WHO GOES THERE? EVALUATING THE PERFORMANCE OF A CNN FOR MONITORING DELPHINID PRESENCE TO THAT OF THE C-POD WITHIN DIVERSE MARINE SOUNDSCAPES

Ellen White, School of Ocean and Earth Science, University of Southampton, UK.
elw1d23@soton.ac.uk
Jonathan M. Bull, School of Ocean and Earth Science, University of Southampton, UK.
Paul R. White, Institute of Sound and Vibration, University of Southampton, UK.
Denise Risch, Marine Science Department, Scottish Association of Marine Science, Oban

1 CONTEXTUAL BACKGROUND

Passive acoustic monitoring (PAM) is a cost-effective method of gathering big data for monitoring the distribution of cetaceans[1,2], yielding important insights into occurrence, movements, and behaviour over space and time, for use in species conservation and habitat management[3,4,5,6,7]. Dolphins are an indicator species, common to the British Isles, with the status and health of regional populations acting as important metrics for conservation and in assessing ecosystem health[8,9]. Acoustically they emit a diverse range of signals, employed in hunting, navigation, and communication, consisting of highly directional transient signals (echolocation clicks) and tonal omni-directional, frequency modulated whistles[10,11]. The presence of these acoustic cues within broadband acoustic recordings can be used as a proxy for true animal presence[12], but manual extraction of signals is laborious and unrealistic on an appropriate timescale for conservation. Automation of the detection of these acoustic signals is therefore an essential component of the analysis pipeline for gathering data on delphinid occurrence.

The application of Deep Learning (DL) algorithms is increasingly prevalent within the field of bioacoustics. Convolutional Neural Networks (CNNs), which use image processing techniques to read and classify acoustic data into predetermined classes, have been shown to excel at detecting signals which vary in respect to time, frequency, and amplitude[13,14,15,16]. Using broadband acoustic data as input, CNNs have achieved high accuracies at the task of detecting dolphin vocalisations within variable marine environments[17,18,19,20]. The use of CNNs for monitoring long-term dolphin occurrence in diverse acoustic habitats, beyond proof of concept, is still in its infancy; for such a tool to become part of the standard analysis toolkit we must demonstrate its reliability compared to established methods of detection.

The C-POD logger (Cetacean – Porpoise Detector, Chelonia Ltd UK) has been used extensively for monitoring the distribution, regional density, and acoustic habits of small cetaceans[21,22,23,24,25], and is used as an integral tool in global monitoring programs for comparing species occurrence over time and space[26,27,28,29,30].The C-POD is a non-archival data logger, it does not store broadband acoustic data, instead autonomously detecting, via on-board processing, the presence of echolocation click trains (regularly spaced series of similar clicks) between 20 – 160 kHz. Proprietary software is provided which contains a custom classifier (KERNO) to classify click trains as originating from either dolphins, porpoises or other cetaceans based on their intensity, duration, frequency content and inter-click intervals (ICI)[31,32]. In this work we evaluate the performance of the C-POD compared to a newly available multi-sound source CNN[16] for the task of monitoring hourly dolphin presence in the waters off the west coast of Scotland (Figure 1), using PAM data collected within the COMPASS array (EU INTERREG COMPASS project). We present the first empirical comparison between a CNN and the C-POD for dolphin detection, highlighting the efficiency of each method as a monitoring tool in diverse soundscape conditions.

2 METHODS

2.1 Data Acquisition

Passive acoustic data was analysed from three locations from within the COMPASS array; Tolsta, Hyskier and Shiant Isles (Figure 1, Table 1). Each mooring consists of a single omnidirectional broadband acoustic recorder, the SoundTrap 300HF (Ocean Instruments Ltd, sensitivity -121 dB re 1V/μPa), located 5m above the seabed. Audio data is sampled at 96 kHz and recorded on a 20/40 minutes on/off duty cycle which commences on the hour. A C-POD logger (Chelonia Ltd) is positioned beneath the SoundTrap on the array (Figure 1b), recording data continuously.

The physical conditions and mooring depths at each location contribute to seasonal variation within the regional ambient soundscape (Figure 2). To evaluate detector performance in diverse acoustic fields eight data periods were identified: five one-week periods were selected as being likely to be representative of seasonality across the year; January 24th – 31st, April 1st – 7th, July 1st – 7th, September 1st – 7th and November 17th – 23rd 2019. Three periods associated with named storms were selected, corresponding to times when low-pressure resulted in elevated ambient noise conditions (Table 1), identified using the UKs MetOffice MIDAS archive weather records for 2019 (Met Office, 2022). The selected storms were Storm Gareth (March 11th – 14th 2019), Storm Hannah (April 25th – 28th 2019), and Storm Atiyah (December 7th – 10th 2019). Data gaps are present at individual sites due to mooring loss or recording failures (Table 1).

