

MEASURING INTER-PULSE INTERVALS IN SPERM WHALE CLICKS:  
DEVELOPMENT OF AN AUTOMATIC METHOD AND ITS POTENTIAL FOR  
IDENTIFYING EASTERN CARIBBEAN SOCIAL UNITS

by

Wilfried A. M. Beslin

Submitted in partial fulfilment of the requirements  
for the degree of Master of Science

at

Dalhousie University  
Halifax, Nova Scotia  
August 2018

© Copyright by Wilfried A. M. Beslin, 2018

*Pour ma famille*

# TABLE OF CONTENTS

|   |             |
|---|-------------|
| <b>LIST OF TABLES</b> .....   | <b>vi</b>   |
| <b>LIST OF FIGURES</b> .....  | <b>vii</b>  |
| <b>ABSTRACT</b> .....   | <b>viii</b> |
| <b>LIST OF ABBREVIATIONS USED</b> .....   | <b>ix</b>   |
| <b>ACKNOWLEDGEMENTS</b> .....   | <b>x</b>    |
| <b>CHAPTER 1 - INTRODUCTION</b> .....   | <b>1</b>    |
| 1.2 SPERM WHALE VOCALIZATIONS .....   | 2           |
| 1.2.1 The Inter-Pulse Interval .....  | 3           |
| 1.2.2 On-Axis versus Off-Axis Clicks.....   | 4           |
| 1.3 SPERM WHALE SOCIAL STRUCTURE .....  | 6           |
| 1.4 EASTERN CARIBBEAN SPERM WHALES .....  | 7           |
| 1.5 THESIS OBJECTIVES.....  | 8           |
| <b>CHAPTER 2 - AUTOMATIC ACOUSTIC ESTIMATION OF SPERM WHALE BODY LENGTHS<br/>ACHIEVED THROUGH MACHINE RECOGNITION OF ON-AXIS CLICKS</b> ..... | <b>10</b>   |
| ABSTRACT .....  | 10          |
| 2.1 INTRODUCTION .....  | 10          |
| 2.2 METHODS .....   | 12          |
| 2.2.1 Data Collection .....   | 12          |
| 2.2.2 Software Design Overview .....  | 14          |
| 2.2.3 Audio Loading and Preprocessing .....   | 15          |
| 2.2.4 Click Detection.....  | 15          |
| 2.2.5 Feature Extraction .....  | 17          |
| 2.2.5.1 <i>Pulse Detection</i> .....  | 22          |
| 2.2.5.2 <i>Frequency Spectrum Calculation</i> .....   | 24          |
| 2.2.5.3 <i>Exponential Fitting</i> .....  | 25          |
| 2.2.6 Click Classification.....   | 25          |
| 2.2.7 IPI Calculation and Validation.....   | 29          |
| 2.2.8 Animal Count and Length Estimation .....  | 31          |
| 2.2.9 Performance Analysis .....  | 32          |
| 2.3 RESULTS.....  | 34          |
| 2.3.1 Classifier Performance.....   | 34          |
| 2.3.2 Output IPI Distributions.....   | 34          |

|   |           |
|---|-----------|
| 2.3.3 Differences between Recording Scenarios.....  | 38        |
| 2.4 DISCUSSION.....   | 41        |
| 2.4.1 Performance.....  | 41        |
| 2.4.1.1 <i>Differences between Recording Scenarios</i> .....  | 45        |
| 2.4.2 Applications.....   | 47        |
| 2.5 CONCLUSIONS.....  | 48        |
| CHAPTER 2 ENDNOTES.....   | 49        |
| <b>CHAPTER 3 - EVALUATION OF A METHOD FOR AUTOMATICALLY IDENTIFYING EASTERN CARIBBEAN SPERM WHALE SOCIAL UNITS FROM THEIR INTER-PULSE INTERVAL DISTRIBUTIONS.....</b> | <b>50</b> |
| ABSTRACT.....   | 50        |
| 3.1 INTRODUCTION.....   | 50        |
| 3.2 METHODS.....  | 53        |
| 3.2.1 Data Collection.....  | 53        |
| 3.2.2 Algorithm.....  | 54        |
| 3.2.2.1 <i>Construction of Unit IPI Profiles</i> .....  | 54        |
| 3.2.2.2 <i>Inferring Unit Presence</i> .....  | 57        |
| 3.2.3 Performance Analysis.....   | 58        |
| 3.3 RESULTS.....  | 60        |
| 3.3.1 Unit IPI Profiles and Ability to Differentiate Between Units.....   | 60        |
| 3.3.2 Overall Performance.....  | 62        |
| 3.3.3 Performance by Unit Presence Scenario.....  | 64        |
| 3.4 DISCUSSION.....   | 68        |
| 3.4.1 Unit Recognition.....   | 69        |
| 3.4.1.1 <i>Problems with Unit A</i> .....   | 69        |
| 3.4.1.2 <i>Problems with IPI Profiles in General</i> .....  | 70        |
| 3.4.2 Algorithm Performance.....  | 71        |
| 3.4.3 Potential Improvements and Alternatives.....  | 72        |
| 3.5 CONCLUSIONS.....  | 75        |
| <b>CHAPTER 4 - CONCLUSION.....</b>  | <b>77</b> |
| 4.1 SUMMARY OF PROJECT OUTCOME.....   | 77        |
| 4.1.1 Automatic IPI Compilation.....  | 77        |
| 4.1.2 Automatic Unit Identification.....  | 79        |

|  |            |
|--|------------|
| 4.2 APPLICATIONS AND FUTURE DIRECTIONS .....   | 80         |
| <b>REFERENCES .....</b>  | <b>82</b>  |
| <b>APPENDIX A - ACTIVITY DIAGRAMS FOR AUTOMATIC IPI COMPILATION ALGORITHM</b><br>.....           | <b>91</b>  |
| <b>APPENDIX B - LIST OF PARAMETERS FOR AUTOMATIC IPI COMPILATION ALGORITHM</b><br>.....          | <b>107</b> |
| <b>APPENDIX C - DATASET FOR TESTING AUTOMATIC UNIT DETECTION .....</b>                           | <b>117</b> |
| <b>APPENDIX D - RESULTS OF AUTOMATIC UNIT DETECTION FOR ALL RECORDING</b><br><b>GROUPS .....</b> | <b>122</b> |
| <b>APPENDIX E – REFERENCES FOR APPENDICES .....</b>  | <b>138</b> |

## LIST OF TABLES

|   |    |
|---|----|
| <b>Table 2.1</b> Description of features used for classifying clicks as being on-axis sperm whale clicks or not .....       | 19 |
| <b>Table 2.2</b> Metrics used to assess the performance of automatic classification, relative to manual classification..... | 27 |

## LIST OF FIGURES

|   |    |
|---|----|
| <b>Figure 1.1</b> Example sperm whale clicks, showing multi-pulsed structure.....   | 6  |
| <b>Figure 2.1</b> Example of features used in the detection of pulses within a click. ....  | 24 |
| <b>Figure 2.2</b> Example click waveforms examined for construction of the click classifier training dataset and their assigned labels. ....  | 28 |
| <b>Figure 2.3</b> Example IPI distributions output from 4 “standard” recordings from Dominica. ....   | 36 |
| <b>Figure 2.4</b> Performance of automatic IPI compilation for 174 “standard” recordings from Dominica. ....  | 37 |
| <b>Figure 2.5</b> Automatic IPI compilation compared to manual compilation for 9 “standard” recordings. ....  | 38 |
| <b>Figure 2.6</b> Example IPI distributions output from 4 “standard” recordings from the Galápagos.....   | 40 |
| <b>Figure 2.7</b> Performance of automatic IPI compilation compared between 174 “standard” recordings from Dominica, and 141 “standard” recordings from the Galápagos....                       | 41 |
| <b>Figure 3.1</b> Example of IPI profile creation using kernel density estimation.....  | 56 |
| <b>Figure 3.2</b> IPI distribution profiles for the generic UNK and MALE groupings. ....  | 57 |
| <b>Figure 3.3</b> IPI distribution profiles for common Dominica sperm whale units. ....   | 61 |
| <b>Figure 3.4</b> Proportion of correct (dark) versus incorrect (light) identifications of the target unit based on the “Singles-Singles” test. ....  | 62 |
| <b>Figure 3.5</b> Cumulative probability distributions of the $\Delta$ BIC scores of the “true” models for the “All-All” test when unit A was considered known (left) and unknown (right). .... | 63 |
| <b>Figure 3.6</b> Precision and recall of unit detection for the “All-All” test when unit A was considered known (left) and unknown (right).....  | 64 |
| <b>Figure 3.7</b> Precision and recall of unit detection for the “All-All” test when unit A was considered known (left) and unknown (right), broken down by scenario. ....                      | 66 |
| <b>Figure 3.8</b> Trend in performance of the automatic unit detection routine with number of units actually present (as determined by photo-ID).....   | 67 |
| <b>Figure 3.9</b> Trend in performance of the automatic unit detection routine with number of IPIs used to make the inference. ....   | 68 |

## ABSTRACT

Sperm whale clicks have a unique multi-pulsed structure, where the inter-pulse interval (IPI) is related to body length. This feature makes passive acoustic monitoring especially informative for sperm whales. I investigated methods for analysing IPIs automatically from raw audio recordings. First, I developed an algorithm for compiling reliable IPI measures. I then used this algorithm in an attempt to identify social units of sperm whales in the Eastern Caribbean.

In practice, IPIs are difficult to measure, because the multi-pulsed structure is clear only when clicks are recorded along the whale's longitudinal axis. To achieve automatic IPI compilation, I trained a support vector machine (SVM) to recognize probable on-axis sperm whale clicks. The compilation algorithm uses the SVM to isolate "Good" clicks, whose IPIs are then validated based on precision and repetition. This method was found to be successful in producing IPI distributions with precise peaks that likely corresponded to individual whales. These peaks could also be resolved using Gaussian mixture models, thereby providing automatic estimates how many whales were present and how large they were. However, the routine rejected a considerable number of clicks (> 99% on average). Nevertheless, since sperm whales click regularly, audio recordings lasting at least 10 minutes are likely to yield adequate IPI distributions for the vocalizing whales.

The second objective of my thesis was to automatically identify permanent sperm whale social units (with about 3-12 members each) based on IPIs. I established IPI profiles for 5 units commonly encountered off Dominica, and used these to infer unit presence from empirical IPI distributions. IPI profiles showed some potential for recognizing individual units, but the routine successfully identified all units present in only about 30% of cases. Groupings of units and the presence of unknown individuals were particularly problematic.

The automatic IPI compilation algorithm developed here should make it possible to assess length distributions of sperm whales from passive acoustic surveys, even when multiple whales are clicking simultaneously. IPIs also have potential for identifying social units, which would be particularly useful for monitoring the declining Eastern Caribbean population. However, the unit detection routine implemented here needs to be improved.

## **LIST OF ABBREVIATIONS USED**

IPI - Inter-pulse interval

ICI - Inter-Click Interval

SNR - Signal-to-noise ratio

RMS - Root mean square

SVM - Support vector machine

GMM - Gaussian mixture model

EM - Expectation-Maximization

KDE - Kernel Density Estimation

BIC - Bayesian Information Criterion

OVL - Overlapping Coefficient

$\Lambda$  – Likelihood ratio

m - metre

km - kilometre

ms - Millisecond

$\mu$ s - Microsecond

F<sub>s</sub> - Sampling Rate

kHz - Kilohertz

dB - Decibel

$\mu$ Pa – Micropascal

FFT – Fast Fourier Transform

ZCR – Zero Crossing Rate

## ACKNOWLEDGEMENTS

Hal: thank you so very much for having faith in me and taking me on as a student. My time in your lab has presented me with invaluable knowledge, connections, and experiences that I have always dreamed of. Thank you for being such a kind, encouraging, and supportive supervisor.

Shane: my time working with you in Dominica was the most memorable experience I have had to date. Thank you so much for letting me participate in your DSWP research, and for introducing me to the island of Dominica and the whales that inhabit its waters. During my honours thesis, my only evidence that these whales existed was from journal articles and pieces of sloughed skin in DMSO vials. Since then, I've always wondered what it would be like to meet them personally, to see them socialize with their companions and watch them do what sperm whales do. You and Hal made it happen. That really means a lot to me.

Thank you to the people I met in the field, both on *Balaena* and *Flying Fish*, for being such great companions: Peter Madsen, Michael Ladegaard, Mafalda de Freitas, and Pernille Tønnesen. Big thanks also to the people of Dominica for being so kind and accommodating, especially Pernell Francis and Captain Dave Fabien.

Luke Rendell, Alex Hay, and Andy Horn: thank you for all your advice, and for believing that I had what it takes to complete this thesis. Thanks also to Luke Rendell, Alex Hay, and Brian Miller for affirming that I am worthy of the title Master of Science, and for helpful comments on the completed manuscript.

Whitehead lab: your friendship and support over the years has been truly invaluable. Mauricio: thank you for being an amazing desk neighbour, for your advice on grad school, and for all the pep talks. Laura: thank you for continuing Mauricio's legacy as both desk neighbour and advisor. Manolo: thank you for the interesting psychological discussions, the much-needed coffee, and for keeping me busy with frogs and plants. Christine: thank you for all the social events, the adventures, and your general companionship throughout the Master's program from the very beginning. Ana: thank

you for sharing lunch breaks with me and being such a good listener! Elizabeth: thank you for sharing your bottomless knowledge and passion on all things cetacean. Taylor: thank you for ensuring the lab stayed organized and for making lab meetings run smoothly. Dara: thank you for sharing your research with us so enthusiastically and unapologetically. I'm going to miss you all!

To the other members of the Biology Organization of Graduate Students: thank you for your friendship, advice, and social events, and for teaching me new skills. I never imagined that I would walk out of the graduate program as a certified bartender.

Thank you to the many people who have provided helpful resources and advice throughout this project. Kristian Beedholm and Peter Madsen: thank you providing code when I was working to understand click detection. Gabrielle Macklin and Marie Ryan: thank you for lending me your time during the early stages of click classification. Stan Matwin and Erico Neves De Souza: thank you for your advice on classification and machine learning. Charlotte Dunn: Thank you for beta-testing my software and providing useful feedback.

To my own family: your support for me is without end. Thank you so much for all your love and help during the completion of my thesis, and for giving me a place to live.

This work would not have been possible without the financial support I received from the Nova Scotia Research and Innovation Graduate Scholarship, the NSERC Canada Graduate Scholarship, and the Patrick F. Lett Fund.

Finally, the sperm whales deserve to be acknowledged too: were it not for you, this thesis would literally not have happened. Thank you for being so patient with us (and for aligning yourselves with our hydrophone from time-to-time). I hope this thesis will help to protect you in the future.

# CHAPTER 1 - INTRODUCTION

## 1.1 PASSIVE ACOUSTIC MONITORING OF CETACEANS

Passive acoustic monitoring (PAM) has proved to be an extremely useful technique for studying marine mammals. While visual surveys provide a great deal of information, cetaceans, which spend most of their lives underwater and sometimes far from shore, can be expensive and challenging to monitor visually. In contrast, these animals are extremely vocal, and sound propagates particularly well through water. Because of this, PAM has enabled researchers to detect cetaceans that may be several kilometres away (Barlow and Taylor, 2005; Stafford *et al.*, 2007), in conditions that may be unfavourable for visual observations (Thomas *et al.*, 1986b). Acoustic monitoring also has the advantage that call detection can often be automated. In the case of fixed PAM, where cabled or autonomous recorders are deployed, data can also be collected continuously or near-continuously over long periods of time (Mellinger *et al.*, 2007).

The most basic information provided by PAM is detection: whether an animal is present or absent. This is frequently used to establish the distribution or seasonal occurrence of a species (e.g. Stafford *et al.*, 2007; Norris *et al.*, 1999; Gordon *et al.*, 2000; Verfuß *et al.*, 2007; Gedamke and Robinson, 2010; Hannay *et al.*, 2013). However, unless complex hydrophone arrays are available, inference from PAM alone can be limited. Thus, PAM surveys must often be supplemented with visual information. Abundance estimation, for example, requires estimates of detection distances and/or group sizes (Buckland *et al.*, 2001; Mellinger *et al.*, 2007), which are most easily obtained visually. Another limitation of PAM is that the recognition of individuals is usually very difficult, if not impossible. For this purpose, the photography of natural markings is a standard protocol for cetaceans (see Hammond *et al.*, 1990).

## 1.2 SPERM WHALE VOCALIZATIONS

The sperm whale (*Physeter macrocephalus*) is particularly well suited to study through passive acoustics. Sperm whales are wide-ranging, pelagic animals that spend most of their time foraging at depths of over 300 m, returning to the surface for about 10 minutes every 30-45 minutes or more (Watkins *et al.*, 1993; Papastavrou *et al.*, 1989; Watwood *et al.*, 2006). When they are out of sight however, sperm whales are extremely conspicuous acoustically. Their vocalizations consist almost entirely of loud, impulsive clicks, which they use to hunt for prey and communicate. With a maximum source level of 236 dB re 1  $\mu$ Pa (RMS, at 1 m), sperm whale clicks hold the record for the loudest biological sound ever recorded (Møhl *et al.*, 2003). Sperm whale clicks have been found to be detectable typically up to 9 km, and sometimes up to almost 40 km (Barlow and Taylor, 2005). Because of the long and frequent dives of sperm whales and the high audibility of their clicks, the incorporation of passive acoustics into sperm whale surveys can significantly increase the number of animals detected (Barlow and Taylor, 2005).

Sperm whale clicks come in different types, which differ based on their structure, rate of production, and the behavioural context in which they are produced. “Usual” clicks (Weilgart and Whitehead, 1988) are by far the most common. These are highly directional clicks (Møhl *et al.*, 2000) emitted in trains at regular rates of about 1-2 clicks/seconds (Backus and Schevill, 1966; Whitehead and Weilgart, 1990). Because of their high directionality and constant rate of production while at depth, usual clicks are presumably used for echolocation (Mullins *et al.*, 1988; Madsen *et al.*, 2002b). Another fairly common type is the “coda” click, which makes up the stereotyped coda calls used by females and juveniles to communicate (Watkins and Schevill, 1977; Mullins *et al.*, 1988; Weilgart and Whitehead 1993; Marcoux *et al.*, 2006; Schulz *et al.*, 2008; Gero *et al.* 2016b), but is of lower source level than the usual click (Madsen *et al.* 2002a). Other types include “slow” clicks, emitted by males presumably for courtship and/or competitive displays (Weilgart and Whitehead, 1988; Madsen *et al.*, 2002b), and “creak” clicks, which are relatively quiet clicks emitted at high rates that may be used

for homing in on prey (Gordon, 1987; Goold and Jones, 1995; Jaquet *et al.*, 2001; Madsen *et al.*, 2002b; Miller *et al.*, 2004).

Coda clicks are the focus of many studies that investigate the social structure and behaviour of sperm whales (e.g. Rendell and Whitehead, 2003; Rendell and Whitehead, 2004; Marcoux *et al.*, 2006; Schulz *et al.*, 2008; Antunes *et al.*, 2011; Schulz *et al.*, 2011; Gero *et al.*, 2016a; Gero *et al.*, 2016b). For survey purposes, however, usual clicks are most important, due to their power and near-omnipresence whenever a sperm whale of either sex is present. Thus, this thesis focuses primarily on usual clicks.

### **1.2.1 The Inter-Pulse Interval**

In addition to their conspicuousness, sperm whale clicks are especially useful to PAM because they possess a unique feature not observed in other cetacean species. In the original description by Backus and Schevill (1966), it was noted that individual clicks are composed of multiple pulses, spaced 2-4 milliseconds apart, where the later pulses progressively decay in amplitude. In 1972, Norris and Harvey presented an explanation for this multi-pulsed structure by proposing that sperm whales produce clicks using their unique nasal structures. The sperm whale's nose is unusual in that it is especially large, occupying most of the head, and it contains two large oil-filled organs stacked upon one another: the spermaceti organ dorsally (i.e. on top), and the "junk" ventrally (i.e. at the bottom). These organs are roughly barrel-shaped and run longitudinally through the nose. Norris and Harvey (1972) theorized that sperm whales produce clicks by pushing air through a lip-like structure of connective tissue (called "museau de singe", French for monkey lips), located at the distal end of the nose. Air sacs located at either end of the spermaceti organ cause this single pulse to reverberate within the organ, thereby appearing as multiple pulses to a receiver. Following an experiment in which artificial sounds were transmitted through the head of a recently deceased sperm whale, Møhl (2001) presented strong evidence supporting the Norris and Harvey theory, with the slight revision that most of the pulse energy is redirected into the junk after reflecting off the frontal (rear) air sac. This revised theory, now called the "bent-

horn” model, has been further supported by the confirmation that sound is produced at the museau de singe (Madsen *et al.*, 2003) and by analysis of the three-dimensional beam pattern of sperm whale clicks (Zimmer *et al.*, 2005b).

An interesting consequence of the Norris and Harvey and bent-horn models is that the inter-pulse interval (IPI) represents the time taken for sound to travel twice through the spermaceti organ. Therefore, whales with larger spermaceti organs are expected to have longer IPIs. Furthermore, Nishiwaki *et al.* (1963) have shown that the length of a sperm whale’s head, which is approximately equal to the length of the spermaceti organ, scales with its total body length. It is therefore possible to estimate a sperm whale’s body length simply by measuring its IPI (Norris and Harvey, 1972; Møhl *et al.*, 1981; Adler-Fenchel, 1980; Gordon, 1991; Rhinelander and Dawson, 2004; Growcott *et al.*, 2011). This has made it possible to acoustically assess the size distributions of sperm whale populations (Adler-Fenchel 1980), measure growth rate (Miller *et al.*, 2013), and to assign coda vocalizations to individuals (Rendell and Whitehead, 2004; Marcoux *et al.*, 2006; Schulz *et al.*, 2008; Antunes *et al.*, 2011; Schulz *et al.*, 2011; Gero *et al.*, 2016b). Also, because sperm whales exhibit extreme sexual dimorphism (Best, 1979; Best *et al.*, 1984), IPIs can be used to distinguish mature males from the much smaller females or juvenile males.

