Collective Anomaly Detection in Fisheries

Bjørnar Tessem[1][0000−0003−2623−2689] and Aida Ashrafi[1][0009−0000−9117−592X]

Dept. of Information Science and Media Studies, University of Bergen, Norway
{bjornar.tessem,aida.ashrafi}@uib.no

Abstract. One of the tasks of the Norwegian Directorate of Fisheries is surveillance of ocean fisheries in Norwegian waters. To discourage illegal, unreported and unregulated fisheries, they collect various types of data about fishing activities, in particular data about each catch operation. However, catch data from fishing activities are by nature unpredictable, and many factors may be causes of variation. This makes it hard to identify single catch operations that are anomalous and perhaps incorrectly reported. In this paper we show how we can use the concept of collective anomalies by looking at collections of catch operation reports and check how they deviate from the expected. We do this by running a machine learning model to predict total catches of trawlers' catch operations, compute the prediction errors from the model, and see how the prediction error distribution of a vessel deviates from the whole set of catch reports. The experiments are promising and we are able to identify deviating vessels in a consistent manner, but the outcomes still need to be evaluated by domain experts.

Keywords: Collective anomaly detection · Sustainability · Fisheries · Catch prediction.

1 Introduction

The sustainability of fisheries as well as biological diversity under water is continuously threatened by illegal, unreported and unregulated (IUU) fishing activities. According to the Food and Agriculture Organization of the United Nations (FAO) this amounts to up to 26 million tonnes of catch every year[7]. This represents about 30% of the total yearly catch of about 90 million tonnes of wild fish captures [12]. As it has been documented that more than 30% of the fish stocks are overexploited [10], we are in danger of destroying one of the main food resources of humans. An example of the serious effect of overfishing were experienced in the early 1990-ies when the cod population around Newfoundland [14] almost completely disappeared.

Fisheries data with high quality and machine learning methods may provide tools to support authorities in their efforts to uncover IUU fishing. The Norwegian Directorate of Fisheries (DoF) has for almost two decades collected daily catch reports from fishing vessels above 15 meters and has also recently included smaller vessels. These are data where vessels report activities when at sea, and in particular data about single fishing activities including location, species caught, and catch size[6].

One tool for reducing IUU could be to uncover unreported fishing by detecting potential deviations in the catch reports, i.e., so called anomaly detection. A first attempt would be look for anomalies in single catches, using unsupervised methods. However, due to the randomness in the amount of fish caught, it is hard to claim that a single activity is a deviation even if being in the outskirts of the total set of reports [3]. Fishing activities under similar conditions may lead to almost no catch, but also catches in ten thousands of kilos.

For this reason we are here exploring the possibility to use collective anomalies to identify vessels that have an overall deviating behaviour compared to the rest. Collective anomalies are observed when related data instances as a collection is anomalous with respect to the entire data set [5]. In our case we consider collections of catch reports from single vessels. The assumption is that vessels that have a IUU fishing behaviour over time will show a deviating reporting practice, even though their single catch reports may be well within what is expected. We combine a machine learning model for predicting total catch, and then use statistical properties of errors in the predictions to identify vessels that have significantly different distributions in errors than the population in total.

We continue the paper with an explanation of collective anomaly detection, continue with a description of the data, our selection of data and the preprocessing steps needed to get our approach working. After having selected a machine learning model to use for catch predictions, we describe how we have used statistical tests to identify candidate deviating vessels. We conclude with a discussion of the validity of the approach and opportunities for future work.

2 Anomaly detection

Anomaly detection is a field of statistics which has the purpose of detecting data points in a data set that for some reason deviates from the statistical distribution of the data [5]. A kind of anomaly detection may be to identify values on single data features that are outside of the acceptable value set for a feature, often named outliers in practical machine learning. But we may also have an anomaly when a data point have a combination of feature values that are not possible or not within the combined statistical distribution. Identifying such data points is more challenging as it may be unclear when a data point is outside a distribution.

