Acoustic data enlighten East Pacific ecosystems

Stanislas Bebin, Daniele Bianchi & Jérôme Guiet
University of California, Los Angeles

The California Current System (CCS) is an upwelling region that hosts large biological communities. But in a context of climate change ~~(CC)~~ and global warming, these ecosystems are under pressure because their environment changes, with shifts in chemical properties of the water and changes in occurrence of upwellings events.

Models are tools to study ~~CC~~ influence over these communities up to high trophic levels. These models need to be constrained by data, which are difficult to get because the ocean is ~~really~~ hard to observe and sample. This is why acoustic data appear to be a ~~large~~ source of information to better constrain these models, because they can give information about the ocean's interior.

Therefore, we implemented a set of tools to clean and analyze ~~these~~ data. These tools have been applied on data coming from several scientific cruises along the CCS and some open-ocean cruises. This paper presents the different steps done ~~in order~~ to obtain usable data. We also present biological indicators we used and the correlations we found between occurrence of biomass and different environmental drivers.

In conclusion, the acoustic data were cleaned, except from one remaining noise: the "false bottom". We present preliminary results relating biomass with SST, chlorophyll, sunlight and moonlight. We point~~ed~~ the ability of acoustic data to identify large scale patterns that could help constrain ecosystem models. However, our data processing should be improved, and acoustic data need to be tackled with care because of ~~their~~ inhomogeneity between cruises and ~~the~~ noise~~s~~ they can contain.

The California Current System (CCS) is an upwelling system located along the West Coast of North America. From a biological point of view, the CCS is one of the most productive ecosystem in the world (Block et al. (2011)). Consequently, this ecosystem is also of great importance for people living in this region. It is the core of an entire economy based on fisheries and marine resources. But with global warming, this ecosystem is very likely to undergo important changes which will affect its own functioning, and therefore the services it provides.

Some models are being implemented (ATLANTIS Marshall et al. (2017) or APECOSM Maury (2010)) because the effects of these changes on marine food-webs are not well know. Their goal is to reproduce the link between the environment, the biomass distribution and the interactions with fisheries. ~~In order to~~ constrain these models at high trophic levels (pelagic fishes, big predators etc.), a first kind of data is based on catches, such as harvest time series of the Sea Around Us Project[1]. A second data type is based on the spatial occurrence of fishing activities (Global Fishing Watch dataset[2]). However, data of biomass distribution are scarce because, except for the upper layer, the ocean is impossible to observe and ~~really~~ difficult to sample.

In this context, acoustic data ~~appear as a data source of value to~~ constrain biological models. On the one hand because they scan the water column and they can provide high resolution information about the biomass (Benoit-Bird and Lawson (2016)). On the other hand, since acoustics are used for different purposes (fisheries, scientific oceanic campaigns such as seismic or bathymetry), a big amount of acoustic data is available.

In this study we describe how we processed the data available for the CCS and also some open-ocean data. Once ~~we processed and cleaned~~ the data from different noises, we implemented indicators that allowed us ~~the~~ to see the ~~occurrence of~~ biomass ~~along~~ the water column, and to compare it with some environmental drivers. The data processing requires an extreme carefulness because the quality of the data is highly variable. However we were able to extract regional maps with consistent patterns which can be useful for ecosystem modeling. Moreover, besides constraining the models, studying this enormous dataset might give us a better understanding of what happens inside the ocean.

We got our data from the Water Column Sonar Data Archive. The data ~~we use for this study~~ come from a specific sonar : Simrad EK60. This echo sounder is a multi-frequency echo sounder, which means that along the track, at each ping (a ping is when the sonar emits a pulse of energy as an acoustic

[1] http://www.seaaroundus.org/
[2] http://globalfishingwatch.org/

wave), the echo sounder can operate several frequencies simultaneously. We used the 18 kHz, 38 kHz, 70 kHz, 120 kHz and 200 kHz frequencies when ~~they were~~ available, ~~so that we could~~ make a multi-frequency analysis. Table 1 lists all the cruises we used, and some of their key characteristics.

The raw data provided by the EK60 sonar are values of volume backscattering ~~strength, whose symbol is~~ S_v~~, and unit is~~ $dB.m^{-1}$. The raw data are called "echograms": they represent S_v values ~~along~~ depth and time. The backscattered signal comes mostly from biological elements or the seabed.

But echograms ~~also~~ contain noises. These noises are generated by various sources - due to the equipment or the environment - and adopt different shapes in the echograms. The two main noises encountered in our data are the impulsive noise (IN) and the background noise (BN). Therefore, before doing any analysis, these data need to be cleaned from these noises.

The IN appears because of an unsynchronized echo sounder. It is a sporadic noise no longer than one ping, at a specific depth and time. Following Ryan et al. (2015) and Anderson et al. (2005), we implemented a filter for IN.

The BN is a noise coming from both the environment and the echo sounder itself. This noise is present at each ping, increases with depth, and is stronger for high frequencies at depth (see MacLennan ~~(~~1986~~)~~). Following De Robertis and Higginbottom (2007), we implemented a filter to remove this noise.

On top of that, we ~~computed~~ an algorithm that detects the bottom in our transects in order to remove its ~~signal~~, ~~so that we only have~~ biological backscatter ~~in our data.~~

After processing the data, we can analyze them with multi-frequency indicators to identify how the biomass is distributed along the water column~~s~~. To do so, we used two multi-frequency indicators: the multi-frequency single-beam imaging (MFSBI) indicator, and the multi-frequency indicator (MFI). These two indicators rely on the fact that we can differentiate organisms from the different intensities they will backscatter under different frequencies.

The MFSBI indicator (described in Wall et al. (2016) and Jech and Michaels (2006)) summarizes on a same echogram the contribution of four frequencies we have on four different echograms (18 kHz, 38 kHz, 120 kHz and 200 kHz). It informs on the dominant backscattered frequency (or frequencies) along the different parts of the echogram. However we do not have the relative contribution of these frequencies. This indicator is convenient to visualize on a single figure the spatial patterns of the biological signal in the water column.