### **1.2.2 On-Axis versus Off-Axis Clicks**

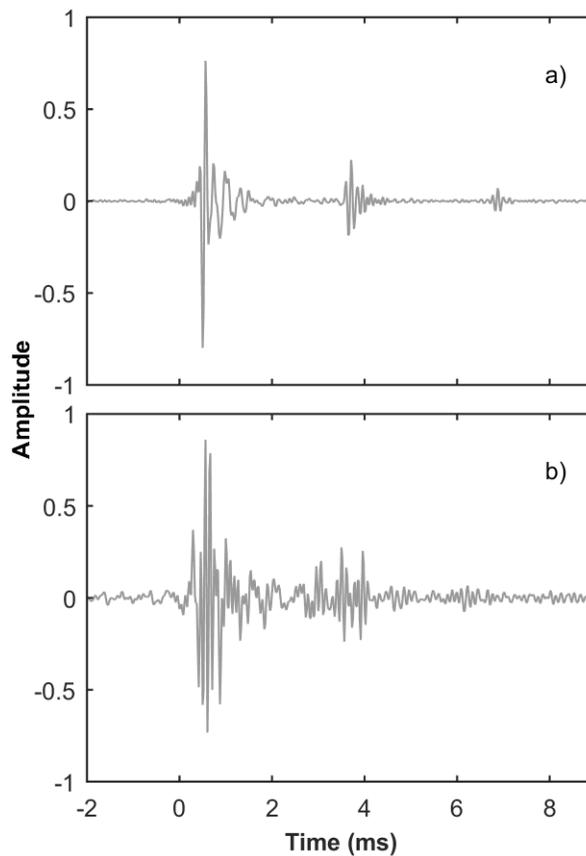
Unfortunately, most sperm whale clicks in typical far-field recordings do not display a clear structure suitable for IPI calculation. They often appear with extra pulses at variable locations, making the pulse interval irregular. These extra pulses can sometimes be explained by coincidence with echoes or other transients (such as clicks from other sperm whales), but they primarily arise from directional effects. The structure of a sperm whale click, both in frequency and time, appears different based on the position of the receiver relative to the whale’s acoustic axis (Møhl *et al.*, 2003; Zimmer *et al.*, 2005a). Only clicks recorded on-axis (i.e. in line in front of, or behind, the emitter) display the characteristic multi-pulse structure representative of the

spermaceti organ size; clicks recorded off-axis are “contaminated” by omnidirectional reflections from the air sacs (Figure 1.1; Zimmer *et al.*, 2005a). Thus, it is actually quite difficult to obtain reliable estimates of a whale’s IPI. There is also some structural variation among on-axis clicks, depending on whether they were recorded in front of a whale (“forward”) or behind (“backward”). Both orientations yield clicks that display the true IPI, but in forward recordings, the first pulse (often called “p0”) has a smaller amplitude than the second. This is because recordings from the front pick up weak energy from the initial pulse produced at the museau de singe, whose subsequent reflections are amplified as they travel forward through the junk (Zimmer *et al.*, 2005a).

Naturally, the window of opportunity where a receiver is considered aligned with a whale’s acoustic axis is very small. However, sperm whales are known to frequently change their orientation at depth, with pitch approaching -90 degrees during descent and varying between  $\pm 50$  degrees while foraging (Zimmer *et al.*, 2003). It is therefore possible to record on-axis clicks occasionally, even from hydrophones fixed at the bottom or near the surface. Using a dataset of approximately 4000 sperm whale usual clicks recorded from the surface, Adler-Fenchel (1980) found that about 11% of clicks were suitable for IPI calculation.

Although it is possible to obtain on-axis clicks, there will almost always be many more off-axis ones, which complicates IPI analysis. Most studies that have used IPIs have worked around this problem by manually searching for and removing off-axis clicks (e.g. Adler-Fenchel, 1980; Gordon, 1991; Drouot *et al.*, 2004; Rendell and Whitehead, 2004; Rhineland and Dawson, 2004; Marcoux *et al.*, 2006; Schulz *et al.*, 2008; Antunes *et al.*, 2011; Schulz *et al.*, 2011; Gero *et al.*, 2016b). This method is effective, but can quickly become impractical. Just one hour of recording can yield over 4000 clicks per whale, making manual triage very time consuming. Another approach is to average every click in a sequence (Teloni *et al.*, 2007; Antunes *et al.*, 2010; Miller, 2010). However, this method assumes that all clicks in the sequence were produced by the same whale. Thus, it is only reliable in situations where only one whale is present, or if individual click trains can be separated. Unfortunately, click train separation is not trivial

and usually requires complex localizing arrays. Click averaging may be a feasible approach for computing IPIs from mature males, which are largely solitary; however, for the highly social females and juveniles, clicks that have a poor multi-pulse structure usually need to be removed.



**Figure 1.1** Example sperm whale clicks, showing multi-pulsed structure. The click in a) was recorded on-axis and displays a clear structure, where the inter-pulse interval (IPI) corresponds to spermaceti organ length. The click in b) was likely recorded off-axis and displays a contaminated multi-pulsed structure, where the true IPI is obscured.

### **1.3 SPERM WHALE SOCIAL STRUCTURE**

A key characteristic of sperm whales that is of both academic and management interest is their complex social structure. Sperm whale society is characterized by sexual segregation, multi-level associations, and cooperation. Females and immature males live in stable groups and reside in tropical or subtropical waters year-round (Best, 1979).

Males, however, migrate to higher latitudes once they mature, where they live alone or in loose aggregations with other males, revisiting the females only briefly to mate (Best, 1979).

The fundamental element of sperm whale society is the social "unit". Units are cohesive groupings of about 3 - 24 females and their immature offspring (Whitehead *et al.*, 1991; Christal *et al.*, 1998, Gero *et al.*, 2014) that are often, but not always, matrilineally related (Richard *et al.*, 1996; Lyrholm and Gyllensten, 1998; Mesnick, 2001; Gero *et al.*, 2008; Konrad, 2017). Members within units form tight social bonds that are stable over long periods of time, on the scale of decades (Christal *et al.*, 1998; Gero *et al.*, 2014). Unit members also exhibit apparently altruistic behaviour, in that they cooperate with one another to care for calves (Arnbom and Whitehead, 1989; Whitehead, 1996; Gero *et al.*, 2009; Gero *et al.*, 2013) and defend themselves against predators (Arnbom *et al.*, 1987; Pitman *et al.*, 2001). A social unit may also temporarily join with other units to form larger groups, which last from a few hours to several days (Whitehead *et al.*, 1991; Gero *et al.*, 2014). When forming groups, units associate exclusively with members of the same "clan", which are distinguished by their coda dialects (Rendell and Whitehead, 2003; Gero *et al.*, 2016a).

#### **1.4 EASTERN CARIBBEAN SPERM WHALES**

The best studied population of sperm whales roams the waters off the Lesser Antilles in the Eastern Caribbean. These whales have been studied off the coast of Dominica almost annually since 2005, as part of a long-term behavioural research project (see Gero *et al.*, 2014). In this population, units are particularly small (about 7-9 animals), and many of them exhibit strong, long-term (i.e. decades-long) site fidelity to the waters off Dominica (Gero *et al.*, 2014). This has made it possible to follow individual whales over time and document their relationships with one another. Consequently, social units are very well characterized in the Eastern Caribbean. Much knowledge on sperm whale communication and the social dynamics within and among social units has originated from this population (see Schulz *et al.*, 2008; Schulz *et al.*,

2011; Antunes *et al.*, 2011; Gero *et al.*, 2016a; Gero *et al.*, 2016b; Gero *et al.*, 2008; Gero *et al.*, 2009; Gero *et al.*, 2013; Gero *et al.*, 2014; Gero *et al.*, 2015).

Eastern Caribbean sperm whales are also of management concern. Because this population is small, relatively isolated, and frequently in close proximity to humans, they are especially vulnerable to anthropogenic impacts (Gero *et al.*, 2007; Gero *et al.*, 2014). Unfortunately, in recent years, this population has begun to show signs of sudden decline (Whitehead and Gero, 2015; Gero and Whitehead, 2016). It is thus imperative that the dynamics and movements of this population be monitored closely. Although much is known about individuals and their relationships, little is known about the large-scale movements of these whales. The overwhelming majority of research on this population has been conducted off the west coast of the island of Dominica, spanning an area of about 2000 km<sup>2</sup>. However, this is a small area for sperm whales, and the units off Dominica are known to frequent the other islands of the Lesser Antilles as well (Gero *et al.*, 2007). With the integration of IPI analysis, passive acoustic monitoring has the potential to help in tracking the movements of these whales across their full range, without having to depend on more invasive technologies such as satellite tags.

## **1.5 THESIS OBJECTIVES**

The central theme of this thesis relates to the use of IPIs to develop automated software tools for highly informative passive acoustic monitoring of sperm whales. Particular interest was placed on the social units of female and immature sperm whales in the Eastern Caribbean. The thesis is divided into two objectives: automatic IPI compilation, and automatic unit identification.

The first objective was the development of an algorithm for automatically estimating the true IPIs of all whales present in an audio recording (Chapter 2). This algorithm is intended to work with only one hydrophone, with no assumptions about which whale each click originated from. To achieve this, I used a machine learning

approach, in which automatically detected clicks are classified as being on-axis or not. IPIs measured from on-axis clicks are then clustered together statistically to infer how many whales are present, and what their IPIs are.

The second objective was the development of an algorithm for automatically inferring the identities of all social units present in an acoustic recording (Chapter 3). This was based on the premise that each social unit has a distinctive distribution of IPIs. To achieve this, I characterized the IPI distributions of units regularly encountered off Dominica as statistical models. The algorithm makes use of the automatic IPI compilation routine from Chapter 2 to obtain empirical IPI distributions from acoustic recordings, which are then compared to the unit IPI profiles.

# **CHAPTER 2 - AUTOMATIC ACOUSTIC ESTIMATION OF SPERM WHALE BODY LENGTHS ACHIEVED THROUGH MACHINE RECOGNITION OF ON-AXIS CLICKS<sup>1</sup>**

## **ABSTRACT**

The waveforms of individual sperm whale clicks often appear as multiple pulses, which are the product of a single pulse reverberating throughout the spermaceti organ. Since there is a relationship between spermaceti organ size and total body size, it is possible to estimate a whale's length by measuring the inter-pulse intervals (IPIs) within its clicks. However, if a click is recorded off-axis, the IPI corresponding to spermaceti organ length is usually obscured. This paper presents an algorithm for automatically estimating the "true" IPIs of sperm whales in a recording by measuring them from on-axis clicks only. The routine works by classifying detected clicks with a support vector machine, assessing the stability of their IPIs, and then clustering the stable IPIs using Gaussian mixture models. Results show that the routine is very accurate in obtaining reliable IPIs, but has a high false negative rate. Nonetheless, since sperm whales click very frequently, it is possible to obtain useful IPI distributions with only a few minutes of recording. 10 minutes is recommended as a minimum. This algorithm makes it possible to estimate the body lengths of multiple sperm whales automatically with only one hydrophone.

## **2.1 INTRODUCTION**

Passive acoustic monitoring (PAM) has become a popular means of studying whales and dolphins over the past several years. With better recording equipment, sound analysis tools, and the realization that cetaceans are more easily observed acoustically than visually, PAM is increasingly being used to supplement or replace traditional visual surveys (Thomas *et al.*, 1986a; Mellinger *et al.*, 2007). This is especially true for sperm whales, since this species spends most of its time foraging at depth

(Watwood *et al.*, 2006), during which it typically produces loud clicks (Backus and Schevill, 1966; Whitehead and Weilgart, 1990). The incorporation of passive acoustics into sperm whale surveys has significantly increased the range and sensitivity of detection (Barlow and Taylor, 2005).

Sperm whale clicks also possess an interesting feature: a single click is composed of multiple pulses (Backus and Schevill, 1966). According to the accepted “bent-horn” model of sperm whale sound production (Norris and Harvey, 1972; Møhl, 2001), these pulses are the product of reverberations between air sacs at the front and back of the spermaceti organ. As a consequence, the inter-pulse interval (IPI) is directly related to the length of the spermaceti organ. Since there is also an allometric relationship between spermaceti organ length and body length (Nishiwaki *et al.*, 1963), it is possible to estimate a whale’s body length simply by measuring its IPI (Norris and Harvey, 1972; Møhl *et al.*, 1981; Adler-Fenchel, 1980; Gordon, 1991; Rhineland and Dawson, 2004; Growcott *et al.*, 2011). This feature makes PAM especially informative for sperm whales.

Unfortunately however, most sperm whale clicks from typical far-field recordings do not display a clear structure suitable for IPI calculation. They often appear with extra pulses at variable locations, making the pulse interval irregular. These extra pulses arise because of directionality: sperm whale clicks are highly directional, and their structure in both frequency and time appears different based on the position of the receiver relative to the whale’s acoustic axis (Møhl *et al.*, 2003; Zimmer *et al.*, 2005a). Only clicks recorded on-axis display the characteristic multi-pulse structure representative of the spermaceti organ size. Clicks recorded off-axis are confounded by omnidirectional reflections from the air sacs (Zimmer *et al.*, 2005a). As a consequence, IPI calculation is actually a difficult task, because it requires that on-axis clicks be separated from the off-axis ones.

Most studies that have used IPIs have worked around the directionality problem by manually searching for and removing off-axis clicks (e.g. Adler-Fenchel, 1980; Gordon, 1991; Drouot *et al.*, 2004; Rendell and Whitehead, 2004; Rhineland and Dawson, 2004; Schulz *et al.*, 2011). This method is effective, but can quickly become

impractical, as just one hour of recording can yield over 4000 clicks per whale. Another approach is to average every click in a sequence (Teloni *et al.*, 2007; Antunes *et al.*, 2010). Averaging works because the stability of the true IPI allows it to emerge above the noise, but this assumes that each click was produced by the same whale. Thus, this method is only reliable if individual click trains can be separated, which is difficult and impractical in many situations.

One approach to IPI compilation that has not been tested until now is using automatic classification to isolate on-axis clicks. A great advantage to this approach is that it does not require knowledge of which whale produced which clicks, so click trains do not need to be resolved. The goal of this research was to produce a software tool capable of compiling reliable IPI distributions automatically, based on machine classification of clicks. This tool is designed to be as simple to use as possible, requiring only a single-channel audio recording file as input. The output consists of filtered IPI distributions, with estimates of how many whales are present, what their true IPIs are, and ultimately their body lengths. Such a tool could greatly enhance the effectiveness of passive acoustic monitoring for sperm whales.

This paper describes how the tool was developed, its underlying algorithms, and its performance.

## **2.2 METHODS**

### **2.2.1 Data Collection**

Female and immature sperm whales were followed off the west coast of the island of Dominica in the Eastern Caribbean, from February-April 2015 (56 days effort) aboard a 12-m auxiliary sailing vessel as a part of a long-term behavioural research program (see Gero *et al.*, 2014). Acoustic recordings were made using a custom-built towed hydrophone (Benthos AQ-4 elements, frequency response 0.1 – 30 kHz) and a filter box with high-pass filters up to 1 kHz. This resulted in a recording chain with a flat frequency response across a minimum of 2 – 20 kHz. Audio data were collected through

a computer-based recording system, with a sampling rate of either 48 kHz or 96 kHz, and 16-bit resolution. All recordings were stored in WAVE format. Recordings were categorized into two types based on how they were obtained, which are referred to as “first-click” and “standard”.

In the “first-click” protocol, acoustic recording was initiated immediately after a whale began a foraging dive. The research vessel remained stationary. In this scenario, since the whale is near and facing almost directly away from the research vessel during its descent, the echolocation clicks it produces are likely to be perceived clearly and on-axis. The purpose of “first-click” recordings was to obtain samples of on-axis sperm whale clicks. A total of 7 “first-click” recordings were used for this purpose. These lasted between about 3.5 – 7.5 minutes (36 minutes total). Due to the social nature of female sperm whales, most “first-click” recordings captured more than one animal diving at the same time. Based on photographic identification of the flukes of diving whales (Arnbom, 1987), these 7 recordings contained clicks from at least 10 adult female-sized individuals in total.

In the “standard” protocol, acoustic recording was initiated at predetermined time periods (usually one-hour intervals) during days when sperm whales were encountered. In some cases, the research vessel was stationary, while in others it was sailing or motoring at low speed. In this scenario, the location, orientation, number, and identity of whales immediately surrounding the hydrophone is usually unknown. Since sperm whales spend most of their time foraging, any whales present during standard recordings are likely to produce echolocation clicks. However, these clicks may be perceived from any angle, and the majority are typically off-axis. “Standard” recordings were normally run for 4 minutes, although on a few occasions, this varied between 3-15 minutes. “Standard” recordings were used to obtain click samples typical of most passive acoustic monitoring situations. A total of 174 “standard” recordings were used, representing 14 hours total.

### 2.2.2 Software Design Overview

This section provides an overview of the proposed algorithm for automatically compiling sperm whale IPIs. Each component of the algorithm is then described in further detail in the sections that follow. A complete graphical representation is also included in Appendix A.

The routine takes digital audio files as input (WAVE format), filters the contents automatically for on-axis sperm whale clicks through a series of steps, and outputs the IPIs of the filtered clicks along with estimates of animal counts, their IPIs, and body lengths. All analysis is conducted at 48 kHz. If the original sampling rate is different, then the recording is upsampled or downsampled automatically as needed. Since the routine is intended to be usable with just one hydrophone, only one audio channel is used. If the input recording has multiple channels, the first channel is retained by default. This program uses MATLAB version R2015a with the Signal Processing Toolbox, the Statistics and Machine Learning Toolbox, the Curve Fitting Toolbox, and the Parallel Computing Toolbox (The MathWorks, Inc., Natick, Massachusetts, USA).

The routine necessarily depends on many parameters that could be adjusted. A full sensitivity analysis of how each parameter affects the output would require a substantial amount of testing, and with the exception of two parameters, was beyond the scope of this work. However, default values were established for all parameters, based on published information, data observations, and/or small heuristic tests. In the case of heuristic determination, defaults were selected based on their robustness to a variety of scenarios (e.g. excellent, poor, or variable signal-to-noise ratio). Thus, using default parameter values, the routine should provide adequate performance for a wide range of recording qualities. Notable exceptions will be discussed in section 2.4 *Discussion*. The full list of parameters, including their default values, are given in Appendix B.

### 2.2.3 Audio Loading and Preprocessing

The routine extracts the recorded sound pressure waveform from one channel of an input audio file. If the waveform needs to be resampled, an FIR anti-aliasing filter is applied. This filter uses a Kaiser window and has an order of  $50 \times \max(p, q)$ , where  $p/q$  is the reduced resampling ratio. The waveform is then noise-filtered using a 2-12 kHz Butterworth bandpass filter, run in both directions to avoid non-linear frequency-dependent delay (i.e. zero-phase filtering). Although sperm whale clicks contain some energy outside this band, IPIs were found to be generally more precise when limited to 2-12 kHz (precision in this sense will be discussed in section 2.2.7 *IPI Calculation and Validation*). Filter order after zero-phase filtering is 12.

### 2.2.4 Click Detection

Candidate sperm whale clicks are detected within the time series by a custom click detection algorithm. This algorithm is based on the Page test (Page, 1954), a method commonly used to isolate cetacean clicks. The particular implementation used here is similar to the ones described by Miller (2010) and Zimmer (2011), and used by the open-source PAM software PAMGUARD (Gillespie *et al.*, 2008). It was adapted to capture the multi-pulsed structure of sperm whale clicks as accurately as possible.

The detector works by examining each sample in the time series in sequence. For any given sample  $i$ , it operates under one of two states: either “noise” (click absent) or “signal” (click present). State changes are mediated by the value of a signal strength statistic  $V_i$  relative to two threshold values,  $T_{on}$  and  $T_{off}$ . If the current state is “noise”, the detector will switch to the “signal” state when  $V_i > T_{on}$ . If the current state is “signal”, the detector will switch to the “noise” state when  $V_i < T_{off}$ . The detector will also be forced back into the “noise” state if it has been in “signal” for too many samples. By default, this limit is set to the number of samples equivalent to 40 ms, a duration that is longer than most sperm whale usual clicks. The detector always begins assuming the “noise” state.

Initial estimates of click temporal range (defined by start and end samples) are established during state transitions. End samples are simply defined as the moments where state changes from “signal” to “noise”. Start samples, however, involve some backtracking: they are defined as the moments where  $V_i > T_{off}$  just before transitions to the “signal” state. In other words, click detection events are triggered by  $T_{on}$  crossings, but click ranges are determined by  $T_{off}$  crossings.

The thresholds  $T_{on}$  and  $T_{off}$  are set to 10 and 1 by default, respectively. The signal strength statistic  $V_i$  is defined as:

$$V_i = \frac{s_i}{n_i} \quad (2.1)$$

where  $s_i$  and  $n_i$  are estimates of instantaneous signal and noise power, respectively. Both of these power estimates are obtained using exponential moving average filters, implemented with the recursive formula:

$$z_i = (1 - \alpha)z_{i-1} + \alpha\psi_i \quad (2.2)$$

where  $\psi$  is some measure of the power of the raw waveform (that is, including both signal and noise). For this detector,  $\psi$  is the square of the waveform envelope, where envelope is computed as the absolute value of the analytic signal (obtained using the Hilbert transform). The parameter  $\alpha$  in equation (2.2) is constrained between 0 and 1 and affects the level of smoothing. For signal power,  $\alpha = 0.2$  by default. Noise power uses one of two  $\alpha$  values depending on the detector state: by default, it is 0.000002 when in the “signal” state, and 0.0002 during the “noise” state. Equation (2.2) is initialized with  $z_0 = \psi_1$  for the signal estimate, and  $z_0 = \psi_{RMS}$  for noise.

After each sample has been processed, a series of post-hoc validation routines are applied to edit the click ranges. The first step immediately removes any clicks shorter than 50  $\mu$ s by default, a lower limit for the duration of a typical sperm whale pulse. The next step translates the start and end samples of each click to account for the delay introduced by the signal power estimation filter. Following this, each click is assessed individually on two criteria. The first criterion checks if clicks can be merged with succeeding ones. This step was implemented to allow for the possibility that individual

pulses were detected, rather than whole clicks. A merge is performed when three conditions are true:

- 1) The interval between the current click and the next click is within the range expected for sperm whale IPIs (2–9 ms by default, as suggested by Marcoux *et al.*, 2006).
- 2) The interval between the start of the current click and the end of the next click is within the maximum expected click duration (40 ms by default).
- 3) The peak amplitude of the next click is not a significant fraction of the peak amplitude of the current click. This condition was implemented to reduce the chance of clicks being merged with their surface reflections, and is based on the fact that true pulses in on-axis clicks progressively decay in amplitude<sup>2</sup> (Zimmer *et al.*, 2005a).

The second criterion ensures that clicks are sufficiently long to be sperm whale clicks (at least 2 ms by default). This prompts the removal of all clicks that are long enough to be pulses, but too short to be multi-pulsed clicks. It is implemented last to allow individual pulses a chance to be merged.

### **2.2.5 Feature Extraction**

After candidate clicks have been detected, the program computes a set of features for each click. These features are described in Table 2.1. The purpose of these features is to provide information from which on-axis sperm whale clicks can be differentiated from off-axis clicks and other transients automatically. How they are used to classify clicks will be discussed in section 2.2.6 *Click Classification*. To a human observer, on-axis sperm whale clicks can be recognized relatively easily by observing their waveforms. However, quantifying the characteristics that define on-axis clicks for automatic classification is difficult. Thus, feature selection required a certain amount of exploratory analysis and trial-and-error.

Among the features listed in Table 2.1, some of them, such as click duration, peak/centroid frequency, and bandwidths, are commonly used to characterize odontocete clicks. Others, such as pulse ZCR variance, and goodness of exponential fit, were conceived in attempt to quantify the unique multi-pulsed structure of on-axis sperm whale clicks. To obtain an initial evaluation of the potential usefulness of each feature, two-sample t-tests were used, in which a set of clicks were manually categorized as being on-axis sperm whale clicks or not (this dataset is explained in section 2.2.6 *Click Classification*). These tests suggested that all features in Table 2.1 differ significantly between the two click types. To further refine the final set of features, the accuracy of automatic classification was examined for various subsets of the feature list in Table 2.1. Accuracy in this sense refers to the total agreement between automatically-assigned labels and manually-assigned labels for a given dataset (explained in Table 2.2). Since the testing of all  $2^{16}$  possible feature subsets was infeasible, only a few strategic sets were examined. This included the removal of potentially correlated features (e.g. retaining only one of -3 dB bandwidth, -10dB bandwidth, or RMS bandwidth), and the removal of entire categories of features (e.g. all spectral features, all pulse-based features). Ultimately, it was found that the entire feature set in Table 2.1 yielded the highest accuracy.