Supervised machine learning approaches can be used to train models for detecting anomalies, but are then dependent on having labeled data sets. In practice, it is difficult to get sufficiently large data sets in many domains, as the domain experts have to prioritise other work than labeling for machine learning, and it may even for them be hard to classify a data points as an anomaly. Anomalies also tend to be rare, which has the implication that data sets may become highly unbalanced when labeled,

The lack of labels is the main motivation for using unsupervised techniques for identifying potential anomalies in a data set. One of the strategies used there is to create a lower-dimensional representation of the data points and then run a reconstruction approach. Large error in the reconstructed data point indicates that the data point is not a result of a normal data point generation. This can for example be done by auto-encoder/decoder approaches [13]. This approach was explored without much success on fisheries data by Ashrafi et al. [3] using a relational autoencoder (RAE) [9], showing that real world practical use of such methods may be challenging. An alternative approach is to apply clustering methods on the data set, and identify the data points that are outside the identified clusters or far away from the centre of a cluster [1]. Ashrafi et al. [3] also explored this approach on fisheries data, and were able to identify some anomalies. However, most of these would also have been possible to find using simple rule based approaches.

2.1 Collective anomaly detection

In many situations, all data points in a collection may be perfectly acceptable, and none can be considered an anomaly. However, there may still be situations where a collection of data points, selected according to some feature value or defined group, has values outside of what we should expect. Such a deviation from the rest of the data is termed a collective anomaly [5]. In a more thorough survey these authors group collective anomalies into three groups, collective anomalies in sequences, geospatial data, and graphs [4]. Recent examples include Li et al. [8] who use deep learning on labeled sequence data, and Viale et al. [15] who look at unsupervised anomaly detection. Qin et al. [11] use statistical properties of reconstruction loss as anomaly measures.

The data set we are working with is not labeled for anomalies. It does also not fit well the three categories Chandola et al [4] describe. Our approach has been to train a supervised machine learning model with fish catch weight as targets on tabular data. The aim has to identify vessels whose reports are deviating from the rest of the data, but the approach can also be used to identify deviation in particular time periods or locations. We calculate individual prediction errors on each data point, and use the statistical distribution of these errors from the whole data set as an expected prediction error distribution. Any selection of collected data that does not have a compatible distribution is considered a potential anomaly.

3 Daily catch reports

The Directorate of Fisheries has several open data sources [6]. The data we are exploring here is so called Daily Catch Activity (DCA) messages from fishing vessels. They report daily about any fishing operations they have done, or if they are not fishing, about their movement (steaming) or other not-fishing activities they are doing at sea. For us it is the fishing operations that are of interest, and we have decided to focus on operations performed with bottom trawls as gear type. For this exploratory analysis we have chosen to work with data from 2018.

The data consist about information about when a fishing operation message is sent, information about when and where the fishing operations started, as well as when it ended. Further, duration of operation and movement distance during operation is added. There is information about the catch of the operation. This includes information about the main species caught, as well as other species caught. The other species than the main species is often termed by-catch. For each species there is information about the total weight (round weight) in kilos. Finally there is information about the vessel, its call signal (unique identifier), name, and various information about its build, including length, width, and gross tonnage (internal volume).

The data is represented in a large .csv file, and consists of large amounts redundant information as they provide various names and codes etc. for the same information. It must also be mentioned that each fishing message may consist of several fish operations with different start and end point. Data for each vessel and message time feature is thus repeated for each of these. In addition each operation again consists of one data row (a line) for each species caught. Each of these lines contains identical information for every feature except for the species and its round weight.

For he purpose of this study we reduce the data so that each row in our prepared dataset for bottom trawling activities in 2018 consists of the following features:

Message ID unique code for each message
Call signal unique ID for vessel
Start time starting time of fishing operation
Start position latitude latitude where the operation starts
Start position longitude longitude where the operation starts
Sea depth, start sea depth where the operation starts
Stop time stopping time of fishing operation
Stop position latitude latitude where the operation stops
Stop position longitude longitude where the operation stops
Sea depth, stop sea depth where the operation stops
Duration length in minutes of the operation
Drag distance distance moved during operation
Main species FAO (code) the main species caught in the operation using FAO[1] code
Vessel length the length of the vessel
Total round weight sum of the weights of the species reproted in the catch

This provides us with data about 33,385 catch operations now prepared for machine learning.

4 Collective anomalies from prediction errors

Fishing operations are highly random when it comes to the total catch. There are many factors that may influence the total catch beyond the features of our data

[1] https://www.fao.org/home/en

set. These include fish migrations, weather, bottom conditions, recent catches at same location etc. A very large span of outcomes in terms of catch may be considered inside what one could expect. In a previous work by Ashrafi et al. on similar data, approaches like reconstruction error on single data points and clustering [3] were used to suggest anomalous catch operations. However, they were mainly able to identify operations that has some kind of errors in data registration. Most of them could have been discovered by a rule based approach.

