the assignment accuracy for West Scotland decreased in this scenario, from 58% to 47% correct assignments. Therefore, the addition of sulfur isotopes improved the assignment to some locations and reduced the assignment success for others.

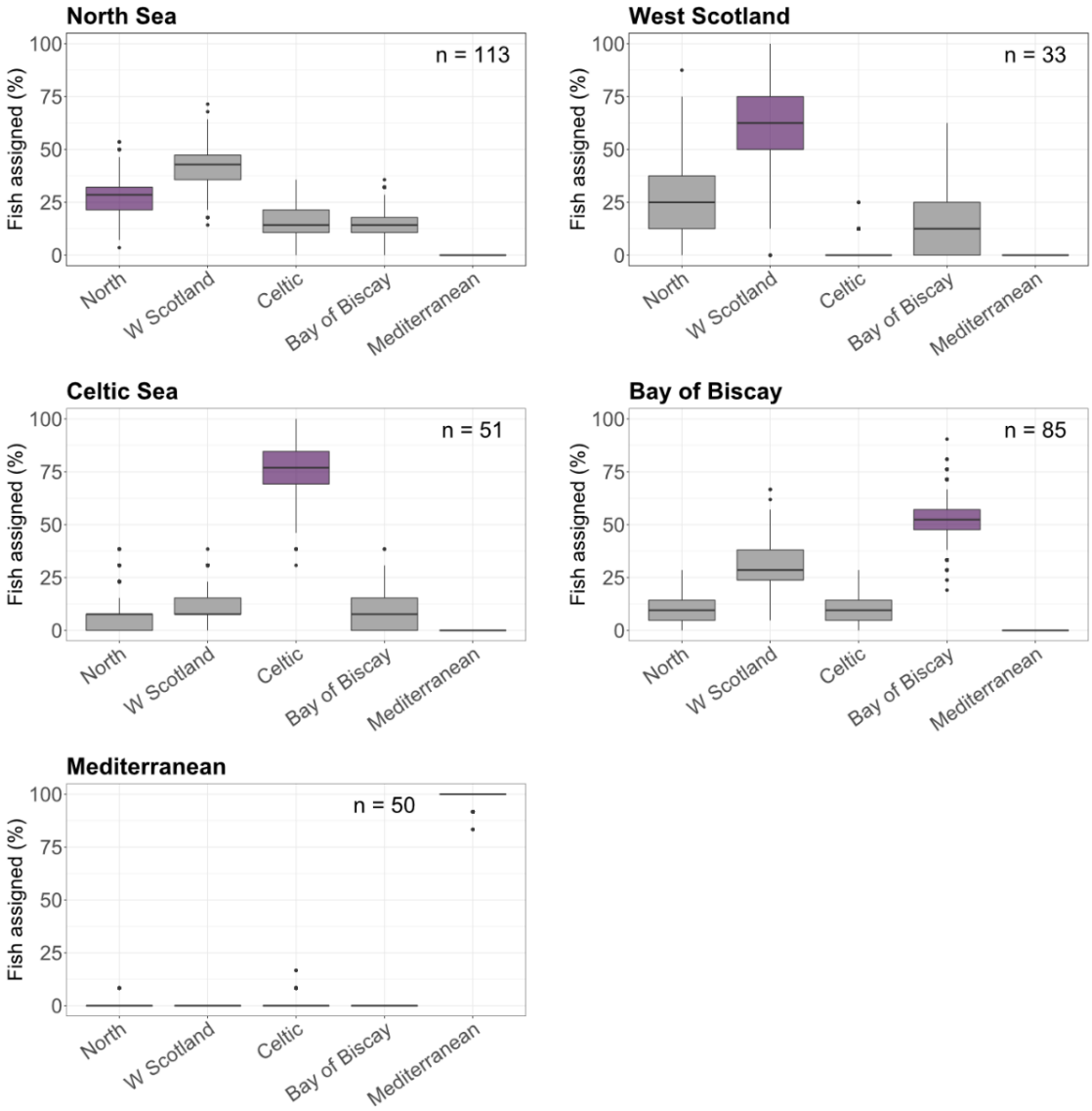


Figure 3.11 Assignment results for hake using carbon, nitrogen and sulfur stable isotope data, showing the percentage of individuals from each known location assigned to all the possible regions over 1000 repeat simulations. The coloured boxes show the correct regions of origin.

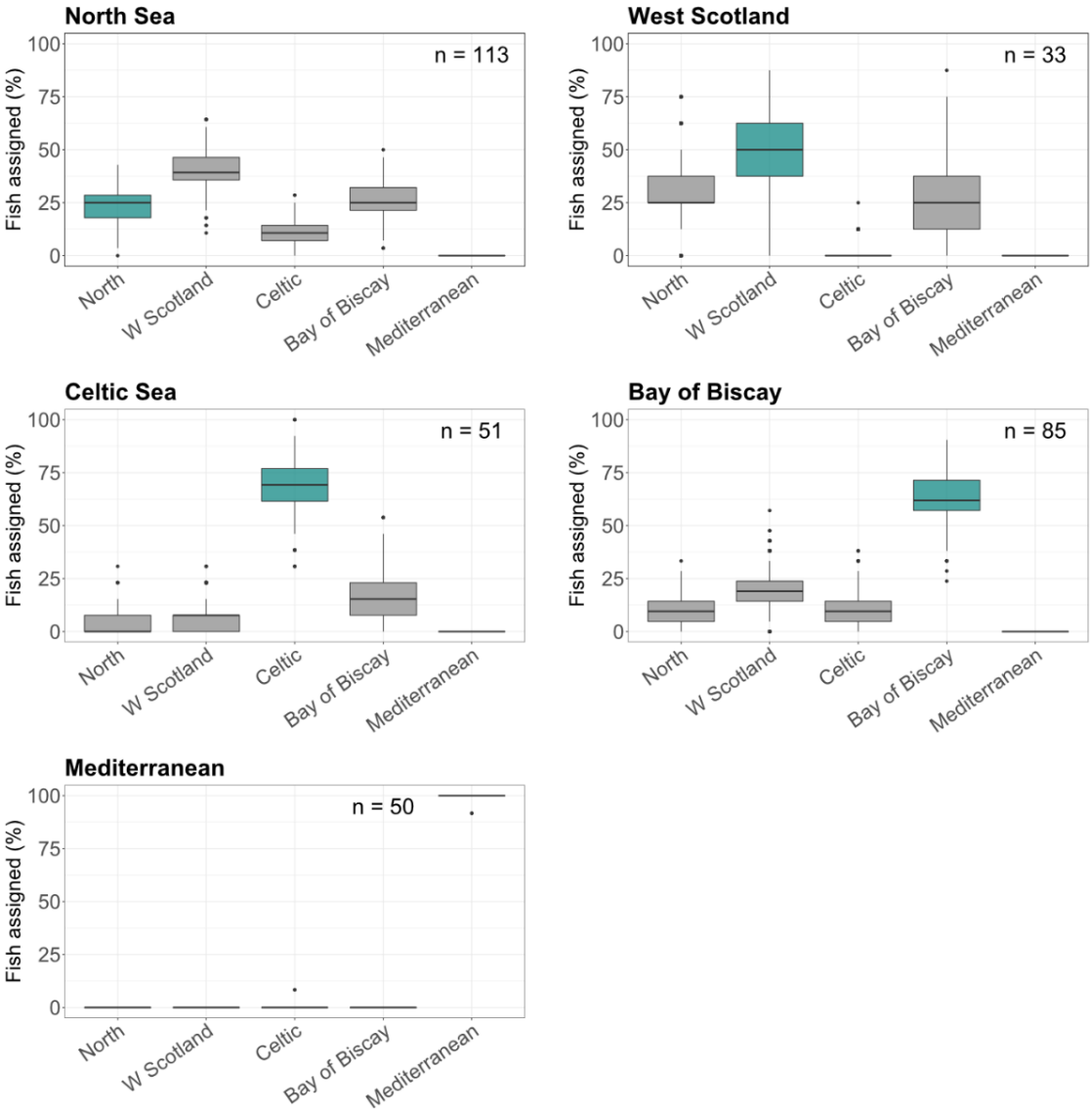


Figure 3.12 Assignment results for hake using only carbon and nitrogen stable isotope data, showing the percentage of individuals from each known location assigned to all the possible regions over 1000 repeat simulations. The coloured boxes show the correct regions of origin.

Table 3.7 Mean percentage of individual hake assigned to the correct origin region over 1000 repeat simulations using three isotopes ($\delta^{13}C$, $\delta^{15}N$ and $\delta^{34}S$) and two isotopes ($\delta^{13}C$ and $\delta^{15}N$).

Region	Mean correct assignments (%)	
	CNS	CN
North Sea	27.4	23.3
West Scotland	58.2	47.0
Celtic Sea	73.3	71.5
Bay of Biscay	52.0	62.0
Mediterranean	99.3	100.0
Total	**62.0%**	**60.8%**

3.4.2.3 Haddock assignment using random forest

Assigning the haddock samples to origin using random forest classification resulted in an overall mean accuracy of 66%. The assignment accuracies achieved for each region are shown in Table 3.8. Haddock from the Norwegian Sea and Rockall could be assigned with high accuracy (92% and 98% respectively), whereas those from the North Sea and West Scotland were more challenging to assign correctly and resulted in lower assignment accuracies of 33% and 31% respectively.

The assignment accuracy using random forest was also investigated with the Bay of Biscay and Celtic Sea samples combined, due to the very low sample numbers collected from the Bay of Biscay and the close proximity of these stations to the Celtic Sea. The results of this assignment are shown in Table 3.9. A mean assignment accuracy of 65% was achieved for all regions, which is very similar to that using random forest with the Bay of Biscay and Celtic Sea separately. However, several regions showed an increased success rate of assignment, for example assignment accuracy for the Icelandic haddock improved from 46% to 56%, and for West Scotland from 31% to 41% accuracy, although these are still relatively low. The combined Bay of Biscay and Celtic Sea group had a success rate of 70%.

Table 3.8 Assignment accuracies achieved for haddock from each sampled location using leave-one-out random forest classification and balanced sample sizes. The correct assignments are shown in bold.

Assigned origin	**True origin region – number assigned**							
	Norwegian	Iceland	Faroes	North Sea	West Scotland	Rockall	Celtic	Bay of Biscay
Norwegian	**46**	0	0	0	0	0	1	0
Iceland	0	**25**	1	29	0	0	3	0
Faroes	0	5	**21**	37	6	0	4	0
North Sea	0	9	5	**53**	7	0	3	0
West Scotland	0	4	5	15	**12**	1	0	0
Rockall	0	0	0	2	10	**59**	0	0
Celtic	3	6	3	12	3	0	**35**	0
Bay of Biscay	1	6	0	12	1	0	6	**5**
Percentage correct	**92%**	**46%**	**60%**	**33%**	**31%**	**98%**	**67%**	**100%**

Table 3.9 Assignment accuracies achieved for haddock from each sampled location, with Celtic Sea and Bay of Biscay merged together as one region, using leave-one-out random forest classification and balanced sample sizes. The correct assignments are shown in bold.

Assigned origin	**True origin region – number assigned**						
	Norwegian	Iceland	Faroes	North Sea	West Scotland	Rockall	Celtic & Bay of Biscay
Norwegian	**47**	0	0	0	0	0	2
Iceland	1	**31**	3	22	2	0	5
Faroes	0	6	**21**	32	6	0	4
North Sea	0	8	6	**63**	8	0	4
West Scotland	0	3	3	25	**16**	3	2
Rockall	0	0	0	1	5	**57**	0
Celtic & Bay of Biscay	2	7	2	17	2	0	**40**
Percentage correct	**94%**	**56%**	**60%**	**39%**	**41%**	**95%**	**70%**

3.4.2.4 Hake assignment using random forest

Random forest classification resulted in an overall assignment accuracy of 72.3% for the hake samples. The assignment accuracies for individual regions ranged from 58% to 100%, as shown in Table 3.10. The Mediterranean hake could be distinguished from all others with very high confidence, and all individuals were assigned correctly (100% accuracy). However, the North Sea had a greater number of incorrect assignments, with individuals being misassigned to West Scotland in particular as well as the other regions in the Atlantic, resulting in a lower assignment accuracy of 58%.

Table 3.10 Assignment accuracies achieved for hake from each sampled location using leave-one-out random forest classification and balanced sample sizes. The correct assignments are shown in bold.

Assigned origin	**True origin region – number assigned**				
	North Sea	West Scotland	Celtic	Bay of Biscay	Mediterranean
North Sea	**66**	7	4	15	0
West Scotland	23	**21**	1	7	0
Celtic	11	1	**38**	8	0
Bay of Biscay	13	4	8	**55**	0
Mediterranean	0	0	0	0	**50**
Percentage correct	**58%**	**64%**	**75%**	**65%**	**100%**

3.4.2.5 Haddock - comparison of assignment methods

Two different assignment techniques were applied to the haddock stable isotope data – the multivariate normal probability distribution technique and random forest classification – since these were found to be the most effective classification techniques for Atlantic cod in Chapter 2. The assignment accuracies obtained using these methods are compared in Table 3.11. Random forest classification gave the highest assignment accuracy overall across all regions (66% compared to 60% using the multivariate method), but for individual regions the results were mixed. Both methods resulted in a high proportion of correct classifications for the Norwegian Sea and Rockall samples, with 92-98% assignment success. The random forest also gave a very high assignment accuracy for the Bay of Biscay (100% correct), compared to only 21% correct using the multivariate analysis. However, caution is needed in interpreting this because the sample size for the Bay of Biscay was very small (n = 5) and a greater number of samples is necessary to draw accurate conclusions. Conversely, samples from other regions, including the Faroes and West Scotland, were less well assigned using the random forest technique and multivariate analysis gave more reliable results (11% and 27% higher assignment success respectively).

Table 3.11 Comparison of assignment accuracy for haddock from each geographic region using multivariate normal probability distributions and random forest classification. For multivariate analysis the mean accuracy over 1000 simulations is shown and the random forest used leave-one-out cross validation.

Location	Assignment accuracy (%)	
	Multivariate analysis	Random forest
Norwegian Sea	98.1	92.0
Iceland	38.4	45.5
Faroes	70.5	60.0
North Sea	37.6	33.1
West Scotland	57.9	30.8
Rockall	92.2	98.3
Celtic Sea	67.1	67.3
Bay of Biscay	20.7	100.0
Mean accuracy	**60.3%**	**65.9%**

3.4.2.6 Hake - comparison of assignment methods

The hake samples were also assigned to origin using both the multivariate analysis technique and random forest classification due to the demonstrated effectiveness of these methods for cod in Chapter 2. The results of this comparison are shown in Table 3.12. Random forest gave the highest mean assignment accuracy of 72%, compared to 62% using the multivariate method. Samples from all five regions could be assigned with a greater accuracy using random forest classification than the multivariate technique. For example, the North Sea had a success rate of 58% using random forest, but only 27% using the multivariate analysis, and the Bay of Biscay hake were assigned with 65% accuracy with random forest compared with 52% using the multivariate method. However, both techniques gave very high success rates of 99-100% correct assignments to the Mediterranean Sea.

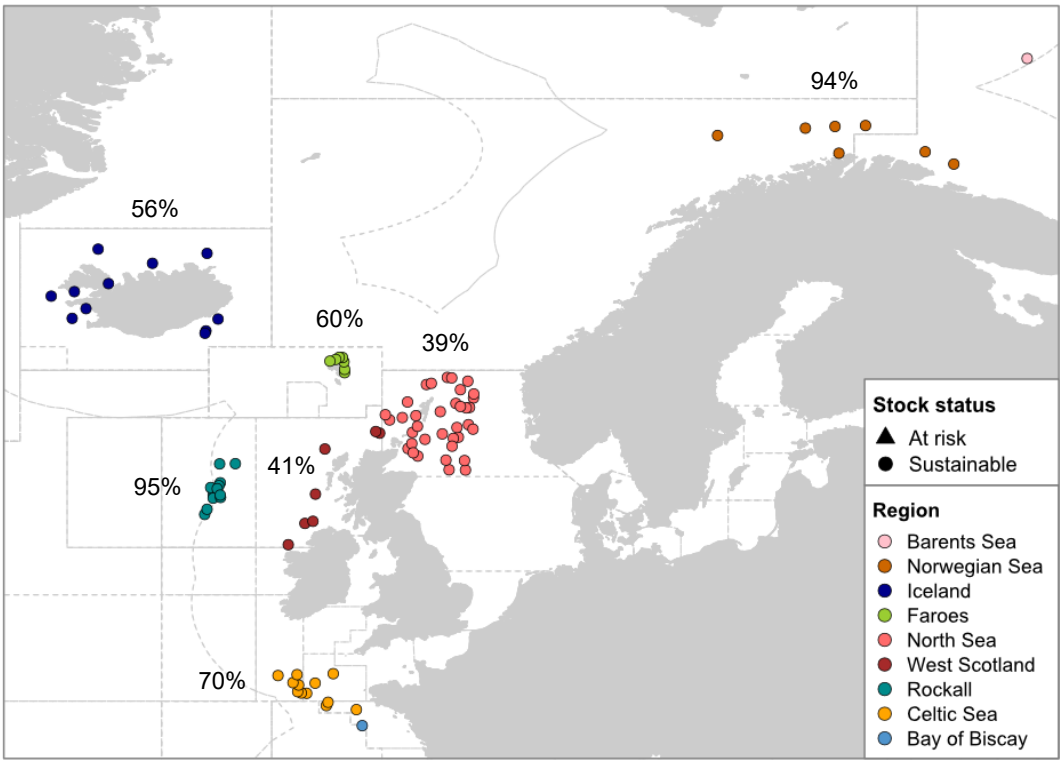
Table 3.12 Comparison of assignment accuracy for hake from each geographic region using multivariate normal probability distributions and random forest classification. For multivariate analysis the mean accuracy over 1000 simulations is shown and the random forest used leave-one-out cross validation.

Location	Assignment accuracy (%)	
	Multivariate analysis	Random forest
North Sea	27.4	58.4
West Scotland	58.2	63.6
Celtic Sea	73.3	74.5
Bay of Biscay	52.0	64.7
Mediterranean	99.3	100.0
Mean accuracy	**62.0%**	**72.3%**

3.4.2.7 Comparison of haddock and hake assignments

Hake could be assigned with greater accuracy across all regions in common (except for the Bay of Biscay, which was combined with the Celtic Sea for haddock due to the very low number of samples). The North Sea samples were significantly better assigned for hake than haddock, since only 39% of haddock were correctly assigned to the North Sea compared with 58% of hake. Hake were also more reliably assigned to West Scotland than haddock, with 64% of hake versus 41% of haddock correctly assigned. However, this higher assignment accuracy for hake may partly be due to the smaller number of possible regions that samples could be assigned to, meaning that there were fewer incorrect options available for assignment. A comparison of the assignment accuracies achieved for each species is shown in Figure 3.13.

a) Haddock



b) European hake

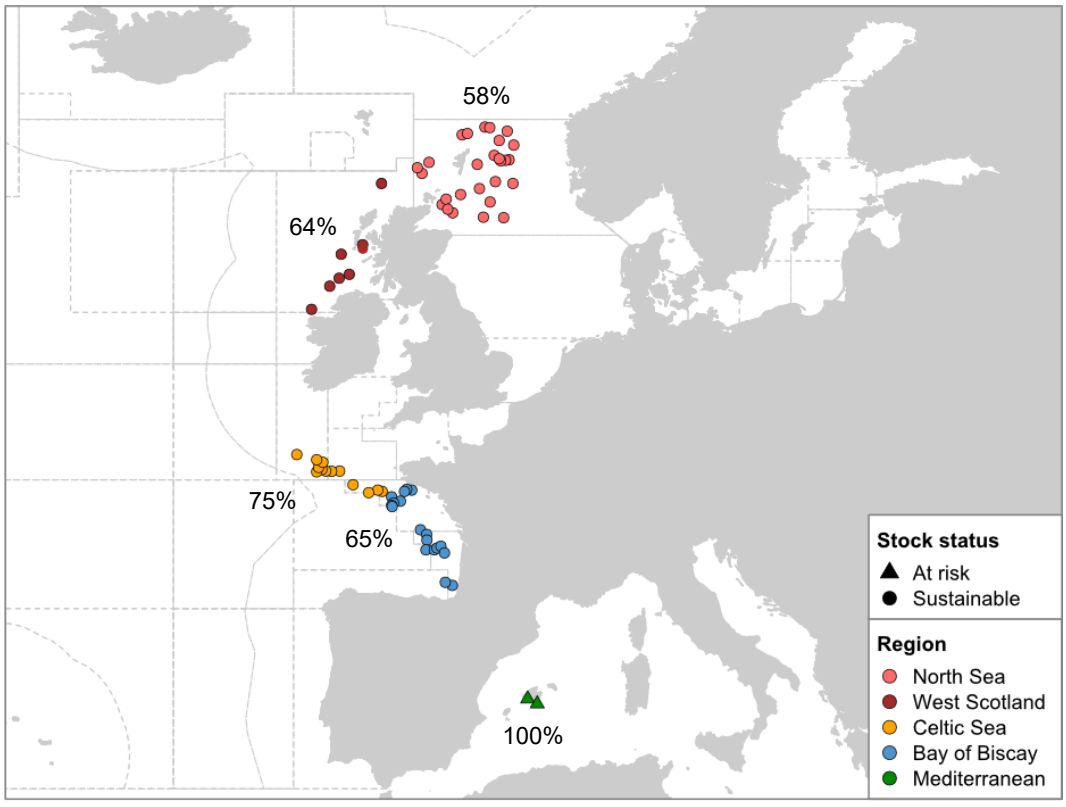


Figure 3.13 Locations of sampling stations for a) haddock and b) European hake within ICES sub area boundaries, and assignment success rates in each region using random forest classification. For haddock, the assignment accuracy for the Celtic Sea and Bay of Biscay samples combined is shown due to the very low sample numbers collected from the Bay of Biscay.

3.4.3 Assigning independent known origin samples to the reference dataset

3.4.3.1 Haddock

The stable isotope composition of independent known origin haddock muscle samples collected in 2002-2017 from the Norwegian Sea and Iceland (samples from both provided by Young's Seafood Ltd.), as well as the North Sea (from Jennings and Cogan (2015)) and the Celtic Sea (data from Jennings and Cogan (2015) and Ifremer (Kopp, 2018), were corrected for the Suess effect and compared to the results from the current study. Figure 3.14 shows the comparison between the $\delta^{13}C$ and $\delta^{15}N$ values from this study (2018) and those of the samples collected between 2002 and 2017. The Norwegian Sea and Iceland samples are both very similar between years, with almost all the 2017 samples being contained within the data ellipses of the 2018 data. The North Sea samples from 2002-2006 have a much wider range of values than those in this study. Although the 2018 samples do show overlap with those from Jennings and Cogan (2015), they are at the higher end of $\delta^{13}C$ and $\delta^{15}N$ values measured in the previous study. The Celtic Sea samples also show good agreement between all sampling years, but the 2010 and 2014 samples have slightly greater variation and so some data points are found outside of the 2018 data ellipse (Figure 3.14).

Data from these independent samples were assigned to their most likely origin based on their carbon and nitrogen isotope composition (after Suess correction), using the known origin samples analysed in this study as a reference dataset, and the results are shown in Table 3.13. The 2017 Norwegian samples could be assigned to the correct origin with high accuracy (100% correct), similar to the self-assignment of the reference samples which resulted in 92-98% accuracy. The Iceland samples were assigned with a moderate accuracy of 50%, compared to 38-46% when assigning the reference samples. However, the Celtic Sea samples had a lower success rate of only 21% compared to 67% with the non-independent samples. A larger proportion (44%) of the Celtic Sea samples were misassigned to Iceland instead. The North Sea samples could be assigned with a very low accuracy of 1.2%, with the majority of samples misassigned to the closest neighbouring regions – West Scotland (53%) and Rockall (29%). A low assignment accuracy was expected for the North Sea based on the self-assignment of the reference samples, which achieved 33-38% accuracy. Therefore, the assignment accuracies were reduced for the North Sea and Celtic Sea compared to the estimates where known origin samples from the current study are self-assigned, whereas the Norwegian Sea and Iceland had similar assignment accuracies using independent or non-independent samples.

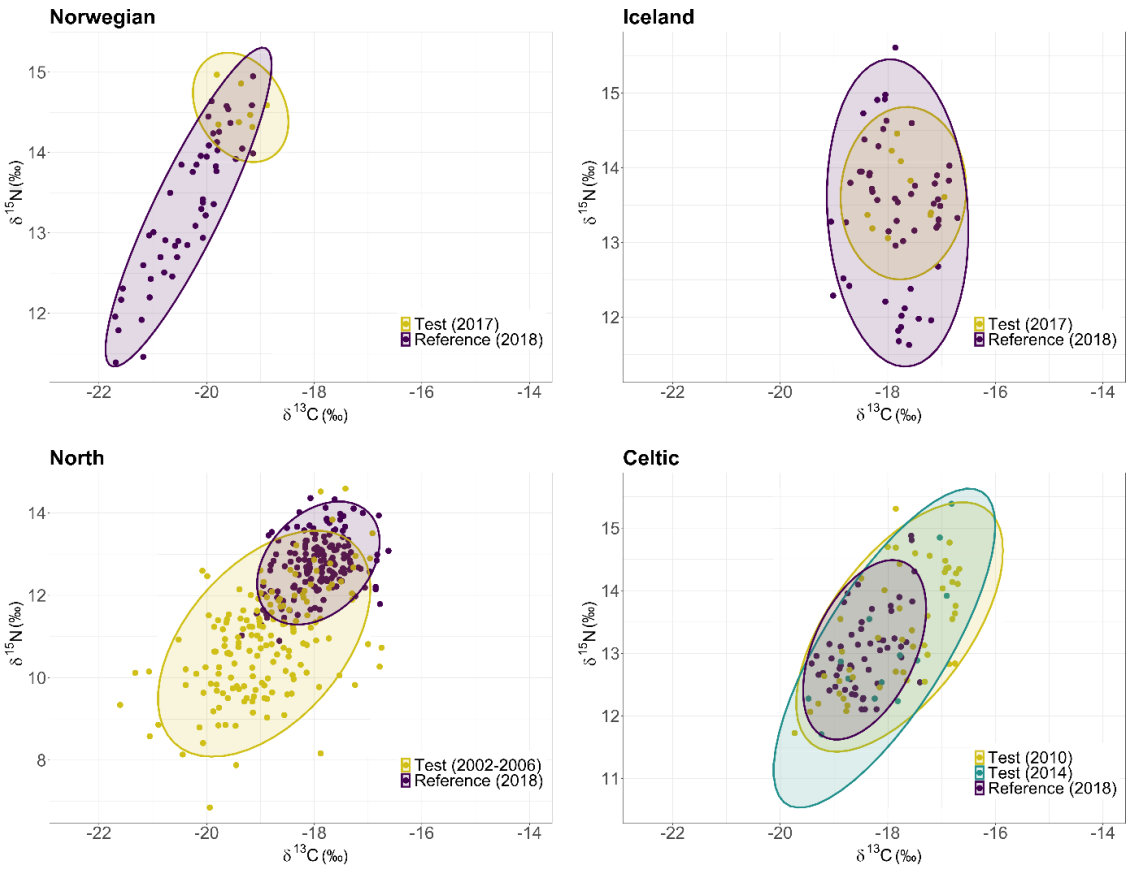


Figure 3.14 Carbon and nitrogen stable isotope values measured in haddock samples collected in this study (reference samples) compared with those collected from previous years and studies (test samples), after applying Suess correction on $\delta^{13}C$ values. Previous data: Norwegian Sea – samples provided by Young's Seafood Ltd.; Iceland – samples provided by Young's Seafood Ltd.; North Sea – data from Jennings and Cogan (2015), Celtic Sea – 2010 data from Jennings and Cogan (2015) and 2014 data provided by Ifremer (Kopp, 2018) from the EVHOE 2014 survey.

Table 3.13 Results of assigning haddock samples from previous studies to the sample data collected in our study as a reference dataset (with a correction for the Suess effect on $\delta^{13}C$ values), using only $\delta^{13}C$ and $\delta^{15}N$. The correct assignments for each region are highlighted in blue. Previous data: Norwegian Sea – samples provided by Young's Seafood Ltd.; Iceland – samples provided by Young's Seafood Ltd; North Sea – data from Jennings and Cogan (2015); Celtic Sea – 2010 data from (Jennings and Cogan, 2015) and 2014 data provided by Ifremer (Kopp, 2018) from the EVHOE 2014 survey.

True origin region	Assigned region	Haddock assigned (count)	Haddock assigned (%)
Norwegian Sea	**Norwegian Sea**	**7**	**100.0**
	Iceland	0	0.0
	Faroes	0	0.0
	North Sea	0	0.0
	West Scotland	0	0.0
	Rockall	0	0.0
	Celtic Sea	0	0.0
	Bay of Biscay	0	0.0
Iceland	Norwegian Sea	0	0.0
	Iceland	**5**	**50.0**
	Faroes	2	20.0
	North Sea	1	10.0
	West Scotland	0	0.0
	Rockall	0	0.0
	Celtic Sea	1	10.0
	Bay of Biscay	1	10.0
North Sea	Norwegian Sea	2	1.2
	Iceland	6	9.5
	Faroes	11	17.5
	North Sea	**2**	**1.2**
	West Scotland	86	52.8
	Rockall	47	28.8
	Celtic Sea	8	4.9
	Bay of Biscay	1	0.6
Celtic Sea	Norwegian Sea	0	0.0
	Iceland	27	43.5
	Faroes	11	17.7
	North Sea	5	8.1
	West Scotland	2	3.2
	Rockall	0	0.0
	Celtic Sea	**13**	**21.0**
	Bay of Biscay	4	6.5

3.4.3.2 Hake

Stable isotope data from independent hake samples caught in the Celtic Sea in 2010 (Jennings et al., 2008) and in 2014 (data from Ifremer (Kopp, 2018) using samples collected during the EVHOE 2014 survey). The $\delta^{13}C$ and $\delta^{15}N$ values of these samples are compared in Figure 3.15 (after Suess correction of the $\delta^{13}C$ values), which illustrates that the Celtic Sea samples from both 2010 and 2014 match well with the data in this study (2018), although the 2010 samples show a wider variation in $\delta^{15}N$ values than those from the other years and so some datapoints fall outside of the 2018 data ellipse.

The independent samples were assigned to their most likely origin using the $\delta^{13}C$ and $\delta^{15}N$ values (after correction for the Suess effect) using the data from this study as a reference dataset (Table 3.14). This resulted in a very high success rate of 93% for the 2014 samples, which is significantly higher than the accuracy achieved using self-assignment of the data from the current study (73-75% accuracy), whereas a lower assignment accuracy of 59% was obtained using the 2010 samples. A relatively large proportion (21%) of the 2010 samples were assigned to the region in closest proximity - the Bay of Biscay.

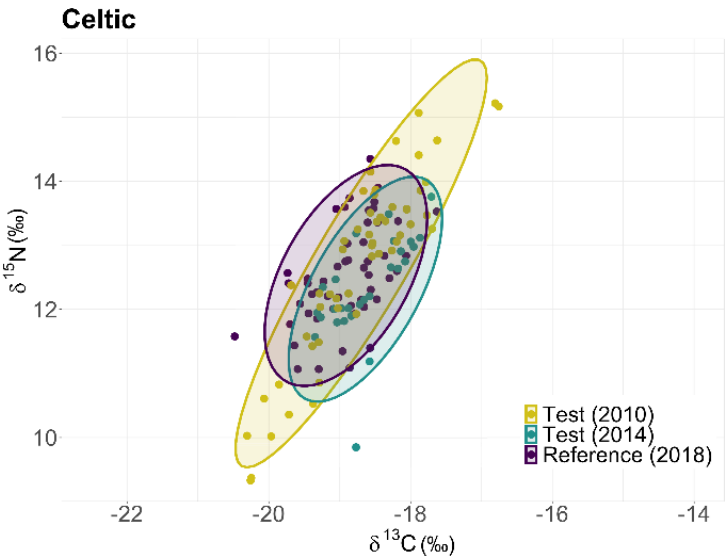


Figure 3.15 Carbon and nitrogen stable isotope values measured in hake samples collected in this study (reference samples) compared with data from Jennings and Cogan (2015) and from Ifremer (Kopp, 2018) from the EVHOE 2014 survey, after applying Suess correction on $\delta^{13}C$ values (test samples).

Table 3.14 Results of assigning hake samples from other studies to the sample data collected in our study as a reference dataset (with a correction for the Suess effect on $\delta^{13}C$ values), using $\delta^{13}C$ and $\delta^{15}N$. The correct assignments for each region are highlighted in blue. Celtic Sea 2010 data was taken from Jennings and Cogan (2015) and 2014 data was provided by Ifremer (Kopp, 2018) from samples collected during the EVHOE 2014 survey.

True origin region	Assigned region	Hake assigned (count)	Hake assigned (%)
Celtic Sea (Ifremer, 2014)	North Sea	0	0.0
	West Scotland	0	0.0
	Celtic Sea	**26**	**93.0**
	Bay of Biscay	1	4.0
	Mediterranean	1	4.0
Celtic Sea (Jennings, 2010)	North Sea	2	3.4
	West Scotland	10	17.2
	Celtic Sea	**34**	**58.6**
	Bay of Biscay	12	20.7
	Mediterranean	0	0.0

3.4.4 Isotopic variation within discrete regions

The isotopic variation within discrete regions was also investigated, since in this study we have collected samples from different stations across the sampled regions. This could allow the use of continuous-surface methods to infer the origin of fish at a higher spatial resolution. Haddock samples were collected from stations all around Iceland except for some areas in the east and south, so the spatial variation in this region could be investigated. In the North Sea, we have collected sufficient samples of haddock and hake to build isoscape models via interpolation between data points.

3.4.4.1 Spatial variation around Iceland

Haddock from Iceland show relatively large variation in $\delta^{13}C$, $\delta^{15}N$ and $\delta^{34}S$ values compared to some other sampled regions (Figure 3.7 earlier on), with a range of 2.4‰, 4.0‰ and 5.4‰ among individuals from Iceland respectively. The spatial variation in haddock isotopic composition around Iceland is displayed in Figure 3.16, showing the mean value measured at each station. There is a pattern of higher $\delta^{13}C$ values to the south and west of Iceland, with a maximum mean station value of -17.0‰, and lower values to the north where the minimum is a mean of -18.7‰. The $\delta^{15}N$ values are higher along the northern coast of Iceland, with intermediate values to the west and the lowest values present in the southeast. The mean station isotope values for nitrogen range from 11.9‰ to 14.9‰. The pattern of variation is less clear in the $\delta^{34}S$ values, since there are higher and lower values on both the north and south coasts, with a minimum station mean of 15.8‰ and a maximum of 18.4‰.

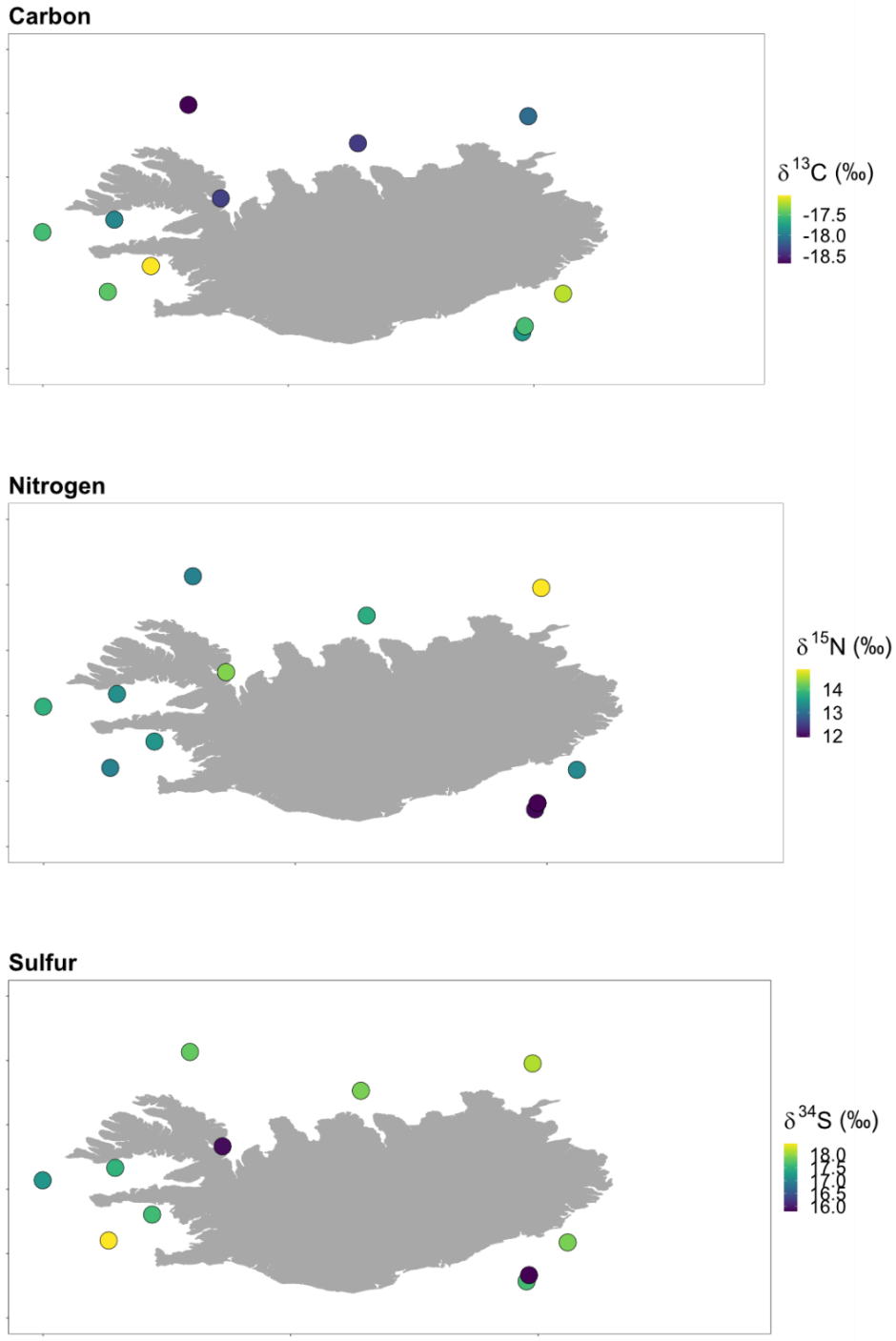


Figure 3.16 Spatial variation in $\delta^{13}C$, $\delta^{15}N$ and $\delta^{34}S$ values measured in haddock from Icelandic waters, showing the mean value at each station.

3.4.4.2 North Sea isoscape

3.4.4.2.1 Haddock

Haddock were sampled from 35 different stations across the northern North Sea, which allows the variation within this region to be investigated. The mean station-level isotope values ranged from -19.0‰ to -17.0‰ for $\delta^{13}C$, from 11.5‰ to 13.7‰ for $\delta^{15}N$ and from 16.3‰ to 18.6‰ for $\delta^{34}S$. Therefore, there were variations of 2‰, 2.2‰ and 2.3‰ for $\delta^{13}C$, $\delta^{15}N$ and $\delta^{34}S$ across the sampled area. Carbon, nitrogen and sulfur isoscapes were created for this region using interpolation by kriging between the sampled stations, and the associated variances are also shown (Figure 3.17). There is a pattern of lower values to the north and west of the isoscape in both $\delta^{13}C$ and $\delta^{15}N$, with higher values to the south and east. The sulfur isoscape shows the inverse of this pattern, with lower values of $\delta^{34}S$ across the south and east and higher values to the north and west. Therefore, for all three isotopes there is a similar pattern where a boundary area occurs at the Shetland Islands, where the isotopic values vary to the southeast and northwest of these islands, in a very similar way to the Atlantic cod isoscapes in Chapter 2.

The variances for all three stable isotopes are similar, with increasing variances at the edges of the isoscapes (Figure 3.17). This is because these areas are outside of the area with measured data points and so interpolation cannot be carried out reliably.

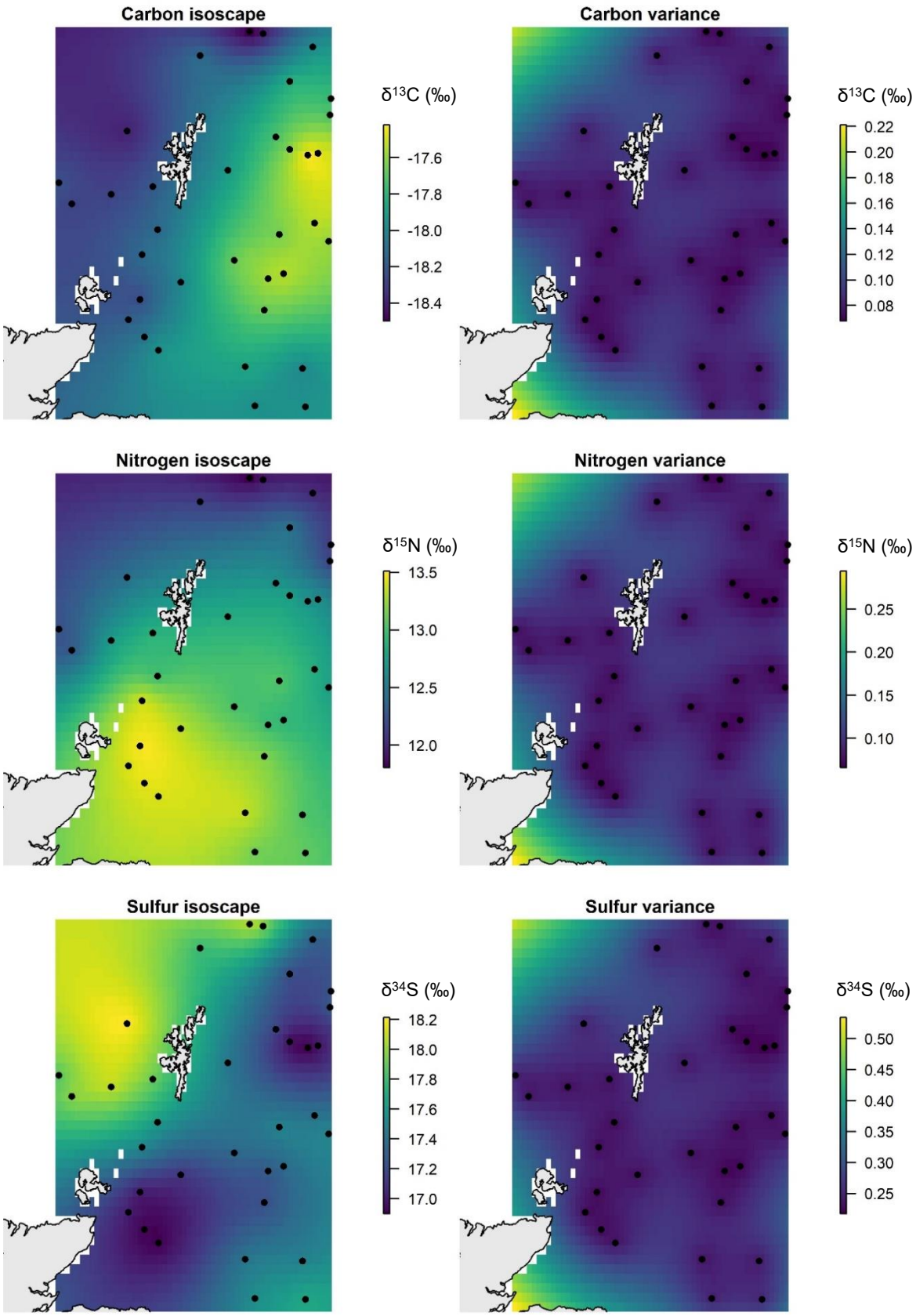


Figure 3.17 Carbon, nitrogen and sulfur isoscapes for the North Sea sampling area derived from haddock muscle tissue (left), with the sampled stations indicated by black circles. The associated spatial variances for each isoscape are shown on the right.

3.4.4.2.2 Hake

European hake were collected from a total of 27 stations in the North Sea as part of this study. The mean station isotopic values varied between -19.9‰ to -18.9‰ for $\delta^{13}C$, 12.7‰ and 14.2‰ for $\delta^{15}N$ and 17.1‰ and 18.7‰ for $\delta^{34}S$, meaning there was a variation of 1‰, 1.5‰ and 1.6‰ among stations for carbon, nitrogen and sulfur stable isotopes respectively. The carbon, nitrogen and sulfur isoscapes for hake, created using interpolation by kriging between the sampled stations, and their associated variances are shown in Figure 3.18. The carbon isoscape does not show much variation across the region. The nitrogen isoscape for hake is similar to that of haddock (Figure 3.17) and cod (Chapter 2), since there are higher $\delta^{15}N$ values in the southeast part of the isoscape and lower $\delta^{15}N$ values to the northwest. The highest $\delta^{34}S$ values are found in the eastern area of the isoscape, and the lowest values in the shallow water to the southwest, close to the Scottish mainland. For hake, the nitrogen isoscape is the only one of the three isoscapes to show the pattern of differing values to the northwest and southeast of the Shetland Islands that was seen in the haddock and cod isoscapes. Such a clear trend is not observed in the carbon and sulfur isoscapes for hake.

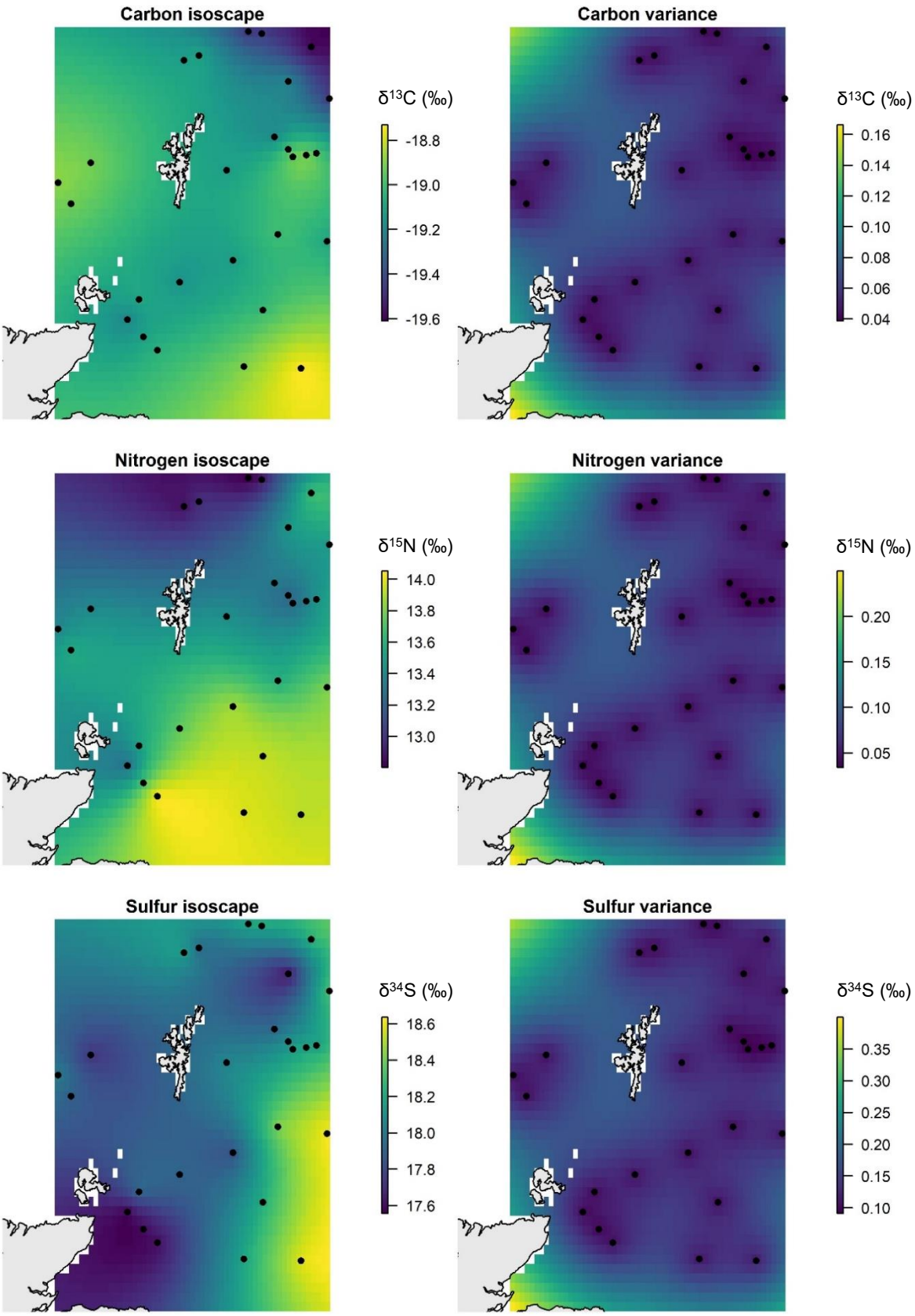


Figure 3.18 Carbon, nitrogen and sulfur isoscapes for the North Sea sampling area derived from hake muscle tissue (left), with the sampled stations indicated by black circles. The associated spatial variances for each isoscape are shown on the right.

3.4.4.2.3 Isoscape variograms

The sample variograms and associated variogram models for kriging interpolation for both species are shown in Figures 3.19 and 3.20. A linear model fitted the variograms best for all three isotopes, although for the sulfur variogram for haddock an exponential model also fitted reasonably well. I used the linear model for consistency, since without additional reference data it cannot be established which model is more accurate. The variogram models for each isotope have a similar gradient between species, suggesting that isotope ratios in the muscle tissue vary over similar spatial scales within this region. However, sulfur isotopes show the most spatial structure over the sampled area and carbon isotopes show the least. Sulfur isotope values also demonstrated more variance than carbon and nitrogen at all distances, since for stations 100km apart, the semivariance of the model is around 0.08-0.15‰ for carbon and nitrogen compared to 0.2-0.3‰ for sulfur in haddock and hake. The scatter around the model is relatively small for haddock in the carbon and nitrogen variogram, and similarly for hake with carbon, but there is much greater scatter on the sulfur variograms for both species as well as in the nitrogen variogram for hake. This suggests that the error is greater in these models with greater variance and therefore any spatial structure predicted in the isoscapes with greater variance is less reliable. The sulfur variogram for hake is particularly variable, to the extent that a horizontal model could have been fitted, so it is likely that there is no spatial structure in sulfur isotope values for hake across the sampled North Sea area.

Compared to Atlantic cod (Chapter 2), haddock shows similar spatial structure in carbon isotope values, whereas nitrogen and sulfur isotope values vary over smaller spatial scales (as indicated by the steeper gradient of the variogram model). Hake do not appear to show such clear spatial trends in isotope values as cod or haddock, due to the large scatter on the variograms (Figure 3.20).

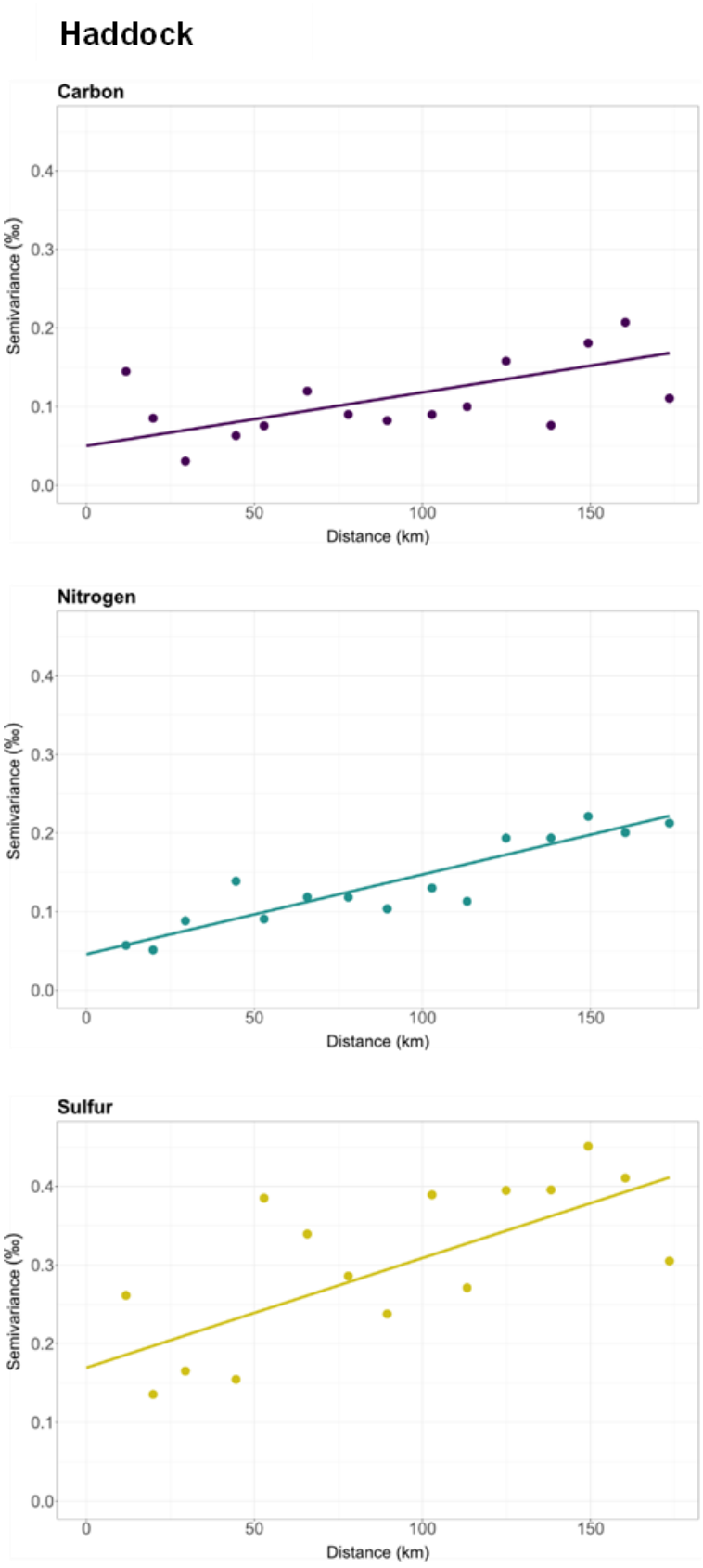


Figure 3.19 Variograms and the associated variogram models for carbon, nitrogen and sulfur stable isotope ratios using haddock from the North Sea, illustrating the variance in isotope values with distance between the sampled stations.

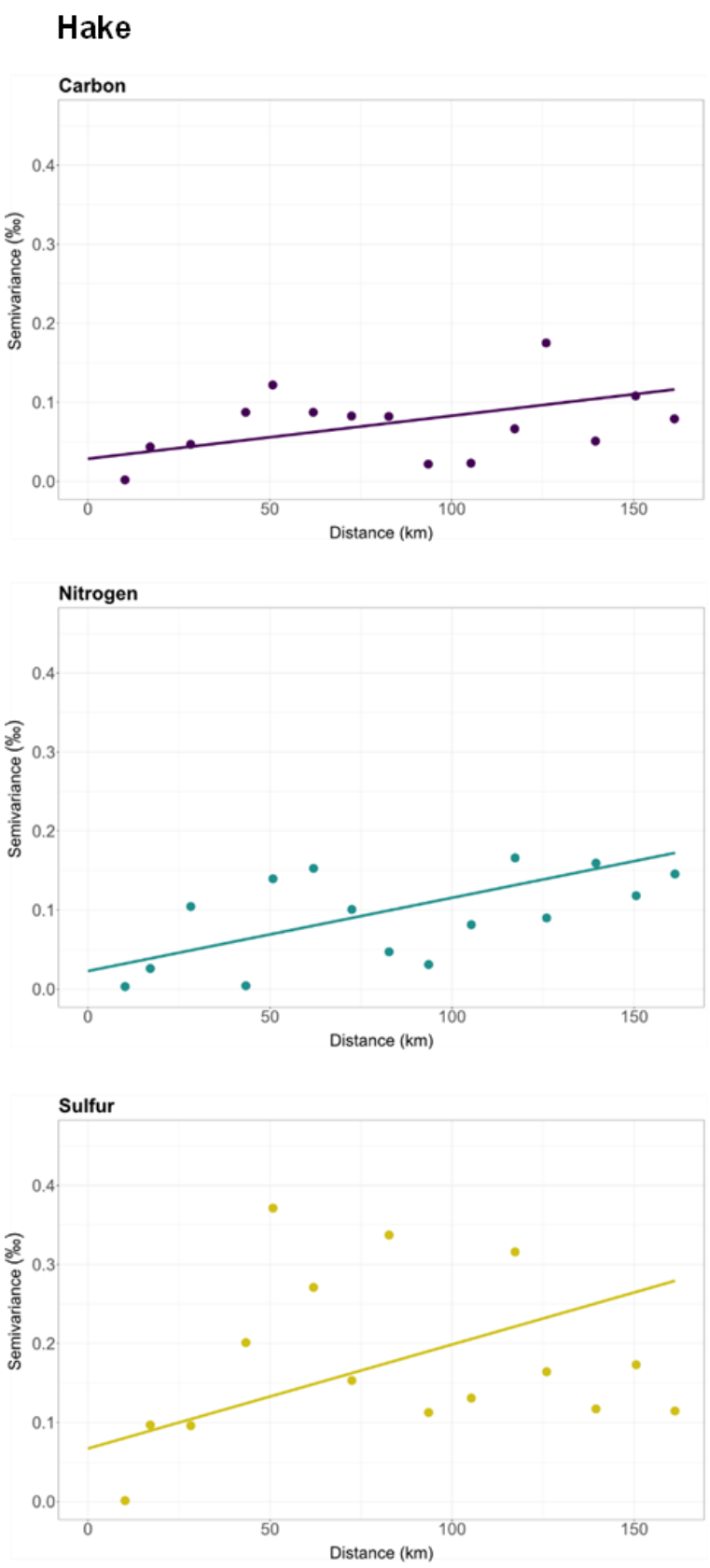


Figure 3.20 Variograms and the associated variogram models for carbon, nitrogen and sulfur stable isotope ratios using hake from the North Sea, illustrating the variance in isotope values with distance between the sampled stations.