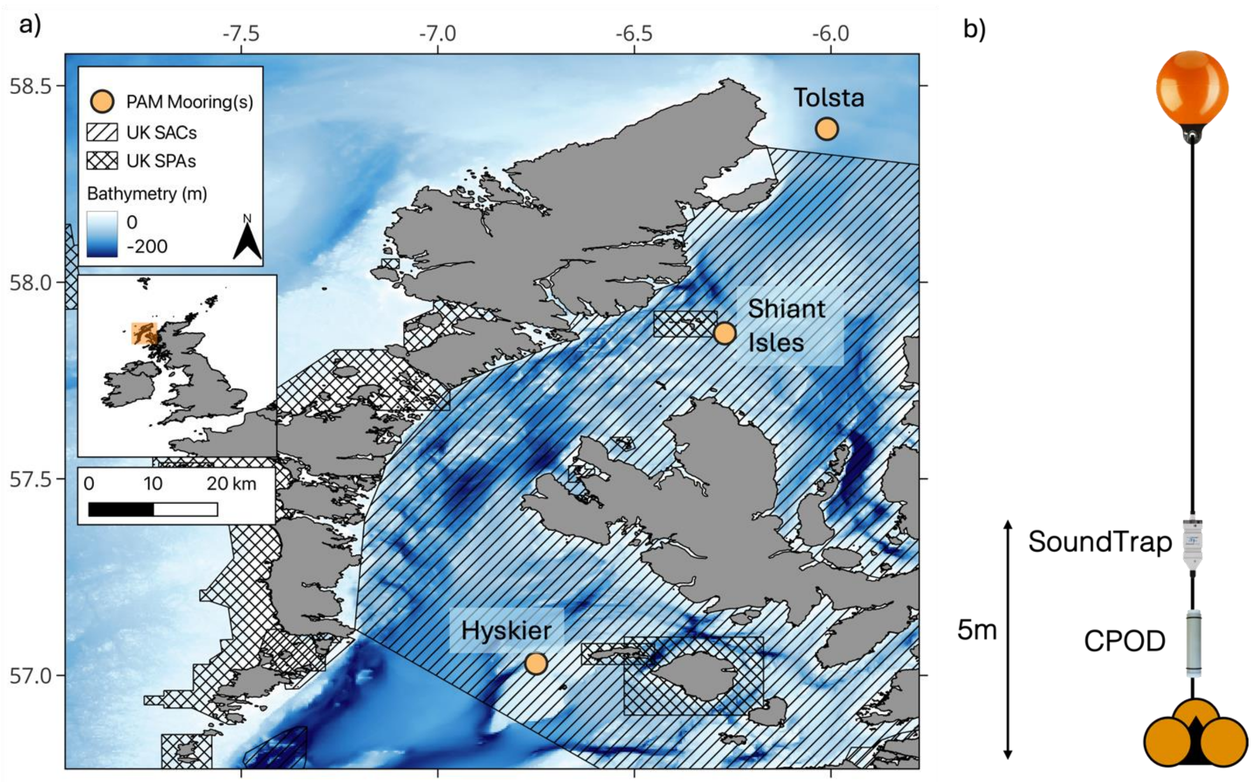

Figure 1 a) Map of the three acoustic mooring locations used within this work, located off of western Scotland as part of the COMPASS array. The regional bathymetry is depicted by the colour scale (m), sourced from GEBCO. b) Schematic depiction of each acoustic mooring consisting of a broadband recorder (SoundTrap 300HF, Ocean Instruments) and a C-POD logger positioned 5m above the seabed.

2.2 Analysis of Acoustic Data

This section describes how the number of detection positive hours (dph) were determined manually, from the CNN, and from the C-POD.

Manual detection

Each of the 20-minute broadband acoustic files were reviewed visually (spectrograms) and aurally in Audacity (version 3.0.02, 2021) to identify the hourly presence of delphinid click trains. Spectrograms were viewed consecutively in 10-second windows, with data displayed linearly between 0 – 48 kHz using a Hanning window of size 2048, and a dynamic range of 0 – (-80) dB. If the manual analysis detected a click train (three or more clicks in series) within the 20-minute period, then it was labelled as a dph (Figure 5.2).

A secondary label was assigned if whistles were detected during the manual review. If a file contained only whistles it would not be labelled a dph.

Table 1. Description of the acoustic data used within this work, recorded at three locations within the COMPASS array. The seasonal periods refer to: J - January 24th – 31st, A - April 1st – 7th, Ju - July 1st – 7th, S - September 1st – 7th and N - November 17th – 23rd 2019, as well as three storm periods: St. G – Storm Gareth, St. H – Storm Hannah, and St.A – Storm Atiyah.

Site	Latitude	Longitude	Seasonal Period	Analysis Hours
Tolsta Recorder depth: 100m	58.39ºN	-6.01 ºW	J / A / Ju / N St. G / St. H / St. A	984
Hyskier Recorder depth: 55m	57.03 ºN	-6.75 ºW	J / A / Ju / S / N St. G / St. H / St. A	1152
Shiant Isles Recorder depth: 73m	57.87 ºN	-6.27 ºW	J / A / Ju / S / N	864
			Total Analysis Hours	3000

CNN detection

The CNN was deployed upon the full set of broadband data to identify delphinid presence, the specific details of the network architecture can be found in White *et al*., 2022. Upon input the network splits each file into 3-second chunk. Each chunk is converted to an input image by computing three individual spectrograms using three FFT windowing sizes: 1024, 2048 and 4096. Each spectrogram uses a Hanning window with 50% overlap and the dynamic range is normalised between -80 – 0 dB. The spectrograms are stacked to form a three-channel matrix, and resized to 224x224 pixels, forming the RGB input image[16]. The network then classifies each chunk into one of four labels: Ambient noise, delphinid tonal (whistles), delphinid clicks (click trains) or vessel noise, with an output file listing the labels in temporal order, per audio file. To determine a dph the sum of delphinid click labels is computed per hour; if the sum is ≥1 a file is labelled a dph (Figure 5.2). No acoustic data used within this study was used to train the original CNN, and no re-training is conducted to adapt the network to any of the locations, representing 'unseen' environments to the CNN.