The randomness of fisheries suggest that for finding anomalies we could investigate a collective anomalies approach. The idea is that even though a fishing vessel may have catch reports that all are acceptable according to models prepared from the data, the vessel's overall behaviour may be different from what one would expect. The most critical feature of our prepared data set in relation to sustainability is of course the total round weight. In conclusion our approach is to construct a machine learning model predicting total catch, use this for prediction, and then test if the prediction errors are acceptable. The steps are as follows:

1. Apply a standard supervised machine learning model to the prepared data. Ensure that data are further adapted to the constraints of the particular model.
2. Run a prediction of total catches on the whole data set. For each data point compute the distance between the predicted catch and the reported catch as the prediction error.
3. For collections of data points test if the statistical properties of the prediction error distribution is compatible with the full distribution. Apply a suitable statistical test for this purpose. If the collection's distribution is significantly different from the full distribution, we have identified a potential collective anomaly

4.1 Choice of machine learning approach

The machine learning approaches we have selected for testing are k-nearest neighbour, standard linear regression, decision tree, multilevel perceptron, random forest, a standard gradient boosting, and the xgboost model. All are provided for use with the scikit-learn machine learning library[2]. We did some further preparation of the data as below:

Main species There is more than 20 main species in the prepared data, many of them with few occurences. We chose to focus on the six most common main species and represented those with one-hot-encoding. Those species are cod, greenland halibut, haddock, pollock(saithe), northern prawn, and (beaked) redfish. All those were represented as main catch in more than 1,000 catch operations.

Time The starting and stopping time were represented by using trigonometric transformations which is common for cyclical features. This gave us four new features (sine and cosine values for the time stamps).

[2] https://scikit-learn.org/stable/

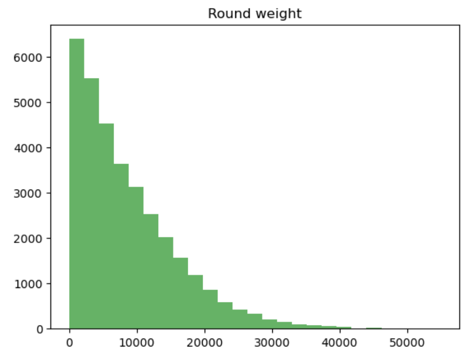

Fig. 1. The distribution of catch sizes. Mean = 8,698 kilos, standard deviation = 7,561 kilos

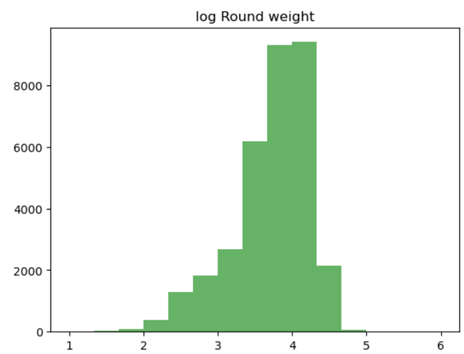

Fig. 2. The distribution of log10 of catch sizes. Mean = 3.722, standard deviation = 0.521

Total catch The total catch has an approximate log-normal (long tail) distribution (see Figure 1). At the same time, distance in catch sizes is dependent not only on the catch sizes them selves, but it needs to be considered in relation to the catch sizes. Hence it was seen as pertinent to use the 10-logarithm (log10) value of the catch size as the target value when assessing machine learning models. The distribution of log10 of catch sizes is given in Figure 2.

The remaining numerical features do not in principle need scaling to work with tested methods. We also chose to omit the vessel call signal from the input features, as we later grouped catches according to call signal. The final prepared data set amounts to 33,385 data points (catch operations) distributed on 66 vessels.

Table 4.1 summarize the performance of the seven tested standard models on the prepared data set. The numbers are average scores from a cross-validation process. We provide R^2-measures on training and test data sets. Xgboost proves itself a strong method giving high scores on both training and test, and not

Table 1. R^2 scores for a selection of standard machine learning model on the prepared data set.