3.5 Discussion

This study demonstrates that the stable isotope composition of both haddock and European hake muscle tissue varies according to their catch region, and therefore stable isotope analysis may have potential use as a forensic tool to verify the spatial origin of haddock and hake products from certain regions. However, some fishery regions were isotopically similar and so isotopic approaches in isolation have limitations. Identifying regions of isotopic distinctiveness allows a priori assessment of whether isotope methods are likely to be appropriate for specific forensic or traceability requirements. An important finding from this study is that, in general, isotopic differences were conserved across whitefish species so that spatial differences inferred from reference data for one species are reliable guides to expected differences in other ecologically similar taxa, at large regional scales at least. This study also contributes a large dataset of stable isotope ratios measured in haddock and European hake muscle tissue from a wide range of regions, and is one of the few studies to have investigated the use of sulfur stable isotope ratios for traceability of marine fish. The discrete regions sampled generally correspond to the ICES subareas within FAO Major Fishing Area 27, and therefore have direct relevance to fisheries management of these species within the Northeast Atlantic.

Haddock from the Norwegian Sea and Rockall could be differentiated from fish caught from other fishery areas with high confidence (98% accuracy) using one of the two assignment methods (either multivariate analysis or random forest classification). The other regions investigated for haddock overlapped more in their isotopic compositions and therefore were not able to be assigned to origin with such a high level of accuracy. Isotopic assignment accuracy rates ranged from 38% for Iceland and the North Sea to 71% for the Faroes. The Bay of Biscay samples showed a widely differing success rate depending on the assignment method, with 21% accuracy using the multivariate method and 100% accuracy using random forest classification. This is likely to be because the random forest model used balanced sample sizes, unlike the multivariate technique, and therefore indicates that there is good potential for haddock from the Bay of Biscay to be accurately distinguished when a greater sample size is used. When the Bay of Biscay and Celtic Sea samples were combined, the assignment accuracy was improved compared to the that achieved for the Celtic Sea alone. This is likely to be because when treated as separate regions, some Celtic Sea individuals were assigned to the nearest geographic region, the Bay of Biscay, due to their close proximity and therefore isotopic similarity and therefore combining these two regions resulted in an increased accuracy of assignment to origin. Assignment to the nearest neighbouring region also occurred for the Rockall haddock, where a certain proportion were assigned to West Scotland instead.

Overall, haddock and hake could be assigned to origin with a similar level of accuracy (60-66% and 62-72% respectively). For both haddock and hake there are one or two very distinctive regions (Norwegian Sea and Rockall for haddock, and the Mediterranean for hake) which increase the overall classification success, but the central Northeast Atlantic fishery areas overlap to greater or lesser degrees. This is also true for cod, where those from regions around the UK shelf seas and the Faroes overlap in their isotopic composition to a certain extent whereas cod from the Northeast Arctic and Baltic Sea are very distinct, resulting in a relatively high assignment accuracy overall (Chapter 2). For regions that were investigated for both haddock and hake, including the North Sea, West Scotland and Celtic Sea, hake samples could be assigned more reliably than haddock. This may be partly due to the smaller number of possible assignment regions for hake, which is further amplified because the Mediterranean hake have a very different isotopic composition to the other sampled regions, leading to fewer incorrect classifications to this region. The Norwegian and Rockall haddock are also isotopically distinct (92-98% assignment accuracy), but not quite to the same extent as the Mediterranean hake (99-100% correct classifications).

Comparing the distinguishing power of stable isotopes for tracing provenance of haddock and hake with that for Atlantic cod from Chapter 2 shows that cod has the highest assignment accuracy overall, although only slightly. Furthermore, cod could also be assigned with higher accuracy than haddock for some individual regions, whereas hake achieved similar or even slightly higher success rates in certain cases. Samples from the Norwegian Sea were assigned with high confidence for both cod and haddock (up to 99%). However, cod from Iceland and the Faroes had a higher assignment accuracy than haddock (70% versus 38% for Iceland, and 86% versus 71% for the Faroes). All three species were sampled from the North Sea, however hake were slightly more isotopically distinct (58% correctly assigned), while for the other two species, North Sea samples could not be assigned effectively using stable isotope methods (classification success of 39% for haddock and 52% for cod). This finding illustrates that the ability to discriminate successfully depends on the isotopic composition of the region compared to that in other sampled areas. For cod and haddock, we identified a larger number of potential discrete fishery sources within the central Northeast Atlantic compared to hake, which is likely to make it more challenging to assign fish correctly.

3.5.1 Comparison of assignment methods

Of the two assignment methods used, random forest classification gave slightly higher assignment accuracies overall for both haddock and hake, although differences for individual regions were generally small. For hake, the only region with a significant difference in assignment accuracy was the North Sea, which showed an improvement of

31% using random forest rather than the multivariate method. Haddock from the Bay of Biscay were also assigned to origin with much greater accuracy using random forest classification (improving from 20% to 100% correct assignments), but this region had a very small sample size. Another difference was seen in the assignment accuracy for haddock from West Scotland, which was reduced from 58% using the multivariate method to 31% using random forest classification. However, both assignment methods gave similar results overall except for regions with very small sample sizes, which gives confidence that the inferences are robust to the data analysis method used.

This study used the stable isotope ratios of sulfur in addition to carbon and nitrogen, which has rarely been used in marine fish origin traceability studies. However, it was found to be a useful tracer in distinguishing among catch locations of scampi (*Nephrops norvegicus*) (Slesser and Trueman, 2021) due to anoxic conditions in seabed sediments creating spatial differences in $\delta^{34}S$. Spatial variations in sulfur stable isotope values were also documented in jellyfish tissues by St. John Glew et al. (2019), demonstrating that $\delta^{34}S$ values also vary in pelagic ecosystems in the Northeast Atlantic. In this study, sulfur stable isotopes contributed additional distinguishing power to the assignments to origin for both haddock and hake from several regions. The largest improvements were in the North Sea and Celtic Sea assignments for haddock, which increased by 20% and 27% respectively with the inclusion of sulfur isotope values, and in the West Scotland assignments for hake, which increased by 11%. Conversely, the Bay of Biscay assignments were reduced significantly for both haddock (40%) and hake (10%) with the addition of sulfur, suggesting that sulfur isotopes are of less importance for distinguishing fish from this region.

These results indicate that $\delta^{34}S$ values in muscle tissue of haddock and hake vary regionally, although the mechanisms underpinning these variations are not understood. Therefore, sulfur stable isotopes were shown to be valuable for origin traceability in haddock and hake, in contrast to Atlantic cod (Chapter 2) where sulfur isotopes had low distinguishing ability in general, although the usefulness of sulfur isotopes may depend on the exact region of interest.

3.5.2 Assignment of independent known origin samples

The assignments using independent known origin samples from the Norwegian Sea, Iceland, North Sea and Celtic Sea for haddock, and those from the Celtic Sea for hake all showed good agreement with the results of this study. Assigning these independent samples to their most likely origin using data from the current study as a reference dataset resulted in varying success rates. Haddock from the Norwegian Sea in the previous dataset could be assigned with a similar accuracy as with self-assignment of the reference data. Hake collected from the Celtic Sea in 2014 (Ifremer dataset) could be assigned with an even higher accuracy of 93% compared to 67% using non-independent samples

(assigning a subset of the dataset from this study to itself). However, for the North Sea and Celtic Sea the accuracy was much lower than the accuracies obtained using non-independent samples. The North Sea had a very low assignment accuracy of 1.2%, and the Celtic Sea also had a much lower assignment accuracy using the independent samples (21% compared with 67% for haddock, and 59% compared to 74% for hake). This lower assignment accuracy is expected when assigning the independent samples, since the datasets were collected in different years (2002-2017 for the test data compared to 2018 for the reference data) and from different sampling locations within each region, and were analysed in different laboratories, all of which introduce additional variance. The assignment accuracies from non-independent reference datasets are best case estimates, so the independent test data is likely to be assigned less accurately. The similar assignment rates achieved for certain locations are therefore encouraging.

There are differences in the exact locations of sampling within each region which may be contributing to the variation between datasets. The samples in Jennings and Cogan (2015) from the North Sea were collected from a similar location to those here, however the stations in our study cover a much wider geographic range. The locations of the Celtic Sea stations also show some geographic overlap, but the stations are generally situated in the more southerly part of the Celtic Sea in this study whereas those in the previous datasets are closer to the UK coast slightly further north, which may have created additional variation between the datasets. When building a reference dataset, it is important to collect samples from across the whole of each region to better encompass the true variation, and therefore to obtain the most reliable assignment results.

Temporal variation in the spatial distribution of $\delta^{13}C$ and $\delta^{15}N$ occurs over seasonal and yearly time scales due to changes in the rate of hydrodynamic and biogeochemical processes with time (Magozzi et al., 2017). This is then transferred up the food web and incorporated into the tissues of consumers. Certain areas may be more susceptible to the temporal effects, because areas with fluctuating environmental conditions are likely to show greater temporal variation, whereas stable isotope ratios in other regions, are relatively stable over time (MacKenzie et al., 2014). This means that collecting samples for a reference collection over the same time period as that of the unknown test samples is likely to result in higher assignment accuracies. This may partly explain the reduced assignment accuracies in the North and Celtic Seas, since sample collection for the independent datasets took place up to 16 years prior to the 2018 reference data for the North Sea and up to 8 years prior for the Celtic Sea. For hake, the 2014 independent dataset could be assigned with considerably greater accuracy than the 2010 dataset. This explanation is also supported by the relatively high assignment accuracies obtained using independent datasets for the Norwegian Sea and Iceland, which were only collected one year before the reference data. Although ideally reference collections should be made each year, in

practice it would likely be necessary to compare across years. This test case showing broad applicability of reference data to samples collected in different years is a valuable case study replicating what might be required in real life traceability cases.

Using independent samples from other datasets to estimate the assignment accuracy is crucial for determining the true effectiveness of the technique in an origin traceability context. The assignment accuracy and precision when assigning other datasets to data from the current study is likely to be lower than assigning a subset of the same data to itself, since a statistical artefact may occur when non-independent samples are used. This would result in a higher success rate than would be obtained if the technique was used in real world traceability situations.

3.5.3 Comparison of stable isotope analysis with genetic techniques

The use of genetic techniques has also been investigated for several marine fish species (Nielsen et al., 2012b). This relies on populations being reproductively isolated over multigenerational timescales, so individuals from certain regions may be clearly distinguished while others may not have sufficient genetic differences to allow discrimination. Stable isotopes indicate the location where an individual was last foraging, integrated over the tissue turnover time, where regions are isotopically distinct. Therefore, genetic tracers reveal the spatial histories of populations whereas stable isotopes identify an individual's location at the point of capture.

Novel genetic methods such as single-nucleotide polymorphism (SNP) markers have recently been utilised to investigate the genetic structuring of haddock populations and has discovered three main genetic clusters, consisting of a Northwest Atlantic, Northeast Arctic and Northeast Atlantic cluster, as well as a genetically distinct fjord population (Berg et al., 2021). This suggests that genetic techniques could be used to differentiate among these three main clusters, but within each cluster it may be challenging to trace haddock to origin due to the relatively high level of genetic mixing. The results of the current study demonstrate that stable isotopes can also reliably distinguish Northeast Arctic haddock from those caught in the Northeast Atlantic (92-98% assignment accuracy).

Nielsen et al. (2012b) found that European hake from the Mediterranean and Atlantic Ocean could be distinguished using SNPs (single-nucleotide polymorphisms) and were assigned to basin of origin with an accuracy of 98%. In the current study, we were able to correctly classify hake to the Mediterranean Sea with a very high level of accuracy of 99-100%, matching that achieved using genetic techniques. However, European hake have been found to show high levels of connectivity among regions within the Atlantic and so have low genetic differentiation (Milano et al., 2014, Pita et al., 2011), although the North Sea and northern Portugal populations show some separation (Milano et al., 2014). This

suggests that within ocean basins on a smaller spatial scale, it may be challenging to use genetics for the traceability of European hake. Stable isotope analysis may be an effective complementary technique, since in this study hake were successfully assigned to origin within the Northeast Atlantic basin with accuracies of between 58% and 75%.

3.5.4 Isotopic variation within discrete regions

Variation in the isotopic composition of haddock and hake muscle tissue within discrete regions was observed, as well as the variation among regions. Haddock caught from around Iceland showed a trend of higher $\delta^{13}C$ values to the south and west of Iceland and lower values to the north. This is very similar to the pattern observed for Atlantic cod in Chapter 2. There is not such a clear pattern in the $\delta^{15}N$ and $\delta^{34}S$ values, although the highest $\delta^{15}N$ values are off the north coast whereas the lowest value is to the south. This variation in isotopic composition between the north and south parts of Iceland may be due to the different currents flowing past. Cold waters of the East Greenland and East Icelandic currents flow southwards from the Arctic, reaching the north and east of Iceland, but the warmer waters of the North Atlantic current and Irminger currents flow around the south and west of Iceland (Oskarsson et al., 2009). Different temperatures and other environmental variables associated with these water masses are likely to have resulted in the haddock muscle isotopic variation from different areas of Icelandic waters. The more mixed sulfur stable isotope values may be due to localised effects on benthic anoxic sediment $\delta^{34}S$ values in different areas.

Variation among stations in the North Sea was also found in haddock, and to a lesser extent, hake isotopic compositions. Haddock showed a clear transition from lower $\delta^{13}C$ and $\delta^{15}N$ values northwest of the Shetland Islands to higher values to the south and east. Sulfur displayed the inverse, since higher $\delta^{34}S$ values were found to the northwest of the Shetlands, indicating that anoxic sediments with lower $\delta^{34}S$ values may be incorporated through their diet in these shallower areas. This pattern for all three isotopes is very similar to that observed in Atlantic cod in Chapter 2, where the Shetland Islands appeared to be a transition zone from high to low isotope values. There is a current eastwards above the Shetlands, where the North Atlantic current follows the edge of the continental shelf, whereas within the North Sea the circulation flows anti-clockwise around the basin (OSPAR Commission, 2000). These different water currents may lead to varying isotopic compositions of fish muscle due to the differing environmental variables associated with each water mass. The pattern observed suggests that haddock do not frequently move across this boundary and generally stay on one side or the other of the Shetland Islands, as the isoscapes also suggested for cod. For hake, the variograms indicated that any spatial structure inferred was not reliable, so there may not be any real spatial trends in

isotope values. Therefore, it is likely that hake undertake larger movements around the sampled area and frequently travel either side of the Shetland Islands.

St. John Glew et al. (2019) developed carbon, nitrogen and sulfur isoscapes for the UK shelf seas using pelagic jellyfish tissues, which provide a comparison for the North Sea isoscapes presented in the current study and in Chapter 2. This reveals differences in the spatial patterns for the same isoscape region, since no north/south variation is observed in the jellyfish isoscapes as it is in those here. However, this may be because the jellyfish isoscapes do not cover the region north of the Shetland Islands where most of the spatial structure is present in the whitefish isoscapes. The $\delta^{13}C$ values in the jellyfish isoscape are lower very close to mainland Scotland, whereas there is a small area of low $\delta^{15}N$ values in the southeast. The sulfur jellyfish isoscape shows lower values along the north edge, which is the inverse of the whitefish isoscapes where the lower $\delta^{34}S$ values are in the southwest. Therefore, there appear to be differences between the pelagic and demersal food webs in this region.

If enough stations were sampled across the smaller-scale regions of isotopic variation, then test samples of haddock or hake could be assigned to a more precise location using isoscapes. Generally, this may not be of great use in fisheries management since the smallest units of spatial management are ICES subareas (e.g. IVa for the northern North Sea), but it could have more importance if No Take Zones or Marine Protected Areas are designated or if local byelaws are introduced, since it would allow regulators to ensure regulations were being adhered to within smaller geographic areas.

3.5.5 Limitations and future research

Limitations in the use of stable isotope analysis for spatial traceability arise from the movement of individuals. Highly migratory species may travel across multiple different regions, resulting in an isotopic signature that is integrated across all of these regions and will be difficult to interpret, especially in species with a slow tissue turnover time. A similar complication results in species or populations that feed on mobile prey that have themselves crossed isotope gradients. European hake are known to feed on pelagic fish such as sardines, anchovies and myctophids (Ferraton et al., 2007), which may move across regional boundaries. It has also been shown that there are frequent migrants of European hake between regions in the Northeast Atlantic (Pita et al., 2011), and haddock also have a relatively high level of connectivity among some areas of their range. Therefore, a recent migrant could be incorrectly assigned to origin. However, stable isotope ratios still have a benefit over genetic techniques in this regard, since genetics would always assign an individual to its historical population of origin whereas stable isotopes have the potential to correctly trace a migrant to the capture region if it has been resident for longer than the tissue turnover time.

Since the heavier isotopes of nitrogen, and to a lesser extent of carbon, increase with trophic level, $\delta^{15}N$ and $\delta^{13}C$ values typically increase systematically with body size due to the typical size structuring of marine food webs (Jennings et al., 2001, Jennings et al., 2008). Therefore, a correction may be needed to account for trophic level effects if individuals of different sizes are assigned to origin using their isotopic compositions. For example, European hake of small or medium size mainly prey on crustaceans, but their diet becomes piscivorous when they grow to a larger size (Ferraton et al., 2007), and therefore move to a higher trophic level with age. In this study, only fish of commercial size were sampled, meaning that all individuals were of at least minimum landing size and smaller juvenile fish at the lowest trophic levels were excluded. However, body size was not recorded since the aim was to develop a tool to be used for provenance of traded fish products, where fish would all be over minimum landing size, but the exact size of the fish is unlikely to be known once processing has been carried out. Furthermore, Ferraton et al. (2007) found that stable isotope ratios of carbon and nitrogen in juvenile hake muscle tissue were not correlated with body size, although further study would be necessary to determine if this holds true for larger hake also. To ensure the best accuracy of assignment, individuals covering the full range of body sizes should be sampled for the reference collection.

Another limitation of the current study is that sufficient samples were not collected from all possible regions across the range of both species to fully encompass the true variance of individuals. This reduces the possibility of accurately assigning individuals to these origins and does not allow the assignment accuracies to be reliably estimated. The Bay of Biscay and Barents Sea for haddock in particular had very small sample sizes, and other regions for hake such as Rockall and the western Iberian coast were not sampled at all in this study. Discrete assignments require a reference dataset of the muscle tissue isotope values for all potential origin regions, since this technique assumes that all possible catch regions are contained in the reference dataset and any outside of this will be incorrectly assigned.

Temporal variation in baseline stable isotope ratios also introduces uncertainty in assignment, since the stable isotope composition of phytoplankton at the base of the food web may vary seasonally or annually. For terrestrial food products, new reference data are collected each year from all geographic regions for this reason (Camin et al., 2017). Therefore, for marine food products reference data should also be acquired during a similar timeframe as the test samples to enable the most reliable assignment to capture region.

3.5.6 Conclusions

In this study, the isotopic compositions of haddock and hake muscle were found to show regional variations and were sufficiently distinct to assign individuals to capture location with >90% accuracy for certain regions. Hake from the Atlantic and Mediterranean could be traced to basin of origin with 100% accuracy and haddock from the Norwegian Sea with 98% accuracy, both of which exceed the 95% confidence limit required to meet industry standards (Camin et al., 2017). However, haddock and hake from other regions were more challenging to assign and could not achieve this level of accuracy. This indicates that stable isotope analysis is unlikely to be used as a stand-alone test for verifying claims of origin for haddock and hake in commercial or legal settings, except for specific regions, but could be used in combination with genetic techniques to result in a higher level of accuracy than either method used alone.

Chapter 4 Tracing the catch location of critically endangered European glass eels using stable isotope and fatty acid analyses

4.1 Abstract

The European eel (*Anguilla anguilla*) is now a critically endangered species following catastrophic declines in populations since 1980, with numbers estimated to be only 5% of 1970s levels. Illegal trade of live European glass eels from Europe into Asia to supply eel farms is one of the largest wildlife crimes in the world, and is a serious threat to the survival of the species. Identifying the geographic source of smuggled eels is crucial for combatting illegal trade, but the complex life history of eels makes this very challenging. In this study, I investigate whether the stable isotope and fatty acid compositions of European glass eels can be used to identify geographic origin. Stable isotope analysis was conducted on 200 glass eels from four European rivers, and a subset of these were also analysed for fatty acid composition. The lipid corrected carbon, nitrogen and sulfur stable isotope ratios as well as fatty acid content of eels were similar among all sampled rivers, however the carbon content and uncorrected $\delta^{13}C$ values did show significant variations with catch location. Glass eels were assigned to known origin with a low overall assignment accuracy of 32-62% using stable isotope and fatty acid tracers independently and combined. However, eels from the River Severn and River Oria were relatively distinct and could be assigned with up to 78% and 89% accuracy respectively, which offers potential discrimination under specific circumstances. Fatty acids do not appear to provide any useful discriminatory power, but the results encourage further investigation into the use of stable isotope analysis as a complementary technique to be used together with other approaches, particularly for certain locations.

4.2 Introduction

The European eel (*Anguilla anguilla*) has been caught for consumption throughout history, giving rise to important commercial eel fisheries across Europe from the late 1800s onwards (Dekker, 2019). By the 1950s European eels supported one of the largest freshwater fisheries in many areas of Europe and northern Africa (FAO, 2019), but populations of European eel have suffered drastic declines since 1980 throughout their distributional range (ICES, 2020). Juvenile eel recruitment reached a minimum in 2011 and

remains very low, with recruitment in 2021 estimated at 0.6% for the North Sea area and 5.5% for elsewhere in Europe compared to 1960-1979 averages (ICES, 2022f). Many reasons have been identified for this decline including overexploitation, habitat loss, river obstruction, pollution, parasitism and changes to ocean environmental conditions at the spawning site (Bonhommeau et al., 2008, Fazio et al., 2012, Belpaire et al., 2019, ICES, 2020). The critical state of European eel populations led to the introduction of the Eel Regulation, requiring EU Member States to implement eel management plans by 2009, with the aim of restoring the spawner run to at least 40% of the level that would have existed in the absence of any anthropogenic impacts (EU, 2007).

The European eel is now classified as a critically endangered species in the IUCN Red List (IUCN, 2018), which led to it being listed under Appendix II of the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES). Consequently all trade of European eels to and from the European Union has been prohibited since 2010 (ICES, 2020). However, there is known to be a significant illegal trade of eels from Europe into some countries in Asia to supply eel farms (Stein et al., 2016, Europol, 2018, UNODC, 2020). Aquaculture of European eels is an important commercial activity - in 2017 over 96% of the global supply of eels was from aquaculture (UNODC, 2020). Eel farming relies on the capture of wild juveniles (glass eels), since adult European eels have never been successfully bred in captivity (Jéhannet et al., 2021). China is by far the largest eel producer, producing 85% of global eel production by weight (all eel species) in 2017 (UNODC, 2020). Since the decline of the Japanese eel, Asian countries have had to source other eel species to supply aquaculture. Demand for wild juveniles has been implicated in the decline of European eels, and the resulting ban on trade of live eels has created a major incentive for eel smuggling. Illegal trafficking of live juvenile eels from Europe to Asia is considered one of the the world's biggest wildlife crimes in terms of both traded individuals and market value (National Wildlife Crime Unit, 2022). The illegal eel trade was estimated to involve approximately 3 billion individual eels at its peak, almost a quarter of the total recruitment of glass eels each year, and was worth 3 billion euros anually (National Wildlife Crime Unit, 2022). In 2017 over 6,800 kg of live glass eels were seized (UNODC, 2020), although trafficking of eels is estimated to have reduced by 50% since 2016 when a European law enforcement operation was launched to combat this crime (Europol, 2022). The leading market for eels is in Asia, where they are often regarded as a delicacy, but with the global spread of Asian food culture they are increasingly being consumed in other countries around the world (Kaifu et al., 2019). This high demand for aquacultured eels, combined with enticing profit margins of up to several thousand euros per kilogram in the illegal supply chain (Europol, 2022), results in a considerable incentive for fraud. There are two sources of glass eels for this trade – either commercial fishers who catch glass eels for the legal market and knowingly or unknowingly supply illegal exporters, or poachers who fish covertly with the intent of supplying the illegal

market (UNODC, 2020). Identifying the source of eels intecepted in transit (or retrospectively for eels being grown on in aquaculture) would dramatically assist policing of the illegal eel trade.

Genetic techniques for species identification have been developed to detect illegal smuggling of European glass eels, which has led to the successful prosecution of eel traffickers (Espiñeira and Vieites, 2016, Cardeñosa et al., 2019). However, this does not give any information on the spatial origin of the eels. Traceability of eels is an obligation of the EU Eel Regulation (EC 1100/2007), which requires the origin of eels to be identified (EU, 2007). Moreover, since the distributional range of European eels extends beyond the borders of the European Union, it is not possible to determine whether eels are of European origin using genetic species identification. Since 2010 there have been CITES certified exports of live European eels from North Africa, with considerable increases in exports from Morocco in particular, which is outside of the EU export restrictions (Pavitt et al., 2021). It is unknown whether these certified exports all originate from North Africa as claimed or whether they include eels of EU origin (UNODC, 2020). There is legal aquaculture of eels in many countries around the world, but some producers are known to use glass eel stock from illegal sources. For example, investigations uncovered illegal fishing of glass eels in Spain that were transported to a legal eel farm in Greece, providing cover for the poaching. Glass eels were then illegally exported to China from the farm (UNODC, 2020). The production of eels in Chinese farms has increased since the 2010 ban on exports from the EU, despite the total value of glass eel imports of all species declining by half since then, so it is unclear how the production is being supported (UNODC, 2020). Previous studies and law enforcement operations have discovered that high proportions of eel products imported from China contain European eel (Richards et al., 2020, Interpol, 2020). Therefore, it is crucial to identify whether confiscated European glass eels originated from within or outside of the EU to enforce the export ban, as well as determining the catch location of eels within the EU to impose quotas and allow sustainable management of fisheries. Establishing the origin of traded eels would also enable authorities to ensure that the claimed origins of glass eels in European eel farms are accurate and legal.

In light of the critical state of European eel stocks, ICES has advised that all anthropogenic mortality of European eel should be kept as close to zero as possible since 2003 and it advises zero catches for 2023, including for aquaculture (ICES, 2022f). Despite this, legal fisheries exist in European countries such as Spain, France and the UK, but is generally restricted to small volumes. In France the quota for glass eels was just under 65 tonnes for 2017-2018, whereas the UK has no national quota. Instead, fishing is restricted to licensed fishers who can only fish within a certain season and in limited areas using only hand nets (GOV.UK, 2022). Commercial landings of glass eels were 52 tonnes in 2021, caught in

four countries – Spain, Portugal, France and the UK – and total commercial adult eel landings were over 2000 tonnes (ICES, 2021e). A significant proportion of the glass eel catches are directed to eel farms, since most eels consumed today come from aquaculture.

European eels have a complex life cycle and undertake impressive migrations of up to 10,000 km or more (Righton et al., 2016). Their geographic range extends from northern Norway southwards to the coast of north Africa in the Atlantic, as well as into the Baltic, Mediterranean and Black Seas (Righton and Metcalfe, 2011). Eels are catadromous, meaning they are born in the ocean and then migrate to freshwater to grow and mature, before returning to the marine environment to spawn. Adult eels spawn in the Sargasso Sea and the eggs hatch into transparent larvae (leptocephali), which cross the Atlantic Ocean towards the European coast mainly by drifting with oceanic currents (Schmidt, 1923). Upon reaching the European continental shelf, the leptocephali undergo their first metamorphosis into transparent glass eels. The glass eels gather into large shoals and migrate towards the mouth of the estuary. Once they have entered freshwater and reach a size of 6-7 cm in length, the glass eels undergo another transformation into elvers (Figure 4.1). The eels are now pigmented and feed on small invertebrates as they continue their journey up-river, searching for a suitable habitat to live (Righton and Metcalfe, 2011). The elvers eventually become yellow eels, which are fully adapted for freshwater, and remain in the river for up to 20 years before they begin their downstream migration back out to sea (Vøllestad, 1992). To prepare for this journey, the yellow eels undergo another major metamorphosis into silver eels, which have a number of specialised physiological and morphological adaptations for the next marine phase of their life (Righton and Metcalfe, 2011). Silver eels migrate down-river and escape into the sea, facing another huge oceanic migration to reach their spawning grounds in the Sargasso Sea (Righton et al., 2016). Due to this complex life cycle, European eels can be exposed to adverse conditions or anthropogenic impacts in multiple different habitats, which makes them particularly susceptible to disturbance (Miller et al., 2016).

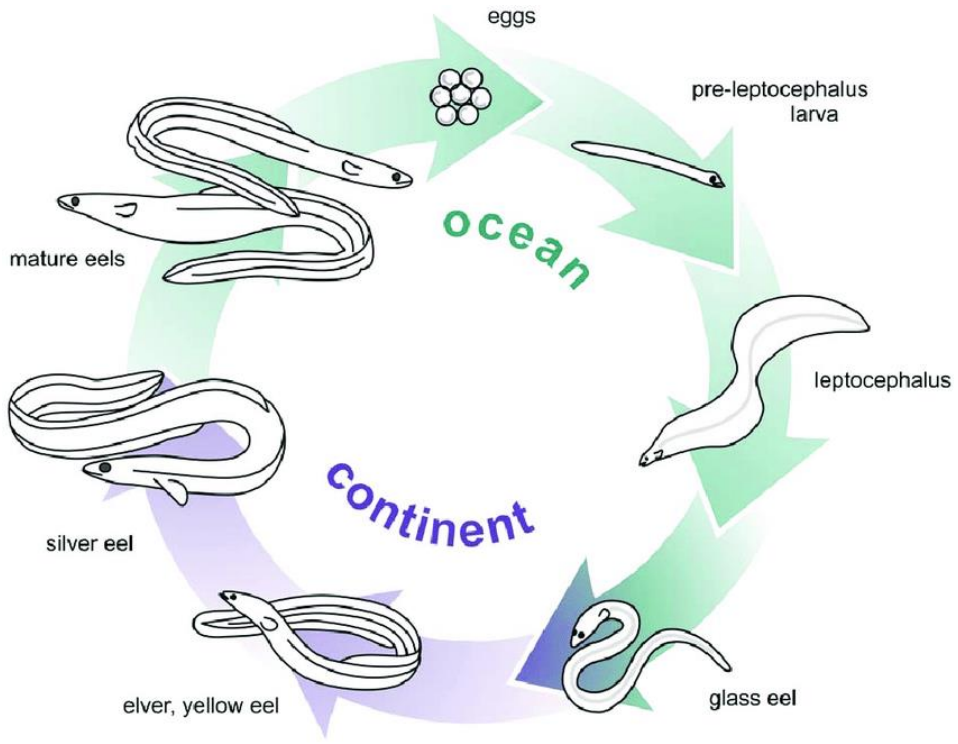


Figure 4.1 Life cycle of the European eel (*Anguilla anguilla*). Image from Henkel et al. (2012)

Stable isotope approaches have been widely used for tracing the movements of migrating animals, such as birds (Kelly et al., 2002, Hobson et al., 2014), butterflies (Hobson et al., 1999), bats (Cryan et al., 2004) and deer (Cormie et al., 1994), as well as for tracing migration of marine species including fish (Trueman et al., 2012a), marine mammals (Witteveen et al., 2009, Trueman et al., 2019), turtles (Vander Zanden et al., 2015) and seabirds (St. John Glew et al., 2018). For example, the at-sea foraging locations of seabirds were identified by analysing the isotopic composition of feathers collected from individuals whilst at their breeding grounds on land (St. John Glew et al., 2018). Similarly, carbon stable isotope values in Atlantic salmon scales were used to determine the locations of open ocean feeding grounds for different populations of salmon (MacKenzie et al., 2011). However, the majority of these studies used the hard calcified structures of animals, such as scales, feathers or otoliths, to derive spatial information since these structures act as a permanent record of isotopic signatures over time. It is much more difficult to track migrating animals using soft tissues because the stable isotope signatures of different regions become integrated over time as the animal moves and assimilates food, meaning that the data are difficult to interpret. Therefore stable isotopes are generally not beneficial for tracking highly migratory species unless the animals spend long enough in one location to incorporate the local isotopic signature. This technique for tracing spatial origin has rarely been investigated for species undergoing migration from marine to freshwater environments.

One example of a catadromous species with a similar life cycle to eels is barramundi, which spend parts of their life in freshwater, estuarine and marine environments (Davis, 1986). Stable isotopes have been used to successfully distinguish between farmed barramundi from different origins. A study by Gopi et al. (2018) found that barramundi from farms in Taiwan and Malaysia had significantly different $\delta^{13}C$ and $\delta^{15}N$ values, but did not investigate whether this technique could be used for tracing provenance in wild-caught barramundi. Another study did include wild-caught barramundi, and were able to reliably determine geographic origin among those caught in Malaysia and two regions of Australia using stable isotopes, achieving an assignment accuracy of 84% (Gopi et al., 2019c). The addition of elemental abundance data measured using XRF resulted in a similar assignment accuracy as stable isotope analysis used alone (81%).

Fatty acid analysis is another promising technique that has been investigated for tracing the geographic origin of a number of aquatic species (Zhang et al., 2017, Gong et al., 2018, Go et al., 2022, Fonseca et al., 2022). The fatty acid composition of fish is strongly affected by diet, and so reflects the fatty acid signature of primary production or prey items (Linko et al., 1985, Shirai et al., 2002, Dwyer et al., 2003). Fatty acids in phytoplankton may show regional variations due to environmental factors and differences in the plankton species composition (Reuss and Poulsen, 2002). These fatty acid signatures are transferred up the food chain through diet and are incorporated into the tissues of fish, therefore acting as a fingerprint indicating the location where individuals were last feeding. However, no studies in published literature could be found where this technique has been applied to catadromous species.

The spatial traceability of European eels has scarcely been studied, despite its importance in the sustainability of the species. The current study contributes to a wider project coordinated by the Sustainable Eel Group, and applies biochemical techniques to this purpose for the first time (Stein et al., 2019). Trace element analysis and metabolomics are explored as part of this project, in addition to stable isotope and fatty acid analyses described here. Prior to this, only one other study could be found in published literature that uses biochemical techniques to trace the origin of eels. Yamashita et al. (2006) applied trace element analysis of the muscle tissue to distinguish among Japanese eels (*Anguilla japonica*) of different origins. Farmed eels from five regions in Asia and wild-caught eels from three rivers in Japan showed clustering by region of origin based on the concentrations of six elements, but no accuracies of assignment to origin were obtained. To my knowledge, no studies prior to the current project have been carried out that focus on tracing the provenance of European glass eels, nor could any published studies be found that assigned eels to geographic origin based on stable isotope and fatty acid compositions. However, Vasconi et al. (2019) used these two techniques to differentiate between European eels of different production methods. The lipid content, fatty acid profile

and stable isotope ratios ($\delta^{13}C$ and $\delta^{15}N$) were found to vary significantly between farmed and wild-caught eels, and therefore eels from the two production methods were chemically distinct. However, it was not possible to discriminate among eels from different farms, likely because the composition of the commercial feed was similar across sites (Vasconi et al., 2019).

For marine species, genetic tracers have been successfully applied to distinguish among individuals from different regions (Nielsen et al., 2012b, del Rio-Lavín et al., 2022a), so this could also be a valuable technique for tracing provenance of catadromous species. Marshall (2005) found that barramundi show significant differences in DNA data among populations in different river drainage basins around Australia, which suggests the possibility of assigning individuals to area of origin using genetics. However, genetic techniques are not likely to be effective for European eels because is it a single panmictic population, where adults from all populations migrate to a common spawning ground, and so there is very low genetic variation in adult eels across their distribution (Palm et al., 2009). Therefore, the differences among eels caught in different locations are not likely to be sufficient to allow discrimination.

Therefore, there is a distinct lack of research into the potential use of biochemical techniques for the spatial verification of eels. The development of forensic tests for provenance, particularly for glass eels destined for farms, would be hugely beneficial for detecting illegal trade that threatens the future survival of the European eel. It would also enable more sustainable management of fisheries and more effective enforcement of quotas. The aim of the current study is to investigate whether the stable isotope and fatty acid composition of European glass eels can provide discrimination based on river of origin, and therefore to assess the potential of these techniques to be used as a forensic tool for detecting illegal glass eel fishing and trade. We measured the carbon, nitrogen and sulfur stable isotope ratios of 200 glass eels from four rivers across Europe, and also analysed the fatty acid profiles of 39 of these same eels. We then used these datasets to determine the accuracy with which individuals can be traced back to their true river of origin, firstly based on the two techniques separately and then both techniques combined.

4.3 Methodology

4.3.1 Sampling

Juvenile European eels (glass eels), *Anguilla anguilla*, were sampled by the Sustainable Eel Group as part of a wider project investigating the use of potential chemical markers for glass eel traceability. Glass eels were caught in 2017 and 2018 from four European rivers near to the mouths – River Severn (UK), River Parrett (UK), River Vilaine (France) and

River Oria (Spain) (Figure 4.2). Fifty eels were collected from each site for analysis (200 eels in total), and were stored frozen at -20°C until processing in the laboratory.

Figure 4.2 Locations of the four rivers sampled for glass eels (*Anguilla anguilla*).

The posterior half of each glass eel (excluding head section) was severed, given a unique ID number and transferred to a separate plastic vial. These samples were then divided into two, allowing one half to be analysed for stable isotope ratios and the other for fatty acid composition. Table 4.1 shows the number of glass eels from each river used for these analyses. Care was taken to avoid sampling the internal organs, such as the gut, stomach and liver, since this would be likely to result in inaccurate measurements. It was not possible to remove the skin due to the small size of the eels (generally ~7cm long), so the skin was included with the muscle tissue for analysis, despite the potential complications in interpretation introduced by the different protein and lipid compositions of the tissues.

Table 4.1 Number of individual glass eels analysed for stable isotope ratios and fatty acid composition at each of the four rivers sampled.

River of origin	Number of samples analysed	
	Stable isotopes	Fatty acids
Severn, UK	50	9
Parrett, UK	50	10
Vilaine, France	50	10
Oria, Spain	50	9
Total	**200**	**38**

4.3.2 Stable isotope analysis

The frozen eel muscle samples were freeze-dried at -55°C for 24 hours, and then homogenised into a powder with scissors. For each sample, approximately 3.5mg was weighed and placed into a tin capsule for analysis. A total of 200 samples were analysed for bulk carbon, nitrogen and sulfur stable isotope ratios using an Elementar vario PYRO cube elemental analyser coupled with an Isoprime visION isotope ratio mass spectrometer, housed at the University of Southampton within the National Oceanography Centre.

Stable isotope compositions are expressed as $\delta^{13}C$, $\delta^{15}N$ and $\delta^{34}S$ values in per mille (‰) relative to the international standards (Vienna-PeeDee Belemnite, Air and Vienna-Canyon Diablo Troilite). Laboratory internal standards (sulfanilamide and a fish muscle standard), an in-house comparison standard (glutamic acid) and blanks were included together with the samples to monitor accuracy and precision of measurements.

The eel muscle samples all had a high lipid content, indicated by the high C:N ratios of between 3.68 and 5.14. Therefore, an arithmetic lipid correction was applied to the $\delta^{13}C$ values using the methodology described by Kiljunen et al. (2006), since lipids are depleted in $\delta^{13}C$ compared to proteins and carbohydrates.

4.3.3 Fatty acid analysis

A subset of the eels analysed for stable isotope composition were also analysed for fatty acid composition. From each of the four sampling sites, ten of the eels were randomly selected for fatty acid analysis, however two eels did not produce useable data due to accidental errors in the laboratory procedure, so data were acquired from 38 samples in total. Analysis of samples was conducted at the University of Southampton based at the National Oceanography Centre.

The frozen eel muscle samples were freeze-dried at -55°C for 24 hours. Each freeze-dried sample was weighed and then placed into a 10ml ASE cell. The cell was topped up with sand until full in order to fill void space within the ASE cell. For each batch of 5 samples, a blank cell containing only sand was prepared to indicate any contamination during the process.

The internal standard selected was 2-methyloctadecanoic acid, due to its similarity to the fatty acids contained in eel muscle and because its peak does not overlap with any peaks in the eel samples. A standard stock solution was created by combining the standard with dichloromethane, giving a stock solution concentration of 3741.21ng/µl. 25µl of standard stock solution was injected into each ASE cell, including the blanks. The amount of standard stock solution added (25µl) was decided upon because this volume represents approximately 100ng/µl of the internal standard in the final sample, which gives a good peak size compared to the peaks in the eel samples.

The samples contained in the ASE cells were extracted using a Thermo 350 Accelerated Solvent Extractor with the following program: preheat = 5 min; heat = 5 min; static = 5 min; pressure = 1500 psi; flush = 70%, purge = 300 s.; cycles = 3; solvent = 9:1 dichloromethane:methanol. Subsequently, the samples were evaporated to a residue in a Genevac EZ-2 vacuum centrifuge. The total lipid extraction for each sample was transferred into a gas chromatograph vial by dissolving consecutively in dichloromethane, methanol and dichloromethane again, before being evaporated to a residue on a hot plate with nitrogen gas directed at the vials. The weight of the total lipid extraction residue was recorded. A mixture of N,O-bis(trimethylsilyl)trifluoroacetamide (BSTFA) and dichloromethane was created and 1ml of this was added to each vial to derivitise the fatty acids. The samples were placed in an oven at 50°C for 30 minutes to allow derivitisation to occur.

All samples were then analysed for their fatty acid composition using a Thermo Trace 1310 gas chromatograph coupled to Thermo TSQ8000 mass spectrometer. The gas chromatograph used DB-5 column (30 m × 0.25 mm i.d, 0.25-µm film thickness) and the oven program was started at 40°C (held for 2 min), increased at a rate of 6°C/min to 310°C, and was then held for 20 minutes. Dichloromethane was included as a blank following every three samples and the sand blanks from the initial extraction process were also analysed.

Fatty acids present were identified using mass spectra, library matches, and by comparison to known standards. The relative peak area of the internal standard was used to calculate the abundance of each fatty acid in the total extract, and this was then converted to mass per milligram of eel tissue using the eel weights recorded prior to analysis.

4.3.4 Statistical analysis

In this study, samples were assigned to one of four discrete possible rivers based on comparing the isotopic or fatty acid composition of each sample to the distribution of values measured within a reference population of known origin.

Two methods of assignment to origin were compared to determine the most effective – a multivariate normal probability technique and random forest classification. This was conducted firstly using the stable isotope data and fatty acid data separately to investigate the ability of each of these techniques to assign eels to river of origin. Then the assignments were carried out using the stable isotope and fatty acid data combined.

The multivariate technique and random forest classification were both carried out as described in Chapter 2 on page 42. All analyses were performed using R version 3.6.2 (R Core Team, 2019).

4.4 Results

4.4.1 Stable isotope analysis

Measurement error associated with the stable isotope analysis of the fish muscle samples is 0.1-0.2‰ for $\delta^{13}C$, 0.1-0.2‰ for $\delta^{15}N$ and 1.2-1.3‰ for $\delta^{34}S$, determined as the standard deviation of replicate measurements of three internal standards measured together with the samples – sulfanilamide, a fish muscle standard and an in-house glutamic acid standard (Table 4.2).

Table 4.2 Comparison of the mean values and uncertainties (standard deviations) in the stable isotope measurements of three internal standards that were analysed together with the eel samples.

Standard	Laboratory	$\delta^{13}C$ (‰)		$\delta^{15}N$ (‰)		$\delta^{34}S$ (‰)	
		Mean	SD	Mean	SD	Mean	SD
Fish muscle	NOCS	-19.23	0.15	11.88	0.18	19.58	1.34
Sulfanilamide	NOCS	-27.38	0.17	-1.58	0.14	-6.43	1.21
Glutamic acid	NOCS	-13.06	0.10	-3.85	0.14	NA	NA

The bulk elemental contents (carbon, nitrogen and sulfur) and C:N ratios of glass eels from all sampled rivers are generally similar (Figure 4.3). The means and standard deviations for each location are listed in Table 4.3. No significant differences were found in the nitrogen or sulfur content or in the C:N ratios (Kruskal-wallis – N(%): χ^2=97.351, d.f.=114,

p=0.8679; S(%): χ^2=38.807, d.f.=34, p=0.2619; C:N: χ^2=92.821, d.f.=91, p=0.4272), however there is a significant difference in the carbon content (ANOVA - f=15.88, d.f.=3, p=2.77e^{-09}). The carbon composition is between 44% and 53%, with a slightly higher carbon content in glass eels from the River Oria (Spain), reflecting higher lipid contents. Nitrogen compositions are relatively similar across all eels, with a mean content of 11.4%. The sulfur content ranges from 0.6% to 1.1%, with the highest content in eels from the River Severn (UK) and the lowest in eels from the River Oria. The C:N ratio varies between 3.68 and 5.14, so a lipid correction was applied to all glass eels. The River Oria eels have the highest mean C:N ratio of 4.41, whereas the Rivers Vilaine and Parrett have relatively low C:N ratios (mean of 4.19 for both), implying a higher lipid content in eels from the River Oria than those from the other rivers.

Stable isotope compositions of the glass eel samples did not differ significantly among the four rivers (ANOVA - δ^{13}C (lipid corrected): f=2.396, d.f.=3, p=0.0695; δ^{15}N: f=1.961, d.f.=3, p=0.121; Kruskal-Wallis - δ^{34}S: χ^2=174.59, d.f.=158, p=0.1737), except for the uncorrected δ^{13}C values which did show significant differences among rivers (ANOVA - f=4.956, d.f.=3, p=0.00245). The δ^{13}C (uncorrected and lipid corrected), δ^{15}N and δ^{34}S values measured in the eel samples from each river are shown in Figure 4.4, and the means and standard deviations for each location are listed in Table 4.4. Overall, lipid corrected δ^{13}C values range from -21.4‰ to -18.1‰ and there is very little variation among rivers, however the mean uncorrected δ^{13}C value is lower in eels from the River Oria, suggesting a higher lipid content (in agreement with the C:N ratios). The δ^{15}N values are also very similar in eels from all rivers, ranging from 4.8‰ to 8.9‰. There is slightly more variation in the δ^{34}S values across the rivers, although these differences are not significant. The River Oria has the lowest mean δ^{34}S value of 20.0‰, but with a relatively wide range of values from 16.6‰ to 22.2‰, and the River Severn has the highest mean of 21.8‰.

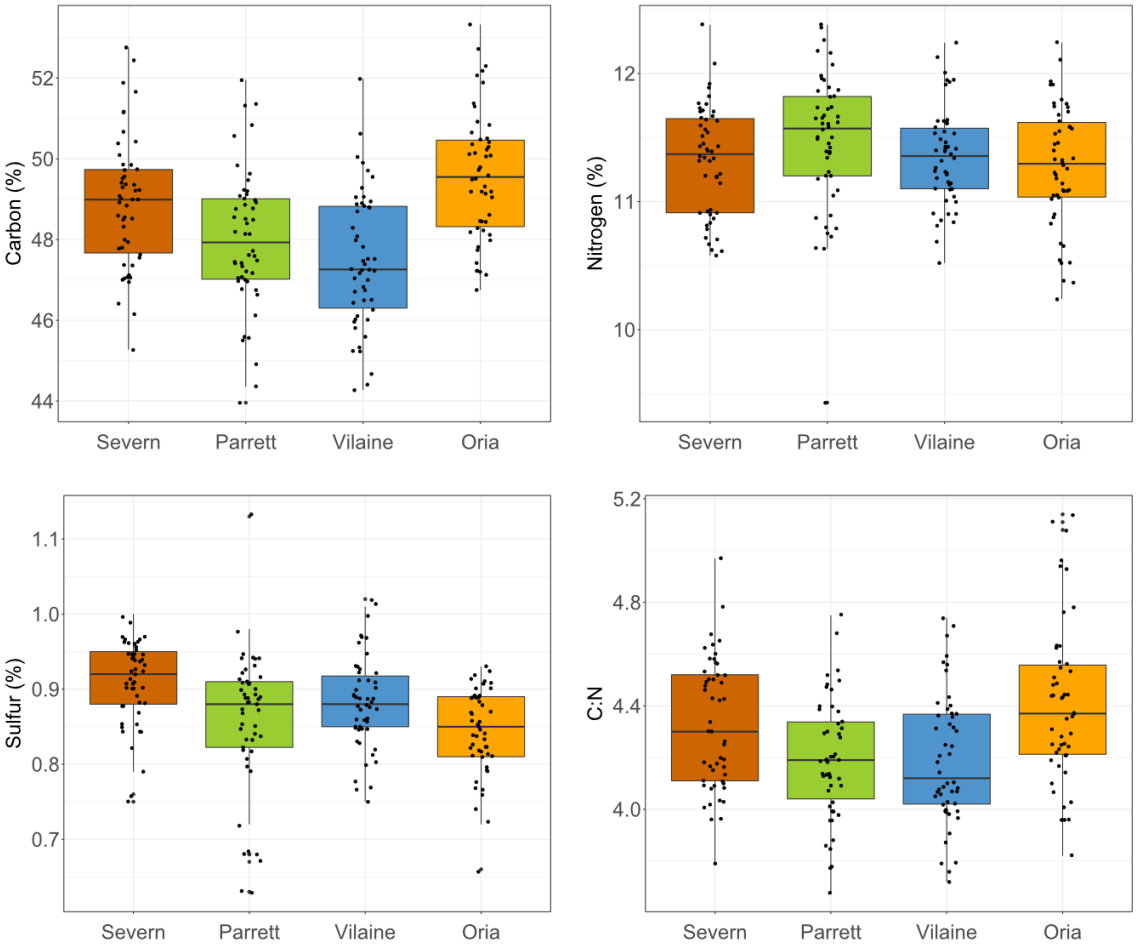


Figure 4.3 Carbon, nitrogen and sulfur tissue contents (%) and C:N ratios of glass eels from the four sampled rivers.

Table 4.3 Mean carbon, nitrogen and sulfur tissue compositions (%) measured in glass eels from each of the four rivers and the calculated C:N ratios, all with associated standard deviations.

River of origin	Carbon (%)		Nitrogen (%)		Sulfur (%)		C:N	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Severn, UK	48.9	1.62	11.3	0.43	0.9	0.06	4.3	0.25
Parrett, UK	48.0	1.73	11.5	0.54	0.9	0.09	4.2	0.23
Vilaine, France	47.5	1.68	11.3	0.38	0.9	0.06	4.2	0.25
Oria, Spain	49.6	1.59	11.3	0.48	0.8	0.06	4.4	0.31

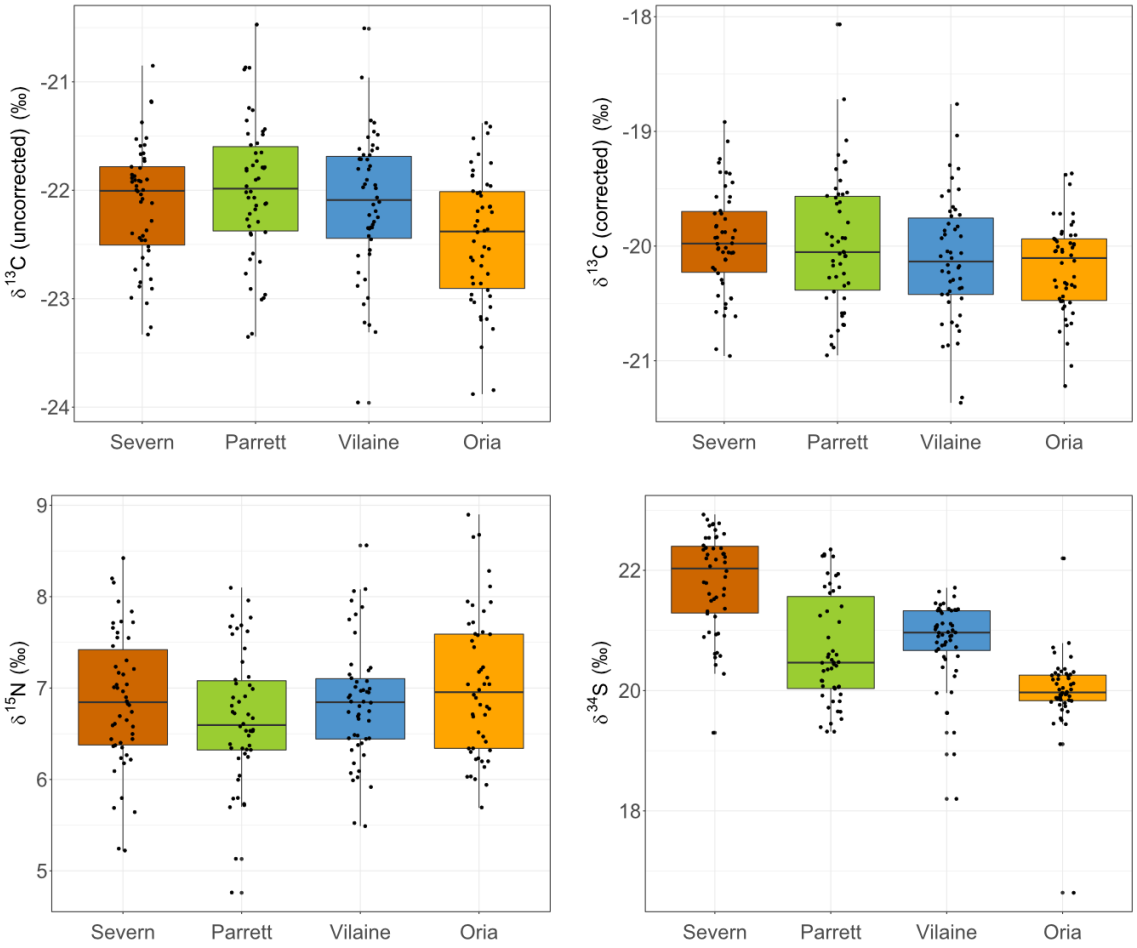


Figure 4.4 Uncorrected and lipid corrected $\delta^{13}C$, $\delta^{15}N$ and $\delta^{34}S$ values measured in glass eels from the four sampled rivers.

Table 4.4 Mean stable isotope ratios of carbon (uncorrected and lipid corrected), nitrogen and sulfur measured in glass eels from each of the four rivers, with the associated standard deviations.

River of origin	$\delta^{13}C$ uncorrected (‰)		$\delta^{13}C$ lipid corrected (‰)		$\delta^{15}N$ (‰)		$\delta^{34}S$ (‰)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Severn, UK	-22.1	0.56	-20.0	0.45	6.9	0.74	21.8	0.82
Parrett, UK	-22.0	0.63	-20.0	0.59	6.7	0.72	20.7	0.91
Vilaine, France	-22.1	0.64	-20.1	0.54	6.8	0.65	20.8	0.68
Oria, Spain	-22.5	0.61	-20.2	0.40	7.0	0.79	20.0	0.66

Figure 4.5 displays histograms for the carbon, nitrogen and sulfur stable isotope composition of eels from each location. The $\delta^{13}C$ and $\delta^{15}N$ values of eels from all four rivers show a normal distribution, whereas the $\delta^{34}S$ values are not normally distributed for any of the sampled rivers.

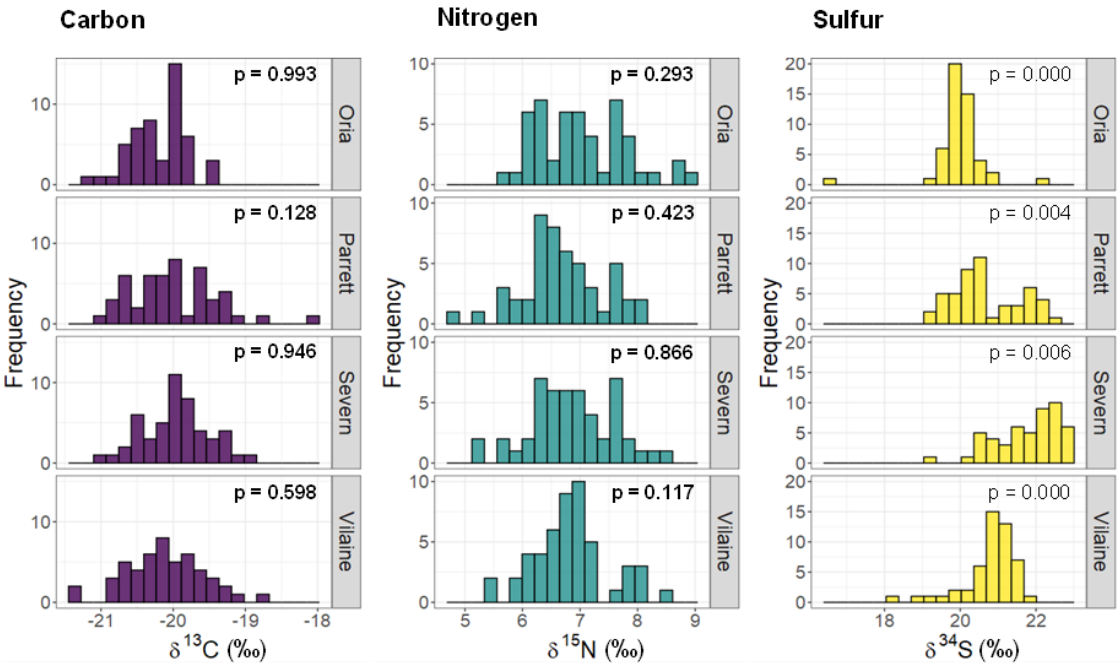


Figure 4.5 Frequency distributions of $\delta^{13}C$, $\delta^{15}N$ and $\delta^{34}S$ values from glass eels from each of the sampled rivers. The p-values from the Shapiro-Wilk tests for normality are shown on each histogram, and those that were shown to be normally distributed ($p>0.05$) are indicated in bold.