Most of these features depend on information that must be computed beforehand. This includes the location of individual pulses within a click, the frequency spectra of clicks, and exponential curve fits. The following subsections describe the calculation of these dependencies in more detail.

**Table 2.1** Description of features used for classifying clicks as being on-axis sperm whale clicks or not

| Feature            | Dependency | Description   |
|--------------------|------------|---|
| Click duration     | None       | The total duration of a click   |
| Click peak SNR     | None       | The instantaneous signal-to-noise ratio of the tallest peak in the click. The signal and noise powers are those measured during the Page test, from exponential moving average filters.   |
| Peak frequency     | Spectrum   | The frequency with the largest amplitude in the power spectrum.   |
| Centroid frequency | Spectrum   | The frequency that divides the power spectrum into two halves with equal energy. It is often used to describe the frequency content of broadband signals (Au, 1993). Centroid frequency was computed as follows:<br>$f_0 = \frac{\sum_{i=1}^n f_i  X_i ^2 \Delta f}{\sum_{i=1}^n  X_i ^2 \Delta f}$ Where $f_i$ is the frequency at sample $i$ , $X_i$ is the frequency domain signal at sample $i$ (obtained using the Fast Fourier Transform), and $\Delta f$ is the increment between frequencies. $n$ is equal to half the number of samples used in the FFT, which is always a power of two. |
| -3 dB bandwidth    | Spectrum   | The width of the frequency band defined by the endpoints about the peak frequency where the peak's power drops by 3 dB (called the "half-power" point). This is a common method to measure the frequency range of odontocete clicks and other transients (Au, 1993).  |
| -10 dB bandwidth   | Spectrum   | The width of the frequency band defined by the endpoints about the peak frequency where the peak's power drops by 10 dB. This method has been used to describe the frequency range of sperm whale clicks (Madsen <i>et al.</i> , 2002b).  |

**Table 2.1 (Continued)**

| Feature                                 | Dependency      | Description   |
|---|-----------------|---|
| Root-mean-square (RMS) bandwidth        | Spectrum        | <p>A measure of frequency bandwidth measured about the centroid frequency. This is often used to measure the bandwidth of broadband transients (Au, 1993), including sperm whale pulses (Møhl <i>et al.</i>, 2000). Based on Au (1993), RMS bandwidth was calculated as follows:</p> $\beta = \sqrt{\frac{\sum_{i=1}^n f_i^2  X_i ^2 \Delta f}{\sum_{i=1}^n  X_i ^2 \Delta f}} - f_0^2$ <p>Where <math>f_i</math> is the frequency at sample <math>i</math>, <math>X_i</math> is the frequency domain signal at sample <math>i</math> (obtained using the Fast Fourier Transform), and <math>\Delta f</math> is the increment between frequencies. <math>n</math> is equal to half the number of samples used in the FFT, which is always a power of two.</p> |
| Pulse count                             | Pulse detection | The number of pulses detected within a click.   |
| Mean pulse duration                     | Pulse detection | The mean duration of all pulses within a click.   |
| Pulse duration variance                 | Pulse detection | The variance of the durations of all pulses within a click.   |
| Pulse zero-crossing rate (ZCR) variance | Pulse detection | <p>The variance of the zero-crossing rates of all pulses within a click.</p> <p>For a given sample <math>i</math>, a zero crossing event is detected if the following condition is true:</p> $x_i x_{i-1} < 0$ <p>where <math>x</math> is the waveform. The ZCR was calculated as the number of zero crossing events divided by the number of samples in a pulse.</p>   |

**Table 2.1 (Continued)**

| Feature                                   | Dependency      | Description  |
|---|-----------------|--|
| Best pulse cross-correlation              | Pulse detection | The maximum correlation value among cross-correlations run between the tallest pulse and every other pulse in the click. It is used as a measure of similarity between pulses in a click. To ensure that each correlation function can be compared with one another and across multiple clicks, each pulse is normalized for cross-correlation (that is, each pulse is scaled such that their peaks equal one). The tallest pulse is lagged against every other pulse. At lag zero, its peak is aligned with the first sample of the target pulse. At the final lag (equal to the duration of the target pulse), its peak is aligned with the last sample of the target pulse. Rather than zero-padding the regions beyond the bounds of the tallest pulse, its range is extended to include the surrounding waveform samples. |
| Goodness of Gaussian fit to tallest pulse | Pulse detection | The r2 coefficient of the best least-squares Gaussian fit to the envelope of the tallest pulse. Gaussian fits are applied based on this equation:<br>$y(t) = \alpha e^{\left[-\frac{(t-\beta)^2}{\gamma}\right]}$ where $y$ is the waveform envelope, and $t$ is time. To avoid spurious fits, the $\alpha$ , $\beta$ , and $\gamma$ coefficients were constrained near the $y$ and $t$ bounds of the pulse, such that the peaks of the fit and the pulse envelope were likely to align.   |
| Goodness of exponential fit               | Exponential fit | The r2 coefficient of the best exponential fit to the peaks of the pulses in a click.  |
| Exponential fit $\alpha$ coefficient      | Exponential fit | The $\alpha$ coefficient in equation (2.3) for the best exponential fit to the peaks of the pulses in a click.   |
| Exponential fit $\beta$ coefficient       | Exponential fit | The $\beta$ coefficient in equation (2.3) for the best exponential fit to the peaks of the pulses in a click.  |

### 2.2.5.1 Pulse Detection

The isolation of individual pulses within a sperm whale click is particularly tricky, because the noise level within clicks is often highly variable, and later pulses may be fainter than the average noise. Thus, conventional click detection does not perform well at this resolution, because a fixed threshold risks rejecting many pulses, or detecting many spurious ones. Therefore, a different approach was used. This approach involves signal smoothing, followed by the detection of local maxima.

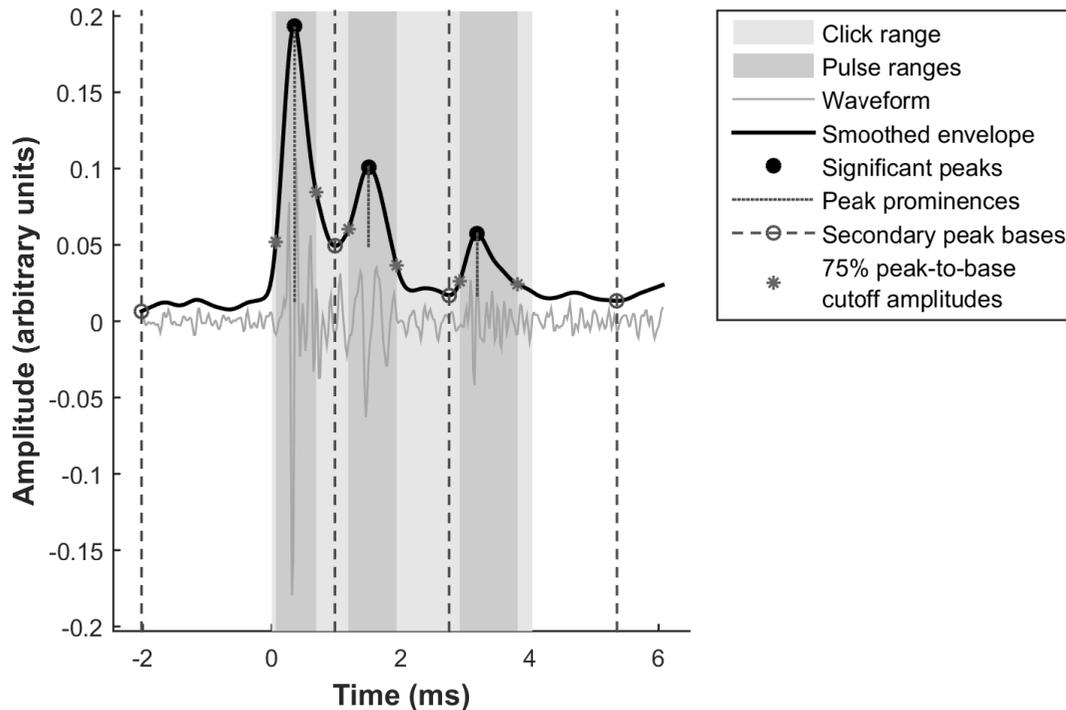
The pulse detector begins by attempting to smooth the waveform envelope around the click of interest in such a way that individual sperm whale pulses might be best represented. It does this by applying a series of increasing smoothing bandwidths until a suitable one is found. Suitability in this sense is determined by peak separation: if the smoothed waveform envelope contains prominent peaks that are too close together (less than 1 ms by default), then greater smoothing is required. Smoothing itself is performed by locally weighted linear regression (“LOWESS”). Based on visual inspection of smoothed click envelopes, this method appeared to define sperm whale pulses slightly better than moving average approaches.

After a suitable smoothed envelope has been established, potential pulses are found by detecting significant peaks, or local maxima, in the envelope. While envelope smoothing is effective for eliminating rapid changes in amplitude, there often remain small peaks that are attributed to slowly-varying changes in noise amplitude, rather than actual sperm whale pulses. To detect significant peaks, the pulse detection routine uses the *findpeaks* function in MATLAB’s Signal Processing Toolbox (The MathWorks, Inc. (2015)). *findpeaks* can assess the significance of a peak based on its “prominence”. According to The MathWorks, Inc. (2015), prominence is measured as the difference between the peak’s amplitude, and the amplitude of its greatest primary base. Primary bases are defined here as being the two points with smallest amplitude, one on either side of the peak, within the interval that spans the first points (as measured from the peak) that are taller than the peak. If no such point exists, then the interval extends to the signal endpoint. A peak is considered significant if its prominence exceeds a certain

threshold. For this application (i.e. pulse detection in sperm whale clicks), the prominence threshold is determined by taking a small fraction (5% by default) of the difference in amplitude between the most prominent and least prominent peaks in the smoothed envelope. Figure 2.1 includes examples of significant peak prominences.

Once significant peaks are found, thereby indicating the presence of pulses, the next step estimates the temporal ranges of each pulse. This is done by first identifying the secondary bases of each peak. Here, secondary bases are defined as being the points along the smoothed envelope that correspond to the nearest significant local minima on either side of the peak (Figure 2.1). Significant local minima are essentially the significant peaks of the negative of the smoothed envelope. Between the peak and its two secondary bases, cutoff points describing the initial pulse ranges are established. These points correspond to where the envelope, as measured from the peak, crosses some reference amplitude. For each secondary base, this amplitude is a fraction (75% by default) of the difference between the peak amplitude and the base amplitude (Figure 2.1).

Before being accepted, a pulse range is assessed based on its duration. If it is too short ( $< 0.05$  ms), the pulse is removed. If it is excessively long ( $> 1$  ms), the algorithm attempts to shorten it by raising the two cutoff amplitudes. In the rare case that a pulse cannot be made shorter, it is removed.



**Figure 2.1** Example of features used in the detection of pulses within a click. The small unmarked peaks and valleys did not have significant prominences, and were essentially treated as noise. Prominence is determined by the *findpeaks* function in MATLAB’s Signal Processing Toolbox (The MathWorks, Inc., 2015).

### 2.2.5.2 Frequency Spectrum Calculation

To compute spectra, a Tukey window is applied to each click, where the flat portion always encompasses the entire click. The Fast Fourier Transform (FFT) is then applied to each windowed click. The number of points used in FFT is set as the smallest power of two that is larger than the number of samples within the window of the longest click in the file. Thus, the number of FFT points is consistent for each click in a file, but can vary between files.

### 2.2.5.3 Exponential Fitting

This process depends on pulse detection and is intended to describe the amplitude decay of pulses in on-axis clicks. It involves the least-squares fitting of exponential curves of the form:

$$y(t) = \alpha e^{\beta t} \quad (2.3)$$

where  $y$  corresponds to the peak pulse amplitudes (measured from the waveform envelope), and  $t$  is time. For every click, this equation is fit to the peaks of all pulses composing the click. To standardize these fits and ensure that the coefficients are comparable across clicks, each click undergoes two transformations before the fit is applied. First, the whole click is scaled in amplitude such that its tallest peak is equal to one. Second, it is scaled along the time axis so that the first pulse's peak occurs at  $t = 0$ , and the mean delay between consecutive peaks is equal to one. This is done in attempt to standardize the IPI in a manner that is robust to variability in the number of pulses detected within clicks.

### 2.2.6 Click Classification

The next step uses the extracted features to automatically classify each click as being an on-axis sperm whale click or not. Classification is performed by a support vector machine (SVM) that uses a quadratic kernel. This particular classification method was selected because it yielded the highest accuracy, compared to other methods including discriminant function analysis, logistic regression, decision trees, and nearest-neighbour classifiers. Accuracy was assessed as in Table 2.2.

SVM training involved the construction of an example dataset from manually labelled clicks. The waveforms of clicks detected in the "first-click" recordings were carefully examined, and assigned to one of three categories: "Good", "Bad", or "Unsure". Clicks labelled "Good" were direct-path sperm whale echolocation clicks with a clear multi-pulse structure characteristic of on-axis clicks, as described by Zimmer *et al.* (2005a). Clicks labelled "Bad" represented everything else, including non-sperm

whale transients, off-axis clicks, confounded on-axis clicks, surface reflections, and coda clicks. Surface reflections were included in the “Bad” category because they often appear noisier than direct-path clicks, resulting in imprecise IPI measurements. Coda clicks are also considered “Bad” because their structure is quite different from the much more common usual clicks (Madsen *et al.*, 2002a), and their measured IPIs are not necessarily identical to those of usual clicks (Schulz *et al.*, 2011). “Unsure” clicks were those that exhibited characteristics of on-axis clicks, but could not confidently be called such. These clicks were usually borderline cases (i.e., partially on-axis), had very low signal-to-noise ratio, or were potentially false alarms created by multiple coinciding transients. Figure 2.2 shows examples of click waveforms that were assigned to each category. All manual labelling was performed by one observer (W.B.).

After label assignment, features were extracted for every “Good” and “Bad” click; “Unsure” clicks were omitted. The SVM was trained to differentiate “Good” clicks from “Bad” ones using the resulting dataset, then evaluated using 10-fold cross-validation. Ultimately, the training dataset consisted of 6986 clicks in total: 487 “Good”, and 6499 “Bad”. The final SVM was made to output the probability of each click being “Good” by using Platt’s (1999) method. This makes it possible to set a threshold for how “Good” each click must be to be accepted.

**Table 2.2** Metrics used to assess the performance of automatic classification, relative to manual classification. Alternative names for these metrics are included in brackets. Each metric depends on some or all of the following variables:

TP = number of true positives

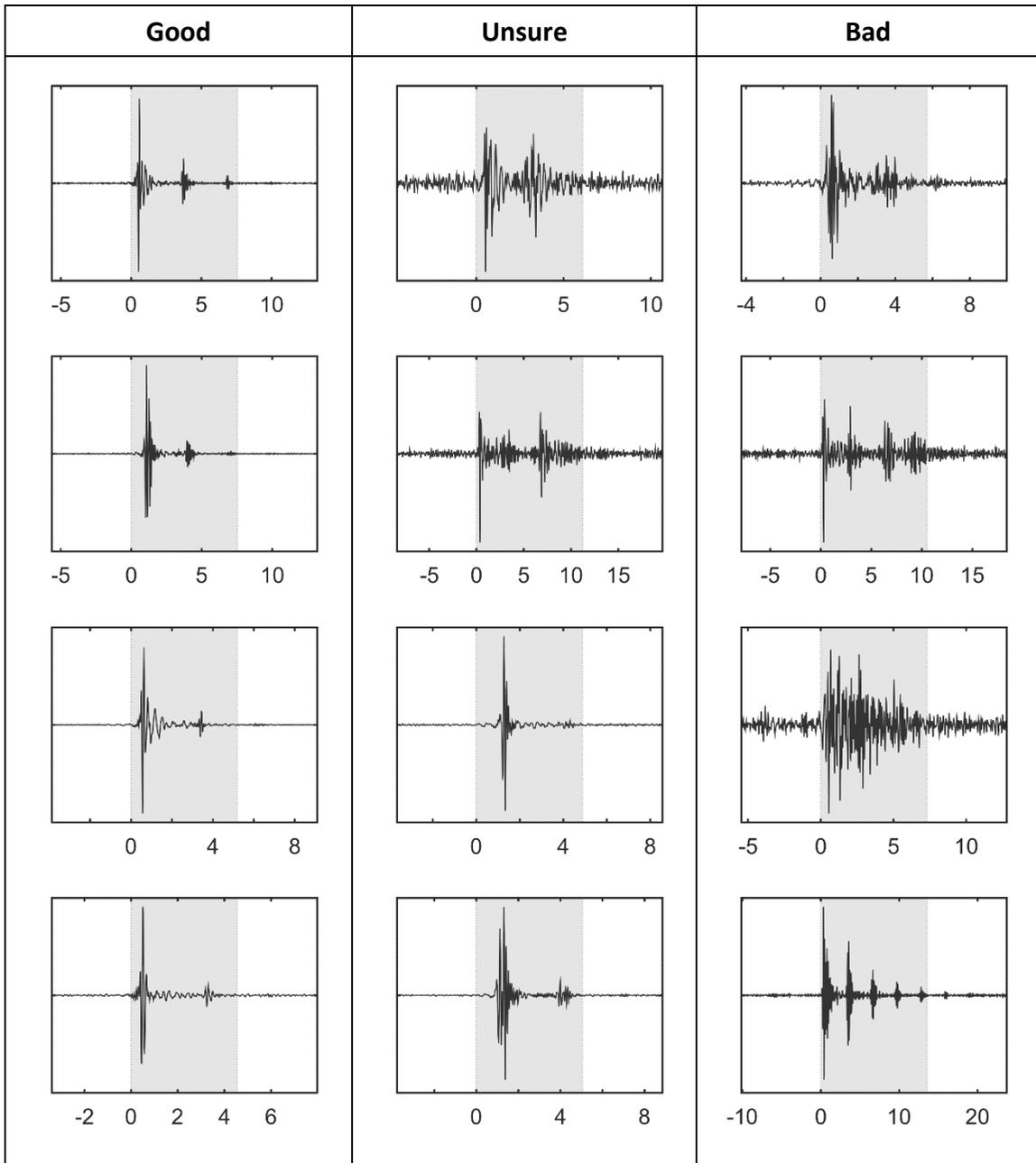
TN = number of true negatives

FP = number of false positives

FN = number of false negatives

Here, the positive class corresponds to “Good” clicks, and the negative class corresponds to “Bad” clicks.

| <b>Metric</b>                                   | <b>Formula</b>                                      | <b>Description</b>   |
|---|---|--|
| Accuracy  | $\frac{TP + TN}{TP + TN + FP + FN}$                 | Total agreement between automatic and manual classification  |
| Adjusted Accuracy                               | $\frac{\frac{TP}{TP + FN} + \frac{TN}{TN + FP}}{2}$ | Total agreement in which each class is weighted equally. This is to take into account the large discrepancy between the number of “Good” and “Bad” clicks. |
| Sensitivity<br>(Recall)<br>(True Positive Rate) | $\frac{TP}{TP + FN}$                                | Proportion of all “Good” clicks that were correctly classified as “Good”   |
| Specificity<br>(True Negative Rate)             | $\frac{TN}{TN + FP}$                                | Proportion of all “Bad” clicks that were correctly classified as “Bad”   |
| Precision                                       | $\frac{TP}{TP + FP}$                                | Proportion of clicks classified as “Good” that were actually “Good”  |



**Figure 2.2:** Example click waveforms examined for construction of the click classifier training dataset and their assigned labels. X-axis is in milliseconds; Y-axis is in arbitrary units. Highlighted regions represent automatically detected click ranges. Clicks were labelled “Good” if they exhibited a clear, decaying multi-pulsed structure characteristic of on-axis sperm whale echolocation clicks, as in Zimmer *et al.* (2005a). “Unsure” Clicks were those that had ambiguous multi-pulsed structures but were potentially on-axis. “Bad” clicks included those that had unclear or absent multi-pulsed structures, because either 1) they were recorded off the acoustic axis, 2) they were contaminated by other signals, 3) they were echoes of on-axis clicks, or 4) they were not sperm whale clicks. Since the goal was recognize on-axis echolocation clicks, the “Bad” category also included coda clicks, even if they had clear structures (e.g. bottom-right).

### 2.2.7 IPI Calculation and Validation

After each click has been automatically classified as “Good” or “Bad”, the routine computes IPIs for all “Good” clicks. This is done using the two methods proposed by Goold (1996): autocorrelation analysis, and cepstral analysis. Thus, each “Good” click initially has two IPI estimates. For both methods, the program constrains IPI calculation between 2 ms, and either 9 ms or the click duration, whichever is shorter. The upper bound of 9 ms is a limit for the IPIs of large male sperm whales, while the lower bound of 2 ms is used to avoid confusion from high correlations within wide first pulses, but may exclude clicks produced by young calves (Marcoux *et al.*, 2006).

For cepstral analysis, the power cepstrum is computed as:

$$C_q = |FFT(\log_{10}(|FFT(x_t)|^2))| \quad (2.4)$$

where  $FFT$  denotes the Fast Fourier Transform. To get a good signal in the cepstrum, it is best if all pulses have the same amplitude. To facilitate this, clicks are windowed before the first  $FFT$ , where the window function consists of chi-squared probability densities as suggested by Goold (1996):

$$f_w(n) = \frac{1}{2^{k/2}\Gamma(k/2)} n^{(k/2)-1} e^{-(n/2)} \quad (2.5)$$

where  $\Gamma$  is the gamma function for positive integers:

$$\Gamma(k) = (k - 1)! \quad (2.6)$$

In all cases,  $k$  is set to 4, as this value appeared most appropriate based on visual inspection of several windowed clicks. The second  $FFT$  uses a Tukey window with the flat part spanning 2-12 kHz. The number of samples used for each window is the maximum of either the smallest power of two that is larger than the number of samples within the longest click, or the smallest power of two such that the “Nyquist quefreny” is greater than the upper IPI limit.

For each click, a final IPI is obtained by averaging the autocorrelation and cepstral IPIs. The point of using both methods is to improve confidence in the IPI

estimate. Neither method on its own is perfect (Antunes *et al.*, 2010), but if they both return the same number, then the final IPI is likely to be reliable. Therefore, the next step in the routine rejects all clicks whose two IPI estimates deviate from the average by more than 0.05 ms by default, as in Schulz *et al.* (2011).