Method	train	test
k-nearest neigbour	0.528	0.523
linear regression	0.557	0.422
multi-level perceptron	0.630	0.618
decision tree	1.000	0.472
random forest	**0.961**	**0.735**
gradient boosting	0.632	0.634
xgboost	**0.827**	**0.716**

over-fitting like decision tree and random forest. Random forest has the highest score on the test data, so we will also try that in the anomaly detection process. Training time were in matter of seconds for all except for random forest that took about a minute to train with 100 trees. As a conclusion we have chosen to try xgboost and random forest for the anomaly analysis. We also try the use of random forest on the catch data themselves as a target. The score of this model on a test set showed an R^2-score of 0.51, which is significantly worse than xgboost and random forest on log10-data, but acceptable for our exploratory investigation.

4.2 Statistical tests for deviating errors

For testing we decided to use a single sample t-test where test the prediction error for each vessel towards the mean of the error of the log10-value of all catches. This distribution of log10-error seems to be close to a normal distribution for all three models. The normal distribution is a requirement for the use of such a t-test. The concrete approach for testing error distribution consisted of these steps:

1. Prepare data for machine learning
2. Build models for
   (a) Xgboost with log10-data as target
   (b) Random forest with log10-data as target
   (c) Random forest with catch-data as target
3. Predict log10 of catches (directly or by log10-transformation) from the three models for all fishing operations
4. Compute prediction errors (prediction-reported) for each log10-catch
5. For each model and vessel perform a t-test on the prediction errors against the mean prediction error (for the particular model)
6. List vessels with $p < 0.01$ in one of the three tests

There may also be other indications from such predictions, for example the total amount of predicted catch compared to the reported catch. The prediction models on the log10-targets are, however, not performing well on the data along

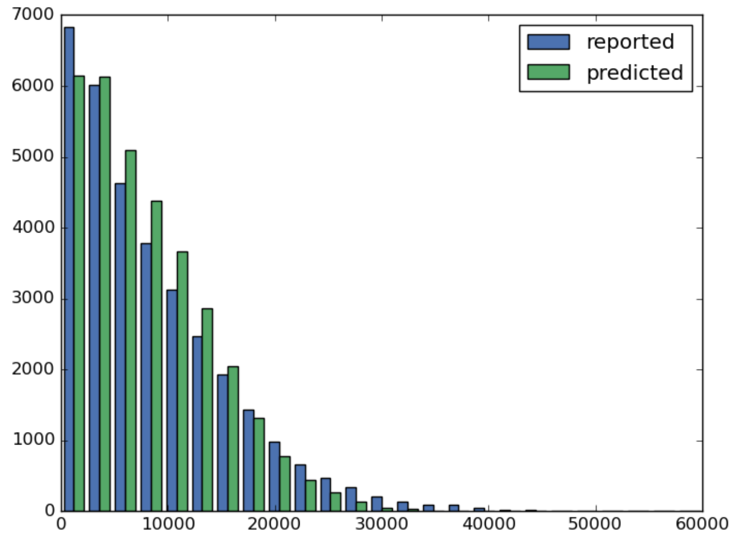

Fig. 3. The distribution of predicted and reported catch sizes on all data.

the long tail. The consequence is that they on average predict less catches than the reported. The ratio between the sum of the predicted catches and the sum of reported catches of a vessel is therefore larger than 1.0 for these two models. For the Xgboost model we computed this ratio (here named rp-ratio) to 1.16 and for the Random forest model to 1.08. If a vessel's reported catch/predicted catch ratio is much less than this ratio, then that is an additional indication that a vessel has an other distribution of catch sizes than the whole population. For the random forest on total catch model the rp-ratio is 0.99.

In table 4.2 we see the vessels that scored with $p < 0.01$ on at least one of the three models. The table includes 29 (anonymised) out of 66 vessels that reported bottom trawl catches in 2018. Only one of the vessels (LDEF) reports significantly less catches than predicted on all models, and also reports less total catch than predicted (rp-ratio). Three vessels (LSTV, LVWX, LWXY) report larger catches than predicted for all models. For two of those (LVWX, LWXY) all rp-ratios are consistent with the t-test outcomes.

To illustrate how predicted and reported catch distributions differ we have transformed the predicted log10 values from the random forest model to catch numbers. Further, we have plotted a histogram showing the distribution of predicted and reported catches of different sizes for various data sets. Figure 3 shows the distribution for all catch reports. The prediction misses by predicting fewer on the large values (the long tail) and predicting fewer also on the smallest catches. It also shows that there is more catch predictions than catch reports in the middle catch size groups (about 2.000 to 15.000 kg).

