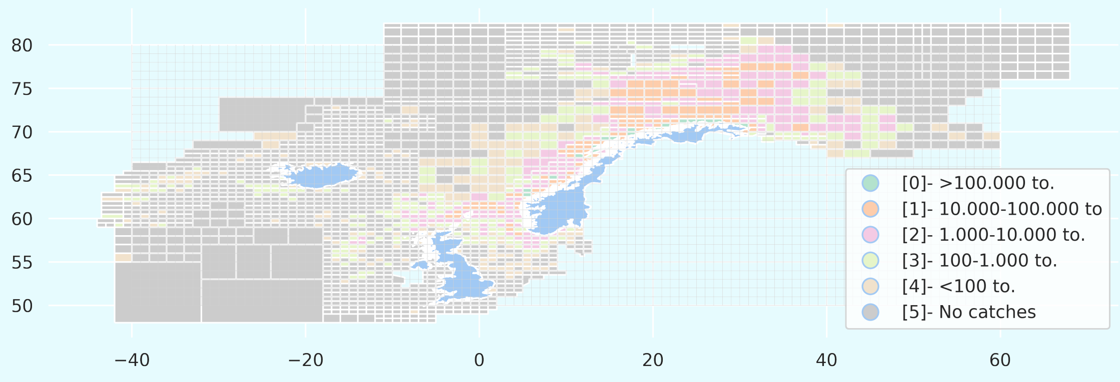
1 Dataset

I start with the presentation of the dataset because being aware of its size is essential to understand some of the decisions made in later chapters. As the method in this paper is *spatial*, the basis of the dataset is a raster. The patches of the raster are defined by Fiskeridirektoratet (218), grouping more than $10 mil.km^2$ of the north-east Atlantic into 2107 patches. The catches are retrieved from Fiskeridirektoratet (2015), the applied data is from 2002-2019, with a period duration of one month. The complete dataset consists of 2107 patches over 208 months, sums up to 404,544 observations.

Both the patches as well as the aggregated catches over all periods can be seen in Figure 1.1.

Figure 1.1: Catches per Patch

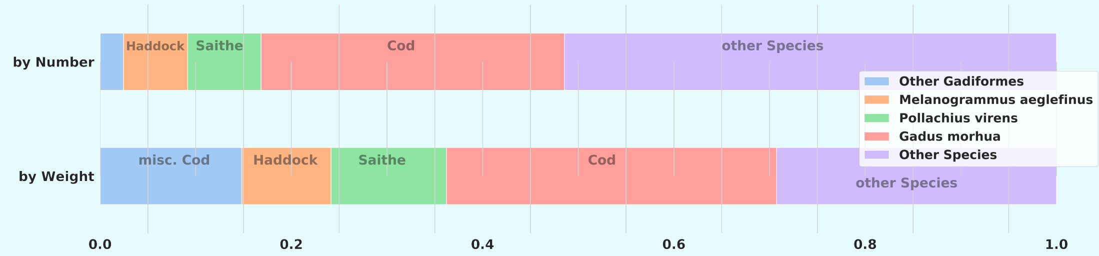


This Map displays the patches, the longitudes between $40°$ West and $60°$ East, the latitudes $50°$ North to $80°$ North. Iceland, GB and Norway are blue-coloured. The catches are aggregated over the years 2002-2019. The majority of catches are close to shore. Remarkable, more than half of the patches have none catches documented.

Figure 1.1: Catches per Patch

I decided to focus on a certain family of fishes, because different species may react contrary to environmental parameters. So the species of choice are members of *Gadioformes*, representing 48.5% of the reported catches and 70.7% of the landed mass. Figure 1.2 illustrates the detailed distribution of *Gadioformes* species. The three most important types are: Haddock (*Melanogrammus aeglefinus*), Saithe (*Pollachius virens*) and Cod (*Gadus morhua*), Cod alone represent a third of the catches. The shares of the species, by number and by weight, are displayed in Figure 1.2.

Figure 1.2: Distribution of Reported Catches



Cod is the most important species, both by weight and number, with shares higher than 30%. While 'other Species' is almost half the number of all reported catches, it is less important for the aggregated weight.

Figure 1.2: Distribution of Reported Catches

The List of *patches* is enriched with observations of static and varying parameters. The six parameters are : *Patch area*, *North Atlantic Oscillator* (NAO), *Sea Surface Temperature* (SST), *Distance*, *Quality* and *Topology*.

A detailed plot of these parameters, which are used as regression variables, is contained in Figure 1.4 at the end of this chapter. This figure furthermore has a table that describes the data.