After each IPI has been calculated and validated for precision, a final validation step is performed. This involves searching for IPI repetitions. Since sperm whales emit usual clicks in trains at short, regular intervals, it is expected that the same IPI will be recorded more than once within a few seconds. The program exploits this property to further validate the IPIs it has measured. For each “Good” click with a precise IPI, the routine scans the time series locally about the click’s time of occurrence, in both directions. The target of this scan is another click with the same IPI as the focal click, within tolerance ( $\pm 0.05$  ms by default). If the scan is successful, then a new scan is performed about the repeated click. This cycle continues for as many repetitions as specified. To reduce confusion, a “repetition” is explicitly defined as being one recurring instance (within tolerance) of an IPI within a neighboring click. Based on this definition, “zero repetitions” means that a click has no neighbors with a similar IPI, “one repetition” means that a click has one neighbor with a similar IPI, and so on.

When searching for the first repetition, the scan is conducted within a broad range of typical sperm whale inter-click-interval (ICI) values from the original click (0.25 - 1.5 seconds by default). For subsequent repetitions, the range is narrowed such that only clicks with the same ICI as that separating the previous two clicks ( $\pm 0.2$  seconds by default) are considered. If more than one click is found within range, then each click is used to search for successive repetitions until the required number of repetitions has been met. All clicks with an insufficient number of successive IPI repetitions within ICI range are removed. Those with enough repetitions contribute to the final IPI distribution.

### 2.2.8 Animal Count and Length Estimation

The number of whales present and their body lengths are estimated through cluster analysis of the filtered IPI distribution. This is accomplished using Gaussian mixture models (GMMs). IPI measurements from individual whales have been found to be quite stable, often within  $\pm 0.05$  ms from the mean (Schulz *et al.*, 2011; Growcott *et al.*, 2011). Therefore, output IPI distributions are expected to contain mixtures of narrow peaks, where each peak corresponds to an individual whale (assuming each whale present in the recording has a distinct IPI). Mixture modelling is thus a suitable approach for resolving the composition of IPI distributions.

GMMs are fitted through the Expectation-Maximization (EM) algorithm, which is an iterative process for estimating the most likely parameter values. It requires that the number of clusters,  $k$ , be specified beforehand, and may also take estimates of cluster means, variances, and proportions to accelerate convergence. In this case, GMMs are initialized based on two Gaussian kernel density estimates (KDEs). One kernel uses a wide bandwidth (0.0333 ms by default), while the other has a narrow bandwidth (0.0167 ms by default). GMMs are run for every  $k$  within the range  $\max(1, N_{wide} - 1)$  to  $N_{narrow} + 1$ , where  $N_{wide}$  and  $N_{narrow}$  represent the number of peaks found within the wide and narrow bandwidth KDE functions, respectively. Initial estimates of cluster means and proportions are also based on KDE peaks. In the case where  $k = N_{wide} - 1$ , a peak is removed at random. Likewise, for  $k = N_{narrow} + 1$ , a peak is added at random. For values of  $k$  in-between, the peaks to use are decided based on sparseness, where the most isolated peaks are added first as  $k$  increases. The initial standard deviation varies based on the value of  $k$  and the bandwidths of the two KDEs. To avoid numerical instabilities, and also account for IPI quantization to some degree, a value of  $(1/Fs)/4$  is added to each standard deviation during the EM estimation, where  $Fs$  is in kHz. By default, standard deviation is constrained to be identical for all clusters in a model. GMMs with different numbers of clusters are compared to one another using the Bayesian Information Criterion (BIC).

The results of GMM clustering provide insight into how many whales are present, and what their IPIs are. The means of each cluster in a GMM are estimates of each whale's true IPI. From these measures, body lengths can be estimated using equations such as those published by Gordon (1991) and Growcott *et al.* (2011). The GMM whose BIC score is smallest is considered to be the most likely scenario. However, it should be noted that BIC is not a perfect model selector, and occasionally there are situations where a slightly lesser supported model may be more accurate. Thus, in practice, the automatic routine returns several GMMs, which are ranked according to their BIC scores.

### **2.2.9 Performance Analysis**

Performance of the classifying SVM was assessed based on the metrics in Table 2.2. To assess the performance of the automatic IPI filtration routine as a whole, its output was examined for "standard" recordings, in which the orientation of whales was unknown. The routine was run multiple times for each "standard" recording, where two parameters were changed for each run: the minimum probability at which a click is considered "Good", and the number of times each IPI must be repeated in succession. Minimum probability was tested for values of 0.1, 0.3, 0.5, 0.7, and 0.9, with IPI repetition fixed at 1. IPI repetition was tested for values of 0 through 4, with minimum probability fixed at 0.7. Ideally, to measure routine performance, the "true" probability distribution of IPIs for each whale present during the "standard" recordings would need to be known. Unfortunately, this information is extremely difficult to obtain in the field, and was not available for this analysis. Thus, performance was assessed using two alternative approaches. One of these is called "peak definition", which measures the stability of variance among clusters of IPIs in a given recording. "Peak" in this sense refers to areas of high density in IPI distributions that appear roughly Normally distributed, and presumably correspond to the IPIs of individual whales. The other measure of performance is referred to as "accuracy", with respect to manually compiled

IPI distributions. Accuracy in this sense quantifies how two probability distributions are similar to one another, and is distinct from the classification accuracy listed in Table 2.2.

To measure peak definition, mixture modelling was applied. For each distribution, two GMMs with an equal number of clusters were compared, where one model required all clusters to have the same variance, and the other did not. The constrained model consisted of the “best” model output by the same procedure used for estimating whale lengths. The unconstrained model was obtained by running the EM algorithm for the same  $k$  as the constrained model, with initial parameter estimates also equal to the constrained model values. Peak definition was measured as the log likelihood ratio between these two models:

$$\log(\Lambda) = \log\left(\frac{\mathcal{L}_{constrained}}{\mathcal{L}_{unconstrained}}\right) \quad (2.7)$$

The argument for using this measure is that individual whales are not expected to differ greatly in their IPI variation, so models with shared variance should fit reasonably well. If the likelihood of unconstrained (and thus potentially overfit) models is considerably better, then this suggests that individual clusters may not be clear. Distributions with well-defined peaks should have a  $\log(\Lambda)$  close to zero. Since  $\log(\Lambda)$  is necessarily zero when  $k = 1$ , those cases were ignored.

To measure accuracy (with respect to manually-compiled IPI distributions), 9 “standard” recordings were selected, which varied in quality from good to poor. Quality was indicated by the mean peak SNR of each detected click. IPIs were compiled manually from each recording using a custom-written MATLAB application. This application allows a user to highlight two pulses in a click waveform, align them, and perform cross-correlation to compute the IPI. To reduce observer bias, it also presents clicks in random order, and obscures IPI values until all clicks have been processed. IPI distributions obtained this way were subsequently screened for potentially erroneous IPIs (indicated by sparse singleton values), which were removed. Manual IPI compilation was performed with no knowledge of the routine-derived distributions for the same files, and was performed by one observer (W.B.). Manual and automatic distributions

were compared through shared-variance GMMs, which were fit using the same process as for length estimation. The “accuracy” of the automatic method was measured as the total overlap in area between the probability density functions of the two GMMs.

Since the automatic routine was developed entirely using clicks obtained from Dominica with a particular recording setup, it is of interest to examine how it performs with different types of recordings. To this end, automatic IPI distributions were also examined separately from 141 recordings made off the Galápagos Islands from January to May 2014. These recordings followed the same “standard” protocol as for Dominica, and were also sampled at 48 kHz with the same research vessel and recording equipment. However, in addition to oceanographic differences, the Galápagos differs significantly from Dominica in that sperm whales are typically grouped in much greater numbers (Whitehead *et al.*, 2012).

## **2.3 RESULTS**

### **2.3.1 Classifier Performance**

Overall accuracy of the SVM in classifying on-axis sperm whale clicks versus other transients, as estimated by 10-fold cross validation before Platt transformation, was 98.8%. When adjusted to account for the imbalance in frequency between each class, accuracy is 94.3%. Sensitivity (a.k.a. true positive rate, or recall) is 89.1%, specificity (a.k.a. true negative rate) is 99.5%, and precision is 92.7%.

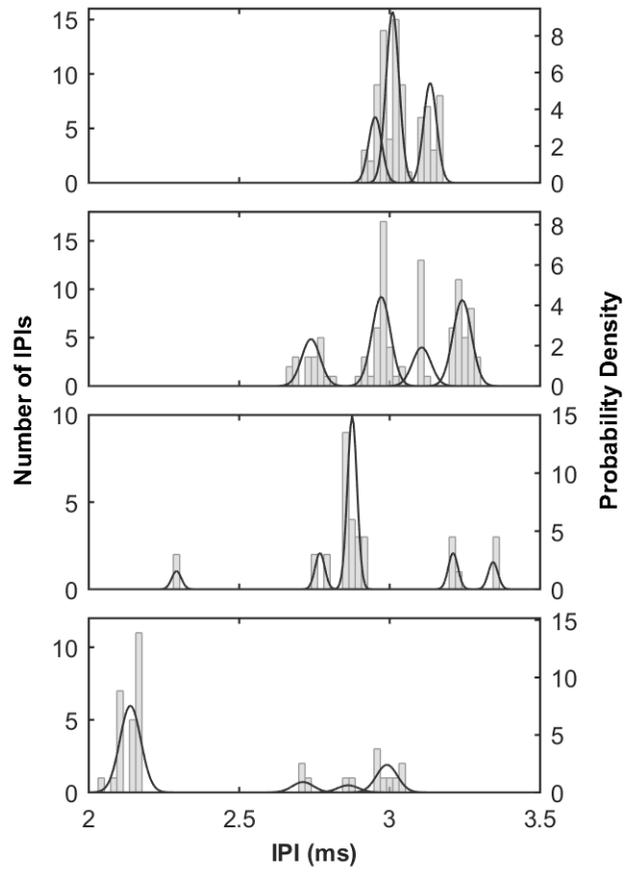
### **2.3.2 Output IPI Distributions**

In many cases, the routine filtered out all IPIs, resulting in empty distributions. This occurred with greatest frequency when the number of required IPI repetitions was high (e.g. 2 repetitions or higher). However, based on visual inspection of each distribution, many of those that were not empty appeared as expected, in the sense that they contained narrow peaks at values appropriate for Caribbean sperm whales (Figure 2.3). Gaussian mixture models also appeared to detect these peaks quite

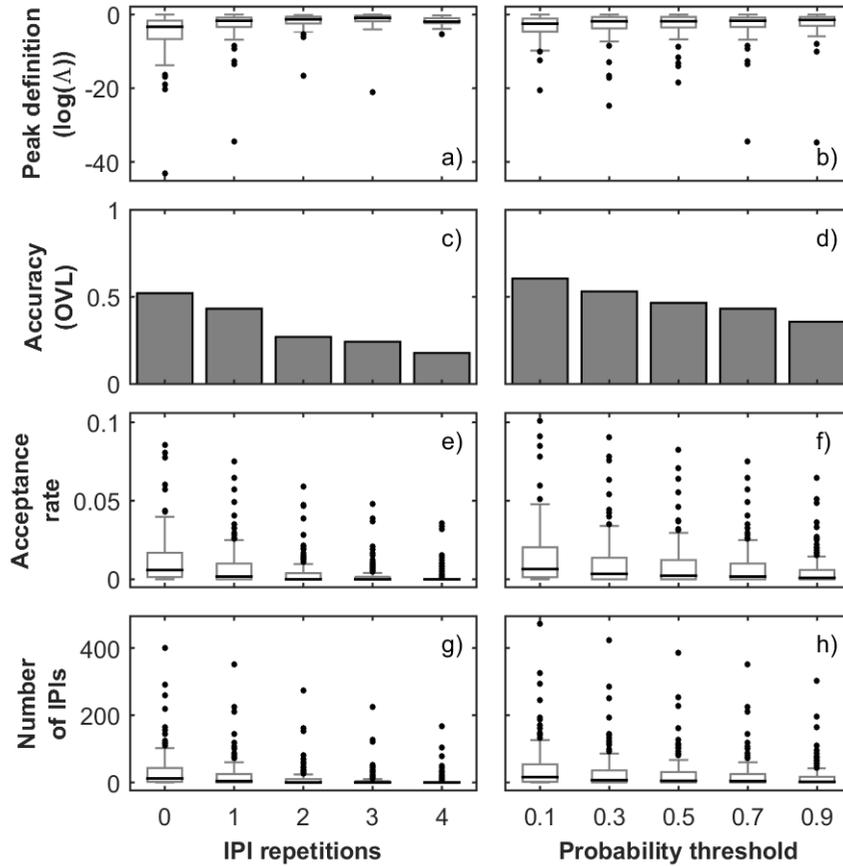
accurately in most cases. Those distributions that were not as clear either had very few IPIs, had many small clusters (most likely noise), or had many peaks very close to each other. The vast majority of noisy distributions occurred when the filter did not require IPIs to be repeated. In general, for distributions with many IPIs, increasing the required number of repetitions resulted in clearer patterns, but at the cost of missing peaks.

Peak definition, as assessed by the likelihood ratio between GMMs with shared and unshared variance, was often worse when IPIs did not need to be repeated. For all cases where IPIs did need to be repeated, peak definition was usually quite good in comparison, and did not change greatly with the number of repetitions (Figure 2.4a) or the probability threshold (Figure 2.4b). Accuracy, as measured by the amount of overlap between the probability density functions of GMMs fit to manually and automatically compiled IPIs, showed a consistent decrease as both filter parameters became more selective (Figures 2.4c, 2.4d). Observing the manual and automatic distributions themselves shows fairly good agreement in the detection of peaks, although some peaks in the automatic distribution appeared to be missing (Figure 2.5a). The acceptance rate is also much lower with automatic filtration (Figure 2.5b). These indicate a high false negative rate.

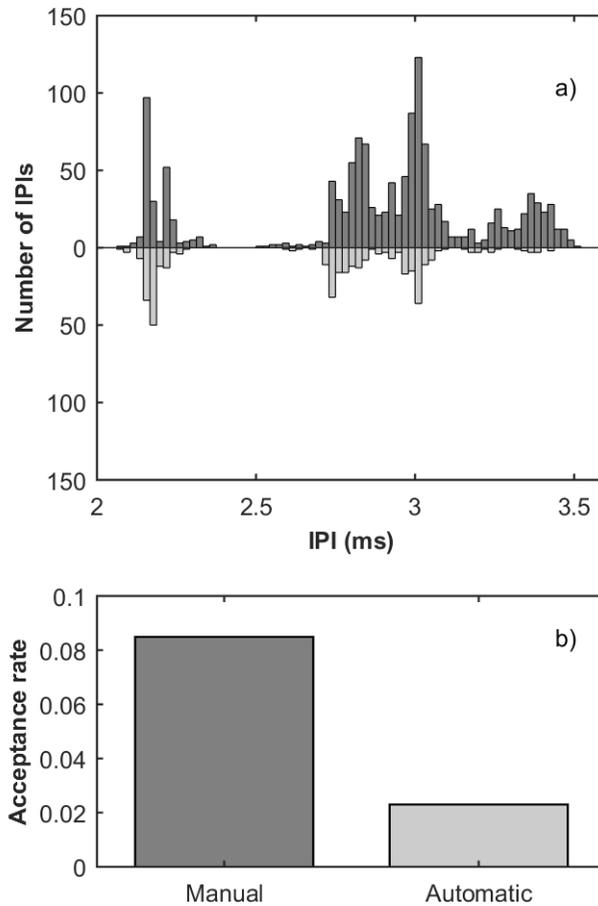
The proportion of detected clicks that are accepted into the final distributions is always very low, below 0.01 in the overwhelming majority of cases, and never above 0.1 (except for one instance). Not surprisingly, it decreases consistently as filter parameters become more selective, but this is much more pronounced with IPI repetition (Figure 2.4e) than with probability threshold (Figure 2.4f). Acceptance rates of 0 are common, and usually represent the majority of cases when IPIs need to be repeated twice or more (depending slightly on probability threshold). However, there was a fair amount of variation in all cases. Though uncommon, it was possible for some distributions produced by the least selective filter to be empty. Likewise, some distributions produced by the most selective filter were larger than average (Figures 2.4g, 2.4h).



**Figure 2.3** Example IPI distributions output from 4 “standard” recordings from Dominica. Filtration parameters were set at 1 IPI repetition, and a “goodness” probability threshold of 0.7. Black lines represent probability density functions of clusters from the best Gaussian mixture models. Bin width =  $1/F_s$ .



**Figure 2.4** Performance of automatic IPI compilation for 174 “standard” recordings from Dominica. The left column shows the effect of variable numbers of IPI repetitions, with the “goodness” probability threshold held constant at 0.7. The right column shows the effect of variable “goodness” probability thresholds, with the number of IPI repetitions held constant at 1.  $\log(\Delta) = \log$  likelihood ratio between mixture models with shared and unshared variance. OVL = overlapping coefficient between probability density functions of mixture models fit to manually and automatically compiled IPI distributions.



**Figure 2.5** Automatic IPI compilation compared to manual compilation for 9 “standard” recordings. Automatic compilation for the case shown here required 1 IPI repetition and a “goodness” probability threshold of 0.7. The dark shade corresponds to manual data. (a) Comparison of IPI distributions, with manual counts on top and automatic counts on the bottom. Bin width =  $1/F_s$ . (b) Click acceptance rates.

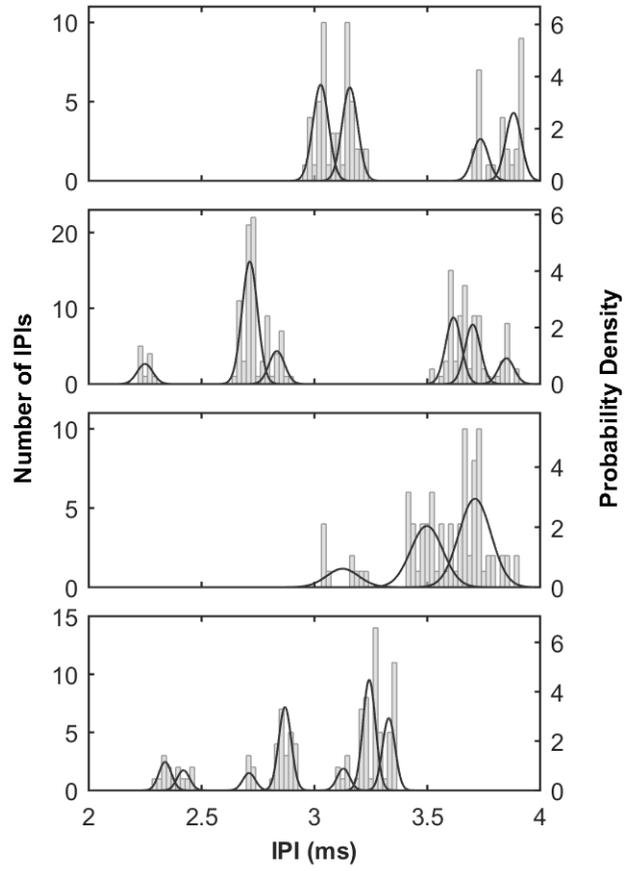
### 2.3.3 Differences between Recording Scenarios

As expected, the frequency of click detections was much greater for the Galápagos (mean = 1390 clicks/minute) than for Dominica (mean = 489 clicks/minute). IPI distributions from the Galápagos were similar to Dominica in that they often contained narrow peaks when enough IPIs were present; however, the density of peaks was generally higher, and individual clusters tended to be more ambiguous (Figure 2.6).

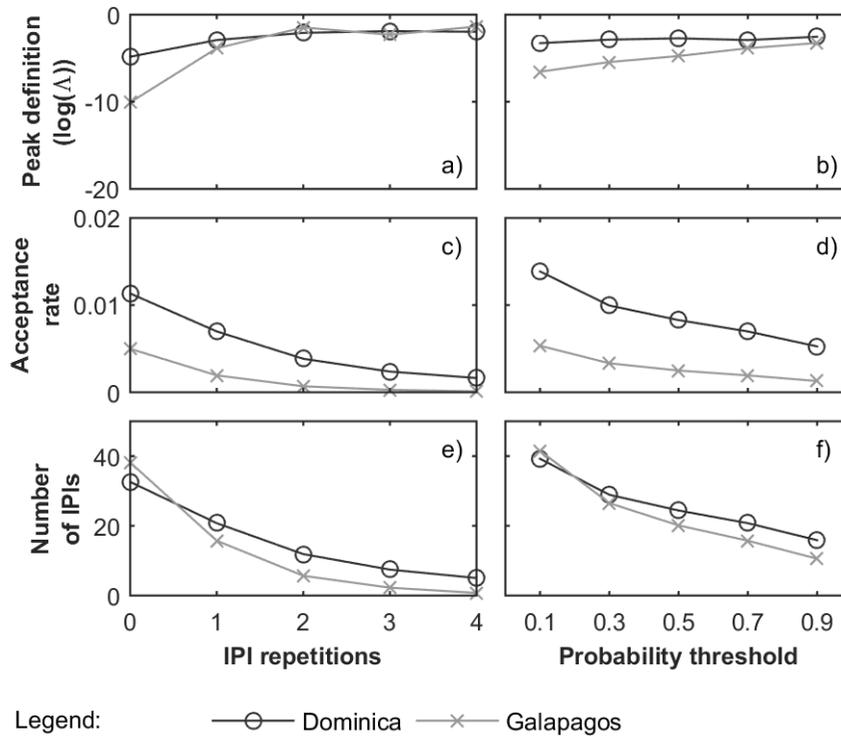
Compared to Dominica, Galápagos distributions showed similar trends in peak definition, but were generally of lower quality. On average, Galápagos peak definition

was comparable to Dominica when IPIs needed to be repeated twice or more, but became progressively worse below two repetitions (Figure 2.7a). When probability threshold was varied, Galápagos distributions showed a slight increase in average peak definition, but the level remained inferior to Dominica (Figure 2.7b). In summary, it appears that peak definition is generally worse for Galápagos distributions, but it improves at a faster rate than for Dominica as filter parameters become more selective.

Regarding click acceptance rate, Galápagos distributions showed the same decreasing trends as for Dominica with both number of IPI repetitions and probability threshold. However, Galápagos distributions consistently had smaller acceptance rates on average than for Dominica (Figures 2.7c, 2.7d). Despite this though, Galápagos distributions contained relatively similar numbers of IPIs as in Dominica (Figures 2.7e, 2.7f).



**Figure 2.6** Example IPI distributions output from 4 “standard” recordings from the Galápagos. Filtration parameters were set at 1 IPI repetition, and a “goodness” probability threshold of 0.7. Black lines represent probability density functions of clusters from the best Gaussian mixture models. Bin width =  $1/F_s$ .



**Figure 2.7** Performance of automatic IPI compilation compared between 174 “standard” recordings from Dominica, and 141 “standard” recordings from the Galápagos. Points represent mean values. The left column shows the effect of variable numbers of IPI repetitions, with the “goodness” probability threshold held constant at 0.7. The right column shows the effect of variable “goodness” probability thresholds, with the number of IPI repetitions held constant at 1.  $\log(\Lambda)$  = log likelihood ratio between mixture models with shared and unshared variance.

## 2.4 DISCUSSION

### 2.4.1 Performance

The automatic IPI compilation algorithm presented here was overall successful in producing IPI distributions straight from single-hydrophone recordings of foraging sperm whales. The support vector machine was shown to be very effective in distinguishing between on-axis sperm whale echolocation clicks and other click types. The full routine, when applied to approximately 4-minute long recordings of sperm whales in Dominica, often produced IPI distributions that contained precise peaks at values that were reasonable for these whales, provided that the IPI filtration parameters were not extreme. Furthermore, the distributions produced by the routine

were similar to those obtained through manual filtration of clicks. Thus, while the absolute accuracy of automatic distributions cannot be known, these results suggest that they are reliable, as far as manual IPI compilation can be considered reliable.