Figures 4, 5, and 6 show the distribution for three different vessels. LCBA (Figure 4) is a vessel that has the expected distribution, and follows neatly

Table 2. Scores for collective anomaly detection

		Xgboost			Random forest			Random forest catch		
Call sign	N	stat	p	rp-ratio	stat	p	rp-ratio	stat	p	rp-ratio
LABC	563	1.391	0.165	1.153	1.581	0.114	1.086	2.775	0.006	0.972
LBCD	966	-0.697	0.486	1.156	-0.029	0.977	1.066	-2.669	0.008	0.986
LCDE	1016	0.603	0.547	1.252	1.055	0.292	1.145	7.248	0.000	0.969
LDEF	**945**	**6.785**	**0.000**	**1.044**	**6.155**	**0.000**	**1.038**	**4.078**	**0.000**	**0.947**
LFGH	833	0.023	0.982	1.205	-0.297	0.766	1.106	3.447	0.001	0.976
LGHI	769	-0.845	0.398	1.137	-2.643	0.008	1.081	-3.063	0.002	1.001
LHIJ	791	1.917	0.056	1.148	5.084	0.000	1.050	4.821	0.000	0.951
LIJK	1029	-0.148	0.882	1.151	-0.466	0.641	1.079	-3.084	0.002	0.999
LJKL	1045	-2.466	0.014	1.147	-3.733	0.000	1.086	-8.587	0.000	1.011
LKLM	515	-0.790	0.430	1.116	-1.389	0.165	1.065	-5.351	0.000	0.996
LLMN	261	-0.673	0.502	1.087	-0.319	0.750	1.043	-3.798	0.000	1.005
LMNO	167	-0.842	0.401	1.090	0.368	0.714	1.031	-2.648	0.009	0.968
LNOP	225	-0.171	0.864	1.077	0.156	0.876	1.037	-4.078	0.000	0.982
LOPQ	281	0.896	0.371	1.062	0.129	0.897	1.043	-5.217	0.000	0.991
LPQR	905	0.418	0.676	1.170	2.936	0.003	1.061	0.584	0.559	0.973
LQRS	58	2.860	0.006	1.013	1.379	0.173	1.050	1.185	0.241	0.955
LRST	22	2.283	0.033	0.928	2.981	0.007	0.953	1.855	0.078	0.870
LSTU	109	-3.590	0.000	1.139	-2.861	0.005	1.060	-7.744	0.000	1.018
LTUV	319	0.877	0.381	1.259	-0.006	0.995	1.184	3.481	0.001	0.987
LUVW	1106	1.442	0.150	1.157	2.834	0.005	1.075	4.855	0.000	0.968
LVWX	**967**	**-3.820**	**0.000**	**1.160**	**-5.589**	**0.000**	**1.097**	**-7.324**	**0.000**	**1.018**
LWXY	**919**	**-7.897**	**0.000**	**1.273**	**-5.159**	**0.000**	**1.093**	**-7.345**	**0.000**	**1.011**
LXYZ	824	1.405	0.160	1.112	0.425	0.671	1.058	-4.102	0.000	0.984
L1AB	922	2.588	0.010	1.127	-0.665	0.506	1.085	-4.055	0.000	0.999
L2CD	253	-0.600	0.549	1.102	-1.314	0.190	1.052	-3.731	0.000	0.994
L3EF	920	-1.776	0.076	1.151	-2.892	0.004	1.076	-7.758	0.000	1.005
L4GH	906	-1.420	0.156	1.143	-2.676	0.008	1.084	-2.908	0.004	0.998
L5IJ	243	-0.149	0.882	1.116	-0.897	0.371	1.072	-2.774	0.006	0.984
L6KL	785	-0.250	0.803	1.288	0.580	0.562	1.092	3.150	0.002	0.994

the distribution we see in figure 3. LWXY (Figure 5) has a distribution less similar to the distribution of all vessels, with two peaks. This vessel has relatively much more catches above the prediction than the others, and reports more large catches, i.e. above 20.000 kg. LDEF (Figure 6) has a distribution with a peak around 1.000 kg and one around 12.000 kg. This vessel reports fewer large catches, and more in the middle.

