

Line by line comments:

Line36: Unless I am misinterpreting your methods, I believe you should change "were seen" to "were visually estimated to be"

Line44-45: I would suggest removing "All" from the beginning of the sentence to read "Female mammals feed their...."; since all female mammals do not end up having offspring. Additionally, is there a publication prior to 2012 that you should reference?

Line73-75: Could you add a citation to support this sentence? The term "large whale" can be ambiguous as it does not always relate to only baleen whales, and odontocetes have a very different social structure and parental care- with max weaning occurring at age 13-15yrs for some species (Tyack 1986). For baleen whales, many species wean their calves before migration back to the breeding grounds, and some wean prior to arriving at the feeding grounds (Clapham 2000; Murphy 1995). Weaning occurs between 6-7months for blues (Tyack 1986), fins (Aguilar and García-Vernet 2013), grays (Sumich 2001), and 4-5 months for minke whales (Murphy 1995). For humpback whales, Chittleborough (1958) is a good reference, although even though they found calves still with milk in their stomachs after returning to the breeding grounds, separation on the feeding grounds has been documented (Baraff and Weinrich 1993; Steiger and Calambokidis 2000). I would suggest clarifying this sentence and adding citations to support your statement.

Line81-83: Could you please add in a citation to support the sentence that milk is crucial in ensuring rapid growth; therefore allowing the start of the migration

Line83-85: Could you please add in a citation to support this sentence? I know there have been several papers on the rate of individuals during migration. How much slower would a mother-calf pair go- especially if the calf was in the slipstream of mom. Also, is there previous literature about the nursing or suckling behavior of humpbacks during their migration?

Line85-88: I would suggest altering these statements to be more inclusive of individuals that wean prior to returning to the calving area and acknowledge that the mother may be helping the calf capture prey. Therefore the calf is not necessarily feeding independently. For, example:

"After its first six months of life, the calf starts to feed on prey but continues to get nutrition from milk until complete weaning, on average at 10 to 12 months old (Clapham, 2018)."

Line109-111: Could you please add a citation to support this statement? I could not find a paper that stated humpback whale calves have tongues with lingual papillae that may seal the tongue into a tube.

Line122: I am not sure what the term "cruising whales" means. Please replace with common terminology. For example, "fast-pasted whales," "sprinting whales," "whales traveling at least 5 km/hour", "whales traveling at a steady pace," etc.

Line160: Rewording the end of the sentence would improve clarity depending on whether the

overall project was looking at mother-calf interactions in general and your data was looking at the interaction on the breeding ground. For example:

"The tag deployments were conducted as part of an ongoing study on humpback whale mother-calf interactions. The tag data analyzed in our study were collected between August and September during the calving seasons of 2018 and 2019."

Vs

"The tag deployments were conducted as part of an ongoing study on humpback whale mother-calf interactions during the calving season and occurred between August and September of 2018 and 2019."

Line162: Could you add in how many individual tags you used? Was it one in 2018 and a different one that was deployed twice in 2019? It would just be interested later on about the machine learning to see if it was also compensating for differences in individual tags vs. just different individual whales.

Line235-237: Could you include a sentence or two here or in the discussion explain why you choose to only look at adjacent non-suckling periods and not a random assortment of non-suckling periods throughout each tag deployment or all 20 second periods throughout the tag deployments? It is a good method to remove the noise of other behaviors (surface activity, resting, etc.). However, it would be good to acknowledge the possible bias in your results to periods including the behavioral transition to or from nursing - not knowing if those are any different from non-suckling behavior that does not occur close to suckling behavior. For example, if you only looked at 20 secs before a whale lunged and 20 sec after, you would only capture the behavior as the whale was attempting to corral or home in on its prey and the time it was processing the food. Possibly inadvertently excluding instances of behavior associated with socializing, rest, or surface activity- behaviors that may or may not have more similar signals to your study behavior.

Line253-255: I found the use of the term "total periods" could be interpreted as the total duration of the tag deployment (data with and without video)- rather than all suckling and non-suckling periods. Possibly slightly rephrasing this sentence would improve clarity. For example:

"Contrary to the methods used to identify the behavioral signatures unique of suckling events, we included all previously identified suckling and non-suckling periods (not just adjacent periods) during the supervised machine learning process."

