Overview:

The manuscript details a large-scale qPCR-based eDNA survey of hake along the U.S. west coast. The authors collected eDNA samples across the region (186 stations, 6 depths) and simultaneously gathered hake biomass data using acoustic and trawl methods. Using collected data, they separately modeled eDNA and acoustic-trawl distributions in this large area, allowing for quantitative method comparisons at management-level scales. This study represents an impressive and novel fish eDNA dataset in the ocean and draws some important conclusions regarding the use of non-invasive qPCR methods to survey fisheries-important taxa, like hake.

This manuscript should be considered for publication in Proceedings B and I think it will be of interest to a wide audience. While I do have a lot of suggestions to improve the manuscript, most of these are minor and I believe all of them can be addressed in a timely manner in a revised draft.

General comments:

I would consider changing the title of the manuscript to better reflect the specific target organism and scope of the study. For instance, something like “Environmental DNA provides quantitative estimates of Pacific hake abundance and distribution in the coastal ocean”. You may or may not want to include the scientific name as well.

There are some missing details that may strengthen the abstract. It would be nice to highlight some quantitative results. For example, is there some number or percent overlap that you can report based on the comparison of qPCR and acoustic-trawl data? What was the level of variation between methods? You could also report a correlation value (Pearson). There is also no reference of the models or eDNA index in the abstract. An important finding you touch on but may expand is that on a sample-to-sample basis, eDNA and other methods (acoustic-trawl) may not correlate due to inherent differences in the method (and in your case the differences in sampling time of day) but they do correlate nicely on large scales that are relevant for fisheries management.

The introduction reads a bit unorganized and is redundant at times. I think you can tighten this up. I would suggest a paragraph that introduces eDNA, its benefits (e.g., non-invasive, sensitive, cost-effective), and applications to management strategies that require quantitative measurements. Important to note, is that applying eDNA to management of ocean resources is still new. Another paragraph could transition to our knowledge of traditional surveys (trawls, visual, acoustic, etc.) in the ocean vs. eDNA and include inherent differences in fish eDNA in the environment compared to a single discrete fish sample collected in a net tow. Then, you can have a paragraph or two that zoom in on your study region, taxa, and question. Somewhere in the introduction, you may also distinguish amplicon metabarcoding vs. qPCR-based eDNA approaches and how they differ in the types of information gained (e.g., metabarcoding promotes monitoring though is only qualitative vs. qPCR that is more targeted yet quantitative).

This is touched on in the discussion but I would mention it earlier. Lastly, I think the introduction would benefit by zooming in on fish and fisheries earlier, establishing the potential for qPCR to inform large scale management in the ocean.

Regarding methods, more detail would be helpful in the main text. A lot of this detail is buried in the Supplemental file. I would pull some of this information into the manuscript and refer to the Supplement for additional details and figures. For instance, more information on your qPCR methods is needed in the main text. I know the emphasis is on the model(s) but it is important to clearly state qPCR eDNA methods, especially for reproducibility (see Langlois et al. 2021 – *Environmental DNA*). The reader may miss these key details. Some qPCR questions that arise: What samples did you use for positive hake controls? Did you use voucher specimens? What primers (and other reagents) did you use? Did you account for false positives or negatives? What about closely-relate taxa? Did you measure bulk DNA concentrations in the samples (e.g., via Qubit)? Or did you just quantify hake-specific DNA concentration? I would also pull-out information on DNA contamination and washing error and touch on these in the methods or results section (see specific points below), referencing the Supplement for more details. Again, some readers may not look at the Supplemental file.

I would be consistent with the way you reference “acoustic-trawl” surveys throughout the manuscript. It is important to emphasize that your biomass model is integrating both acoustic and trawl data, and so, referring to this as “acoustic-trawl” seems accurate.

More specific comments:

Line 23: Environmental DNA does not necessarily reflect species present in a given habitat. Something like, “…and collecting this environmental DNA can provide information on species that may be present in a given habitat” may be more appropriate.

Line 29: Spatially smooth signature? Sounds cryptic. Do you mean a wide spatial coverage? I would clarify here and in other parts where this is mentioned.

Line 71: You could also include information about distribution of eDNA from its source location due to movement by currents or tides. Though not a concern for your study, it is important to note that fish eDNA can also be patchy in coastal areas (see Kumar et al. 2021 – *Environmental DNA*). When you integrate over wide areas this becomes less of an issue. I would make this clear.

Line 75: I would be careful saying this is the most spatially extensive eDNA survey of the oceans to date. It may be more accurate to say “fish eDNA” survey. Though not sampling eDNA per se, the Tara Oceans expedition performed DNA barcoding of microbes-protists on a global scale. I am not diminishing your study; I would just clarify. You may also say “qPCR-based eDNA” survey. Both would be accurate.

Line 76: I would remove the statement “an area of ocean approximately equivalent to Portugal’s land area”. This seems extraneous.

Line 81: You mention there are rich datasets of hake biomass, presumably from trawls. Have there been many studies targeting hake with qPCR-based eDNA methods? Have others conducted trawl vs. eDNA of important fisheries taxa? What have they found?

Line 87: Has this type of model been used in terrestrial applications? Any papers you could cite?

Lines 96-99: I would move this material to the introduction.

Line 101: You reference a related paper with more details. Still, I think it would be nice to point out lat-long range here or geographic sites that cover the study area.

Line 114: Here, you could briefly mention that you found some DNA in the negative controls but it was mostly detected below xxx copies L^{-1}. Then reference the Supplemental file for more information.

Lines 116-117: Is this the filtering protocol used for controls and discrete seawater samples? I would specify. How long were the filters stored before extraction?

Lines 119-120: Citation associated with the extraction method and the increase in yield?