C-POD detection

The C-POD detection files were downloaded and analysed using C-POD.exe, proprietary software provided by Chelonia Ltd. The KERNO classifier, the base classifier provided for C-POD delphinid detection, was used to identify delphinid click trains. The software provides filter settings for the click detection of 'High', 'Moderate' or 'Low' quality trains[23]. Click trains of all three quality levels were accepted and any hour in which a click train was detected was labelled as being dph (Figure 5.2). We note, the C-POD continuously recorded for the full hour, in contrast to the broadband data which only captures the first 20-minutes of each hour.

2.3 Evaluation

To quantify the performance of each detection method Accuracy (*A*), Precision (*P*), Recall (*R*) and the D-1 score were computed (equations 1 – 4):

$$A = \frac{N_{TP} + N_{TN}}{Nall}, \quad (1)$$

$$P = \frac{N_{TP}}{N_{TP}+N_{FP}}, \quad (2)$$

$$R = \frac{N_{TP}}{N_{TP} + N_{FN}}, \quad (3)$$

$$F1 - Score = \frac{N_{TP}}{N_{TP}+\frac{1}{2}(N_{FP}+N_{FN})}. \quad (4)$$

Where N_{TP} is the number of true positives, correct classifications with respect to the manual validation set; N_{TN} is the number of true negatives, the number of true hours without an animal present; N_{FP} is the number of false positives, incorrect classifications with respect to the manual validation set; and N_{FN} is the number of false negatives, also described as the number of missed detections. The false-positive rate (FPR) and false-negative rate (FNR) are computed to determine the ratio of false-positives and missed detections output, respectively:

$$FPR = \frac{N_{FP}}{N_{FP} + N_{TN}}, \tag{5}$$

$$FNR = \frac{N_{FN}}{N_{FN} + N_{TP}}. \tag{6}$$

3 RESULTS

The overall performance of both the CNN and C-POD were evaluated with respect to the manual detections, for a total of 3000 hours of audio data (Table 1). Manual analysis of the 20-minute files reported a total of 452 dph at Tolsta, 313 dph at Hyskier and 298 dph at Shiant Isles (Table 2), with Tolsta reporting the highest dph for all analysis periods. Analysis of the acoustic data by the C-POD is rapid, taking approx. 1 hour to acquire the detection data. The CNN was run through the audio data in 17 hours, 1.5 hours per week of data, using MATLAB on a standard laptop. In contrast it took approximately one month to manually label periods of delphinid activity within the 3000 hours of audio data.

Table 2. The total true positive (N_{TP}) and true negative (N_{TN}) delphinid hours detected during the manual analysis of each temporal period and site.

Site		Overall	January	April	July	September	November	Gareth	Hannah	Atiyah
Tolsta	N_{TP}	452	122	33	56		116	33	22	70
	N_{TN}	532	70	135	112		52	63	74	26
Hyskier	N_{TP}	313	33	10	52	57	115	7	5	34
	N_{TN}	839	159	158	116	111	53	89	91	62
Shiant Isles	N_{TP}	298	39	43	46	93	77			
	N_{TN}	566	153	125	122	75	91			

The performance of the CNN and the C-POD was markedly different in comparison to the manually validated dolphin clicks, quantified as detection positive hours (dph). The CNN outperforms the C-POD for all seasonal periods, identifying 228 more true positive (TP) hours than the C-POD, accurately identifying 92% of the manually identified dph (Table 3). The C-POD has a lower overall FPR (0.02) compared the to the CNN (0.07), resulting in a higher precision score (Table 3). The overall accuracy (A) and recall (R) scores are higher for the CNN due to the C-PODs very high missed detection rate (FNR), Table 3. The quantity of missed detections is the most significant difference between the two algorithms with the CNN reporting higher recall scores across the three sites, for all seasonal periods (Table 3). There is a strong influence of seasonality and weather on the performance of each detection method (Figures 2 & 3). The C-POD struggles to detect delphinid click trains in the ambient soundscape during storm events, missing 67 – 93 % of the true delphinid positive hours, with FNRs ranging between 0.63 – 0.93 during the storm events, September and November compared to 0.27 – 0.43 for the CNN (Table 3). Overall the CNN consistently outperforms the C-POD across the three locations as an automated tool for delphinid echolocation click detection, with both platforms outputting low FPR.