The size of the *Patch area* can be calculated (*see*: Appendix 2.A.1) from the border coordinates. Their size varies by two factors; coastal patches are more productive and higher frequented, so they are smaller to have a more detailed classification. And the raster that shapes the patches is constructed alongside geographical coordinates, thus patches farther north become smaller as the spatial distance between *Longitudes* shrinks.

The *NAO* is an index (*details*: Appendix 2.A.2) that incorporates the general weather conditions of a month, as the sea conditions have an impact on fishing vessels.

For similar reasons, the *SST* is added, yet it is measured for each patch to capture local differences. I expect the local *SST* to have less influence on the vessels itself but rather to indicate the biological conditions beneath the surface. A detailed description of how that data is prepared, including the technical specifications, is presented in Appendix 2.A.3.

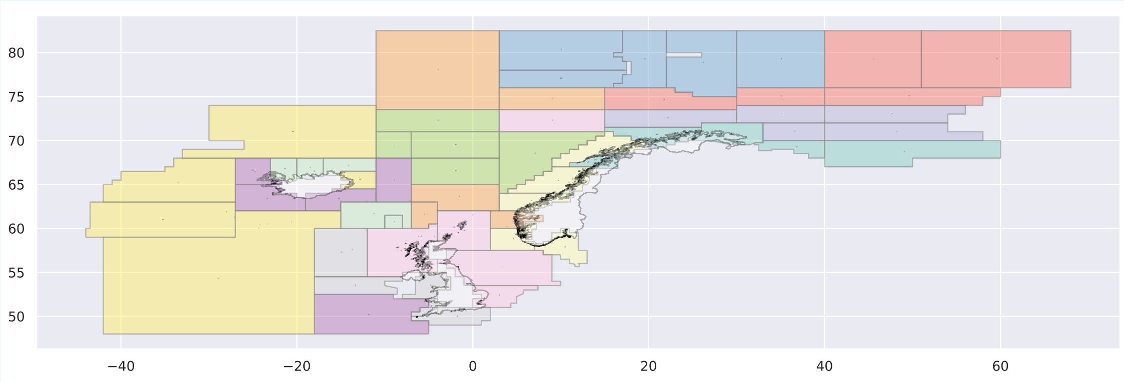
Many indicators involve *Distance*, but I neglected to use the distance between a patch and the landing station. That is because a vessel can visit a lot of different patches, but only unload at one station. Further, the decision for one station is influenced by factors that are not linked with the catch situation of the patches, e.g. the station's current workload, local market situations, or long-term contracts of the companies operating the vessels. To establish a reference point, I weighted each patch on how often it was frequented and then calculated the place with the shortest overall distance - that is the municipality of *Steigen*. Appendix 2.A.4 illustrates how to calculate distances between coordinate points on a sphere.

Quality is an index between 0 and 1 that should give evidence about the quality of catches in a patch. The problem is that quality is an ordinal scale, so the average is only a rough indicator. To have a parameter that is better manageable, the average quality is normalised to 1 for the highest value and 0 for the worst.

For the variable *Topology*, I created a parameter with the lowest average patch depth as 0 and the deepest as 1. The data is extracted from NOAA (2019).

I further signed the patches to greater areas, seen on Figure 1.3, to be able to apply them as dummies in some of the models.