Line253-261: I found this section confusing, and likely a minor rewording will quickly clear up my confusion. This paragraph leads me to believe you took the 20 sec non-suckling periods and split them into 2 s blocks. Then took the suckling periods of different durations, split them into 2 s blocks, and discarded the < 2 s windows remaining in each suckling period. Additionally, and remaining track data that did not fall into the original 20 s periods were discarded. Therefore, there were no 2 s windows that contained part of a suckling and non-suckling behavior. If I am interpreting that correctly, I think it is important to include those details so that someone replicating this study with another dataset can follow your protocols correctly. If you took the whole track, split it into 2 s windows, and then determine which behavioral period each window

fell under (suckling, non-suckling, or combination of suckling and non-suckling), that would be important to state and how you handled the windows that capture the transition to or from suckling.

Line257-261: Could you briefly include why you selected the 43 features you computed for each window? A reference to your supplemental data here seems appropriate. However, I did not see 43 features listed on the spreadsheet labeled "peerj-66429-nursings_features2s_new_stat.csv", so maybe I am misinterpreting what part of the study that excel sheet relates to.

Line263-264: Did you happen to run this without excluding these blocks to see how the machine learning was affected? Since one of your suckling events occurred between 1.6 and 2.3 m, with an average of 2.0 m, I would be concerned that future tag deployments could miss suckling near the surface if they followed this model or protocol. For example, you mention the possibility that nursing may be occurring more often close to the surface at night. Additionally, nursing during migration has yet to be studied; therefore, we do not know how speed may be influenced by age or migration. It would be good to see how your models run with "noisier" data since being more inclusive may allow future studies to deal with new situations.

Line279: Based on the results, I understood that you used all three deployments for this part of the analysis and just kept sections of each deployment to be used for the 40% test set. Was there a reason you did not use just two deployments and kept the third deployment as unseen data? Therefore, all of the different models could be tested on their performance in detecting suckling despite individual whale and individual tag variations. I understand you tested this with the best model you found during this analysis, but since it was found using those three tag deployments, is it possible that using three other tag deployments may not result in the same "best" model being selected?

Line329-333: I think the clarity and flow of this sentence could be improved by splitting it into two statements. Additionally, I think you should include that you ran this analysis three times, each time leaving out a different deployment. For example:

"To assess whether the model can generalize across different individual whale calves, we also tested a leave-one-out split. We trained the model with the highest performance on $N-1$ calves' data. We then tested its performance on the unseen remaining calf data, establishing if the tag data obtained from individuals could be used to identify the suckling behavior of a different whale. We repeated this analysis three times, leaving out a different deployment each time, following the same machine learning workflow outlined in Fig. 1.

Line355: You state here that none of your calves were "newborn calves" (Line355 & 710) based on Cartwright & Sullivan (2009) findings; however, your first sentence in your abstract refers to newborns (Line23), and newborns are mentioned again in your discussion bout marginal papillae on the calf's rostral tongue (Line619-623). Clarity could be improved if you use the term neonate when referring to unfurled dorsal fins if in other sections of your manuscript you are using the term newborn to classify calves during their natal season on the breeding grounds.

Line366-368: I am slightly confused by these sentences, and likely the removal of unclear antecedents and some rewording would help prevent any confusion. What does "them" (Line366) represent? Based on the previous sentence, I would assume "them" means "suckling events." Therefore, the sentence would read: "Most of the suckling events represented a sequence of 2-6 successive events spaced less than a minute apart (N= 30, dispersed in 10 series)." Furthermore, why is N= 30 when you stated there were 34 suckling events earlier in the same paragraph? The following sentence states, "They represent 1.7% of the total observation time". Does the 1.7% represent the 30 suckling events dispersed in 10 series, including the ~1min between individual suckling events? Or do the following sentences capture what you meant? :

"We detected 34 suckling events, which lasted on average 18.8+-8.8 s, and represented 1.7% of the total video deployment. Most suckling events were clustered in a series of 2-6 suckling events, occurring less than a minute apart. Eight-eight percent (N=30) of the suckling events were recorded during ten different series (or clusters of suckling events)."