The MFI indicator (described in Wall et al. (2016), using Trenkel and Berger (2013)) computes more accurately than the MFSBI the relative backscatter of the different frequencies at a given position/pixel in the echogram (depth and ping). At each pixel, a MFI value is computed ~~and it~~ characterize~~s~~ the shape of S_v as a function of echo sounder frequencies (Figure 7, on the left). These S_v shapes can be linked to four classes of organisms (Figure 7, on the right). A MFI ranging between 0 and 0.4 corresponds to large gas-bearing organisms (fish such as hake, gas-filled-swim-bladder fish); a MFI between 0.4 and 0.6 corresponds to small bubbles (can either be small perturbations in the water column, or either an indicator for larval fish or phytoplankton); a MFI between 0.7 and 0.8 is fluid-like zooplankton (euphausiids, copepods or shrimps); and a MFI between 0.8 and 1 would indicate non-gas bearing organisms (such as mackerel, non-swim-bladdered fish).

Then, for each of the four MFI classifications, we integrated S_v on a given depth layer: the obtained value is called S_L. For each MFI, we computed maps of S_L for the different layers we used: for a given MFI, we could then see the spatial occurrence of the corresponding backscattered signal.

The noise filters and the bottom detection were successfully implemented. The impulsive noise removal works under any circumstances. ~~Figure 5~~ is an example of this removal. Our background noise filter works successfully too. ~~Figure 6~~ shows the noise removal for a 120 kHz echogram. In this example, the noise removal even allows us to see a new layer of signal. This filter is extremely important, especially for the highest frequencies. Our algorithm for the bottom detection works successfully too: we removed the bottom from the echograms. Therefore, we were able to remove the main noises from our data as well as the non biological backscatter from the seabed.

However, despite all our data processing, a noise called "false bottom" or "ghost bottom" remains in our echograms. This noise is due to multiple reflections between the bottom and the surface of the water from previous pings, detected by the current record. Nevertheless, this noise was usually located deep enough not to be taken into account in our analysis.

Once processed, we could start analyzing our data with the multi-frequency indicators. The MFSBI indicator pointed out several different typical patterns in our cruises.

Figure 1 illustrates the kind of transect we obtained. First of all, we ~~could~~ notice layers of low frequencies dominant backscatter (18 and 38 kHz, grey and blue layers). ~~These layers didn't fit into a very well organized pattern in all our cruises, but we noticed some recurrence of these layers. There was~~ two main depths ~~at which we could see them~~ : the first one around 50-200 meters deep, and the second one between 200 and 350 meters. These two main layers ~~are very~~ likely ~~to~~ indicate daily migration of fishes.

~~Another pattern we observed in the data is an~~ aggregation of high frequency backscatter organisms around shelf breaks (see orange patches around the shelf on

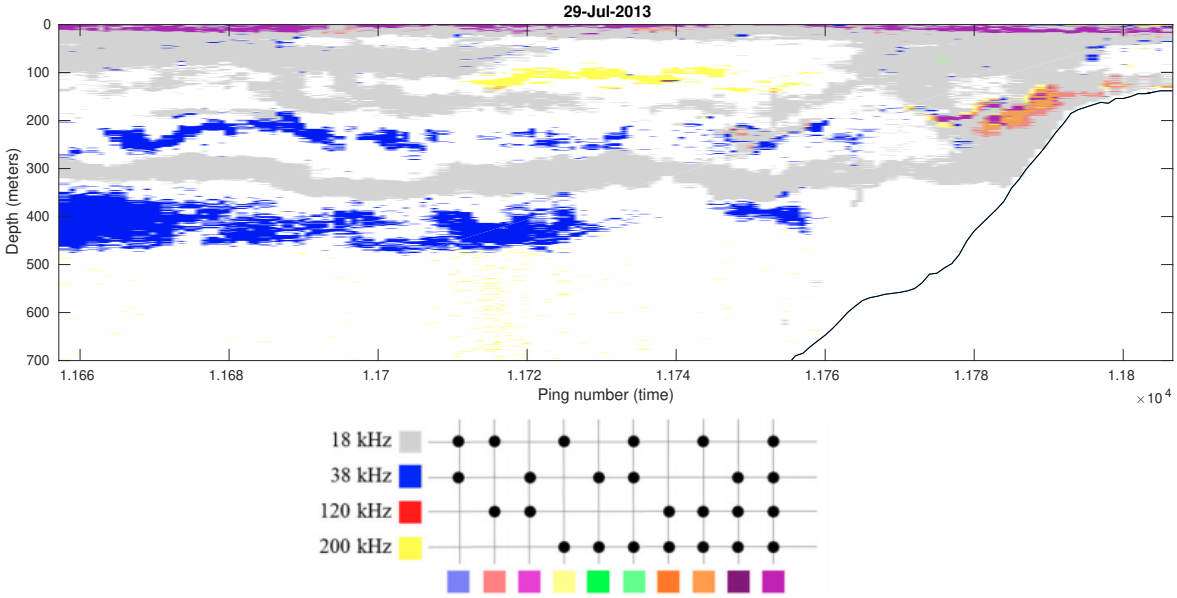

Figure 1: MFSBI transect of SH1305 cruise. The color table below explains the meaning of the different colours for the MFSBI indicator. The black line is the bottom. This table is taken from Wall et al. (2016).

Figure 1), most of the time between 100 and 200 meters deep. This pattern was strong in the data from the cruises located in the CCS. We think that these aggregations are due to the ascent of nutrient-rich anoxic deep waters along the shelf slope, meeting oxygenated waters and resulting in highly productive areas. These high frequency backscatter organisms can be fish without swim-bladder, or layers of zooplankton.
Finally, in all the transects there were patches in the surface layer (above 50 meters) with strong backscatter at all the frequencies (see the purple surface layer on Figure 1). These are probably very productive waters, full of zooplankton and different kind of fishes. The occurrence of these patches should be linked to the observed chlorophyll concentrations or the SST. However, our initial comparisons were not that satisfactory. The observed environmental drivers were maybe not precise enough, or the MFSBI indicator is a too crude estimate of the occurrence of living organisms.

The MFI allowed us to begin a finer analysis of our dataset. After calculating the MFI for each pixel of our echograms, we used the layers previously found with the MFSBI indicator (0-50 m, 50-200 m, and 200-350 m) to integrate S_v and obtain S_L values for each MFI groups. We did~~n't do any~~ analysis~~is~~ below 350 meters because the 200 kHz echograms were not reliable ~~any more~~ below that depth. Figures 2 (below), 8, 9 and 10 (Supplementary Material) illustrate the obtained maps indicating the spatial occurrence of the different MFI groups and the corresponding S_L values. To compare with environmental drivers, we also plotted SST and chlorophyll density values[3]. On Figure 2, it seems that there is a correlation between S_L maps, SST (more biomass occurrence for low SST) and chlorophyll for the three different depths. The fact that we can see a strong S_L at all the depths is probably due to the daily migration we referred to before. Swim-bladdered fishes are thus influenced by surface drivers. On Figure 10 which corresponds to the non-swim-bladder fishes (~~so~~ fishes that barely migrate), most of the signal is in the lowest layer (200-350 meters). This is coherent because this kind of fishes stay deep.
These results are a starting point and illustrate large scale variations in the occurrence of distinct groups of organisms linked to environmental drivers. Finer analysis are required to better quantify these links and use them to constrain marine ecosystem models.