The results of the stable isotope analysis do not reveal any clear clustering of the eel samples by river of origin, as shown in Figure 4.6. The samples from each river are generally grouped together, but the data ellipses for all four rivers overlap almost entirely. Therefore, the eels caught in different locations cannot be distinguished on the basis of their stable isotope compositions.

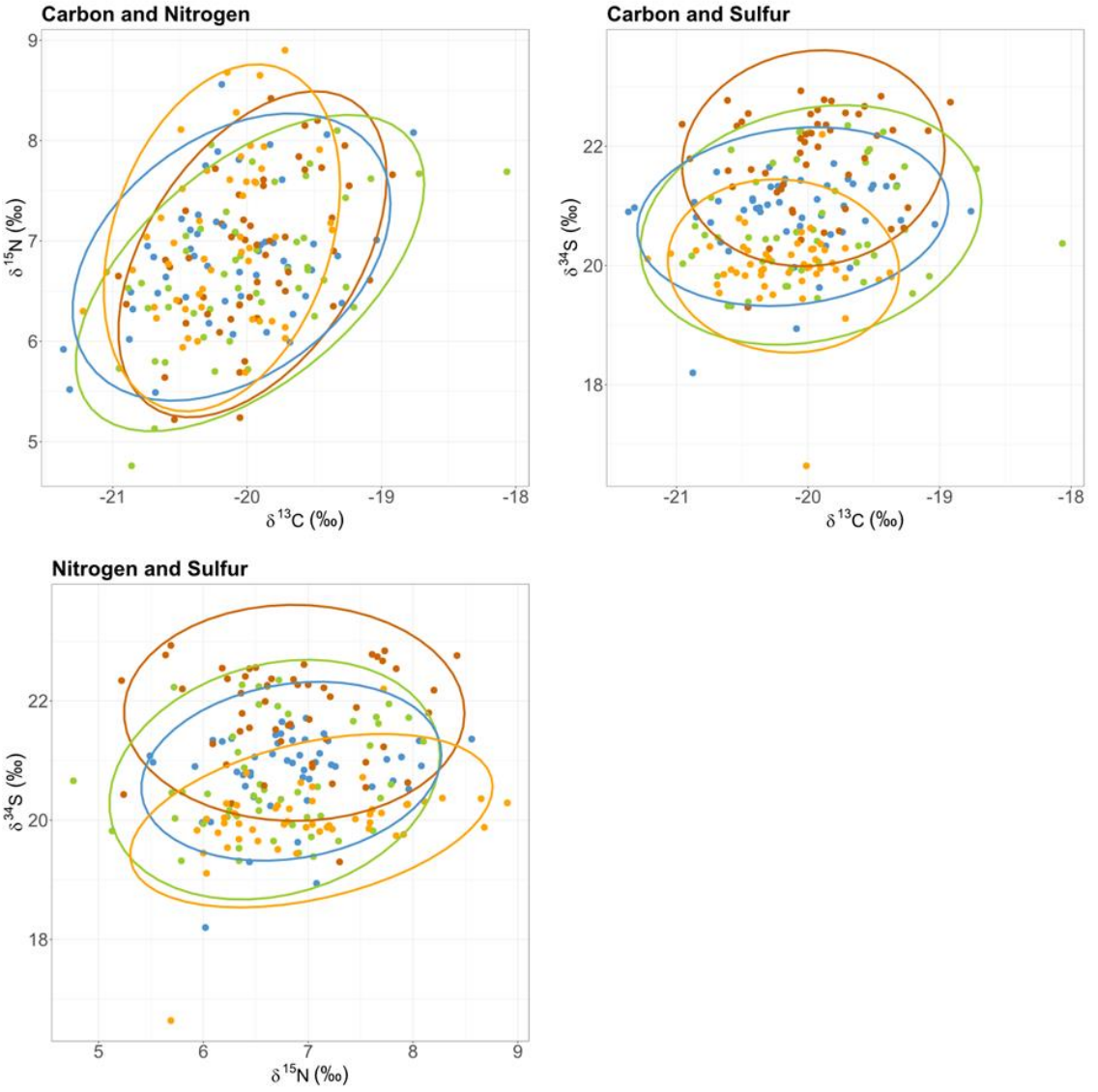


Figure 4.6 Carbon, nitrogen and sulfur stable isotope values for each individual glass eel sampled, coloured by region of origin. The 90% data ellipses are also shown for each geographic region.

4.4.2 Fatty acid analysis

A total of 19 fatty acids were identified in the glass eel samples, including saturated, monounsaturated and polyunsaturated (PUFA) fatty acids. The positions of double bonds along the chain of the molecule could not be determined using the analytical method applied, so only the number of carbon atoms and double bonds is recorded. The abundances of all these fatty acids in eels from each location are displayed in Figure 4.7, and Table 4.5 lists the means and standard deviations by river of origin. There are no significant differences in the fatty acid compositions of eels among the four rivers (ANOVA - d.f.=3, p>0.513), except for cholesterol (ANOVA - f=4.025, d.f.=3, p=0.0149). The

abundance of cholesterol is significantly lower in glass eels from the River Oria, where the mean is 3.62 µg per mg of eel, whereas the River Vilaine has the highest mean cholesterol concentration of 6.29 µg per mg of eel. The fatty acid present in the highest abundances was palmitic acid (16:0), with mean concentrations between 10.46 µg per mg of eel in the River Oria and 12.68 µg per mg of eel in the River Parrett. Glycerol was the second most abundant fatty acid, which was found in the glass eels in concentrations of between 3.75 and 16.11 µg per mg of eel. A monounsaturated fatty acid (18:1) was also present in relatively high concentrations (up to 14.3 µg per mg of eel), but all other measured fatty acids had low concentrations of below 5 µg per mg of eel in general.

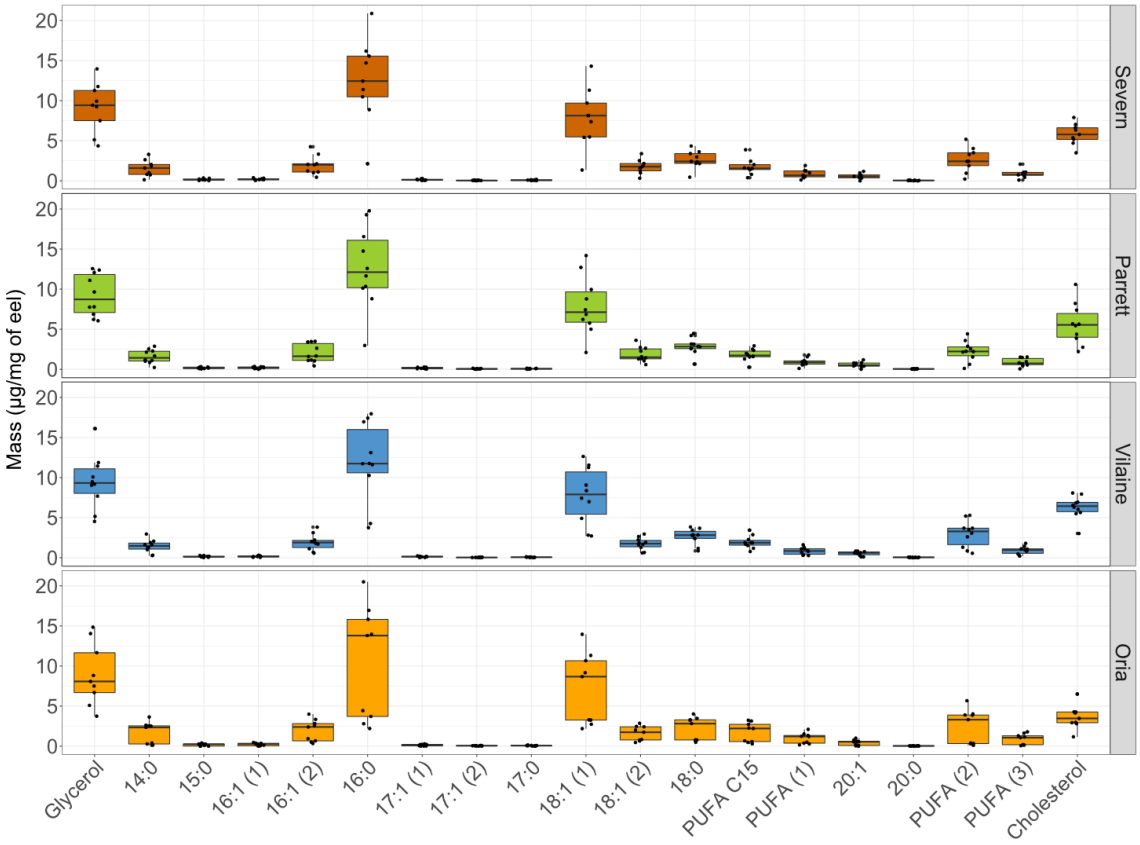


Figure 4.7 Abundance of each of the identified fatty acids in eels from the four sampled rivers. The numbers in brackets indicate more than one fatty acid with the same number of carbon atoms and double bonds where the exact structure could not be identified.

Table 4.5 Means and standard deviations of fatty acid abundances (µg/mg of eel) for each of the four rivers. Fatty acids are listed as C:n, where C is the number of carbon atoms and n is the number of double bonds. Where there are more than one fatty acid with the same number of carbon atoms and double bonds, this is because the exact structure could not be identified.

Fatty acid	Mean (µg/mg of eel)				SD (µg/mg of eel)			
	Severn	Parrett	Vilaine	Oria	Severn	Parrett	Vilaine	Oria
Glycerol	9.18	9.24	9.46	8.95	3.10	2.62	3.34	3.84
Myristic acid (14:0)	1.57	1.59	1.44	1.62	1.00	0.85	0.81	1.37
Pentadecylic acid (15:0)	0.17	0.18	0.14	0.18	0.11	0.10	0.08	0.15
16:1	0.20	0.19	0.17	0.22	0.11	0.11	0.09	0.17
16:1	1.95	2.00	1.91	1.96	1.20	1.15	1.01	1.36
Palmitic acid (16:0)	12.52	12.68	11.88	10.46	5.29	5.12	4.96	7.10
17:1	0.14	0.15	0.13	0.14	0.08	0.09	0.06	0.10
17:1	0.06	0.06	0.04	0.06	0.03	0.04	0.02	0.03
Margaric acid (17:0)	0.09	0.07	0.07	0.07	0.05	0.04	0.03	0.05
18:1	7.91	7.90	7.78	7.25	3.73	3.63	3.51	4.42
18:1	1.77	1.85	1.74	1.58	0.89	0.93	0.77	0.91
Stearic acid (18:0)	2.66	2.87	2.63	2.17	1.11	1.05	0.98	1.45
PUFA C15 unidentified	1.74	1.79	1.94	1.73	1.00	0.74	0.78	1.22
PUFA unidentified	0.85	0.90	0.84	0.97	0.56	0.49	0.45	0.66
20:1	0.59	0.57	0.54	0.43	0.36	0.33	0.27	0.35
Arachidic acid (20:0)	0.05	0.05	0.05	0.03	0.02	0.02	0.03	0.02
PUFA unidentified	2.66	2.23	2.98	2.44	1.54	1.28	1.66	2.14
PUFA unidentified	0.88	0.86	0.92	0.84	0.54	0.50	0.49	0.69
Cholesterol	5.82	5.62	6.29	3.62	1.32	2.56	1.43	1.46

Principal Components Analysis (PCA) demonstrates that the eel samples do not cluster according to their capture river, since no structure is observed and data points from each river are widely distributed across the plot (Figure 4.8). The first and second principal components explain 76.2% and 7.3% of the variation among eel samples respectively. Despite cholesterol being the only fatty acid to show significant differences with river of origin using ANOVA, it contributed least to the variation among samples in the PCA (Figure 4.9).

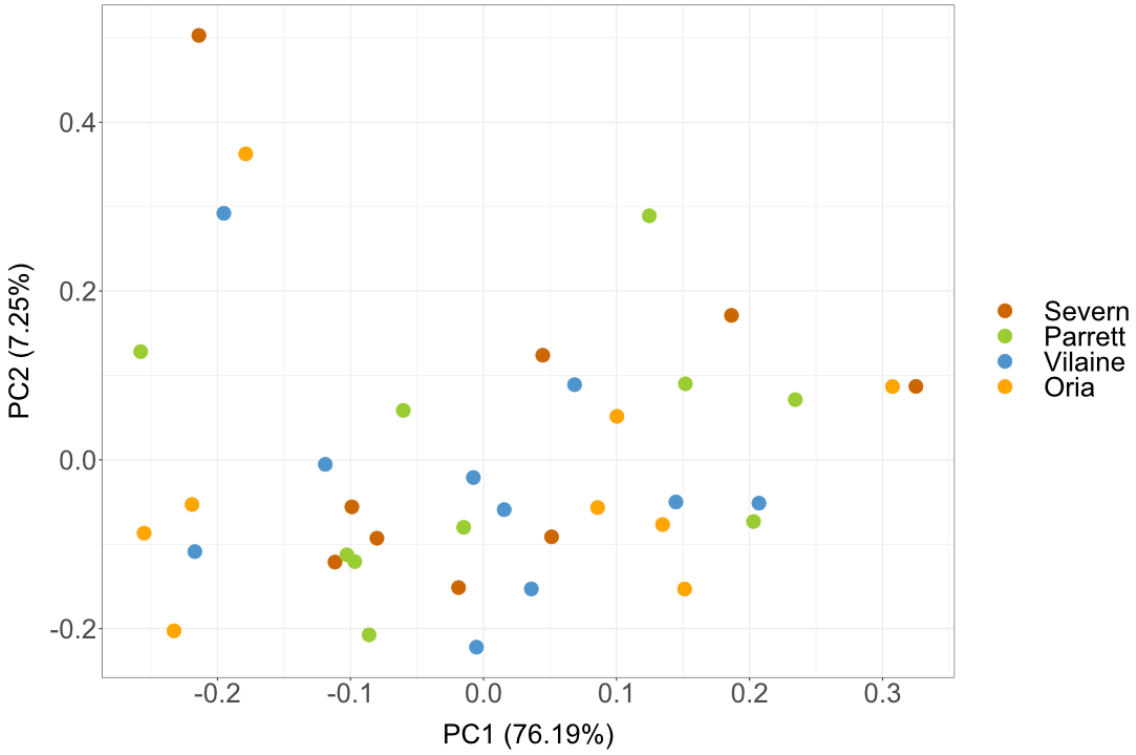


Figure 4.8 Principal components 1 and 2 using the fatty acid data to show variability among the eel samples by river of origin with Principal Components Analysis (PCA).

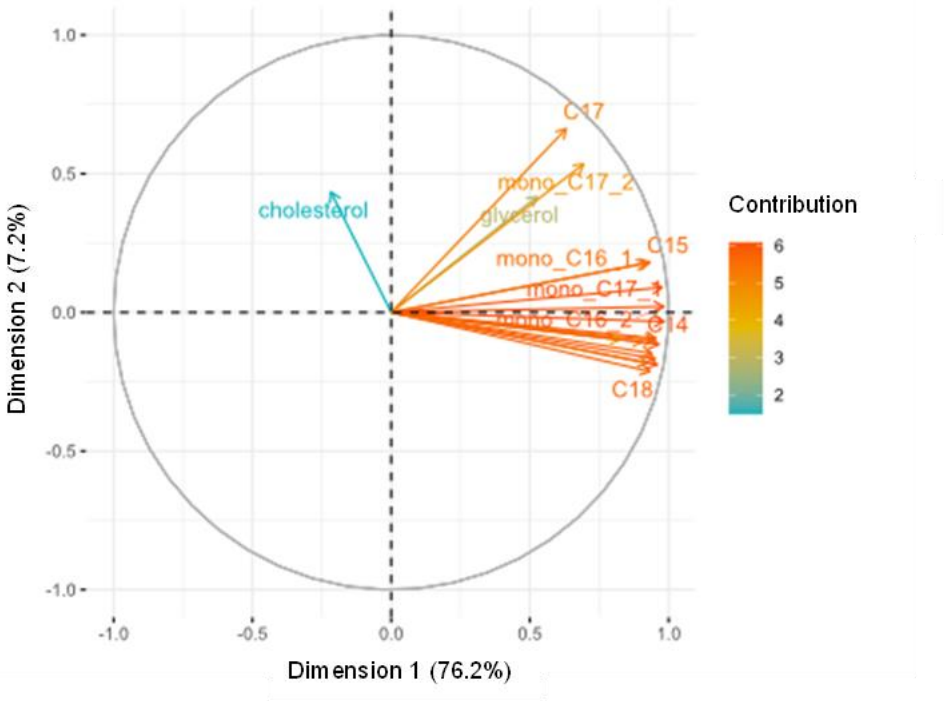


Figure 4.9 Principal Component Analysis variables plot, showing the fatty acids and their contribution to the principal components.

4.4.3 Assignment to river of origin

4.4.3.1 Assignment using stable isotopes

The mean assignment accuracy to known origin for all rivers was 49.5% using the multivariate technique with all three stable isotopes (carbon, nitrogen and sulfur), and the success rates for each location are shown in Table 4.6. The assignment accuracies with repeated simulations where different subsets of the data were selected for the training and test datasets are illustrated in Figure 4.10. Glass eels from the River Oria (Spain) were the most accurately assigned, with a mean correct assignment rate of 70.8%, whereas those from the River Parrett (UK) had the lowest success rate of only 25.5%, which is the same as random assignment (since there are four sites) and indicates that stable isotopes gave no extra information in this case. The River Vilaine (France) also had a relatively low percentage of correct assignments (37.7%). For eels from all rivers, incorrect assignments were classified to all three of the remaining locations (Figure 4.10).

The assignment accuracies using only $\delta^{13}C$ and $\delta^{15}N$ values for the multivariate assignment technique over repeated simulations are shown in Figure 4.11. Excluding sulfur isotope values and using only carbon and nitrogen resulted in an overall success rate of 30.9%. This is significantly reduced compared to using all three isotopes, and again is not much better than random assignment, suggesting that $\delta^{13}C$ and $\delta^{15}N$ values contain no geographic information (as implied by the scatterplots in Figure 4.6). The success rate was lower for all locations except the River Parrett when $\delta^{34}S$ values were excluded, with a difference of up to 33% compared to the accuracy using three isotopes (Table 4.6). This demonstrates the importance of sulfur isotope values in distinguishing among glass eels from these four rivers.

Table 4.6 Mean percentage of individuals assigned to the correct origin region over 1000 repeat simulations using three isotopes ($\delta^{13}C$, $\delta^{15}N$ and $\delta^{34}S$) and two isotopes ($\delta^{13}C$ and $\delta^{15}N$).

Region	Mean correct assignments (%)	
	CNS	CN
Severn, UK	63.8	30.5
Parrett, UK	25.5	36.3
Vilaine, France	37.7	12.7
Oria, Spain	70.8	43.9
Total	**49.5%**	**30.9%**

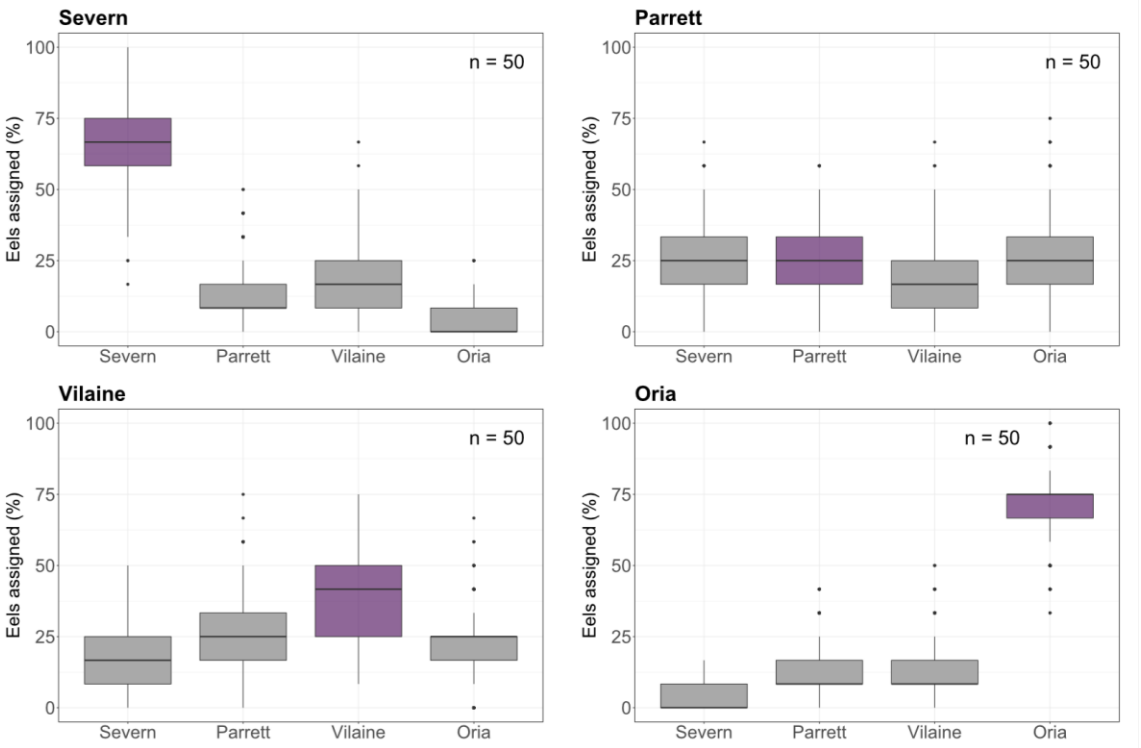


Figure 4.10 Assignment results using carbon, nitrogen and sulfur stable isotope data, showing the percentage of individuals from each known location assigned to all the possible regions over 1000 repeat simulations. The coloured boxes show the correct regions of origin.

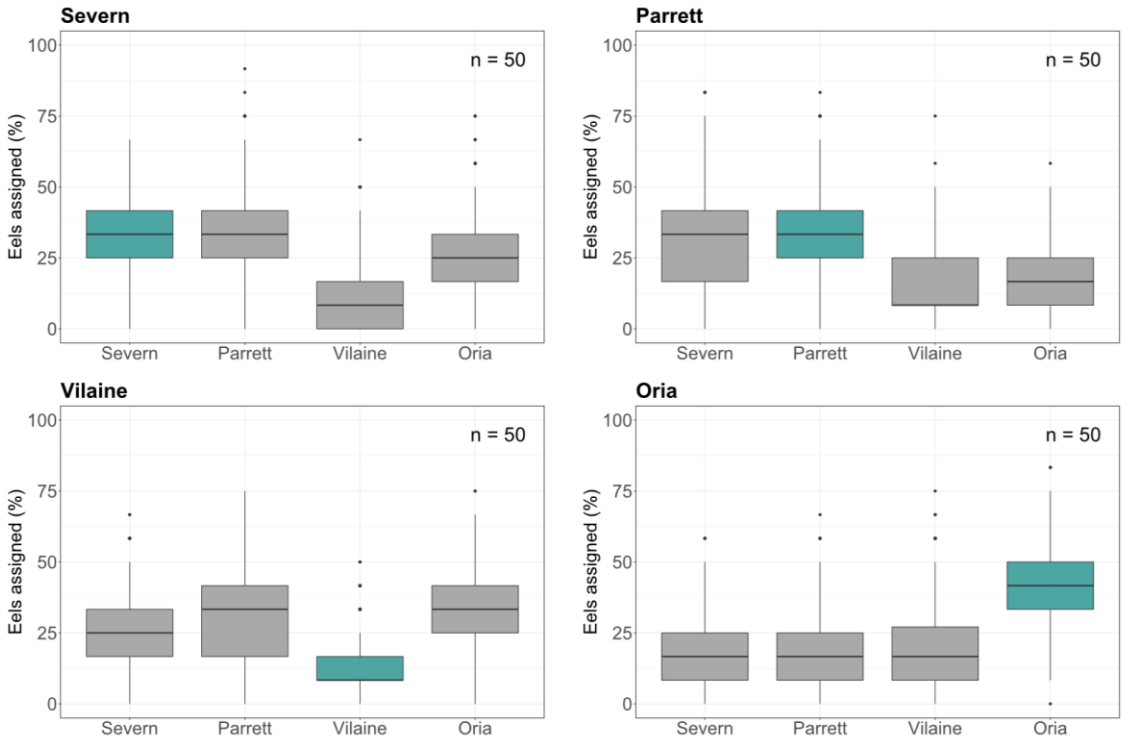


Figure 4.11 Assignment results using only carbon and nitrogen stable isotope data, showing the percentage of individuals from each known location assigned to all the possible regions over 1000 repeat simulations. The coloured boxes show the correct regions of origin.

Glass eels were assigned to their most likely river of origin based on the carbon, nitrogen and sulfur stable isotope composition using three assignment techniques - linear discriminant analysis and random forest classification as well as the multivariate normal probability distribution technique. The assignment accuracies achieved for each river of origin using these two additional techniques are detailed in Tables 4.8 and 4.9, and the accuracies obtained using the three approaches are compared in Table 4.7. The overall mean assignment accuracy was very similar using all the techniques, although multivariate analysis gave a slightly lower success rate (50% compared with 52% for the other methods). Linear discriminant analysis (LDA) and random forest classification both achieved a relatively high assignment accuracy of 74% for the glass eels from the River Oria, and it was only slightly lower using the multivariate analysis (71%). LDA performed best in the assignment of eels from the River Severn, achieving 71% assignment accuracy. The Rivers Severn and Oria were the most distinct from one another, with very few samples being misassigned to each other. Linear discriminant analysis using the stable isotope data resulted in some evidence of clustering by river of origin, but samples from all rivers show large amounts of overlap with each other (Figure 4.12). Therefore, it was challenging to assign the glass eels to river of origin correctly with all three assignment techniques and relatively low success rates were obtained, though eels from certain locations were more easily distinguished from one another.

Table 4.7 Comparison of assignment accuracy achieved for each river of origin based on the stable isotope data using multivariate normal probability distributions, linear discriminant analysis and random forest classification. For multivariate analysis and linear discriminant analysis, the mean accuracy over 1000 simulations is shown, whereas for random forest leave-one-out cross validation was used.

River of origin	Assignment accuracy (%)		
	Multivariate analysis	Linear Discriminant Analysis	Random forest
Severn, UK	63.8	71.0	46.0
Parrett, UK	25.5	31.6	32.0
Vilaine, France	37.7	32.2	54.0
Oria, Spain	70.8	74.1	74.0
Mean accuracy	**49.5%**	**52.2%**	**51.5%**

Table 4.8 Assignment accuracies achieved with linear discriminant analysis (LDA) and 1000 repeat simulations, using carbon, nitrogen and sulfur stable isotopes for each sampled river. The correct assignments to origin are shown in bold.

Assigned origin	**True origin river – percentage assigned (%)**			
	Severn, UK	Parrett, UK	Vilaine, France	Oria, Spain
Severn, UK	**71.0**	27.7	20.4	2.0
Parrett, UK	14.7	**31.6**	27.2	17.4
Vilaine, France	10.0	13.5	**32.2**	6.6
Oria, Spain	4.4	27.2	20.2	**74.1**

Table 4.9 Assignment accuracies achieved with leave-one-out random forest classification, using carbon, nitrogen and sulfur stable isotope values for each sampled river. The correct assignments to origin are shown in bold.

Assigned origin	**True origin river – number assigned**			
	Severn, UK	Parrett, UK	Vilaine, France	Oria, Spain
Severn, UK	**23**	9	10	2
Parrett, UK	10	**16**	7	6
Vilaine, France	16	11	**27**	5
Oria, Spain	1	14	6	**37**
Percentage correct	**46%**	**32%**	**54%**	**74%**

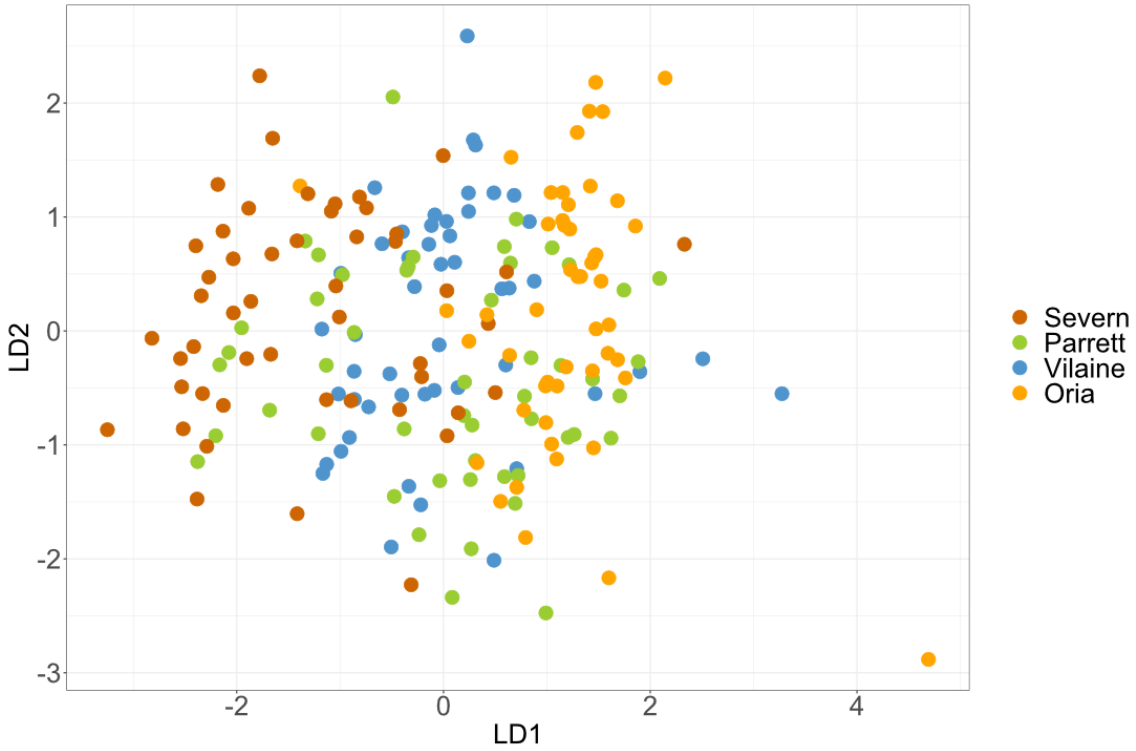


Figure 4.12 Linear discriminant analysis (LD1 and LD2) using the carbon, nitrogen and sulfur stable isotope compositions measured in eel muscle tissue from each of the four sampled rivers.

Since the lipid uncorrected $\delta^{13}C$ values and carbon content showed variation by river of capture (Figures 4.4 and 4.3), assignments to origin were also conducted using these variables in addition to or instead of the lipid corrected $\delta^{13}C$, $\delta^{15}N$ and $\delta^{34}S$ values. Firstly, random forest classification using the uncorrected $\delta^{13}C$, $\delta^{15}N$ and $\delta^{34}S$ values was performed, which resulted in an overall mean classification success of 57%, slightly higher than with the lipid corrected $\delta^{13}C$ values. The assignment accuracies achieved for each river are shown in Table 4.10. Assignment accuracies for the Rivers Severn and Vilaine were improved using the uncorrected $\delta^{13}C$ values compared to using those that had been lipid corrected (increases of 10-16% to 62% and 64% accuracy respectively), whereas the accuracy for the River Oria was reduced to 66% using the uncorrected $\delta^{13}C$ values.

Following this, the assignment accuracy was investigated using the carbon content of the eels as well as the $\delta^{13}C$ (lipid corrected), $\delta^{15}N$ and $\delta^{34}S$ values, and the results are shown in Table 4.11. The overall mean assignment accuracy was 62%, higher than previously achieved using other combinations of variables. Assignment accuracies for the individual rivers were generally similar to those using the uncorrected $\delta^{13}C$, $\delta^{15}N$ and $\delta^{34}S$ values in Table 4.10, but the accuracy for the River Oria was significantly improved, reaching a classification success of 82% (an increase of 16%).

Table 4.10 Assignment accuracies achieved with leave-one-out random forest classification, using the carbon (uncorrected), nitrogen and sulfur stable isotope values for each sampled river. The correct assignments to origin are shown in bold.

Assigned origin	**True origin river – number assigned**			
	Severn, UK	Parrett, UK	Vilaine, France	Oria, Spain
Severn, UK	**31**	6	5	3
Parrett, UK	6	**18**	8	11
Vilaine, France	11	10	**32**	3
Oria, Spain	2	16	5	**33**
Percentage correct	**62%**	**36%**	**64%**	**66%**

Table 4.11 Assignment accuracies achieved with leave-one-out random forest classification, using carbon (lipid corrected), nitrogen and sulfur stable isotope values as well as the carbon content (%) for each sampled river. The correct assignments to origin are shown in bold.

Assigned origin	**True origin river – number assigned**			
	Severn, UK	Parrett, UK	Vilaine, France	Oria, Spain
Severn, UK	**34**	10	8	1
Parrett, UK	7	**18**	9	6
Vilaine, France	8	10	**30**	2
Oria, Spain	1	12	3	**41**
Percentage correct	**68%**	**36%**	**60%**	**82%**

4.4.3.2 Assignment using fatty acids

Glass eels were assigned to their most likely origin based on the fatty acid composition using random forest classification and linear discriminant analysis (LDA). The results of these assignments are compared in Table 4.12 and listed in more detail in Tables 4.13 and 4.14. The overall assignment accuracies using both of these techniques were low – 32% correct for random forest and 31% for LDA, which are not much better than random assignment, suggesting that fatty acids contain little geographic information. Using both techniques, eels from the River Oria could be assigned with the greatest accuracy, although random forest classification achieved a significantly higher success rate (78%) compared with LDA (59%). Random forest classification was also able to assign a greater proportion of eels from the River Vilaine correctly, achieving a success rate of 40%, whereas only 21% were correctly assigned using LDA. LDA performed better for eels from the River Severn and Parrett, since this method assigned 13% and 31% correctly respectively compared to 0% and 10% using random forest. However, these assignment

accuracies are still very low, indicating that fatty acid analysis cannot reliably distinguish among glass eels from these rivers. The only exception is those from the River Oria when assigned using random forest classification, for which the assignment accuracy was 78%.

Linear discriminant analysis resulted in clear clustering of the eel samples by river of origin (Figure 4.13) despite the low assignment accuracy. Eels from the Rivers Parrett and Oria are distinct and do not overlap any other clusters, but the Severn and Vilaine clusters overlap completely and cannot be distinguished.

Table 4.12 Comparison of assignment accuracy for each river using the fatty acid data with linear discriminant analysis and random forest classification. For linear discriminant analysis the mean accuracy over 1000 simulations is shown, whereas random forest used leave-one-out cross validation.

River of origin	Assignment accuracy (%)	
	Linear discriminant analysis	Random forest
Severn, UK	12.9	0.0
Parrett, UK	30.9	10.0
Vilaine, France	20.5	40.0
Oria, Spain	59.3	77.8
Mean accuracy	**30.9%**	**32.0%**

Table 4.13 Assignment accuracies achieved with linear discriminant analysis (LDA) and 1000 repeat simulations using the fatty acid data for each sampled river. The correct assignments to origin are shown in bold.

Assigned origin	True origin river – percentage assigned (%)			
	Severn, UK	Parrett, UK	Vilaine, France	Oria, Spain
Severn, UK	**12.9**	25.5	31.9	13.8
Parrett, UK	22.0	**30.9**	29.1	14.4
Vilaine, France	37.0	27.2	**20.5**	12.6
Oria, Spain	28.2	16.5	18.5	**59.3**

Table 4.14 Assignment accuracies achieved with leave-one-out random forest classification, using the fatty acid data for eels from each sampled river. The correct assignments to origin are shown in bold.

Assigned origin	**True origin river – number assigned**			
	Severn, UK	Parrett, UK	Vilaine, France	Oria, Spain
Severn, UK	**0**	7	2	1
Parrett, UK	5	**1**	2	0
Vilaine, France	3	0	**4**	1
Oria, Spain	1	2	2	**7**
Percentage correct	**0%**	**10%**	**40%**	**78%**

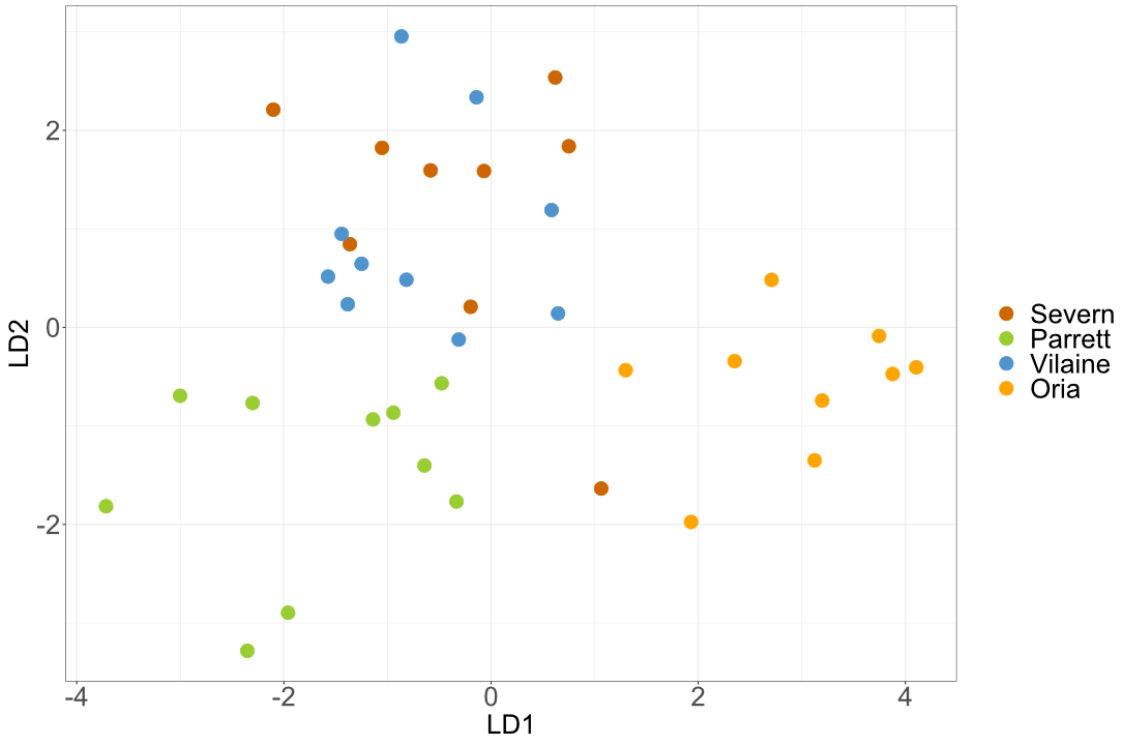


Figure 4.13 Linear discriminant analysis (LD1 and LD2) using the fatty acid compositions measured in eel muscle tissue from each of the four sampled rivers.

The variable importance plot in Figure 4.14 illustrates that cholesterol, pentadecylic acid (15:0) and an unidentified polyunsaturated fatty acid were the most important for accurate assignment of eels to origin using random forest classification, resulting in a mean decrease in accuracy of 10.9, 6.5 and 5.5 respectively when each of these fatty acids were excluded. Therefore, there were relatively low mean decrease accuracies for all fatty acids, indicating that no individual fatty acid contributed that significantly to the classification accuracy. There was also no observable trend in either saturated, monounsaturated or polyunsaturated (PUFA) fatty acids being of more importance for discrimination. Several

fatty acids show negative importance, meaning that the assignment accuracy decreases overall when these fatty acids are included. The monounsaturated fatty acid 18:1 causes the greatest decrease in accuracy.

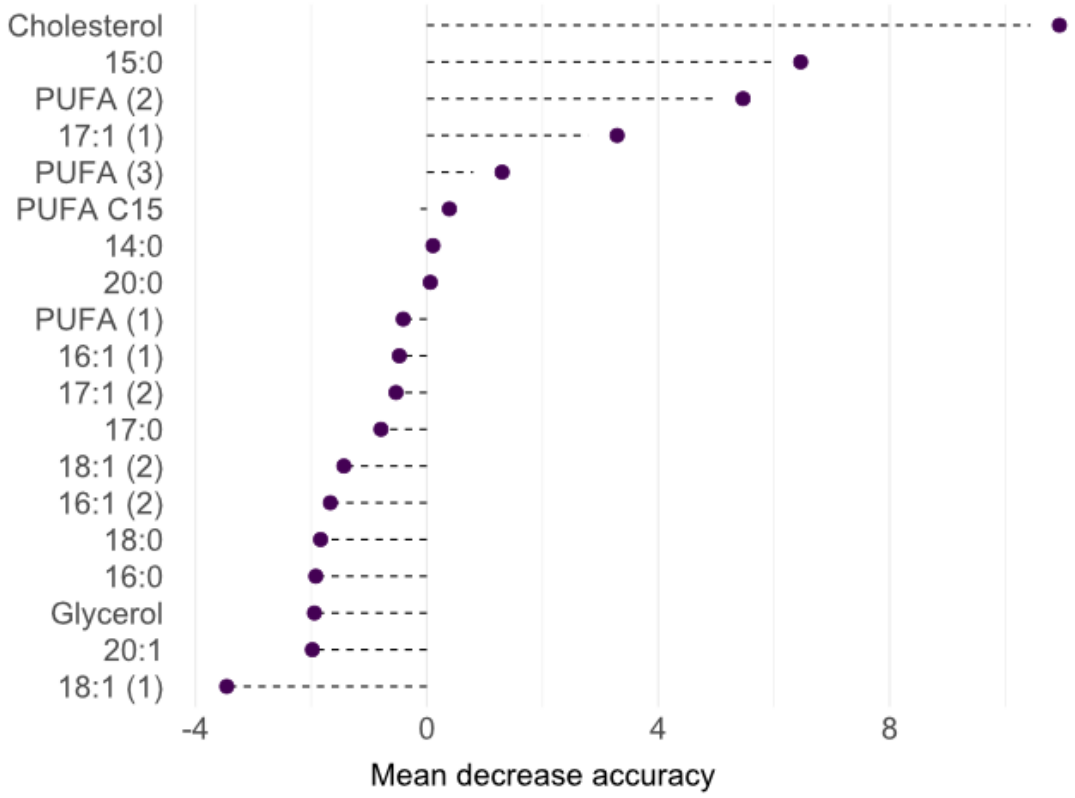


Figure 4.14 Random forest variable importance plot, illustrating which fatty acids contribute most to the correct classification of eels to their capture river. Mean decrease accuracy is the measure of performance of the model without each variable, so a higher value indicates that the fatty acid is important in predicting the river of origin and removal of that fatty acid causes the model to lose accuracy of assignment.

It was also investigated whether the assignment accuracy is increased when the River Severn and River Parrett are combined into one UK group, since these locations are in close geographic proximity and a number of glass eels were incorrectly assigned to the other UK river. Glass eels were assigned to their most likely origin among the UK rivers, River Vilaine and River Oria using random forest classification, and the results are shown in Table 4.15. The accuracy of assignment was increased for the UK rivers combined, achieving 74% correct classification. However, the success rate for the River Vilaine and River Oria were reduced compared to the previous assignments – only 10% of eels from the River Vilaine were assigned correctly and 67% of those from the River Oria. Therefore, combining the two UK rivers into one group increases the assignment accuracy for eels from the UK, but at the detriment of the other two locations.

Table 4.15 Assignment accuracies achieved with leave-one-out random forest classification, using the fatty acid data for eels from each sampled river when the River Severn and River Parrett are combined into one group. The correct assignments to origin are shown in bold.

Assigned origin	True origin river – number assigned		
	UK (Severn & Parrett)	Vilaine, France	Oria, Spain
UK (Severn & Parrett)	**14**	9	3
Vilaine, France	2	**1**	0
Oria, Spain	3	0	**6**
Percentage correct	**74%**	**10%**	**67%**

4.4.3.3 Assignment using stable isotopes and fatty acids combined

Lastly, the accuracy of assignment to origin using stable isotope and fatty acid data combined was investigated. Both LDA and random forest classification were used to assign the eels to most likely river of origin, and the results of this are compared in Table 4.16. Tables 4.17 and 4.18 show these assignment results in more detail. The overall mean assignment accuracy was low using both classification techniques (44% for random forest and 34% for LDA). Random forest classification failed to correctly assign any eels from the River Parrett. However, this method achieved a high success rate of 89% for the River Oria eels. The River Severn eels were assigned with moderate accuracy (56% correct), with the majority of incorrect assignments being made to the other UK river, the River Parrett. Few misassignments occurred between the Rivers Severn and Oria, indicating that these locations are more easily distinguished from each other. Using LDA, the highest assignment accuracy was 57% for the River Oria, although this is still significantly lower than that achieved using random forest. However, LDA assigned eels from the River Parrett with 32% accuracy where random forest was not able to assign any eels correctly. Therefore, even when stable isotope and fatty acid data are used in conjunction glass eels from these rivers cannot be reliably differentiated, with the exception of eels from the River Oria which can be classified to origin with high accuracy using random forest.

The variable importance plot in Figure 4.15 illustrates that $\delta^{34}S$, cholesterol and $\delta^{15}N$ contributed the most to the accurate assignment of eels to origin using random forest classification, resulting in a mean decrease in accuracy of 39.5, 10.1 and 7.4 respectively when each of these variables were excluded. Therefore, sulfur stable isotope values were the most important variable by far for the correct classification of eels. However, several variables show negative importance, indicating that the assignment accuracy decreases

overall when these are included. The greatest decrease in accuracy is caused by including lipid corrected $\delta^{13}C$ values and glycerol.

Table 4.16 Comparison of assignment accuracy for each river using random forest classification and linear discriminant analysis, with both stable isotope and fatty acid data combined. Random forest used leave-one-out cross validation and for linear discriminant analysis the mean accuracy over 1000 simulations is shown.

River of origin	Assignment accuracy (%)	
	Linear Discriminant Analysis	Random forest
Severn, UK	28.9	55.6
Parrett, UK	31.9	0.0
Vilaine, France	18.8	30.0
Oria, Spain	57.1	88.9
Mean accuracy	**34.2%**	**43.6%**

Table 4.17 Assignment accuracies achieved with linear discriminant analysis (LDA) and 1000 repeat simulations using the fatty acid and stable isotope data combined for each sampled river. The correct assignments to origin are shown in bold.

Assigned origin	True origin river – percentage assigned (%)			
	Severn, UK	Parrett, UK	Vilaine, France	Oria, Spain
Severn, UK	**28.9**	27.3	34.1	7.7
Parrett, UK	25.6	**31.9**	32.4	16.3
Vilaine, France	32.4	23.8	**18.8**	19.0
Oria, Spain	13.2	17.1	14.8	**57.1**

Table 4.18 Assignment accuracies achieved with leave-one-out random forest classification, using the fatty acid and stable isotope data combined for eels from each sampled river. The correct assignments to origin are shown in bold.

Assigned origin	True origin river – number assigned			
	Severn, UK	Parrett, UK	Vilaine, France	Oria, Spain
Severn, UK	**5**	4	0	0
Parrett, UK	3	**0**	7	0
Vilaine, France	1	4	**3**	1
Oria, Spain	0	2	0	**8**
Percentage correct	**56%**	**0%**	**30%**	**89%**

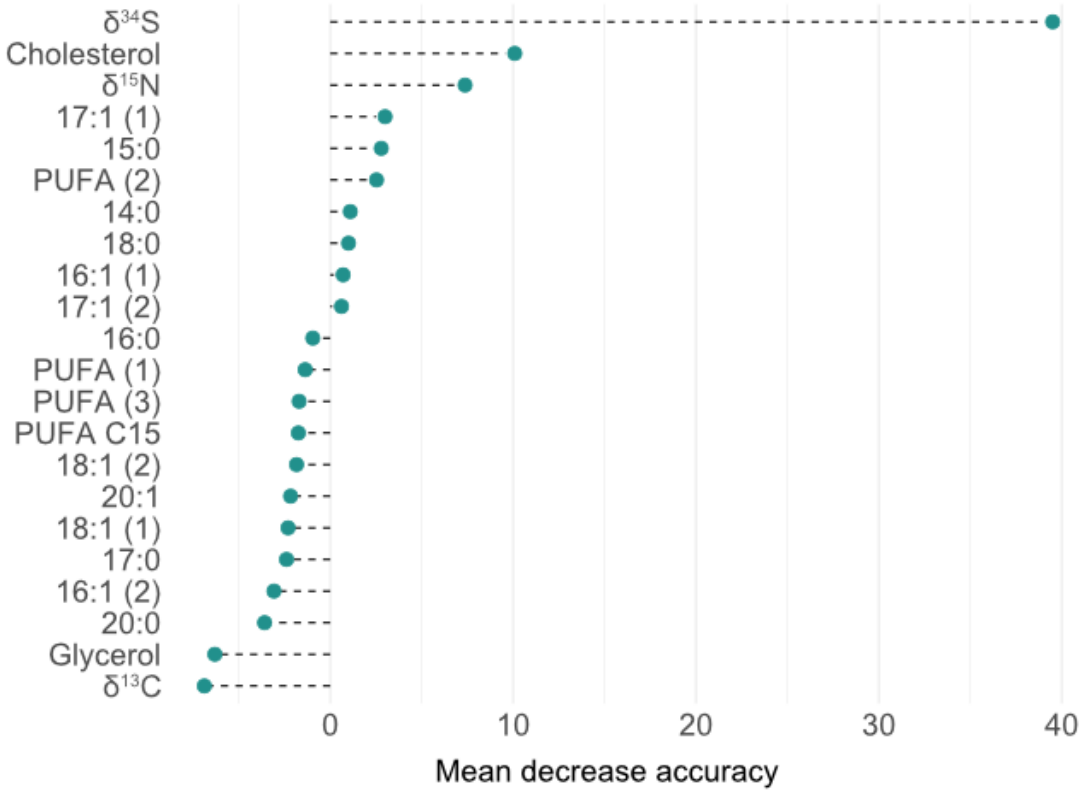


Figure 4.15 Random forest variable importance plot, illustrating which stable isotopes and fatty acids contribute most to the correct classification of eels to their capture river. Mean decrease accuracy is the measure of performance of the model without each variable, so a higher value indicates that the variable is important in predicting the river of origin and removal of that variable causes the model to lose accuracy of assignment.

Since a number of variables were found to have a negative mean decrease accuracy using random forest classification (Figure 4.15), it was investigated whether removing these variables from the assignment would improve the classification success. The carbon content was found to be a useful variable in the assignment to origin previously (Table 4.11), so this was also included in the random forest classification. The results of this assignment using the selected fatty acids (14:0, 15:0, 16:1 (1), 17:1 (1), 17:1 (2), 18:0, PUFA 2, cholesterol), stable isotope data ($\delta^{15}N$ and $\delta^{34}S$) and carbon content (%) showed that overall 54% of eels could be correctly assigned to river of origin. The results for each of the rivers individually are shown in Table 4.19. For this assignment, $\delta^{34}S$ values were even more important for the correct classification to origin, with a mean decrease accuracy of 57.9. The $\delta^{15}N$ values and cholesterol abundance were also important, but the mean decrease accuracies were nowhere near as high as that for $\delta^{34}S$ (9.9 and 8.4 respectively). The assignment accuracy for eels from the River Severn was significantly higher than the assignment using all the fatty acids and stable isotopes in Table 4.18, with a success rate of 78% compared to 56% (Table 4.19). The assignment accuracy for the River Vilaine was also increased (50% compared to 30% in the previous assignment). Eels from the River

Oria could be assigned with high confidence in both assignments (89% accuracy). However, none of the River Parrett eels were correctly assigned, as was the case in the assignment using all the fatty acids and stable isotopes.

Table 4.19 Assignment accuracies achieved with leave-one-out random forest classification, using the fatty acid, stable isotope and carbon content data combined for eels from each sampled river, but with the variables found to have a negative mean decrease accuracy removed. The correct assignments to origin are shown in bold.

Assigned origin	**True origin river – number assigned**			
	Severn, UK	Parrett, UK	Vilaine, France	Oria, Spain
Severn, UK	**7**	4	0	0
Parrett, UK	2	**0**	4	0
Vilaine, France	0	4	**5**	1
Oria, Spain	0	2	1	**8**
Percentage correct	**78%**	**0%**	**50%**	**89%**

Finally, assignment to origin was conducted using the River Severn and River Oria eels only, since these were the most distinct. Random forest classification was performed based on the fatty acid, stable isotope and carbon content data combined (using only selected fatty acids from importance plot in Figure 4.15) for the eels from these two rivers. Eels could be assigned with 100% correct classification to either the River Severn or River Oria (Table 4.20).

Table 4.20 Assignment accuracies achieved with leave-one-out random forest classification, using the fatty acid, stable isotope and carbon content data combined (using only selected fatty acids from importance plot in Figure 4.15) for eels from the River Severn and River Oria only. The correct assignments to origin are shown in bold.

Assigned origin	**True origin river – number assigned**	
	Severn, UK	Oria, Spain
Severn, UK	**9**	0
Oria, Spain	0	**9**
Percentage correct	**100%**	**100%**

4.4.3.4 Comparison of stable isotopes and fatty acids for spatial traceability of eels

Table 4.21 shows a comparison of the assignment accuracies achieved with random forest classification using the stable isotope data (including carbon content) and fatty acid compositions individually, as well as using both approaches in combination (using only selected fatty acids from importance plot in Figure 4.15). This demonstrates that stable isotope analysis gave the highest overall assignment accuracy of 62%, while fatty acid analysis gave the lowest accuracy of 32%. In all three cases, the eels from the River Oria could be assigned most reliably, with a success rate of at least 78%. Combining both approaches further increased the success rate for the River Oria, reaching a high assignment accuracy of 89%. The River Severn eels could be assigned to origin most successfully using the combination of stable isotope and fatty acid data, achieving a success rate of 78%, whereas using either fatty acid analysis or stable isotope approaches individually resulted in lower assignment accuracies of 68% at most. For eels from the remaining locations (Parrett and Vilaine), relatively low assignment accuracies were obtained using all techniques and classification methods. A maximum of 60% assignment accuracy was achieved for the River Vilaine and 36% for the River Parrett.

Within the context of low overall differentiation, stable isotope analysis with carbon content data performed the best on average, since this method achieved more similar accuracies among regions, whereas fatty acids or the combination of approaches had a larger range of success rates from 0% to ~80-90%. However, for certain rivers such as the River Oria and River Severn it is best to combine stable isotope and fatty acid analysis to achieve the highest success rates. Although the results demonstrate that there is little opportunity to identify the origin of eels across multiple sites, under certain circumstances there may be value in these traceability approaches, for example if the claimed origin of glass eels is the River Severn or Oria then it may be possible to test for provenance. The Rivers Oria and Severn were the most distinct from each other, with low numbers of incorrect assignments to each other, so this shows more promise in traceability and may achieve a high enough accuracy to have practical use.

Table 4.21 Comparison of assignment accuracy for each river based on the stable isotope composition, fatty acid composition and the combination of both approaches combination (using only selected fatty acids from importance plot in Figure 4.15), using random forest classification with a leave-one-out approach.

River of origin	Assignment accuracy (%)		
	Stable isotope analysis with carbon content	Fatty acid analysis	Stable isotopes with carbon content & fatty acids
Severn, UK	68.0	0.0	77.8
Parrett, UK	36.0	10.0	0.0
Vilaine, France	60.0	40.0	50.0
Oria, Spain	82.0	77.8	88.9
Mean accuracy	**61.5%**	**32.0%**	**54.2%**

4.5 Discussion

The findings of this study demonstrate that the stable isotope and fatty acid compositions of European glass eels are similar among rivers of origin, and therefore eels at this stage of migration generally cannot be reliably assigned to capture river using these techniques. However, eels from the River Oria and the River Severn were relatively distinct due to differing $\delta^{34}S$ values and cholesterol abundances, achieving up to 89% assignment accuracy for the River Oria and up to 78% accuracy for the River Severn, which gives potential for discrimination under specific circumstances. Eels caught in the other two rivers had low assignment accuracies of 0-60%, depending on the river and assignment method applied.

These results suggest that the eels had not been present in the river of origin for long enough for the stable isotope or fatty acid signatures to be incorporated into the tissues or that they do not feed within the estuary. Their catadromous life history means that, after hatching in the Sargasso Sea, the eel larvae travel across the Atlantic Ocean and enter river systems as glass eels (Schmidt, 1923). It was upon entering the river mouths that glass eels were collected for this study, and therefore may have only spent a short time relative to the tissue turnover time inhabiting the river system before being caught. The similar isotopic and fatty acid compositions of the eels from all sampled rivers suggest that the muscle tissue and lipid content were largely synthesised during growth in the open ocean before entering the distinct estuaries, while eels from all four locations were sharing a common migration route.

A study by Bardonnet and Riera (2005) measured the stable isotope composition of glass eels to investigate their feeding behaviour in estuaries. The carbon isotopic values of glass

eels in Bardonnet and Riera (2005) were -21.8 to -21.0‰ for those caught in marine waters, indicating a diet of primarily oceanic plankton. Glass eels sampled in estuaries had very similar $\delta^{13}C$ values to this, between -22 and -20.5‰, which suggests that the glass eels did not feed significantly since entering the estuary. Bardonnet and Riera (2005) discovered that only 15% of eels appeared to have ingested food from terrestrial sources, with the majority of glass eels retaining their isotopic signature from marine waters. In the current study, the average $\delta^{13}C$ value of glass eels from all rivers was around -20‰, with most values between -19 and -21‰. The similarity of these values to those of the marine and estuarine eels in Bardonnet and Riera (2005) further supports the hypothesis that the glass eels did not feed since they left oceanic waters and had retained their marine isotopic signatures, creating challenges the use of stable isotope analysis for tracing the origin of glass eels. Therefore, it may not be possible to use stable isotope analysis for traceability of glass eels from estuaries, but if eels were sampled after travelling further up-river then biochemical techniques may provide better distinguishing power among those from different locations.