The total delphinid present hours is computed in hourly bins, per site, per season, to reveal seasonal and diurnal patterns in animal activity as detected by each of the analysis methods (Figure 2). Detections output from the CNN strongly correlate to the temporal patterns described by the manual validation set, indicating delphinids are present within the waters around Tolsta for a greater proportion of time than the other two sites (Figure 2). Seasonally, delphinid presence is highest in November and January at Tolsta, November at Hyskier and September and November at Shiant Isles (Figure 2). Both the CNN and the manual labels determine a diurnal pattern in delphinid occurrence, with animals vocalising at higher rates between 16:00

and 04:00. Acoustic presence is lowest at around midday for all sites and seasons, except for July at Tolsta (Figure 2). The C-POD displays a weak correlation to the other methods of detection, under-representing delphinid presence temporally (Figure 2), failing to identify almost all periods of activity at Hyskier and during storm events (Figure 3). Periods of stormy weather indicate elevated ambient noise conditions decrease the detection range of the C-POD, the CNN also suffers reporting the highest missed detection rate during Storm Hannah, with a FNR of 0.55 (Table 3, Figure 3).

Table 3. Performance metrics for the CNN and C-POD for each analysis period, with respect to the manually validated data at Tolsta, Hyskier and Shiant Isles combined. Metrics reported are true-positives (TP), true-negatives (TN), false-negatives (FN), false-positives (FP), accuracy (A), precision (P), recall (R), F1-score, false-positive rate (FPR) and the false-negative rate (FNR).

	Period	**N_{TP}**	**N_{TN}**	**N_{FN}**	**N_{FP}**	**A**	**P**	**R**	**F1**	**FPR**	**FNR**
CNN	**Overall**	**978**	**1459**	**430**	**133**	**0.82**	**0.89**	**0.70**	**0.78**	**0.07**	**0.30**
	Jan	242	234	64	36	0.83	0.87	0.80	0.83	0.13	0.20
	Apr	133	291	51	29	0.84	0.82	0.73	0.77	0.09	0.27
	July	164	250	78	12	0.82	0.93	0.68	0.78	0.04	0.32
	Sep	90	157	65	24	0.74	0.79	0.58	0.67	0.13	0.42
	Nov	253	138	97	16	0.78	0.94	0.73	0.82	0.10	0.27
	St.G	25	145	15	7	0.89	0.78	0.62	0.70	0.04	0.37
	St.H	12	158	15	7	0.89	0.63	0.45	0.52	0.04	0.55
	St.A	59	86	45	2	0.76	0.97	0.57	0.72	0.02	0.43
CPOD	**Overall**	**750**	**1380**	**838**	**36**	**0.71**	**0.96**	**0.47**	**0.62**	**0.02**	**0.53**
	Jan	217	226	128	5	0.77	0.98	0.62	0.76	0.02	0.37
	Apr	131	290	78	5	0.84	0.96	0.63	0.76	0.01	0.37
	July	144	233	122	5	0.75	0.96	0.54	0.69	0.02	0.46
	Sep	77	124	131	4	0.59	0.95	0.37	0.53	0.03	0.63
	Nov	138	109	251	6	0.49	0.95	0.35	0.51	0.05	0.65
	St.G	7	151	33	1	0.82	0.87	0.18	0.29	0.01	0.82
	St.H	2	161	25	4	0.84	0.33	0.07	0.12	0.02	0.93
	St.A	34	86	70	2	0.62	0.94	0.32	0.48	0.02	0.67

4 DISCUSSION

Acoustically reliable methods of automated signal detection, robust to fluctuations in seasonal and regional ambient noise levels, are required to effectively monitor dolphin populations. Here we evaluate the capabilities of a multi-sound source CNN, trained to detect dolphin acoustic signals as well as anthropogenic sound sources across a wide frequency spectrum (0 – 48 kHz), to detect delphinid presence in broadband acoustic data in comparison to the C-POD. Both detectors are empirically evaluated with respect to a ground-truthed dataset to determine their efficiency as tools for long-term management and conservation of dolphins in Scottish waters. The CNN detect higher numbers of delphinid positive hours than the C-POD, for all analysis periods, with the degree of discrepancy between the two detectors varying seasonally. The CNN achieved high accuracy overall (0.82) which varied in diverse ambient noise environments (0.74 – 0.89), with a relatively short analysis time (1.5 hours per week) compared to manual labelling. Seasonal and diel patterns in animal presence were demonstrated by the CNN, with temporal patterns correlating to that of the manual analysis (Figure 3) highlighting the model as a reliable indicator of dolphin occurrence throughout the year, suitable for long-term monitoring within variable ambient noise conditions. We do note that the sensitivity of the CNN, albeit higher than the C-POD, did suffer during adverse weather periods and further work will be conducted to estimate the effect of ambient noise on the detection range of the CNN. Fine-tuning a CNN to the ambient soundscape in which it will be deployed can improve the performance of signal detection by a large degree[20]. The results of the CNN in this work on unseen data are positive and demonstrate a higher degree of accuracy than the C-POD with no re-training. To further increase performance, and reduce the number of false detections we recommend, where possible, fine-tuning the algorithm[20], on a subset of diverse ambient data frames extracted from the region of deployment.