Beside the CCS cruises, we also processed open-ocean cruises. These only provide the 18 kHz frequency so we could~~n't apply the previously described~~ multi-frequency indicators. However, they revealed interesting patterns. Figure 3 illustrates the daily migration of large populations from the deep sea to the surface, driven by the sun elevation angle. These daily migrations are well documented by Bianchi et al. (2013), but this example is interesting since we can distinguish different layers of organisms that migrate. For instance, during the second daily migration on this figure, we can observe at least five different migrating layers. The processes ~~controlling~~ these different layers ~~can be very interesting to study~~.
Another point is the difference between the East and West deep scattering layer (DSL): the DSL occurs deeper towards Hawaii (left of the figure) than towards California. This might be due to differences of the water temperature, but further analysis are required.

[3]We got chlorophyll concentration data from GlobColour Project, and the sea surface temperature (SST) from the OSTIA dataset

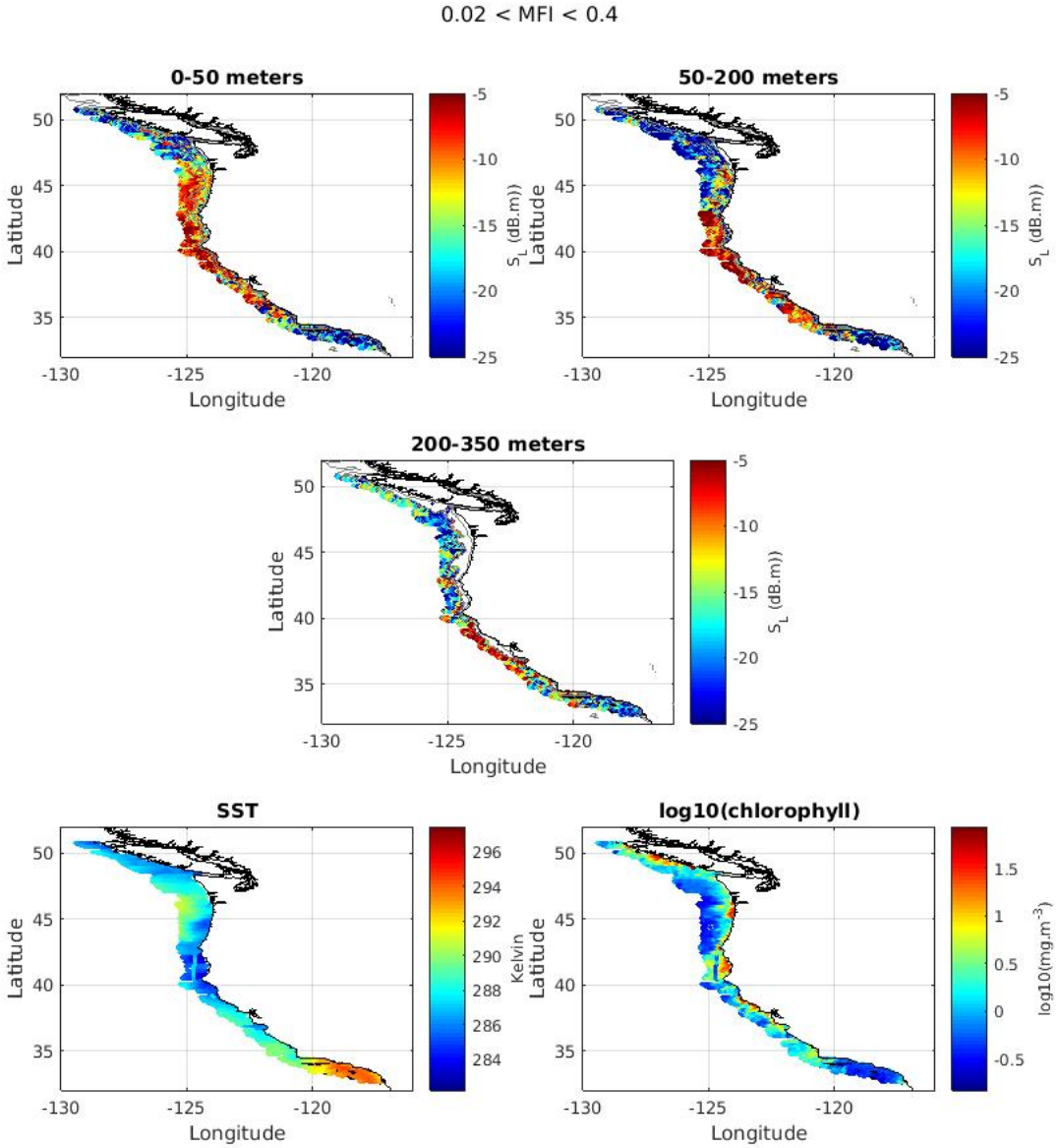

Figure 2: For a specific MFI range of values (here $0.02 < \text{MFI} < 0.4$, swim-bladder fishes and large bubbles), maps of S_L values at three different layers. We also plotted the SST and chlorophyll maps. Cruise RL1606.

Interestingly, we also observed night migrations on open-ocean cruises. Indeed, Figure 4 shows that a very strong layer of backscatter, around 50 meters deep, moves downward during nighttime. This migration seems caused by the variation of the moonlight since we see a link between the elevation of the moon and the depth of the layer. The dotted lines (indicating when the moon rises above the horizon) seem to correspond to the beginning of these night migrations. Though, we have to be careful because what is plotted is the moon elevation, not the moonlight. To have it, we need the moon phase as well. During this cruise at this specific time (from January 28$^{\text{th}}$ to February 1$^{\text{st}}$ 2016), the moon phase was around a full moon period. So the light intensity coming from the moon was important during the night, which might explain these migrations.