Lines 120-123: Expand on qPCR details. Pull out from Supplemental file and reference additional details and figures.

Line 126: Confused by this sentence. What is the “process”?

Line 134: I would refer to specific figures in the Supplemental file that are associated with model optimization mentioned in text.

Line 140: I would briefly expand on the wash error and refer to the Supplemental material for relevant figures. Otherwise, the reader will have no idea why you included this in your model.

Line 147: Could say “indicates” instead of “indexes”.

Line 153: Could say “simultaneously” instead of “contemporaneously”.

Lines 159-161: This is a key aspect of the study. The reader should be aware of this somewhere in the introduction. Not necessarily the day vs. night differences between methods that preclude direct site level comparisons but rather that fact that you are focusing on broad-scale and not sample-to-sample comparisons.

Lines 216-217: Remove “genetic signatures of hake” and replace with “hake eDNA”. When you mention detected hake eDNA, are you implying this is above your qPCR detection limit (e.g., 20 copies L^{-1}). I would mention this detection limit somewhere in the qPCR methods section. May be nice to report the range (and mean) of hake DNA concentrations you observed in the study.

Line 218: Spatial smoothing sounds odd. Maybe, spatial coverage?

Lines 220-222: Give a range of values and average.

Line 223: Results are vague. What is generally low? <50 copies L^{-1}? What is meant by higher and more heterogenous? Give us some values (average and range).

Lines 226-228: Confusing and seems like a discussion topic. Why do you think there was more variability between replicates at the surface? Just a more environmentally heterogeneous sample compared to deeper waters. Hake would be more abundant at the surface when you sampled eDNA at night, so this may reflect feeding behaviors on patchy prey. I wonder if including environmental variables (via CTD) would support your data or provide more context. Did you explore this?

Line 229: Still not liking the sound of “spatially smooth”.

Lines 231-232. Move to introduction.

Line 233: What is meant by strong spatial patterning?

Line 235: Move some of this to discussion.

Line 237: You present Pearson correlation data between the two indices but do not explain this in the methods. You should also report significance of the correlations (p-values).

Lines 247-248: This statement seems redundant after you provide strong Pearson correlation data.

Lines 249-251: Redundant with information in methods.

Lines 256-257: Move to discussion. You may consider combining the last two paragraphs of the results.

Lines 259-263: I may be a bit more direct here in terms of what you did and the major findings. For instance, “ Our study assessed the efficacy of eDNA sampling to quantitatively estimate hake concentration and distribution along the US west coast, a well-established fisheries habitat for this species. We sampled eDNA at 186 stations, with depth and latitude, and used a model to estimate and compare with traditional acoustic-trawl data. We found these methods were comparable across wide geographic scales (Pearson of xx), despite less eDNA samples being collected compared to acoustic transects.”

Lines 271-272: Okay, so this type of model has been used elsewhere?

Lines 275-279: You hint at this but I would emphasize that with eDNA you are non-invasively collecting seawater and applying more sensitive methods to target hake. This can be applied to other fisheries-important species within the same samples and supplement existing fisheries operations and stock assessment data.

Line 291: Do you report environmental metadata collected with the water samples (CTD data)? This may be nice to include, if not for the standard practice of reporting

metadata. Including extra variables in the models is likely beyond the scope of this study, which was aimed at comparing sampling methods, but it may be useful to interpret spatial trends in hake eDNA.

Line 292: Typical to reference figures in the discussion for this journal?

Line 307: I agree it is worth mentioning barcoding here. I would expand a bit. Incorporating both approaches would be useful, allowing for biomonitoring of the food web (barcoding), as well as more targeted approaches that inform fisheries (qPCR). Both methods can be exploited from the same collected sample and DNA samples can be archived for long periods of time, preserved or frozen. There is thus a lot of potential and upside to collecting water samples in conjunction with traditional surveys and would not add too much more effort (e.g., filtering water and storing it).

Lines 318-321: Add more detail here about what this information can provide to fisheries management and stakeholders. Expand or wrap this into a Conclusions section. What is the next step? Do you need to complete more eDNA surveys over different years/seasons and compare to trawl data? Are there other types of surveys we could compare with in this region, like camera-based monitoring? How do we translate a strong correlation between eDNA index and fish biomass to an actual fisheries-reliable metric (akin to landings or stock assessments)? Can we use this to model future populations? Add more variables to the model? Mentioning next steps would be helpful.

Lines 322-324: All files are on the GitHub page! I would think about adding more to the description (readme) to indicate what types of data files and R scripts are in the zip file. It will make it easier for those trying to access the data in the future.

Figure 1: Overall, the figures look good. The bars in panel G are a bit hard to interpret.

Figure 2: Panel letters are on the inside and smaller compared to Figure 1. I would try to be consistent. I would reference that latitudinal bins are denoted by dotted lines.

Figure 4: Panel C is hard to interpret. The colors are faint in the graph and everything is blended. You may try faceting panel C based on bottom depth. This figure or parts of it could also go into the Supplemental section. I think Figures 1-3 are most central to the story.

Supplemental Information:

Standards: Include information on qPCR standards and controls in the main text.

Investigation of contamination: Some of this information (sentence 2) is redundant with main text. I would briefly mention this in the methods/results section and refer to the Supplemental figures or additional details. Any speculation as to the sources of the contamination? I would move some of your explanation regarding the contamination to the main text. This is important to report.

Inhibition: Why so much inhibition in surface samples?

Ethanol wash error: Some of this should be in main text to provide context for the model.

eDNA model – Measures of fit and uncertainty: Regarding the 3m samples. Was the ships intake line properly sanitized or flushed? Was there a visual difference in the flow rate that may have caused DNA to shear?