Glass eels from the River Oria, Spain, had the highest mean carbon content and mean C:N ratio of all the eels, combined with the lowest $\delta^{13}C$ values (not lipid corrected), likely reflecting higher lipid contents that could possibly be related to a shorter migration from the Sargasso Sea. River Oria eels also had lower cholesterol abundance and $\delta^{34}S$ values than those from the other three sampled rivers. This may indicate that glass eels from the River Oria followed a different path across the Atlantic than other eels and hence incorporated stable isotope signatures from their diet in different locations. Alternatively, they may have followed the same path initially but then their route split from that of other eels due to the more southerly location of the River Oria and so a portion of the journey was spent travelling independently from the eels destined for the UK or France before entering the estuary. However, it is also possible that the greater differentiation of the River Oria eels is because of different dietary preferences.

Cholesterol, the only fatty acid found to significantly differ in the glass eels across locations, has a vital role in maintaining the integrity and fluidity of cell membranes, and reducing membrane permeability. It is also a key structural component of muscle, the nervous system and brain (Zampelas and Magriplis, 2019, Norambuena et al., 2013). Animals, including marine teleost fish, synthesise cholesterol in the liver as well as obtaining it from the diet, since it is present in food items of animal origin (Leaver et al., 2008, Xu et al., 2020). The lower abundance of cholesterol present in the River Oria eels could therefore be a result of differences in diet, although this seems unlikely since glass eels are thought to feed on primary producers (plankton and POM) (Bardonnet and Riera, 2005), which only contain very small amounts of cholesterol (Behrman and Gopalan, 2005). An alternative

explanation is that variation in the physiology of eels from different populations, and therefore from different gene pools, causes the differing cholesterol levels.

4.5.1 Comparison between stable isotope and fatty acid analyses

Stable isotope analysis was overall more effective than fatty acid analysis at distinguishing among glass eels caught in different locations, achieving a mean assignment accuracy of 62% compared with 32% using fatty acids. For the rivers Severn, Parrett and Vilaine, stable isotope compositions were able to assign eels to origin with a significantly higher success rate than using fatty acids. However, both techniques could more reliably assign eels from the River Oria to origin (78-82% accuracy). Fatty acids have been found to have limited success for spatial traceability in several other studies, for example Busetto et al. (2008) were not able to differentiate between wild turbot from Denmark and the Netherlands using fatty acids alone and Turchini et al. (2008) also found that fatty acids could not distinguish among Murray cod from different farms. Fonseca et al. (2022), however, had more success and classified the bivalve *Scrobicularia plana* and European seabass to three locations within an estuary with a classification accuracy of 80-100% using fatty acid profiling. Therefore, it may be that fatty acids alone are not a suitable tool for distinguishing the origin of eels. It appears that stable isotope ratios in glass eels, particularly $\delta^{34}S$, show greater spatial variation than fatty acids for the sampled rivers. An alternative explanation is that the fatty acid signature took longer to be incorporated into the eels' tissues, and so still reflected that of the open ocean prior to entering the estuaries.

Combining the two approaches resulted in greater assignment accuracy for the River Oria (89% correct), which is high enough that it could have tangible use in traceability applications. The assignment accuracy for the River Severn also increased to 78% using both techniques together. Sulfur isotope values were by far the most important variable for correct classification. Assigning only the River Oria and River Severn eels to origin using the techniques combined resulted in a success rate of 100% for both rivers, further suggesting that these approaches could offer distinguishing power for certain more distinct rivers. Therefore, there is benefit in the simultaneous use of fatty acids and stable isotopes for the spatial traceability of glass eels in this study. This was also demonstrated by Go et al. (2022), who discovered that for Manila clams collected at different sites around Korea, the use of fatty acids in conjunction with stable isotope analysis improved assignment accuracies compared with using stable isotopes alone (success rates of 98% and 77% respectively). However, for the River Parrett and River Vilaine the combined use of these two techniques did not offer any advantage. In fact, for these two rivers the assignment accuracies were decreased compared to those using stable isotopes independently, suggesting that eels become more similar across rivers when both datasets are merged and that some distinguishing ability is lost.

4.5.2 Comparison with other techniques

This study formed part of a wider project trialling a range of different techniques to trace the spatial origin of glass eels, including trace element analysis, strontium isotope analysis and metabolomics in addition to the stable isotope and fatty acid analyses used here (Stein et al., 2019). Results of this project demonstrated that using multi-elemental analysis of the eel bodies, 75.8% of individuals could be correctly assigned to river of origin. Strontium isotope ratios ($^{87}Sr/^{86}Sr$) in eel bodies were also significantly different among locations, with the highest values in those from the River Vilaine, but the differences were within measurement uncertainties. When the strontium isotope values were combined with the elemental data, the assignment accuracy increased to 91.4% and reached almost 99% if the two UK rivers were grouped together. However, when the UK rivers were kept separate, the eels from the River Severn had a low classification success of 42.9%. Using the eel otoliths, discrimination using trace element analysis was not possible and no significant difference was observed in the strontium isotope ratios, likely because the residence time in the estuary was not long enough for the incorporation of local chemical signatures into the otolith. Metabolomics data revealed good separation of eels from the four rivers using discriminant analysis, and 100% correct classification was achieved with only three variables using stepwise discriminant analysis, although this was based on very low sample numbers (five eels from each location). Both trace element and metabolomic approaches therefore appear more effective than stable isotope or fatty acid analyses when used independently, although the accuracy for eels from the River Oria came close using stable isotopes and fatty acids combined (89% correct). Fatty acid analysis was one of the least successful methods trialled. However, carbon stable isotope ratios added discriminatory power when combined with the elemental data, increasing the overall mean assignment accuracy from 75.8% to 80.1%, indicating that stable isotopes may have value when used in conjunction with other methods. Even using metabolomics, which had the highest assignment accuracy overall, it was more challenging to differentiate between eels from the River Severn and River Vilaine without including the carbon stable isotope data. This suggests that stable isotopes may have potential importance in contributing to the spatial assignment of glass eels when combined with other approaches despite it not being the most successful method when used alone.

4.5.3 Limitations and future research

Limitations in the use of stable isotope analysis for spatial traceability arise from the movement of individuals. Highly migratory species may travel across multiple different regions, resulting in an isotopic signature that is integrated across all of these regions and will be difficult to interpret, especially in species with a slow tissue turnover time. This means that if glass eels are caught during the migration phase, just after entering the

estuary, their isotopic signature will be a mixture of all recent feeding locations. Therefore, ideally glass eels should be allowed to remain in the river for a period of time before sampling so that their tissue isotopic composition reflects that of their current location, otherwise it will be very difficult to assign eels reliably. Fatty acids also require time to become incorporated into synthesised tissues and so may not show geographic variation if sampled too soon upon entering the estuary. However, in practice glass eels are caught on their entry migration into the river for use in aquaculture, and so would not have any extra time to acquire the local biochemical signatures than the eels in this study had. Moreover, traceability can be further complicated if glass eels are fed while in transit in holding tanks, since the stable isotope or fatty acid composition of the eels will be altered from their riverine signature and eels from all locations will become more similar based on the composition of the feed. Consequently, the potential to infer origin will reduce with time since capture. These limitations mean that unless a more rapidly assimilated tracer with larger variations among locations can be found, it is unlikely that discrimination at a suitable level of accuracy can be achieved, except in specific cases such as differentiating the River Oria from the River Severn. Since trace element concentrations of glass eel bodies gave promising results in Stein et al. (2019), it may be beneficial to investigate the use of hand-held X-ray fluorescence devices to analyse elemental abundance (as demonstrated for cod in Chapter 5).

Since the heavier isotopes of nitrogen, and to a lesser extent of carbon, increase with trophic level, $\delta^{15}N$ and $\delta^{13}C$ values typically increase systematically with body size due to the typical size structuring of marine food webs (Jennings et al., 2001, Jennings et al., 2008). Therefore, for stable isotope analysis to be applied reliably for spatial traceability, eels of unknown origin should be of the same body size as those in the reference dataset, or a correction may be needed to account for trophic level effects if individuals of different sizes are used.

Temporal variation in baseline stable isotope ratios also introduces uncertainty in assignment, since the stable isotope composition of phytoplankton at the base of the food web may vary seasonally or annually (Magozzi et al., 2017). Similarly, fatty acids in phytoplankton also show temporal variations due environmental factors and changes in the plankton species composition (Reuss and Poulsen, 2002), which result in variations in the fatty acid signatures of fish over time (Budge et al., 2002). For terrestrial food products, new reference data are collected each year from all geographic regions for this reason (Camin et al., 2017). Therefore, for glass eels reference data should also be acquired during the same timeframe as the test samples to enable the most reliable assignment to capture region.

A further limitation of the current study is that whole glass eel bodies were analysed, including the muscle tissue and skin (although avoiding the internal organs), since the small

body size of all the eels made removing the skin impractical. This potentially increased the complexity of interpretation due to the differing protein or lipid compositions of the skin and muscle tissues.

In this study, glass eels from Morocco were sampled but could not be provided to the laboratory for analysis due to logistical constraints. European eels have a distributional range from northern Norway to the coast of north Africa in the Atlantic, as well as in the Baltic, Mediterranean and Black Seas (Righton and Metcalfe, 2011). Future investigation into the ability to discriminate among glass eels from a wider range of locations would be beneficial, since locations that are situated greater distances apart than those in this study may be more biochemically distinct. This may be of particular importance if glass eels from the EU could be distinguished from those of non-EU origin, since the trade of European eels to and from the EU is prohibited due to its critically endangered status (European Commission, 2019). A forensic test to verify the provenance of eels claimed to originate from outside of Europe would allow illegal trade from the EU to be detected.

4.5.4 Conclusions

The illegal trade of European glass eels (*Anguilla anguilla*) is a serious threat to the survival of the species. A forensic test for the geographic origin of glass eels would be highly beneficial for the detection of illegal exports and the enforcement of regulations, as well as enabling sustainable management of fisheries. The findings of this study demonstrate that neither the fatty acid nor stable isotope composition of glass eels varied significantly with river of origin, with the exception of those from the River Oria in Spain. Overall assignment accuracies were low using both techniques individually and combined together (32-62%), although stable isotope analysis was the most effective technique of the two and combining approaches increased the accuracy further for certain rivers. Glass eels from the River Oria were sufficiently distinct to be assigned with up to 89% accuracy, and those from the River Severn could be assigned with 78% accuracy. These results encourage further investigation into the use of stable isotope analysis as a complementary technique, since it has potential to provide additional discriminatory power to other more successful approaches, particularly for certain locations, even though it is not effective enough to be used as a stand-alone test.

Chapter 5 Can high-throughput XRF approaches successfully distinguish among cod of different origins: a comparison of two instruments

5.1 Abstract

Fish and seafood are known to be at risk of fraud and mislabelling, which undermines sustainable management efforts and causes economic losses. Identifying the geographic origin of products is vital for combatting this, but tracing the catch location of fish and seafood is much more challenging than identifying the species in products. Trace element analysis has been demonstrated to successfully distinguish among marine species from different origins. However, so far very few studies have investigated the use of x-ray fluorescence (XRF), particularly hand-held devices, for analysing the elemental composition of fish and seafood for spatial traceability, despite the clear advantages of this technology. In this study, the potential for using XRF technology for tracing the provenance of Atlantic cod (*Gadus morhua*) is explored, based on the multi-elemental profile of the muscle tissue. I analysed the elemental composition of 225 cod muscle samples from ten catch regions across the Northeast Atlantic using two different XRF instruments - the Itrax core scanner and the Vanta handheld analyser - opportunistically using cod samples collected for a previous chapter. Individual cod could be traced back to known origin with relatively high accuracy using both instruments (83-86%), with potassium and arsenic found to be the most important elements for correct classification. Cod from the Barents Sea and Faroes were the most distinct and were assigned to origin with a 100% success rate. Using a combination of XRF and stable isotope data measured in the same cod samples (from Chapter 2), the assignment accuracy increased to 91% which is greater than using either of the methods independently. The results of this study demonstrate that XRF instruments have promising potential to differentiate among whitefish muscle samples sourced from different catch regions, encouraging further investigation into its use as a forensic test for origin. The Vanta hand-held analyser shows particular promise for use in commercial settings due to its portability and real-time results.

5.2 Introduction

Demand for fish and seafood products is rising globally, which means traceability of marine products throughout retail chains is becoming ever more important as trade networks increase in complexity and fish stocks face increased exploitation pressure (FAO, 2020b).

Several techniques have been explored to determine the catch region of fish and seafood products with varying success (Gopi et al., 2019a), but currently no widely accepted test for spatial origin exists. Genetic markers have been shown to be a powerful tool for discriminating among reproductively isolated populations and has been successfully applied to determine the origin of cod from several populations on a relatively large spatial scale. Stable isotope analysis has also been established as an effective technique for assigning a range of wild-caught species to their geographic origin (Carrera and Gallardo, 2017, Trueman et al., 2017, Zhang et al., 2017, Gong et al., 2018, Zhang et al., 2019, Slesser and Trueman, 2021, Cusa et al., 2021), including Atlantic cod (as in Chapter 2) for which individuals could be traced to origin with a similar accuracy as with genetic techniques in Nielsen et al. (2012b) for the same regions. Stable isotope compositions also have the potential to differentiate among regions on a smaller spatial scale than genetic techniques in certain cases.

Multi-elemental analysis provides a potential alternative or complimentary tool for geographic origin traceability. In contrast to stable isotopes, there are no a priori spatial predictions for trace element compositions, so spatial assignment depends on identifying multivariate clustering to which samples can be assigned, without prior expectation or understanding of the mechanisms leading to elemental differences among regions. This technique has been successfully applied to several marine species and taxonomic groups including fish (Yamashita et al., 2006, Gopi et al., 2019c), shrimps (Smith and Watts, 2009, Li et al., 2014, Carter et al., 2015, Ortea and Gallardo, 2015, Li et al., 2017, Gopi et al., 2019b), bivalves (Bennion et al., 2019, Morrison et al., 2019), octopus (Martino et al., 2022a) and sea cucumbers (Liu et al., 2012), but has more often been investigated using individuals from aquaculture farms than wild-caught samples. One of the relatively few studies to trace the origin of wild-caught as well as farmed individuals is that by Ortea and Gallardo (2015), where 93.5% of shrimps were assigned to their correct region or farm of origin based on the concentration of five elements (As, Cd, Pb, P and S). Morrison et al. (2019) also extended the use of trace element analysis to wild-caught samples and was able to classify king scallops to their site of harvest with a success rate of 82.75%, even though two of the sites were within the same bay and only 6km apart.

However, to have practical use on a large scale in the seafood industry, provenance methods must be rapid, high-throughput and portable. Although stable isotopes offer good regional separation for many marine species, the sample preparation and analysis are

time-consuming processes, demanding expertise both for the analytical and data handling components. The analysis also requires the use of a mass spectrometer, so samples must be sent to a laboratory for analysis to be undertaken by a trained technician. Multi-elemental analysis also traditionally suffers from these disadvantages, since the elemental concentrations are determined by ICP-MS or ICP-AES following digestion, which also both involve relatively laborious sample preparation and must be carried out in a laboratory by highly trained staff. The development of X-ray fluorescence (XRF) instruments has allowed much more rapid assessment of the elemental composition of samples and can acquire data for a large number of different elements simultaneously. They also confer the advantage of non-destructive sample analysis, and now include hand-held instruments that are portable and can be used at locations outside of the laboratory. This would allow analyses to be conducted at the point of interception of a product by less experienced personnel and with rapid results, thereby enabling a high throughput of samples. Therefore XRF technology has the potential to open the way for routine monitoring of seafood products in situ at any point through the supply chain.

Atlantic cod (*Gadus morhua*) is a species of significant economic importance and supports large fisheries across the North Atlantic, leading to a global capture production of over 1.1 million tonnes in 2020 (FAO, 2022a). In the UK, landings of Atlantic cod had the highest value of all demersal species in 2020, worth £46.9 million (MMO, 2021). The Northeast Arctic fisheries in the north Norwegian and Barents Sea and the fisheries around Iceland provide the largest catches of cod, with over 690,000 and 270,000 tonnes respectively landed in 2020 (ICES, 2021c, ICES, 2021a), and these are certified as sustainable by the Marine Stewardship Council (MSC) (Marine Stewardship Council, 2020b). However, cod fisheries in other regions of the north Atlantic have varying stock statuses and fishing pressures, making some more vulnerable to over-exploitation. For example, the North Sea cod fishery was awarded certification as sustainable by the MSC in 2017, but only two years later this was withdrawn due to the stock declining to below safe biological limits (Marine Stewardship Council, 2020b). The eastern Baltic cod stock is also at risk and ICES has recommended zero catch for 2022 (ICES, 2021d). The differing catch limits and regulations within each of the ICES subareas creates an incentive for false claims of origin to conceal cod caught above the quota or in illegal fishing areas. This risk of fraud combined with the possibility of accidental mislabelling means that confirming the true catch region of products is essential for enabling sustainable management of fish stocks under increasing demand, particularly to protect and rebuild those that are depleted but also to ensure that the health of larger stocks is maintained. This also allows compliance with regulations to be enforced and acts as a deterrent for fraud. Furthermore, traceability is necessary to maintain consumer confidence in products that are certified as sustainable and to protect honest producers, since these products often have a higher value.

Given the demonstrated effectiveness of multi-elemental profiling for determining provenance of marine products, the use of XRF equipment for spatial traceability is therefore an attractive prospect. Gadd et al. (2018) was the first study to explore the possibility of using the Itrax XRF core scanner to analyse soft biological tissues, rather than the typical application of geological cores, and developed a suitable method to implement this. Following this, the Itrax scanner has been tested for origin traceability of tiger prawns (Gopi et al., 2019b), Asian seabass (Gopi et al., 2019c) and octopus (Martino et al., 2022a). Using the Itrax to analyse elemental abundance, Gopi et al. (2019b) was able to assign farmed and wild-caught tiger prawns to geographic origin with correct classification rates of 98% and 100% using random forest classification and linear discriminant analysis respectively, and found that using only seven essential elements (Mg, Cu, Ca, K, Fe, Se and Zn) gave a similar accuracy to using all 31 recorded elements. Asian seabass (*Lates calcarifer*) could not be traced to origin with this same level of accuracy, but achieved 72% correct classification with random forest using elemental analysis by the Itrax (Gopi et al., 2019c). Martino et al. (2022a) combined the isotopic composition of the statoliths with multi-elemental analysis of the soft tissue using the Itrax to determine provenance of wild-caught octopus from southeast Asia and Australia, which resulted in 94.7% of octopus samples being correctly assigned to region of origin using these two techniques together. In this study, $\delta^{18}O$ and $\delta^{13}C$ tracers were responsible for the majority of the variation among regions, followed by Br, K and As. Apart from these few studies there has been scarcely any investigation into the use of the Itrax scanner for analysing the elemental composition of fish or seafood samples. The Olympus Vanta analyser is another XRF instrument that can quickly determine the abundance of a wide range of elements in a sample, but to my knowledge has never been trialled for use on fish or seafood products. The major benefit of the Vanta analyser is that it is portable and can be taken into the field (Olympus, 2017), so could be used at seafood landing ports, fish markets, or at retailers for verification of provenance.

X-ray fluorescence (XRF) instruments assess the abundance of elements within samples by producing an x-ray beam that irradiates the sample surface. This ionises the atoms, causing inner-shell electrons to be ejected. Higher energy outer-shell electrons then move inwards to replace them, thereby releasing energy in the form of fluorescent x-rays which are detected by the XRF instrument (Brouwer, 2006). The radiation emitted when an atom is ionised is unique to each element, since all elements have a different electron configuration. Therefore, a sample emits radiation that is characteristic of the atoms present, and the peak area for each element reflects the abundance within the sample (Brouwer, 2006). The sensitivity of the XRF instruments to different elements varies, with light elements more poorly detected because they have lower fluorescence energies that are more susceptible to scatter and absorption effects, so may be unable to reach the detector at low concentrations (Potts, 1992). However, the Vanta analyser can detect

certain light elements more reliably than the Itrax scanner, since the Vanta makes contact with the sample surface unlike the Itrax. The Vanta also internally calibrates measurements and so provides data on the absolute elemental concentrations in parts per million (Olympus, 2017), whereas the Itrax acquires data as elemental counts (Gadd et al., 2018), best expressed using log ratios (Weltje and Tjallingii, 2008)

The aim of this study is to investigate whether there is potential to use x-ray fluorescence (XRF) instruments to trace the catch region of Atlantic cod, and therefore provide a forensic technique for verifying the spatial origin of traded fish products that could be hand-held and portable for quick in-situ analysis. Cod muscle tissue samples collected in 2018 to characterise stable isotope compositions (Chapter 2) were used opportunistically for this study. We sought to compare two XRF technologies for determining the geographic origin of cod - the Itrax core scanner and the Vanta analyser. Firstly, suitable methods were developed for the sample preparation and analysis of powdered fish muscle tissue using both instruments, as well as for the data processing required after initial acquisition. These methods were applied to record the abundances of a wide range of elements present in 225 Atlantic cod samples from ten catch regions across the Northeast Atlantic Ocean using both XRF instruments. We then estimated the accuracies with which individuals can be traced back to their true origin region, which provided a comparison of the effectiveness of the Itrax and Vanta XRF for spatial traceability of cod. This study analysed freeze-dried fish samples with the aim of exploring whether there is value in investigating further with fresh fish in the future.

5.3 Methodology

5.3.1 Sample collection

This study built opportunistically on samples of Atlantic cod (*Gadus morhua*) muscle tissue collected in 2017 and 2018 to characterise stable isotope compositions (described in Chapter 2). Therefore, cod samples from a total of ten regions across the Northeast Atlantic Ocean were used, from multiple different stations within each region. Table 5.1 lists the number of cod samples analysed from each region and the organisations who provided the samples. Not all samples collected for Chapter 2 were analysed for multi-elemental abundance, since in some cases the amount of sample material available was not sufficient to obtain a reliable measurement. In total, 225 individual Atlantic cod samples were analysed for this study.

Since the samples were obtained from research fisheries surveys or fish tagging studies carried out by research organisations, the authenticity of the catch locations was ensured. However, this excludes the Barents Sea samples provided by Young's Seafood, for which

no exact locations are known. The regions sampled cover the main commercial fishery areas supplying the trade network, and broadly correspond to the ICES subareas within FAO major fishing area 27, except for the region referred to as the Norwegian Sea which consists of stations falling within both ICES subareas 1 and 2 (Barents and Norwegian Sea subareas) (Figure 5.1).

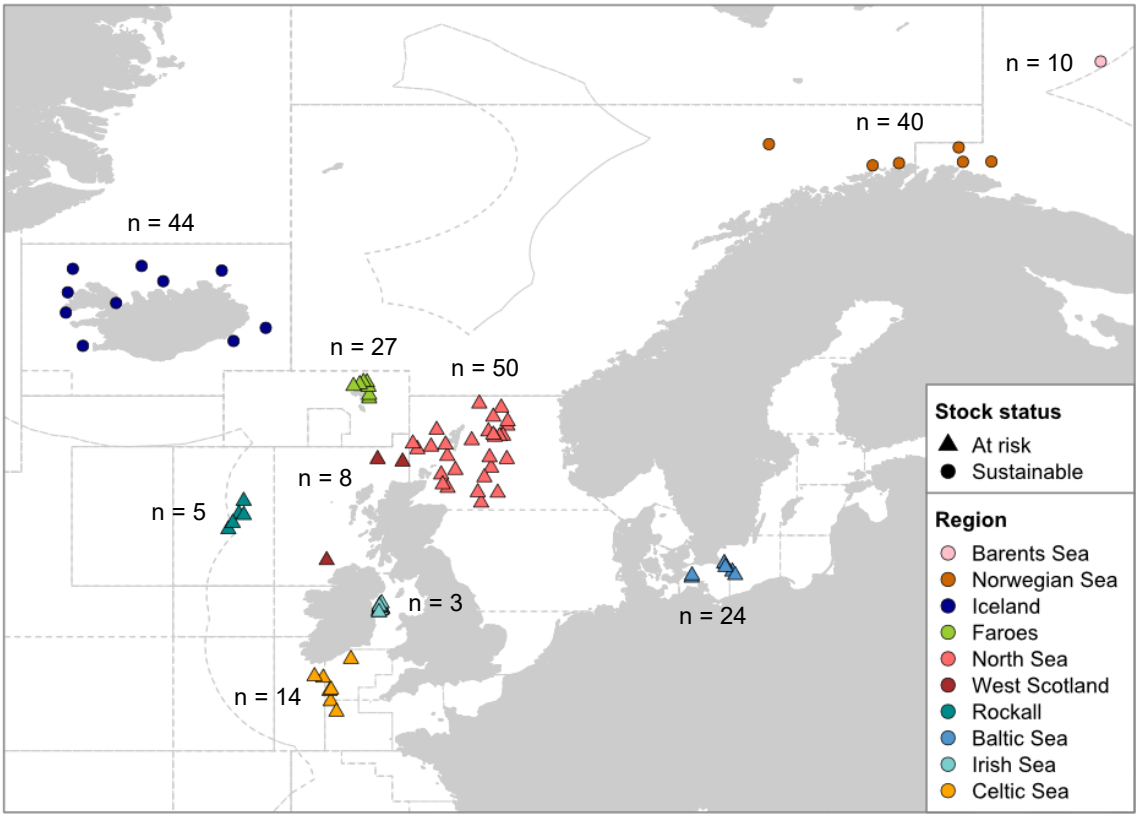


Figure 5.1 Locations of all stations where samples of Atlantic cod were obtained, coloured according to the geographic regions. The total number of samples (n) collected from each region are shown and the symbols indicate whether the stock is sustainable or at risk (as determined from the ICES 2022 advice for each stock (ICES, 2022h), based on the spawning stock biomass relative to the limit where reproduction of the stock is impaired (B_{lim})). ICES subarea boundaries are indicated by the dotted grey lines.

Table 5.1 Total number of individual Atlantic cod analysed from each geographic region, the number of stations sampled per region, and the source from which samples were obtained.

Catch region	Number of cod analysed	Number of stations	Year samples collected	Source of samples and organisation name
Barents Sea	10	Unknown	2017	Young's Seafood Ltd.
Norwegian Sea	40	6	2018	Annual fisheries survey - Institute of Marine Research
Iceland	44	10	2018	Annual fisheries survey - Marine and Freshwater Research Institute
Faroe Islands	27	7	2018	Annual fisheries survey - Faroe Marine Institute
North Sea	50	24	2018	Annual fisheries survey - Marine Scotland
West Scotland	8	3	2018	Annual fisheries survey - Marine Scotland
Rockall	5	5	2018	Annual fisheries survey - Marine Scotland
Baltic Sea	24	4	2018	Fish tagging survey - Technical University of Denmark
Irish Sea	3	3	2018	Fish tagging survey - Marine Institute
Celtic Sea	14	7	2018	Annual fisheries survey - Ifremer
Total	**225**	**69**		

5.3.2 Sample analysis

On collection all samples were stored frozen at -20°C, with the exception of the Faroe Islands samples that were preserved in 70% ethanol for transport to the UK before being frozen. All samples were freeze-dried and ground to a uniform powder using a metal grinder that was manually operated using a twist action. The same grinder was used throughout to ensure consistent grain sizes across all samples as much as possible, and care was taken not to include any skin with the muscle tissue. Compressed air was used to clean the grinder between samples to avoid cross-contamination.

To prepare the samples for analysis by the XRF instruments, the samples were arranged in a grid along a wooden plank, following the method described by Gadd et al. (2018) for the use of Itrax for soft biological tissues. A strip of double-sided sticky tape was attached

along the length of the plank to secure the samples, and masking tape was used to divide the strip of tape into 2x2cm sections. Masking tape was also placed along the top and bottom of the sticky tape to form a grid of 4cm^2 squares (Figure 5.2). One sample was placed onto each square of sticky tape using a spatula, with enough powder to cover most of the square as evenly as possible, and with a thickness of approximately 1-2mm. For some samples where only a very small amount of powder was available, the width of the square was reduced to 1cm to ensure there was an appropriate sample thickness. The spatula was cleaned with ethanol between samples to avoid cross-contamination. Each sample was pressed down to compress the powder and create a flat surface, and then labelled with its unique sample number onto the masking tape.

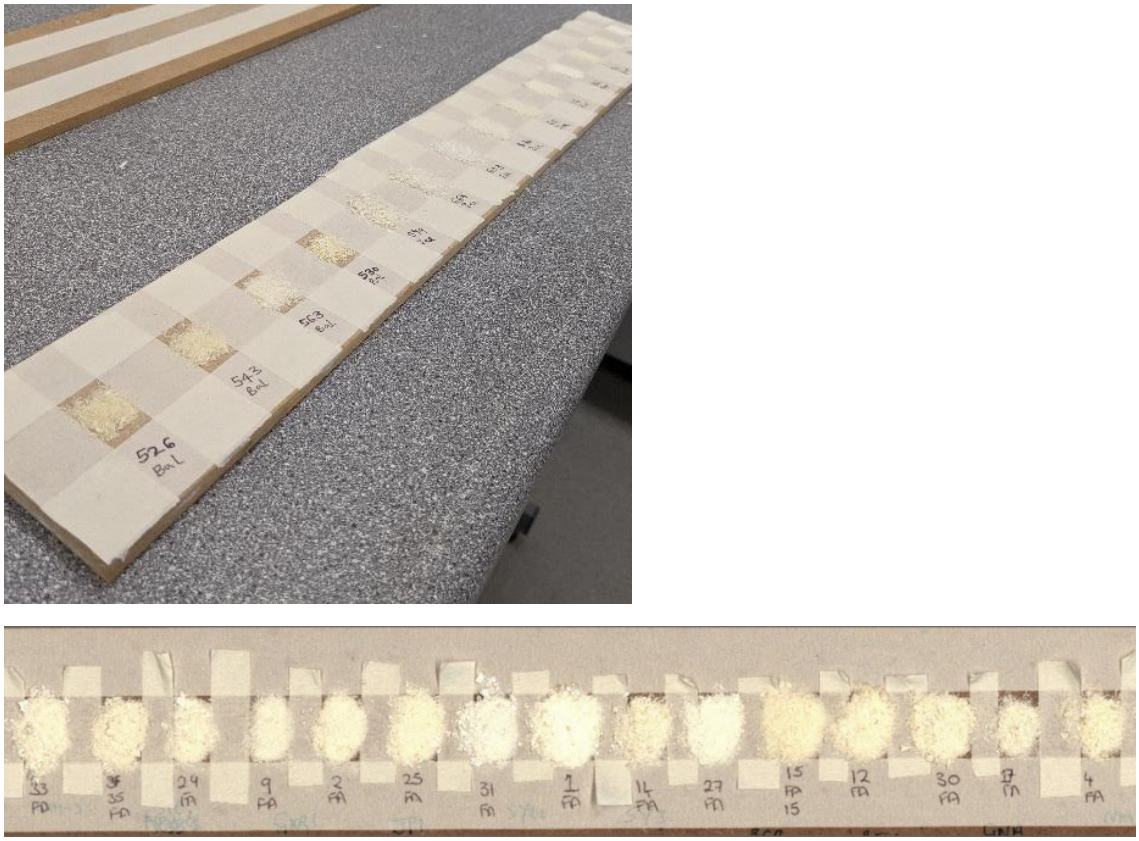


Figure 5.2 The sample grid used for analysis by the Itrax and Vanta instruments, with a powdered sample placed in each 4cm^2 grid square.

5.3.2.1 Itrax core scanner

The Itrax scanner is primarily designed for split sediment core samples, so a different process for scanning samples and data processing had to be developed to give reliable data for the many separate fish muscle samples that have a different composition to sediment samples.

The samples were analysed using the COX Analytical Systems Itrax X-ray Fluorescence (XRF) Scanner housed at the British Ocean Sediment Core Research Facility (BOSCORF) in the National Oceanography Centre Southampton. The grid of powdered samples was loaded into the Itrax and scanned at 0.2mm intervals with a 60 second exposure time. A molybdenum tube was used with a current of 30 mA and a voltage of 50 kV. The settings were chosen to allow for detection of heavier elements, while the count time was selected to allow for those in low abundance to be detected. However, the Itrax is a non-contact XRF scanner with the detector held above the surface, which limits the detection of lighter elements. The sample height was also recorded at each 0.2mm interval, measured by the height of the scanner above the sample surface, and an optical image of each sample surface was captured.

The Itrax core scanner determines the elemental composition of a sample at high resolution using micro x-ray fluorescence. The elemental concentrations are recorded in counts per second (cps), so are considered semi-quantitative. The Itrax scanner was programmed to record counts for a standard suite of 40 elements developed primarily for sediment analyses (Al, Si, P, S, Cl, Ar, K, Ca, Sc, Ti, Cr, Mn, Fe, Ni, Cu, Zn, As, Se, Br, Sr, Zr, Pd, Cd, Sn, I, Cs, Sm, Eu, Ta, W, Ir, Pt, Au, Pb, Bi, Rn, Ra, Th, Pa, U). It is recognised that many elements will not be present in fish muscle at detectable concentrations, but the element suite was not modified a priori. Replicates were measured for 21 samples to determine consistency, where each replicate used a different subset of the powdered sample placed onto the sample grid.

Samples of the wooden board and double-sided tape were also measured using the Itrax to assess whether the instrument was recording any counts from these materials beneath the fish muscle samples, and masking tape was analysed to investigate whether it contained any indicator element that could be used to distinguish the sample boundaries.

A new tool named ‘ITRAX SumSpectra’ was developed by Cox Analytical to batch reprocess the data, allowing the best fit to be calculated for the fish samples using the entire dataset. Each dataset was then reviewed individually, to ensure the Mean Squared Error was minimised, and an assessment made of the spectra for each sample.

5.3.2.2 Vanta analyser

The Vanta analyser is a portable instrument that uses x-ray fluorescence to estimate elemental abundance within samples. It can be used in either fixed or handheld mode. The potential to operate in handheld mode and as a portable instrument is particularly relevant to application in food forensics settings. The Vanta has internal calibration and therefore provides estimates of absolute elemental concentrations in parts per million (ppm). During sample analysis, the Vanta analyser makes contact with the sample, so it can measure lighter elements more reliably, and it is more suited to discrete samples than

the Itrax scanner. It can be used for a variety of materials, but is often used for rock samples and sediment cores. Since no published studies could be found where the Vanta has been used to analyse fish muscle samples, the process for sample analysis and data processing had to be developed for this study.

The samples were also analysed using the Olympus Vanta handheld X-ray Fluorescence (XRF) Analyser housed at the British Ocean Sediment Core Research Facility (BOSCORF) in the National Oceanography Centre Southampton. The same grids of powdered samples from the Itrax were presented to the Vanta analyser, which was mounted on a core scanner, and the Vanta was manually driven to each sample (Figure 5.3). The three-beam option was selected on the Vanta to maximise heavy element detection. Samples were analysed using a count length of 60 seconds and the number of repeats was set to three (i.e. each measurement was repeated three times and an average was produced by the instrument). Replicate measurements were performed for 32 samples by repeat analysis of the same square on the sample grid (with a set of three measurements for each replicate) to assess consistency of measurement. The abundance of 39 elements were measured by the Vanta analyser (Mg, Al, Si, P, S, K, Ca, Ti, V, Cr, Mn, Fe, Co, Ni, Cu, Zn, As, Se, Rb, Sr, Y, Zr, Nb, Mo, Ag, Cd, Sn, Sb, Ba, La, Ce, Pr, Nd, W, Hg, Pb, Bi, Th, U).

The Vanta automatically performs internal validation of the measurements using Olympus data that has been tested under a range of conditions, so no processing of the raw data is required after acquisition.

Figure 5.3 Olympus Vanta mounted on a core scanner to analyse the powdered fish muscle samples, which are presented to the instrument in a grid arrangement.

5.3.3 Data processing

5.3.3.1 Itrax

The reliability of measurements using both the Itrax and Vanta instruments may be affected by the angle and roughness of the sample surface presented, and by spurious counts from inter-element interactions and spectral interferences. We were not aware of previous work assessing element interactions specifically relating to fish muscle. Given that our interest primarily lies in the ability to differentiate among matrix-matched samples based on known origin rather than quantifying absolute element concentrations, we ignored counts for elements known to occur in fish muscle at sub ppm concentrations (Ar, Ti, Zr, Pd, Sm, Eu, Ta, W, Ir, Pt, Au, Bi, Rn, Ra, Th, Pa, U). We did not apply any other matrix-specific arithmetic corrections.

Sample roughness was addressed by selecting only the data where there was a flat, uniform surface and excluding any unreliable data from the sloping sides of the sample mounds. To achieve this, firstly a profile of the sample surface height was plotted for each sample. In this way it was noted that for the majority of samples the central 60% portion (i.e. 30% above and below the halfway distance along the sample) would include only the flat surface on top of the sample mounds, although there was a minority of samples that contained a peak or very uneven surface within that section. Therefore, the time resolved elemental counts for each sample were individually assessed to determine whether the central 60% portion contained only measurements on the upper flat surface of the sample

mound or whether the surface appeared sloping or very rough on visual inspection, so that the data could be filtered accordingly. In the latter case, the section containing a flat surface was manually selected using the surface height profiles and distance measurements associated with each datapoint. Diagrams illustrating a good sample with a flat upper surface and those that were peaked or rough, and therefore excluded, are illustrated in Figure 5.4. This data selection also ensured that no measurements from the masking tape were included, since the samples had a much greater elevation on the surface height profiles than the tape.

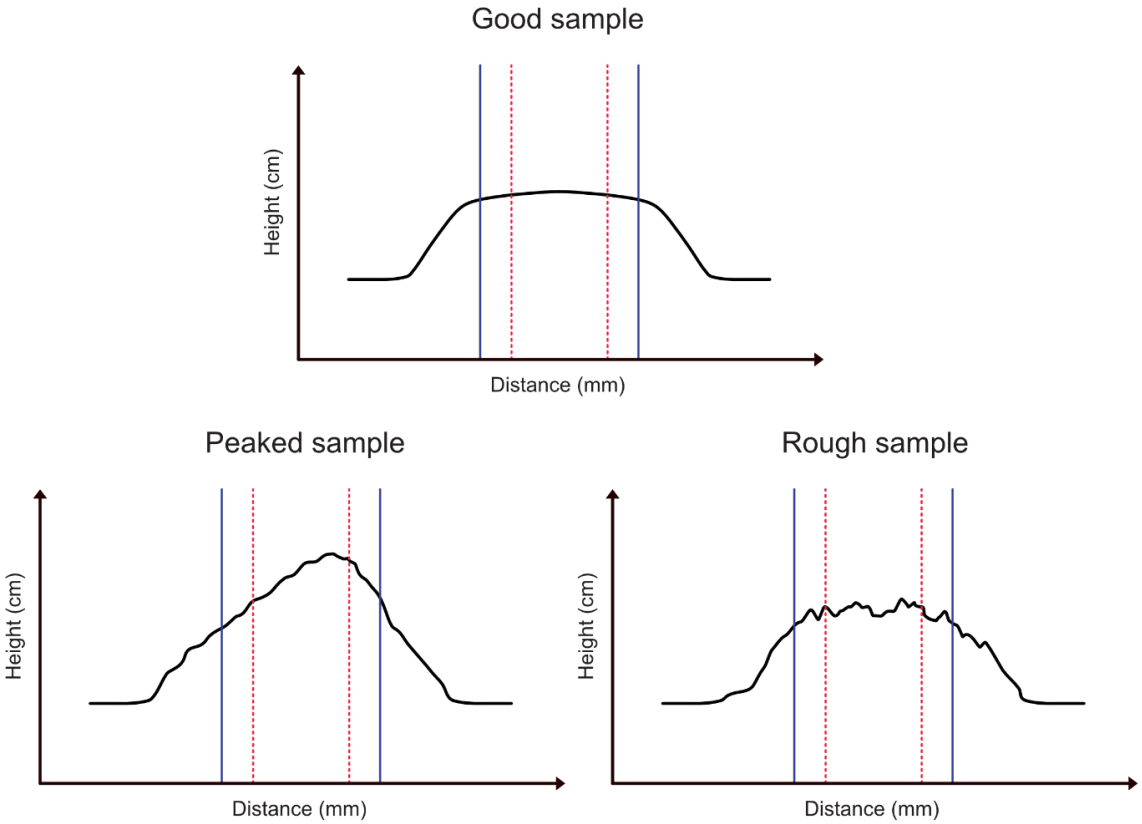


Figure 5.4 Schematic of the sample height profile for a good sample in comparison to a peaked or rough sample, showing the full extent of the Itrax scan (blue lines) and the central 60% portion that was selected (red dashed lines).

The thickness of each sample was estimated by subtracting the height at the bottom of the sample mound from the minimum height of the flat upper part of the sample. The roughness of each sample was also calculated, defined as the sum of the height differences between individual datapoints across the flat section of the sample selected for the final dataset.

To assess whether the data were more unreliable for samples that were thinner or those with a more uneven surface, and therefore whether any further samples should be

excluded from the analysis, the mean squared error was plotted separately against sample thickness and roughness.

The data were then filtered to remove measurements that had a validity of zero (indicating that the return signal is not stable or valid as determined by the internal quality assessment made by the Itrax) and those with a mean squared error (MSE) of greater than 5. This MSE threshold was selected with the aim of identifying parts of the data that were uncharacteristically variable, but without removing whole samples. Using this filtered dataset, the mean counts per second per sample and standard deviation were calculated for each element measured by the Itrax scanner. As explained above, element selection for data analysis was carried out whereby lighter elements that cannot be reliably measured by the Itrax scanner were discarded (considered to be those consistently recording less than 150 cps) along with elements that are not typically present in seawater or fish muscle tissue.

5.3.3.2 Vanta

All the data acquired by the Vanta analyser were retained in the dataset with no filtering by error values, however samples without any useable flat or uniform surface were removed (8 in total) in the same way as with the Itrax data. For each element measured by the Vanta analyser, the one sigma error was plotted separately against the sample thickness and roughness (both calculated from the Itrax dataset) to identify any covariance and therefore indicate whether any further samples should be removed for the data analysis.

The mean values of elemental abundance and standard deviations across replicate counts were calculated for each sample. Elements were selected for data analysis by excluding the lighter elements that could not be reliably measured by the Vanta and those that are not expected to be found in seawater or fish muscle tissue.

5.3.4 Statistical analysis

All statistical analyses were performed using R version 3.6.2 (R Core Team, 2019). Co-variance among elements was assessed by creating a matrix of pairwise scatterplots, which also indicates the correlation coefficients between all combinations of variables, and only one element of any strongly co-varying element pair was retained for subsequent statistical analyses.

The abundance of each element by region of origin for both the Itrax and Vanta datasets was initially graphically analysed using boxplots. To assess any significant differences among regions ANOVA tests were conducted. To determine the consistency of measurement by both XRF instruments, a comparison of replicate results was carried out

for samples that were analysed more than once by calculating the mean difference between replicates as a proportion of the mean concentration.

XRF data are usually regarded as semi-quantitative for most applications, and element ratios (log ratios) are recommended for data exploration. To select element ratio pairs likely to provide informational value, raw element counts were entered into Principal Component Analysis (PCA). Element locations on two dimensional plots (PC1 and PC2) were used to identify element pairs falling in different quadrants, and therefore likely to show limited covariance. The final combination of element pairs was chosen by plotting a number of PCA biplots with elements selected in this way and observing subjectively which gave the best separation. In the case of having an odd number of elements, the element that is found to contribute least to the assignments using random forest classification with the raw data will be omitted from the ratio pairs.

PCA was then carried out using these selected element log ratios to determine graphically whether samples separate according to region of origin, and to determine which elements are most important in distinguishing cod from different regions. Any zeros in the element log ratio data were replaced with 0.0001 to avoid these samples being omitted from the PCA due to the resultant infinity values. The PCA analyses using raw and log ratio element data were first calculated on the Itrax dataset and subsequently on the Vanta dataset separately.

5.3.4.1 Assignment to region of origin

Cod samples were assigned to their most likely origin using random forest classification, which is a machine learning based classification algorithm that relies on a multitude of individual decision trees operating as an ensemble. Random forest classification is suited to assigning samples using elemental abundance data because it can make use of a large number of different variables and does not rely on the data having a normal distribution.

Assignment to region of origin using random forest was conducted using the methodology described in Chapter 2 on page 44, first with the raw element data and then with the element log ratios selected using PCA. The assignment was carried out using the Itrax and Vanta data separately.

Assignment to origin using random forest classification was also conducted without filtering the data (leaving all data in the dataset regardless of the roughness or unevenness of the sample surface), to investigate the importance of the cleaning protocol prior to data analysis. However, the datapoints with a validity of zero and those with mean squared error of 5 or greater were still removed.

The importance of reprocessing the data using the ‘ITRAX SumSpectra’ tool was also investigated, since this was a very time intensive process and it would be beneficial to

assess the improvement this step provided. Therefore, random forest classification was performed in the same way as previously described using the original data from the Itrax, before reprocessing was carried out, and the assignment accuracy was compared to that using the reprocessed data.

5.3.4.1.1 Combining XRF data and stable isotope data for assignment to origin

Finally, the stable isotope ratios measured in cod muscle tissue in Chapter 2 were combined with the XRF elemental abundance data, since both these datasets were obtained by analysing the same cod samples. The stable isotope data for each sample were matched to their associated elemental data using the unique ID number assigned to every sample. The most successful of the two XRF methods (Itrax or Vanta) was selected for this analysis. The cod samples were assigned to region of origin using random forest classification based on this combined dataset, to investigate whether the use of two approaches simultaneously improves the ability to distinguish individuals from different origins compared to using the methods independently.

5.4 Results

5.4.1 Itrax

A total of 19 elements detected by the Itrax scanner were selected to be included in the data analysis and origin assignment – S, Cl, K, Ca, Sc, Cr, Mn, Fe, Ni, Cu, Zn, As, Se, Br, Sr, Cd, I, Cs and Pb. Phosphorus and tin both co-vary with potassium (phosphorus: $r = 0.962$, d.f. = 214, $p<0.001$; tin: $r = 0.846$, d.f. = 214, $p<0.001$), so phosphorus and tin were excluded. Aluminium and silicon were removed because these light elements cannot be reliably measured using the Itrax. A further group of elements were excluded because while counts above detection limits were returned, these elements are typically found in very low concentrations seawater or fish muscle (e.g. argon and titanium), and counts assigned to these elements are likely to reflect spectral artefacts associated with element interactions.

To assess consistency of measurement, a group of 21 tissue samples were analysed two or three times by the Itrax using different sub-samples of the powdered tissue. The associated measurement precision of the Itrax measurements, represented by the mean difference between replicates expressed as a proportion of the mean concentration, ranged from ± 0.21% for zinc to 38.6% for potassium (Figure 5.5). Where the same sample was measured multiple times, only one replicate was kept in the dataset for subsequent analysis. In this case, the thickest sample was selected from the replicates to allow the best possible accuracy of measurement.

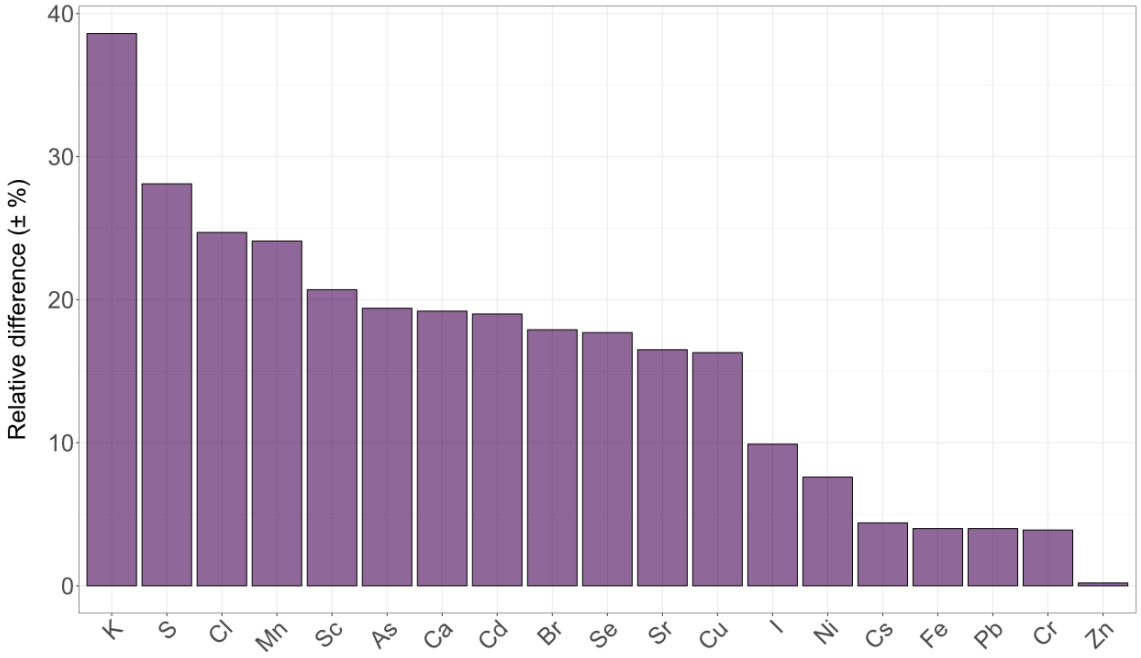


Figure 5.5 The average measurement precision for each element measured by the Itrax scanner, defined as the mean relative difference between the two replicate measurements for each sample. The relative difference was calculated as the difference between the mean sample values for both replicates, divided by the mean of the two samples and expressed as a percentage.

The effect of the coarseness of grinding on recovered counts for the powdered samples was investigated by measuring four samples twice – once ground coarsely and then repeated after further grinding to a uniform powder using the same grinder for all samples. The mean relative difference between these two replicates ranged from ±2.3% for calcium and arsenic to ±101.3% for chromium, as shown in Figure 5.6. The relatively low precision for certain elements suggests that there may be an impact on data quality if samples are not ground finely and to the same degree, as consistent with similar studies from sediment cores (MacLachlan et al., 2015). Therefore, all samples for the final dataset were ground finely using the same grinder, as described in the methods. Samples that were found to have a rough surface were discarded or the flattest part of the sample surface was selected.

After observing the height profiles for each sample, data from eight samples were discarded completely because they contained no sufficiently flat surface upon visual inspection, and therefore may not have been measured accurately by the Itrax scanner. Datapoints were manually selected from the flattest upper portion for 36 samples that had a non-uniform surface, and for the remaining 181 samples datapoints were retained from the central 60% portion to coincide with the flattest surface (as described in the methodology). The mean squared error for the remaining data does not increase with either the thickness

of the sample or the roughness of the sample surface (after removing those with MSE > 5), therefore no further samples were removed from the dataset. This is interpreted as an indication of data quality for the chosen data points.

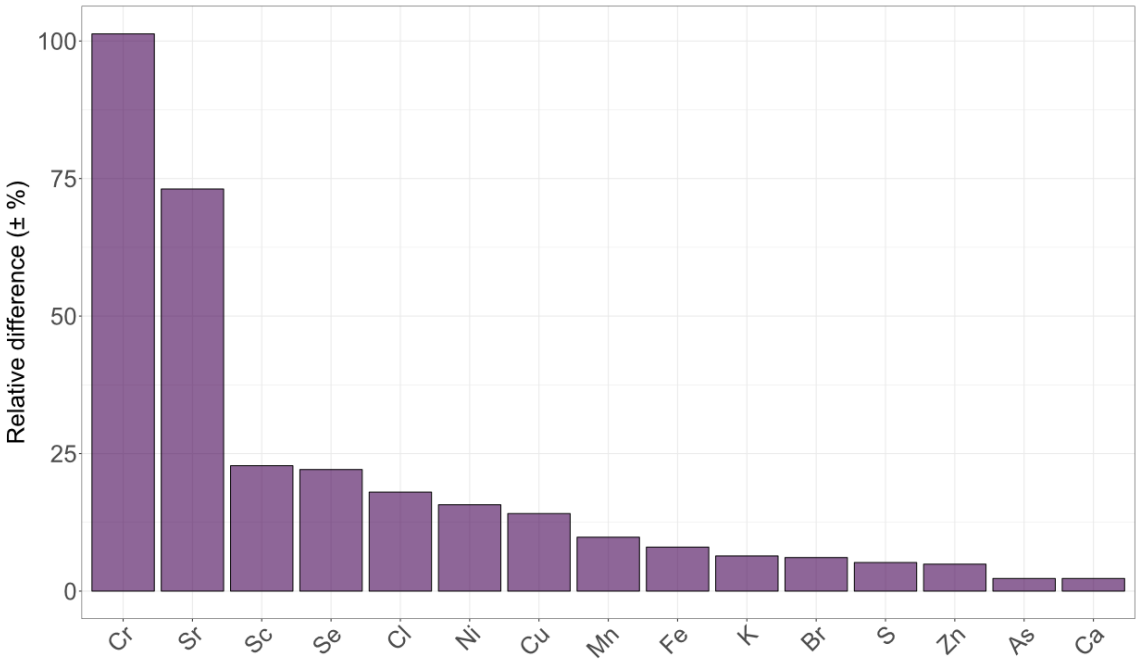


Figure 5.6 Comparison between samples ground to a uniform fine powder using the grinder and samples that were ground more coarsely. The average measurement precision for each element is shown, as measured by the Itrax scanner, defined as the mean relative difference between the two replicate measurements for each sample. The relative difference was calculated as the difference between the mean sample values for both replicates, divided by the mean of the two samples and expressed as a percentage.

The mean counts per second (cps) for the selected elements and those that co-vary are shown in Figure 5.7, ordered by mean counts per second from highest to lowest. Analysis of the cod samples using the Itrax scanner gave much higher counts per second for potassium in almost all samples than for any other element and has the largest range - potassium has a mean value of 180,832 cps, which is over 12 times greater than that of sulfur (the element with the second highest mean counts per second). The Certified Reference Material for cod muscle (EVISA, 2022) also lists potassium as being the element present in the highest concentration by far of all those measured and sulfur as being the second most abundant (Table 5.2), so supports the Itrax counts recorded in this study. Calcium also has high counts from the cod samples, with a mean of 10,922 cps, and is therefore in agreement with the cod reference material as being a relatively abundant element. Whereas caesium has the lowest counts from the Itrax of all the selected elements (mean of 24 cps). Other elements with low counts are chromium, iodine and

scandium. Most elements have a relatively small range of counts per second among the samples, but a few including chlorine, potassium and calcium show greater variation. Arsenic also has a large range of values, with very low counts in many cases, even 0 cps in over 40% of samples, and yet a minority of other samples with much higher counts reaching 97,261 cps. Histograms in Figure 5.8 illustrate the frequency distributions for concentrations of the 19 selected elements. Only scandium shows a normal distribution ($p > 0.05$).

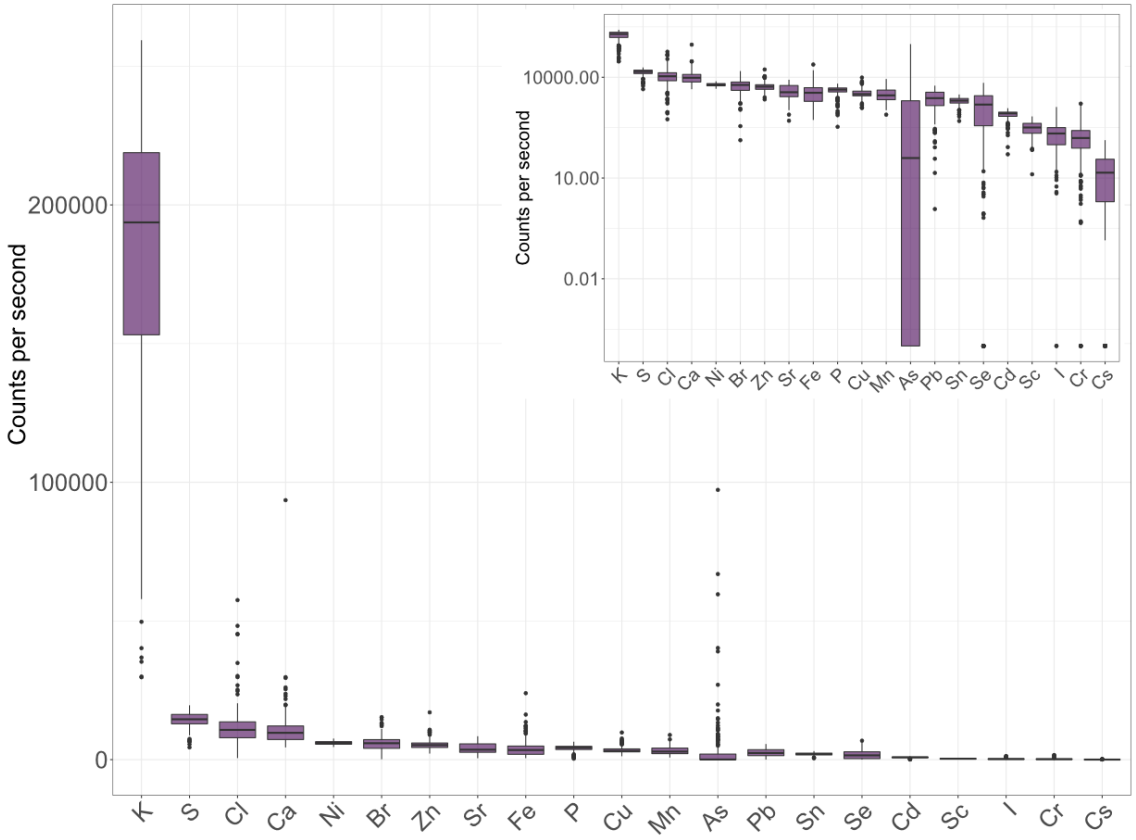


Figure 5.7 Mean counts per second for each of the selected elements and co-varying elements in individual cod muscle tissue samples as recorded by the Itrax scanner. Insert plot shows the same results on a log 10 scale for counts per second. Elements are ordered according to mean counts per second from highest to lowest.