In all our data processing, a lot of improvements are possible. First of all, beside the IN and the BN, there are a few other noises we didn't remove such as the Attenuated Signal and the Transient Noise (Ryan et al. (2015)). We implemented filters to remove them but we decided not to use them for different reasons. Both noises were rare in our data (and deep enough for the Transient Noise), so it shouldn't affect our analysis. But still it would be better to remove them.

Moreover, we didn't remove the "false bottom"

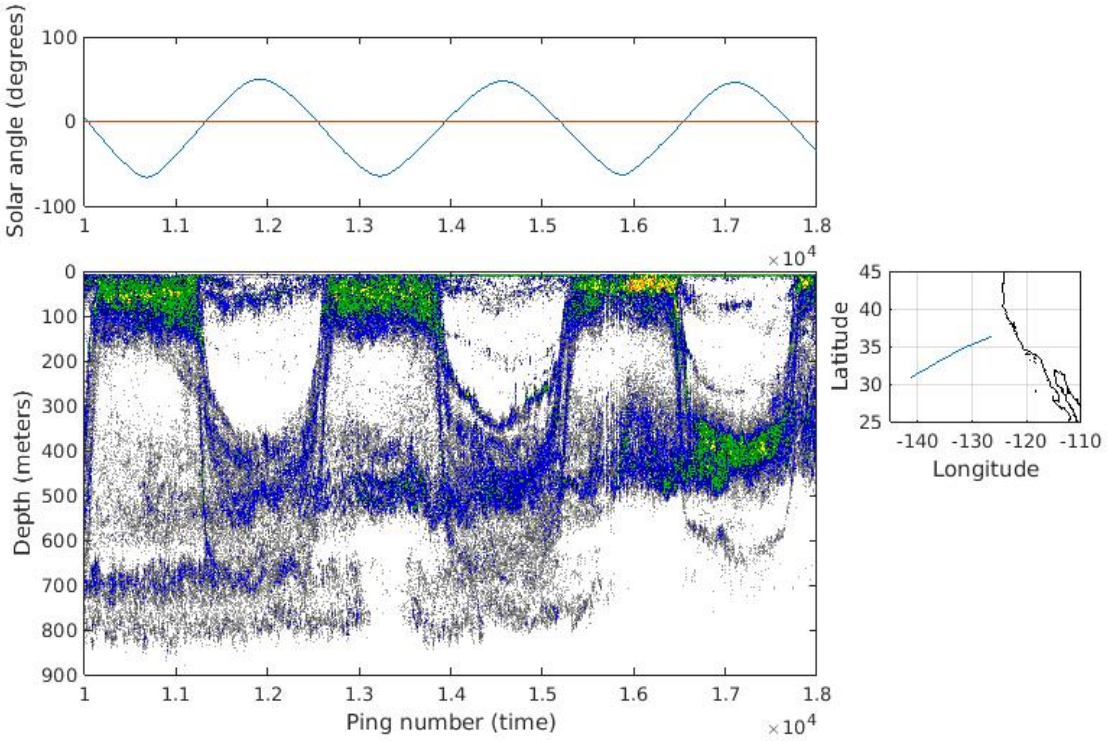

Figure 3: Transect of the cruise EX1505. This is the 18 kHz data, showing three daily migrations of the biomass from the deep sea to the surface layer, plotted with the solar elevation angle on top. A map of this cruise portion is shown on the right.

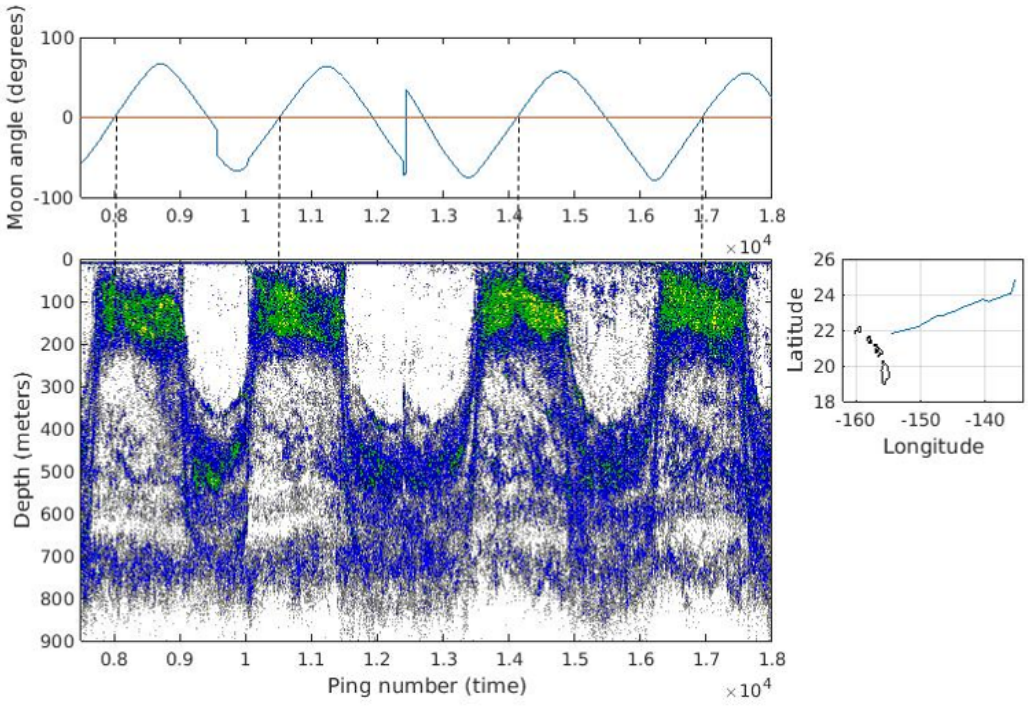

Figure 4: Transect of the cruise EX1601. This is the 18 kHz data, showing four daily migrations from the deep sea to the surface layer, but also four nocturnal migrations. We plotted our data with the moon elevation angle on top. A map of this cruise portion is shown on the right (islands are Hawaii).

from our data either. This noise was more recurrent, so it would be important to remove it. To tackle this problem, while other echogram users correct this noise manually (going throughout the data and removing suspicious patterns), we suggest to adapt the bottom detection algorithm for false bottom identification.

Again to improve our data processing, the parameters of the different filters may require a finer tuning for the different transects. Even if parameters are suggested in the literature and because each echogram is made under varying conditions, an adjustment of parameters is required to better process the echograms and therefore improve the detection of biological patterns.