IPI repetition was shown to be a very influential parameter. When IPIs do not need to be repeated, the resulting distributions are likely to contain more IPIs, but in many cases, they are unsuitable for analysis. This is evident from the “peak definition” measure. Most distributions produced without requiring IPI repetition had a distinctively noisy appearance, so it is likely that this noise is responsible for poor peak definition. GMMs, as implemented here, attempt to cluster every sample, so outliers can be problematic. Outliers may be grouped into very small, possibly single-sample clusters, or they may be included with other samples to form wide clusters. In either case, models with shared variance usually do not fit well, because the erroneous clusters are likely to have large differences in variance.

Comparing automatic IPI distributions with manually compiled ones showed that the automatic method becomes less accurate as filter parameters become more selective. This may seem surprising, but it is easily explained. Increasing filter selectiveness results in fewer false positives, but this comes at the cost of more false negatives, or misses. The rate at which the number of misses increases is much greater than the rate at which the number of false positives decreases, which causes overall accuracy to decrease. A consequence of this imbalance is that it is not necessarily desirable to attain maximum accuracy. As shown by peak definition, a modest number of false positives can make it difficult to analyze an IPI distribution. In contrast, false negatives are a nuisance, but they do not complicate analysis to the same degree as false positives. Therefore, a balance needs to be found between the two, with greater weight placed on reducing false positives. In light of this, disabling IPI repetition checks is still not a good option, even though this yields the highest overall accuracy.

This brings up the greatest weakness of automatic IPI compilation: acceptance rate. For all recordings, even the least selective filtration criteria resulted in very small distributions, relative to the total number of clicks that were detected. To a certain

extent, this is expected, given the rarity of clicks with clear multi-pulse structures. Essentially, individual clicks are only suitable for IPI calculation if three criteria are met: 1) the hydrophone must be aligned with the whale's acoustic axis; 2) the click must not coincide with other clicks or echoes; and 3) the click must be significantly louder than background noise. Clearly, the probability that all of these conditions will be true for any click is small, particularly with far-field PAM recordings. Some recordings may be more likely to meet them than others, depending on factors such as noise level, distance from the whales, number of whales, and reflective profile of the environment. Whale orientation, however, is a more random factor, and some recordings may simply be more fortunate than others in the amount of time that whales are aligned with the hydrophone; this would explain why acceptance rate is so variable. However, the rarity of on-axis clicks alone does not explain the routine's particularly low acceptance rate. Recall that there is a large discrepancy in acceptance rate between manually and automatically compiled IPI distributions (Figure 2.5b), which reflects a highly aggressive filter.

While it is not ideal to reject so many positives, this should not be debilitating in practice. The main reason for this is because of the high click rates of sperm whales: 1.2 clicks/second while foraging, according to Whitehead and Weilgart (1990). Thus, good IPI distributions can successfully be obtained from just a few minutes of recording, as is evident from many of the distributions obtained here (e.g. Figure 2.3). Another reason is that it does not take many IPIs to resolve peaks. Since there are few false positives, each IPI in a distribution is very likely to be a true one, at least when each IPI is required to be repeated at least once. Thus, automatic IPI distributions are likely to be reliable even when clusters contain few samples, in that those clusters likely represent the IPIs of some whales in the area. As a recommendation, 5 samples per peak should be a good minimum to indicate the peak's validity (assuming that IPI repetition is enforced). Based on a click acceptance rate of 0.7% (the average across all standard recordings with 1 IPI repetition) and a click production rate of 1.2 clicks per second per whale, it would take about 10 minutes of recording to meet this criterion. However, it is important to

remember that click acceptance rate is extremely variable, so a 10 minute recording could still yield many fewer IPIs, and occasionally many more. If possible, SNR should be maximized to improve the chances of accepting many clicks. Maintaining distance between the hydrophone and reflective surfaces (notably the sea surface) may also improve acceptance, since direct-path clicks will be less likely to overlap with their echoes. However, unless recordings are made specifically after dives, whale orientation will always remain a significant random factor.

On the software end, the only way to substantially improve acceptance rate, aside from relaxing filtration criteria, is by improving the classifier. With an estimated sensitivity of 89.1%, the SVM seems quite good at recognizing on-axis clicks, but this is perhaps not enough. Since each IPI must be repeated to be valid, any on-axis clicks missed by the SVM can further invalidate surrounding clicks by creating gaps in repetition chains. This explains why acceptance rate decreases rapidly as the number of required repetitions increases. Thus, improving the SVM's sensitivity would greatly reduce this problem. However, it should not be done with detriment to specificity, otherwise peak definition may decrease. This may seem difficult, but it should be possible. One factor that likely contributes to classifier confusion is binary classification. This is a problem for two reasons. First, there is no hard separation between on-axis and off-axis clicks. Ideas for solving this problem include fuzzy labelling (i.e. weighting) of training instances, or using a semi-supervised learning approach where "Unsure" clicks are included as unlabelled instances (Schwenker and Trentin, 2014). The second problem is that there are several click types which may exhibit features that overlap with those of the targeted type: for example, on-axis coda clicks, and surface-reflected on-axis clicks. This can be addressed by using more than two classes. In this case, it might work best if classification is done hierarchically, where clicks are given multiple labels: for example, each click could be classified as being a usual, coda, or other click type, being on-axis or not, and being a direct-path or reflected click. This would be especially useful for studies that focus on codas (e.g. Rendell and Whitehead, 2004;

Marcoux *et al.*, 2006; Schulz *et al.*, 2008; Antunes *et al.*, 2011; Schulz *et al.*, 2011; Gero *et al.*, 2016b).

Another factor that could contribute to low acceptance rate is if the dataset used to train the SVM does not fully capture the complete range of possibilities. The “Good” clicks in the training dataset consisted mainly of clicks from whales that started a foraging dive, in the first few minutes of their dive. However, some features, notably spectral ones, are known to change with depth (Thode *et al.*, 2002). Thus, the SVM might potentially have difficulty recognizing clicks from deeper whales. Individual variation in click features might also cause some difficulty, as there were only 10 diving whales in the training dataset. However, the high accuracy reported by cross-validation suggests that this is a relatively minor problem.

Using the current classifier, checking for one repetition is likely the best trade-off in most cases, at least for Dominica surface recordings. If many IPIs are available, the number of repetitions may be increased to further improve the quality of the distribution. In contrast, if few IPIs are available, one technique to improve acceptance rate might be to disable IPI repetition, and then ignore potentially erroneous IPIs (i.e. very small clusters). This would be possible, for example, by applying mixture modelling techniques that are robust to outliers (McNicholas, 2016). Of course though, these distributions would likely not be as precise as when IPI repetition is enforced.

As evident from Figure 2.4, the “goodness” probability threshold does not impact peak definition or acceptance rate as drastically as IPI repetition (between 0.1 and 0.9 at least). This might be a consequence of the binary nature of the SVM, and also the fact that ambiguous clicks were not used to train it. Nevertheless, a value of 0.7 is recommended as a default.

#### *2.4.1.1 Differences between Recording Scenarios*

Compared to Dominica, IPI distributions from the Galápagos generally had more peaks, which were often closely spaced and more ambiguous to interpret. This fits with

groups being considerably larger off the Galápagos (Whitehead *et al.*, 2012). As for the generally poorer peak definition and lower acceptance rate, the most likely explanation is that clicks recorded in the Galápagos were of poorer quality overall, in the sense that few of them had clear multi-pulsed structures. More poor quality clicks would necessarily result in a lower acceptance rate. Peak definition could also be impacted, due to a higher number of false positives: if more poor-quality clicks are present, the SVM has more opportunities to misclassify “Bad” clicks as “Good”. A likely reason why the Galápagos might have poorer clicks is because of its higher click density: when more whales are clicking together, the clicks have a higher chance of overlapping with one another, resulting in a greater proportion of unusable clicks.

Another explanation for the apparent inferiority of Galápagos IPI distributions could be that the SVM does not recognize on-axis clicks from the Galápagos as easily as it does for Dominica. This could happen if the distribution of classifying features differs somehow between Dominica and Galápagos clicks. Such a difference might occur if, for example: sound does not propagate the same way between regions, or the recording setup differed in some way that was not identified. Another plausible cause is that the “voices” of the whales encountered are significantly different between regions. Noise is not likely a factor, since the recordings analyzed here did not differ significantly between regions in this regard.

Ideally, to deal with potential differences between recording scenarios, the classifier should be trained using clicks from each scenario. Unfortunately, this is a time-consuming procedure that must be done by someone who is skilled at recognizing on-axis sperm whale clicks. A simpler but less effective workaround is to try adjusting the IPI filtration parameters. Parameter values could be increased if there appear to be many false positives for example, or perhaps decreased if the number of accepted IPIs is overwhelmingly low.

## 2.4.2 Applications

The ability to measure IPIs automatically should be a great addition to sperm whale PAM. One of the primary goals of marine mammal passive acoustic surveys is abundance estimation, since this is essential for ecosystem and management studies (Mellinger *et al.*, 2007). Abundance estimates depend on the number of animals detected, which can be difficult to obtain through acoustics alone. For sperm whales, counting the number of peaks in IPI distributions could be one way of doing this, with the caveat that only whales of different size will be detected. If similar-sized whales are present, then this count would represent an underestimate, unless additional information is available (e.g. bearing or location). IPIs would also provide information that is usually impossible to get from standard passive acoustic surveys, notably the size of each animal, and to a certain extent, sex (mature males can be identified). This important information must usually be obtained from visual surveys, which are expensive and prone to limitations such as weather and time of day. Thus, through IPIs, acoustic information could be used to compare length and sex distributions between areas, seasons, and different time periods, as well as between social units (Best, 1979; Whitehead *et al.*, 1991) and clans (Rendell and Whitehead, 2003; Gero *et al.*, 2016a).

Since IPI is variable between individuals, this measure could also be used to some extent to track the movements of individual whales or social units. For example, if multiple sensors with the ability to determine IPIs are deployed in an area, IPI “hits” could be compared between sensors over time. If a particular IPI peak is detected at some hydrophone X, and again later at another hydrophone Y, then one could infer that the same whale has traveled from X to Y. Of course, this kind of IPI-based telemetry would be limited by the number of whales in an area that have similar IPIs. It could be particularly useful, though, in areas where whales travel in social units with stable memberships. In this case, the signature of a unit would be a set of IPI peaks. These peak distributions might contain a fair amount of information that could be used to discriminate units with some confidence.

## 2.5 CONCLUSIONS

On-axis sperm whale clicks can quite accurately be recognized by an automatic classifier. From this, an algorithm capable of automatically compiling and analyzing reliable sperm whale IPI distributions directly from acoustic recordings has been developed. The method works with only one audio channel, and does not require knowledge of how many whales are present, or how they are oriented with respect to the hydrophone. Examination of the output IPI distributions show that they often contain clear peaks, and are comparable with manually compiled IPIs. However, the method rejects many more clicks than expected by manual compilation. Fortunately, given the high click rates of sperm whales, recordings of at least 10 minutes are likely to yield enough IPIs to produce clear distributions, although the actual number of IPIs is highly variable and hard to predict. Recordings with high SNR are more likely to accept more IPIs, but whale orientation is a significant source of variation. Based on the current implementation, filtration parameters may need to be adjusted to accommodate different recording scenarios. In the long term, expansion of the classifier's training dataset with a wider variety of clicks (e.g. from whales at greater depths, mature males, bottom-mounted hydrophones, new regions, etc.) may enable it to perform better under a wider variety of scenarios. Modifications to the classification model might also improve acceptance rate.

The software should be a useful extension to sperm whale PAM. The ability to obtain IPIs, and consequently body length estimates from sperm whales without the need to tag or even see them, should be a great advantage for studying their abundance, movements, and behaviour.

## CHAPTER 2 ENDNOTES

<sup>1</sup>The work presented in this chapter was submitted for publication in the Journal of the Acoustical Society of America. The manuscript is authored by Wilfried A. M. Beslin (WB), Hal Whitehead (HW), and Shane Gero (SG). Author contributions: WB designed and wrote the algorithm, performed all analyses, and wrote the manuscript; HW supervised the project, edited the manuscript, and contributed funds for field data collection; SG contributed funds for and coordinated the field data collection in Dominica, and edited the manuscript. All authors collaborated in the conception of the project, participated in data collection, and gave their final approval for submission of the manuscript. The manuscript was first submitted on June 20, 2018.

<sup>2</sup>There is a slight difference depending on whether the click is recorded in front (forward) or behind the whale (backward). According to Zimmer et al. (2005), the first pulse in forward clicks has a much smaller amplitude than the second. This type of click is typically less common than backward clicks, and was not present in the “first-click” recordings used here. However, since the click detector breaks up areas that do not follow a progressive decay pattern, this first pulse should be omitted. Since the remaining parts of forward clicks resemble backward ones, it is likely that the routine will still recognize forward clicks as being on-axis.

# **CHAPTER 3 - EVALUATION OF A METHOD FOR AUTOMATICALLY IDENTIFYING EASTERN CARIBBEAN SPERM WHALE SOCIAL UNITS FROM THEIR INTER-PULSE INTERVAL DISTRIBUTIONS**

## **ABSTRACT**

Passive acoustic monitoring (PAM) is a very effective technique for studying cetaceans, but the information it provides is often limited. For sperm whales, PAM can be especially informative: sperm whale clicks are composed of multiple pulses, where the inter-pulse interval (IPI) bears a relationship with body length. The goal of this project was to develop a PAM-based method for automatically identifying individual sperm whale social units in the Eastern Caribbean, based on the premise that each unit has a distinctive IPI distribution. To test this, IPIs were measured automatically from 225 recordings taken off the coast of Dominica, where unit identity was known from photo-ID. IPI profiles were constructed for 5 units, based on kernel density estimates of IPIs from recordings with only one unit present. These were then fit to IPIs from other recordings. Results showed that IPI profiles could distinguish between units in most cases (60%-100%), except for one unit (0%), which had a poorly characterized IPI profile. Unit detection overall was fully successful about 30% of the time, and was most accurate when only one unit was present. As it stands, this method is not reliable, but it has the potential to be improved.

## **3.1 INTRODUCTION**

The re-sighting of known individuals across space and time can provide important information on population biology (Lebreton *et al.*, 1992; Schwarz and Seber, 1999), movements and habitat use (Whitehead, 2001; Ovaskainen *et al.*, 2008), and social structure (Whitehead, 2008). In turn, this knowledge is needed to make effective management decisions. For whales and dolphins, the photography of natural markings

has proved to be an extremely effective technique for recognizing individuals, and is now a standard protocol for many field studies (see Hammond *et al.*, 1990). However, photo-identification depends on visual surveys, which are often expensive and challenging to conduct for cetaceans, since these animals spend most of their lives underwater and can live far from shore. Another method frequently used to study cetaceans is passive acoustic monitoring (PAM). Owing to the excellent propagation of sound through water and the highly vocal nature of cetaceans, PAM has the advantage that it can detect animals that may be several kilometers away (Barlow and Taylor, 2005; Stafford *et al.*, 2007), in conditions that may be unfavourable for visual observations (Thomas *et al.*, 1986b). However, the information that can be obtained through PAM is typically limited to presence/absence of a species; individual identification through PAM is usually very difficult, if not impossible.

The sperm whale is particularly well suited to study through passive acoustics. Sperm whales are deep divers that spend most of their time foraging at depth (Watwood *et al.*, 2006), during which they regularly produce loud clicks that can be heard from several kilometers away (Backus and Schevill, 1966; Whitehead and Weilgart, 1990; Barlow and Taylor, 2005). Additionally, these clicks exhibit a unique structure that has not been observed in other cetacean species. This structure reveals morphometric information that is ordinarily not possible to obtain through PAM. A single sperm whale click is composed of multiple pulses (Backus and Schevill, 1966), where the inter-pulse interval (IPI) corresponds to the time taken for sound to travel back and forth through the spermaceti organ (Norris and Harvey, 1927; Møhl, 2001). Since there is a relationship between the size of the spermaceti organ and total body size (Nishiwaki *et al.*, 1963; Clarke, 1978; Gordon, 1991), it follows that total body length can be estimated by measuring the IPI (Norris and Harvey, 1972; Adler-Fenchel, 1980; Møhl *et al.*, 1981; Gordon, 1991). Clear relationships between IPI and body length have been demonstrated for both female and male sperm whales (Gordon, 1991; Rhinelander and Dawson, 2004; Growcott *et al.*, 2011). While not an individually

distinctive signature, body length, or IPI, can provide useful information for the purpose of identification, especially for groups of individuals.

Sperm whale social structure is characterized by sexual segregation. As a species, they are globally distributed, but males and females live separate lives. Females and their calves live in tropical and subtropical waters year-round, while mature males reside in higher latitudes, occasionally revisiting the warmer waters to mate (Best, 1979). Females are especially social and exhibit multiple levels of association. The fundamental element of female society is the social "unit", which is a small (3-24 member) grouping of individuals that are often matrilineally related (Best, 1979; Whitehead *et al.*, 1991; Christal *et al.*, 1998, Gero *et al.*, 2014). Social relationships among members of a unit are very stable, lasting on the scale of decades (Christal *et al.*, 1998; Gero *et al.* 2014; Gero *et al.*, 2015). When they are not foraging, unit members can often be found socializing at the surface in small clusters (Whitehead and Weilgart, 1991). Unit members also travel together over large distances in search of food, often 1000 km or more (Dufault and Whitehead, 1995; Whitehead *et al.*, 2008), and occasionally associate with other units for a few hours or days (Whitehead *et al.*, 1991; Gero *et al.*, 2014). Monitoring the movements of individual units can provide a more biologically relevant basis on which to manage them, allowing for more effective conservation (Gero *et al.*, 2007).

The purpose of this project is to evaluate a method for detecting and identifying individual sperm whale units through PAM, based on inter-pulse intervals. The premise is that each unit in a population is likely to exhibit a distinctive distribution of IPIs, which may be used to identify them. In particular, a small population of female and juvenile sperm whales roaming the Lesser Antilles in the Eastern Caribbean is the focus of this project. These whales have been studied off the coast of Dominica regularly since 2005 as part of a long-term behavioural research project (see Gero *et al.*, 2014), making them the best studied sperm whales in the world. Sperm whale units off Dominica are small (about 7-9 animals) and exhibit high site fidelity; as of 2012, 17 units have been re-identified across multiple years, including one which was first documented in 1984

(Gero *et al.*, 2014). Individual whales off Dominica and their unit memberships are very well characterized, making this population ideal for testing the potential of automatic unit recognition.

## **3.2 METHODS**

### **3.2.1 Data Collection**

The data used in this study come from the long-term behavioural research program described by Gero *et al.* (2014). To avoid changes in IPI due to growth, only one year of data was used. Specifically, this study used data collected off the west coast of the island of Dominica (approximately 2000 km<sup>2</sup> coverage) from February-April 2015 (56 days effort). Research was conducted aboard a 12-m auxiliary sailing vessel. The field protocol involved tracking female and immature sperm whales both visually (searching for blows) and acoustically (listening for clicks). When clusters of whales were spotted at the surface, they were gently approached and followed. Once the whales prepared to dive, photographs of their exposed flukes were taken for individual identification. The date and time of each photograph was saved. Photographs were only used for analysis if the identity of the target individual was certain.

Acoustic recordings usually lasting about 3-7 minutes were made immediately after each whale fluked, thereby capturing their first echolocation clicks from behind. Recordings were also made periodically (usually 1-hour intervals) when whales were foraging. These usually lasted for about 4 minutes. The recording equipment consisted of a custom-built towed hydrophone (Benthos AQ-4 elements, frequency response 0.1 – 30 kHz) and a filter box with high-pass filters up to 1 kHz. This resulted in a recording chain with a flat frequency response across a minimum of 2 – 20 kHz. Audio data were collected through a computer-based recording system, with a sampling rate of either 48 kHz or 96 kHz, and bit depth of 16 bits per sample. All recordings were stored in WAVE format. In total, 225 recordings were used, representing a total duration of about 20 hours and 24 minutes. Both of the recording types typically included the echolocation clicks of some, but not all, sperm whales in the social unit(s) being tracked. Some of

these recorded clicks will have been on-axis (especially those from a diving whale with the hydrophone in the dive slick) and so potentially yielding useful IPIs. However many of the recorded clicks are off-axis and so do not give reliable IPIs (see Chapter 2).

Individuals that were photographed at least once within the interval spanning 2 hours prior to the start of a recording, and 2 hours after the end of the recording, were considered to be present during that recording. This information was used to determine which units were present during each recording. Dependent calves, which usually do not fluke, were not included. Unit membership was assigned as in Gero *et al.* (2014), with the exception that units “F” and “U” were treated as one unit (see Appendix C). In addition to the commonly encountered units, two extra groupings were designated: “UNK”, to which all unknown, unidentified, or uncommon individuals were assigned; and “MALE”, to which all mature males were assigned. Mature males are clearly distinguishable from females based on their considerably larger size (Best, 1979; Best *et al.*, 1984) and distinctive acoustics (Weilgart and Whitehead, 1988). All individuals encountered off Dominica in 2015 are listed in Appendix C.

### **3.2.2 Algorithm**

The algorithm for detecting which units are present based on IPIs requires two steps. First, a “profile” of the IPI distribution of each unit must be constructed. Once these are available, the profiles are compared to IPI values extracted from audio recordings.