These visual presentations of the distributions may have limited value as an explanation themselves for the deviation in prediction error. However, they do show how predicted and reported catches varies in a different way.

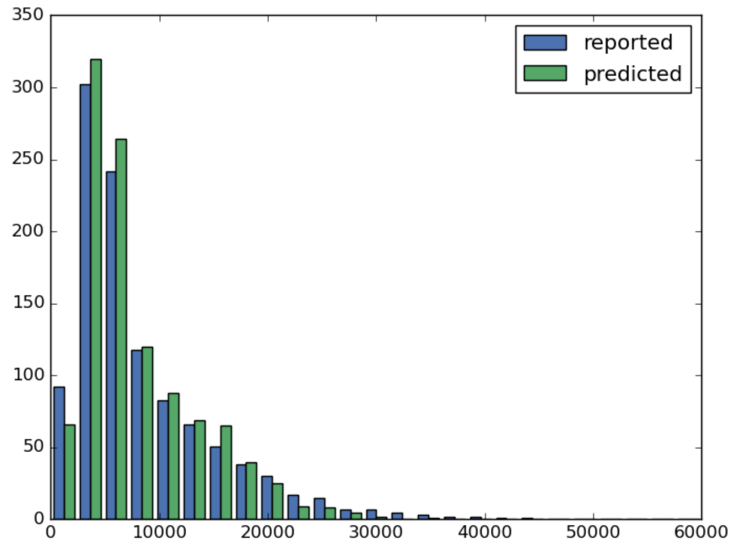

Fig. 4. The distribution of predicted and reported catch sizes for vessel LCBA.

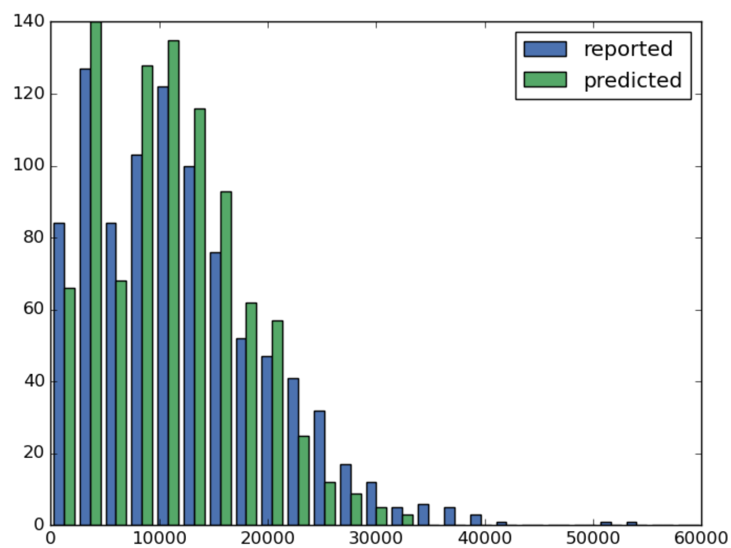

Fig. 5. The distribution of predicted and reported catch sizes for vessel LWXY.

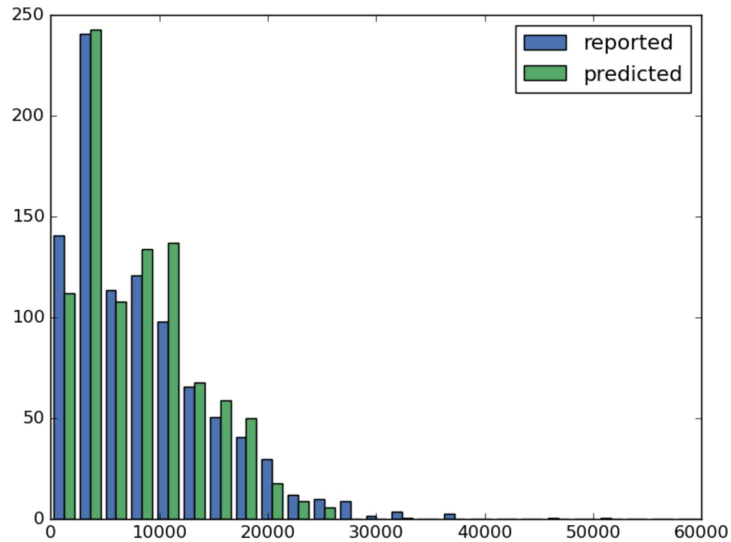

Fig. 6. The distribution of predicted and reported catch sizes for vessel LDEF.

