**Post Publication work:**

**Last Updated 18/08/2017**

This document explores some additional analysis I’ve undertaken with the classification system since we submitted the work to JASA. It’s possible that the work presented will eventually be made into a JASA note to follow on our initial work. However, probably won’t have time to write it up properly and I’d rather the information be available than not. So, take note, this work has not been peer reviewed and very possibly contains a few mistakes.

One of the important criticisms of this work is that we could not test the system performance on out-of-sample data for the acoustic encounters. In 2013 we simply did not have enough acoustic encounters to run this analysis. However, in the interim the ECoMMAS array has been deployed each year and we have data from 2014-2016 (the 2017 data are still being collected). Since the initial submission of this manuscript I’ve gone through the 2014 and 2015 deployments looking for good examples of broadband and frequency banded click train encounters that could be used to test out-of-sample performance. The results of this work are represented in the confusion matrix, **Table 1**. If you aren’t familiar with confusion matrices, the rows represent the number of true observations (here there were 71 verified broadband encounters) and the columns represent how the observations were classified (15 of the broadband encounters were incorrectly classified as frequency banded).

**Table 1**- Confusion matrix for out-of-sample categorisation performance

|  |  |  |  |
| --- | --- | --- | --- |
|  | GAM/Likelihood Categorization (Out of Sample) | | |
| Verified Category | Broadband | Frequency Banded | Unknown |
| Broadband | 56 | 15 | 0 |
| Frequency Banded | 2 | 4 | 0 |

A big problem that we had in 2014 and 2015 were the low numbers of frequency banded click trains that could be detected by both the C-POD and the adjacent SM unit. This was the opposite problem that we had in the 2013 data where we could find lots of frequency banded click train encounters but relatively few broadband click train encounters. So, it’s still difficult to determine exactly how well the categorisation system will perform on out-of-sample frequency banded encounters.

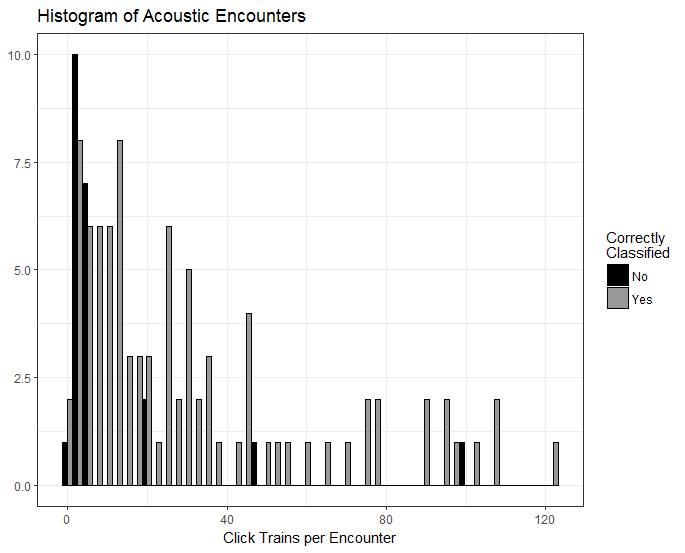
If we combine the in-sample (e.g. 2013 data) and out of sample (2014 and 2015) this is how the classifier did with in and out of sample data together (still less than ideal). **Table 2** shows the confusion matrix for all available acoustic encounters.

**Table 2**: Confusion matrix for all available acoustic encounters

|  |  |  |  |
| --- | --- | --- | --- |
|  | GAM/Likelihood Categorization (All) | | |
| Verified Category | Broadband | Frequency Banded | Unknown |
| Broadband | 66 | 17 | 0 |
| Frequency Banded | 2 | 23 | 3 |

Another thought raised by the reviewers is that the classifier might perform better with a greater number of click trains in the encounter. We weren’t able to properly address the concern in the paper but I’ve since looked at this using the 2013-2015 dataset. Again, these results have the same issue as above: most of the frequency banded encounters represent in-sample performance.

**Figure 1** Histogram of the number of trains per encounter and the whether the encounter was correctly classified using this system



I haven’t done extensive analysis on this but hopefully **Figure 1** shows that, yes, the performance of the categorisation system generally improves with the number of click trains in the encounter. Below, ~5 trains per encounter it’s a toss-up with whether or the system will correctly classify the encounter. Above, 10 click trains per encounter the system performs considerably better.