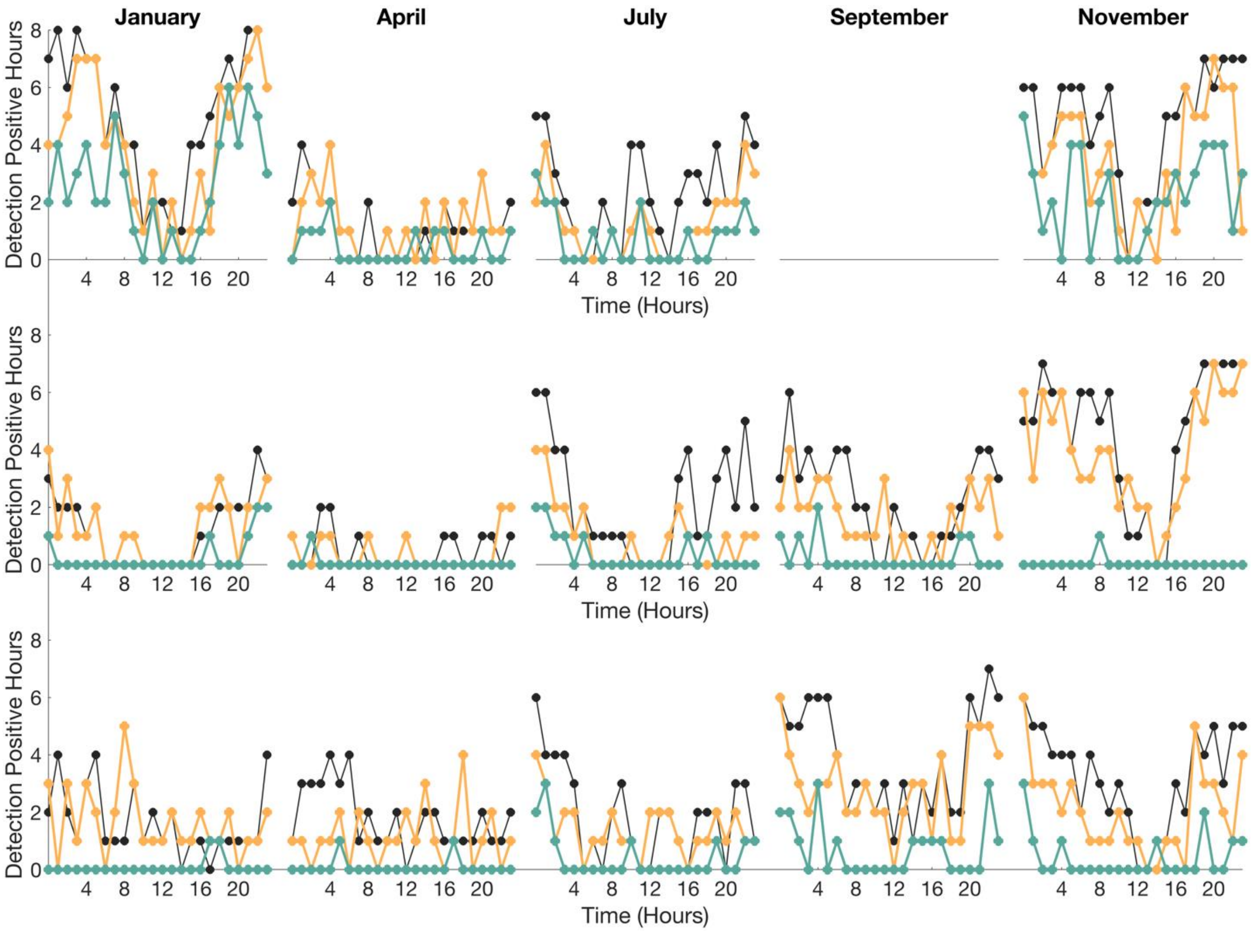

Figure 2. The total number of delphinid detections summed per hourly bins, per season, for the CNN (Orange line), C-POD (blue line) and manual labels (black line) at a) Tolsta, b) Hyskier and c) Shiant Isles. The CNN provides an accurate but conservative depiction of seasonal and diel patterns in dolphin occurrence, highlighting the advantages of its use as a monitoring tool. The C-POD fails to identify temporal variability, with a particularly poor performance at Hyskier.

There are inconsistencies between the results from the C-POD and the other detection methods, despite the systems being moored in proximity within the water column. Although we found the C-POD to output a very low rate of false detections, it failed to identify 36% of the manually labelled positive hours despite using the least conservative detection filters available. The experimental set up was that CNN operated on 20-minutes of acoustic data per hour due to the recorder duty cycle, but the C-POD operated on the full analysis hour. The broadband acoustic was set to duty cycle to achieve the same deployment longevity as the C-POD, which must be considered when selecting a suitable tool for long-term monitoring of dolphin presence. Given the extra temporal window to identify animal presence the rate of missed detections is concerning for robust monitoring of delphinid occurrence, particularly at our shallow site Hyskier where 93% of overall true detections were not detected. Our results are consistent with existing studies which have compared C-POD performance to other automated approaches (e.g., PAMGuard[33]), concluding that although the C-POD produces very low false-positives, it under reports delphinid presence[29,31]. Dolphin echolocation clicks occur between 20 – 100 kHz but research on C-POD detection ranges found the sensitivity of the logger to be weakest below 80 kHz[28]. The reduction in sensitivity coupled with adverse ambient noise conditions present a difficult operational environment for the C-POD. We must note that the

C-POD has been superseded by the F-POD (Full waveform capture POD)[30] and so the issue of missed detections may be reduced if studies can acquire the new system.