5 Discussion

In the process describes above we found 3 candidate vessels that has provides potentially anomalous data. The reasons for the outcomes showed in the previous section may be the vessel crew's ability as fishermen, negligence when entering data into reports, and in worst case intended irregular reporting. Also the models may not be able to catch the different outcomes that comes from going for different species in different areas. The experts in the Directorate of Fisheries would be the right persons to assess such outcomes, and decide if there is need for some kind of follow up activities.

In general, anomaly detection is a difficult issue in many domains, and also here. We have not yet evaluated the approach with fisheries experts, and confirming a collective anomaly would need to be verified through some kind of additional investigation. Still, it may valuable for busy domain experts to get help from such models. They can identify vessels that need attention in the future. If one also runs anomaly tests on area, time periods or combinations one may identify who, when, and where inspections may take place.

The test for statistical deviation was a simple t-test for testing the mean value for a data set vs. a specified value, here the average error on the whole data set. t-tests depend on a symmetric distribution, and the prediction errors were not entirely symmetric. This is because the reported catch distributions were not perfectly log-normal. We have, however, assumed that the distribution was close enough to symmetric for our purposes. There may be better statistical tests that would fit the distributions we have here. However, the simplicity of the t-test makes it a very good candidate.

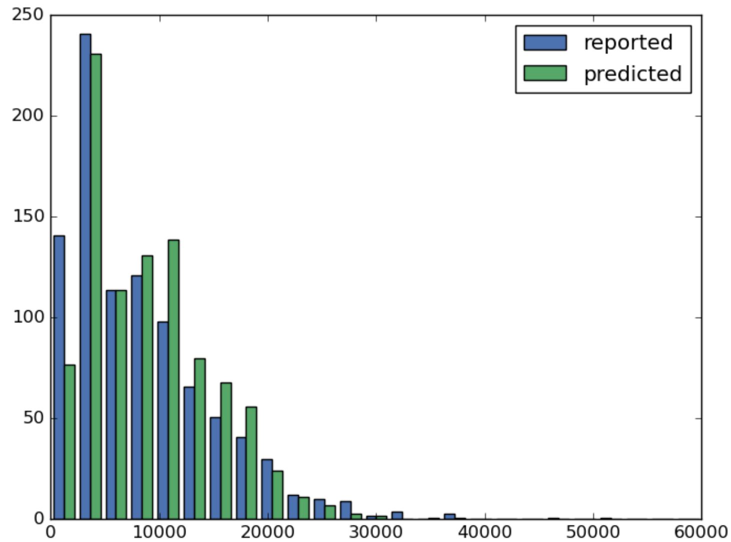

Fig. 7. The distribution of predicted and reported catch sizes for vessel LDEF based on a random forest model trained on non-transformed catch data.

One could also run two-sample t-tests that compared not the average, but the general distribution of two data samples, i.e., the complete sample and the vessel sample. We did some experiments with this, but got mainly the same results for these data as when using the one sample t-test.

One may argue that the log10 transformation will change the data and introduce potential anomalies. This is why we also tried a random forest model on the non-transformed data. The LDEF vessel was seen to have a significantly different error distribution also with this model, and the predicted vs reported catch distribution is very close to the one for the model made for the transformed data (see Figure 7 and compare with Figure 6).

6 Conclusions and further work

Uncovering anomalies in fisheries data is difficult, since fisheries inherently have much randomness. We have tried to identify collective anomalies by testing the distribution of the prediction error from three machine learning models. The results of the approach on trawlers from 2018 seem promising. We identified three vessels out of 66 that had prediction error distributions significantly different from the full distribution. These could be suggested as vessels with deviating behaviour.

The results and evaluation so far is mainly a proof-of-concept evaluation, as they are not assessed by potential users in the Directorate of Fisheries. Running evaluations with domain experts will be a task for future research. We will also in the future experiment with alternative machine learning models. Further, predictions of other targets from the data set is highly relevant for these data, for example, catches from selected species, by-catches, catches with other fishing gear, etc.

Finally, we have previously developed machine learning models that detect discrepancies in reporting about when fishing is going on [2]. Combining collective anomaly detection and such report discrepancies may provide experts with new tools for identifying vessels to focus on in surveillance. Visualisations that support understanding of the machine learning models' outcomes is an obvious way to go.