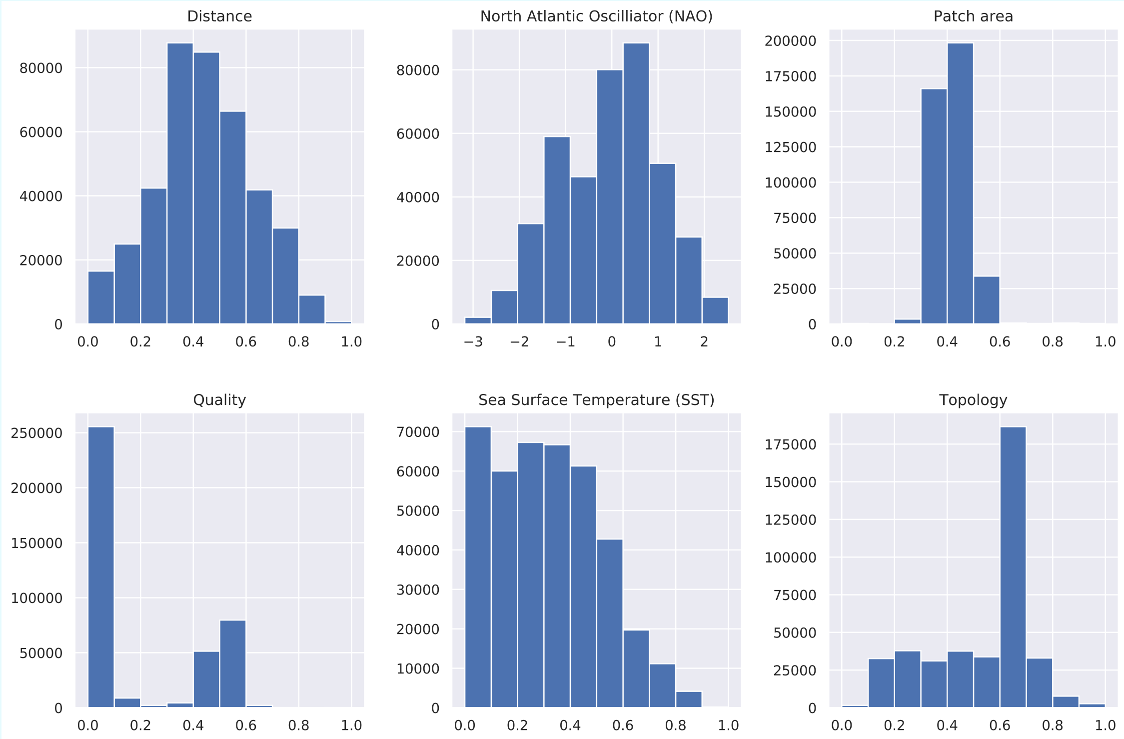
Figure 1.3: Greater Areas for Dummies



This Map displays the composed Greater Areas, the longitudes between $40°$ West and $60°$ East, the latitudes $50°$ North to $80°$ North. Iceland, GB and Norway are blank space. Using these 11 areas as dummies is more practicable than using a dummy for each patch.

Figure 1.3: Greater Areas for Dummies

Figure 1.4: Histogram of the regressed Variables



	Patch area	Nor. Atlant. Oscilliator (NAO)	Sea Surface Temp. (SST)
count	404544	404544	404544
mean	0.414200	-0.041198	0.315674
std	0.061314	1.086941	0.200369
min	0.000000	-3.179000	0.000000
25%	0.380338	-0.910747	0.153191
50%	0.406394	0.050781	0.306383
75%	0.430537	0.706248	0.459574
max	1.000000	2.521300	1.000000

	Distance	Quality	Topology
count	404544	404544	404544
mean	0.442189	0.171718	0.536186
std	0.185740	0.233802	0.191387
min	0.000000	0.000000	0.000000
25%	0.324077	0.000000	0.395069
50%	0.435326	0.000000	0.621671
75%	0.568107	0.490000	0.673566
max	1.000000	1.000000	1.000000

Figure 1.4: Histogram of the regressed Variables

2 Appendix

Appendix 2.A Dataset

2.A.1 Patch Area

The area of each patch can be calculated from the shape data. There is only to consider to use an appropriate *EPSG* code for the *coordinate reference system (CRS)*. The graphics in this paper all show the patches with *EPSG* 4326, that is the style known from familiar world maps. But due to earth curvature, the distance between two longitudes is not the same at different latitudes, therefore it is important to use a Mercator-like projection (*EPSG* 32633) when calculating the areas.

2.A.2 North Atlantic Oscillation

The *North Atlantic Oscillation (NAO)* is a weather index that summarises the anomaly of atmospheric pressure above the North Atlantic [Barnston and Livezey (1987)], more precisely the difference between the atmospheric pressure of Iceland and the Azores at sea level. While in meteorology *NAO* is primarily used to predict the weather in Europe and North America, it is used in this paper as an indicator of stormy weather - a variable that has an impact on fishing decisions and landing success. Data of the monthly mean value of the *NAO* is retrieved from NOAA/NationalWeatherService (2019) and integrated into the dataset. Even though this is a coarse variable not specified on the location of the patches, it indicates the general naval conditions. Papers such as Gutiérrez-Estrada et al. (2009) use a similar index when predicting landings in the pacific.

2.A.3 Sea Surface Temperature

To obtain data for the local *Sea Surface Temperature (SST)*, two data sources are used, both based on the *Advanced Microwave Scanning Radiometer-EOS (AMSR-E)*. The older dataset relies completely on *AMSR-E*, while the newer one uses additional sources (AVHRR-3, GOES-11). To translate the raw satellite data into the *SST*, *AMSR-E* uses observations from a range of five frequencies (between 7 and 37 GHz) [Wentz and Meissner (2007)]. The method generally contains a model of atmospheric processes which get Monte-Carlo-simulated and trained by shore/ship-based *SST* measurements; as described by Wentz and Meissner (2000).
From 2002 the Wentz (2002) dataset with a resolution of one degree is used, the Naval Oceanographic Office (2008) from 2009 on has a spatial resolution of 0.1 degrees. The *SST* from each month middle and from the grid point which is closest to each patch is integrated into the dataset. As the variable ranges between 271.15 and 293.85 Kelvin, it is advisable to regress it only *normalised*.