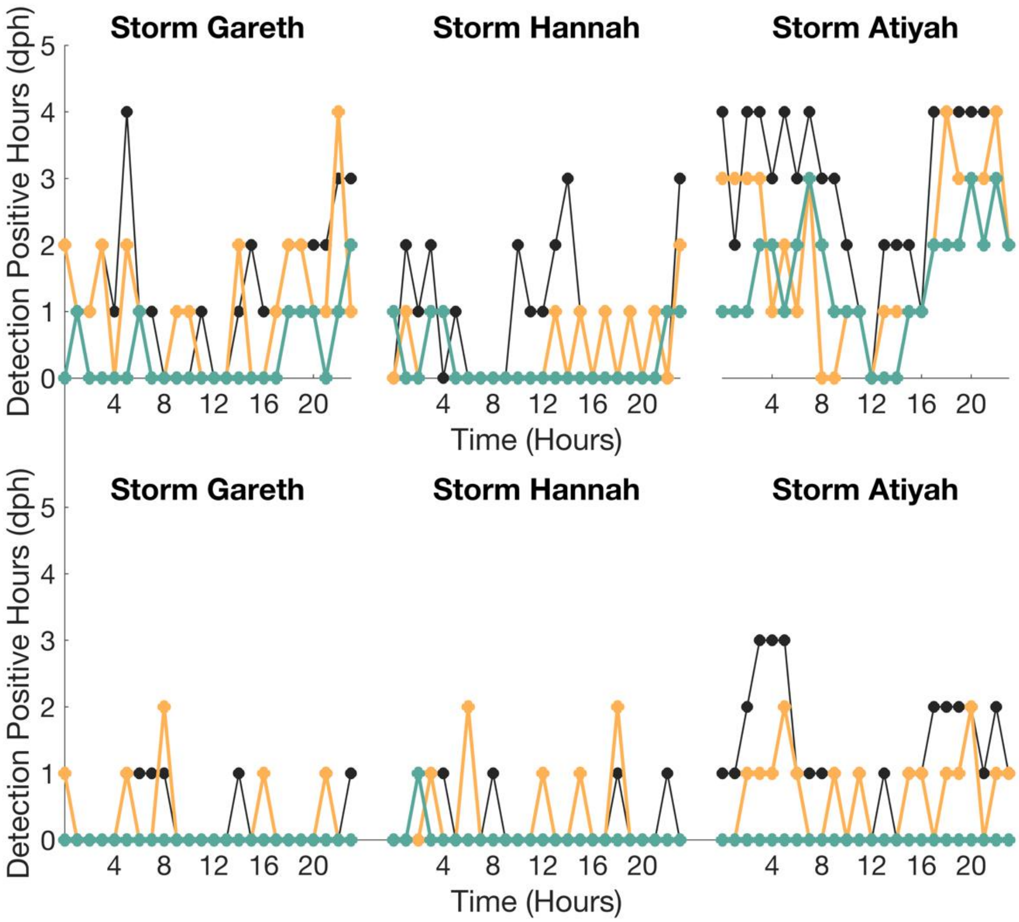

Figure 3. The total number of delphinid detections summed per hourly bins during each storm event, for the CNN (orange lines), C-POD (Blue lines) and manual labels (black lines) at a) Tolsta and b) Hyskier. The performance of the CNN and CPOD reduce during storm events, with the CPOD failing to identify all positive delphinid hours at Hyskier.

More work is necessary to determine why the difference between the two detection methods are so great. The ambient environment surrounding both detectors is the same, but the operation of the two methods are very different. The C-POD uses a zero-crossing algorithm[34], which may be more affected by ambient noise than the visual input to the CNN. An extension of this work will address detection on a finer-scale, computing detection positive minutes (dpm) per hour. The detections will be statistically correlated with ambient noise measurements to provide a clearer understanding of the environmental conditions which affect the performance of each method.

The dolphin populations on the west coast of Scotland are understudied with respect to other populations in the British Isles. Here we use the CNN to uncover seasonal trends in dolphin presence at each mooring. At Tolsta presence is high between November and January, dropping off during the summer seasons, but at Hyskier and Shiant Isles the dolphin ‘season’ appears earlier, with presence increasing from June through to November (Figure 2). This inherent site shift is interesting and further work can use the CNN to map distribution across the annual dataset and attempt to understand dolphin movements in the west of Scotland. With respect to diel presence our work illustrates echolocation click signals occur most frequently between sunrise and sunset (Figure 2). The echolocation click is used primarily for hunting and navigation, suggesting the locations are important foraging sites throughout specific seasons. The CNN used in this work is trained to detect dolphin broadband clicks as well as tonal signals. An extension of this work will annotate tonal signals for an empirical comparison between the two signal types which can be used to further interpret the diel pattern and site use by the animals, important for conservation work.