In addition, a big issue with these data is to interpret them as biomass. From this work, we can interpret backscattered signals only as a proof of occurrence, but definitely not as an absolute biomass (weight by volume unit). This interpretation is extremely delicate because the backscatter from an individual can differ depending on where this individual is in the water column (deep or shallow). This is especially true for species carrying gas-filled-swim-bladders (contraction of gas with depth). Moreover within same species groups at a given frequency, the size of the individuals can generate different backscatter strengths as well (Davison et al. (2015)). This why MFI groups have big error bars on Figure 7, right. We have to be careful interpreting our results.

Current attempts to model the CCS biological component at high resolution need a finer understanding of how different organisms are spatially distributed. Although these results are preliminary, we paved the way to reach this goal.

First, we developed tools allowing the processing of echograms, from their extraction from the website to their analysis with multi-frequency biological indicators. A set of filters has been implemented ~~in order~~ to remove the main source of noise in the data, impulsive noise and background noise ~~being the most important~~. A functional bottom detection algorithm also allowed us to remove the strong signal backscattered by the seabed. The main remaining problem ~~in the data~~ is the false bottom. That~~'s~~ why we need to be careful about the interpretations we could make with the biological indicators. While this noise can be very frequent in some parts of coastal transects, it mostly occurs at depths we did~~n't~~ take into account in our MFI analysis (below 350 m~~eters~~).

Secondly, our goal to constrain marine ecosystem models with acoustic data, seems to be possible. Indeed, the computed muti-frequency biological indicators show large scale patterns. Although finer analysis are required to clearly reveal the drivers, comparisons with SST and chlorophyll show a link between these latter and the occurrence of biomass for distinct MFI groups. We should keep in mind that these results can't be linked to accurate biomass estimates because it is difficult to link backscattered signal strength to biomass without on site trawl sampling.

Finally, these acoustic data are really promising for ~~having~~ a better understanding of ecosystems dynamic in the ocean, illustrated by the example of the night migrations that might be explained by moonlight intensity. ~~These data are a source of information of inside the ocean that we cannot neglect.~~

Supplementary Material

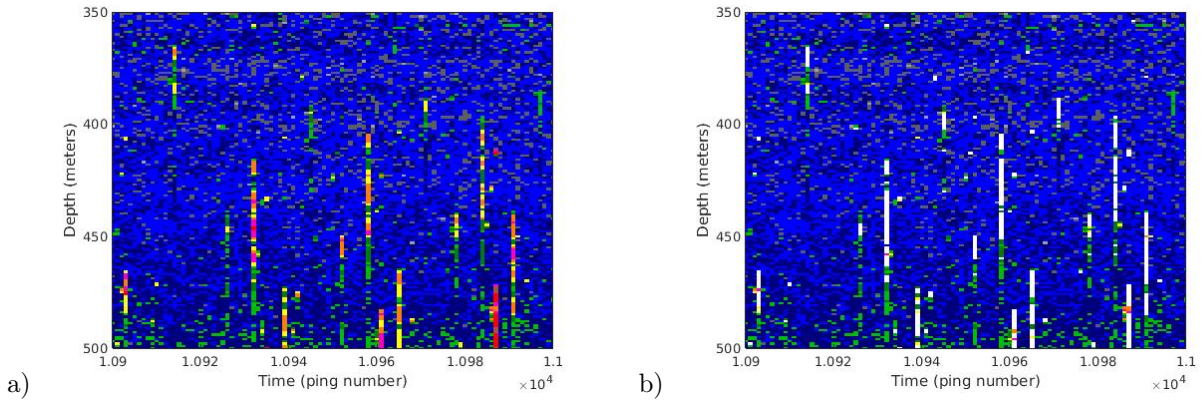

Figure 5: Example of impulsive noise removal. Time is represented as a ping number. Figure (a) is the data with the noise (very strong signals, red lines). Figure (b) is the same data after applying our impulsive noise removal function.

Cruise ID	Dates	Frequencies available (kHz)	References
SH1305	2013-05-21 to 2013-08-29	18, 38, 70, 120, 200	NCEI, NOAA (2013)
RL1606	2016-06-28 to 2016-09-23	18, 38, 70, 120, 200, 333	NCEI, NOAA (2016)
EX1601	2016-01-20 to 2016-02-07	18, 38, 70, 120, 200	NGDC, NOAA (2016)
EX1505	2015-10-05 to 2015-10-16	18	NGDC, NOAA (2015)

Table 1: List of all the cruises presented in this article, with their dates, frequencies available, and references

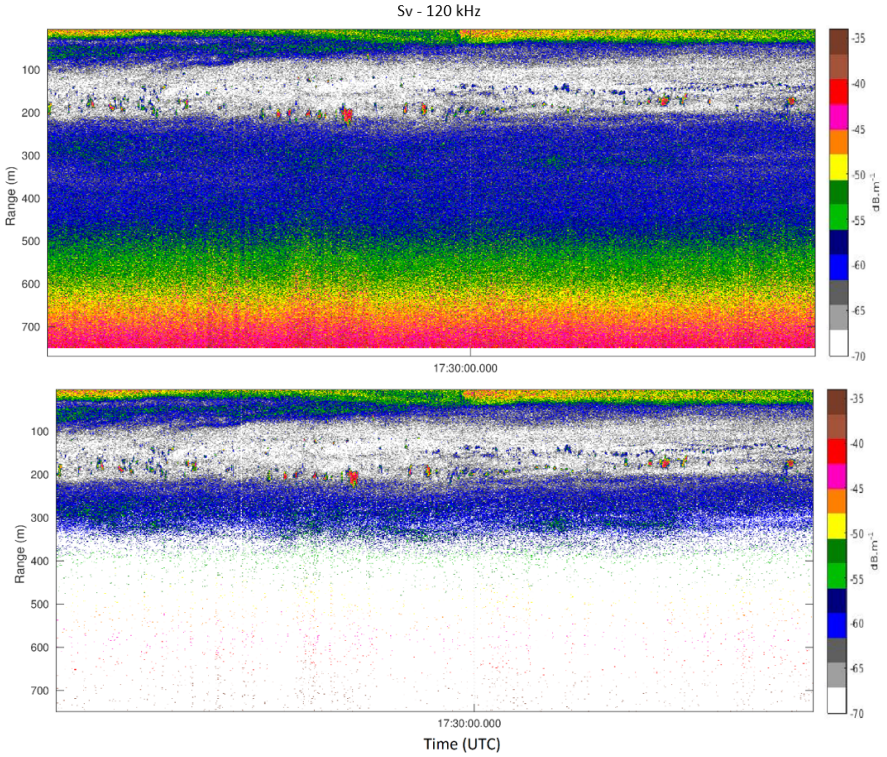

Figure 6: Example of background noise removal. The upper figure is the data with the background noise. We can see on it a surface layer of backscatter, and some aggregations, strong red signal (probably fish schools). The figure below is the same echogram after applying our noise removal function. A deep scattering layer we couldn't see on the upper figure appears.