Line369-372: Is there a reason why you combined the depth results for descent and bottom instead of reporting them separately? Especially since you have them separate in the statistical analysis. Additionally, I assumed the min and max in parenthesis reported the absolute minimum and maximum depth suckling recoded at each phase. However, it appears the min and max mean depth was reported. Please clarify which results you are reporting. For example, for non-diving phase I suggest either stating "(min average = 2.0 m, max average = 8.4 m)" or "(min = 1.6 m, max = 9.2 m)".

Bottom: aver 20.2, min =10.6, max= 34.6, aver min =11.1, aver max= 32.6

Descending: : aver 16.2, min= 8.5, max= 28.4, aver min =8.9, aver max= 26.9

Line374-375: Since this was just a visual estimate from the calf's perspective, how could you tell the midline of the ventral side of the mother was a good representation for the pitch of the mother? Is there a paper you could cite that references a tagged mother's actual pitch based on the camera angle of the calf?

Line375-376: It would be beneficial, and strengthen your argument, to split the recorded pitch (pitch angle relative to the horizontal) up into the average pitch for the 3 phases rather than just reporting it as "fairly low on average (< 15°)", especially since the two more horizontal phases (bottom and nontdiving) are more different (-7.9 vs. -15.12) than the descending phase (-13.4). Not that you have to analyze it further- but it can provide a focus for future studies. Additionally, a quick look at your data, and it seems like the descend of a calf while nursing compared to non-nursing dives is quite different from adjacent periods. Therefore, you may lose some of the signal or possible significant results if you are only comparing periods of adjacent behavior to nursing events. Again, I am not suggesting a reanalysis, but it would be good to acknowledge that adjacent and non-adjacent periods are not necessarily the same.

Phase	Behavior	Total #	Aver. pitchmean	Aver. pitchmin	Aver. pitchmax
Descending	Nursing	9	1.6	-13.4	15.6
Descending	Nursing dive	15	-15.5	-38.7	9.3

Descending	Non nursing dive	5	-32.9	-53.7	-4.2
Bottom	Nursing	20	7.9	-7.9	24.0
Bottom	Nursing dive	7	6.3	-9.7	22.6
Bottom	Non nursing dive	17	6.9	-12.8	27.7
notdiving	Nursing	5	5.1	-15.12	23.8
notdiving	surface	29	2.6	-27.6	33.2

Line378-381: Did calves always start a series of suckling by rolling to the right first, as the wording of this sentence suggests? Or was the first side not consistent? That information would be interesting to clarify since handedness has been suggested/proven for this species, with righthandedness the dominant side (Zoidis and Lomac-MacNair 2017; Canning et al. 2011).

Line386: Is there a reason for the use of "leaning" vs. "rolling." Does lean describe visual observations in the video, whereas roll references the tag data? It would be helpful if there is a difference that you include clarify that in the methods.

Line389-390: This sentence could use some minor rewording to improve clarity, such as:

"During our tag deployments, non-suckling dives in which the calf remained beneath the mother for at least consecutive 5 s were more common than suckling dives".

Line389-391: It would also make this section stronger if you could report the frequency, or percentage of total tag deployment, of these close-proximity events when suckling did and did not occur. For example, the calf was in close proximity to the mother ventral side without nursing for XX% of the total tag deployment. The percentage would allow for the comparison you reference in your discussion (Line531-533), which currently is not comparing apples to apples (# of dives vs. % of the time)- see my comment below.

Line390-391: This sentence should be reworded. Currently, it reads as if the positioning of the calf does not necessarily relate to suckling; however, your results state that the calf's position while suckling is unique enough to be used in machine learning. If I interpret your results correctly, I think rewording it would increase consistency throughout your manuscript. For example,

"In other words, using calf proximity to the mother's ventral side alone does not indicate that it is suckling is occurring. Additional accelerometry singles, which are described below, are necessary to determine the occurrence of the behavior."

Line392-430: You refer to Table 2 throughout this section; however, it appears that it is supposed to be Table 3.

Line411-412: I am slightly confused by this sentence. Is it supposed to read, "In other words,

when suckling, the calves displayed a FSR and an ODBA that were comparable to adjacent non-suckling periods during descent, bottom, and surface phases"?