2.A.4 Distances

The most elaborative adjustment of the dataset is the computation of the distances. In the original dataset contains the coordinates of each catch as a point in the middle of the reported patch. Rather than the landing position, the dataset provides information about the landing station, landing municipal and landing region. The stations are managed by the *Norwegian Food Safety Authority*, which only published the list of recently active stations, thus the list of station addresses is incomplete. Instead, the chosen landing coordinates are extracted from the *OpenCage Geocoder API*, returning a point in the middle of each municipal. The little inaccuracy seems bearable, as the patches also have a variance of an arcminute.
An issue to mention is that vessels are not bound to a particular station. Therefore, modelling predictions for the same vessel and patch at different points in time may have different distances as an input. Fortunately, two facts hamper that problem. First, vessels that operate costal are generally faithful to one location, while almost all vessels that change their landing location operate in pelagic patches. This sorting, in conjunction

with the geographical characteristics of Norway, makes the distance still a useful indicator. The Norwegian coast can be roughly seen as an axis from north to south, where all possible fishing grounds lie west of. So for long distances, far from shore, the distance between stations has less impact.

The distance between two points on the surface of a globe can not be directly extracted from their coordinates as the width of one longitude is diminishing when the latitude is higher. For the actual computation of the distances the so-called *haversine* equation from Inman (1849) is applied:

$$d = 2r\ arcsin(\sqrt{sin^2(\frac{\varphi_2 - \varphi_1}{2}) + cos(\varphi_1)cos(\varphi_2))sin^2(\frac{\lambda_2 - \lambda_1}{2})}$$

$\varphi_1, \varphi_2:$ *latitude*

$\lambda 1, \lambda 2:$ *longitude*

Bibliography

Barnston, A. G. and Livezey, R. E. (1987). Classification, seasonality and persistence of low-frequency atmospheric circulation patterns. *Monthly weather review*, 115(6):1083–1126.

Fiskeridirektoratet, N. (2000-2015). Åpne data: fangstdata koblet med fartøydata. *https://fiskeridir.no/Statistikk/AApne-data/AApne-datasett/Fangstdata-koblet-med-fartoeydata*. retrieved: 12-01-2018.

Fiskeridirektoratet, N. (218). Lokasjoner t.o.m. 2017. *http://opendatafiskeridirektoratet-testkommune.hub.arcgis.com/datasets/lokasjoner-t-o-m-2017*. retrieved: 13-01-2020.

Gutiérrez-Estrada, J. C., Yáñez, E., Pulido-Calvo, I., Silva, C., Plaza, F., and Bórquez, C. (2009). Pacific sardine (sardinops sagax, jenyns 1842) landings prediction. a neural network ecosystemic approach. *Fisheries Research*, 100(2):116–125.

Inman, J. (1849). *Navigation and Nautical Astronomy, for the Use of British Seamen*. F. & J. Rivington.

Naval Oceanographic Office, P. (2008). Ghrsst level 4 k10 sst global 1 meter sea surface temperature analysis. $http : //podaac.jpl.nasa.gov/dataset/NAVO - L4HR1m - GLOB - K10_SST$. retrieved: 08-01-2019.

NOAA (2019). National oceanic and atmospheric administration - north atlantic topological data. *https://maps.ngdc.noaa.gov/viewers/wcs-client/*. retrieved: 14-11-2019.

NOAA/NationalWeatherService (1950-2019). Monthly mean nao index since january 1950. *https://www.cpc.ncep.noaa.gov/products/precip/CWlink/pna/nao.shtml.* retrieved: 07-31-2019.

Wentz, F. and Meissner, T. (2000). Amsr ocean algorithm theoretical basis document, version 2, report, remote sens. *Syst., Santa Rosa, Calif.*

Wentz, F. and Meissner, T. (2007). Algorithm theoretical basis document: Amsr-e ocean algorithms supplement 1.

Wentz, F.J., T. M. C. G. M. (2002). Ghrsst level 3u global subskin sea surface temperature from the advanced scanning microwave radiometer - earth observing system (amsr-e) on the nasa aqua satellite. *https://doi.org/10.5067/ghams-3gr7a.* retrieved: 08-01-2019.