Table 5.2 Trace element concentrations in cod muscle Certified Reference Material BCR-422 as determined by spectrometry, originally certified by Quevauviller et al. (1992) and produced by the Joint Research Centre (EVISA, 2022). Elements are listed from highest to lowest concentration.

Element	**Certified value** ($\mu g\ g^{-1}$)	**Uncertainty** ($\mu g\ g^{-1}$)
Potassium	21700	800
Sulfur	11500	600
Sodium	2190	420
Calcium	330	90
Arsenic	21.1	0.5
Zinc	19.6	0.5
Bromine	17.0	0.9
Iron	5.46	0.30
Iodine	4.95	0.49
Selenium	1.63	0.07
Magnesium	1.37	0.07
Copper	1.05	0.07
Strontium	0.66	0.12
Mercury	0.559	0.016
Manganese	0.543	0.028
Lead	0.085	0.015
Cadmium	0.017	0.002
Cobalt	0.015	0.002

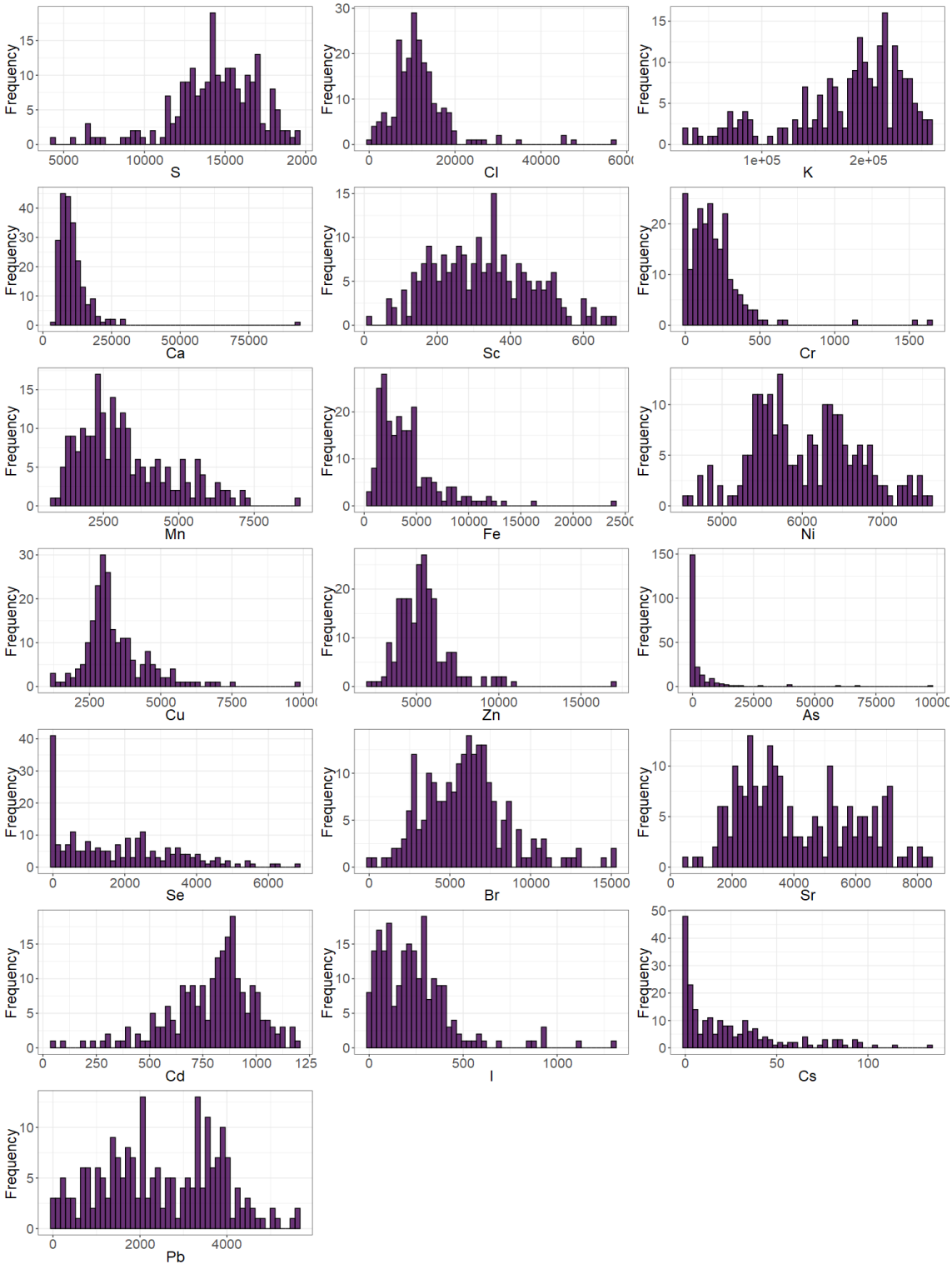


Figure 5.8 Frequency distributions of the counts per second recorded for all selected elements measured by the Itrax scanner in Atlantic cod muscle tissue from all regions.

The mean counts per second for the selected elements in cod from each region as recorded by the Itrax scanner are shown in Table 5.3, together with the standard deviations. The elemental composition of cod muscle tissue varies among the 10 regions for the sampled individuals, with all selected elements demonstrating a significant effect of region of origin (Kruskal-Wallis, $p < 0.05$). Figure 5.9 displays the variation in elemental counts per second among different regions of origin for a few example elements of interest. Cod samples caught in the Barents Sea have much higher counts per second for arsenic than cod from all other regions, with a mean of 35,817 cps and some samples recording almost 100,000 cps, compared with means of 0 – 4,815 cps for the other catch regions. Iodine was found in the highest abundance in the Baltic Sea cod where sample counts range from 219 to 1,301 cps, whereas the iodine counts at all other locations are between 0 cps and a maximum of 546 cps. The Faroes samples have significantly lower abundances of potassium (mean 6,657 cps) than any of the other regions, and samples from Rockall also have relatively low potassium concentrations (mean 11,5535 cps), compared with a mean of 22,3452 cps in the North Sea. Iron was found in the lowest concentration in the most northerly locations of the Barents and Norwegian seas. Although some elements are highly variable among regions, other elements including lead and strontium are similar across all locations with a wide range of counts within each (Figure 5.9).

Table 5.3 Mean elemental counts per second recorded in the cod muscle tissue samples from each geographic region using the Itrax scanner, and the associated standard deviations (SD).

Region	Counts per second (cps)									
	S		Cl		K		Ca		Sc	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Barents	13337	1182	11227	1656	186409	24045	7566	2080	220	47
Norwegian	14673	1944	8499	2021	217047	28305	7253	1614	211	103
Iceland	14650	2574	17089	12238	179806	32009	8924	2989	282	117
Faroes	11696	3258	3795	2212	66657	18457	10948	4046	272	81
North Sea	16630	1363	12066	2264	223452	18488	14237	12791	454	122
WScotland	14056	1145	16939	1769	166276	40510	8824	806	368	94
Rockall	9862	3075	11300	5792	115535	18272	8909	2335	233	49
Baltic	13001	1291	11283	1721	192399	17539	15366	5301	439	75
Irish	13931	219	15648	5070	177508	12835	10572	2201	491	24
Celtic	15704	1968	16365	5320	180907	26628	13284	6313	309	57

Region	Counts per second (cps)									
	Cr		Mn		Fe		Ni		Cu	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Barents	293	68	2173	294	1463	348	6239	433	3838	553
Norwegian	245	129	2438	853	1873	791	6193	695	3306	880
Iceland	152	96	2539	1217	3114	1817	6009	724	3150	672
Faroes	148	135	2956	1729	3214	1533	5987	867	2763	787
North Sea	126	126	3608	1081	5591	2326	5939	438	3275	600
WScotland	143	157	4357	703	7234	2096	5924	637	3184	1149
Rockall	26	19	2941	1028	5074	4047	5856	790	2330	462
Baltic	212	126	5674	869	6257	4887	6318	514	5342	1288
Irish	303	191	5987	779	6247	2735	6677	334	6552	695
Celtic	436	555	2390	762	4856	3843	5672	374	2778	266

Region	Counts per second (cps)									
	Zn		As		Se		Br		Sr	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Barents	5554	916	35817	30960	2704	1090	9870	1539	4255	1577
Norwegian	4597	797	4815	6059	757	889	4170	1529	5299	2105
Iceland	4989	1508	225	740	778	961	6149	3039	4102	1507
Faroes	5060	1165	0	0	2112	1820	3934	1986	4197	1522
North Sea	5679	657	1788	2166	2404	1214	6024	1301	3588	1777
WScotland	8616	1351	64	180	2624	1282	7346	1292	3218	2217
Rockall	6773	843	0	1	805	895	4746	1361	3559	727
Baltic	5434	3056	53	79	3187	1971	6793	2232	3755	2133
Irish	7819	1899	485	354	4550	914	9206	1534	5083	347
Celtic	6127	1408	4427	4207	1574	963	9006	3032	3876	1559

Region	Counts per second (cps)							
	Cd		I		Cs		Pb	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Barents	805	87	151	74	4	4	2451	1102
Norwegian	840	130	86	85	8	16	1723	1036
Iceland	754	156	151	102	24	28	1892	1155
Faroes	524	174	204	99	13	26	2765	1498
North Sea	922	229	288	118	31	27	3100	997
WScotland	807	74	246	82	45	20	3141	1385
Rockall	580	76	114	69	62	34	2823	1216
Baltic	898	102	618	285	40	29	2725	1826
Irish	921	46	340	44	39	25	3594	691
Celtic	771	85	283	88	16	8	2626	948

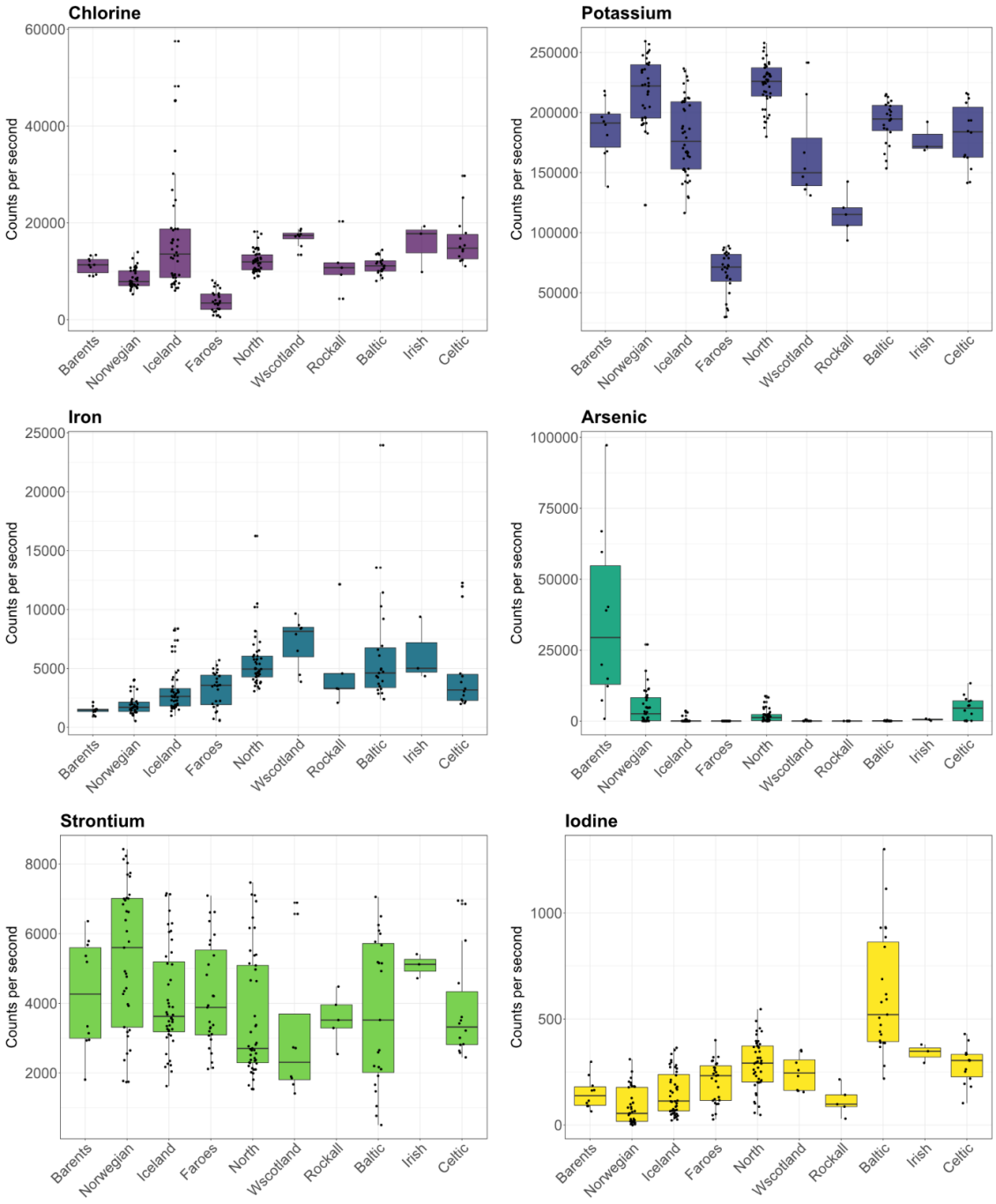


Figure 5.9 Elemental counts per second measured in the Atlantic cod muscle tissue samples from each geographic region using the Itrax scanner, for six example elements – chlorine, potassium, iron, arsenic, strontium and iodine.

5.4.2 Vanta

A total of 13 elements measured by the Vanta analyser were selected to be included in the data analysis and origin assignment – Si, S, K, Ca, Mn, Fe, Cu, Zn, As, Sr, Ba, La, U. Phosphorus co-varies with potassium, so was excluded ($r = 0.930$, d.f. = 186, $p<0.001$). Magnesium and aluminium were also removed because these light elements cannot be reliably measured using the Vanta, and a further group of elements were excluded because they are not typically found in seawater or fish muscle (e.g. titanium and silver). A few elements that are known to be present in fish muscle, including cobalt, selenium, cadmium and lead (EVISA, 2022), were not included because the concentrations recorded by the Vanta were below the 5 ppm detection limit of the instrument (Olympus, 2022) and in many cases were zero ppm, therefore could not be utilised as reliable data.

Replicates were measured for 32 samples using the same sub-samples of powdered tissue each time and the results were compared to ensure consistency of measurement. The precision of measurement, defined as the mean difference between replicates as a percentage of the mean value, was between ± 2.4% for potassium and ± 25.1% for copper, apart from barium and lanthanum which had much higher values of ± 98.1% and ± 92.6% respectively (Figure 5.10). All replicate results were retained in the final dataset and used to calculate the mean elemental concentrations for each sample.

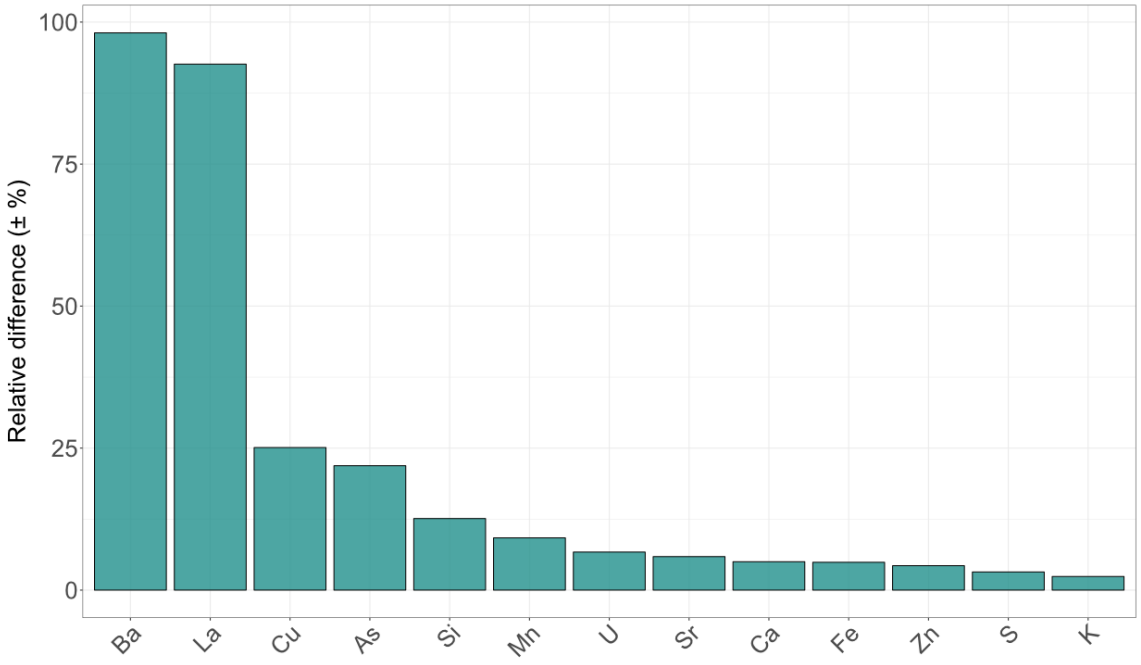


Figure 5.10 The average measurement precision for each element measured by the Vanta analyser, defined as the mean relative difference between the two replicate measurements for each sample. The relative difference was calculated as the difference between the mean sample values for both replicates, divided by the mean of the two samples and expressed as a percentage.

The same eight samples as in the Itrax data were discarded, since they were discovered to have a very uneven surface on the Itrax height profiles and may have given unreliable results. The individual error for any element does not show a positive correlation with either the thickness or the roughness of the sample, therefore no further samples were removed from the dataset. However, the error for element abundance estimates were sometimes high, with one sigma errors of 0-79 ppm (the mean error representing between 0% and 17% of the mean concentrations recorded), except for barium and lanthanum which had very high errors (mean values of 200 ppm and 368 ppm respectively) due to the many zero concentration values recorded.

Estimates of mean abundances in parts per million (ppm) for the selected elements and co-varying phosphorus recorded in all samples by the Vanta analyser are shown in Figure 5.11. The elements present in the highest concentrations within the cod muscle tissue are potassium, sulfur and phosphorus. Potassium has the highest mean abundance overall with an average of 21,589 ppm in the samples, and a very wide range from 5,016 ppm to 32,057 ppm. The second most abundance element is sulfur (mean of 16,266 ppm), but its range overlaps with that of potassium so that for certain samples the concentration of sulfur is greater than potassium. Phosphorus has a mean abundance of 9,906 ppm, and its range overlaps with both potassium and sulfur, to the extent that its minimum concentration is similar to that of potassium despite its mean being much lower. Cod muscle samples also showed relatively high concentrations of calcium with a mean of 717 ppm and a maximum of 4,565 ppm. However, strontium has the lowest abundance of all the reliably detected elements with an overall mean of 4 ppm, and a restricted range of 3 - 8 ppm. Other elements found in low abundances are copper, barium, lanthanum and uranium with mean values below 10 ppm. Arsenic also has a mean slightly below 10 ppm due to a large number of samples with zero ppm, but certain samples have much higher concentrations of arsenic up to 184 ppm. Histograms in Figure 5.12 illustrate the frequency distributions for the 13 selected elements. Sulfur is the only element to show a normal distribution (Shapiro-Wilk, $p = 0.6967$).

A comparison of the Vanta results with the published cod muscle Certified Reference Material BCR-422 (EVISA, 2022) generally reveals good agreement for the determined trace elements. The cod samples analysed by the Vanta were found to contain a mean potassium concentration of 21,589 ppm and the certified value of potassium in the cod reference material is 21,700 ppm (Table 5.2). Likewise, sulfur concentrations are comparable, with a sample mean of 16,266 ppm from the Vanta and a value of 11,500 ppm in the certified reference material. Calcium concentrations are also on the same order of magnitude, since the mean value of samples from the Vanta is 717 ppm compared to 330 ppm in the cod reference material. An average of 28 ppm zinc was recorded in the cod samples and in the reference material the certified value for zinc is 19.6 ppm, again

showing good agreement. However, iron and manganese concentrations in the cod muscle appear to be higher when estimated by the Vanta than would be expected from the certified reference material – means of 57 ppm and 75 ppm in the cod samples compared with 5.46 ppm and 0.543 ppm in the reference material for iron and manganese respectively. This may be because the detection limit of the Vanta is around 5 ppm for these elements (Olympus, 2022), and they may have been present in too low concentrations to be accurately estimated. Therefore, the overall similarities between elemental abundances given by the Vanta and the certified values for BCR-422 suggest that the Vanta gives a reliable estimate of the concentration for many elements, where they are present above the detection limits of the instrument.

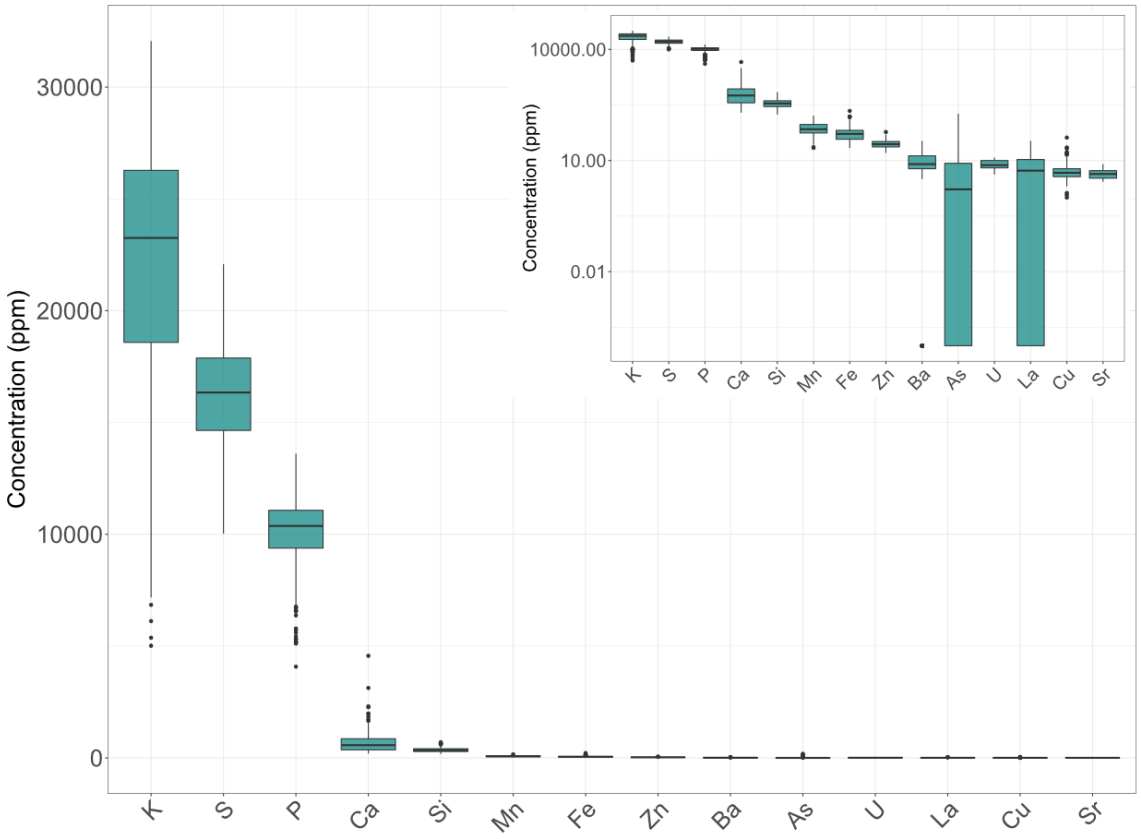


Figure 5.11 Mean elemental concentration (ppm) in individual cod muscle tissue samples as recorded by the Vanta analyser for the selected elements and co-varying elements. Insert plot shows the same results on a log 10 scale for elemental abundance. Elements are ordered according to mean concentrations from highest to lowest.

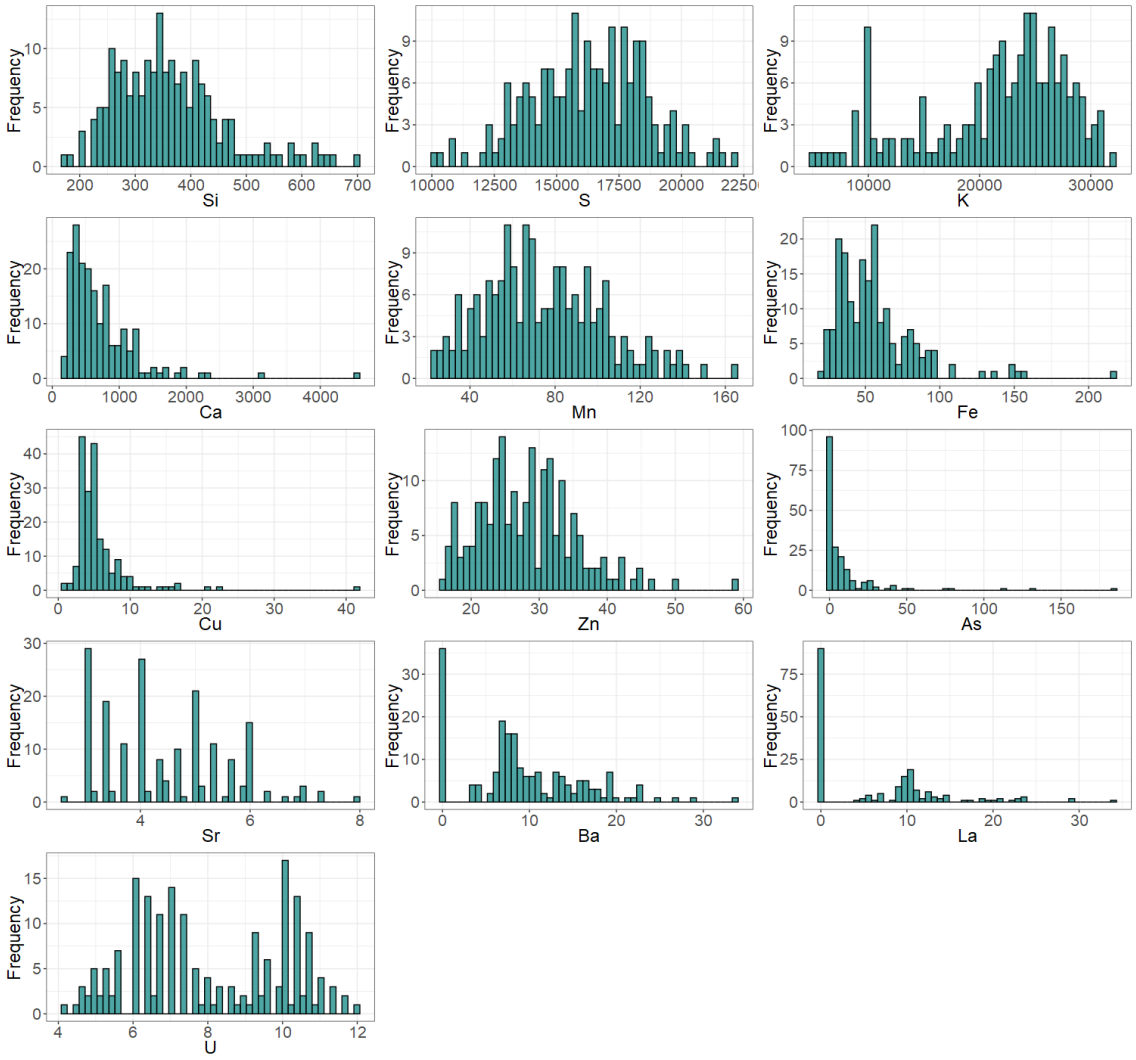


Figure 5.12 Frequency distributions of the elemental concentration (ppm) recorded for all selected elements measured by the Vanta analyser in Atlantic cod muscle tissue from all regions.

The mean concentrations (ppm) for each selected element per region recorded by the Vanta analyser are shown in Table 5.4, together with the standard deviations. Table 5.4 also contains the elemental concentrations found in the cod muscle Certified Reference Material BCR-422 (EVISA, 2022) to allow comparison with the Vanta data. The elemental composition of cod muscle tissue varies among the ten regions for the sampled individuals, with all selected elements except for lanthanum demonstrating a significant effect of region of origin (Kruskal-Wallis, d.f. = 9, $p < 0.001$). The variation in elemental concentrations among different regions for six selected elements is demonstrated in Figure 5.13. Silicon concentrations in the Irish Sea cod were higher than the majority of samples from other locations, with a mean of 623 ppm, which also distinguishes those individuals from their closest neighbours, the Celtic Sea cod, that have a mean concentration only slightly above half that of the Irish Sea. The Faroes samples have significantly lower concentrations of

potassium, with a mean of 9,581 ppm compared to 14,282 – 26,921 ppm in the other regions. Manganese shows a rough trend of increasing concentrations from north to south, since the most northerly samples in the Barents Sea have the lowest mean (29 ppm) and rising to 104 ppm further south in the Irish Sea. The Celtic Sea samples are the exception to this trend, which have a mean abundance more similar to the Norwegian Sea. The cod caught in the Baltic Sea in general have the lowest concentrations of zinc (mean of 23 ppm), but they also have the highest recorded zinc concentrations with very few outliers up to 59 ppm. Most of the cod from the Barents Sea contain a significantly higher abundance of arsenic than cod from other regions, since the mean concentration is 70 ppm compared to 0 – 15 ppm elsewhere. Cod from the Barents Sea have a large range of arsenic concentrations, from a low of 8 ppm reaching up to 184 ppm in one individual (Figure 5.13). Certain elements, such as barium and lanthanum, are present in low concentrations so show only slight variations in their mean abundance among regions and have overlapping ranges of values.

Although in general the elemental concentrations given by the Vanta analyser are in good agreement with the Certified Reference Material (CRM), as discussed previously, there are some variations among regions (Table 5.4). The estimates of potassium abundance by the Vanta are similar to the CRM in many regions, however the Faroes and Irish Sea samples have lower concentrations than the CRM (9,581 ppm and 14,282 ppm respectively) but are still of the same order of magnitude. Similarly, the calcium concentrations in the samples are in good agreement with the CRM overall, but a few regions including the Faroes, Baltic Sea and Irish Sea have higher recorded abundances of calcium in the samples than the CRM (up to 1,018 ppm in the samples compared with 330 ppm in the CRM). The concentrations of manganese, iron, copper and strontium are higher than the CRM in all regions, especially in the Rockall samples for iron and copper where the concentrations are 112 ppm and 13 ppm respectively compared to 5.46 ppm and 1.05 ppm in the CRM. The mean regional abundance of manganese is between 29 ppm and 104 ppm in the samples, significantly higher than 0.543 ppm as was measured in the CRM. Arsenic concentrations vary widely among regions, with some close to zero ppm on average and up to 70 ppm in the Barents Sea samples. The Norwegian and Celtic Sea samples have arsenic concentrations most similar to the CRM with means of 14 and 15 ppm.

Table 5.4 Mean abundance of the 13 selected elements (parts per million) for each region and the standard deviations (SD), measured in the cod muscle tissue samples using the Vanta analyser. The concentration of elements and their uncertainty in the cod muscle Certified Reference Material (CRM) BCR-422 (Quevauviller et al., 1992, EVISA, 2022), as determined by spectrometry, are also listed as a comparison where available (NA indicates that an element was not measured in the CRM).

	Elemental abundance (ppm)									
Region	**Si**		**S**		**K**		**Ca**		**Mn**	
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
Barents	298	71	15210	1093	22514	1561	342	141	29	4
Norwegian	382	68	16063	1734	26921	2970	398	200	61	28
Iceland	298	114	16628	1541	21590	3725	590	348	68	33
Faroes	453	83	16650	2055	9581	2178	1018	335	94	28
North Sea	311	76	17286	2201	25430	3805	791	728	78	18
WScotland	310	27	16266	1497	20747	2090	324	58	75	7
Rockall	340	61	15700	3013	21095	1225	506	133	65	13
Baltic	404	88	13629	1391	20054	3323	1007	548	96	14
Irish	623	78	11450	1233	14282	627	983	242	104	3
Celtic	327	37	18720	2092	24122	3070	884	548	63	17
CRM	**Value**	**+/-**	**Value**	**+/-**	**Value**	**+/-**	**Value**	**+/-**	**Value**	**+/-**
	NA	NA	11500	600	21700	800	330	90	0.543	0.028

	Elemental abundance (ppm)									
Region	**Fe**		**Cu**		**Zn**		**As**		**Sr**	
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
Barents	26	4	6	4	27	3	70	58	4	0
Norwegian	38	9	6	3	27	6	14	15	5	1
Iceland	44	20	5	3	26	7	1	2	4	1
Faroes	51	6	5	1	33	4	0	0	5	1
North Sea	69	13	6	4	29	6	6	5	4	1
WScotland	81	13	4	1	29	5	1	1	3	0
Rockall	112	72	13	16	28	8	0	1	4	1
Baltic	64	32	7	3	23	11	0	0	4	1
Irish	66	14	6	1	24	5	1	1	4	0
Celtic	71	45	4	1	33	4	15	11	6	1
CRM	**Value**	**+/-**	**Value**	**+/-**	**Value**	**+/-**	**Value**	**+/-**	**Value**	**+/-**
	5.46	0.30	1.05	0.07	19.6	0.5	21.1	0.5	0.66	0.12

Region	Elemental abundance (ppm)					
	Ba		La		U	
	Mean	SD	Mean	SD	Mean	SD
Barents	11	6	5	6	5	0
Norwegian	9	6	7	8	7	2
Iceland	9	6	4	6	7	2
Faroes	6	4	5	6	10	1
North Sea	8	5	8	7	9	2
WScotland	10	7	9	12	7	0
Rockall	4	6	0	0	7	1
Baltic	16	8	7	10	7	1
Irish	12	11	7	6	7	0
Celtic	8	5	8	6	9	1
CRM	**Value**	**+/-**	**Value**	**+/-**	**Value**	**+/-**
	NA	NA	NA	NA	NA	NA

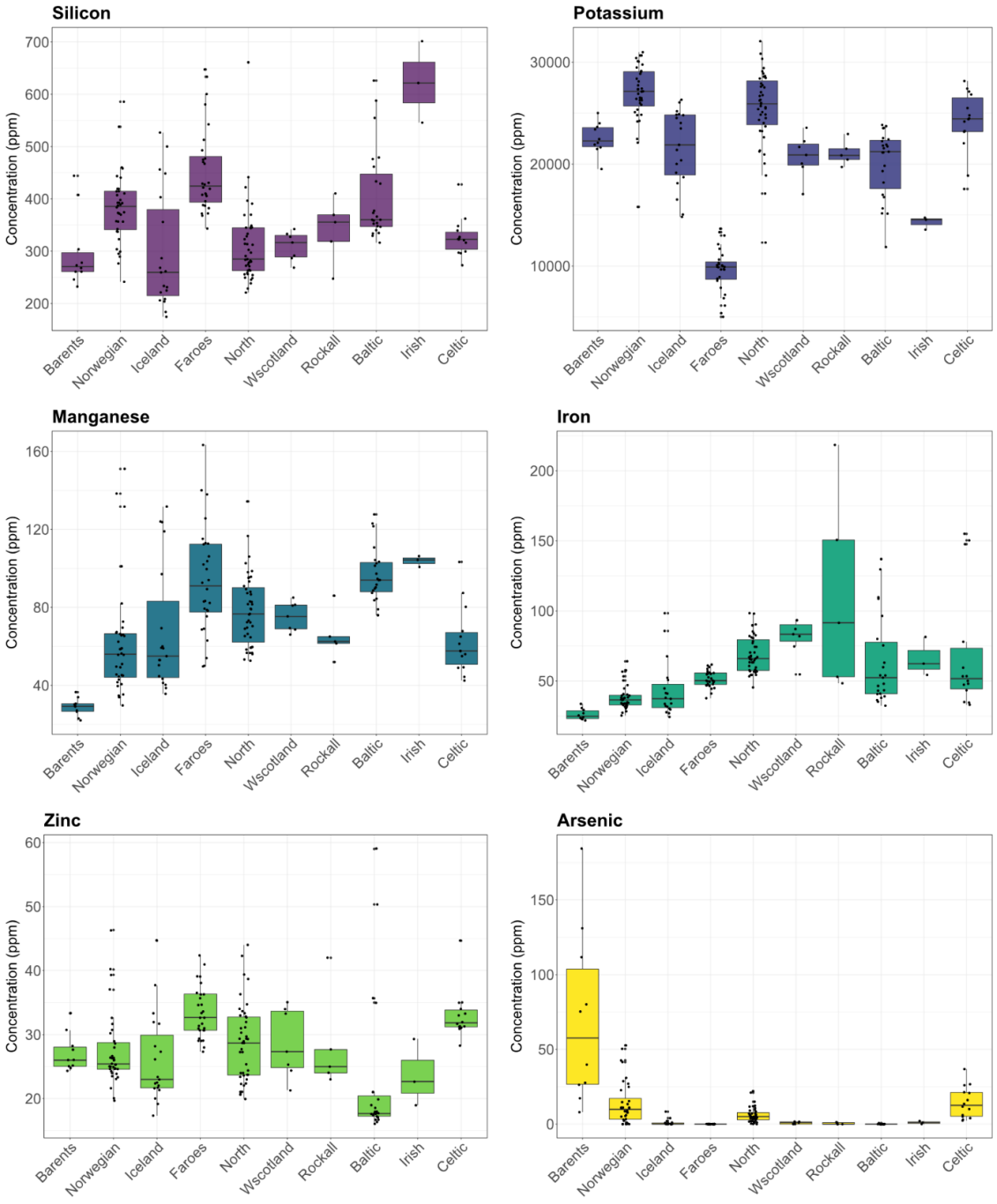


Figure 5.13 Elemental concentrations measured in the Atlantic cod muscle tissue samples from each geographic region, for six example elements – silicon, potassium, manganese, iron, zinc and arsenic.

5.4.3 Comparison between Itrax and Vanta results

The results from the Itrax scanner and Vanta analyser cannot be directly compared, because the Itrax records data as counts per second and is therefore semi-quantitative whereas the Vanta estimates concentrations in parts per million. However, the relative regional variations in a particular element can be compared between the two instruments.

There are many similarities between the results from the Itrax and Vanta, and in general the Vanta gave both absolute and relative concentrations that also agree with those in the cod muscle Certified Reference Material BCR-422 (EVISA, 2022). Both instruments recorded a much lower abundance of potassium in the Faroes samples with very few or no samples from other regions showing a concentration as low (Figure 5.14). The measurements of arsenic with both instruments are also in agreement, since the relative abundance of arsenic in cod caught in the Barents Sea is very much higher than any other sampled location – the mean concentration is at least 7.4 and 4.7 times greater than that of other regions recorded by the Itrax and Vanta respectively. Calcium measurements also show similarities, with both instruments recording a few outliers in the North Sea that contain much higher abundances than all other samples. The maximum concentration of calcium in the Baltic and Celtic Sea samples also reach higher values than most in both cases. Manganese has a general trend of increasing abundance from northern to southern regions, with the exception of the Celtic Sea samples which have a decreased manganese content compared to the Irish Sea and so appear similar to the more northerly locations, as recorded by both the Itrax and Vanta (Figure 5.14).

However, there are also several differences between the data recorded by the Itrax and Vanta. For example, higher counts per second of copper were recorded in cod from the Baltic and Irish Seas using the Itrax but with the Vanta these regions have similar mean concentrations to all other locations, and strontium shows regional variations in concentration with the Vanta whereas relatively constant counts per second across regions were recorded by the Itrax (Figure 5.14). Further differences appear in the presence or absence of some elements within the cod muscle tissue. The Itrax scanner recorded the presence of elements such as chromium, nickel and tin, but these were measured as zero ppm by the Vanta analyser. Selenium, cadmium and lead were also recorded as zero ppm by the Vanta in the vast majority of samples, and in the remaining few the concentrations were very low, whereas the Itrax scanner indicated that these elements are present in the cod samples. These differences in the presence of elements are likely due to detection limits of the instruments rather than true absence, or due to artefacts in the Itrax data from the user selecting elements to improve the modelled best fit and minimise the mean squared error. The expected concentrations of selenium, cadmium and lead in cod muscle are 1.63 ppm, 0.017 ppm and 0.085 ppm respectively based on the certified values in the cod muscle reference material BCR-422 (EVISA, 2022), which indicates that these elements are below the detection limits for the Vanta (5 ppm) (Olympus, 2022) and therefore suggests that the zero ppm Vanta records may be justified for these elements. Silicon was recorded with a mean concentration of 361 ppm by the Vanta but not recorded as being present by the Itrax. The Itrax cannot reliably obtain data for lighter elements such as silicon, since the fluorescent radiation does not have sufficient energy to reach the detector and leads to low or zero counts. However, the Vanta can better estimate

concentrations lighter elements because the detector makes contact with the sample surface, so in this case the data for silicon produced by the Vanta is more reliable than that from the Itrax.

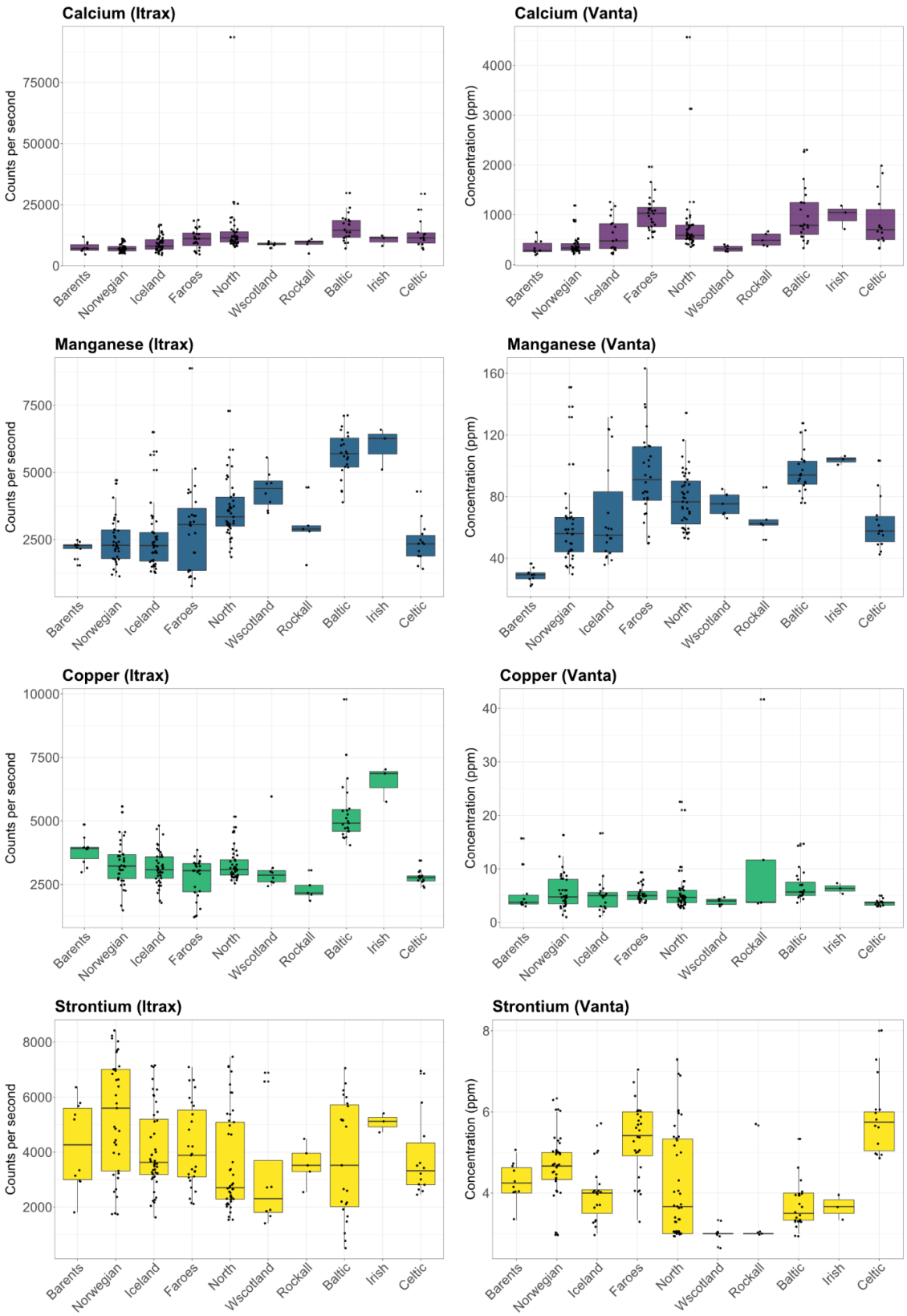


Figure 5.14 Comparison of the elemental abundance by region recorded using the Itrax scanner and Vanta analyser in the cod samples for calcium, manganese, copper and strontium.

5.4.4 Clustering by region of origin

5.4.4.1 Itrax

Principal Component Analysis (PCA) was first conducted using the raw element data for the 19 selected elements recorded by the Itrax scanner. This reveals evidence of clustering by region of origin, but there is a large amount of overlap among regions and the clusters are not clearly defined (Figure 5.15). Samples from the same location often display wide variation or outliers. The first and second principal components explain 26.8% and 14.1% of the variation among cod samples respectively. The Baltic Sea and Irish Sea cod form one of the slightly more distinct clusters, and the Rockall samples are generally fairly separate from the remaining locations on these principal components. However, the majority of samples are interspersed together and overlap with one or more other regions.

The PCA variables plot in Figure 5.16 was used to decide which elements to pair together when calculating the log ratios (since data from the Itrax scanner is semi-quantitative). Elements from different quadrants from the variables plot were used in combination to maximise the variation. Figure 5.16 also shows that scandium, selenium, lead and nickel contribute most to the variation among samples for these principal components, whereas chromium, chlorine and arsenic contribute the least.

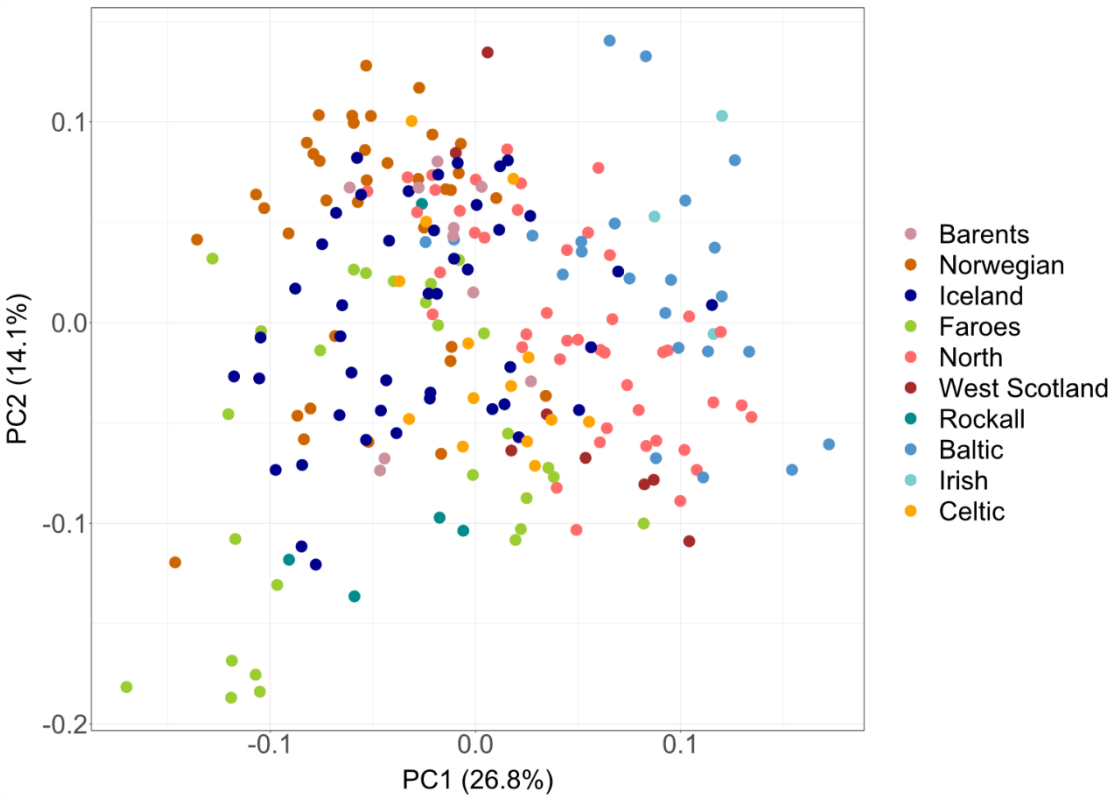


Figure 5.15 Principal components 1 and 2 using the raw elemental data from the Itrax scanner to show variability among the cod samples by region with Principal Component Analysis (PCA).

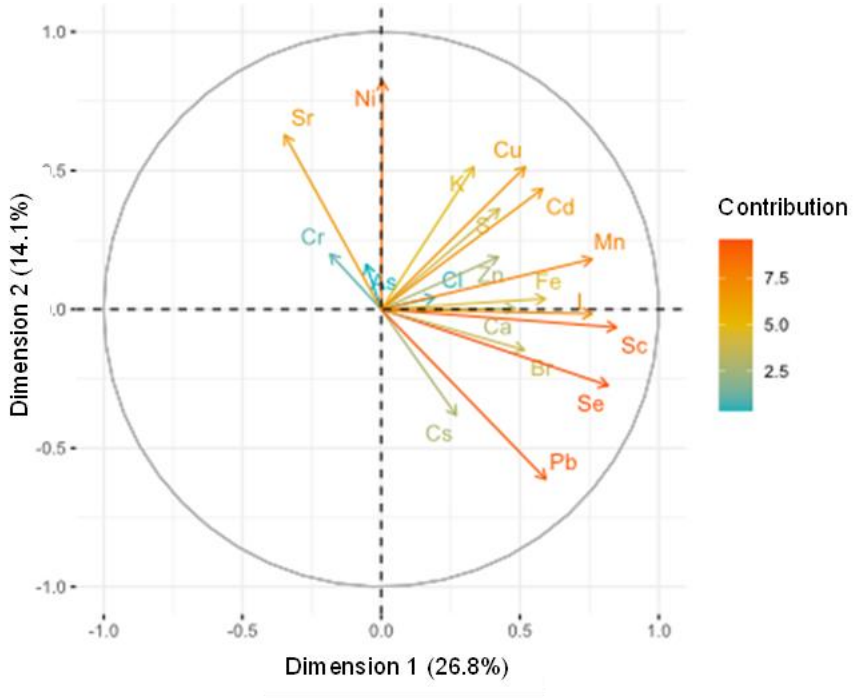


Figure 5.16 Principal Component Analysis variables plot, showing the elements and their contribution to the principal components.

Principal Component Analysis was also conducted using the log ratios of element pairs, selected based on the PCA using raw data combined with trial and error to achieve the best possible clustering by region. The element pairs are listed in Figure 5.18. Using these element log ratios, the first and second principal components explain 24.5% and 19.1% of the variation among cod samples respectively. Clustering by catch region is slightly improved compared to that with the raw element data, since the Norwegian and Faroes cod form clear clusters that are more distinct (Figure 5.17). However, some samples from these regions still overlap significantly with other clusters, and for other locations including the Barents, Iceland, Rockall and Celtic Sea the samples overlap completely with one another and cannot be differentiated.

The PCA variables plot in Figure 5.18 reveals that log ratio 3 (Cu:K) was the most important for separating the samples, and log ratio 8 (Cr:Sc) contributed least to the variation.

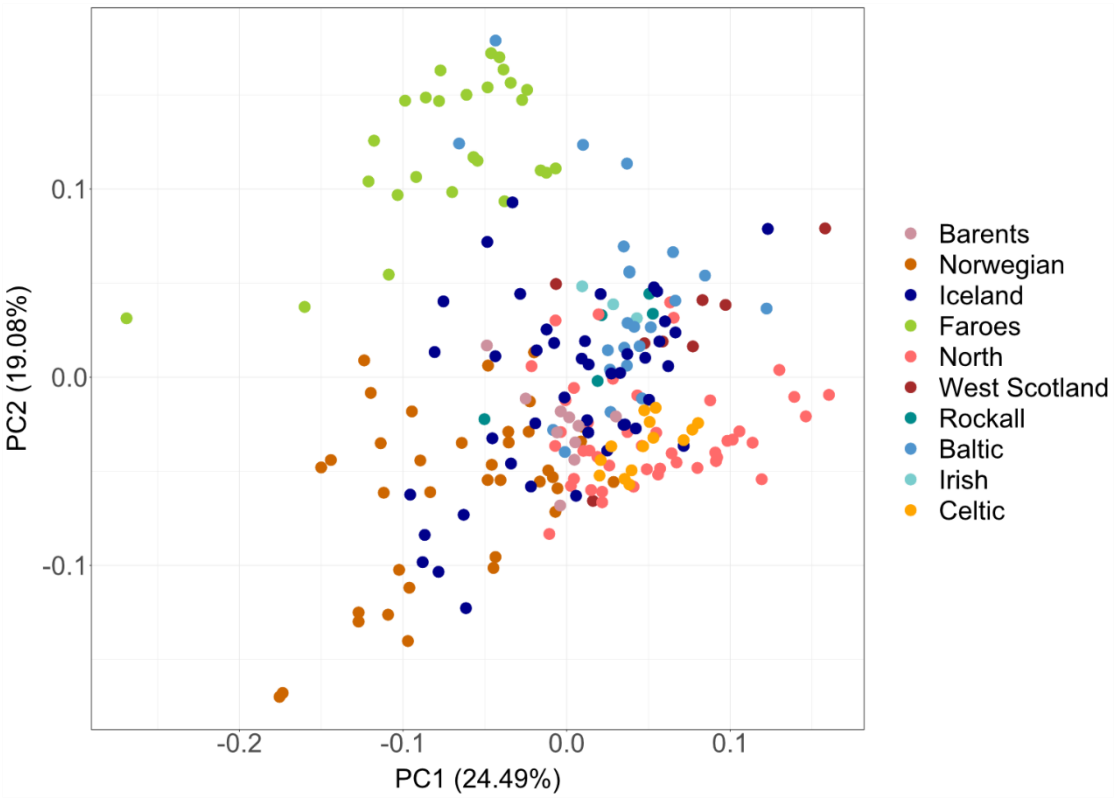


Figure 5.17 Principal components 1 and 2 using the log ratio elemental data from the Itrax scanner to show variability among the cod samples by region with Principal Component Analysis (PCA).

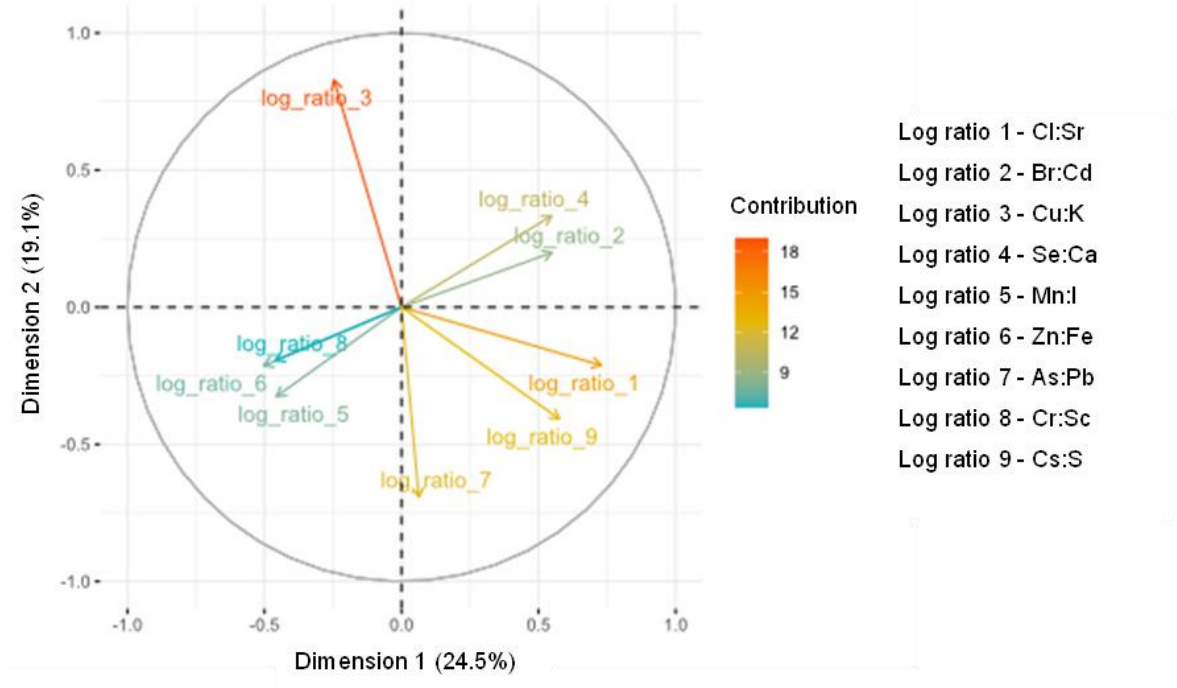


Figure 5.18 Principal Component Analysis variables plot, showing the element log ratios and their contribution to the principal components.