Line429-430: Similar comment as above. I am confused by the part of the sentence which says, "comparable to adjacent bottom non-suckling during descent, bottom, and surface suckling." Is it supposed to say, "the roll rate during adjacent bottom non-suckling events were comparable to descent, bottom, and surface suckling events? Also, why are not adjacent surface non-suckling events and adjacent descent non-suckling events absent in the reporting of these statistical tests?

Line456-457: If I understand your results correctly, your results suggest the AdaboostM1 model may incorrectly label 30% of the suckling events as non-suckling events. Since suckling appears to be a somewhat rare behavior, underestimating it by 30% could be significant for future studies would need to be addressed/accounted for; therefore, I would suggest removing the word "slightly" in Line 457.

Line531-533: I would suggest rewording this sentence. Your study only determined the number of dives in which the calf was under the mother for at least 5 s while not suckling, not the frequency of the calf positioning under the mother or the amount of time the calf spent under their mother while not suckling as in Tackaberry et al. (2020). Unless you report on the frequency of time the calf spent in close proximity to the mother without nursing, I think it is difficult to compare these results to Tackaberry et al. (2020).

Line533-536: I would suggest making sure you are clear when you use the term "positioning." You use it to say the calf's position does not necessarily relate to suckling (Line532 & Line535-536), but at the same time, you are describing the calf's position (Line582) in reference to the pitch angle. Being more specific (i.e., making sure to reference the calf's position concerning the mother vs. the calf's position in the water column) will improve clarity and remove room for misinterpretation.

Line663-665: Why would this be costlier? The question that the model is being used to answer will influence if an overestimation is worse than an underestimation. Underestimating the amount of suckling occurring by 30% would lead to a miss conception of the amount of suckling required for a growing calf or the amount of milk transferred per suckling event to explain the calf growth rate. Also, casting a wider net is more efficient in some cases, allowing someone to manually audit the data and correct the model's mistakes. I would recommend softening this sentence to be more inclusive of future uses of your model.

Line713-728: I would soften your conclusions a little. In Line723 you say the behavior is "highly stereotyped," yet you only found that absolute role and lower speed with equivalent FSR and ODBA were significant when comparing your data to adjacent non-suckling data that including the transition to and from nursing (it did not include all non-suckling periods in your dataset which may have similar signals). Additionally, you say it can be detected automatically, yet your best model missed 30 % of the suckling periods. I think your conclusion would be more robust and accurate if you acknowledged that your results are only based on three tag deployments on humpback whales in the breeding ground. I think additional videos from tag deployments in the breeding grounds are required to strengthen your model before you could confidently apply your

model to tags without the ability to verify the behaviors. I would suggest highlighting that your study is proof of method and that other studies can now build upon your results. I would hesitate to say, "...and will be useful in studying other baleen whales for which such vital behavior is still undiscovered", but maybe soften it to say this study should be replicated with other baleen whale species.

Figures & Tables:

Figure1- It would be nice for the caption to have a little more detail, such as how the suckling and non-suckling behavior is actually 2 s windows of track data. Providing the sample size of each section of the three pie charts would also strengthen the figure's impact.

Figure2- I like this figure and all the information you were able to include. It might not be possible, but increasing the font size of the labels would be a bonus. For example, the seconds might not be necessary on the x-axis, so possibly just having hh: mm would allow for slightly larger text. It would also be an added bonus if you could show the different dive phases (descent, bottom, ascent, surface) in the depth plots using a shaded window or having a yellow box designate suckling instead of the red line and use line color to designate dive phase.

Figure4- In the caption, it would be good to use the exact wording used in the text- suckling and adjacent non-suckling periods, rather than events. It would be good to include the average duration of the suckling periods and 20 s for the adjacent suckling periods to ensure the reader is not confusing these periods with the 2 s windows used in the machine learning. For example: "Comparison of suckling periods (average 18.8 s) and adjacent non-suckling periods (20 s) with respect to activityphases." In the figure, it is crucial to alter the y-axis labels to include mean, max, or min, depending on what tag data you used to make these plots.

Figure5 &6: I suggest including the detail that all three tag deployments were used in the 40:60 holdout training-test split.