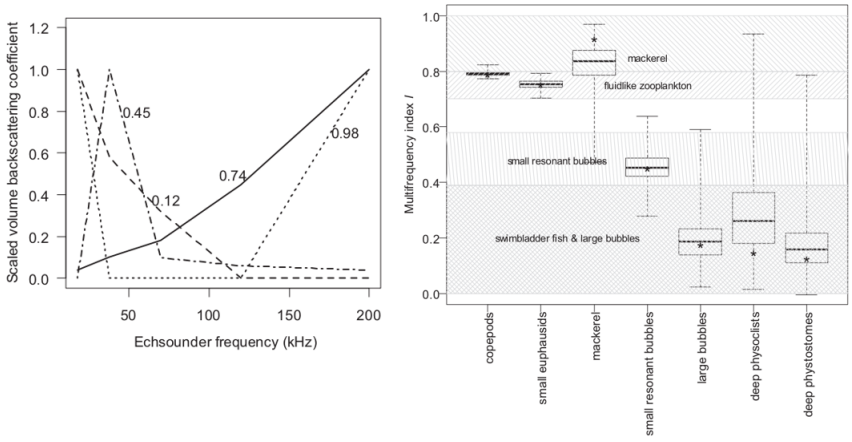

Figure 7: On the left, graphic showing the shape of the volume backscattering coefficient (S_v) curves (scaled between 0 and 1) according to the frequency for four characteristic MFI values. On the right, graphic relating MFI values to the corresponding classification types. Graphics from Trenkel and Berger (2013).

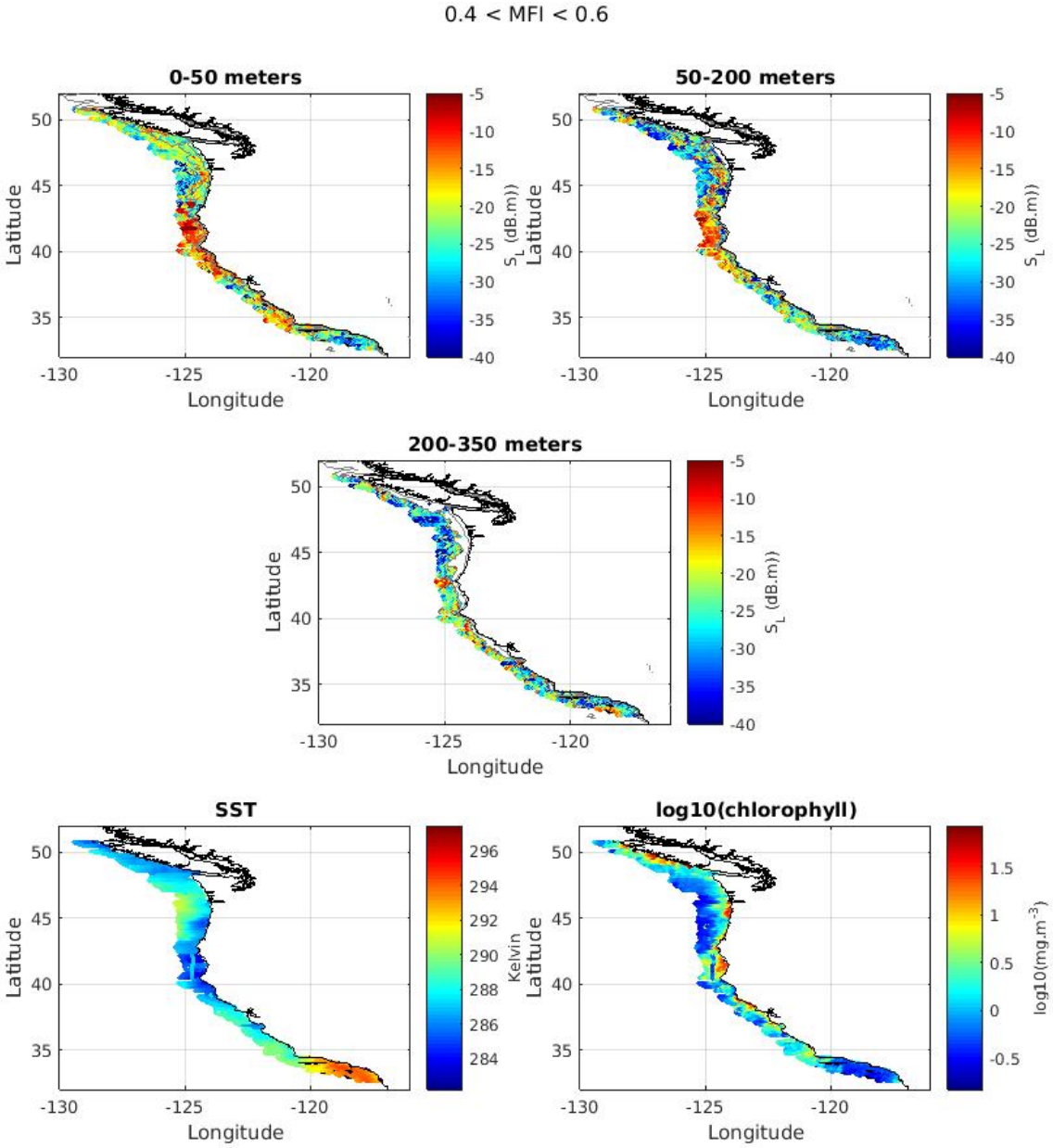

Figure 8: For a specific MFI range of values (here $0.4 < \text{MFI} < 0.6$, small resonant bubbles), maps of S_L values at three different layers. We also plotted the SST map and chlorophyll map. Cruise RL1606.