5.4.4.2 Vanta

Principal Component Analysis was carried out using the raw element data for the 13 selected elements measured by the Vanta analyser. Clustering by region of origin can be clearly observed, but the clusters overlap significantly, often with those of several other regions (Figure 5.19). Some regions, including the North Sea, vary widely between individuals on the first and second principal components. Other regions, such as the Norwegian Sea and Rockall, contain outliers that are far from the main cluster. The Iceland cod appear to be split into two smaller clusters in different parts of the plot. However, the Barents, Faroes and Celtic Sea all form relatively defined clusters, albeit with samples from other regions overlapping. The first principal component explains 27.6% of the variation among cod samples and the second principal component explains 16.4%. Manganese, uranium, silicon and sulfur contribute the most to the variation on principal components 1 and 2, whereas copper and lanthanum contribute the least (Figure 5.20).

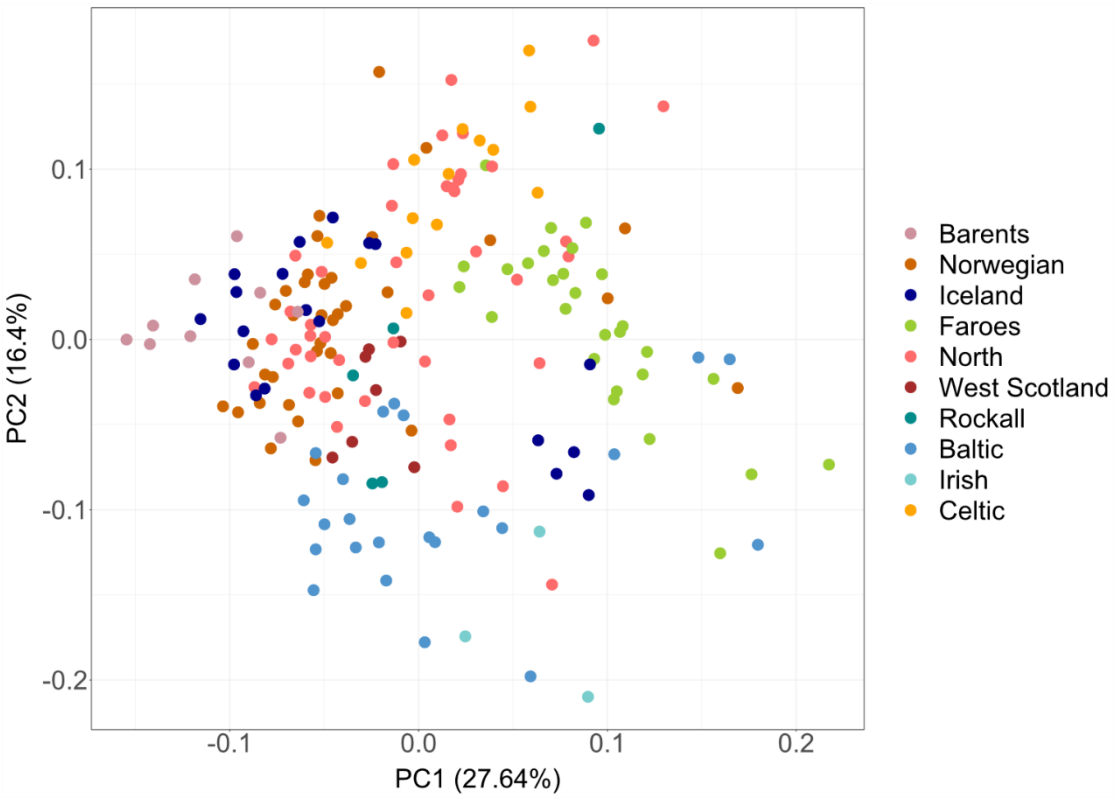


Figure 5.19 Principal components 1 and 2 using the raw elemental data from the Vanta analyser to show variability among the cod samples by region with Principal Component Analysis (PCA).

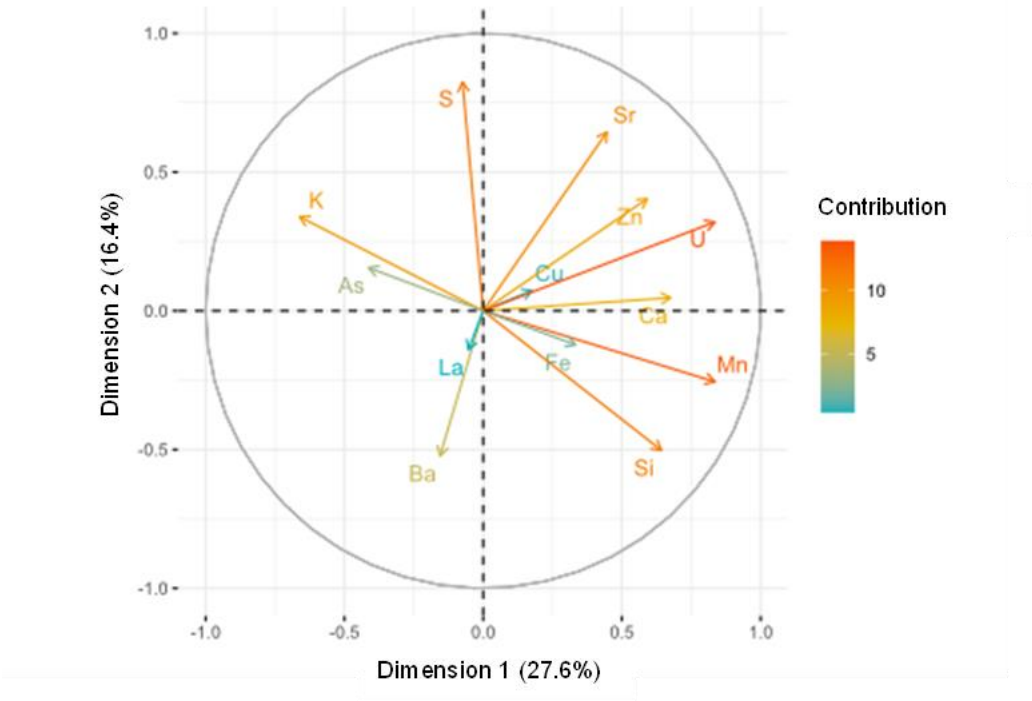


Figure 5.20 Principal Component Analysis variables plot, showing the elements and their contribution to the principal components.

Principal Component Analysis was then conducted using element log ratio data from the Vanta analyser. Log ratio element pairs were selected by trial and error to achieve the optimum clustering by region, matching elements from different quadrants in the raw PCA variables plot (Figure 5.20) as a starting point in order to maximise the variation among samples. Again, some clustering by catch region is visible, but the samples appear to be divided into two groups and samples from many regions straddle both groups (Figure 5.21). The clustering by region of origin appears to be less well defined compared with the PCA using raw element data, which indicates that it may be better to use the raw data for distinguishing among regions. All clusters show substantial overlap with several other regions on the first and second principal components. The Barents Sea, Faroes and Baltic Sea samples show the most distinct clusters, although these are still split across the two halves of the PCA plot.

The first principal component explains 44.9% of the variation among cod samples and the second principal component explains 17.6%. The first principal component using log ratio data therefore contains significantly more explained variance than the PCA using raw data, which explained 27.6% of the variation. The PCA variables plot in Figure 5.22 reveals that the log ratios U:Ba and Mn:K are most important for differentiating among the cod samples, whereas the log ratio Si:Zn contributes least to the variation. Potassium, as well as being one of the most important elements in the PCA, showed good agreement with the Certified

Reference Material (CRM) concentrations, whereas copper was one of the least important elements and did not have such good agreement with the CRM. However, although manganese appears to be an important element in the PCA using both raw and log ratio data, the concentrations recorded by the Vanta were significantly higher than in the CRM. Furthermore, zinc was in the least important log ratio pair, but had good agreement with the CRM concentrations. Therefore, the most important elements contributing to the variation among samples do not appear to necessarily be those that show good agreement between the Vanta and CRM concentrations.

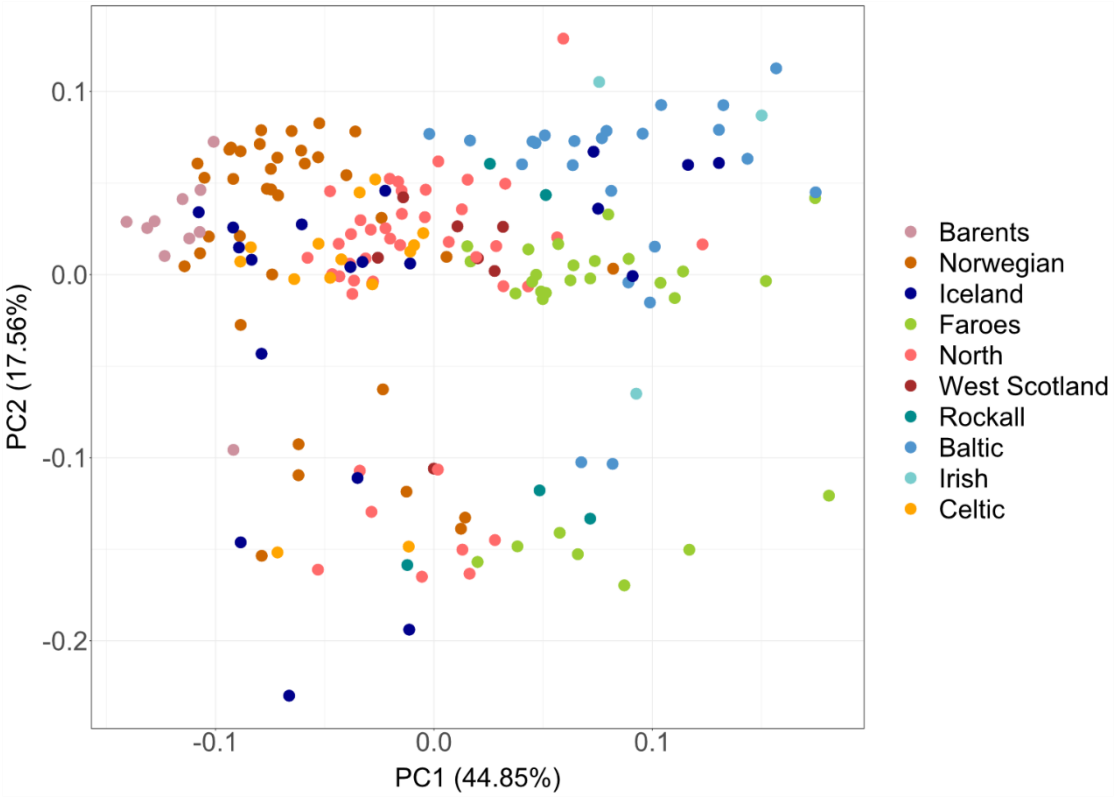


Figure 5.21 Principal components 1 and 2 using the log ratio elemental data from the Vanta analyser to show variability among the cod samples by region with Principal Component Analysis (PCA).

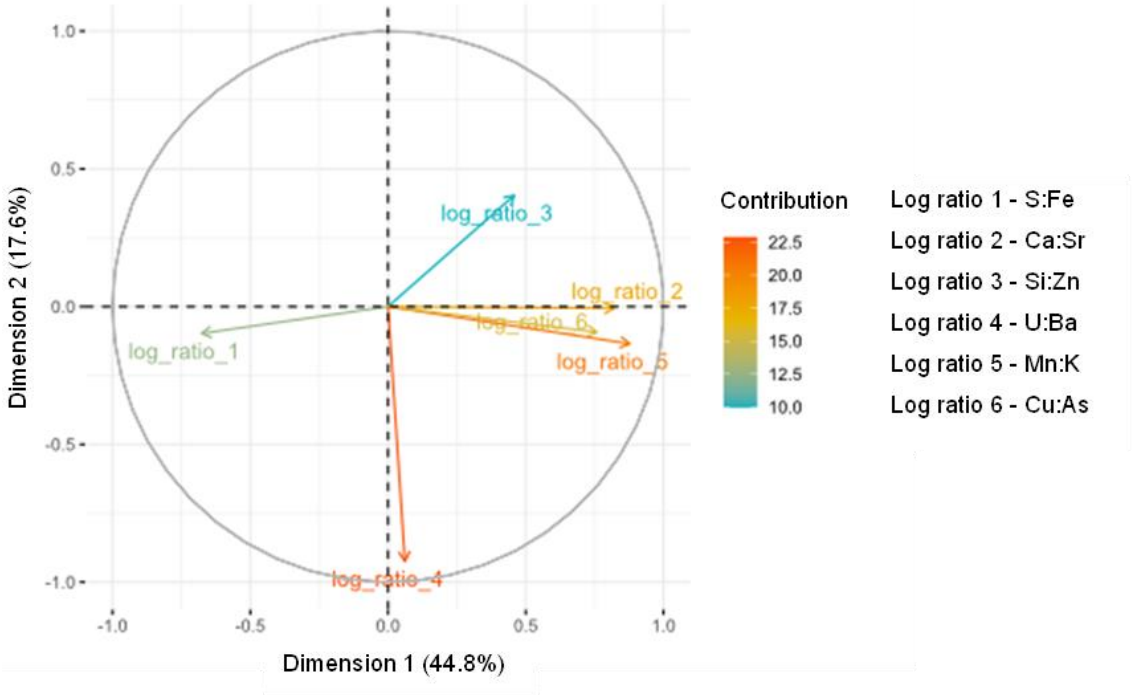


Figure 5.22 Principal Component Analysis variables plot, showing the element log ratios and their contribution to the principal components.

5.4.5 Assignment to region of origin

5.4.5.1 Itrax scanner

Random forest classification using the raw element data from the Itrax scanner resulted in an overall mean assignment accuracy of 83% for all regions. All the Barents Sea and Faroes samples were assigned correctly to origin (100%), and the Baltic cod also had a high assignment accuracy of 96% (Table 5.5). The cod from Iceland were more challenging to assign with only 48% correct, since a number of samples were incorrectly classified to several other regions, including 18% assigned to the Norwegian Sea. Although the Rockall samples could be assigned with relatively high accuracy (80%) and therefore indicates good potential for reliably tracing cod to this region, only five samples were collected and so a greater sample size is necessary to investigate this further. If Rockall was excluded from the assignment, similar assignment accuracies were obtained for all regions. Therefore, Rockall was included here to demonstrate the possible success rate that could be achieved. The Irish Sea cod were omitted from the assignment because only three were sampled and this was not sufficient to create a training subset for assigning a test sample.

Conducting the random forest classification using out-of-bag samples rather than leave-one-out cross validation (where each sample is removed sequentially and used as a test sample) gave an out-of-bag (OOB) error rate of 22.5% and so the error was slightly higher

than that from the cross-validation. The regions that were classified with the highest confidence using the out-of-bag samples were the Barents Sea and Faroes, and the least accurately assigned were samples from Iceland (Table 5.6), meaning both techniques were in close agreement overall. However, this method also gave a relatively high class error of 40% for the Rockall samples, which is higher than the error using leave-one-out cross validation and so indicates that a larger sample size is necessary to more accurately estimate the assignment accuracy.

The most important elements for the accurate assignment of cod to their origin were potassium and chlorine, which resulted in a mean decrease in accuracy of 46.5 and 33.4 respectively when each element is removed from the random forest model (Figure 5.23). Nickel and strontium contributed least to the assignment, with mean decrease accuracies of below 10.

Table 5.5 Assignment accuracies achieved for each sampled region (excluding Irish Sea) using raw elemental data recorded by the Itrax scanner, using random forest classification with leave-one-out cross validation and balanced sample sizes. The correct assignments are shown in bold.

Assigned origin	**True origin region – number assigned**								
	Barents	Norwegian	Iceland	Faroes	North Sea	West Scotland	Rockall	Baltic	Celtic
Barents	**10**	1	0	0	0	0	0	0	0
Norwegian	0	**28**	8	0	1	0	0	0	0
Iceland	0	4	**21**	0	5	0	0	1	2
Faroes	0	0	0	**27**	0	0	0	0	0
North Sea	0	3	6	0	**35**	1	0	0	1
West Scotland	0	0	3	0	0	**7**	1	0	0
Rockall	0	1	1	0	0	0	**4**	0	0
Baltic	0	0	2	0	3	0	0	**22**	0
Celtic	0	0	3	0	1	0	0	0	**11**
Percentage correct	**100%**	**76%**	**48%**	**100%**	**78%**	**88%**	**80%**	**96%**	**79%**

Table 5.6 Confusion matrix from random forest classification performed using the out-of-bag samples, based on the raw elemental data recorded by the Itrax scanner (excluding Irish Sea). The error for each class and the correct assignments are shown in bold.

Assigned origin	**True origin region – number assigned**								
	Barents	Norwegian	Iceland	Faroes	North Sea	West Scotland	Rockall	Baltic	Celtic
Barents	**10**	1	0	0	0	0	0	0	0
Norwegian	0	**28**	8	0	1	0	0	0	0
Iceland	0	4	**21**	0	5	0	0	1	2
Faroes	0	0	0	**27**	0	0	1	0	0
North Sea	0	3	6	0	**36**	1	0	0	1
West Scotland	0	0	3	0	0	**7**	1	0	0
Rockall	0	1	1	0	0	0	**3**	0	0
Baltic	0	0	2	0	2	0	0	**22**	0
Celtic	0	0	3	0	1	0	0	0	**11**
Class error	**0.000**	**0.243**	**0.523**	**0.000**	**0.200**	**0.125**	**0.400**	**0.043**	**0.214**

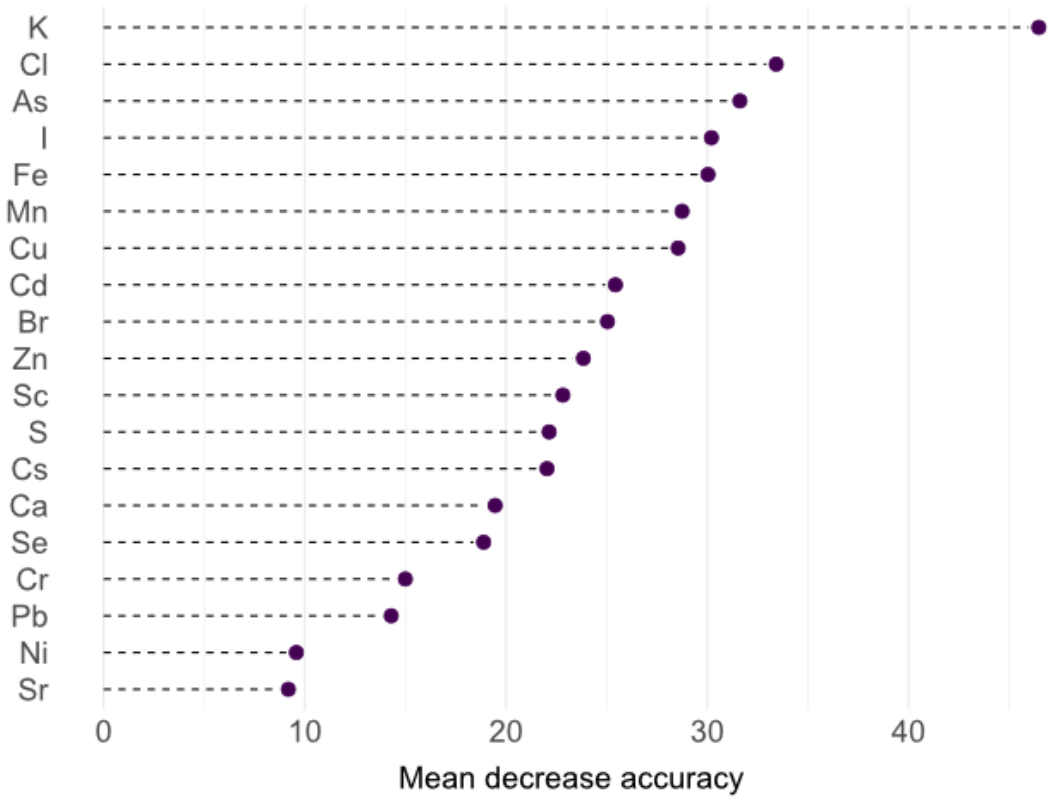


Figure 5.23 Random forest variable importance plot using the raw Itrax data, illustrating which elements contribute most to the accurate classification of cod to geographic origin. Mean decrease accuracy is the measure of performance of the model without each element, so a higher value indicates that the element is important in predicting the region of origin and removal of that element causes the model to lose accuracy of assignment.

When the elemental log ratio data were used for the random forest classification, the overall mean correct classification for all regions was 79%, which is slightly lower than that with the raw elemental data. Cod from several regions were assigned with the same accuracy as when the raw data were used, but the Celtic Sea samples had a 7% increase in correct assignments, making the success rate 86%. However, cod from West Scotland and the Baltic Sea showed a significant decrease in assignment accuracy (25% and 13% respectively) when the element log ratios were used, and therefore resulted in only 63% of West Scotland and 83% of Baltic cod being correctly assigned (Table 5.7).

The Rockall and Irish Sea cod were excluded from the assignment, because these had only five or fewer samples and this was not considered sufficient to represent the elemental composition of cod from these regions to create the training dataset. If included, the resulting assignment accuracies for Rockall and the Irish Sea were very low for this reason (0-20% correct).

The OOB error rate from random forest classification using out-of-bag samples was 26.0%, which is slightly higher than estimated using leave-one-out cross validation. However, the two methods were in close agreement on the assignment accuracies by region, with the highest accuracy for the Barents Sea and Faroes samples, and the lowest accuracy for those from Iceland (Table 5.8).

The variable importance plot in Figure 5.24 reveals that the log ratio of Cu:K contributed the most to correct classification, whereas the log ratios of Cr:Sc and Se:Ca were the least important for assignment to origin. Potassium is therefore an important element for classification of both raw and log ratio data from the Itrax, whereas calcium, selenium and chromium were relatively unimportant for both.

Table 5.7 Assignment accuracies achieved for each sampled region (excluding Rockall and Irish Sea) using log ratio elemental data from the Itrax scanner, using random forest classification with leave-one-out cross validation and balanced sample sizes. The correct assignments are shown in bold.

Assigned origin	**True origin region – number assigned**							
	Barents	Norwegian	Iceland	Faroes	North Sea	West Scotland	Baltic	Celtic
Barents	**10**	2	1	0	0	0	0	1
Norwegian	0	**27**	8	0	1	0	0	0
Iceland	0	5	**20**	0	3	1	0	0
Faroes	0	0	0	**27**	0	0	3	0
North Sea	0	3	4	0	**36**	1	0	1
West Scotland	0	0	6	0	1	**5**	1	0
Baltic	0	0	2	0	1	1	**19**	0
Celtic	0	0	3	0	3	0	0	**12**
Percentage correct	**100%**	**73%**	**46%**	**100%**	**80%**	**63%**	**83%**	**86%**

Table 5.8 Confusion matrix from random forest classification performed using the out-of-bag samples, based on the log ratio elemental data from the Itrax scanner (excluding Rockall and Irish Sea). The error for each class and the correct assignments are shown in bold.

Assigned origin	**True origin region – number assigned**							
	Barents	Norwegian	Iceland	Faroes	North Sea	West Scotland	Baltic	Celtic
Barents	**10**	2	1	0	0	0	0	1
Norwegian	0	**28**	10	0	1	0	0	0
Iceland	0	4	**18**	0	3	1	0	0
Faroes	0	0	0	**27**	0	0	3	0
North Sea	0	3	5	0	**35**	1	0	1
West Scotland	0	0	4	0	1	**5**	1	0
Baltic	0	0	3	0	1	1	**19**	0
Celtic	0	0	3	0	4	0	0	**12**
Class error	**0.000**	**0.243**	**0.591**	**0.000**	**0.222**	**0.375**	**0.174**	**0.143**

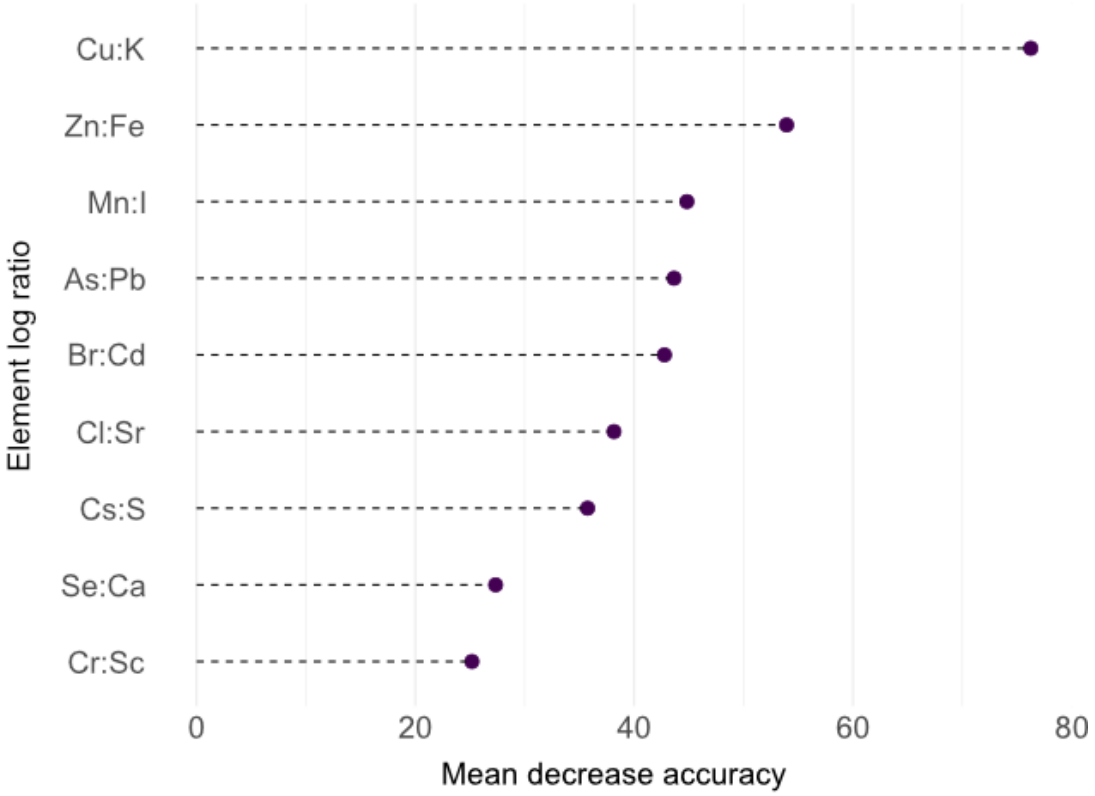


Figure 5.24 Random forest variable importance plot using the Itrax log ratio data, illustrating which log ratios of element pairs contribute most to the accurate classification of cod to geographic origin. Mean decrease accuracy is the measure of performance of the model without each log ratio pair, so a higher value indicates that the log ratio is important in predicting the region of origin and removal of that log ratio causes the model to lose accuracy of assignment.

5.4.5.2 Vanta analyser

Random forest classification using the raw elemental data from the Vanta analyser gave an overall mean assignment accuracy of 86%. Therefore the assignment accuracies were generally high, with all but two regions achieving a correct assignment rate of over 80% (Table 5.9). The Barents Sea, Faroes, North Sea and West Scotland samples were all assigned with 90-100% accuracy, but the cod from Iceland and the Norwegian Sea were more challenging to assign and had lower success rates of 68% and 69% respectively. In some cases, the incorrect assignments were assigned instead to the nearest neighbouring region, for example Barents and Norwegian Sea cod were assigned to each other, but on other occasions samples were assigned to more distant regions.

Due to their small sample sizes, the Rockall and Irish Sea were excluded from the assignment, but if included neither had any samples assigned correctly. A larger sample

size is necessary to determine the accuracy with which cod from these two regions can be traced back to their origin.

Random forest classification with out-of-bag samples resulted in an OOB error rate of 16.1%, which is in good agreement with the leave-one-out cross validation method. The error rates for all individual regions are also very similar to the cross-validation, and it supports the regions with the lowest assignment accuracies as being Iceland and the Norwegian Sea (Table 5.10).

Figure 5.25 shows the most important elements for the accurate assignment to origin using the raw Vanta data. Potassium and arsenic give the greatest contribution to accurate assignment of the cod samples, resulting in a mean decrease in accuracy of 62.4 and 43.5 respectively when each element is removed from the random forest model. Whereas copper, barium and lanthanum are of least importance, with mean decrease accuracies of less than 20. In fact, lanthanum shows negative importance (-1.8), meaning that it worsens the assignment accuracy overall. However, there are certain regions (Norwegian Sea, Faroes and Iceland) where lanthanum does improve the success rate so it was not removed from the assignment.

Table 5.9 Assignment accuracies achieved for each sampled region (excluding Rockall and Irish Sea) using raw elemental data recorded by the Vanta analyser, using random forest classification with leave-one-out cross validation and balanced sample sizes. The correct assignments are shown in bold.

Assigned origin	**True origin region – number assigned**							
	Barents	Norwegian	Iceland	Faroes	North Sea	West Scotland	Baltic	Celtic
Barents	**9**	4	1	0	0	0	0	0
Norwegian	1	**25**	0	0	0	0	0	1
Iceland	0	0	**13**	0	1	0	2	0
Faroes	0	1	0	**28**	1	0	2	0
North Sea	0	1	0	0	**39**	0	0	1
West Scotland	0	1	0	0	2	**7**	0	0
Baltic	0	1	4	0	0	0	**19**	0
Celtic	0	3	1	0	0	0	0	**12**
Percentage correct	**90%**	**69%**	**68%**	**100%**	**91%**	**100%**	**83%**	**86%**

Table 5.10 Confusion matrix from random forest classification performed using the out-of-bag samples, based on the raw elemental data recorded by the Vanta analyser (excluding Rockall and Irish Sea). The error for each class and the correct assignments are shown in bold.

Assigned origin	**True origin region – number assigned**							
	Barents	Norwegian	Iceland	Faroes	North Sea	West Scotland	Baltic	Celtic
Barents	**9**	4	1	0	0	0	0	0
Norwegian	1	**24**	0	0	0	0	0	1
Iceland	0	1	**13**	0	1	0	2	0
Faroes	0	1	0	**28**	1	0	1	0
North Sea	0	1	0	0	**38**	0	0	1
West Scotland	0	0	0	0	3	**7**	0	0
Baltic	0	2	4	0	0	0	**20**	0
Celtic	0	3	1	0	0	0	0	**12**
Class error	**0.100**	**0.333**	**0.316**	**0.000**	**0.116**	**0.000**	**0.130**	**0.143**

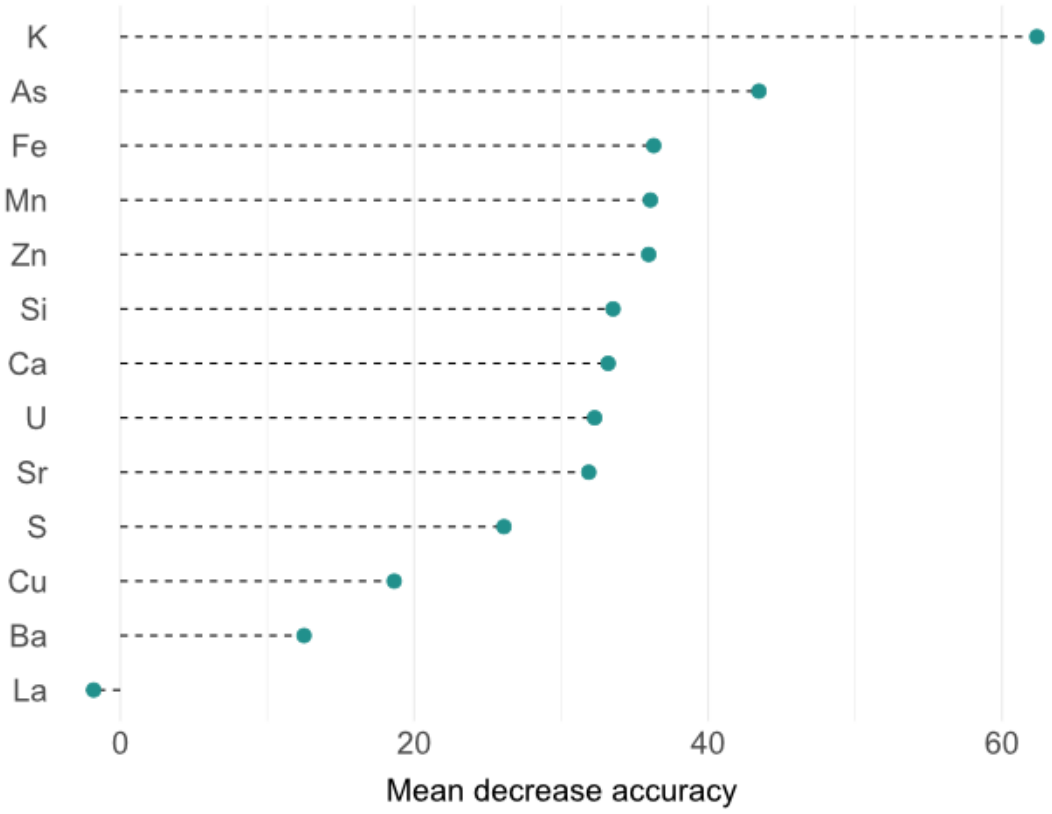


Figure 5.25 Random forest variable importance plot using raw data from the Vanta analyser, illustrating which elements contribute most to the accurate classification of cod to geographic origin. Mean decrease accuracy is the measure of performance of the model without each element, so a higher value indicates that the element is important in predicting the region of origin and removal of that element causes the model to lose accuracy of assignment.

Random forest classification was also conducted using the element log ratio data from the Vanta analyser, which resulted in an overall mean assignment accuracy of 73% for all regions using leave-one-out cross validation. Therefore, using the log ratio Vanta data resulted in a reduced assignment accuracy compared to using the raw data. With the exception of one region, the assignment accuracy for individual regions is also lower than using the raw data – a decrease of between 10% (Barents) and 26% (Iceland) when using the element log ratios. Only the Baltic Sea increased by 4% to a success rate of 87% using the log ratio data. However, both are in agreement that the Icelandic and Norwegian Sea cod were the most difficult to assign reliably, since these regions also had the lowest success rates using the log ratio data (42% and 50% respectively) (Table 5.11). The regions to which cod could be traced with the highest accuracy were similar to using the raw data - Faroes, Baltic Sea and West Scotland (89-86%) – although the assignment accuracies for the Faroes and West Scotland were slightly lower using log ratios. In the

same way as when using the raw Vanta data, if the Rockall and Irish Sea samples were included none were correctly assigned using the log ratio data.

The OOB error rate from random forest classification conducted with out-of-bag samples was 28.9%, which is similar to the estimate using random forest with cross validation. Furthermore, Iceland and the Norwegian Sea had the highest error rates of the individual regions, also in agreement with the cross validation (Table 5.12).

The variable importance plot in Figure 5.26 demonstrates that the log ratios of Cu:As and Mn:K are the most important in correctly classifying the cod samples to catch region, with mean decrease accuracies of 64.4 and 61.5 respectively when these ratios were excluded from the assignment. The log ratio of U:Ba contributes least to the assignment accuracy. Potassium and arsenic were also the most important elements using the raw data, and similarly barium was one of the least important, whereas copper was important as part of a log ratio pair but contributed relatively little as a single element.

Table 5.11 Assignment accuracies achieved for each sampled region (excluding Rockall and Irish Sea) using log ratio elemental data from the Vanta analyser, using random forest classification with leave-one-out cross validation and balanced sample sizes. The correct assignments are shown in bold.

Assigned origin	**True origin region – number assigned**							
	Barents	Norwegian	Iceland	Faroes	North Sea	West Scotland	Baltic	Celtic
Barents	**8**	8	2	0	0	0	0	0
Norwegian	2	**18**	3	0	0	0	0	0
Iceland	0	4	**8**	1	0	0	0	0
Faroes	0	2	2	**25**	1	0	3	0
North Sea	0	3	0	0	**32**	1	0	3
West Scotland	0	0	0	0	6	**6**	0	1
Baltic	0	0	4	2	0	0	**20**	0
Celtic	0	1	0	0	4	0	0	**10**
Percentage correct	**80%**	**50%**	**42%**	**89%**	**74%**	**86%**	**87%**	**71%**

Table 5.12 Confusion matrix from random forest classification performed using the out-of-bag samples, based on the log ratio elemental data from the Vanta analyser (excluding Rockall and Irish Sea). The error for each class and the correct assignments are shown in bold.

Assigned origin	**True origin region – number assigned**							
	Barents	Norwegian	Iceland	Faroes	North Sea	West Scotland	Baltic	Celtic
Barents	**8**	5	2	0	0	0	0	0
Norwegian	2	**20**	3	0	0	0	0	0
Iceland	0	4	**8**	0	0	0	0	0
Faroes	0	2	2	**26**	3	0	3	0
North Sea	0	3	0	0	**30**	1	0	3
West Scotland	0	0	0	0	5	**6**	0	1
Baltic	0	0	4	2	1	0	**20**	0
Celtic	0	2	0	0	4	0	0	**10**
Class error	**0.200**	**0.444**	**0.579**	**0.071**	**0.302**	**0.143**	**0.130**	**0.286**

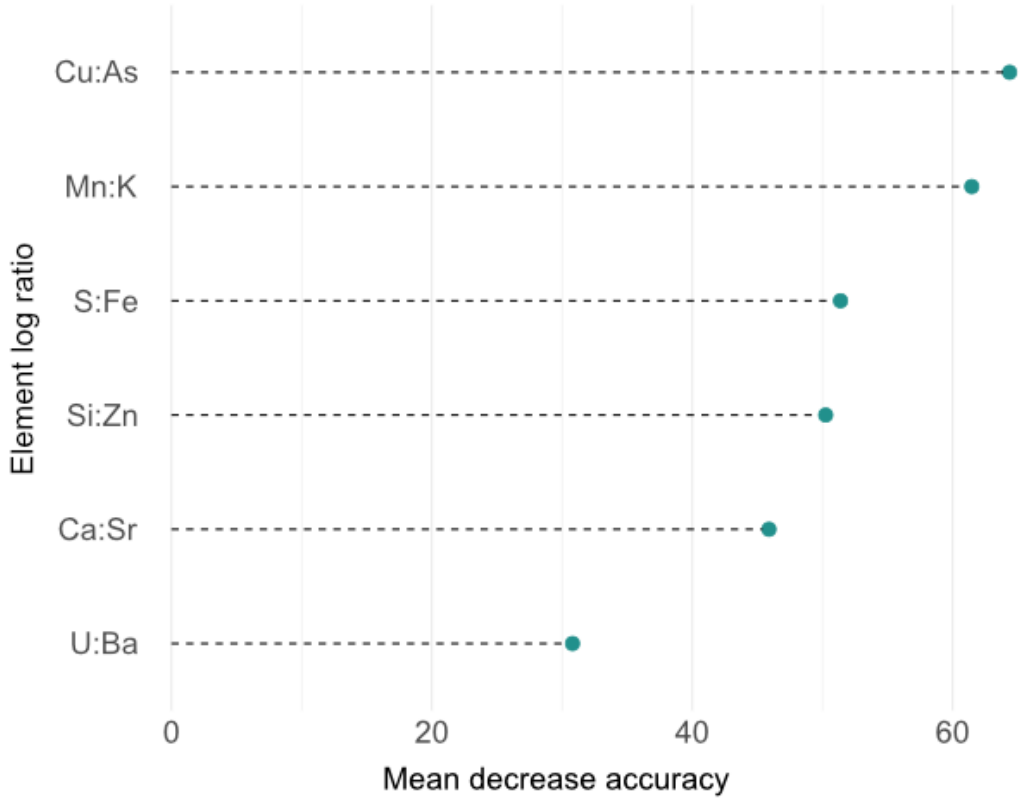


Figure 5.26 Random forest variable importance plot using log ratio data from the Vanta analyser, illustrating which log ratio element pairs contribute most to the accurate classification of cod to geographic origin. Mean decrease accuracy is the measure of performance of the model without each log ratio element pair, so a higher value indicates that the log ratio is important in predicting the region of origin and removal of that log ratio causes the model to lose accuracy of assignment.

5.4.5.3 Comparison between Itrax scanner and Vanta analyser

The overall mean success rate of assignment to origin using the Itrax scanner was slightly higher with the raw data rather than the log ratio data (83% and 79% respectively). The Vanta analyser also performed better using the raw data (86% compared to 73% with the log ratio data) (Table 5.13). The Itrax and Vanta each gave high accuracies of assignment using raw data, although the success rate was slightly higher using the Vanta analyser. Considering individual regions, the Itrax more reliably assigned the Barents, Norwegian and Baltic Sea cod to origin (an increase of 10%, 7% and 9% respectively) whereas the Vanta assigned the Icelandic, North Sea and West Scotland cod more accurately (between 12% and 20% greater success rate) (Table 5.13). The Faroes and Celtic Sea cod were assigned with identical accuracy using both instruments (100% and 86% respectively). Therefore, overall the Vanta analyser using raw data gave the highest assignment accuracy, although the Itrax may perform better for certain individual regions.

Table 5.13 Comparison between the assignment accuracies achieved by the Itrax scanner and Vanta analyser using both raw and element log ratio data with random forest classification.

Origin region	Itrax - correct assignments (%)		Vanta - correct assignments (%)	
	Raw data	Log ratio data	Raw data	Log ratio data
Barents	100	100	90	80
Norwegian	76	73	69	50
Iceland	48	46	68	42
Faroes	100	100	100	89
North Sea	78	80	91	74
West Scotland	88	63	100	86
Baltic	96	83	83	87
Celtic	79	86	86	71
Overall mean	**83%**	**79%**	**86%**	**73%**

5.4.5.4 Results with unfiltered data

To investigate the importance of filtering the data prior to spatial assignment, random forest classification was also conducted using the processed data from the Itrax and Vanta without data cleaning, where data were not removed from the dataset regardless of the roughness or unevenness of the sample surface (although datapoints with a validity of zero and those with mean squared error of 5 or greater were still removed). The results demonstrate that the overall accuracy of assignment to region of origin using raw data (rather than log ratio data) is very similar with and without prior filtering of the data (Table 5.14), provided that all samples are ground to a homogeneous powder. The success rates for individual regions were also very similar, with only slight differences for certain locations, for example the assignment accuracy decreased from 96% to 83% for the Baltic Sea using the unfiltered data from the Itrax. This suggests that the surface roughness or how flat the sample surface is may not have a significant effect on the reliability of data recorded by the Itrax or Vanta when samples to be analysed are ground to a fine powder. However, there may be a greater effect when more coarsely ground samples are used, as described previously (see Figure 5.6).

Table 5.14 Comparison between the assignment accuracies achieved by the Itrax scanner and Vanta analyser with both raw filtered and raw unfiltered data using random forest classification.

Origin region	Itrax - correct assignments (%)		Vanta - correct assignments (%)	
	Filtered data	Unfiltered data	Filtered data	Unfiltered data
Barents	100	100	90	90
Norwegian	76	83	69	72
Iceland	48	45	68	68
Faroes	100	100	100	100
North Sea	78	82	91	92
West Scotland	88	88	100	100
Baltic	96	83	83	91
Celtic	79	79	86	79
Overall mean	**83%**	**82%**	**86%**	**87%**

5.4.5.5 Combining XRF data with stable isotope ratios for assignment to origin

The ability of two contrasting techniques used in combination to assign cod to origin was also investigated to determine whether the assignment accuracy increases when more than one technique is used at once. The raw elemental concentration data acquired by the Vanta analyser was combined with stable isotope data measured as part of Chapter 2 (lipid corrected $\delta^{13}C$, $\delta^{15}N$ and $\delta^{34}S$ values) using the same Atlantic cod samples. The decision was made to use the Vanta for this assignment rather than the Itrax data, because the assignment accuracies obtained using the Vanta were slightly higher and since it is portable it is more promising for widespread use on traded fish products. Random forest classification using the XRF and stable isotope datasets combined gave an overall mean assignment accuracy of 91% (excluding Irish Sea), which is higher than the accuracy obtained using either of the two methods independently (86% using the Vanta and 79% using stable isotopes). The assignment accuracies for each region using XRF and stable isotope techniques together are listed in Table 5.15. All regions except for Iceland had an assignment accuracy greater than 80%, and six of the nine regions had an accuracy over 90%. The Faroes, West Scotland and Rockall cod could be assigned with the highest accuracy (100%), whereas cod from Iceland were still the most challenging to assign. Icelandic cod as well as those from the Celtic Sea could be assigned more reliably using both techniques in combination than either technique used separately, showing an improvement of between 6% and 18% (Table 5.13). Other regions, including the Norwegian, North and Baltic Seas, had a higher success rate using one of the techniques independently, but overall for all regions the assignment accuracies were more consistently

high using the combination of both approaches. Therefore, the ability to reliably assign cod to origin was improved by combining XRF and stable isotope approaches, since the best distinguishing power of each technique was incorporated to result in higher assignment accuracies overall for all the sampled regions.

Conducting the random forest classification using out-of-bag samples, rather than leave-one-out cross validation (where each sample is removed sequentially and used as a test sample) gave an OOB error rate of 12.4% and so the error was slightly higher than that from the cross-validation. The regions that were classified with the highest confidence were the Faroes, West Scotland and Rockall, and the least accurately assigned were samples from Iceland (Table 5.16), so these results are in close agreement overall with the cross-validation.

The $\delta^{13}C$ values (lipid corrected) and potassium concentration were found to be the most important variables for accurately assigning cod to region of origin, resulting in a mean decrease in accuracy of 58.1 and 56.5 respectively when each element is removed from the random forest model (Figure 5.27). Copper, barium and lanthanum contributed least to the assignment, with mean decrease accuracies of below 10.

Table 5.15 Assignment accuracies achieved for each sampled region (excluding Irish Sea) using the raw elemental data recorded by the Vanta analyser combined with the stable isotope data (carbon, nitrogen and sulfur) from Chapter 2 for the same samples, using leave-one-out cross validation and balanced sample sizes. The correct assignments are shown in bold.

Assigned origin	**True origin region – number assigned**								
	Barents	Norwegian	Iceland	Faroes	North Sea	West Scotland	Rockall	Baltic	Celtic
Barents	**9**	4	2	0	0	0	0	0	0
Norwegian	1	**30**	0	0	0	0	0	0	0
Iceland	0	0	**14**	0	1	0	0	0	0
Faroes	0	0	1	**28**	1	0	0	1	0
North Sea	0	0	0	0	**37**	0	0	0	1
West Scotland	0	0	0	0	2	**7**	0	0	0
Rockall	0	0	0	0	0	0	**5**	1	0
Baltic	0	1	1	0	0	0	0	**21**	0
Celtic	0	1	1	0	2	0	0	0	**13**
Percentage correct	**90%**	**83%**	**74%**	**100%**	**86%**	**100%**	**100%**	**91%**	**93%**

Table 5.16 Confusion matrix from random forest classification performed using the out-of-bag samples, based on the raw elemental data recorded by the Vanta analyser combined with stable isotope data (carbon, nitrogen and sulfur) from Chapter 2 for the same samples (excluding Irish Sea). The error for each class and the correct assignments are shown in bold.

Assigned origin	**True origin region – number assigned**								
	Barents	Norwegian	Iceland	Faroes	North Sea	West Scotland	Rockall	Baltic	Celtic
Barents	**9**	4	2	0	0	0	0	0	0
Norwegian	1	**30**	0	0	0	0	0	0	0
Iceland	0	0	**13**	0	1	0	0	0	0
Faroes	0	0	1	**28**	1	0	0	1	0
North Sea	0	0	0	0	**36**	0	0	0	1
West Scotland	0	0	0	0	3	**7**	0	0	0
Rockall	0	0	0	0	0	0	**5**	1	0
Baltic	0	2	2	0	0	0	0	**21**	0
Celtic	0	0	1	0	2	0	0	0	**13**
Class error	**0.100**	**0.167**	**0.316**	**0.000**	**0.163**	**0.000**	**0.000**	**0.087**	**0.071**

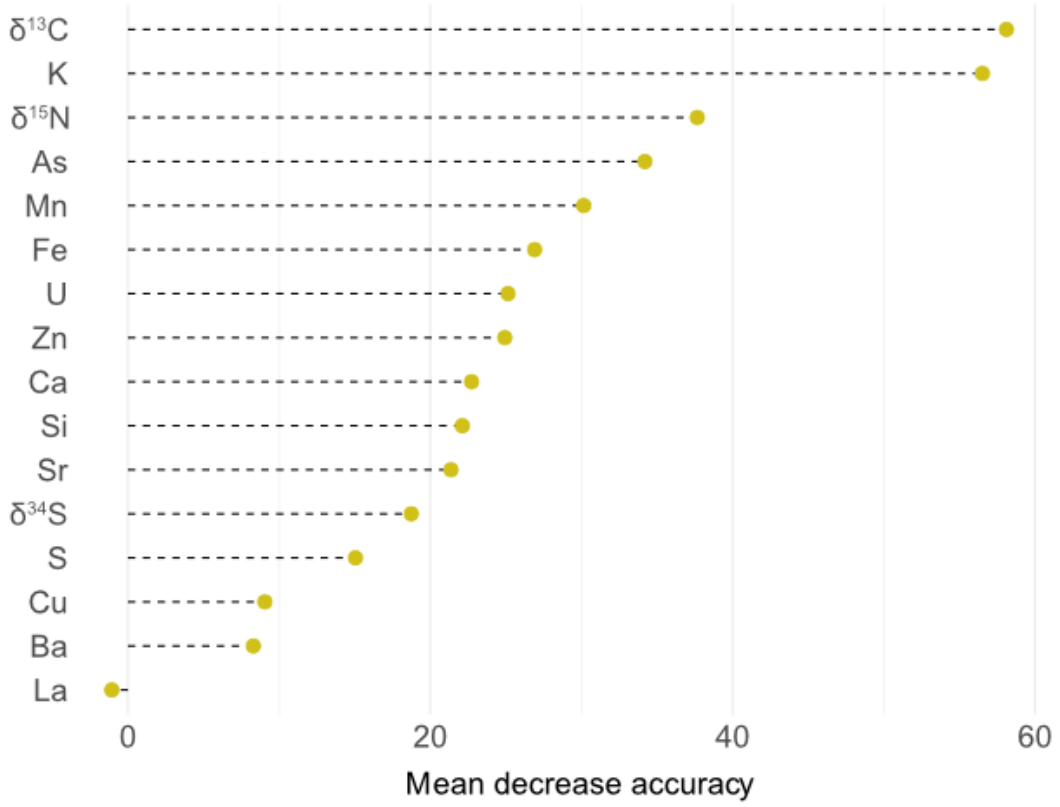


Figure 5.27 Random forest variable importance plot using raw XRF data from the Vanta analyser combined with stable isotope data from Chapter 2 using the same cod samples, illustrating which elements contribute most to the accurate classification of cod to geographic origin. Mean decrease accuracy is the measure of performance of the model without each element, so a higher value indicates that the element is important in predicting the region of origin and removal of that element causes the model to lose accuracy of assignment.

5.5 Discussion

The results of this study demonstrate that XRF instruments (Itrax scanner and Vanta analyser) have promising potential to differentiate whitefish muscle samples sourced from different catch regions. The methodology used in this opportunistic pilot study imposed some analytical challenges that may be avoided with fresh muscle samples, nevertheless, optimisation to improve the success rate of assignment is needed to establish XRF as a more viable option for widespread use. In our pilot study, the elemental composition of cod muscle tissue varied with spatial origin encouraging further investigation into whether XRF analysis could be used as a forensic test for the verification of provenance for traded whitefish products.

5.5.1 Reliability and accuracy of trace element data

The elemental concentrations recorded in the cod muscle samples show good agreement with those reported for cod muscle Certified Reference Material (CRM) BCR-422 (Quevauviller et al., 1992, EVISA, 2022). Elements found in the highest abundance were potassium, sulfur and calcium and the concentrations of these elements as estimated by the Vanta analyser were of similar magnitude to those reported in the CRM. Manganese and iron concentrations were significantly higher in the Vanta data than reported in the CRM, but as we were not able to analyse the CRM directly, it is not clear if these differences reflect systematic analytical error, or relatively high Fe and Mn concentrations in our wild fish. Broadly, however, both relative and absolute concentration estimates recovered from the Itrax and Vanta were consistent with certified values in cod muscle, giving confidence in the results from both XRF instruments.

Concentrations of trace metals in Atlantic cod muscle tissue have been reported in several other published studies (Franklin, 1987, Hellou et al., 1992, Brown and Balls, 1997, Henry et al., 2004, Julshamn et al., 2004, Julshamn et al., 2013). Zinc was reported to be 2-8 µg/g wet weight in cod muscle (Rainbow, 2018), which is equivalent to 10–40 µg/g dry weight, since dry weight concentrations are approximately five times wet weight concentrations (Rainbow, 2018). In our study, zinc concentrations in cod muscle were 23–33 ppm on average depending on the region, which fits well within the range in Rainbow (2018). Hellou et al. (1992) measured concentrations of arsenic in Atlantic cod from the northwest Atlantic and found a range of 4-52 µg/g dry weight, so the concentrations recorded in this study of up to 191 ppm are higher in some cases. However, the vast majority of samples were below 50 ppm with only very few exceeding this. It was also noted by Neff (1997) that wide ranges of arsenic concentrations were present within a single taxon, similar to the results for Atlantic cod obtained here. Copper concentrations in cod were reported to be 0.5–2 µg/g dry weight by Rainbow (2018), which is slightly lower than the Vanta estimates in our study. The CRM is in agreement with Rainbow (2018) in stating a copper concentration of 1.05 µg/g, so the estimates from the Vanta analyser may be slightly elevated. However, the estimates from the Vanta of zero ppm for the concentrations of cadmium, lead and mercury are supported by the CRM and Rainbow (2018), since these were measured as below 1 µg/g dry weight in both sources and are therefore below the detection limit of the Vanta. The results in Rainbow (2018) are sourced from a number of studies, including Franklin (1987), Brown and Balls (1997) and Julshamn et al. (2013).

5.5.2 Assignment accuracies achieved using the two XRF instruments

Among sampled sites, cod from the Barents Sea, Faroes, West Scotland and Baltic Sea were the most chemically distinct and an assignment accuracy to known origin of >95% was achieved using either or both the Itrax and Vanta data. Furthermore, cod from all regions except Iceland could be assigned to origin with 75% or greater accuracy with at least one of the XRF instruments. The Icelandic cod were the most challenging to assign and resulted in the lowest success rate, with only 48% and 68% of individuals correctly traced to origin using the Itrax and Vanta respectively. The PCA plots showed that the Icelandic cod were split into two groups, spanning a relatively large geographic range (Figure 5.19). Many of the individuals in the smaller group were caught in the eastern half of Iceland, so a higher assignment accuracy could possibly be achieved by dividing the Iceland samples into two groups – western and eastern Iceland. This variation in elemental composition amongst cod from different parts of Iceland may be due to different ocean currents flowing around Iceland. The cold waters of the East Greenland and East Icelandic currents flow southwards from the Arctic, flowing past the north and east of Iceland. To the west and south of Iceland, the North Atlantic current and Irminger current bring warmer water (Oskarsson et al., 2009). Therefore, the environmental characteristics associated with different water masses may have led to differences in the elemental composition of cod muscle tissue from the west and east of Iceland. This pattern was also revealed in the stable isotope data from cod around Iceland in Chapter 2, further supporting potential for finer-scale geographic differentiation within Icelandic whitefish.

Using both the Itrax scanner and the Vanta analyser, higher assignment accuracies to geographic origin were achieved using the raw elemental data compared to element ratio data, particularly for the Vanta. This suggests that it is not time efficient to calculate the log ratios for this purpose, especially considering the challenge of objectively selecting pairs of elements for the log ratios, and so it may be better to use the raw element data for both instruments. However, the Itrax records intensity in counts per second and the data acquired are semi-quantitative due to the net sum effect (since the Itrax cannot measure all elements), so comparing log ratios is generally regarded as best practice (Weltje and Tjallingii, 2008). Furthermore, raw counts from the Itrax are not directly comparable among elements due to its varying ability to detect different elements and the effects of variables such as porosity, roughness and moisture content of the sample, further necessitating the use of ratios. In the future, machine learning approaches may give the best results since they can draw on a larger range of element pair combinations than is possible or time effective with a human operator, and can therefore find the most powerful combination for element log ratios. The Vanta provides quantitative data in parts per million (ppm), since it conducts internal calibration, so it is not necessary to calculate ratios.

5.5.3 Variations in muscle tissue elemental abundance among regions

Trace elements are taken up by marine fish such as Atlantic cod both from the surrounding water and from their diet (Rainbow, 2018). Therefore, the elemental abundances in the muscle tissue vary due to the environmental concentrations in the surrounding water as well as due to the physiology, trophic habits and potentially genetic stock of the fish. Concentrations of elements within the water column may vary among regions because of local anthropogenic contamination, including from contaminated rivers, direct discharges, terrestrial run off and atmospheric deposition, and due to variations in the seabed geology (Rainbow, 2018). Furthermore, individuals or populations in different regions may take up and accumulate trace elements at different rates, or may have different diets that contain varying concentrations of elements. Some trace elements are essential for living organisms to survive, such as iron, copper and zinc (Watanabe et al., 1997), whereas other elements have no known metabolic function, including mercury, lead and silver, and so these accumulate in the tissues or are excreted (Sunda, 1989). We did not find any clear differences in either the essential or non-essential elements being more important for distinguishing among regions of origin, and therefore it is not clear whether or to what extent physiological factors or environmental factors have the most influence on the ability to distinguish cod from different locations.

Potassium was the most important element for distinguishing among samples for both the Itrax and Vanta raw data, which could partly be because its presence in high concentrations means it is easy to measure and so is less influenced by noise. The main sources of potassium in seawater are continental river runoff and hydrothermal activity at mid-ocean ridges (Li et al., 2019, Elderfield and Schultz, 1996), which create local variations in abundance. Potassium is an essential element for living organisms and has many vital functions including the regulation of intracellular osmotic pressure, transmission of nerve impulses, muscular contraction and maintaining a regular heartbeat as well as being vital for protein synthesis (Harvard Health Publishing, 2019). In addition to obtaining potassium from their diet, fish can also uptake potassium from the water through the gills (Presas-Basalo, 2022). Therefore, potassium levels in cod muscle tissue may be influenced by their dietary preferences as well as ambient availability of potassium in the surrounding water. It is also possible that potassium concentrations in cod muscle are sensitive to physiological and ecological variation, where differences in genetics and performance of fish across regions exaggerate the chemical differences present in seawater.