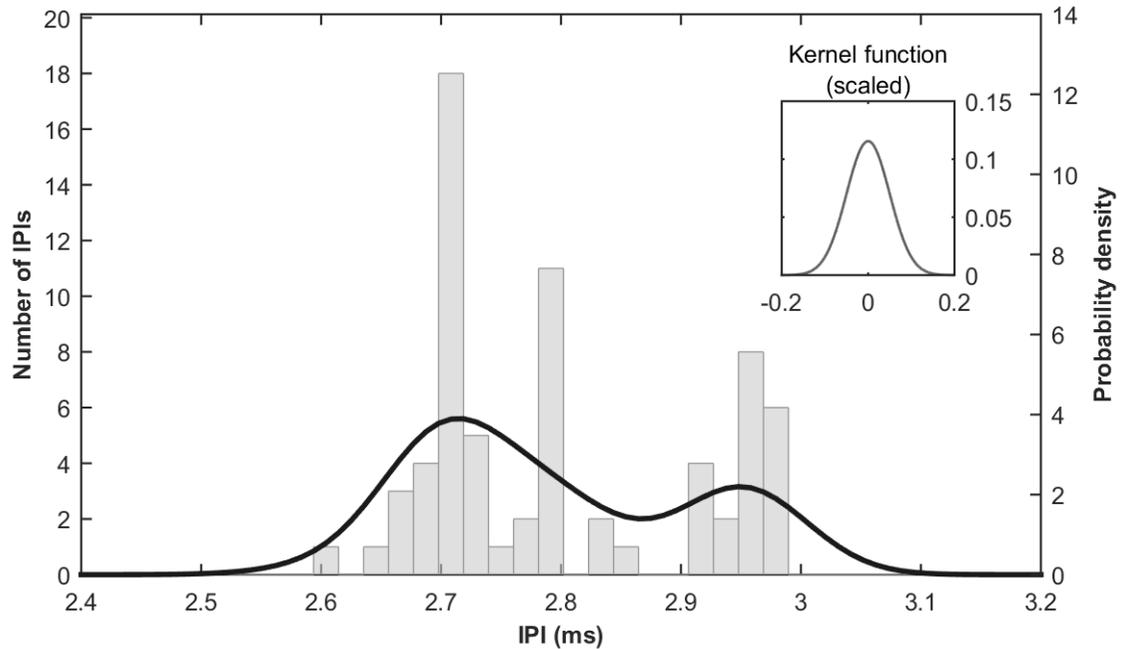
#### *3.2.2.1 Construction of Unit IPI Profiles*

A unit’s IPI profile consists of a probability density function describing the likelihood of obtaining a particular IPI from that unit. An ideal way to represent this is as a Gaussian mixture model, where each component corresponds to an individual whale. In this case, the means of each component correspond to the “true” IPIs of each whale, and the standard deviations represent the variation in measurement of an individual’s

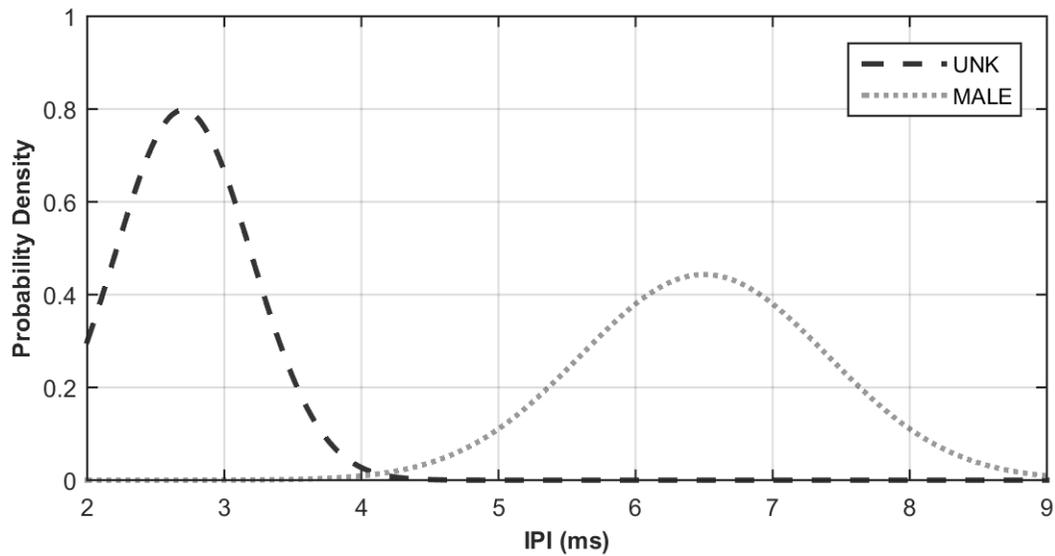
IPI. However, this requires knowledge of the “true” IPIs of every whale in a unit. While it is possible to estimate this (see Chapter 2), a high degree of confidence in the IPI measures is needed to accurately represent units as mixture models. This is very difficult to obtain, especially when individuals are of similar size. Therefore, for this analysis, an alternative approach was used. IPI profiles were constructed as Gaussian kernel density estimates, applied to IPI values obtained from all recordings where only one unit was found to be present based on photo-ID (Figure 3.1). One disadvantage of this approach is that component proportions cannot be standardized as they could (and should be) with GMMs. In other words, if more IPIs are available for some unit members than others, then this discrepancy will be incorporated into the unit profile when using KDE. Consequently, the unit may be harder to identify when only lesser represented members are present. This would not be a problem using GMMs, since in this case the relative contribution of each cluster (corresponding to one whale) can easily be adjusted.

The kernel bandwidth (equal to standard deviation for a Gaussian kernel) was set to 0.05 ms, a value that approximates the measurement error of a whale’s IPI. (Schulz *et al.*, 2011; Chapter 2). Generic profiles were also created for the *UNK* and *MALE* groupings. Both of these consisted of Normal distributions, with parameters designed to span the entire range of IPIs for their respective sex/age classes. For *UNK*, which represents females and immature males, mu and sigma were set to 2.7 and 0.5 ms, respectively. For *MALE*, which represents mature males, mu and sigma were set to 6.5 and 0.9 ms, respectively (Figure 3.2). These parameters were determined based roughly on IPIs observed from whales off Dominica and from Growcott *et al.* (2011). There are two advantages to representing unknown individuals and mature males as additional independent profiles. One advantage is that these distinct sex/age classes can be identified separately. For example, if a mature male enters an area, the routine could automatically identify him as being a mature male. The other advantage of generic profiles is that they potentially enable the simultaneous detection of known and unknown whales. For example, if a known and unknown unit are interacting, both

groupings could conceivably be identified at once, rather than detecting only the known unit, or marking all whales as unknown.



**Figure 3.1** Example of IPI profile creation using kernel density estimation. Histogram shows the distribution of IPI values extracted from a group of recordings. The thick solid line is the IPI profile, which is a kernel density estimate of the IPI distribution. The kernel density estimate was created by adding several Gaussian probability density functions (the kernel function, top-right inset), such that the mean of each Gaussian corresponded to one of the extracted IPI values. In other words, a Gaussian function was centred over every IPI value, and the sum of each Gaussian produced the density estimate. Kernel bandwidth ( $\sigma$ ) = 0.05 ms.



**Figure 3.2** IPI distribution profiles for the generic *UNK* and *MALE* groupings. These are designed to represent the full range of IPIs for female/immature sperm whales, and mature male sperm whales, respectively. The 2 and 9 ms cutoff points are based on Marcoux *et al.* (2006). Parameters were determined in consideration of observed IPI values from whales in Dominica (both male and female) and from Growcott *et al.* (2011).

### 3.2.2.2 Inferring Unit Presence

Once profiles have been established, the algorithm uses them to construct a set of “presence models”. A presence model consists of one or more unit IPI profiles combined together. In the implementation presented here, since IPI profiles consist of kernel density functions, this means that a presence model is effectively a mixture of kernel density distributions, where each mixture component represents the IPI distribution of a particular unit. All components in a presence model are scaled such that the combined probability density function integrates to 1. The complete set of presence models consists of every possible combination of profiles, including singletons. So, for example, if 5 profiles are available, the total number of presence models would be  $2^5 - 1 = 31$ .

Presence models are used to infer which units are present in an audio recording. After stable IPIs have been measured from the recording, the likelihood of every model as a fit for the IPI distribution is calculated. Models are then ranked according to the Bayes Information Criterion (BIC), which penalizes the likelihood for the number of components (i.e. units) included in the model. The model with lowest BIC is the most supported. Each model is assigned a  $\Delta$ BIC score, which represents the deviation in BIC from the best supported model. Thus, the best supported model has a  $\Delta$ BIC of 0. Admittedly, one limitation of this approach is that there can be a certain degree of uncertainty in which model is actually “correct”, and it may not always be the case that this model is the one with the absolute lowest BIC score. This is a drawback of automated inference using BIC. Nevertheless, BIC remains an effective method that is capable of selecting appropriate models in most cases. For the purpose of automatic unit identification, all units whose profiles compose the best supported model, based on BIC, are considered to be “detected”.

### **3.2.3 Performance Analysis**

To construct unit IPI profiles, recordings with only one unit present, based on photo-IDs, were isolated and grouped together by unit. To test the performance of the unit detection algorithm, all recordings were grouped based on the day in which they were recorded, and the units known to be present. Refer to Appendix C for information on these groups and how they were used. Stable IPIs were extracted automatically from all recordings using the approach outlined in Chapter 2, where the number of required IPI repetitions was set to 1, and the probability threshold for “goodness” of clicks was set to 0.7. The smallest IPI that can be measured by this routine is 2 ms, which excludes dependent calves (Marcoux *et al.*, 2006). This routine also fits Gaussian mixture models to the output IPIs. These models were used to eliminate potential false positives by removing all IPIs that were assigned to very small clusters (i.e. Gaussian components containing < 5 IPIs). IPI distributions from recordings that were part of the same group (i.e. recorded on the same day, with the same combination of units present) were

merged together. The automatic unit detection algorithm was run on each of the resulting IPI distributions.

Automatic unit detection was tested in two ways, for two different purposes. The first test sought to examine how well units can be recognized, or differentiated from one another, based on IPI profiles. To do this, only models consisting of one unit were compared, using only those recordings where just one unit was known to be present. This will be referred to hereafter as the “Singles-Singles” test. The second test sought to examine how well the algorithm performs in a broader context when several units may be present. This involved comparing all presence models, using all recording groups. Results of this test were also broken up to examine the performance of the algorithm under different scenarios. Three scenarios were recognized:

- 1) “Singles”: only one familiar unit is present
- 2) “Multiples”: two or more familiar units are present
- 3) “Unknowns”: any case where unknown individuals or mature males are present.

This second test as a whole will be referred to as the “All-All” test.

In both tests, the analysis of groups with only one unit present required an additional step. Since all recordings in these groups were used to create IPI profiles, care needed to be taken to ensure that the same recordings were not used for both the creation of a unit profile, and the testing of that profile’s ability to identify the unit. This was achieved using a cross-validation approach. When testing on each group of recordings with only one unit present, an incomplete profile was used. Incomplete profiles were created using all recordings containing only the unit of interest, except for those recordings being tested.

Performance was assessed in two ways. The first measure was the  $\Delta$ BIC score of the “true” model (that is, the model that contained all units known to be present from photo-ID). The second measure was the routine’s ability to correctly detect units. Unit detection was formulated as a binary classification problem, where the dataset consisted of individual units that could belong to one of two categories: “present” or

“absent”. A unit was considered *actually* present if one or more members were photo-identified around the time of recording, and was *predicted* to be present (or “detected”) if its IPI profile was part of the best-supported presence model. In this framework, the accuracy of unit detection can be summarized as *precision* and *recall*. These were computed for the “All-All” test, both as a whole and for each scenario separately. For each of these analyses, the datasets of all relevant groups of recordings were combined as one dataset. Thus, precision and recall were computed as follows:

$$precision = \frac{\sum_i TP_i}{\sum_i TP_i + \sum_i FP_i} \quad (3.1)$$

$$recall = \frac{\sum_i TP_i}{\sum_i TP_i + \sum_i FN_i} \quad (3.2)$$

where:

- $TP_i$  = the number of units in group of recordings  $i$  that were *predicted to be present*, and were *actually present* (i.e. correct detections)
- $FP_i$  = the number of units in group of recordings  $i$  that were *predicted to be present*, but were *actually absent* (i.e. incorrect detections)
- $FN_i$  = the number of units in group of recording  $i$  that were *predicted to be absent*, but were *actually present* (i.e. misses)

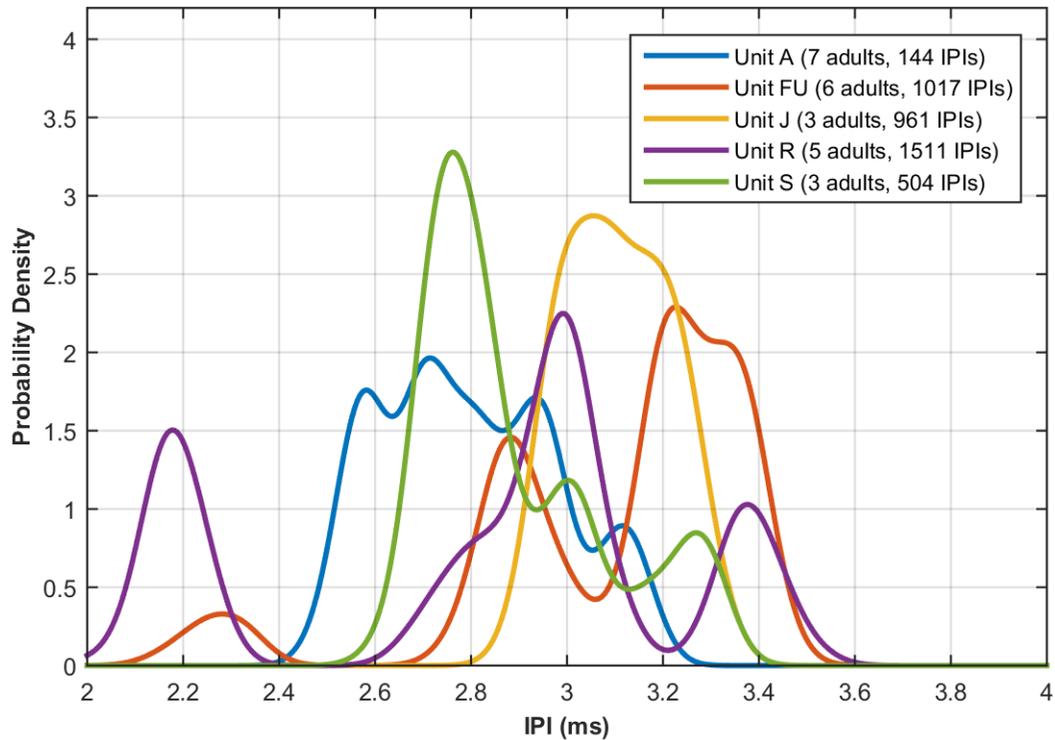
In other words, *precision* is the proportion of detections reported by the routine that were correct, and *recall* is the proportion of units actually present that were successfully detected by the routine.

### 3.3 RESULTS

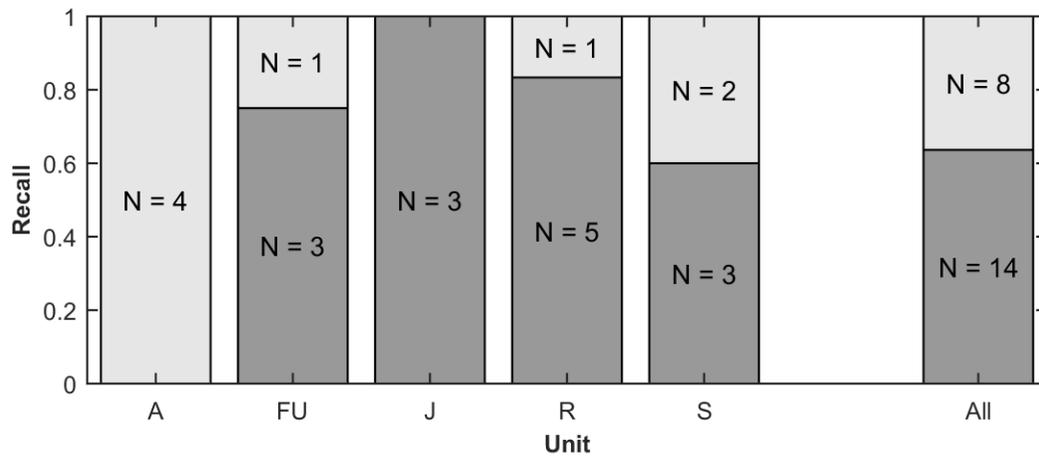
#### 3.3.1 Unit IPI Profiles and Ability to Differentiate Between Units

Unit IPI profiles could be obtained for 5 common sperm whale units off Dominica:  $A$ ,  $FU$ ,  $J$ ,  $R$ , and  $S$ . Profiles overlapped considerably in the IPI values they

covered, but each exhibited a unique distribution (Figure 3.3). Based on the “Singles-Singles” test, IPI profiles constructed this way were capable of identifying the correct unit from extracted IPI values about 64% of the time. However, success was not uniform across units. In particular, the test consistently failed to recognize unit A. Success in identifying all other units varied between 60% and 100% (Figure 3.4).



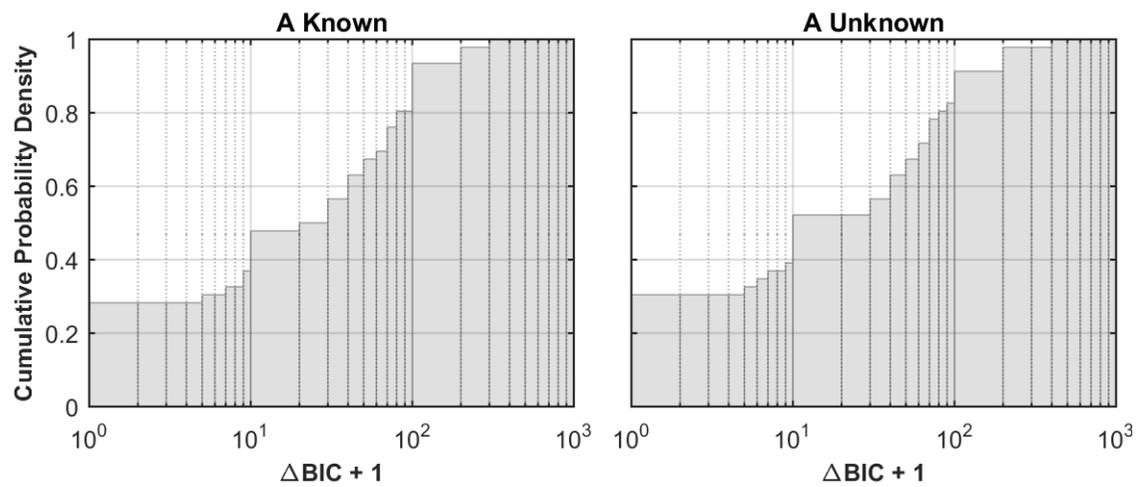
**Figure 3.3** IPI distribution profiles for common Dominica sperm whale units. These profiles were obtained by performing kernel density estimates on stable IPIs extracted from all recordings in 2015 where only the unit of interest was known to be present. Kernel bandwidth = 0.05. Number of individuals represents the number of unit members present during the recordings, excluding dependent calves, who’s IPIs are not expected to have been measured. Individual presence was based on whales photo-identified within  $\pm 2$  hours of the start and end times of each recording.



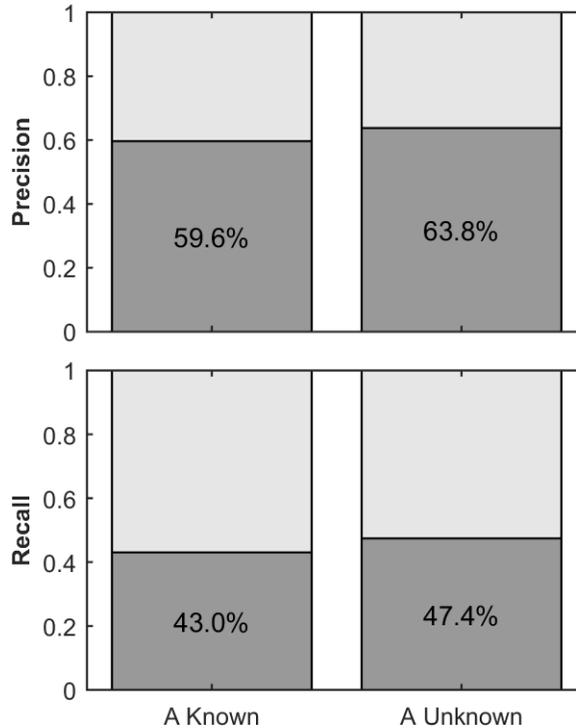
**Figure 3.4** Proportion of correct (dark) versus incorrect (light) identifications of the target unit based on the “Singles-Singles” test. This test involved running the automatic unit detection algorithm on only those recordings where just one unit was known to be present, and only models consisting of one unit were considered. Combinations of units and unknown/male individuals were not part of this analysis. “N” corresponds to the number of recording groups.

### 3.3.2 Overall Performance

In light of the apparent inability to recognize unit A, the “All-All” test was performed under two different cases: one where all unit profiles were used, and another where unit A was considered to be an unknown unit. However, when examined as a whole, the routine did not show a notable difference in performance between the two cases. The distribution of  $\Delta$ BIC scores for the “true” model was similar whether unit A was treated as known or not (Figure 3.5). In both cases, the routine detected all units correctly around 29% of the time. Regarding the ability to detect individual units correctly, the routine performed slightly better when unit A was unknown, but not by very much (Figure 3.6).



**Figure 3.5** Cumulative probability distributions of the  $\Delta\text{BIC}$  scores of the “true” models for the “All-All” test when unit A was considered known (left) and unknown (right). The “All-All” test involved running the automatic unit detection algorithm on all recording groups, with models for every possible combination of units considered, including unknown and/or male groupings.



**Figure 3.6** Precision and recall of unit detection for the “All-All” test when unit *A* was considered known (left) and unknown (right). This represents the routine’s ability to detect a unit when that unit is present (including *UNK* and *MALE*), and to not detect it when it is not.

Precision = proportion of all reported detections that are correct;

Recall = proportion of units actually present that were successfully detected.

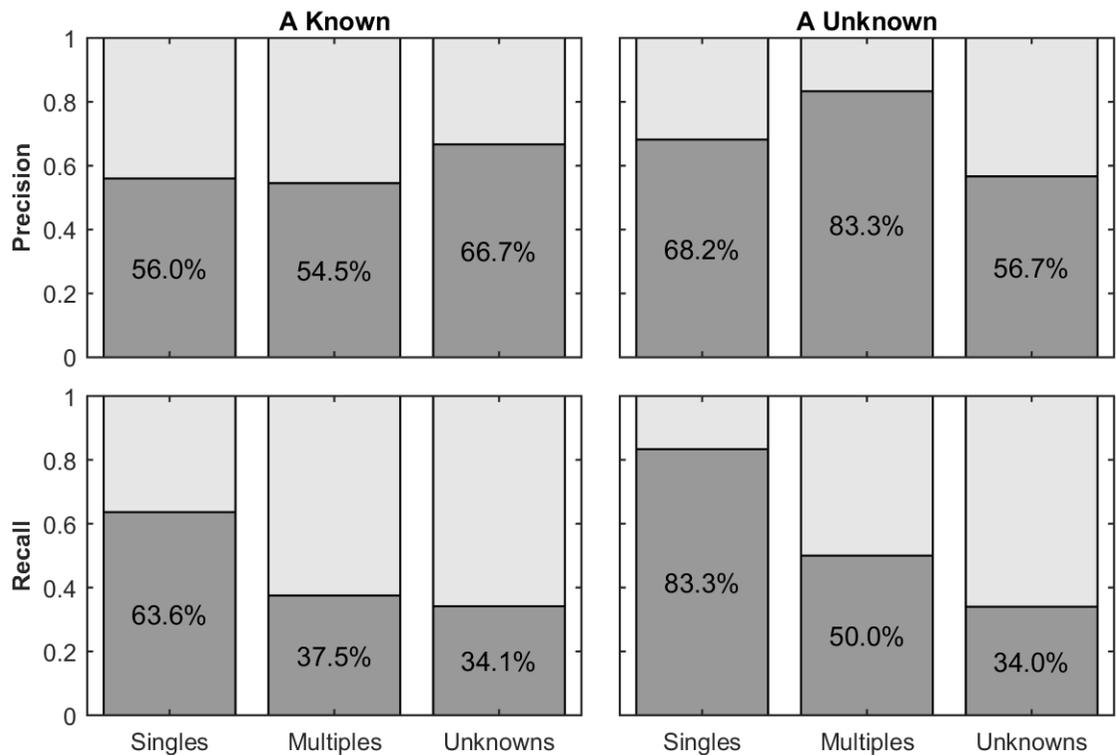
### 3.3.3 Performance by Unit Presence Scenario

Precision and recall for each scenario (“Singles”, “Multiples”, and “Unknowns”) are summarized in Figure 3.7. When broken down by scenario, the “All-All” test exhibited more pronounced differences depending on whether unit *A* was treated as a known or unknown unit. In both cases, the “Singles” scenario had better recall than “Multiples” and “Unknowns”, indicating that the routine could correctly detect present units more easily when they were alone. Recall for the “Singles” and “Multiples” scenarios was higher when unit *A* was considered to be unknown, indicating that the routine correctly detected a greater proportion of known units that were present when unit *A* was unknown. However, recall for the “Unknowns” scenario did not show a great

difference whether unit *A* was known or unknown. This indicates that the routine correctly detected approximately the same proportion of units and/or unknown whales that were present in both cases, for those situations where unknown whales were present.