In conclusion, we find the CNN to be an exciting tool for long-term monitoring of dolphin activity, providing higher accuracy than the C-POD but achieving a similar low rate of false positives. Notably we are impressed by the CNNs rate of missed detections compared to the C-POD. Fine-tuning the network to the specific soundscape in which it is deployed is expected to improve performance, and monitoring projects can make use of delphinid tonal signals which the model can also extract from the soundscape. This work encourages the collection of broadband acoustic to provide context to detector performances and the ambient environment in which the animals are residing. As marine habitats get noisier a range of national and international regulations[35,36,37] have arisen to understand and quantify noise levels regionally. Using a CNN as part of the analysis pipeline allows automated signal extraction to occur in a timely manner and encourages researchers to gather large PAM datasets which can be used to address a wealth of ecological questions.

5 REFERENCES

1. Clark, C.W., Brown, M.W. and Corkeron, P., 2010. Visual and acoustic surveys for North Atlantic right whales, *Eubalaena glacialis*, in Cape Cod Bay. *Mar Mammal Sci*, *26*(4), pp.837-854
2. Wimmer, J., Towsey, M., Planitz, B., Roe, P. and Williamson, I., 2010. Scaling acoustic data analysis through collaboration and automation. In *2010 IEEE Sixth International Conference on e-Science*. *IEEE,* pp. 308-315.
3. Lammers, M.O., Au, W.W. and Herzing, D.L., 2003. The broadband social acoustic signalling behaviour of spinner and spotted dolphins. *JASA*, *114*(3), pp.1629-1639.
4. Mellinger, D.K., Nieukirk, S.L., Matsumoto, H., Heimlich, S.L., Dziak, R.P., Haxel, J., Fowler, M., Meinig, C. and Miller, H.V., 2007. Seasonal occurrence of North Atlantic right whale (*Eubalaena glacialis*) vocalizations at two sites on the Scotian Shelf. *Mar Mammal Sci*, *23*(4), pp.856-867.
5. Brookes, K.L., Bailey, H. and Thompson, P.M., 2013. Predictions from Harbor Porpoise habitat association models are confirmed by long-term passive acoustic monitoring. *JASA*, *134*(3), pp.2523-2533.
6. Van Geel, N.C., Risch, D., Benjamins, S., Brook, T., Culloch, R.M., Edwards, E.W., Stevens, C. and Wilson, B., 2022. Monitoring cetacean occurrence and variability in ambient sound in Scottish offshore waters. *Front. Remote sens*, *3*:934681.
7. Fleishman, E., Cholewiak, D., Gillespie, D., Helble, T., Klinck, H., Nosal, E.M. and Roch, M.A., 2023. Ecological inferences about marine mammals from passive acoustic data. *Biol. Rev, 898,* pp. 1633-1647.
8. Moore, P.W., Dankiewicz, L.A. and Houser, D.S., 2008. Beamwidth control and angular target detection in an echolocating bottlenose dolphin (Tursiops truncatus). *JASA*, *124*(5), pp.3324-3332.
9. Bossart, G.D., 2011. Marine mammals as sentinel species for oceans and human health. *Vet. Pathol*, *48*(3), pp.676-690.
10. Caldwell, M.C., Caldwell, D.K. and Tyack, P.L., 1990. Review of the signature-whistle hypothesis for the Atlantic bottlenose dolphin. *The bottlenose dolphin*, pp.199-234.
11. Janik, V.M. and Slater, P.J., 1998. Context-specific use suggests that bottlenose dolphin signature whistles are cohesion calls. *Anim. Behav*, *56*(4), pp.829-838.
12. Gibb, R., Browning, E., Glover-Kapfer, P. and Jones, K.E., 2019. Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring. *Methods Ecol. Evol*, *10*(2), pp.169-185.
13. Shiu, Y., Palmer, K.J., Roch, M.A., Fleishman, E., Liu, X., Nosal, E.M., Helble, T., Cholewiak, D., Gillespie, D. and Klinck, H., 2020. Deep neural networks for automated detection of marine mammal species. *Sci. Rep*, *10*(1), pp.1-12.
14. Allen, A.N., Harvey, M., Harrell, L., Jansen, A., Merkens, K.P., Wall, C.C., Cattiau, J. and Oleson, E.M., 2021. A convolutional neural network for automated detection of humpback whale song in a diverse, long-term passive acoustic dataset. *Front. Mar. Sci*, *8*, pp.165.