Iron was also an important element for origin differentiation using both instruments. The dominant input of iron to open ocean waters is atmospheric dust deposition, mainly by aeolian transport from deserts (Jickells et al., 2005), but other sources, such as riverine input and resuspension of sediments, also contribute in coastal regions (Wetz et al., 2006,

Johnson et al., 1999). Differences in the rate of these inputs result in regional variations in iron availability. In the Atlantic Ocean north of 51°N, atmospheric dust deposits were found to be low (mean ≤0.2gm^{-2}a^{-1}) compared to other regions of the north Atlantic (Measures et al., 2008). The results of the current study revealed that the lowest concentrations of iron were found in cod muscle from the Barents and Norwegian Sea, which may be explained by the smaller amount of dust deposition in these northerly regions. The more coastal regions sampled may have had a greater contribution from river or sediment inputs. Iron is an essential nutrient for all living organisms (Morel and Price, 2003), and can be a limiting nutrient for photosynthesis in phytoplankton (De Baar et al., 2005, Boyd et al., 2007). In addition to affecting productivity, iron availability also influences the community structure of phytoplankton (Jickells et al., 2005). Certain species possess adaptations to low iron levels, including a reduction of cell size, minimising the number of iron-containing enzymes (Palenik et al., 2003, Jickells et al., 2005) and luxury iron uptake which allows some phytoplankton to benefit from a sporadic atmospheric dust supply (Kustka et al., 2003), whereas others such as diatoms require a higher iron concentration (Turner and Hunter, 2001, Jickells et al., 2005). In general, open ocean species of phytoplankton have a lower cellular requirement for iron and so are able to grow more rapidly at low iron concentrations than coastal species, due to the significantly lower availability of trace metals in oceanic seawater (Sunda, 1989). Consequently, the species composition of phytoplankton and hence iron content varies spatially, which is transferred up the food chain and incorporated into the muscle tissue of fish. There may also be a link between the iron concentration in cod muscle and the extent of iron utilisation by the phytoplankton bloom at the base of the food chain, since iron limitation can occur post bloom and may lead to a change in the community structure present. This has been observed in the Iceland basin after the spring bloom (Nielsdóttir et al., 2009).

Arsenic was another important element for the differentiation of cod samples by region. Concentrations of arsenic are generally higher in estuaries and coastal waters than in the open ocean (Austin and Millward, 1986), due to both natural and anthropogenic sources. Inputs of arsenic to the ocean include atmospheric deposition, land runoff and sediment resuspension (Byrd, 1990, Cutter et al., 2001), all of which may show regional variations. Therefore, the higher arsenic concentrations in cod from the Barents Sea, and to a certain extent the Norwegian Sea, do not fit with this expectation of lower arsenic levels in the open ocean, indicating that an alternative mechanism may be involved. These higher arsenic concentrations may reflect the differing sediment source for ocean basins adjoined to high latitude continents due to the extensive ice scour, since sediments can be sourced from glacial fan deposits containing a variety of initial source material, often derived from old continental crust and metamorphic zones. Sediments from the Old Upper Continental Crust, such as those found in the Barents Sea, are known to be very distinct in their geochemical composition (McLennan et al., 1990). Furthermore, Demina et al. (2020)

found high arsenic concentrations in Barents Sea sediments (19 ppm) that were double those in the Upper Continental Crust, suggested to be from peat and sedimentary rocks inland containing coal, which has a high arsenic content.

Since arsenic is a heavy metal, it can also display bioaccumulation within marine organisms (Neff, 1997). Inorganic arsenic in seawater is accumulated by marine algae and bacteria, and then converted to organoarsenic compounds. These organic compounds are efficiently bioaccumulated by marine animals and transferred through marine food webs, whereas they only accumulate inorganic arsenic from seawater to a small extent (Neff, 1997). This means that the highest concentrations of arsenic are found in tissues of marine animals that feed primarily on phytoplankton or macroalgae, and concentrations are lower in those that feed at a higher tropic level (Neff, 1997). Polychaetes, bivalves and other herbivorous molluscs therefore often have high concentrations of arsenic, and some accumulate extremely high levels (Gibbs et al., 1983, Phillips and Depledge, 1986). Spatial variation in arsenic concentrations of cod muscle in this study may therefore partly arise from differences in the diet of individuals from different regions, depending on whether the diet mainly consists of marine invertebrates or is primarily piscivorous. Bioaccumulation of elements is also affected by the metabolic activity of the individual (Vieira et al., 2011) and so possibly could also contribute to differences among regions, since the metabolic rate may vary due to either the genetics of the population or environmental factors. Pacific cod were demonstrated to show a negative correlation between arsenic levels and age or length, which was likely a result of dilution with growth, as tissues are synthesised at a faster rate than the element is accumulated (Burger et al., 2007). Therefore, individuals with a lower metabolic rate (such as in colder Arctic regions) would synthesise tissues more slowly and therefore accumulate arsenic to a greater extent. The wide variation among individuals from the same region (some with zero ppm arsenic and others with much higher concentrations) may also be due to this relationship between bioaccumulation and age, since cod of different sizes were sampled in the study, or perhaps due to individual variation in metabolic rate.

For the Itrax data, the additional elements of chlorine and iodine also provided an important contribution to the separation by region, but these elements were not recorded by the Vanta analyser. The least important elements for spatial assignment were lead, strontium and nickel for the Itrax data, whereas for the Vanta data lanthanum, barium and copper were relatively unimportant. These differences between the two XRF instruments are likely due to differing detector thresholds and arrays, as well as the higher degree of operator input required for reprocessing of the Itrax data resulting in some additional elements being included to improve the model's goodness of fit. Rainbow (2018) reports that copper concentrations are relatively stable in Atlantic cod regardless of the region of origin, typically with concentrations in a fairly narrow range of 0.5-2 ppm in dried muscle tissue.

Fish are known to regulate concentrations of essential elements, including copper, in their muscle tissue to relatively constant levels, irrespective of local differences in bioavailability (Rainbow, 2018). This may explain why copper contributed little to the variation among cod in the Vanta data here.

The Faroes samples were stored in 70% ethanol for transport to the UK, whereas all other samples were transported frozen. Possible elemental contamination from the ethanol may have occurred, as we had no control over the preservation method. Preservation effects may have had an influence as the Faroes samples were the most accurately assigned (i.e. the most chemically distinct) of all regions and 100% of individuals were correctly traced to origin using both instruments. However, the main differences in the Faroes samples compared to those from other regions are the lower concentrations of potassium and chlorine, which suggests the distinctness of the Faroes cod is not due to contamination. The oceanic crust in the Faroes region is composed of Mid Ocean Ridge Basalt (MORB) which has been found to have a very distinct elemental composition (McLennan et al., 1990), so the observed differences in Faroes cod muscle tissue may reflect actual elemental properties of the geology of the area. However, it cannot be established whether the distinctiveness of the Faroes cod is due to environmental characteristics or the sample storage method without further research. Therefore, investigation into the true accuracy with which cod from the Faroes can be assigned to origin is necessary using samples transported frozen rather than in ethanol. Future studies should ensure common sample treatment protocols, ideally with minimal contact between fish samples and external chemicals (including water).

5.5.4 Comparison between Itrax scanner and Vanta analyser

The highest rate of correct assignments to origin was obtained from the raw Vanta data (mean of 86% correct overall). The Itrax also gave a good assignment accuracy of 83% using the raw element data, but was not quite as high as using the Vanta data. Both instruments resulted in correct assignments of over 90% for some regions, such as the Barents Sea and Faroes, and over 75% for all other regions except for the Norwegian Sea and Iceland. Therefore, both the Itrax and Vanta show good potential for use in spatial traceability, although the Vanta gave a slightly higher success rate overall. Differences between the results obtained using the Itrax scanner and Vanta analyser may be caused by uncertainty due to the sensitivity to rough sample surfaces, particularly for the Itrax.

Data produced by the Itrax scanner required extensive reprocessing before data analysis could be carried out, which was a substantial time investment (approximately 10 days for the full dataset in this study) and needed a very experienced Itrax operator to reduce the error. The data cleaning protocol for the Itrax was also fairly time consuming, since each sample had to be inspected individually to ensure the correct portion of data was selected

and to check that the surface of the sample was sufficiently flat. However, when comparing assignment accuracies achieved using this filtered data and the original data there was no significant difference. This suggests that the time taken to conduct the data cleaning is not beneficial to the overall result and it may be more efficient (and less subjective) to use the unfiltered data in future. However, the time investment for reprocessing of the Itrax data is unavoidable and so the Itrax method would still have the disadvantage of being much more involved and complex than that of the Vanta.

The Itrax scanner and Vanta analyser each have their own benefits. However, the Vanta is potentially more suitable for testing traded fish products, because it analyses samples more quickly compared to the Itrax (around three minutes rather than over an hour per sample). Additionally, the Vanta has the added benefit of internal calibration which provides elemental concentration in ppm, allowing raw data from the instrument to be used and removing the need to calculate log ratios as with the Itrax. This further simplifies and increases the speed of data collection and analysis. Critically, the Vanta instrument can also be used in hand-held mode which eliminates the time taken to prepare samples on a grid for analysis as with the Itrax, and is portable so can be easily transported to any location throughout the supply chain. This technique could also be applied to fresh or frozen fish fillets, avoiding the uncertainty associated with surface roughness of powdered samples and further reducing the sample preparation time. However, it would be necessary to investigate the influence of the water content of fresh or frozen samples before use on traded fish products, since this could introduce additional variation (Tjallingii et al., 2007, MacLachlan et al., 2015).

Therefore, considering the ease of use of the equipment and lack of data processing required, as well as the good assignment accuracy achieved, the Vanta analyser may be more beneficial than the Itrax for use in verifying the catch regions of fish products through the supply chain. Although this relies on the ability of the Vanta to accurately estimate the abundance of elements in fresh or frozen fish rather than freeze-dried samples.

5.5.5 Comparison with other assignment techniques

Several other natural tracer methods that have been successfully applied to trace the catch region of fish and seafood. Stable isotope analysis is widely used to verify the origin of terrestrial food products (West et al., 2007, Valenti et al., 2017, Rummel et al., 2010, Perini et al., 2016, Camin et al., 2010, Bontempo et al., 2017), and has more recently also been investigated for marine food products (Ortea and Gallardo, 2015, Carrera and Gallardo, 2017, Zhang et al., 2017, Gong et al., 2018, Varrà et al., 2019, Zhang et al., 2019, Kang et al., 2020, de Aquino Ferreira et al., 2021, Slesser and Trueman, 2021, Won et al., 2021). Stable isotope approaches have been demonstrated to reliably assign the same Atlantic cod samples to broad fishery areas with an accuracy of over 80%, and greater than 90% in

certain cases, although isotopically similar regions are more challenging to assign (Chapter 2). Assignment accuracies for the more isotopically distinct regions are similar to those achieved using XRF techniques, although the overall assignment success was not quite as high using stable isotopes (mean of 79%). One major benefit of XRF equipment is the availability of portable devices, which eliminate the time and effort involved in transporting samples to a laboratory for analysis. The disadvantage of stable isotope analysis is that it must be undertaken in a laboratory with a mass spectrometer, which also involves more sample preparation and a longer analysis time. In preparation for isotope analysis, powdered samples must be weighed into tin capsules, which takes a few minutes per sample and therefore allows around 180 samples to be prepared each day, whereas the sample preparation for XRF analysis simply involves placing the powdered sample onto a board. Elemental analysis with XRF equipment not only reduces the amount of sample preparation required, but samples can also be analysed in only 60 seconds (although repeats may be necessary). The mass spectrometer can analyse approximately 90 samples a day, meaning that stable isotope analysis has a lower sample throughput, and it also requires calibration by an experienced user following data acquisition. However, the mechanisms driving spatial variation in stable isotopes are well understood and are therefore predictable, whereas less is known regarding the factors causing regional variations in tissue element abundances making it difficult to predict the species or regions that this technique may be useful for.

Genetic techniques have also been demonstrated to be effective in determining the geographic origin of fish, including Atlantic cod. Cod could be assigned to the North Sea, Baltic Sea and Northeast Arctic populations with a success rate of 98-100% using single nucleotide polymorphisms (SNPs) (Nielsen et al., 2012b). The Itrax and Vanta achieved a similar assignment success for the Barents and Baltic Seas, but the North Sea had a slightly lower proportion of correctly assigned individuals (91%). However, although genetic markers perform well for these specific populations on a relatively large scale, SNP techniques may lack the ability to discriminate among regions that have not remained reproductively isolated, or for catch regions that sample populations with geographically separate spawning grounds, but mixed foraging regions. The accuracy of assignment is therefore reduced for mixed populations, at common feeding grounds used by multiple different populations or at the borders of discrete populations. Trace element analysis using XRF may have the potential to trace fish to their catch region on a smaller spatial scale than genetic methods in some cases, since the abundance of certain elements may differ even among areas in close proximity. Moreover, the analysis of samples using genetic techniques usually must be undertaken in a laboratory and requires significant sample preparation, resulting in a much more time-consuming process than using XRF instruments, particularly the Vanta or other portable XRF equipment. However, the recent development of genetic technologies such as nanopore sequencing, offers more rapid

processing of samples, real-time data acquisition and portable devices (Oxford Nanopore Technologies, 2022). This means that genetic techniques such as these could match the benefits of the Vanta analyser if proven to be effective for spatial traceability of fish products in the future, and so could be beneficial as a complementary technique.

Therefore, XRF techniques for tracing the catch origin of cod not only provide assignment accuracies similar to stable isotope and genetic techniques, but also confer several additional advantages including on-site analysis with no specialist knowledge required, time saved on sample preparation and the ability to obtain results very rapidly. This would allow a much higher throughput of fish products for provenance verification in trade or commercial settings, and so XRF technology still has the potential to play an important part in ensuring the authenticity of traded fish products.

5.5.6 Combining trace element and stable isotope data for assignment

A number of published studies have found that combining approaches improves the ability to distinguish among regions of origin. For example, farmed Mediterranean mussels could be assigned to origin with 97% accuracy using stable isotopes and trace elements together, whereas using the techniques separately resulted in lower success rates of 81% and 88% respectively (del Rio-Lavín et al., 2022b). Similarly, Ortea and Gallardo (2015) found that using stable isotopes and multi-elemental analyses in combination gave a higher assignment accuracy than either method used alone for shrimps from a range of locations. Using stable isotope analysis alone achieved correct classification of 70%, whereas the addition of multi-element analysis increased the correct classification to 100%. Other studies, such as Gopi et al. (2019b), Gopi et al. (2019c) and Martino et al. (2022a), investigated the combination of XRF techniques with stable isotope analysis for tracing the origin of seafood, as in the current study. Gopi et al. (2019b) found that the overall assignment accuracy using both techniques was the same as using the Itrax XRF data alone. However, the results of Gopi et al. (2019c) and Martino et al. (2022a) supported the findings of our study, demonstrating that the isotopic and multi-elemental compositions of marine food products used together are an effective tool for determining origin, and that in many cases the use of two approaches combined increases the distinguishing power.

5.5.7 Limitations and future research

In this study, XRF techniques were only used as an initial investigation into their potential for discriminating among regions of origin for cod muscle tissue, and there was not sufficient time to develop an optimised methodology. Therefore, there are certain limitations of the XRF data acquired that require further investigation. Firstly, the surface roughness or grain size as well as compaction of the powdered samples may affect the

elemental abundance estimates due to increased scatter, as observed by Croudace et al. (2006) and MacLachlan et al. (2015). The roughness or grain size of the samples may cause the x-ray beam from the XRF instrument to hit the sample surface at an angle and could result in the radiation produced not reaching the detector. When samples were ground coarsely it resulted in a relatively low precision for certain elements, suggesting there may be an effect on the data recorded if the sample surface is very rough. To account for any potential effects of this, we selected a sample grinder that created a relatively fine powder and ensured the grain sizes were as homogenous and consistent as possible by grinding all samples with the same grinder. Any data produced from very rough parts of samples' surfaces were excluded from the data analysis in case there was an effect due to surface roughness. Each sample was also compressed with a spatula prior to analysis to try and reduce the effect of variations in the powder compaction. Only a very limited investigation into the effect of sample coarseness was conducted, since very few samples were analysed for the comparison (four samples measured with both grain sizes). A more thorough investigation into the effects of surface roughness and compaction of samples analysed using both the Itrax and Vanta are necessary before this technique could be more widely used. An improvement to the method would be to use a mechanical grinder, since this would produce a more uniform and standardised powder. Furthermore, all data points provided by the Vanta were included in the data analysis even though the error was sometimes high, which may have been a result of uneven sample surfaces. The data analysis protocol should be improved in future to determine a maximum threshold for acceptable error and filter out data points from the final dataset that exceed this value, so that only reliable data are included in the data analysis.

It was also observed that several elements not typically present in fish muscle were recorded by the Itrax and Vanta in measurable concentrations. For example, the Itrax recorded mean sample counts of up to 9500 cps for argon and up to 16,000 cps for zirconium. Similarly, the Vanta gave mean sample abundances of up to 2480 ppm for titanium, although overall the Vanta resulted in fewer elements with spuriously high concentrations than the Itrax. These elements would all be expected to be below detectable concentrations, and any counts assigned to these are likely to reflect artefacts associated with inter-element interactions and spectral interferences. For this study, the main aim was to differentiate among samples based on spatial origin for which the matrix was similar across all samples. For this purpose, the relative abundances of elements were sufficient rather than the absolute element concentrations, so counts for elements known to occur at sub-ppm concentrations were disregarded. If the elemental interactions that occur in the analysis of fish muscle matrix could be better understood, this would allow matrix-specific arithmetic corrections to be applied.

Another limitation in the current study, were the small sample numbers for certain catch regions. For the Irish Sea, Rockall and West Scotland, only between three and eight samples were analysed. To reliably assign fish to their geographic origin, the reference dataset must fully encompass the true variation of individuals from that region. For this reason, the ability to correctly trace cod to origin could not be accurately assessed for regions with few samples. Further research using a greater number of samples is necessary to determine whether XRF techniques work effectively for all fishery areas. It would also be beneficial to sample over a wider range of locations within each region to better represent the population present, since in some cases the sampling stations were clustered together in relatively close proximity and did not cover the full extent of the fishery region.

Investigation into the best combination of elements to use in the log ratio calculations for the Itrax data would also be advantageous to maximise the success rate of assignment, since different log ratio pairs have varying distinguishing power. This study selected the element pairs based on graphical observation of the regional clustering, and therefore only limited combinations could be trialled. However, machine learning could test every possible combination, of which there are billions. This would reveal the most effective element pairs for distinguishing among fishery regions for Atlantic cod, but separate investigations would be necessary for other species where different element combinations may give the best separation.

There are a wide range of factors that can affect the elemental composition of fish muscle, including both factors that affect the ambient concentrations of elements in the surrounding water, such as local anthropogenic sources, release from seabed sediments by resuspension and atmospheric deposition, and factors affecting the uptake and accumulation by fish, as well as variations in diet (Rainbow, 2018). However, the mechanisms underlying these variations are not well understood. Therefore, the elemental abundance present in fish muscle from different regions cannot be predicted a priori. There may also be temporal variations in tissue element concentrations, both seasonal and inter-annual, so research into the extent to which this affects the accuracy of determining spatial origin is essential before the technique could be applied to traded fish products. It is therefore necessary to create a new reference dataset each year to account for temporal variations in the fish muscle composition in order to maintain the same level of assignment accuracy over time.

One of the major advantages of XRF analysis for spatial traceability over other forensic techniques is that XRF equipment can be hand-held and portable, so could be taken to any location along the supply chain such as ports where fish catches are landed, in fish markets or at retailers, for rapid verification of fish product provenance on-site without involving the extra time necessary for laboratory analysis. The Itrax scanner is too large to be

transported to other locations, but was used here as a proof of concept, whereas the Vanta analyser is portable and can be used in hand-held mode so would be ideal for on-site use. During this study the Vanta was only used in fixed mode. Future use of the Vanta in hand-held mode would be an improvement for spatial traceability purposes, however this would not have any effect on data quality. Another improvement for XRF techniques to be used in real world scenarios would be the ability to test fresh fish products. The research here focussed only on freeze-dried fish as a first exploratory investigation, because the analysis of wet fish samples would be likely to introduce additional variation and could reduce detection limits due to the water content (Croudace et al., 2006). The requirement to freeze-dry fish samples prior to analysis is time consuming and would counteract the benefits of the rapid analysis by XRF. Therefore, future research into the analysis of fresh or frozen fish would allow XRF techniques to have maximum benefit in the spatial traceability of traded fish products.

5.5.8 Conclusions

The current study has demonstrated that trace elemental composition of Atlantic cod muscle tissue as recorded by XRF instruments has good potential for use in spatial traceability of cod products throughout the supply chain. An assignment accuracy of 95% would be required for this to be used as a forensic tool for traded fish products in a commercial or legal setting. Both the Itrax scanner and Vanta analyser showed mean assignment accuracies of 83-86% to region of origin, so did not quite meet this requirement. However, assignment accuracies exceeding this level were achieved for certain regions, such as the Barents and Baltic Seas. Combining XRF techniques with stable isotope tracers increased the mean assignment accuracy to 91%, achieving a higher level of accuracy than either method used independently, so the addition of another alternative approach may increase the assignment accuracy further to reach the required 95% threshold for all regions.

Overall, the Vanta analyser may show better potential for use in commercial settings for verifying the catch region of traded fish products because of its portability and rapid results. However, more investigation is necessary into the possibility of validating the origin of fresh or frozen fish products directly to reveal the true opportunity this technique presents.

Chapter 6 Final discussion and summary

The oceans provide a vital food source to over 3 billion people globally and the demand for seafood is continually rising, yet more than a third of fish stocks are exploited at unsustainably high levels (FAO, 2022c). Food fraud is a serious global issue (Visciano and Schirone, 2021), and spatial mislabelling of marine food products threatens the sustainability or recovery of fish stocks as well as deceiving customers (FAO, 2018). With fisheries and aquaculture production now at a record high and continuing to increase (FAO, 2022c), traceability is becoming ever more important for consumers, producers and regulators alike. Highly complex and globalised supply chains mean that mislabelling goes undetected (Leal et al., 2015), since currently there are no widely-accepted forensic tests for the geographic origin of traded fish and seafood products. The aim of the research in this thesis was to address this lack of established tools for verifying the claimed catch location of fish products by exploring potential approaches for spatial traceability.

6.1 Summary of findings

Chapters 2 and 3 investigate the potential for stable isotope ratios in fish muscle tissue to provide geographic discrimination on a scale appropriate for fisheries management, focusing on three commercially important whitefish species – Atlantic cod (*Gadus morhua*), haddock (*Melanogrammus aeglefinus*) and European hake (*Merluccius merluccius*).

In Chapter 2, an extensive dataset is presented consisting of carbon, nitrogen and sulfur stable isotope ratios measured in Atlantic cod muscle tissue from the main fishery areas within the Northeast Atlantic. Significant variations in isotopic composition with area of capture were observed, which allowed individual cod to be assigned to most likely origin with over 90% accuracy for certain regions. However, other regions were isotopically similar and were therefore more challenging to assign correctly. The performance of three different assignment algorithms were compared: linear discriminant analysis, a multivariate normal probability technique and random forest classification. The number of samples analysed enabled a comprehensive investigation into the accuracy with which cod could be traced to their region of origin, employing a discrete assignment method where samples were assigned to pre-defined areas roughly equivalent to ICES subareas. Spatial differentiation in cod using the isotopic composition of muscle tissue was compared with that found using genetic markers, as previously assessed in Nielsen et al. (2012b), revealing that stable isotope tracers can provide discrimination with a similar level of accuracy to genetic techniques for the same regions (96-100% correct assignments). To demonstrate the true effectiveness of the methodology in a traceability context, the known-origin samples collected in the study were used as a reference dataset to assign

independent samples to their most likely origin. This achieved a lower assignment accuracy than self-assignment of the reference dataset, as expected for samples that are truly independent.

Chapter 3 investigates the spatial traceability of haddock and European hake using stable isotope analysis. Here a large dataset of carbon, nitrogen and sulfur stable isotope ratios measured in the muscle tissue of haddock and hake from a wide range of regions is presented, including many of the most important fishery areas. The isotopic composition of muscle tissue was found to vary significantly among regions for both species. The two most effective classification methods from Chapter 2 (multivariate normal technique and random forest) were applied to assign samples to most likely origin, achieving an average accuracy of 66% and 72% for haddock and hake respectively. Fish from certain regions were highly distinctive, such as Mediterranean hake, and could be traced to origin with over 90% accuracy, whereas those from other regions were more challenging to assign due to overlaps in isotopic compositions. In a similar way to Chapter 2, independent samples were assigned to origin using the dataset from the current study as a reference in order to estimate the true effectiveness of the method, which again resulted in reduced assignment accuracies compared to using non-independent test samples.

The findings of Chapters 2 and 3 demonstrate that stable isotope techniques show good potential for verifying the claimed origin of whitefish products, since fish from certain discrete regions could be assigned to origin with a high level of confidence. However, those from other regions were more isotopically similar and so could not be reliably traced to catch location. This indicates that stable isotope tracers are unlikely to provide stand-alone tests for proof of origin with accuracy suitable for legislative action, but could be used in conjunction with other techniques to result in a higher level of accuracy than could be achieved using either method independently.

Chapter 4 applies fatty acid and stable isotope analyses for spatial traceability of the critically endangered European eel, *Anguilla anguilla*. The findings indicate that these techniques cannot be used to reliably assign glass eels to river of origin, since the stable isotope and fatty acid compositions of glass eels were very similar among those from all sampled locations. In many cases, the assignment accuracy was little better than if glass eels had been assigned at random. It is hypothesised that this similarity in composition is due to the glass eels sharing a common migration route across the ocean until close to the time of entering estuaries. However, glass eels from the River Oria, Spain, were more distinct and could be assigned to origin with 89% accuracy using stable isotope and fatty acid analyses combined. Therefore, stable isotope and fatty acid tracers may be capable of distinguishing glass eels caught in certain more distinctive rivers from all others, but for eels from the majority of locations it appears that alternative markers are required, perhaps those that are incorporated into the muscle tissue more rapidly.

In Chapter 5, an alternative approach to spatial traceability is investigated. The use of X-ray fluorescence (XRF) to determine the catch location of Atlantic cod is explored, with opportunistic use of the samples collected for stable isotope analysis in Chapter 2. This was achieved by recording the elemental composition of the muscle tissue using two different XRF instruments - first with the Itrax core scanner, which is a fixed laboratory-based instrument, and then with the portable Olympus Vanta analyser. This study presents a novel use for the Vanta analyser, which to my knowledge has not been applied to fish or seafood previously, whereas the Itrax scanner has only rarely been used for this purpose (Gadd et al., 2018, Gopi et al., 2019c, Gopi et al., 2019b, Martino et al., 2022a). A major part of this chapter was therefore focused on method development, which when applied to the Atlantic cod samples, demonstrated significant regional differentiation based on the elemental concentrations using both the Itrax and Vanta. Assignment accuracies to known origin of over 90% were achieved for certain regions using random forest classification, and cod from all regions with the exception of Iceland could be assigned with at least 75% accuracy with at least one of the XRF instruments. The ability to distinguish among regions using raw element data and log ratio elemental data were compared, revealing that higher assignment accuracies could be obtained with raw data. This research demonstrates that the trace elemental composition of Atlantic cod muscle tissue as recorded by XRF instruments shows good potential for verifying the catch region of cod products throughout the supply chain and encourages further research into this method. Overall, the Vanta analyser may be better suited for use in commercial settings because of its portability and rapid results.

6.2 Comparison of spatial traceability approaches

In this thesis, three different approaches were explored for tracing fish to their geographic origin, comprising stable isotope analysis, trace element analysis using XRF and fatty acid analysis. Stable isotope and trace element analyses were both conducted using common samples of the same species, Atlantic cod, allowing their effectiveness to be directly compared. The assignment accuracy to known origin was comparable using stable isotopes and the Itrax XRF scanner (79% and 82% mean respectively), whereas a slightly higher accuracy was achieved using the Vanta XRF analyser (86% mean). The ability to correctly assign cod to origin varied based on the individual regions, with those from some regions more reliably assigned using stable isotopes and others using XRF. Therefore, both of these approaches were successfully able to distinguish among cod from different regions to a certain extent, depending on the specific region, although the Vanta analyser appeared to be slightly more effective overall. XRF instruments are able to record data on numerous trace elements, whereas stable isotopes in the marine environment usually rely on just three dimensions, providing a benefit of elemental profiling over stable isotope

analysis that may have led to the increased assignment accuracy using the Vanta. Fatty acid analysis also produces data on a large number of dimensions, but was not found to be very successful in discriminating among glass eels from different locations. However, fatty acid analysis may be more effective for other species that do not undergo extensive migrations, as demonstrated by Fonseca et al. (2022) and Go et al. (2022) for bivalves and European seabass.

During my investigation into these different approaches for spatial traceability, several benefits and drawbacks of the techniques became apparent. Stable isotope ratios are somewhat predictable spatially and temporally since the mechanisms underpinning these variations are well understood (Somes et al., 2010, Magozzi et al., 2017), whereas there are many complex factors affecting the elemental composition of living organisms that make it very challenging to predict spatial and temporal variations in trace element concentrations (Rainbow, 2018). The effects of variables such as age, sex and size of fish as well as the sampling procedures are also unpredictable and may influence the elemental composition of the muscle tissue (Li et al., 2016, Rainbow, 2018), masking any variations due to geographic origin. This drawback also applies to fatty acid analysis, since there are many complex variables affecting the phytoplankton community composition and therefore fatty acid content of consumers (Reuss and Poulsen, 2002). This inability to predict spatial variations in tracers means that the potential for distinguishing among fish from specific regions cannot be determined a priori, as it can be with stable isotope compositions for which it is a great benefit in selecting the most suitable tool for the purpose.

The Vanta XRF analyser also offers major advantages in terms of ease and speed of obtaining results, as well as the possibility of using the equipment on-site at any location throughout the supply chain. The Vanta is a portable, handheld device that could easily be transported to landing ports, processing or distribution centres, warehouses, fish markets and retailers for analysing test samples in situ, which would provide a significant time and cost saving compared to laboratory-based methods. Stable isotope analysis, Itrax XRF analysis and fatty acid analysis all require large and heavy equipment that must be used in a fixed location by experienced operators, and so are restricted to use in laboratories. The Vanta is simpler to use and does not require specialist knowledge, so can be operated by inexperienced people following brief training. A further important benefit of the Vanta analyser is the considerable time saving on sample processing and analysis. Fatty acid analysis requires lengthy processing of samples in the laboratory with multiple stages, each taking minutes or hours using different pieces of equipment. There is also a limit on the number of samples that can be processed simultaneously because of available space in the equipment. Therefore, I do not think fatty acid analysis is a viable technique for use on a commercial scale, as it very time consuming and has a low throughput of samples, as well as requiring the purchase of several different instruments for processing and analysis.

Stable isotope analysis also involves fairly time-consuming sample processing, although not to the same extent as fatty acid analysis, while the Itrax scanner requires samples to be arranged onto a grid for presentation to the instrument and then involves extensive reprocessing of the data. All these methods also take several minutes to hours for analysis per sample. Conversely, the Vanta analyser does not necessitate any sample preparation, it can analyse a sample in seconds and data is provided in parts per million (ppm) ready to use. Therefore, the Vanta is by far the most rapid method for acquiring data and the most flexible option for use in a variety of settings. Considering the relatively high assignment accuracies obtained using the Vanta XRF analyser, I believe there is great potential for this technique to have more widespread use in fish and seafood supply networks, despite the lack of predictability in trace element signatures. There is also the possibility of using XRF techniques on fresh or frozen fish products without the need to freeze-dry and grind samples, which would enable even more rapid analysis and a much higher throughput of test samples, although the viability of this has not yet been researched. However, stable isotope analysis also has potential to make an important contribution to the spatial traceability of fish and seafood products, due to our understanding of stable isotope ecology and the demonstrated ability to distinguish individuals with high confidence from certain isotopically distinct regions, as shown in Chapters 2 and 3.

The research presented in this thesis combined with findings of several other published studies such as Ortea and Gallardo (2015), Gong et al. (2018), Cusa et al. (2021), del Rio-Lavín et al. (2022b) and Martino et al. (2022a) confirms that a combination of techniques applied simultaneously is more reliable for assigning origin than any individual method used alone. Although no single technique has been capable of achieving a level of accuracy suitable for verifying the claimed origin of traded fish products from all fisheries, combining approaches may have the potential to meet this requirement and facilitate widespread use. Genetic techniques may also be valuable in contributing to spatial traceability where the approaches studied in this thesis do not provide sufficient discrimination, since genetic tracers reveal the spatial histories of populations over multi-generational timescales rather than an individual's location at the point of capture as the other methods do, and have proven to be effective in published studies using SNPs (Nielsen et al., 2012b, del Rio-Lavín et al., 2022a).

6.3 Future applications

6.3.1 Supply chain traceability of marine food products

The development of a reliable and effective test for the geographic origin of traded fish products would be hugely beneficial to the seafood industry, both for improving customer confidence in products traded as sustainable or from a desirable origin and for upholding

the reputation of honest producers and retailers. Therefore, a test for provenance would allow profits to be maximised as well as benefitting customers through improving satisfaction with products purchased.

A rapid and relatively low-cost method would allow routine testing for the catch location of fish and seafood products at all points throughout the supply chain. This would require a suitable method to be scaled up for widespread use. The Vanta XRF analyser has opportunities available for automation, which combined with an efficient method for presenting samples to the instrument and its rapid analysis time, could allow a high throughput of test samples. Alternatively the Vanta could be used in handheld mode, allowing selective testing of products by carrying the device to the place needed, rather than transferring samples to a fixed setup for automatic analysis. Therefore, elemental profiling using the Vanta seems to offer a convenient and practical solution to the challenge of seafood origin traceability if the methodology can achieve a suitable level of accuracy.

For the authenticity testing of terrestrial food products in commercial or legal settings, confidence limits of 95% are imposed (Camin et al., 2017). Any forensic tool developed to verify the claimed catch location of fish and seafood products must meet these accuracy and precision standards, which may only be feasible by combining two or more techniques. In many cases 95% accuracy may still not be attainable, but it would be most beneficial to focus research on those situations where the required level for regulatory or legal use is possible and in other regions where biochemical approaches can help to focus management attention.

To achieve the ambition of widespread use in traceability, a reference database must be created each year containing data for the selected tracer measured in muscle tissue of the target species from all possible fisheries, as is done for certain terrestrial products (Camin et al., 2017). A large enough sample size from each fishery must be collected to encompass the true variance of the individuals in that region. The data acquired as part of this thesis is valuable as a proof of concept, however some regions were undersampled and others not sampled at all, so future work is necessary to expand the reference dataset to cover all regions within the distributional range of the species. Another future task would be to research the effectiveness of different tracers for determining the origin of other commercially important species that were not investigated here. However, my research highlighted to me the challenge of acquiring a large collection of fish samples, since it took me significant effort to collect those analysed for this thesis. To build a reference collection for a much wider variety of species and regions would be a large undertaking and would need significant investment and collaboration, but based on my research so far, I think that biochemical tracers have valuable information to offer and the benefits to consumers, producers and retailers would outweigh the cost and time invested.

If a methodology can be operationalised and can achieve the required assignment accuracy it could be adopted by a wide range of businesses, from producers and processors to the point of sale. Producers of premium brand products, which are often in demand from fine dining restaurants and other high-end retailers, could ensure their reputation is upheld and not tarnished by lower quality substitutions. It would also prevent others from fraudulently making profits from their brand name while the honest producer suffers economic loss, and even has the potential to be used as evidence in cases of legal disputes in these situations. The premium salmon producer, Loch Duart, has recently introduced an initiative whereby the origin of their products is tested using elemental fingerprinting to prevent fraudulent trading of products under their brand name (Loch Duart, 2022). High-end restaurants or those selling popular local produce, such as Cornish hake, may also benefit from a forensic test for the catch location of their products since it improves the reputation of the business, prevents customers from being deceived in their purchases of often more expensive items, and it protects honest fishers and retailers from competing with fraudulent traders. Finally, supermarkets may be keen to incorporate routine testing into their supply chain to increase sales by giving customers confidence in the products they buy and improving the reputation of the supermarket. This is especially true for those that pride themselves on selling sustainable fish and seafood, which often sell for a higher price, or premium products such as Icelandic cod.

6.3.2 Sustainability certification

The certification of seafood products as sustainable, for example by the Marine Stewardship Council (MSC), encourages producers to engage in sustainable fishing and gives consumers the opportunity to make a responsible choice on the products they buy. However, certified products often fetch a higher price at point of sale and therefore create a greater incentive for fraud. Certification bodies rely on following a chain of custody to ensure products truly originate from a sustainable source, which although greatly improves transparency, still leaves potential for uncertified fish to be introduced into the supply chain through deliberate or inadvertent mislabelling. The MSC has recently been involved with research into the use of DNA and stable isotope techniques to confirm the origin of MSC certified fish and seafood (Marine Stewardship Council, 2020a). This led to a publication by Cusa et al. (2021), which demonstrated that for a number of species either stable isotope or genetic tools could distinguish between certified and non-certified products, although this was based on predictions using a mechanistic approach. Real life applications are severely lacking (Cusa et al., 2021), so although the research in this thesis is a good contribution, considerably more research is needed into the effectiveness of different traceability approaches for specific species and regions before it can be applied to the verification of sustainably sourced products.

6.3.3 Fisheries management

Another valuable application of spatial traceability tools is the sustainable management of fisheries. Global fisheries are managed spatially in discrete geographic areas, each with different quotas and regulations (FAO, 2022c), which means that for effective management it is crucial to know the true catch region of fish and seafood. However, this spatial management also creates an incentive for fraud or illegal fishing, since fish that are not permitted to be landed from a certain region may be legal in another and hence are deliberately mislabelled (Sumaila et al., 2006). An accurate and reliable test for the geographic origin of fish and seafood products would enable management authorities and regulators to identify fraudulent claims of catch location at the point of landing. This would allow the enforcement of regulations and, if accuracy requirements are met, could be used as evidence in court (Camin et al., 2017). Furthermore, routine testing at ports and harbours would deter potential offenders and may reduce the prevalence of illegal fishing. Even the awareness that a scientific method is available for verifying claims of origin can act as an effective deterrent (Sumaila et al., 2006). In addition to protecting honest fishers from economic losses, this would benefit fish stocks since illegal fishing is known to harvest significant amounts of fish each year (Agnew et al., 2009, Sumaila et al., 2020) and hinders the sustainable management of fisheries by leading to inaccurate catch recordings (Sumaila et al., 2006). More effective implementation of regulations would allow at risk stocks to recover and ensure that healthy stocks continue to be fished at sustainable levels.

A further benefit of a test for origin traceability could be in verifying the catch region of EU versus non-EU fish and seafood, or to determine whether the catch landed by a fishing vessel originated within the territorial waters of a certain country for purposes of quota sharing or complying with other legislation.

The Marine Management Organisation (MMO) has recently recognised the potential importance of biochemical techniques in validating the origin of fish for management and enforcement purposes, leading them to commission studies assessing the use of stable isotope (MMO, 2019c), trace element (MMO, 2019d), fatty acid (MMO, 2019b) and genetic techniques (MMO, 2019a, MMO, 2019e). With further research and refinement of the methods, combinations of these approaches could be adopted by the MMO and other management authorities as valuable tools for widespread and routine testing of fish and seafood provenance.

6.3.4 Conclusion

Currently there is no defined confidence threshold for provenance authentication of marine food products in European legislation. Anecdotally a 95% threshold for analytical precision is in place for terrestrial use (Camin et al., 2017), although it is difficult to find a defined

standard in written documents or legislation. The confidence limit must always be higher in legal or forensic cases than for academic inference, so the analytical precision required is expected to be at least 95% in a legislative context for marine food products (ISO 5725-4, 2020). The research in this thesis highlights the need for an agreed threshold to be defined for use in commercial and legal cases, so that forensic tests for origin could become established into widespread use for traded fish and seafood products. For real world applications, aiming to develop tests with 95% accuracy could be recommended to be in line with those used for terrestrial food products, recognising that the level of accuracy required to be of evidential value may vary from case to case. The research presented here demonstrates that this level of accuracy is achievable for certain regions and species, and combining techniques is likely to make this the case for many other fisheries with further research.

Maintaining sustainable fish stocks is increasingly challenging, with continued growth in demand for seafood and a rising proportion of stocks fished at unsustainable levels, yet food and economic security for many people as well as the health of marine ecosystems depend on this. It would be hugely beneficial to the sustainable management of fisheries to have a reliable method of detecting mislabelling and illegal fishing, thereby enabling more effective implementation of regulations and conservation measures. Forensic biochemical techniques show excellent potential to provide such a tool, and so may be crucial for the retrospective authentication of provenance for fish and seafood products in global supply chains. Further research and optimisation would pave the way for these methods to have widespread use in fisheries management and the seafood industry alike.

List of References

ABBOTT, A. & PEARSON, H. 2004. Fear of human pandemic grows as bird flu sweeps through Asia. *Nature,* 427**,** 472.

AGNEW, D. J., PEARCE, J., PRAMOD, G., PEATMAN, T., WATSON, R., BEDDINGTON, J. R. & PITCHER, T. J. 2009. Estimating the worldwide extent of illegal fishing. *Plos one,* 4**,** e4570.

ARU, V., MOTAWIE, M. S., KHAKIMOV, B., SØRENSEN, K. M., MØLLER, B. L. & ENGELSEN, S. B. 2020. First-principles identification of C-methyl-scyllo-inositol (mytilitol) – A new species-specific metabolite indicator of geographic origin for marine bivalve molluscs (*Mytilus* and *Ruditapes* spp.). *Food Chemistry,* 328**,** 126959.

ASCHE, F. 2014. Exchange rates and the seafood trade, FAO/GLOBEFISH. Available: https://www.fao.org/in-action/globefish/publications/details-publication/en/c/347668/.

AUSTIN, L. & MILLWARD, G. 1986. Atmosphere-coastal ocean fluxes of particulate arsenic and antimony. *Continental Shelf Research,* 6**,** 459-474.

BADER-EL-DEN, M., TEITEI, E. & PERRY, T. 2018. Biased random forest for dealing with the class imbalance problem. *IEEE transactions on neural networks and learning systems,* 30**,** 2163-2172.

BARDONNET, A. & RIERA, P. 2005. Feeding of glass eels (*Anguilla anguilla*) in the course of their estuarine migration: new insights from stable isotope analysis. *Estuarine, Coastal and Shelf Science,* 63**,** 201-209.

BARNES, C., JENNINGS, S. & BARRY, J. T. 2009. Environmental correlates of large-scale spatial variation in the $\delta^{13}C$ of marine animals. *Estuarine, Coastal and Shelf Science,* 81**,** 368-374.

BEHRMAN, E. & GOPALAN, V. 2005. Cholesterol and plants. *Journal of Chemical Education,* 82**,** 1791.

BELPAIRE, C., HODSON, P., PIERRON, F. & FREESE, M. 2019. Impact of chemical pollution on Atlantic eels: Facts, research needs, and implications for management. *Current Opinion in Environmental Science & Health,* 11**,** 26-36.

BENNION, M., MORRISON, L., BROPHY, D., CARLSSON, J., ABRAHANTES, J. C. & GRAHAM, C. T. 2019. Trace element fingerprinting of blue mussel (*Mytilus edulis*) shells and soft tissues successfully reveals harvesting locations. *Science of The Total Environment,* 685**,** 50-58.

BERG, P. R., JORDE, P. E., GLOVER, K. A., DAHLE, G., TAGGART, J. B., KORSBREKKE, K., DINGSØR, G. E., SKJÆRAASEN, J. E., WRIGHT, P. J. & CADRIN, S. X. 2021. Genetic structuring in Atlantic haddock contrasts with current management regimes. *ICES Journal of Marine Science,* 78**,** 1-13.

BERRY, D. 2020. Cod Moratorium of 1992. *The Canadian Encyclopedia,.*

BLAHA, F. & KATAFONO, K. 2020. Blockchain application in seafood value chains. *FAO Fisheries and Aquaculture Circular***,** 1-43.

BONG, Y. S., SHIN, W. J., LEE, A., KIM, Y. S., KIM, K. & LEE, K. S. 2010. Tracing the geographical origin of beefs being circulated in Korean markets based on stable isotopes. *Rapid Communications in Mass Spectrometry,* 24**,** 155-159.

BONHOMMEAU, S., CHASSOT, E. & RIVOT, E. 2008. Fluctuations in European eel (*Anguilla anguilla*) recruitment resulting from environmental changes in the Sargasso Sea. *Fisheries Oceanography,* 17**,** 32-44.

BONTEMPO, L., CAMIN, F., ZILLER, L., PERINI, M., NICOLINI, G. & LARCHER, R. 2017. Isotopic and elemental composition of selected types of Italian honey. *Measurement,* 98**,** 283-289.

BOYD, P. W., JICKELLS, T., LAW, C., BLAIN, S., BOYLE, E., BUESSELER, K., COALE, K., CULLEN, J., DE BAAR, H. J. & FOLLOWS, M. 2007. Mesoscale iron enrichment experiments 1993-2005: synthesis and future directions. *Science,* 315**,** 612-617.

BROUWER, P. 2006. Theory of XRF. *Almelo, Netherlands: PANalytical BV*.

BROWN, F. & BALLS, P. 1997. Trace metals in fish and shellfish from Scottish waters. *Scottish Fisheries Research Report (United Kingdom)*.

BUDGE, S. M., IVERSON, S. J., BOWEN, W. D. & ACKMAN, R. G. 2002. Among-and within-species variability in fatty acid signatures of marine fish and invertebrates on the Scotian Shelf, Georges Bank, and southern Gulf of St. Lawrence. *Canadian Journal of Fisheries and aquatic sciences,* 59**,** 886-898.

BURGER, J., GOCHFELD, M., SHUKLA, T., JEITNER, C., BURKE, S., DONIO, M., SHUKLA, S., SNIGAROFF, R., SNIGAROFF, D. & STAMM, T. 2007. Heavy metals in Pacific cod (*Gadus macrocephalus*) from the Aleutians: location, age, size, and risk. *Journal of Toxicology and Environmental Health, Part A,* 70**,** 1897-1911.

BUSETTO, M. L., MORETTI, V. M., MORENO-ROJAS, J. M., CAPRINO, F., GIANI, I., MALANDRA, R., BELLAGAMBA, F. & GUILLOU, C. 2008. Authentication of Farmed and Wild Turbot (*Psetta maxima*) by Fatty Acid and Isotopic Analyses Combined with Chemometrics. *Journal of Agricultural and Food Chemistry,* 56**,** 2742-2750.

BYRD, J. T. 1990. Comparative geochemistries of arsenic and antimony in rivers and estuaries. *Science of the Total Environment,* 97**,** 301-314.

CAMBIÈ, G., KAISER, M. J., MARRIOTT, A. L., FOX, J., LAMBERT, G., HIDDINK, J. G., OVERY, T., BENNET, S. A., LENG, M. J. & MCCARTHY, I. D. 2016. Stable isotope signatures reveal small-scale spatial separation in populations of European sea bass. *Marine Ecology Progress Series,* 546**,** 213-223.

CAMIN, F., BONER, M., BONTEMPO, L., FAUHL-HASSEK, C., KELLY, S., RIEDL, J. & ROSSMANN, A. 2017. Stable isotope techniques for verifying the declared geographical origin of food in legal cases. *Trends in Food Science & Technology,* 61**,** 176-187.

CAMIN, F., LARCHER, R., NICOLINI, G., BONTEMPO, L., BERTOLDI, D., PERINI, M., SCHLICHT, C., SCHELLENBERG, A., THOMAS, F., HEINRICH, K., VOERKELIUS, S., HORACEK, M., UECKERMANN, H., FROESCHL, H., WIMMER, B., HEISS, G., BAXTER, M., ROSSMANN, A. & HOOGEWERFF, J. 2010. Isotopic and Elemental Data for Tracing the Origin of European Olive Oils. *Journal of Agricultural and Food Chemistry,* 58**,** 570-577.

CARDEÑOSA, D., GOLLOCK, M. J. & CHAPMAN, D. D. 2019. Development and application of a novel real-time polymerase chain reaction assay to detect illegal trade of the European eel (*Anguilla anguilla*). *Conservation Science and Practice,* 1**,** e39.

CARRERA, M. N. & GALLARDO, J. M. 2017. Determination of the geographical origin of all commercial hake species by stable isotope ratio (SIR) analysis. *Journal of agricultural and food chemistry,* 65**,** 1070-1077.

CARTER, J., TINGGI, U., YANG, X. & FRY, B. 2015. Stable isotope and trace metal compositions of Australian prawns as a guide to authenticity and wholesomeness. *Food chemistry,* 170**,** 241-248.

CARTER, J. F. & CHESSON, L. A. 2017. *Food Forensics: Stable Isotopes as a Guide to Authenticity and Origin*, CRC Press.

CHATTERJEE, N. S., CHEVALLIER, O. P., WIELOGORSKA, E., BLACK, C. & ELLIOTT, C. T. 2019. Simultaneous authentication of species identity and geographical origin of shrimps: Untargeted metabolomics to recurrent biomarker ions. *Journal of Chromatography A,* 1599**,** 75-84.

CHOUVELON, T., BRACH-PAPA, C., AUGER, D., BODIN, N., BRUZAC, S., CROCHET, S., DEGROOTE, M., HOLLANDA, S. J., HUBERT, C. & KNOERY, J. 2017. Chemical contaminants (trace metals, persistent organic pollutants) in albacore tuna from western Indian and south-eastern Atlantic Oceans: trophic influence and potential as tracers of populations. *Science of the Total Environment,* 596**,** 481-495.

CORMIE, A. B., LUZ, B. & SCHWARCZ, H. P. 1994. Relationship between the hydrogen and oxygen isotopes of deer bone and their use in the estimation of relative humidity. *Geochimica et Cosmochimica Acta,* 58**,** 3439-3449.

CROUDACE, I. W., RINDBY, A. & ROTHWELL, R. G. 2006. ITRAX: description and evaluation of a new multi-function X-ray core scanner. *Geological Society, London, Special Publications,* 267**,** 51-63.

CRYAN, P. M., BOGAN, M. A., RYE, R. O., LANDIS, G. P. & KESTER, C. L. 2004. Stable hydrogen isotope analysis of bat hair as evidence for seasonal molt and long-distance migration. *Journal of Mammalogy,* 85**,** 995-1001.

CUSA, M., ST JOHN GLEW, K., TRUEMAN, C., MARIANI, S., BUCKLEY, L., NEAT, F. & LONGO, C. 2021. A future for seafood point-of-origin testing using DNA and stable isotope signatures. *Reviews in Fish Biology and Fisheries***,** 1-25.

CUTTER, G. A., CUTTER, L. S., FEATHERSTONE, A. M. & LOHRENZ, S. E. 2001. Antimony and arsenic biogeochemistry in the western Atlantic Ocean. *Deep Sea Research Part II: Topical Studies in Oceanography,* 48**,** 2895-2915.

DAVIS, T. 1986. Migration patterns in barramundi, *Lates calcarifer* (Bloch), in Van Diemen Gulf, Australia, with estimates of fishing mortality in specific areas. *Fisheries Research,* 4**,** 243-258.

DE AQUINO FERREIRA, K., BRAGA, A. A. & DI BENEDITTO, A. P. M. 2021. Can stable isotopes be applied to determine shrimp stocks origin in SE Brazil? An approach for utilization in fishery management. *Ocean & Coastal Management,* 205**,** 105500.

DE BAAR, H. J., BOYD, P. W., COALE, K. H., LANDRY, M. R., TSUDA, A., ASSMY, P., BAKKER, D. C., BOZEC, Y., BARBER, R. T. & BRZEZINSKI, M. A. 2005. Synthesis of iron fertilization experiments: from the iron age in the age of enlightenment. *Journal of Geophysical Research: Oceans,* 110**,** C09S16.

DEKKER, W. 2019. The history of commercial fisheries for European eel commenced only a century ago. *Fisheries Management and Ecology,* 26**,** 6-19.

DEL RIO-LAVÍN, A., DÍAZ-ARCE, N., LARRAÍN, M. A., ARANEDA, C., RODRÍGUEZ-EZPELETA, N., JIMÉNEZ, E. & PARDO, M. Á. 2022a. Population structure and geographic origin assignment of *Mytilus galloprovincialis* mussels using SNPs. *Aquaculture,* 550**,** 737836.

DEL RIO-LAVÍN, A., WEBER, J., MOLKENTIN, J., JIMÉNEZ, E., ARTETXE-ARRATE, I. & PARDO, M. Á. 2022b. Stable isotope and trace element analysis for tracing the geographical origin of the Mediterranean mussel (*Mytilus galloprovincialis*) in food authentication. *Food Control,* 139**,** 109069.

DEMINA, L. L., DARA, O., ALIEV, R., ALEKSEEVA, T., BUDKO, D., NOVICHKOVA, E., POLITOVA, N., SOLOMATINA, A. & BULOKHOV, A. 2020. Elemental and mineral composition of the Barents Sea recent and Late Pleistocene− Holocene sediments: A correlation with environmental conditions. *Minerals,* 10**,** 593.

DFO 2022. Stock assessment of Northern cod (NAFO Divisions 2J3KL) in 2021. *Canadian Science Advisory Secretariat Science Advisory Report 2022/041*.

DRINAN, D. P., GRUENTHAL, K. M., CANINO, M. F., LOWRY, D., FISHER, M. C. & HAUSER, L. 2018. Population assignment and local adaptation along an isolation-by-distance gradient in Pacific cod (*Gadus macrocephalus*). *Evolutionary Applications,* 00**,** 1-17.

DWYER, K., PARRISH, C. & BROWN, J. 2003. Lipid composition of yellowtail flounder (*Limanda ferruginea*) in relation to dietary lipid intake. *Marine Biology,* 143**,** 659-667.

DYCK, A. J. & SUMAILA, U. R. 2010. Economic impact of ocean fish populations in the global fishery. *Journal of Bioeconomics,* 12**,** 227-243.

EL SHEIKHA, A. F. & MONTET, D. 2016. How to Determine the Geographical Origin of Seafood? *Critical Reviews in Food Science and Nutrition,* 56**,** 306-317.

EL SHEIKHA, A. F. & XU, J. 2017. Traceability as a key of seafood safety: reassessment and possible applications. *Reviews in Fisheries Science & Aquaculture,* 25**,** 158-170.

ELDERFIELD, H. & SCHULTZ, A. 1996. Mid-ocean ridge hydrothermal fluxes and the chemical composition of the ocean. *Annual Review of Earth and Planetary Sciences,* 24**,** 191-224.

ELLIOTT, C. 2014. Elliott Review into the integrity and assurance of food supply networks-Final report: A national food crime prevention framework. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/350726/elliot-review-final-report-july2014.pdf.

ESPIÑEIRA, M. & VIEITES, J. M. 2016. Genetic system for an integral traceability of European eel (*Anguilla anguilla*) in aquaculture and seafood products: authentication by fast real-time PCR. *European Food Research and Technology,* 242**,** 25-31.

EU. 2007. Council Regulation (EC) No. 1100/2007 of 18 September 2007 establishing measures for the recovery of the stock of European eel. *Official Journal of the European Union* [Online], 248. Available: https://eur-lex.europa.eu/eli/reg/2007/1100/oj.

EUROPEAN COMMISSION. 2018. FISHPOPTRACE Report Summary. Available: https://cordis.europa.eu/result/rcn/56107_en.html.

EUROPEAN COMMISSION. 2019. Directorate-General for Maritime Affairs and Fisheries, MacNab, S., Luchetta, G., Nimmo, F., et al.,. *Evaluation of the Eel Regulation: final report* [Online]. Available: https://data.europa.eu/doi/10.2771/679816.

EUROPEAN COMMISSION. 2022. The EU Fish Market, 2022 Edition. Available: https://data.europa.eu/doi/10.2771/716731.

EUROPEAN UNION. 2013. *Regulation (EU) No 1379/2013 of the European Parliament and of the Council of 11 December 2013 on the common organisation of the markets in fishery and aquaculture products, amending Council Regulations (EC) No 1184/2006 and (EC) No 1224/2009 and repealing Council Regulation (EC) No 104/2000* [Online]. Available: http://extwprlegs1.fao.org/docs/pdf/eur141697.pdf [Accessed 14/06/18].

EUROPOL. 2018. *Glass eel traffickers earned more than Eur 37 million from illegal exports to Asia* [Online]. Available: https://www.europol.europa.eu/newsroom/news/glass-eel-traffickers-earned-more-eur-37-million-illegal-exports-to-asia [Accessed 16/06/18].

EUROPOL. 2022. *49 individuals across Europe arrested in major blow to eels trafficking* [Online]. Available: https://www.europol.europa.eu/media-press/newsroom/news/49-individuals-across-europe-arrested-in-major-blow-to-eels-trafficking [Accessed 04/08/2022].

EVISA. 2022. *Material Database - BCR-422: Trace elements in cod muscle* [Online]. Available: http://www.speciation.net/Database/Materials/JRC--Joint-Research-Centre/BCR422-Trace-elements-in-cod-muscle-;i196 [Accessed 01/04/2022].

FAO 2016. The State of World Fisheries and Aquaculture 2016. Contributing to food security and nutrition for all. Rome. 200 pp.

FAO. 2018. Overview of food fraud in the fisheries sector, by Alan Reilly. *Fisheries and Aquaculture Circular No. 1165* [Online]. Available: www.fao.org/3/i8791en/I8791EN.pdf.