Table1: I would clarify that you are reporting the minimum mean and the maximum mean, in your format (mean $\pm$ SD (min, max)) compared the absolute min and absolute max recorded for the features for each calf. See my similar comment for Line369-372

Table2: I suggest using the phrase or one similar to: "calf observed staying in close proximity beneath the mother for at least five consecutive seconds during the dive." I believe the phrase "at least 5 s at least once" is slightly confusing, and the phrase "five consecutive seconds" may prevent misinterpretation by the reader.

Table3: Like my comment for Figure 4, please include mean, max, or min for each signal, depending on the data you used for this analysis.

Minor suggestions:

1. Unclear antecedent. Many sentences begin with "this" and "it," but it is unclear what that refers to based on the previous sentence. I would suggest replacing "this" or "it" with a subject that improves the clarity of the sentence. I have included a few examples in the

more detailed line references below, but I did not include all of them.

2. Use of parentheses. There are many instances when you place important or helpful information in parentheses; however, it can muddy the overall point of the sentence or increase its length to become more complex and harder to read. I would suggest reviewing the sentences to determine if the information in the parentheses is essential and whether it should receive its own sentence. Alternatively, a glossary of terms would allow you to provide definitions of the terms you are using without cluttering important sentences with additional information.
3. I found that there were several sentences in which unnecessary commas appeared to be used. I understand if you prefer to use them for emphasis. However, I would suggest you take a second look to determine which commas are unnecessary grammatically, and their removal could improve flow or sentence structure.

Line23-26: I would suggest splitting the second sentence into two statements to improve clarity and flow:

"Despite its importance, nursing has been poorly documented in humpback whales (*Megaptera novaeangliae*). Nursing is difficult to observe underwater without disturbing the whales and usually impossible to observe from a ship."

Line29: Unclear antecedent- "This," and remove "any":

"The use of bio-loggers minimized potential disturbance from human presence."

Line41-42: To improve clarity, I would suggest changing "and in baleen whales in general." to "and the baleen whale species."

Line46: This sentence could use some rewording to improve clarity and make sure the punctuation is in the correct order: ['.] Vs ['.]. For example:

"The maternal behavior associated with the transfer of milk to its offspring is referred to as 'nursing.'"

Line49-53: To improve flow and clarity, I suggest replacing the unclear antecedent- "This" with "Suckling" and splitting the sentence into two statements.

"Suckling can involve an active role of mouth movements from the young to stimulate nipple erection, compress the breast's lactiferous sinuses, and generate negative pressure (suction) along with a pulling motion on the teat to withdraw the milk. Alternatively, suckling can involve a passive role in which milk is squirted into the open mouth with no stimulating movements of the offspring's mouth."

Line61-64: This sentence is slightly confusing as you say many animals have two rows of paired breasts, yet your examples have four-seven pairs. Also, the use of "while" in reference to location sounds like you are comparing the number to the location. The previous sentence provides the reader with enough information (breasts are always in pairs and can range from the axilla to the inguinal region. However, if you want to show the variation among species, I suggest splitting or rewording the sentence to improve clarity. For example:

"Among different species of mammals, the number of rows of paired breasts ranges from one to seven. These rows are concentrated in the chest region of some species (e.g., primates, elephants, manatees) while they are concentrated in the inguinal region in other species (e.g., horses, giraffes, hippopotamuses)."

Line 118-120: Split this sentence into two statements may improve clarity and flow. For example:

"ROV's and AUV's have potential, but presently vehicles are not yet agile enough to track whales. Presently, they are too slow and difficult to maneuver, with ROV's also being limited by cable length."

Line 129-132: I would suggest splitting the sentence into two statements to improve clarity and flow. For example:

"The development of animal-borne multi-sensor tags, such as Customized Animal Tracking Solutions (CATS) cam tags, provided a way for remote and accurate recording of whale behavior (Cade et al., 2016). Bio-loggers can be equipped with a suite of sensors including cameras, accelerometers, gyroscopes, magnetometers, pressure (depth) sensor, temperature sensor, light sensor, and hydrophone."

Line 133-135: Please clarify "these models." Possibly rephrasing the beginning of the sentence would help, for example:

"CATS tags are particularly well suited to study suckling behavior since, in addition to kinematic sensors, they"

Line 185-186: The use of the word "immediately" may lead readers to think that you resumed following the pair at some point later on. If this is not true, I would suggest removing the word "immediately."