Precision, which represents the proportion of all detections reported by the routine that were correct, showed a slightly different pattern than recall. In particular, when unit *A* was unknown, the routine was 10% less precise during the “Unknowns” scenario. In other words, for situations in which unknown and/or mature male whales were present, a smaller proportion of the detections reported by the routine were correct when individuals in unit *A* were considered to be unknown. Precision was comparable between “Singles” and “Multiples” scenarios when unit *A* was known, suggesting that particular units can be successfully detected at the same rate whether they are alone or in groups. When unit *A* was unknown, the “Multiples” scenario appeared to have greater precision than “Singles”. However, this should be interpreted with caution. It must be noted that the “Multiples” scenario had few occurrences, particularly when *A* was unknown, making it difficult to assess precision with confidence in this case. The total number of detections for the “Multiples” groups was 11 when unit *A* was known, and 6 when it was unknown. This is compared to 25 and 21 when *A* was known, and 22 and 30 when *A* was unknown, for the “Singles” and “Unknowns” scenarios, respectively.

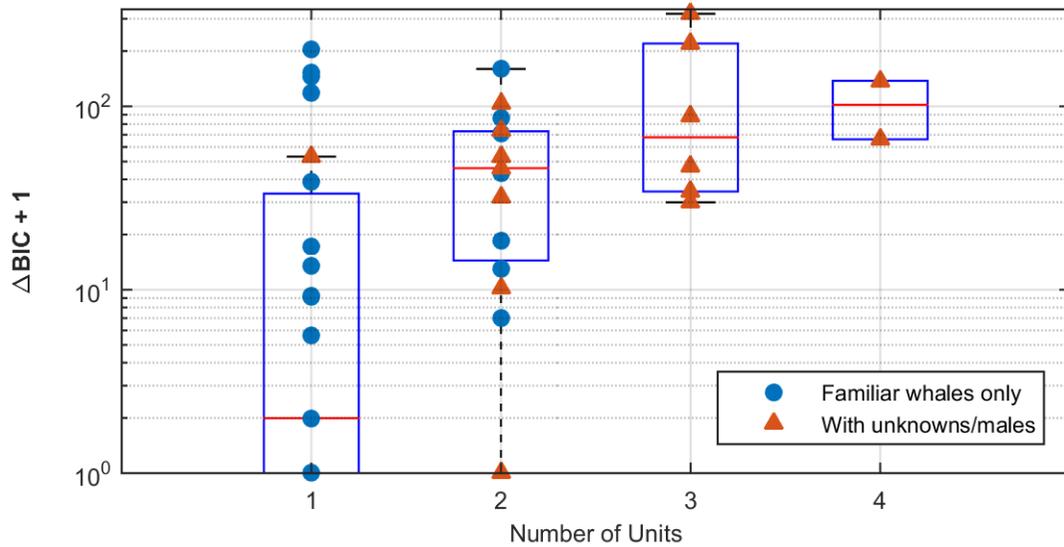
To obtain more insight into other factors that may affect performance, the  $\Delta$ BIC scores of the “true” models were compared against number of units present, and also the number of IPIs available for inference. This showed that the routine generally had more difficulty identifying all units present as more units were present together (Figure 3.8). There was also a slight indication that the “true” model was more likely to be selected when more IPIs were available, though this was apparent mostly when only one unit was present (Figure 3.9).



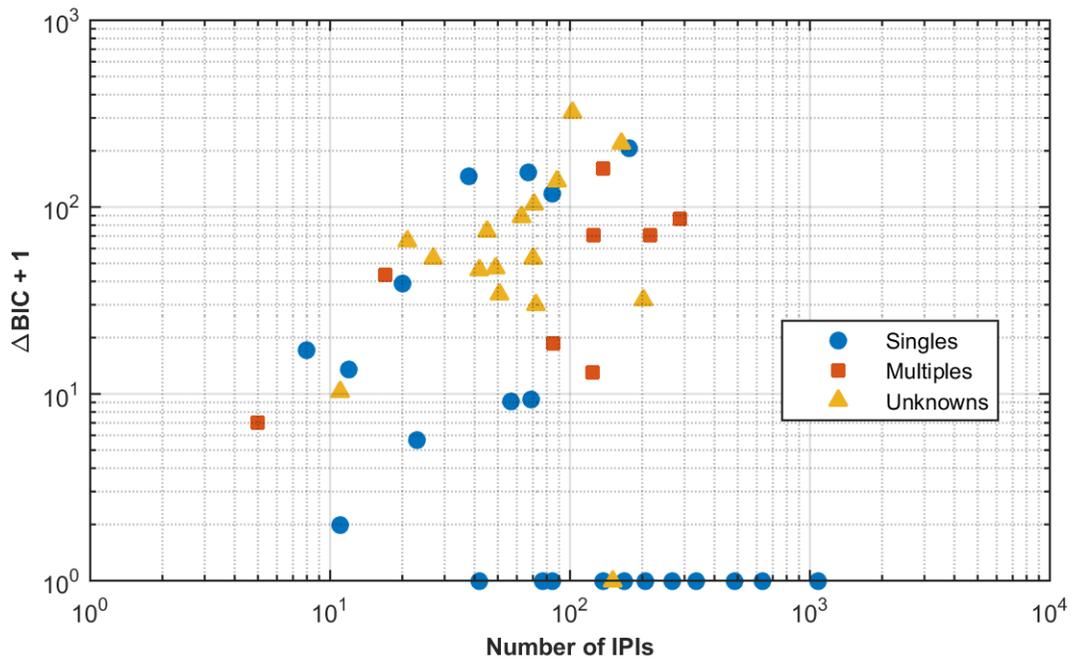
**Figure 3.7** Precision and recall of unit detection for the “All-All” test when unit *A* was considered known (left) and unknown (right), broken down by scenario. This represents the routine’s ability to detect a unit when that unit is present (including *UNK* and *MALE*), and to not detect it when it is not, for the cases where 1) only one known unit is present (“Singles”), 2) two or more known units are present (“Multiples”), and 3) unknown and/or male whales are present (“Unknowns”).

Precision = proportion of all reported detections that are correct;

Recall = proportion of units actually present that were successfully detected.



**Figure 3.8** Trend in performance of the automatic unit detection routine with number of units actually present (as determined by photo-ID). Unknown whales are treated as one unit. In this case, unit *A* is considered known. Circles represent the “Singles” and “Multiples” scenarios (i.e. known whales only), whereas triangles represent the “Unknowns” scenarios (where unknown whales are present, either alone or with known units). Performance is measured as the  $\Delta BIC$  score of the “true” model, where  $\Delta BIC = 0$  indicates that the routine has successfully selected the “true” model as being most likely. Thus, large  $\Delta BIC$  values indicate poor performance. The “true” model is that which includes all units known to be present based on photo ID. Boxplots illustrate the density distribution of  $\Delta BIC$  values for each group composition (i.e. number of units).



**Figure 3.9** Trend in performance of the automatic unit detection routine with number of IPIs used to make the inference. In this case, unit A is considered known. Performance is measured as the  $\Delta\text{BIC}$  score of the “true” model (i.e. the model containing all units actually present, based on photo-ID).  $\Delta\text{BIC} = 0$  indicates that the routine has successfully selected the “true” model as being most likely. Thus, large  $\Delta\text{BIC}$  values indicate poor performance. However, it should be noted that BIC is inherently dependent on the amount of data available, such that differences in BIC between models tend to become larger as more data are available. Thus, the apparent increase in  $\Delta\text{BIC}$  with number of IPIs for less supported models (represented by the cloud on the left) likely does not represent a decrease in performance with increasing number of IPIs. The notable trend in this figure is the difference in number of IPIs between cases where  $\Delta\text{BIC}$  equals 0, and where it does not.

### 3.4 DISCUSSION

IPI profiles appear to have some potential for detecting units, although there are some issues with the current approach. A notable result is that, based on the IPI profiles obtained here, the ability to recognize units was not uniform across units. Furthermore, the accuracy of unit detection appeared to be impacted by how many units were present, and whether or not unknown animals were present. A discussion of these observations follows.

### **3.4.1 Unit Recognition**

#### *3.4.1.1 Problems with Unit A*

Unit IPI profiles constructed from kernel density estimates of observed IPI distributions were more or less capable of recognizing units from new IPI measures, with the exception of unit *A*. One explanation for the poor performance with unit *A* may be differences between recording groups. In fact, each of the four recording groups used to test recognition of unit *A* had a different composition of whales, based on photographs taken within  $\pm 2$  hours of each recording. In other words, different members of unit *A* were present for each recording group. Because of the cross-validation procedure used to ensure the separation of data for profile construction and testing, incomplete profiles of *A* may have been unable to recognize the composition of individuals in the held-out group. Some evidence of this is apparent in Appendix D.

It is perhaps not surprising that unit *A*, of all units, seems to have the greatest issues with regards to consistency of composition between recording groups. Unit *A* is a large unit with particularly heterogeneous patterns of associations among its members. Recently, it was discovered that this unit is actually composed of two separate matriline, with members of the same matriline preferring to form clusters with one another (Konrad, 2017).

In addition to the problem of differing member composition between recording groups, it is likely that not all individuals present had their IPIs represented. It must be noted that the number of peaks in the IPI distributions from each recording group containing unit *A* did not always match the number of individuals that were deemed to be present based on photo-ID. As discussed in Chapter 2, the routine used to extract IPIs from a recording has a tendency to reject many IPIs. Because this routine depends on clicks with a clear multi-pulsed structure, which is only apparent when clicks are recorded on-axis (Zimmer *et al.*, 2005a), this high rejection rate is in large part related to the whales' orientation with respect to the hydrophone. Whales that never or infrequently had their longitudinal axis aligned with the hydrophone were unlikely to

have their IPIs captured. It should be noted that the recordings used to construct the unit A profile consisted mainly of ambient recordings taken when animals were foraging at depth; only one recording in which animals were diving (and thus facing away from the hydrophone) was available. Note also that unit A had the fewest IPIs available compared to the other units (144 versus 504 – 1511), so individuals without their IPIs represented may have been a particular problem for this unit.

#### *3.4.1.2 Problems with IPI Profiles in General*

The problems with unit A reveal that the reliability of an IPI profile is particularly sensitive to the data available. This might be especially true when the profile is constructed as a kernel density estimate. For a unit IPI profile to be reliable, all non-calf members of the unit must be represented. Should a unit member who's true IPI is distinct from the other members be missed during profile construction, then that profile will not fit well to subsequent IPI observations from that unit, if those observations include IPIs from the missed individual. One of the principles behind IPI profiles is that if unrecognized IPI values have been observed, then those IPIs could not have originated from the unit. Thus, if IPIs from a missed individual have been subsequently observed, then the likelihood of the profile may be close to zero, and the unit will unlikely be detected. A related but less problematic issue is if all unit members have been accounted for in the profile, but not all members were present or detected in subsequent observations. In this case, the unit still has a good chance of being recognized, but the relative likelihood of the profile would be lower. It should be noted that this is not an uncommon situation, as units are not always represented as a whole in one recording or a set of recordings. Unit A is an extreme example of this.

Unfortunately, using far-field recordings such as the ones in this study, it is very difficult to be certain of whose IPIs have been observed. As discussed previously, not all whales present during a recording may be accounted for because IPI measurement requires high-quality on-axis clicks. In addition to this, there is some uncertainty as to which whales are actually present. Here, presence of whales was determined based on

photographs taken within 2 hours prior to the start of a recording, and 2 hours after the end. However, not all whales photographed within this interval may have actually had an acoustic presence: some individuals could have been silent at the surface or away from the research vessel, and those photographed near the limits of the interval may have left the area, or not yet entered it during the time of recording. There is also the possibility that not all whales were sighted or photographed. If IPIs from an unseen whale are captured during construction of a unit IPI profile, then that profile may be contaminated, if the unseen individual is not a member of the unit. Contaminated profiles would have a reduced likelihood when fit to new IPI observations from the unit. Contaminated profiles would also increase the chance of the unit being falsely detected, because non-members with IPI signatures similar to the “impostor” IPIs will register as being potential members of the unit.

### **3.4.2 Algorithm Performance**

Overall, the unit detection algorithm successfully identified the correct combination of units about 30% of the time. When the correct combination was not selected, the amount of support for it was highly variable. However, performance was not uniform across all cases. In particular, units were more likely to be missed when they were grouped with other units and/or unknown whales, and there is evidence that this becomes worse as more units are present. Fortunately, such scenarios are relatively uncommon. Depending on the temporal scale of observation, units are most often sighted alone (Gero *et al.*, 2014; Gero *et al.*, 2015). This is also reflected by the fact that the “Multiples” scenario had the fewest occurrences within the dataset used here.

The slight increase in performance for the “Singles” scenario, and possibly “Multiples”, when unit A was treated as unknown, was to be expected. Since the routine had clear difficulties recognizing unit A, then treating unit A as unknown, which effectively removed it from analysis for the “Singles” and “Multiples” scenarios, should naturally have increased unit detection accuracy for those scenarios. However, removing unit A also caused a slight performance decrease for the “Unknowns”

scenario, and overall performance did not change greatly. This suggests that the routine has difficulty recognizing unknown whales as such. Since the probability density function of the *UNK* profile is spread out over a very wide range of IPIs, its likelihood when fit to IPI distributions with highly precise peaks is often much lower than for unit profiles. In other words, the routine has a tendency to associate observed IPIs with known units rather than considering them to be unknown. This is desirable to a certain extent, because unknown whales are not frequently encountered. However, the current approach may not be effective enough for recognizing true unknowns. A consequence of this is that treating poorly characterized units (such as unit A) as unknown is not a viable solution. In cases where the data-deficient unit is present, detection accuracy will still be compromised. Thus, it is important that as many units as possible be well characterized.

In addition to differences between scenarios, there is some evidence that the routine is more likely to correctly detect all units when more IPIs are available for inference. However, it is not clear from the current data if there is a limit at which performance can be deemed to be reliable, within some range of confidence. This needs to be investigated further, using recordings that last for longer durations and/or that are made more frequently. In particular, it would be interesting to test how well the algorithm performs with a fixed PAM setup, such as with bottom-mounted autonomous recorders, or moored hydrophones with a radio link (Mellinger *et al.*, 2007). Automatic unit detection would actually be most beneficial with fixed PAM, since researchers do not need to be present for this type of data collection. Since it is relatively (and perhaps more) easy to identify sperm whale units visually, mobile PAM surveys would be of limited use for this purpose, at least off Dominica.

### **3.4.3 Potential Improvements and Alternatives**

While the unit detection algorithm had some success, there is clearly a need for improvement. Some improvements could perhaps be achieved by making small modifications to the analytical process. For example, if constructing IPI profiles using

KDE, alternative bandwidth values could be investigated. A model selection criterion other than BIC might also be more appropriate. With KDE-based profiles, penalizing the number of units through BIC does have some advantages: since KDE may not represent all unit members equally, models can fit more easily to observations of that unit if they include other units with IPI peaks that overlap with those of the underrepresented whales (see for example Figure D.36 in Appendix D). Also, if unknown whales are present, then the model with greatest likelihood may be one that includes many units with peaks matching those of the unknown whales, rather than the *UNK* profile. However, for known units, if profiles are accurate and empirical IPI distributions are relatively free of noise, then models that include absent units should actually have a lower likelihood than the “true” model. If multiple units are interacting, but not all members’ IPIs have been detected, then BIC may incorrectly favour a simpler model.

One alternative analytical method worth trying is to use Bayesian inference. With this approach, unit presence models would each be assigned a prior probability of occurring, based on what is known from visual surveys, or from previous automatic detections. Units are known to differ in how often they visit the waters off Dominica, and as discussed previously, they are most often seen alone (Gero *et al.*, 2014). Furthermore, when units do associate with one another, they usually do so with preferred companions (Gero *et al.*, 2015). With a Bayesian approach, it would be possible to take all of these differential rates of occurrence into account. Since priors would be based on well-documented trends, this could prove to be more effective for model selection than penalizing likelihood through BIC, which only considers how many units are included in each model. Using priors also has the potential to improve the routine’s handling of unknown whales. As mentioned previously, one issue with the generic *UNK* profile is that it does not fit well to precise IPI peaks compared to other unit profiles. With priors, a possible alternative may be to use unique *UNK* profiles, which consist of models (GMMs or KDEs) that are fit directly to the observed IPIs. Naturally, such models would have a higher likelihood than all other candidates, but this

likelihood could be scaled down by a prior that reflects how often unknown individuals are encountered.

An important issue that Bayesian inference alone would not resolve, however, is the accuracy of IPI profiles. As shown by the problems with unit A, it is crucial that a unit IPI profile be well representative of all non-calf members of the unit. Ideally, profiles should be constructed as mixture models rather than kernel density estimates as used here. Unfortunately, this is challenging, because of the difficulties of associating IPIs with particular individuals. The most realistic approach for achieving this may be to use only dive recordings, in which the hydrophone was immediately placed in the slicks of diving whales. Because sperm whales expose their flukes when diving, individuals can be photo-identified just before the start of these recording. Their clicks are also more likely to be recorded on-axis, increasing the chance for useful IPI measures. The use of multiple hydrophones (i.e. towed arrays) would also be very useful here, as this would facilitate the tracking of individuals underwater, and thus the assignment of click trains to individuals.

For the purpose of profile construction, recordings are probably best analyzed manually, unless many of them are available. The automatic IPI compilation routine from Chapter 2 was not designed specifically for this purpose, and using it as such can be problematic. It is problematic primarily for two reasons. The first reason is that the routine's high click rejection rate can cause the IPI signals of the targeted whales to be missed or poorly characterized. While this is less likely to happen with diving whales, it still can and does happen. The second problem is that it is possible for the IPI signals of other whales foraging in the background to be picked up. Processing dive recordings automatically to construct IPI profiles as mixture models was the original intention for this analysis; however, this method did not produce sufficiently clear IPI distributions for each unit. Another factor that complicates the use of IPI profiles is that these profiles are not static. Due to growth of individual whales and potential changes in unit composition (i.e. from births, deaths, or emigration in the case of males), IPI profiles would need to be updated regularly, perhaps on an annual basis. When considering

these facts, it seems that the proposed approach for recognizing units remotely may be difficult to use in practice.

A radically different solution might be to use coda clicks, either as a replacement to or in tandem with IPIs. Coda clicks are stereotyped patterns of clicks used mainly by female sperm whales for communication (Watkins and Schevill, 1977; Weilgart and Whitehead, 1993; Marcoux *et al.*, 2006). It has been shown that subtle differences in the inter-click intervals of certain coda types can convey information on identity at several levels, including unit and individual identity (Antunes *et al.*, 2011; Gero *et al.*, 2016b; Oliveira *et al.*, 2016). Thus, automatic detection of units, and perhaps even individuals, may be possible by training a classifier to recognize the differences within codas. Estimates of identity based on both IPI profiles and coda signatures could be combined to obtain a single accurate estimate. Alternatively, measures of IPIs from codas could be used to adjust the assessment of unit identity from coda signatures. The incorporation of information from both IPIs and codas has the potential to provide highly accurate automatic estimates of sperm whale identity.

### **3.5 CONCLUSIONS**

A routine for automatically detecting Eastern Caribbean sperm whale units through PAM based on inter-pulse interval distributions was proposed and tested under a wide variety of scenarios. While the method showed potential, detection accuracy was rather poor, for both the most complex situations (mixtures of units and/or unknown individuals) and the simplest (only one unit present). However, it would be interesting to investigate how well the method works with a fixed PAM implementation, in which more recording time is available. To recognize an individual unit based on its IPI profile, the profile model must accurately represent all members of the unit. Unfortunately, due to the difficulties of associating IPI measures with individual whales, IPI profile construction is a difficult task. IPI profiles also need to be updated regularly to account for whale growth and changes in unit composition. Because of these issues, the proposed method is not practical or reliable enough to detect units automatically as it

stands. However, the method could be improved, for example by using Bayesian inference, and by obtaining better IPI profiles. The latter is particularly important and could be achieved by recording whales as they dive, but this would have to be done rigorously, such that IPIs can be obtained for all unit members with high confidence. A particularly promising extension may be to incorporate unit identity information from codas. Depending on how well an automatic classifier could recognize differences in the inter-click interval patterns of coda clicks, this has the potential to identify units or individuals with great accuracy, especially if combined with IPIs.

Automatic passive acoustic detection of sperm whale units would be highly desirable, especially for the Eastern Caribbean. Much of what is known about the whales in this location has been determined almost exclusively from visual surveys conducted off the west coast of Dominica during the months of February through May. While the west coast of Dominica is clearly an important location for these whales, their range extends well beyond this. However, very little information is available on how often they frequent the neighbouring islands (Gero *et al.*, 2007). Automatic detection of units through PAM would facilitate data collection in less studied areas, since the logistical challenges of deploying autonomous recorders in many places are less restrictive than for conducting mobile surveys. Autonomous recorders would also enable year-round monitoring, during periods when it would be undesirable to have observers physically present (e.g. hurricane season).

The sperm whales of the Eastern Caribbean form a small population that is relatively isolated from the rest of the Atlantic (Gero *et al.*, 2007). In recent years, this population has shown signs of poor health and appears to be declining (Whitehead and Gero, 2015; Gero and Whitehead, 2016). To ensure their survival, it is particularly important that the movements and behaviour of social units be closely monitored throughout their inter-national range (Gero *et al.*, 2007; Gero and Whitehead, 2016). Monitoring techniques should also be as non-invasive as possible, to reduce placing greater risks on an already vulnerable population. If it can be perfected, the automatic detection of units through passive acoustic monitoring could be the solution.

## **CHAPTER 4 - CONCLUSION**

### **4.1 SUMMARY OF PROJECT OUTCOME**

In this thesis, I have developed an algorithm capable of automatically compiling stable inter-pulse interval estimates from the clicks of multiple sperm whales present in an acoustic recording (Chapter 2), thereby enabling automatic acoustically-derived estimates of body length (Gordon, 1991; Growcott *et al.*, 2011). Secondly, I have used this algorithm in attempt to identify sperm whale social units automatically from their distinctive IPI distributions (Chapter 3). While the former objective was largely successful, the second showed promise, but was not accurate enough for practical use.