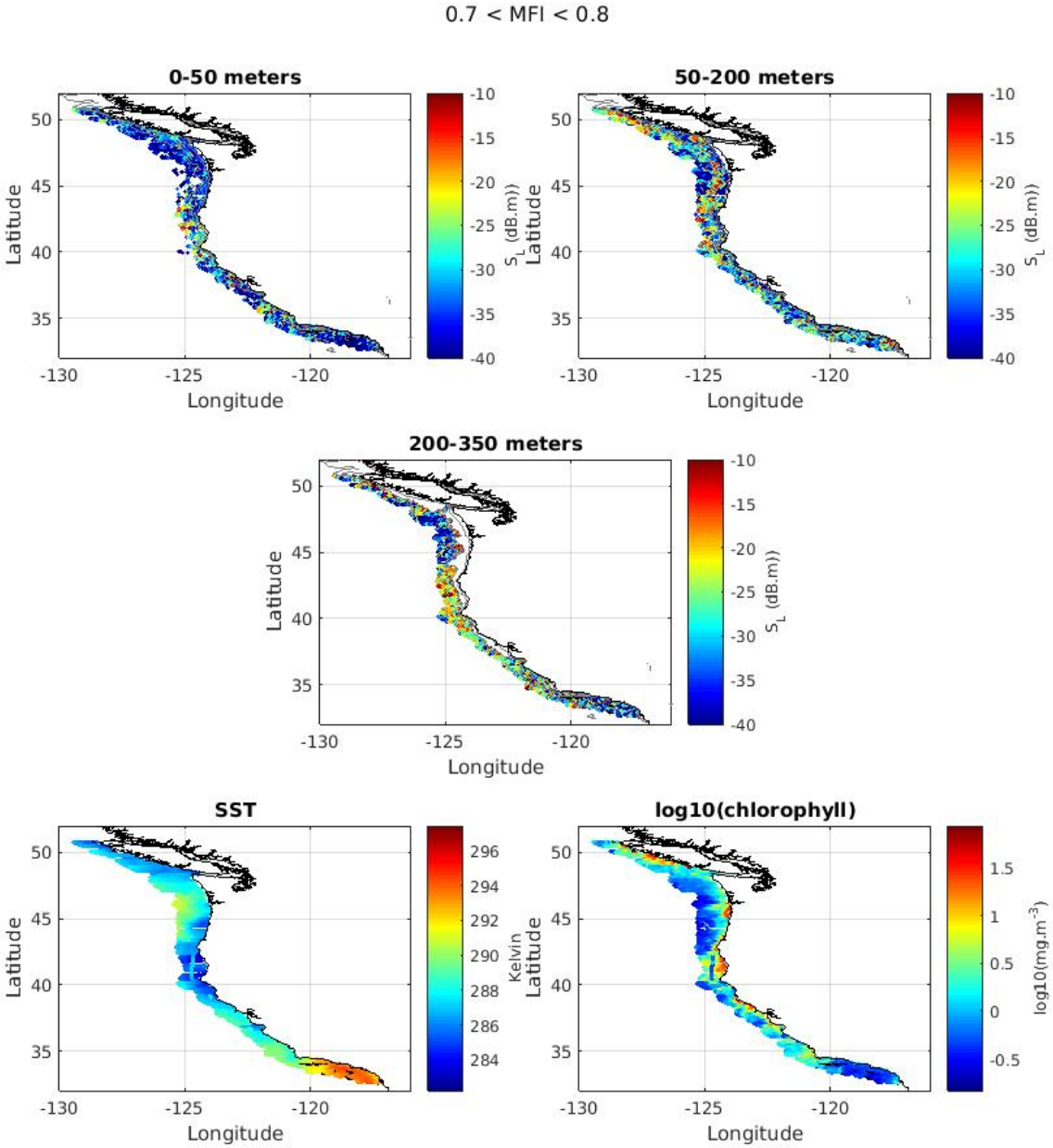

Figure 9: For a specific MFI range of values (here $0.7 < \text{MFI} < 0.8$, fluidlike zooplankton), maps of S_L values at three different layers. We also plotted the SST map and chlorophyll map. Cruise RL1606.