15. Stowell, D., 2022. Computational bioacoustics with deep learning: a review and roadmap. *PeerJ*, *10*:13152.
16. White, E.L., White, P.R., Bull, J.M., Risch, D., Beck, S., Edwards, E.W.J., 2022. More than a whistle: Automated detection of marine sound sources with a convolutional neural network. *Front. Mar. Sci, 9*:879145.
17. Bergler, C., Schröter, H., Cheng, R.X., Barth, V., Weber, M., Nöth, E., Hofer, H. and Maier, A., 2019. ORCA-SPOT: An automatic killer whale sound detection toolkit using deep learning. *Sci. Rep*, *9*(1), pp.1-17.
18. Jin, C., Kim, M., Jang, S. and Paeng, D.G., 2022. Semantic segmentation-based whistle extraction of indo-pacific bottlenose dolphin residing at the coast of Jeju island. *Ecol. Indic*, *137*:108792.
19. Frainer, G., Dufourq, E., Fearey, J., Dines, S., Probert, R., Elwen, S. and Gridley, T., 2023. Automatic detection and taxonomic identification of dolphin vocalisations using convolutional neural networks for passive acoustic monitoring. *Ecol. Inform*, *78*, pp.102-291.
20. White, E., Klinck, H., Bull, J., White, P. and Risch, D., 2023. One size fits all? Adaptation of trained CNNs to new marine acoustic environments. *Ecol. Inform*, 78:102363.
21. Bailey, H., Clay, G., Coates, E.A., Lusseau, D., Senior, B. and Thompson, P.M., 2010. Using T-PODs to assess variations in the occurrence of coastal bottlenose dolphins and harbour porpoises. *Aquat. Conserv.: Mar. Freshw. Ecosyst*, *20*(2), pp.150-158.
22. Zimmer, W.M., 2011. *Passive acoustic monitoring of cetaceans*. Cambridge University Press
23. Nuuttila, H.K., Thomas, L., Hiddink, J.G., Meier, R., Turner, J.R., Bennell, J.D., Tregenza, N.J. and Evans, P.G., 2013. Acoustic detection probability of bottlenose dolphins, Tursiops truncatus, with static acoustic dataloggers in Cardigan Bay, Wales. *JASA*, *134*(3), pp.2596-2609.
24. Carlén, I., Thomas, L., Carlström, J., Amundin, M., Teilmann, J., Tregenza, N., Tougaard, J., Koblitz, J.C., Sveegaard, S., Wennerberg, D. and Loisa, O., 2018. Basin-scale distribution of harbour porpoises in the Baltic Sea provides basis for effective conservation actions. *Biol. Conserv*, *226*, pp.42-53.
25. Palmer, K.J., Brookes, K.L., Davies, I.M., Edwards, E. and Rendell, L., 2019. Habitat use of a coastal delphinid population investigated using passive acoustic monitoring. *Aquat. Conserv.: Mar. Freshw. Ecosyst*, *29*, pp.254-270.
26. Carstensen, J., Henriksen, O.D. and Teilmann, J., 2006. Impacts of offshore wind farm construction on harbour porpoises: acoustic monitoring of echolocation activity using porpoise detectors (T-PODs). *Mar. Ecol. Prog*, *321*, pp.295-308.
27. Tougaard, J., Carstensen, J., Teilmann, J., Skov, H. and Rasmussen, P., 2009. Pile driving zone of responsiveness extends beyond 20 km for harbor porpoises (*Phocoena phocoena*). *JASA*, *126*(1), pp.11-14.
28. Dähne, M., Katharina Verfuß, U., Brandecker, A., Siebert, U. and Benke, H., 2012. Methodology and results of calibration of tonal click detectors for small odontocetes (C-PODs). *JASA*, *134*(3), pp.2514-2522.
29. Garrod, A., Fandel, A.D., Wingfield, J.E., Fouda, L., Rice, A.N. and Bailey, H., 2018. Validating automated click detector dolphin detection rates and investigating factors affecting performance. *JASA*, *144*(2), pp.931-939.
30. Todd, N.R.E., Kavanagh, A.S., Rogan, E. and Jessopp, M.J., 2023. What the F-POD? Comparing the F-POD and C-POD for monitoring of harbor porpoise (*Phocoena phocoena*). *Ecol. Evol*, *13*(6):10186.
31. Sarnocinska, J., Tougaard, J., Johnson, M., Madsen, P.T. and Wahlberg, M., 2016, July. Comparing the performance of C-PODs and SoundTrap/PAMGUARD in detecting the acoustic activity of harbor porpoises (*Phocoena phocoena*). In *POMA*, 27(1):070013.
32. Tregenza, N., Dawson, S., Rayment, W. and Verfuss, U., 2016. Listening to echolocation clicks with PODs. *Listening in the Ocean*, pp.163-206.
33. Gillespie, D., Mellinger, D.K., Gordon, J., McLaren, D., Redmond, P., McHugh, R., Trinder, P., Deng, X.Y. and Thode, A., 2009. PAMGUARD: Semiautomated, open source software for real-time acoustic detection and localization of cetaceans. *The Journal of the Acoustical Society of America*, *125*, pp.2547-2547.
34. Tregenza, N., 2014. The software user guide for CPOD.EXE.

35. Merchant, N.D., Brookes, K.L., Faulkner, R.C., Bicknell, A.W., Godley, B.J. and Witt, M.J., 2016. Underwater noise levels in UK waters. *Sci. Rep*, *6*(1):36942.
36. Merchant, N.D., Putland, R.L., André, M., Baudin, E., Felli, M., Slabbekoorn, H. and Dekeling, R., 2022. A decade of underwater noise research in support of the European Marine Strategy Framework Directive. *Ocean Coast. Manag*, *228*, pp.106-299.
37. Petrizzo, A., Barbanti, A., Barfucci, G., Bastianini, M., Biagiotti, I., Bosi, S., Centurelli, M., Chavanne, R., Codarin, A., Costantini, I. and Cukrov Car, M., 2023. First assessment of underwater sound levels in the Northern Adriatic Sea at the basin scale. *Sci. Data*, *10*(1), pp.137.