FAO 2019. Regional Conference on river habitat restoration for inland fisheries in the Danube river Basin and adjacent Black Sea Areas. *Conference Proceedings, 13–15 November 2018, Bucharest, Romania. FAO Fisheries and Aquaculture Proceedings No. 63. Rome*.

FAO. 2020a. The State of Mediterranean and Black Sea Fisheries 2020. Available: https://doi.org/10.4060/cb2429en.

FAO. 2020b. The State of World Fisheries and Aquaculture 2020. Sustainability in Action. Rome. Available: https://doi.org/10.4060/ca9229en.

FAO. 2022a. *Fisheries and Aquaculture, Global capture production Quantity (1950 - 2019)* [Online]. Available: https://www.fao.org/fishery/statistics-query/en/capture/capture_quantity [Accessed 29/03/2022].

FAO. 2022b. *Fisheries and Aquaculture: Merluccius merluccius (Linneaus, 1758)* [Online]. Available: https://www.fao.org/fishery/en/aqspecies/2238 [Accessed 01/07/2022 2022].

FAO. 2022c. The State of World Fisheries and Aquaculture 2022. Towards Blue Transformation. Rome. Available: https://doi.org/10.4060/cc0461en.

FARABEGOLI, F., PIRINI, M., ROTOLO, M., SILVI, M., TESTI, S., GHIDINI, S., ZANARDI, E., REMONDINI, D., BONALDO, A. & PARMA, L. 2018. Towards the authentication of European sea bass origin through a combination of biometric measurements and multiple analytical techniques. *Journal of agricultural and food chemistry,* 66**,** 6822−6831.

FAZIO, G., SASAL, P., MOUAHID, G., LECOMTE-FINIGER, R. & MONÉ, H. 2012. Swim bladder nematodes (*Anguillicoloides crassus*) disturb silvering in European eels (*Anguilla anguilla*). *Journal of Parasitology,* 98**,** 695-705.

FERRATON, F., HARMELIN-VIVIEN, M., MELLON-DUVAL, C. & SOUPLET, A. 2007. Spatio-temporal variation in diet may affect condition and abundance of juvenile European hake in the Gulf of Lions (NW Mediterranean). *Marine Ecology Progress Series,* 337**,** 197-208.

FONSECA, V. F., DUARTE, I. A., MATOS, A. R., REIS-SANTOS, P. & DUARTE, B. 2022. Fatty acid profiles as natural tracers of provenance and lipid quality indicators in illegally sourced fish and bivalves. *Food Control,* 134**,** 108735.

FOOD STANDARDS AGENCY. 2022. *Shellfish classification* [Online]. Available: https://www.food.gov.uk/business-guidance/shellfish-classification [Accessed 11/07/2022].

FORLEO, T., ZAPPI, A., MELUCCI, D., CIRIACI, M., GRIFFONI, F., BACCHIOCCHI, S., SIRACUSA, M., TAVOLONI, T. & PIERSANTI, A. 2021. Inorganic elements in *Mytilus galloprovincialis* shells: Geographic traceability by multivariate analysis of ICP-MS data. *Molecules,* 26**,** 2634.

FRANKLIN, A. 1987. The concentration of metals, organochlorine pesticide and PCB residues in marine fish and shellfish: Results of MAFF fish and shellfish monitoring programmes, 1977-84. *Science series aquatic environment monitoring report. Centre for Environment, Fisheries and Aquaculture Science. 1987.*

FRY, B. 2006. *Stable isotope ecology*, Springer.

GADD, P., GOPI, K., SAMMUT, J., SAINTILAN, N., CRAWFORD, J. & MAZUMDER, D. 2018. Itrax micro X-ray fluorescence (μXRF) for soft biological tissues. *MethodsX,* 5**,** 1267-1271.

GEE, J., JACK, L. & BUTTON, M. 2014. Minimising fraud and maximising value in the UK food and drink sector: how the UK food and drink sector can improve profitability by cutting the cost of fraud. Available: http://www.port.ac.uk/media/contacts-and-departments/icjs/ccfs/PKF_LittleJohn_-_Fraud_in_the_food_sector_report.pdf.

GENZ, A., BRETZ, F., MIWA, T., MI, X., LEISCH, F., SCHEIPL, F. & HOTHORN, T. 2020. mvtnorm: Multivariate Normal and t Distributions. R package version 1.1-1. Available: http://CRAN.R-project.org/package=mvtnorm.

GEPHART, J. A., HENRIKSSON, P. J. G., PARKER, R. W. R., SHEPON, A., GOROSPE, K. D., BERGMAN, K., ESHEL, G., GOLDEN, C. D., HALPERN, B. S., HORNBORG, S., JONELL, M., METIAN, M., MIFFLIN, K., NEWTON, R., TYEDMERS, P., ZHANG, W., ZIEGLER, F. & TROELL, M. 2021. Environmental performance of blue foods. *Nature,* 597**,** 360-365.

GFCM. 2016. REC.CM-GFCM/40/2016/5 establishing a minimum conservation reference size for European hake in the Mediterranean Sea. Available: https://www.fao.org/gfcm/decisions/ru/.

GIBBS, P., LANGSTON, W., BURT, G. & PASCOE, P. 1983. *Tharyx marioni* (Polychaeta): a remarkable accumulator of arsenic. *Journal of the Marine Biological Association of the United Kingdom,* 63**,** 313-325.

GO, Y.-S., WON, E.-J., KIM, S.-H., LEE, D.-H., KANG, J.-H. & SHIN, K.-H. 2022. Stepwise Approach for Tracing the Geographical Origins of the Manila Clam *Ruditapes philippinarum* Using Dual-Element Isotopes and Carbon Isotopes of Fatty Acids. *Foods,* 11**,** 1965.

GONG, Y., LI, Y., CHEN, X. & CHEN, L. 2018. Potential use of stable isotope and fatty acid analyses for traceability of geographic origins of jumbo squid (*Dosidicus gigas*). *Rapid Communications in Mass Spectrometry,* 32**,** 583-589.

GONZALEZ, E. G., KREY, G., ESPINEIRA, M., DIEZ, A., PUYET, A. & BAUTISTA, J. M. 2010. Population proteomics of the European Hake (*Merluccius merluccius*). *Journal of proteome research,* 9**,** 6392-6404.

GOPI, K., MAZUMDER, D., SAINTILAN, N., YUSOFF, F. M. & SAMMUT, J. 2018. Stable isotope analysis as a tool for determining seafood provenance. *Journal of Environmental Biology,* 39**,** 741-746.

GOPI, K., MAZUMDER, D., SAMMUT, J. & SAINTILAN, N. 2019a. Determining the provenance and authenticity of seafood: a review of current methodologies. *Trends in Food Science & Technology,* 91**,** 294–304.

GOPI, K., MAZUMDER, D., SAMMUT, J., SAINTILAN, N., CRAWFORD, J. & GADD, P. 2019b. Combined use of stable isotope analysis and elemental profiling to determine provenance of black tiger prawns (*Penaeus monodon*). *Food Control,* 95**,** 242-248.

GOPI, K., MAZUMDER, D., SAMMUT, J., SAINTILAN, N., CRAWFORD, J. & GADD, P. 2019c. Isotopic and elemental profiling to trace the geographic origins of farmed and wild-caught Asian seabass (*Lates calcarifer*). *Aquaculture,* 502**,** 56-62.

GOV.UK. 2022. *Guidance: Permission to trap crayfish, eels, elvers, salmon and sea trout* [Online]. Available: https://www.gov.uk/guidance/permission-to-trap-crayfish-eels-elvers-salmon-and-sea-trout [Accessed 05/08/2022].

GRÄLER, B., PEBESMA, E. & HEUVELINK, G. 2016. Spatio-Temporal Interpolation using gstat. *The R Journal,* 8(1)**,** 204-218.

HARVARD HEALTH PUBLISHING. 2019. *The importance of potassium* [Online]. Available: https://www.health.harvard.edu/staying-healthy/the-importance-of-potassium [Accessed 01/08/2022].

HELLOU, J., WARREN, W., PAYNE, J., BELKHODE, S. & LOBEL, P. 1992. Heavy metals and other elements in three tissues of cod, *Gadus morhua* from the Northwest Atlantic. *Marine Pollution Bulletin,* 24**,** 452-458.

HELYAR, S. J., LLOYD, H. A. D., DE BRUYN, M., LEAKE, J., BENNETT, N. & CARVALHO, G. R. 2014. Fish product mislabelling: failings of traceability in the production chain and implications for illegal, unreported and unregulated (IUU) fishing. *PloS one,* 9 (6)**,** e98691.

HENKEL, C. V., BURGERHOUT, E., DE WIJZE, D. L., DIRKS, R. P., MINEGISHI, Y., JANSEN, H. J., SPAINK, H. P., DUFOUR, S., WELTZIEN, F.-A. & TSUKAMOTO, K. 2012. Primitive duplicate Hox clusters in the European eel's genome. *PloS one,* 7**,** e32231.

HENRY, F., AMARA, R., COURCOT, L., LACOUTURE, D. & BERTHO, M. L. 2004. Heavy metals in four fish species from the French coast of the Eastern English Channel and Southern Bight of the North Sea. *Environment International,* 30**,** 675-683.

HOBSON, K. A., VAN WILGENBURG, S. L., WESOOWSKI, T., MAZIARZ, M., BIJLSMA, R. G., GRENDELMEIER, A. & MALLORD, J. W. 2014. A multi-isotope ($\delta^{2}H$, $\delta^{13}C$, $\delta^{15}N$) approach to establishing migratory connectivity in palearctic-afrotropical migrants: an example using wood warblers *Phylloscopus sibilatrix*. *Acta Ornithologica,* 49**,** 59-71.

HOBSON, K. A., WASSENAAR, L. I. & TAYLOR, O. R. 1999. Stable isotopes (δD and $\delta^{13}C$) are geographic indicators of natal origins of monarch butterflies in eastern North America. *Oecologia,* 120**,** 397-404.

HORACEK, M. & MIN, J.-S. 2010. Discrimination of Korean beef from beef of other origin by stable isotope measurements. *Food Chemistry,* 121**,** 517-520.

ICES. 2005. ICES FishMap species factsheet - Cod. Available: https://www.ices.dk/about-ICES/projects/EU-RFP/EU%20Repository/ICES%20FIshMap/ICES%20FishMap%20species%20factsheet-cod.pdf.

ICES. 2013. Report of the Working Group on the Assessment of Southern Shelf Stocks of Hake, Monk and Megrim (WGHMM). 10 - 16 May 2013, ICES Headquarters, Copenhagen. ICES CM 2013/ACOM:11A. 11 pp. Available: https://archimer.ifremer.fr/doc/00165/27664/25840.pdf.

ICES 2020. Joint EIFAAC/ICES/GFCM Working Group on Eels (WGEEL). *ICES Scientific Reports,* 2**,** 223.

ICES 2021a. Cod (*Gadus morhua*) in Division 5.a (Iceland grounds). *Report of the ICES Advisory Committee, 2021. ICES Advice 2021, cod.27.5a*.

ICES 2021b. Cod (*Gadus morhua*) in Subarea 4, Division 7.d, and Subdivision 20 (North Sea, eastern English Channel, Skagerrak). *Report of the ICES Advisory Committee, 2021. ICES Advice 2021, cod.27.47d20*.

ICES 2021c. Cod (*Gadus morhua*) in subareas 1 and 2 (Northeast Arctic). *Report of the ICES Advisory Committee, 2021. ICES Advice 2021, cod.27.1-2*.

ICES 2021d. Cod (*Gadus morhua*) in subdivisions 24–32, eastern Baltic stock (eastern Baltic Sea). *Report of the ICES Advisory Committee, 2021. ICES Advice 2021, cod.27.24-32*.

ICES 2021e. European eel (*Anguilla anguilla*) throughout its natural range *Report of the ICES Advisory Committee, 2021*.

ICES 2021f. Haddock (*Melanogrammus aeglefinus*) in Division 5.b (Faroes grounds). *Report of the ICES Advisory Committee, 2021. ICES Advice 2021, had.27.5b*.

ICES 2021g. Haddock (*Melanogrammus aeglefinus*) in Division 6.b (Rockall). *Report of the ICES Advisory Committee, 2021. ICES Advice 2021, had.27.6b*.

ICES 2021h. Haddock (*Melanogrammus aeglefinus*) in subareas 1 and 2 (Northeast Arctic). *Report of the ICES Advisory Committee, 2021. ICES Advice 2021, had.27.1-2*.

ICES 2022a. Cod (*Gadus morhua*) in Division 6.a (West of Scotland). ICES Advice: Recurrent Advice. Report

ICES 2022b. Cod (*Gadus morhua*) in divisions 7.e–k (western English Channel and southern Celtic Seas). ICES Advice: Recurrent Advice. Report.

ICES 2022c. Cod (*Gadus morhua*) in Subarea 4, Division 7.d, and Subdivision 20 (North Sea, eastern English Channel, Skagerrak). ICES Advice: Recurrent Advice. Report.

ICES 2022d. Cod (*Gadus morhua*) in subdivisions 22–24, western Baltic stock (western Baltic Sea). ICES Advice: Recurrent Advice. Report.

ICES 2022e. Cod (*Gadus morhua*) in subdivisions 24–32, eastern Baltic stock (eastern Baltic Sea). ICES Advice: Recurrent Advice. Report.

ICES 2022f. European eel (*Anguilla anguilla*) throughout its natural range. ICES Advice: Recurrent Advice. Report.

ICES 2022g. Hake (*Merluccius merluccius*) in subareas 4, 6, and 7, and in divisions 3.a, 8.a–b, and 8.d, Northern stock (Greater North Sea, Celtic Seas, and the northern Bay of Biscay). *Report of the ICES Advisory Committee, 2022. ICES Advice 2022, hke.27.3a46-8abd*.

ICES. 2022h. ICES Advice 2022. ICES Advice Publications. Collection. Available: 10.17895/ices.pub.c.5796935.v91

IGUCHI, J., ISSHIKI, M., TAKASHIMA, Y., YAMASHITA, Y. & YAMASHITA, M. 2014. Identifying the origin of *Corbicula* clams using trace element analysis. *Fisheries science,* 80**,** 1089-1096.

INTERPOL. 2020. Contribution on Eel-Icit Trade. Wildlife Crime Working Group Newsletter / 7. January 2020, 2020. . Available: https://www.sustainableeelgroup.org/global-police-operation-tackles-illicit-trade-in-eel-meat/.

INTERPOL/EUROPOL. 2017. Operation Opson VI: Targeting counterfeit and substandard foodstuff and beverage. Available: www.europol.europa.eu/sites/default/files/documents/opson_vi_report.pdf [Accessed December 2016 - March 2017].

INTERPOL/EUROPOL. 2018. Operation OPSON VII Analysis Report. Available: https://www.europol.europa.eu/publications-events/publications/operation-opson-vii-analysis-report.

ISO 5725-4. 2020. Accuracy (trueness and precision) of measurement methods and results - Part 4: Basic methods for the determination of the trueness of a standard measurement method. Available: https://www.iso.org/standard/69421.html.

IUCN. 2018. *The IUCN Red List of Threatened Species - Anguilla anguilla* [Online]. Available: http://www.iucnredlist.org/details/60344/0 [Accessed 15/06/18].

IUCN. 2022. *The IUCN Red List of Threatened Species - European Hake: Merluccius merluccius* [Online]. Available: https://www.iucnredlist.org/species/198562/9031395 [Accessed 01/07/2022 2022].

JÉHANNET, P., PALSTRA, A. P., HEINSBROEK, L. T., KRUIJT, L., DIRKS, R. P., SWINKELS, W. & KOMEN, H. 2021. What goes wrong during early development of artificially reproduced European eel *Anguilla anguilla*? Clues from the larval transcriptome and gene expression patterns. *Animals,* 11**,** 1710.

JENNINGS, S., BARNES, C., SWEETING, C. J. & POLUNIN, N. V. 2008. Application of nitrogen stable isotope analysis in size-based marine food web and macroecological research. *Rapid Communications in Mass Spectrometry,* 22**,** 1673-1680.

JENNINGS, S. & COGAN, S. 2015. Nitrogen and carbon stable isotope variation in northeast Atlantic fishes and squids. *Ecology,* 96**,** 2568-2568.

JENNINGS, S., PINNEGAR, J. K., POLUNIN, N. V. & BOON, T. W. 2001. Weak cross-species relationships between body size and trophic level belie powerful size-based trophic structuring in fish communities. *Journal of Animal Ecology,* 70**,** 934-944.

JENNINGS, S. & VAN DER MOLEN, J. 2015. Trophic levels of marine consumers from nitrogen stable isotope analysis: estimation and uncertainty. *ICES Journal of Marine Science,* 72**,** 2289-2300.

JICKELLS, T., AN, Z., ANDERSEN, K. K., BAKER, A., BERGAMETTI, G., BROOKS, N., CAO, J., BOYD, P., DUCE, R. & HUNTER, K. 2005. Global iron connections between desert dust, ocean biogeochemistry, and climate. *Science,* 308**,** 67-71.

JOHNSON, K. S., CHAVEZ, F. P. & FRIEDERICH, G. E. 1999. Continental-shelf sediment as a primary source of iron for coastal phytoplankton. *Nature,* 398**,** 697-700.

JULSHAMN, K., DUINKER, A., NILSEN, B. M., NEDREAAS, K. & MAAGE, A. 2013. A baseline study of metals in cod (*Gadus morhua*) from the North Sea and coastal

Norwegian waters, with focus on mercury, arsenic, cadmium and lead. *Marine Pollution Bulletin,* 72**,** 264-273.

JULSHAMN, K., LUNDEBYE, A. K., HEGGSTAD, K., BERNTSSEN, M. H. G. & BOE, B. 2004. Norwegian monitoring programme on the inorganic and organic contaminants in fish caught in the Barents Sea, Norwegian Sea and North Sea, 1994–2001. *Food Additives & Contaminants,* 21**,** 365-376.

KAIFU, K., STEIN, F., DEKKER, W., WALKER, N., ANDREW, C., DOLLOFF, K. S., AGUIRRE, A. A., NIJMAN, V. & SASAL, P. 2019. Global exploitation of freshwater eels (genus *Anguilla*): fisheries, stock status and illegal trade. *Eels Biology, Monitoring, Management, Culture and Exploitation.* 5m Books Ltd.

KANG, X., ZHAO, Y., SHANG, D., ZHAI, Y., NING, J., DING, H. & SHENG, X. 2020. Identification of the geographical origins of sea cucumbers in China: The application of stable isotope ratios and compositions of C, N, O and H. *Food Control,* 111**,** 107036.

KELLY, J. F., ATUDOREI, V., SHARP, Z. D. & FINCH, D. M. 2002. Insights into Wilson's Warbler migration from analyses of hydrogen stable-isotope ratios. *Oecologia,* 130**,** 216-221.

KILJUNEN, M., GREY, J., SINISALO, T., HARROD, C., IMMONEN, H. & JONES, R. I. 2006. A revised model for lipid-normalizing $\delta^{13}C$ values from aquatic organisms, with implications for isotope mixing models. *Journal of Applied Ecology,* 43**,** 1213-1222.

KIM, H., KUMAR, K. S., HWANG, S. Y., KANG, B.-C., MOON, H.-B. & SHIN, K.-H. 2015a. Utility of stable isotope and cytochrome oxidase I gene sequencing analyses in inferring origin and authentication of hairtail fish and shrimp. *Journal of agricultural and food chemistry,* 63**,** 5548-5556.

KIM, H., KUMAR, K. S. & SHIN, K.-H. 2015b. Applicability of stable C and N isotope analysis in inferring the geographical origin and authentication of commercial fish (Mackerel, Yellow Croaker and Pollock). *Food chemistry,* 172**,** 523-527.

KOPP, D. 2018. Personal communication.

KURLANSKY, M. 1998. *Cod: A Biography of a Fish that Changed the World*, Penguin.

KUSTKA, A. B., SAÑUDO-WILHELMY, S. A., CARPENTER, E. J., CAPONE, D., BURNS, J. & SUNDA, W. G. 2003. Iron requirements for dinitrogen-and ammonium-supported growth in cultures of *Trichodesmium* (IMS 101): Comparison with nitrogen fixation rates and iron: Carbon ratios of field populations. *Limnology and Oceanography,* 48**,** 1869-1884.

LE NGUYEN, D. D., NGOC, H. H., DIJOUX, D., LOISEAU, G. & MONTET, D. 2008. Determination of fish origin by using 16S rDNA fingerprinting of bacterial communities by PCR-DGGE: An application on *Pangasius* fish from Viet Nam. *Food control,* 19**,** 454-460.

LEAL, M. C., PIMENTEL, T., RICARDO, F., ROSA, R. & CALADO, R. 2015. Seafood traceability: current needs, available tools, and biotechnological challenges for origin certification. *Trends in Biotechnology,* 33**,** 331-336.

LEAVER, M. J., VILLENEUVE, L. A., OBACH, A., JENSEN, L., BRON, J. E., TOCHER, D. R. & TAGGART, J. B. 2008. Functional genomics reveals increases in cholesterol biosynthetic genes and highly unsaturated fatty acid biosynthesis after dietary substitution of fish oil with vegetable oils in Atlantic salmon (*Salmo salar*). *Bmc Genomics,* 9**,** 299.

LEWIS, S. G. & BOYLE, M. 2017. The expanding role of traceability in seafood: tools and key initiatives. *Journal of food science,* 82**,** A13-A21.

LI, L., BOYD, C. E. & ODOM, J. 2014. Identification of Pacific white shrimp (*Litopenaeus vannamei*) to rearing location using elemental profiling. *Food control,* 45**,** 70-75.

LI, L., BOYD, C. E., RACINE, P., MCNEVIN, A. A., SOMRIDHIVEJ, B., MINH, H. N., TINH, H. Q. & GODUMALA, R. 2017. Assessment of elemental profiling for distinguishing geographic origin of aquacultured shrimp from India, Thailand and Vietnam. *Food Control,* 80**,** 162-169.

LI, L., BOYD, C. E. & SUN, Z. 2016. Authentication of fishery and aquaculture products by multi-element and stable isotope analysis. *Food chemistry,* 194**,** 1238-1244.

LI, W., LI, S. & BEARD, B. L. 2019. Geological cycling of potassium and the K isotopic response: Insights from loess and shales. *Acta Geochimica,* 38**,** 508-516.

LIAW, A. & WIENER, M. 2002. Classification and Regression by randomForest. *R News,* 2(3)**,** 18-22.

LINKO, R. R., KAITARANTA, J. K. & VUORELA, R. 1985. Comparison of the fatty acids in Baltic herring and available plankton feed. *Comparative Biochemistry and physiology. Part B, Comparative Biochemistry,* 82**,** 699-705.

LIU, X., GUO, B., WEI, Y., SHI, J. & SUN, S. 2013. Stable isotope analysis of cattle tail hair: A potential tool for verifying the geographical origin of beef. *Food chemistry,* 140**,** 135-140.

LIU, X., XUE, C., WANG, Y., LI, Z., XUE, Y. & XU, J. 2012. The classification of sea cucumber (*Apostichopus japonicus*) according to region of origin using multi-element analysis and pattern recognition techniques. *Food Control,* 23**,** 522-527.

LOCH DUART. 2018. *Loch Duart steps up the fight against food fraudsters* [Online]. Available: https://lochduart.com/loch-duart-steps-fight-food-fraudsters/ [Accessed 13/07/2022].

LOCH DUART. 2022. *Unrivalled Scottish salmon: Origin protected* [Online]. Available: https://lochduart.com/our-salmon/#traceability [Accessed 13/07/2022].

LORRAIN, A., GRAHAM, B. S., POPP, B. N., ALLAIN, V., OLSON, R. J., HUNT, B. P., POTIER, M., FRY, B., GALVÁN-MAGAÑA, F. & MENKES, C. E. 2015. Nitrogen isotopic baselines and implications for estimating foraging habitat and trophic

position of yellowfin tuna in the Indian and Pacific Oceans. *Deep Sea Research Part II: Topical Studies in Oceanography,* 113**,** 188-198.

LUQUE, G. M. & DONLAN, C. J. 2019. The characterization of seafood mislabeling: A global meta-analysis. *Biological Conservation,* 236**,** 556-570.

MA, C., VANDER ZANDEN, H. B., WUNDER, M. B. & BOWEN, G. J. 2020. assignR: An r package for isotope-based geographic assignment. *Methods in Ecology and Evolution,* 11**,** 996-1001.

MACKENZIE, K., LONGMORE, C., PREECE, C., LUCAS, C. & TRUEMAN, C. 2014. Testing the long-term stability of marine isoscapes in shelf seas using jellyfish tissues. *Biogeochemistry,* 121**,** 441-454.

MACKENZIE, K. M., PALMER, M. R., MOORE, A., IBBOTSON, A. T., BEAUMONT, W. R., POULTER, D. J. & TRUEMAN, C. N. 2011. Locations of marine animals revealed by carbon isotopes. *Scientific Reports,* 1**,** 21.

MACLACHLAN, S. E., HUNT, J. E. & CROUDACE, I. W. 2015. An empirical assessment of variable water content and grain-size on X-ray fluorescence core-scanning measurements of deep sea sediments *Micro-XRF studies of sediment cores.* Springer.

MAGOZZI, S., YOOL, A., VANDER ZANDEN, H., WUNDER, M. & TRUEMAN, C. 2017. Using ocean models to predict spatial and temporal variation in marine carbon isotopes. *Ecosphere,* 8(5)**,** e01763.

MARINE STEWARDSHIP COUNCIL. 2018. *How does the blue MSC label compare?* [Online]. Available: https://20.msc.org/what-we-are-doing/our-approach/how-does-the-blue-msc-label-compare#studies [Accessed 03/01/18].

MARINE STEWARDSHIP COUNCIL. 2019. *Press release: North Sea cod to lose sustainability certification* [Online]. Available: https://www.msc.org/media-centre/press-releases/press-release/north-sea-cod-to-lose-sustainability-certification [Accessed 25/01/2023].

MARINE STEWARDSHIP COUNCIL. 2020a. Celebrating and supporting sustainable fisheries. *The Marine Stewardship Council Annual Report 2019-20* [Online]. Available: https://www.msc.org/docs/default-source/default-document-library/about-the-msc/msc-annual-report-2019-2020.pdf.

MARINE STEWARDSHIP COUNCIL. 2020b. *Track a fishery* [Online]. Available: https://fisheries.msc.org/en/fisheries/ [Accessed 30/04/2020].

MARINE STEWARDSHIP COUNCIL. 2022. *Track a Fishery* [Online]. Available: https://fisheries.msc.org/en/fisheries/ [Accessed 01/07/2022 2022].

MARKO, P. B., NANCE, H. A. & GUYNN, K. D. 2011. Genetic detection of mislabeled fish from a certified sustainable fishery. *Current Biology,* 21**,** R621-R622.

MARSHALL, C. R. E. 2005. *Evolutionary genetics of barramundi (Lates calcarifer) in the Australian region.* PhD thesis, Murdoch University.

MARTINO, J. C., MAZUMDER, D., GADD, P. & DOUBLEDAY, Z. A. 2022a. Tracking the provenance of octopus using isotopic and multi-elemental analysis. *Food Chemistry,* 371**,** 131133.

MARTINO, J. C., TRUEMAN, C. N., MAZUMDER, D., CRAWFORD, J. & DOUBLEDAY, Z. A. 2022b. The universal imprint of oxygen isotopes can track the origins of seafood. *Fish and Fisheries,* 00**,** 1–14.

MCLENNAN, S., TAYLOR, S., MCCULLOCH, M. T. & MAYNARD, J. 1990. Geochemical and Nd Sr isotopic composition of deep-sea turbidites: crustal evolution and plate tectonic associations. *Geochimica et cosmochimica acta,* 54**,** 2015-2050.

MEASURES, C., LANDING, W., BROWN, M. & BUCK, C. 2008. High-resolution Al and Fe data from the Atlantic Ocean CLIVAR-CO2 Repeat Hydrography A16N transect: Extensive linkages between atmospheric dust and upper ocean geochemistry. *Global biogeochemical cycles,* 22**,** GB1005.

MEIER-AUGENSTEIN, W. 2017. *Stable isotope forensics: methods and forensic applications of stable isotope analysis*, John Wiley & Sons.

MELLO, L. & ROSE, G. 2005. Seasonal growth of Atlantic cod: effects of temperature, feeding and reproduction. *Journal of Fish Biology,* 67**,** 149-170.

MILAN, M., MAROSO, F., DALLA ROVERE, G., CARRARO, L., FERRARESSO, S., PATARNELLO, T., BARGELLONI, L., CARDAZZO, B. & FARISELLI, P. 2019. Tracing seafood at high spatial resolution using NGS-generated data and machine learning: Comparing microbiome versus SNPs. *Food chemistry,* 286**,** 413-420.

MILANO, I., BABBUCCI, M., CARIANI, A., ATANASSOVA, M., BEKKEVOLD, D., CARVALHO, G. R., ESPIÑEIRA, M., FIORENTINO, F., GAROFALO, G. & GEFFEN, A. J. 2014. Outlier SNP markers reveal fine-scale genetic structuring across European hake populations (*Merluccius merluccius*). *Molecular ecology,* 23**,** 118-135.

MILLER, D. D. & MARIANI, S. 2010. Smoke, mirrors, and mislabeled cod: poor transparency in the European seafood industry. *Frontiers in Ecology and the Environment,* 8**,** 517-521.

MILLER, M. J., FEUNTEUN, E. & TSUKAMOTO, K. 2016. Did a “perfect storm” of oceanic changes and continental anthropogenic impacts cause northern hemisphere anguillid recruitment reductions? *ICES Journal of Marine Science,* 73**,** 43-56.

MINAGAWA, M. & WADA, E. 1984. Stepwise enrichment of ^{15}N along food chains: further evidence and the relation between $\delta^{15}N$ and animal age. *Geochimica et cosmochimica acta,* 48**,** 1135-1140.

MMO. 2018. *Statutory guidance Minimum Conservation Reference Sizes (MCRS) in UK waters* [Online]. Available: https://www.gov.uk/government/publications/minimum-conservation-reference-sizes-mcrs/minimum-conservation-reference-sizes-mcrs-in-uk-waters [Accessed 01/07/2022].

MMO. 2019a. Evaluating and advancing methodologies to validate the origin of capture of Angler fish (*Lophius piscatorius*). A report produced for the Marine Management Organisation, MMO Project No: MMO1167, December 2019, 25pp. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/871013/MMO1167_revised_report_Nov19_v1_RH.pdf.

MMO. 2019b. Evaluating and advancing methodologies to validate the origin of capture of fish / seafood. A report produced for the Marine Management Organisation, MMO Project No: 1193. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/871015/MMO1193_Lipids_report_MMO_Format_v5_final.pdf.

MMO. 2019c. Evaluating and advancing methodologies to validate the origin of capture of fish: Stable Isotope Technologies. MMO Project No: 1192, September 2019, 31pp. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/871008/20190702_Stable_Isotope_technologies_report_MMO_CT.2_Final_CS.pdf.

MMO 2019d. Evaluating and advancing methodologies to validate the origin of capture of fish: Trace Element technologies, MMO Project No: 1194, September 2019, 28pp. 28.

MMO 2019e. Validating the origin of capture of fish - Genetic methodologies (MMO1191). *MMO Project No: 1191*, 21.

MMO. 2021. UK Sea Fisheries Statistics 2020. Available: https://www.gov.uk/government/statistics/uk-sea-fisheries-annual-statistics-report-2020.

MONTES, I., LACONCHA, U., IRIONDO, M., MANZANO, C., ARRIZABALAGA, H. & ESTONBA, A. 2017. Reduced Single Nucleotide Polymorphism Panels for Assigning Atlantic Albacore and Bay of Biscay Anchovy Individuals to Their Geographic Origin: Toward Sustainable Fishery Management. *Journal of agricultural and food chemistry,* 65**,** 4351-4358.

MOREL, F. M. & PRICE, N. 2003. The biogeochemical cycles of trace metals in the oceans. *Science,* 300**,** 944-947.

MORRISON, L., BENNION, M., GILL, S. & GRAHAM, C. T. 2019. Spatio-temporal trace element fingerprinting of king scallops (*Pecten maximus*) reveals harvesting period and location. *Science of The Total Environment,* 697**,** 134121.

NAAUM, A. M., WARNER, K., MARIANI, S., HANNER, R. H. & CAROLIN, C. D. 2016. Seafood Mislabeling Incidence and Impacts. *Seafood Authenticity and Traceability: A DNA-based Pespective*.

NATIONAL FEDERATION OF FISH FRIERS. 2023. *Everything you need to know about fish and chips* [Online]. Available: https://www.nfff.co.uk/pages/fish-and-chips [Accessed 13/01/2023].

NATIONAL FOOD CRIME UNIT. 2016. Food Crime Annual Strategic Assessment: A 2016 Baseline. Available: https://www.food.gov.uk/sites/default/files/.../fsa-food-crime-assessment-2016.pdf.

NATIONAL WILDLIFE CRIME UNIT. 2022. *NWCU and EUROPOL meet to combat illegal trade in eels* [Online]. Available: https://www.nwcu.police.uk/news/nwcu-police-press-releases/nwcu-and-europol-meet-to-combat-illegal-trade-in-eels/ [Accessed 04/08/2022].

NEAT, F. C., BENDALL, V., BERX, B., WRIGHT, P. J., CUAIG, M. O., TOWNHILL, B., SCHÖN, P. J., LEE, J. & RIGHTON, D. 2014. Movement of Atlantic cod around the British Isles: implications for finer scale stock management. *Journal of Applied Ecology,* 51**,** 1564-1574.

NEFF, J. M. 1997. Ecotoxicology of arsenic in the marine environment. *Environmental Toxicology and Chemistry: An International Journal,* 16**,** 917-927.

NIELSDÓTTIR, M. C., MOORE, C. M., SANDERS, R., HINZ, D. J. & ACHTERBERG, E. P. 2009. Iron limitation of the postbloom phytoplankton communities in the Iceland Basin. *Global Biogeochemical Cycles,* 23**,** GB3001.

NIELSEN, E., HEMMER-HANSEN, J. & BEKKEVOLD, D. 2012a. Development and application of molecular tools to investigate the mislabeling of cod sold in Sweden *Case Studies in Food Safety and Authenticity.* Elsevier.

NIELSEN, E. E., CARIANI, A., MAC AOIDH, E., MAES, G. E., MILANO, I., OGDEN, R., TAYLOR, M., HEMMER-HANSEN, J., BABBUCCI, M. & BARGELLONI, L. 2012b. Gene-associated markers provide tools for tackling illegal fishing and false eco-certification. *Nature communications,* 3**,** 851.

NIELSEN, E. E., HANSEN, M. M., SCHMIDT, C., MELDRUP, D. & GRØNKJAER, P. 2001. Population of origin of Atlantic cod. *Nature,* 413**,** 272.

NORAMBUENA, F., LEWIS, M., HAMID, N. K. A., HERMON, K., DONALD, J. A. & TURCHINI, G. M. 2013. Fish oil replacement in current aquaculture feed: is cholesterol a hidden treasure for fish nutrition? *PLoS One,* 8**,** e81705.

OCEANA. 2021. *Wild seafood has a lower carbon footprint than red meat, cheese, and chicken, according to latest data* [Online]. Available: https://oceana.org/blog/wild-seafood-has-lower-carbon-footprint-red-meat-cheese-and-chicken-according-latest-data/ [Accessed 06/02/2023].

OLYMPUS. 2017. Vanta for Geochemistry: Handheld X-ray Fluorescence. Available: www.olympus-ims.com.

OLYMPUS. 2022. *Vanta Periodic Table - Limits of Detection* [Online]. Available: https://www.olympus-ims.com/en/downloads/detail/?0[downloads][id]=276827648 [Accessed 12/04/2022].

ORTEA, I. & GALLARDO, J. M. 2015. Investigation of production method, geographical origin and species authentication in commercially relevant shrimps using stable isotope ratio and/or multi-element analyses combined with chemometrics: An exploratory analysis. *Food chemistry,* 170**,** 145-153.

OSKARSSON, G. J., GUDMUNDSDOTTIR, A. & SIGURDSSON, T. 2009. Variation in spatial distribution and migration of Icelandic summer-spawning herring. *ICES Journal of Marine Science,* 66**,** 1762-1767.

OSPAR COMMISSION 2000. *Quality Status Report 2000, Region II – Greater North Sea*, OSPAR Commission, London. 136 + xiii pp.

OXFORD NANOPORE TECHNOLOGIES. 2022. Available: https://nanoporetech.com/ [Accessed 21/07/2022].

PALENIK, B., BRAHAMSHA, B., LARIMER, F., LAND, M., HAUSER, L., CHAIN, P., LAMERDIN, J., REGALA, W., ALLEN, E. & MCCARREN, J. 2003. The genome of a motile marine *Synechococcus*. *Nature,* 424**,** 1037-1042.

PALM, S., DANNEWITZ, J., PRESTEGAARD, T. & WICKSTRÖM, H. 2009. Panmixia in European eel revisited: no genetic difference between maturing adults from southern and northern Europe. *Heredity,* 103**,** 82-89.

PARDO, M. Á., JIMÉNEZ, E. & PÉREZ-VILLARREAL, B. 2016. Misdescription incidents in seafood sector. *Food Control,* 62**,** 277-283.

PARKER, R. W., BLANCHARD, J. L., GARDNER, C., GREEN, B. S., HARTMANN, K., TYEDMERS, P. H. & WATSON, R. A. 2018. Fuel use and greenhouse gas emissions of world fisheries. *Nature Climate Change,* 8**,** 333-337.

PAVITT, A., MALSCH, K., KING, E., CHEVALIER, A., KACHELRIESS, D., VANNUCCINI, S. & FRIEDMAN, K. 2021. CITES and the sea: Trade in commercially exploited CITES-listed marine species. *FAO Fisheries and Aquaculture Technical Paper No. 666. Rome, FAO*.

PEBESMA, E. J. 2004. Multivariable geostatistics in S: the gstat package. *Computers & Geosciences,* 30**,** 683-691.

PERINI, M., BONTEMPO, L., ZILLER, L., BARBERO, A., CALIGIANI, A. & CAMIN, F. 2016. Stable isotope composition of cocoa beans of different geographical origin. *Journal of Mass Spectrometry,* 51**,** 684-689.

PETERSON, B. J. & FRY, B. 1987. Stable isotopes in ecosystem studies. *Annual review of ecology and systematics,* 18**,** 293-320.

PHILLIPS, D. & DEPLEDGE, M. 1986. Distribution of inorganic and total arsenic in tissues of the marine gastropod *Hemifusus ternatanus*. *Marine Ecology Progress Series,* 34**,** 261-266.

PIMENTEL, T., MARCELINO, J., RICARDO, F., SOARES, A. M. V. M. & CALADO, R. 2017. Bacterial communities 16S rDNA fingerprinting as a potential tracing tool for cultured seabass *Dicentrarchus labrax*. *Scientific Reports,* 7**,** 11862.

PITA, A., PEREZ, M., CERVINO, S. & PRESA, P. 2011. What can gene flow and recruitment dynamics tell us about connectivity between European hake stocks in the Eastern North Atlantic? *Continental Shelf Research,* 31**,** 376-387.

POORE, J. & NEMECEK, T. 2018. Reducing food's environmental impacts through producers and consumers. *Science,* 360**,** 987-992.

POTTS, P. J. 1992. *A handbook of silicate rock analysis*, Blackie & Son Ltd.

PRESAS-BASALO, F. X. 2022. Potassium Homeostasis and Fish Welfare in Coupled Aquaponic Systems. *Fisheries and Aquaculture Journal,* 13**,** 290.

PUSTJENS, A. M., BOERRIGTER-EENLING, R., KOOT, A. H. & RUTH, S. 2018. Food authenticity: provenancing. A case study of fish. *Descriptive food science.* IntechOpen London, UK.

QUEVAUVILLER, P., KRAMER, G. & GRIEPINK, B. 1992. A new certified reference material for the quality control of trace elements in marine monitoring: cod muscle (CRM 422). *Marine pollution bulletin,* 24**,** 601-606.

QUINN, B. 2013. *Horse DNA found in beefburgers from four major supermarkets* [Online]. The Guardian. Available: https://www.theguardian.com/world/2013/jan/15/horse-dna-found-supermarket-beefburgers [Accessed 03/01/18].

R CORE TEAM. 2019. R: A language and environment for statistical computing. Available: https://www.R-project.org/.

RAINBOW, P. S. 2018. *Trace metals in the environment and living organisms: the British Isles as a case study*, Cambridge University Press.

RAMPAZZO, F., TOSI, F., TEDESCHI, P., GION, C., ARCANGELI, G., BRANDOLINI, V., GIOVANARDI, O., MAIETTI, A. & BERTO, D. 2020. Preliminary multi analytical approach to address geographic traceability at the intraspecific level in Scombridae family. *Isotopes in Environmental and Health Studies,* 56(3)**,** 260-279.

REIS-SANTOS, P., GILLANDERS, B. M., STURROCK, A. M., IZZO, C., OXMAN, D. S., LUEDERS-DUMONT, J. A., HÜSSY, K., TANNER, S. E., ROGERS, T. & DOUBLEDAY, Z. A. 2022. Reading the biomineralized book of life: expanding otolith biogeochemical research and applications for fisheries and ecosystem-based management. *Reviews in Fish Biology and Fisheries***,** 1-39.

REUSS, N. & POULSEN, L. 2002. Evaluation of fatty acids as biomarkers for a natural plankton community. A field study of a spring bloom and a post-bloom period off West Greenland. *Marine Biology,* 141**,** 423-434.

RICARDO, F., PIMENTEL, T., GÉNIO, L. & CALADO, R. 2017. Spatio-temporal variability of trace elements fingerprints in cockle (*Cerastoderma edule*) shells and its relevance for tracing geographic origin. *Scientific reports,* 7**,** 1-9.

RICHARDS, J. L., SHENG, V., YI, C. W., YING, C. L., TING, N. S., SADOVY, Y. & BAKER, D. 2020. Prevalence of critically endangered European eel (*Anguilla anguilla*) in Hong Kong supermarkets. *Science advances,* 6**,** eaay0317.

RIGHTON, D., WESTERBERG, H., FEUNTEUN, E., ØKLAND, F., GARGAN, P., AMILHAT, E., METCALFE, J., LOBON-CERVIA, J., SJÖBERG, N. & SIMON, J. 2016. Empirical observations of the spawning migration of European eels: The long and dangerous road to the Sargasso Sea. *Science Advances,* 2**,** e1501694.

RIGHTON, D. A. & METCALFE, J. D. 2011. Eel Migrations. *Fish Migrations.* San Diego: Academic Press.

ROCHFORT, S. J., EZERNIEKS, V., MAHER, A. D., INGRAM, B. A. & OLSEN, L. 2013. Mussel metabolomics - species discrimination and provenance determination. *Food research international,* 54**,** 1302-1312.

RUMMEL, S., HOELZL, S., HORN, P., ROSSMANN, A. & SCHLICHT, C. 2010. The combination of stable isotope abundance ratios of H, C, N and S with $^{87}Sr/^{86}Sr$ for geographical origin assignment of orange juices. *Food Chemistry,* 118**,** 890-900.

SANDOVAL-CASTELLANOS, E., URIBE-ALCOCER, M. & DÍAZ-JAIMES, P. 2010. Population genetic structure of the Humboldt squid (*Dosidicus gigas* d'Orbigny, 1835) inferred by mitochondrial DNA analysis. *Journal of Experimental Marine Biology and Ecology,* 385**,** 73-78.

SCHMIDT, J. 1923. The Breeding Places of the Eel. *Philosophical Transactions of the Royal Society of London. Series B, Containing Papers of a Biological Character,* 211**,** 179-208.

SEAFISH. 2019. Market Insight Factsheet: Cod 2019. Available: https://www.seafish.org/document/?id=5ee125a5-e7cb-4215-ab04-80e174a5098e.

SEAFISH. 2020. Seafood in multiple retail (2020 update): A market insight analysis. Available: https://www.seafish.org/document/?id=e87f48a6-9997-4c82-b4f9-d1bedf98996a.

SEAFISH. 2022. UK Seafood in Numbers 2021. Available: https://www.seafish.org/document/?id=a42c3cf8-b072-4ebe-a100-61661174a0d3.

SHIRAI, N., TERAYAMA, M. & TAKEDA, H. 2002. Effect of season on the fatty acid composition and free amino acid content of the sardine *Sardinops melanostictus*. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology,* 131**,** 387-393.

SKINNER, M. M., MARTIN, A. A. & MOORE, B. C. 2016. Is lipid correction necessary in the stable isotope analysis of fish tissues? *Rapid Communications in Mass Spectrometry,* 30**,** 881-889.

SLESSER, H. W. & TRUEMAN, C. N. 2021. Traceability of the Norway Lobster *Nephrops norvegicus* in UK Shelf Seas: A Stable Isotope Approach. *Journal of Shellfish Research,* 40**,** 153-160.

SMITH, R. G. & WATTS, C. A. 2009. Determination of the country of origin of farm-raised shrimp (Family Penaeide) using trace metal profiling and multivariate statistics. *Journal of agricultural and food chemistry,* 57**,** 8244-8249.

SOMES, C. J., SCHMITTNER, A., GALBRAITH, E. D., LEHMANN, M. F., ALTABET, M. A., MONTOYA, J. P., LETELIER, R. M., MIX, A. C., BOURBONNAIS, A. & EBY, M. 2010. Simulating the global distribution of nitrogen isotopes in the ocean. *Global Biogeochemical Cycles,* 24**,** GB4019.

SPINK, J., EMBAREK, P. B., SAVELLI, C. J. & BRADSHAW, A. 2019. Global perspectives on food fraud: results from a WHO survey of members of the International Food Safety Authorities Network (INFOSAN). *Science of Food,* 3**,** 1-5.

ST. JOHN GLEW, K., GRAHAM, L. J., MCGILL, R. A. & TRUEMAN, C. N. 2019. Spatial models of carbon, nitrogen and sulphur stable isotope distributions (isoscapes) across a shelf sea: An INLA approach. *Methods in Ecology and Evolution,* 10**,** 518-531.

ST. JOHN GLEW, K., WANLESS, S., HARRIS, M. P., DAUNT, F., ERIKSTAD, K. E., STRØM, H. & TRUEMAN, C. N. 2018. Moult location and diet of auks in the North Sea inferred from coupled light-based and isotope-based geolocation. *Marine Ecology Progress Series,* 599**,** 239-251.

STEIN, F., ZITEK, A., TRUEMAN, C. & BUNT, D. 2019. Determining the potential of otolith and tissue microchemistry for the traceability of glass eels. Available: https://www.sustainableeelgroup.org/seg_conservation/determining-the-potential-of-otolith-and-tissue-microchemistry-for-the-traceability-of-glass-eels-2/.

STEIN, F. M., WONG, J. C. Y., SHENG, V., LAW, C. S. W., SCHRÖDER, B. & BAKER, D. M. 2016. First genetic evidence of illegal trade in endangered European eel (*Anguilla anguilla*) from Europe to Asia. *Conservation Genetics Resources,* 8**,** 533-537.

SUMAILA, U. R., ALDER, J. & KEITH, H. 2006. Global scope and economics of illegal fishing. *Marine Policy,* 30**,** 696-703.

SUMAILA, U. R., ZELLER, D., HOOD, L., PALOMARES, M., LI, Y. & PAULY, D. 2020. Illicit trade in marine fish catch and its effects on ecosystems and people worldwide. *Science advances,* 6**,** eaaz3801.

SUNDA, W. G. 1989. Trace metal interactions with marine phytoplankton. *Biological Oceanography,* 6**,** 411-442.

TANAKA, H., OHSHIMO, S., TAKAGI, N. & ICHIMARU, T. 2010. Investigation of the geographical origin and migration of anchovy *Engraulis japonicus* in Tachibana Bay, Japan: A stable isotope approach. *Fisheries Research,* 102**,** 217-220.

TJALLINGII, R., RÖHL, U., KÖLLING, M. & BICKERT, T. 2007. Influence of the water content on X-ray fluorescence core-scanning measurements in soft marine sediments. *Geochemistry, Geophysics, Geosystems,* 8**,** Q02004.

TRUEMAN, C. N., JACKSON, A. L., CHADWICK, K. S., COOMBS, E. J., FEYRER, L. J., MAGOZZI, S., SABIN, R. C. & COOPER, N. 2019. Combining simulation modeling and stable isotope analyses to reconstruct the last known movements of one of Nature's giants. *PeerJ,* 7**,** e7912.

TRUEMAN, C. N., MACKENZIE, K. & PALMER, M. 2012a. Identifying migrations in marine fishes through stable-isotope analysis. *Journal of Fish Biology,* 81**,** 826-847.

TRUEMAN, C. N., MACKENZIE, K. M. & PALMER, M. R. 2012b. Stable isotopes reveal linkages between ocean climate, plankton community dynamics, and survival of two populations of Atlantic salmon (*Salmo salar*). *ICES Journal of Marine Science,* 69**,** 784-794.

TRUEMAN, C. N., MACKENZIE, K. M. & ST. JOHN GLEW, K. 2017. Stable isotope-based location in a shelf sea setting: accuracy and precision are comparable to light-based location methods. *Methods in Ecology and Evolution,* 8**,** 232-240.

TRUEMAN, C. N. & ST. JOHN GLEW, K. 2019. Isotopic tracking of marine animal movement. *Tracking animal migration with stable isotopes.* Elsevier.

TURCHINI, G. M., QUINN, G. P., JONES, P. L., PALMERI, G. & GOOLEY, G. 2008. Traceability and discrimination among differently farmed fish: a case study on Australian Murray cod. *Journal of agricultural and food chemistry,* 57**,** 274-281.

TURNER, D. R. & HUNTER, K. A. 2001. *The biogeochemistry of iron in seawater*, Wiley Chichester.

UNODC. 2020. European Glass Eels, World Wildlife Crime Report 2020: Trafficking in protected species. Available: https://www.unodc.org/documents/data-and-analysis/wildlife/2020/WWLC20_Chapter_7_Eels.pdf.

VALENTI, B., BIONDI, L., CAMPIDONICO, L., BONTEMPO, L., LUCIANO, G., DI PAOLA, F., COPANI, V., ZILLER, L. & CAMIN, F. 2017. Changes in stable isotope ratios in PDO cheese related to the area of production and green forage availability. The case study of Pecorino Siciliano. *Rapid Communications in Mass Spectrometry,* 31**,** 737-744.

VANDER ZANDEN, H. B., TUCKER, A. D., HART, K. M., LAMONT, M. M., FUJISAKI, I., ADDISON, D. S., MANSFIELD, K. L., PHILLIPS, K. F., WUNDER, M. B. & BOWEN, G. J. 2015. Determining origin in a migratory marine vertebrate: a novel method to integrate stable isotopes and satellite tracking. *Ecological Applications,* 25**,** 320-335.

VARRÀ, M. O., GHIDINI, S., HUSÁKOVÁ, L., IANIERI, A. & ZANARDI, E. 2021a. Advances in troubleshooting fish and seafood authentication by inorganic elemental composition. *Foods,* 10**,** 270.

VARRÀ, M. O., GHIDINI, S., ZANARDI, E., BADIANI, A. & IANIERI, A. 2019. Authentication of European sea bass according to production method and

geographical origin by light stable isotope ratio and rare earth elements analyses combined with chemometrics. *Italian Journal of Food Safety,* 8**,** 7872.

VARRÀ, M. O., HUSÁKOVÁ, L., PATOČKA, J., GHIDINI, S. & ZANARDI, E. 2021b. Classification of transformed anchovy products based on the use of element patterns and decision trees to assess traceability and country of origin labelling. *Food Chemistry,* 360**,** 129790.

VARRÀ, M. O., HUSÁKOVÁ, L., PATOČKA, J., GHIDINI, S. & ZANARDI, E. 2021c. Multi-element signature of cuttlefish and its potential for the discrimination of different geographical provenances and traceability. *Food Chemistry,* 356**,** 129687.

VASCONI, M., LOPEZ, A., GALIMBERTI, C., ROJAS, J. M. M., REDONDO, J. M. M., BELLAGAMBA, F. & MORETTI, V. M. 2019. Authentication of farmed and wild european eel (*Anguilla anguilla*) by fatty acid profile and carbon and nitrogen isotopic analyses. *Food control,* 102**,** 112-121.

VENABLES, W. N. & RIPLEY, B. D. 2002. Modern Applied Statistics with S. Fourth Edition. Springer, New York. ISBN 0-387-95457-0.

VIEIRA, C., MORAIS, S., RAMOS, S., DELERUE-MATOS, C. & OLIVEIRA, M. 2011. Mercury, cadmium, lead and arsenic levels in three pelagic fish species from the Atlantic Ocean: intra-and inter-specific variability and human health risks for consumption. *Food and Chemical Toxicology,* 49**,** 923-932.

VILLACORTA-RATH, C., ILYUSHKINA, I., STRUGNELL, J. M., GREEN, B. S., MURPHY, N. P., DOYLE, S. R., HALL, N. E., ROBINSON, A. J. & BELL, J. J. 2016. Outlier SNPs enable food traceability of the southern rock lobster, *Jasus edwardsii*. *Marine biology,* 163**,** 223.

VISCIANO, P. & SCHIRONE, M. 2021. Food frauds: Global incidents and misleading situations. *Trends in Food Science & Technology,* 114**,** 424-442.

VØLLESTAD, L. A. 1992. Geographic variation in age and length at metamorphosis of maturing European eel: environmental effects and phenotypic plasticity. *Journal of Animal Ecology,* 61**,** 41-48.

WARNER, K., MUSTAIN, P., LOWELL, B., GEREN, S. & TALMAGE, S. 2016. Deceptive dishes: seafood swaps found worldwide. Available: https://usa.oceana.org/publications/reports/deceptive-dishes-seafood-swaps-found-worldwide.

WATANABE, T., KIRON, V. & SATOH, S. 1997. Trace minerals in fish nutrition. *Aquaculture,* 151**,** 185-207.

WELTJE, G. J. & TJALLINGII, R. 2008. Calibration of XRF core scanners for quantitative geochemical logging of sediment cores: Theory and application. *Earth and Planetary Science Letters,* 274**,** 423-438.

WEST, J. B., EHLERINGER, J. R. & CERLING, T. E. 2007. Geography and vintage predicted by a novel GIS model of wine $\delta^{18}O$. *Journal of Agricultural and Food Chemistry,* 55**,** 7075-7083.

WETZ, M. S., HALES, B., CHASE, Z., WHEELER, P. A. & WHITNEY, M. M. 2006. Riverine input of macronutrients, iron, and organic matter to the coastal ocean off Oregon, USA, during the winter. *Limnology and Oceanography,* 51**,** 2221-2231.

WITTEVEEN, B. H., WORTHY, G. A. & ROTH, J. D. 2009. Tracing migratory movements of breeding North Pacific humpback whales using stable isotope analysis. *Marine Ecology Progress Series,* 393**,** 173-183.

WON, E.-J., KIM, S. H., GO, Y.-S., KUMAR, K. S., KIM, M.-S., YOON, S.-H., BAYON, G., KIM, J.-H. & SHIN, K.-H. 2021. A Multi-Elements Isotope Approach to Assess the Geographic Provenance of Manila Clams (*Ruditapes philippinarum*) via Recombining Appropriate Elements. *Foods,* 10**,** 646.

WUNDER, M. B. 2010. Using isoscapes to model probability surfaces for determining geographic origins. *Isoscapes.* Springer.

XU, H., ZHANG, Q., KIM, S.-K., LIAO, Z., WEI, Y., SUN, B., JIA, L., CHI, S. & LIANG, M. 2020. Dietary taurine stimulates the hepatic biosynthesis of both bile acids and cholesterol in the marine teleost, tiger puffer (*Takifugu rubripes*). *British Journal of Nutrition,* 123**,** 1345-1356.

YAMANAKA, T., MIZOTA, C. & SHIMOYAMA, S. 2003. Sulfur isotopic variations in soft tissues of five benthic animals from the reductive, tidal-flat sediments in northern Kyushu, Japan. *Marine Biology,* 142**,** 327-331.

YAMASHITA, Y., OMURA, Y. & OKAZAKI, E. 2006. Distinct regional profiles of trace element content in muscle of Japanese eel *Anguilla japonica* from Japan, Taiwan, and China. *Fisheries Science,* 72**,** 1109-1113.

YOUNG, J., BRUGGEMAN, J., RICKABY, R., EREZ, J. & CONTE, M. 2013. Evidence for changes in carbon isotopic fractionation by phytoplankton between 1960 and 2010. *Global Biogeochemical Cycles,* 27**,** 505-515.

ZAMPELAS, A. & MAGRIPLIS, E. 2019. New insights into cholesterol functions: a friend or an enemy? *Nutrients,* 11**,** 1645.

ZANDEN, M. J. V. & RASMUSSEN, J. B. 2001. Variation in $\delta^{15}N$ and $\delta^{13}C$ trophic fractionation: implications for aquatic food web studies. *Limnology and oceanography,* 46**,** 2061-2066.

ZHANG, X., CHENG, J., HAN, D., ZHAO, X., CHEN, X. & LIU, Y. 2019. Geographical origin traceability and species identification of three scallops (*Patinopecten yessoensis*, *Chlamys farreri*, and *Argopecten irradians*) using stable isotope analysis. *Food chemistry,* 299**,** 125107.

ZHANG, X., LIU, Y., LI, Y. & ZHAO, X. 2017. Identification of the geographical origins of sea cucumber (*Apostichopus japonicus*) in northern China by using stable isotope ratios and fatty acid profiles. *Food chemistry,* 218**,** 269-276.

ZHAO, X., LIU, Y., LI, Y., ZHANG, X. & QI, H. 2018. Authentication of the sea cucumber (*Apostichopus japonicus*) using amino acids carbon stable isotope fingerprinting. *Food Control,* 91**,** 128-137.