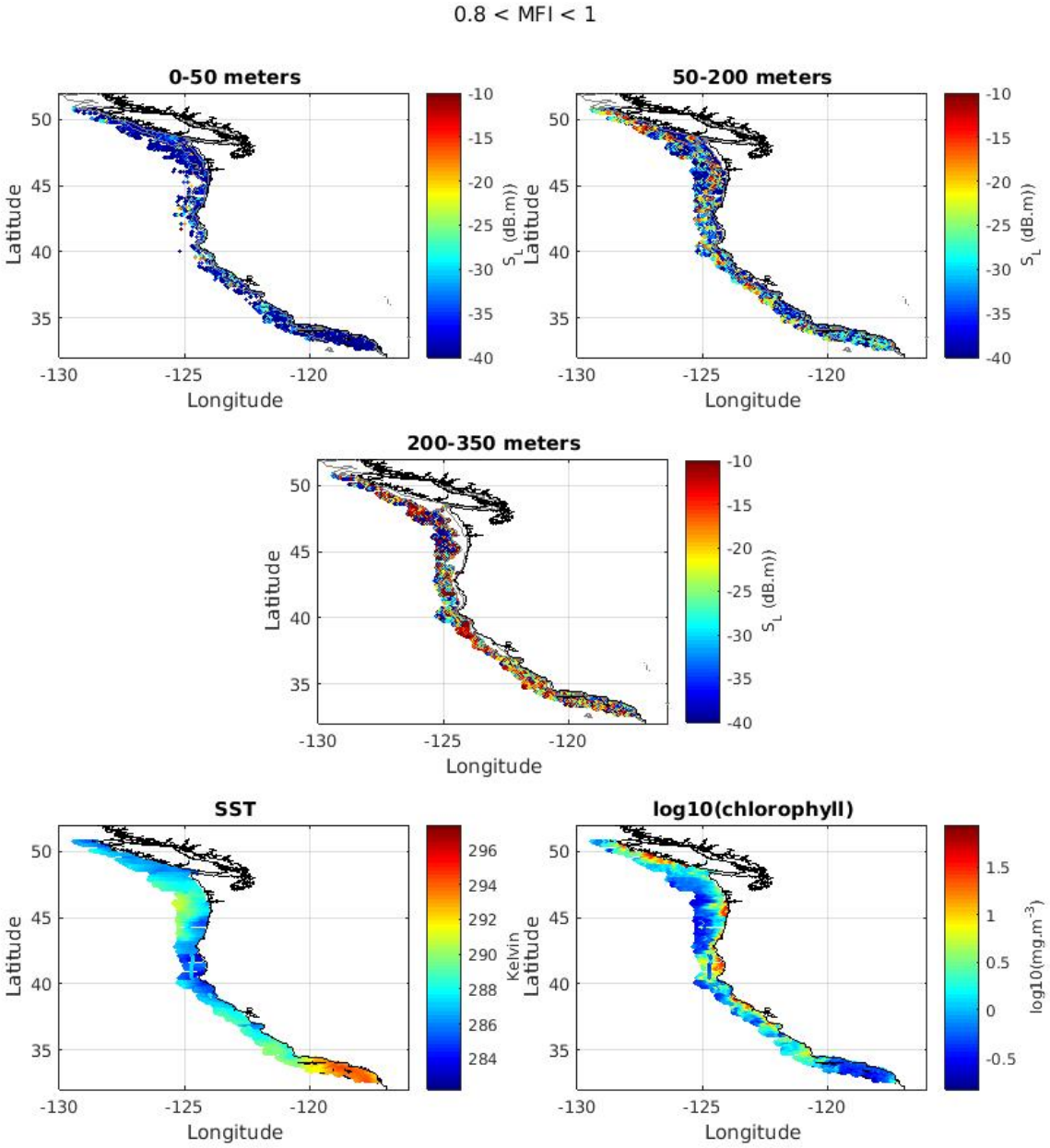

Figure 10: For a specific MFI range of values (here $0.8 < \text{MFI} < 1$, non-swim-bladder fishes), maps of S_L values at three different layers. We also plotted the SST map and chlorophyll map. Cruise RL1606.

References

Anderson, C. I. H., Brierley, A. S., and Armstrong, F. (2005). Spatio-temporal variability in the distribution of epi- and meso-pelagic acoustic backscatter in the irminger sea, north atlantic, with implications for predation on calanus finmarchicus. *Marine Biology*, 146:1177–1188.

Benoit-Bird, K. J. and Lawson, G. L. (2016). Ecological insights from pelagic habitats acquired using active acoustic techniques. *Annual Review of Marine Science*, 8(1):463–490.

Bianchi, D., Galbraith, E. D., Carozza, D. A., Mislan, K. A. S., and Stock, C. A. (2013). Intensification of open-ocean oxygen depletion by vertically migrating animals. *Nature Geoscience*, 6:545–548.

Block, B., Jonsen, I., Jorgensen, S., Winship, A., Shaffer, S., Bograd, S., Hazen, E., G Foley, D., Breed, G., Harrison, A.-L., Ganong, J., Swithenbank, A., Castleton, M., Dewar, H., R Mate, B., Shillinger, G., M Schaefer, K., Benson, S., J Weise, M., and Costa, D. (2011). Tracking apex marine predator movements in a dynamic ocean. 475:86–90.

Davison, P. C., Koslow, J. A., and Kloser, R. J. (2015). Acoustic biomass estimation of mesopelagic fish: backscattering from individuals, populations, and communities. *ICES Journal of Marine Science*, 72(5):1413–1424.

De Robertis, A. and Higginbottom, I. (2007). A post-processing technique to estimate the signal-to-noise ratio and remove echosounder background noise. *ICES Journal of Marine Science*, 64(6):1282–1291.

Jech, J. M. and Michaels, W. L. (2006). A multifrequency method to classify and evaluate fisheries acoustics data. *Journal canadien des sciences halieutiques et aquatiques*, 63:2225–2235.

MacLennan, D. N. (1986). Time varied gain functions for pulsed sonars. *Journal of Sound and Vibration*, 110:511–522.

Maclennan, D. N., Fernandes, P. G., and Dalen, J. (2002). A consistent approach to definitions and symbols in fisheries acoustics. *ICES Journal of Marine Science*, 59(2):365–369.

Marshall, K. N., Kaplan, I. C., Hodgson, E. E., Hermann, A., Busch, S., McElhany, P., Essington, T. E., Harvey, C. J., and Fulton, E. A. (2017). Risks of ocean acidification in the california current food web and fisheries: ecosystem model projections. *Global Change Biology*, 23(4):1525–1539.

Maury, O. (2010). An overview of apecosm, a spatialized mass balanced "apex predators ecosystem model" to study physiologically structured tuna population dynamics in their ecosystem. *Progress in Oceanography*, 84(1):113–117.

McKinnell, S. M. and Dagg, M. J. (2010). Marine Ecosystems of the North Pacific Ocean 2003-2008. Technical report, North Pacific Marine Science Organization (PICES).

NCEI, NOAA (2013). Northwest Fisheries Science Center (2013): Water Column Sonar Data Collection (SH1305, EK60). https://data.nodc.noaa.gov/cgi-bin/iso?id=gov.noaa.ngdc.mgg.wcd:SH1305_EK60.

NCEI, NOAA (2016). Southwest Fisheries Science Center (2016): Water Column Sonar Data Collection (RL1606, EK60). https://data.nodc.noaa.gov/cgi-bin/iso?id=gov.noaa.ngdc.mgg.wcd:RL1606_EK60.

NGDC, NOAA (2015). NOAA Office of Ocean Exploration and Research (2015): Water Column Sonar Data Collection (EX1505, EK60). https://data.nodc.noaa.gov/cgi-bin/iso?id=gov.noaa.ngdc.mgg.wcd:EX1505_EK60.

NGDC, NOAA (2016). NOAA Office of Ocean Exploration and Research (2016): Water Column Sonar Data Collection (EX1601, EK60). https://data.nodc.noaa.gov/cgi-bin/iso?id=gov.noaa.ngdc.mgg.wcd:EX1601_EK60.

Ryan, T. E., Downie, R. A., Kloser, R. J., and Keith, G. (2015). Reducing bias due to noise and attenuation in open-ocean echo integration data. *ICES Journal of Marine Science*, 72(8):2482–2493.

Simmonds, J. and MacLennan, D. (2006). *Fisheries Acoustics*. Blackwell Science.

Trenkel, V. M. and Berger, L. (2013). A fisheries acoustic multi-frequency indicator to inform on large scale spatial patterns of aquatic pelagic ecosystems. *Ecological Indicators*, 30:72–79.

Wall, C. C., Jech, J. M., and McLean, S. J. (2016). Increasing the accessibility of acoustic data through global access and imagery. *ICES Journal of Marine Science*, 73(8):2093–2103.