Line 189: I may suggest changing the word "jump" to something more general or possibly translates better between study regions, such as "surface active behavior."

Line 376-378: To improve clarity, I would suggest rewording this section. For example:

"During all the observed suckling events, rolls were recorded in the data as a sustained deviation from zero during the event (Fig. 2). In all but one instance, the calf alternation between suckling with a right roll then rotating to suckle with a left roll."

Line 527-549: I think rearranging this section would improve flow and clarity. Starting the section with 537-549 would lead nicely into your discussion about your results Line 527-536.

Line 599-601: I would suggest moving these sentences to the second line of the previous paragraph (Line 590) when you are talking about leaning while suckling compared to the altering of sides, which you discuss in the current paragraph.

Line630-633: This sentence may benefit from some rewording to improve clarity. If I am interpreting it correctly:

"The comparison of the data derived from the accelerometer and depth sensor for suckling and adjacent non-suckling events revealed mainly two distinctive characteristics of the suckling behavior: a high absolute roll and a lower speed despite having similar FSR and ODBA as adjacent non-suckling events."

References

- Aguilar, Alex, and Raquel García-Vernet. 2013. "Fin Whale: *Balaenoptera Physalus*." In *Encyclopedia of Marine Mammals*, edited by Bernd Würsig, J.G.M. Thewissen, and Kit M. Kovacs, Third, 368–71. Academic Press. <https://doi.org/https://doi.org/10.1016/B978-0-12-804327-1.00128-X>.
- Baraff, Lisa, and Mason T. Weinrich. 1993. "Separation of Humpback Whale Mothers and Calves on a Feeding Ground in Early Autumn." *Marine Mammal Science* 9 (4): 431–34. <https://doi.org/10.1111/j.1748-7692.1993.tb00476.x>.
- Canning, Claire, Danielle Crain, T. Scott Eaton, Kathryn Nuessly, Ari Friedlaender, Tom Hurst, Susan Parks, Colin Ware, David Wiley, and Mason Weinrich. 2011. "Population-Level Lateralized Feeding Behaviour in North Atlantic Humpback Whales, *Megaptera Novaeangliae*." *Animal Behaviour* 82 (4): 901–9. <https://doi.org/10.1016/j.anbehav.2011.07.031>.
- Chittleborough, RG. 1958. "The Breeding Cycle of the Female Humpback Whale, *Megaptera Nodosa* (Bonnaterre)." *Marine and Freshwater Research* 9 (1): 1–18. <https://doi.org/https://doi.org/10.1071/MF9580001>.
- Clapham, Phillip J. 2000. "The Humpback Whale." In *Cetacean Societies, Field Studies of Dolphin and Whales*, edited by Janet Mann, Richard C Connor, Peter L Tyack, and Hal Whitehead, 173–96. Chicago: University of Chicago Press.
- Murphy, Margaret A. 1995. "Occurrence and Group Characteristics of Minke Whales , *Balaenoptera Acutorostrata* , in Massachusetts Bay and Cape Cod Bay" 93 (3): 577–85.
- Steiger, Gretchen H, and John Calambokidis. 2000. "Reproductive Rates of Humpback Whales off California." *Marine Mammal Science* 16 (1): 220–39. <https://doi.org/https://doi.org/10.1111/j.1748-7692.2000.tb00914.x>.
- Sumich, J L. 2001. "Growth of Baleen of a Rehabilitating Gray Whale Calf." *Aquatic Mammals* 27 (y): 234–38.
- Tyack, Peter. 1986. "Population Biology, Social Behavior and Communication in Whales and Dolphins." *Trends in Ecology & Evolution* 1 (6): 144–50. [https://doi.org/https://doi.org/10.1016/0169-5347\(86\)90042-X](https://doi.org/https://doi.org/10.1016/0169-5347(86)90042-X).
- Zoidis, Ann M., and Kate S. Lomac-MacNair. 2017. "A Note on Suckling Behavior and Laterality in Nursing Humpback Whale Calves from Underwater Observations." *Animals* 7 (7): 1–12. <https://doi.org/10.3390/ani7070051>.