References

1. Ahmed, M., Naser Mahmood, A., Hu, J.: A survey of network anomaly detection techniques. Journal of Network and Computer Applications **60**, 19–31 (Jan 2016). https://doi.org/10.1016/j.jnca.2015.11.016, https://www.sciencedirect.com/science/article/pii/S1084804515002891
2. Ashrafi, A., Tessem, B., Enberg, K.: Detection of Fishing Activities from Vessel Trajectories. In: Nurcan, S., Opdahl, A.L., Mouratidis, H., Tsohou, A. (eds.) Research Challenges in Information Science: Information Science and the Connected World. pp. 105–120. Springer Nature Switzerland, Cham (2023). https://doi.org/10.1007/978-3-031-33080-3_7
3. Ashrafi, A., Tessem, B., Enberg, K.: Analysing Unlabeled Data with Randomness and Noise: The Case of Fishery Catch Reports. Swedish Artificial Intelligence Society pp. 204–212 (Jun 2024). https://doi.org/10.3384/ecp208023, https://ecp.ep.liu.se/index.php/sais/article/view/1015
4. Chandola, V., Banerjee, A., Kumar, V.: Anomaly Detection: A Survey. Tech. rep., University of Minnesota Twin Cities (Aug 2007), https://hdl.handle.net/11299/215731
5. Chandola, V., Banerjee, A., Kumar, V.: Anomaly detection: A survey. ACM Computing Surveys **41**(3), 1–58 (Jul 2009). https://doi.org/10.1145/1541880.1541882, https://dl.acm.org/doi/10.1145/1541880.1541882
6. Directorate of Fisheries: Electronic Reporting Systems, https://www.fiskeridir.no/English/Fisheries/Electronic-Reporting-Systems
7. Food and Agriculture Organization of the United Nations: Illegal, Unreported and Unregulated (IUU) fishing, https://www.fao.org/iuu-fishing/en/
8. Li, Z., Xiang, Z., Gong, W., Wang, H.: Unified model for collective and point anomaly detection using stacked temporal convolution networks. Applied Intelligence **52**(3), 3118–3131 (Feb 2022). https://doi.org/10.1007/s10489-021-02559-0, https://doi.org/10.1007/s10489-021-02559-0
9. Meng, Q., Catchpoole, D., Skillicorn, D., Kennedy, P.J.: Relational Autoencoder for Feature Extraction. In: 2017 International Joint Conference on Neural Networks (IJCNN). pp. 364–371 (May 2017). https://doi.org/10.1109/IJCNN.2017.7965877, http://arxiv.org/abs/1802.03145, arXiv:1802.03145 [cs, stat]
10. Our World in Data: Status of the world's fish stocks, https://ourworldindata.org/grapher/fish-stocks-within-sustainable-levels
11. Qin, H., Zhan, X., Zheng, Y.: Cscad: Correlation structure-based collective anomaly detection in complex system. IEEE Transac-

tions on Knowledge and Data Engineering **35**(5), 4634–4645 (2023). https://doi.org/10.1109/TKDE.2022.3154166

12. Ritchie, H.: The world now produces more seafood from fish farms than wild catch. Our World in Data (2019), https://ourworldindata.org/rise-of-aquaculture
13. Ruff, L., Kauffmann, J.R., Vandermeulen, R.A., Montavon, G., Samek, W., Kloft, M., Dietterich, T.G., Müller, K.R.: A unifying review of deep and shallow anomaly detection. Proceedings of the IEEE **109**(5), 756–795 (2021). https://doi.org/10.1109/JPROC.2021.3052449
14. Steele, D.H., Andersen, R., Green, J.M.: The Managed Commercial Annihilation of Northern Cod. Newfoundland & Labrador Studies **8**(1) (Jan 1992), https://journals.lib.unb.ca/index.php/NFLDS/article/view/919, number: 1
15. Vitale, F., De Vita, F., Mazzocca, N., Bruneo, D.: A Process Mining-based unsupervised Anomaly Detection technique for the Industrial Internet of Things. Internet of Things **24**, 100993 (Dec 2023). https://doi.org/10.1016/j.iot.2023.100993, https://www.sciencedirect.com/science/article/pii/S2542660523003165