#### **4.1.1 Automatic IPI Compilation**

Since IPIs are only clear when clicks are recorded on-axis (Zimmer *et al.*, 2005a), their automatic extraction is difficult. While this can be overcome by averaging every click in a sequence (Teloni *et al.*, 2007; Antunes *et al.*, 2010), this technique assumes that all clicks originated from the same whale. Thus, for whales that travel, dive, and forage together as a cohesive social unit, IPI estimation by click averaging is often not feasible. The method I have developed solves this problem by taking a machine learning approach, in which clicks that display a clear multi-pulse structure (“Good” clicks) are separated from those that do not (“Bad” clicks) by a support vector machine. By training an SVM to classify clicks from surface recordings of female and juvenile sperm whales, I showed that it is possible to recognize “Good” on-axis sperm whale clicks automatically, with very high accuracy (94.3% based on cross-validation, with the imbalance between “Good” and “Bad” classes taken into account). Furthermore, by computing IPIs from only “Good” clicks and subsequently validating them for precision and repetition, I have shown that it is possible to obtain distributions of stable IPIs from recordings where multiple whales are foraging simultaneously. These IPIs can then be clustered fairly

accurately using Gaussian mixture models, providing automatic estimates of the number of whales present and the value of their true IPIs.

Based on comparison with manually compiled IPIs, my routine outputs reliable results, in that both methods produce distributions that contain very similar peaks. However, the automatic method does have a disadvantage. Based on approximately 4-minute long recordings, automatic IPI compilation rejects an overwhelming number of clicks: more than 99% on average, even with relatively passive filtration parameter values. Since on-axis clicks are typically rare and a good signal-to-noise ratio is needed for a clear multi-pulse structure, a high rejection rate is partly to be expected; however, the automatic routine does reject more clicks than expected, compared to manual compilation. Nevertheless, since sperm whales produce many clicks in a short amount of time, it is not particularly difficult to obtain reliable IPI distributions. Based on a click rate of 1.2 clicks/second (Whitehead and Weilgart, 1990), a rejection rate of 99.3%, and a minimum threshold of 5 IPIs per whale, 10 minutes were recommended as a minimum recording time.

A second potential concern with my automatic IPI compilation routine is that the SVM was trained using only clicks from female and juvenile sperm whales recorded near the sea surface off Dominica. To investigate potential overfitting, I applied the method to recordings of sperm whale units from the Galápagos. This produced IPI distributions that were generally similar to the ones from Dominica, except that there were often more peaks, and a greater proportion of clicks were rejected. I concluded that these results are consistent with what would be expected, given that sperm whale groups encountered in the Galápagos are typically much larger than those in the Atlantic (Whitehead *et al.*, 2012). Thus, the method appears to be useful for populations beyond Dominica; however, it is not known how well it performs for mature males or other types of recordings, notably recordings from bottom-mounted sensors.

### **4.1.2 Automatic Unit Identification**

Since the collective body length distribution of whales in a group is likely to differ from group to group, each social unit is expected to have a distinct IPI profile. Based on this premise, one can envision a system in which units may be identified from IPI values extracted from one or more audio recordings. In Chapter 3, I implemented and tested such a system, using the automatic IPI compilation routine from Chapter 2. By compiling IPIs from recordings where only one unit was known to be present, I established IPI profiles for 5 of the most common units encountered off Dominica, based on Gaussian kernel density estimation. I examined the potential of IPI profiles to recognize units through a cross-validation analysis, in which partial profiles were fitted to subsets of the IPI distributions used to construct the full profiles. This showed that IPI profiles may be useful, but only if all unit members are taken into account. If a unit member with a distinct IPI is missing from its unit's profile, then the likelihood of that profile may be extremely small when fit to an IPI distribution that includes the missing individual.

While IPI profiles showed promise for recognizing units individually, the routine as a whole did not perform well. When tested under several different scenarios (one unit present, multiple units present, or unknown whales present), the routine correctly identified all units in only about 30% of cases. Detection accuracy was best when only one unit was present; mixtures of units and the presence of unknown individuals were particularly problematic, with performance tending to decrease as more units were present together. For cases where only one unit was present, there was an indication that the routine performs better as more IPIs are available. Thus, it would be interesting to test this method on longer recordings, particularly from a fixed PAM implementation. However, based on the current assessment, this routine is not reliable enough for most practical uses.

## 4.2 APPLICATIONS AND FUTURE DIRECTIONS

The automatic processing of IPIs in sperm whale clicks has the potential to greatly improve passive acoustic monitoring for this species. With IPIs, it is possible to estimate body length, which is important for many ecological studies. Measures of body length in cetaceans have been used to estimate population parameters (Waters and Whitehead, 1990), to examine geographic variation in morphology of a species (Perryman and Lynn, 1993), to quantify growth rates (Kasuya, 1991; Miller *et al.*, 2013), and to examine segregation or behavioural/ecological differences between size classes (Cubbage and Calambokidis, 1987; Drouot *et al.*, 2004). The ability to do this acoustically for sperm whales using IPIs is a considerable advantage, due to the numerous challenges of photogrammetric length estimation with large whales. My work extends this even further by enabling the process to be automated. With automatic IPI processing, it may also be possible to estimate minimum group sizes remotely.

While the automatic IPI compilation method developed in this thesis has been shown to work, there is undoubtedly room for improvement. In particular, the high click rejection rate can be limiting when long recordings are not available, and it is not clear how well the routine would work with types of recordings other than from the surface-towed hydrophones that I used. Ideally, the best solution to both problems is likely to redevelop the click classifier. Most importantly, a more diverse set of clicks should be used in training, which includes for example clicks from males, and clicks from bottom-mounted recorders. Secondly, it might be possible to reduce the number of false negatives by 1) using fuzzy labelling (Schwenker and Trentin, 2014), in which training clicks are weighted based on the clarity of their multi-pulsed structure, and/or 2) by finding new features that are better able to differentiate between on-axis and off-axis clicks. A larger training dataset would also be welcome, particularly for on-axis clicks. Another approach may be to use multi-class or hierarchical classification, in which on-axis coda clicks can be included as a positive class. This would have the advantage that IPIs from both “usual” (echolocation) and coda (communication) clicks can be analyzed separately.

If the automatic IPI compilation routine can be improved, this would be beneficial to unit detection as well. Automatic unit detection would benefit most from a fixed PAM implementation, in which autonomous recorders are placed in several locations (Mellinger *et al.*, 2007). This would enable tracking of units, and would also provide more IPIs to work with. For this to work, however, the IPI compilation routine must be capable of recognizing clicks from bottom-mounted recorders. Additionally, since female sperm whales frequently produce codas, the ability to recognize on-axis codas would further assist in unit detection.

Automatic unit detection would be highly desirable, especially for the Eastern Caribbean population. If made to work, this technique could enable continuous non-invasive tracking of units, in areas or during periods where they cannot be followed by a research team. In turn, this information could be used to make more effective management decisions, which are very much needed for this population (Gero *et al.*, 2007; Gero *et al.*, 2014; Gero and Whitehead, 2016). However, the unit detection routine developed in this thesis is not reliable enough for this kind of use and would need to be improved. One of the most important issues may be that the current IPI profiles do not accurately represent the IPIs of all members in their units. Because of the few IPIs available and the difficulty in assigning them to individuals, I resorted to kernel density estimation to construct profiles. However, to develop this method further, a Gaussian mixture modelling approach is strongly preferred. To achieve this, it will be necessary to obtain more IPI estimates from all whales in each unit. Furthermore, the process by which units are detected from IPI profiles would likely benefit greatly from a Bayesian approach, in which all the potential combinations of units are weighted by priors. It is well known that all possible cases do not occur at the same rate, as some units are spotted more frequently than others, and there are strong preferred associations between units (see Gero *et al.*, 2014; Gero *et al.*, 2015). Ultimately however, the incorporation of unit-distinctive and individually-distinctive coda signatures (Gero *et al.*, 2016b; Oliveira *et al.*, 2016) may have the greatest potential for robust automatic unit detection.

## REFERENCES

- Adler-Fenchel, H. S. (1980). "Acoustically derived estimate of the size distribution for a sample of sperm whales (*Physeter catodon*) in the Western North Atlantic," *Can. J. Fish. Aquat. Sci.* **37**, 2358–2361.
- Antunes, R., Rendell, L., and Gordon, J. (2010). "Measuring inter-pulse intervals in sperm whale clicks: Consistency of automatic estimation methods," *J. Acoust. Soc. Am.* **127**, 3239–3247.
- Antunes, R., Schulz, T., Gero, S., Whitehead, H., Gordon, J., and Rendell, L. (2011). "Individually distinctive acoustic features in sperm whale codas," *Anim. Behav.* **81**, 723–730.
- Arnbom, T. (1987). "Individual identification of sperm whales," *Rep. Int. Whal. Commn.* **37**, 201–204.
- Arnbom, T., Papastavrou, V., Weilgart, L. S., and Whitehead, H. (1987). "Sperm whales react to an attack by killer whales," *J. Mamm.* **68**, 450–453.
- Arnbom, T., and Whitehead, H. (1989). "Observations on the composition and behaviour of groups of female sperm whales near the Galapagos Islands," *Can. J. Zool.* **67**, 1–7.
- Au, W. W. L. (1993). *The Sonar of Dolphins* (Springer, New York).
- Backus, R. H., and Schevill, W. E. (1966). "Physeter clicks," in *Whales, Dolphins, and Porpoises*, edited by K. S. Norris (University of California Press, Berkeley), pp. 510–527.
- Barlow, J., and Taylor, B. L. (2005). "Estimates of sperm whale abundance in the northeastern temperate Pacific from a combined acoustic and visual survey," *Mar. Mammal Sci.* **21**, 429–445.
- Best, P. B. (1979). "Social organization in sperm whales, *Physeter macrocephalus*," in *Behaviour of marine animals*, edited by H. E. Winn, and B. L. Olla (Plenum Press, New York), Vol. **3**, pp. 227–290.
- Best, P. B., Canham, P. A. S., and MacLeod, N. (1984). "Patterns of reproduction in sperm whales, *Physeter macrocephalus*," *Rep. Int. Whal. Commn.* (Special Issue) **6**, 51–79.
- Buckland, S. T., Anderson, D. R., Burnham, K. P., Laake, J. L., Borchers, D. L., and Thomas, L. (2001). *Introduction to Distance Sampling: Estimating Abundance of Biological Populations* (Oxford University, Oxford).

- Christal, J., Whitehead, H., and Lettevall, E. (1998). "Sperm whale social units: variation and change," *Can. J. Zool.* **76**, 1431–1440.
- Clarke, M. R. (1978). "Structure and proportions of the spermaceti organ in the sperm whale," *J. Mar. Biol. Ass. U.K.* **58**, 1–17.
- Cubbage, J. C., and Calambokidis, J. (1987). "Size-class segregation of bowhead whales discerned through aerial stereophotogrammetry," *Mar. Mammal Sci.* **3**, 179–185.
- Drouot, V., Gannier, A., and Goold, J. C. (2004). "Diving and feeding behavior of sperm whales (*Physeter macrocephalus*) in the northwestern Mediterranean Sea," *Aquatic Mammals* **30**, 419–426.
- Dufault, S., and Whitehead, H. (1995). "The geographic stock structure of female and immature sperm whales in the South Pacific," *Rep. Int. Whal. Commn.* **45**, 401–405.
- Gedamke, J., and Robinson, S. M. (2010). "Acoustic survey for marine mammal occurrence and distribution off East Antarctica (30-80°E) in January-February 2006," *Deep Sea Res. Part II* **57**, 968–981.
- Gero, S., Bøttcher, A., Whitehead, H., and Madsen, P. T. (2016a). "Socially segregated, sympatric sperm whale clans in the Atlantic Ocean," *R. Soc. Open Sci.* **3**: 160061.
- Gero, S., Engelhaupt, D., Rendell, L., and Whitehead, H. (2009). "Who cares? Between-group variation in alloparental caregiving in sperm whales," *Behav. Ecol.* **29**, 838–843.
- Gero, S., Engelhaupt, D., and Whitehead, H. (2008). "Heterogeneous social associations within a sperm whale, *Physeter macrocephalus*, unit reflect pairwise relatedness," *Behav. Ecol. Sociobiol.* **63**, 143–51.
- Gero, S., Gordon, J., Carlson, C., Evans, P., and Whitehead, H. (2007). "Population estimate and inter-island movement of sperm whales, *Physeter macrocephalus*, in the Eastern Caribbean," *J. Cetacean Res. Manage.* **9**, 143–150.
- Gero, S., Gordon, J., and Whitehead, H. (2013). "Calves as social hubs: dynamics of the social network within sperm whale units," *Proc. R. Soc. B* **280**, 20131113.
- Gero, S., Gordon, J., and Whitehead, H. (2015). "Individualized social preferences and long-term social fidelity between social units of sperm whales," *Anim. Behav.* **102**, 15–23.

- Gero, S., Milligan, M., Rinaldi, C., Francis, P., Gordon, J., Carlson, C., Steffen, A., Tyack, P., Evans, P., and Whitehead, H. (2014). "Behavior and social structure of the sperm whales of Dominica, West Indies," *Mar. Mammal Sci.* **30**: 905–922.
- Gero, S., and Whitehead, H. (2016). "Critical decline of the Eastern Caribbean sperm whale population," *PLoS ONE* **11**, e0162019.
- Gero, S., Whitehead, H., and Rendell, L. (2016b). "Individual, unit and vocal clan level identity cues in sperm whale codas," *R. Soc. Open Sci.* **3**, 150372.
- Gillespie, D., Mellinger, D. K., Gordon, J., McLaren, D., Redmond, P., McHugh, R., Trinder, P. W., Deng, X. Y., and Thode, A. (2008). "PAMGUARD: semiautomated, open-source software for real-time acoustic detection and localisation of cetaceans," in *Proceedings of the Institute of Acoustics*, Vol. 30, Pt. 5.
- Goold, J. C. (1996). "Signal processing techniques for acoustic measurement of sperm whale body lengths," *J. Acoust. Soc. Am.* **100**, 3431–3441.
- Goold, J. C., and Jones, S. E. (1995). "Time and frequency domain characteristics of sperm whale clicks," *J. Acoust. Soc. Am.* **98**, 1279–1291.
- Gordon, J. C. D. (1987). "The behaviour and ecology of sperm whales off Sri Lanka," Ph.D. thesis, Darwin College, Cambridge.
- Gordon, J. C. D. (1991). "Evaluation of a method for determining the length of sperm whales (*Physeter catodon*) from their vocalizations," *J. Zool. Lond.* **224**, 301–314.
- Gordon, J. C. D., Matthews, J. N., Panigada, S., Gannier, A., Borsani, J. F., and Notarbartolo di Sciara, G. (2000). "Distribution and relative abundance of striped dolphins, and distribution of sperm whales in the Ligurian Sea cetacean sanctuary: results from a collaboration using acoustic monitoring techniques," *J. Cetacean Res. Manage.* **2**, 27–36.
- Growcott, A., Miller, B., Sirguyev, P., Slooten, E., and Dawson, S. (2011). "Measuring body length of male sperm whales from their clicks: The relationship between inter-pulse intervals and photogrammetrically measured lengths," *J. Acoust. Soc. Am.* **130**, 568–573.
- Hammond, P. S., Mizroch, S. A., and Donovan, G. P. (eds.) (1990). *Individual Recognition of Cetaceans: Use of Photo-Identification and Other Techniques to Estimate Population Parameters*. Rep. Int. Whal. Commn. (Special Issue) **12**.
- Hannay, D. E., Delarue, J., Mouy, X., Martin, B. S., Leary, D., Oswald, J. N., and Vallarta, J. (2013). "Marine mammal acoustic detections in the northeastern Chukchi Sea, September 2007–July 2011," *Cont. Shelf Res.* **67**, 127–146.

- Jaquet, N., Dawson, S., and Douglas, L. (2001). "Vocal behaviour of male sperm whales: Why do they click?," *J. Acoust. Soc. Am.* **109**, 2254–2259.
- Kasuya, T. (1991). "Density dependent growth in North Pacific sperm whales," *Mar. Mammal Sci.* **7**, 230–257.
- Konrad, C. M. (2017). "Kinship in sperm whale society: Effects on association, alloparental care and vocalizations," M.Sc. thesis, Dalhousie University, Halifax, Nova Scotia, Canada.
- Lebreton, J., Burnham, K. P., Colbert, J., and Anderson, D. R. (1992). "Modeling survival and testing biological hypotheses using marked animals: A unified approach with case studies," *Ecol. Monogr.* **62**, 67–118.
- Lyrholm, T., and Gyllensten, U. (1998). "Global matrilineal population structure in sperm whales as indicated by mitochondrial DNA sequences," *Proc. R. Soc. Lond. B* **265**, 1679–1684.
- Madsen, P. T., Carder, D. A., Au, W. W. L., Nachtigall, P. E., Møhl, B., and Ridgway, S. H. (2003). "Sound production in neonate sperm whales (L)," *J. Acoust. Soc. Am.* **113**, 2988–2991.
- Madsen, P. T., Payne, R., Kristiansen, N. U., Wahlberg, M., Kerr, I., and Møhl, B. (2002a). "Sperm whale sound production studied with ultrasound time/depth-recording tags," *J. Exp. Biol.* **205**, 1899–1906.
- Madsen, P. T., Wahlberg, M., and Møhl, B. (2002b). "Male sperm whale (*Physeter macrocephalus*) acoustics in a high latitude habitat: Implications for echolocation and communication," *Behav. Ecol. Sociobiol.* **53**, 31–41.
- Marcoux, M., Whitehead, H., and Rendell, L. (2006). "Coda vocalizations recorded in breeding areas are almost entirely produced by mature female sperm whales (*Physeter macrocephalus*)," *Can. J. Zool.* **84**, 609–614.
- McNicholas, P. D. (2016). "Model-Based Clustering," *J. Classif.* **33**: 331–373.
- Mellinger, D. K., Stafford, K. M., Moore, S. E., Dziak, R. P., and Matsumoto, H. (2007). "An overview of fixed passive acoustic observation methods for cetaceans," *Oceanography* **20**, 36–45.
- Mesnick, S. L. (2001). "Genetic relatedness in sperm whales: Evidence and cultural implications," *Behav. Brain. Sci.* **24**, 346–347.
- Miller, B. S. (2010). "Acoustically derived growth rates and three-dimensional localisation of sperm whales (*Physeter macrocephalus*) in Kaikoura, New Zealand," Ph.D. thesis, University of Otago, New Zealand.

- Miller, B. S., Growcott, A., Sooten, E., and Dawson, S. M. (2013). "Acoustically derived growth rates of sperm whales (*Physeter macrocephalus*) in Kaikoura, New Zealand," *J. Acoust. Soc. Am.* **134**, 2438–2445.
- Miller, P. J. O., Johnson, M. P., and Tyack, P. L. (2004). "Sperm whale behaviour indicates the use of echolocation click buzzes 'creaks' in prey capture," *Proc. R. Soc. Lond. B* **271**, 2239–2247.
- Møhl, B. (2001). "Sound transmission in the nose of the sperm whale *Physeter catodon*. A post mortem study," *J. Comp. Physiol. A* **187**, 335–340.
- Møhl, B., Larsen, E., and Amundin, M. (1981). "Sperm whale size determination: Outlines of an acoustic approach," *FAO Fisheries Ser.* **5**, 327–332.
- Møhl, B., Wahlberg, M., Madsen, P. T., Heerfordt, A., and Lund, A. (2003). "The monopulsed nature of sperm whale clicks," *J. Acoust. Soc. Am.* **114**, 1143–1154.
- Møhl, B., Wahlberg, M., Madsen, P. T., Miller, L. A., and Surlykke, A. (2000). "Sperm whale clicks: Directionality and source level revisited," *J. Acoust. Soc. Am.* **107**, 638–648.
- Mullins, J., Whitehead, H., and Weilgart, L. S. (1988). "Behaviour and vocalizations of two single sperm whales, *Physeter macrocephalus*, off Nova Scotia," *Can. J. Fish. Aquat. Sci.* **45**, 1736–1743.
- Nishiwaki, N., Oshumi, S., and Maeda, Y. (1963). "Changes in form of the sperm whale accompanied with growth," *Sci. Rep. Wh. Res. Inst. Tokyo* **17**, 1–13.
- Norris, K. S., and Harvey, G. W. (1972). "A theory for the function of the spermaceti organ of the sperm whale (*Physeter catodon* L.)," in *Animal Orientation and Navigation*, edited by S. R. Galler, K. Schmidt-Koenig, G. J. Jacobs, and R. E. Belleville, SP-262 (NASA, Washington, DC), pp. 397–417.
- Norris, T. F., McDonald, M., and Barlow, J. (1999). "Acoustic detections of singing humpback whales (*Megaptera novaeangliae*) in the eastern North Pacific during their northbound migration," *J. Acoust. Soc. Am.* **108**, 506–514.
- Oliveira, C., Wahlberg, M., Silva, M. A., Johnson, M., Antunes, R., Wisniewska, D. M., Fais, A., Gonçalves, J., and Madsen, P. T. (2016). "Sperm whale codas may encode individuality as well as clan identity," *J. Acoust. Soc. Am.* **139**, 2860–2869.
- Ovaskainen, O., Rekola, H., Meyke, E., and Arjas, E. (2008). "Bayesian methods for analysing movements in heterogeneous landscapes from mark-recapture data," *Ecology* **89**, 542–554.

- Page, S. E. (1954). "Continuous inspection schemes," *Biometrika* **41**, 100–115.
- Papastavrou, V., Smith, S. C., and Whitehead, H. (1989). "Diving behavior of the sperm whale, *Physeter macrocephalus*, off the Galapagos Islands," *Can. J. Zool.* **67**, 839–846.
- Perryman, W. L., and Lynn, M. S. (1993). "Identification of geographic forms of common dolphin (*Delphinus delphis*) from aerial photogrammetry," *Mar. Mammal. Sci.* **9**, 119–137.
- Pitman, R. L., Ballance, L. T., Mesnick, S. I., and Chivers, S. J. (2001). "Killer whale predation on sperm whales: observations and implications," *Mar. Mammal Sci.* **17**, 494–507.
- Platt, J. (1999). "Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods," in *Advances in Large Margin Classifiers* (MIT Press), pp. 61–74.
- Rendell, L., and Whitehead, H. (2003). "Vocal clans in sperm whales (*Physeter macrocephalus*)," *Proc. R. Soc. Lond. B* **270**, 225–231.
- Rendell, L., and Whitehead, H. (2004). "Do sperm whales share coda vocalizations? Insights into coda usage from acoustic size measurements," *Anim. Behav.* **67**, 865–874.
- Rhineland, M. Q., and Dawson, S. M. (2004). "Measuring sperm whales from their clicks: Stability of interpulse intervals and validation that they indicate whale length," *J. Acoust. Soc. Am.* **115**, 1826–1831.
- Richard, K. R., Dillon, M. C., Whitehead, H., and Wright, J. M. (1996). "Patterns of kinship in groups of free-living sperm whales (*Physeter macrocephalus*) revealed by multiple molecular genetic analyses," *Proc. Natl. Acad. Sci. USA* **93**, 8792–8795.
- Schulz, T. M., Whitehead, H., Gero, S., and Rendell, L. (2008). "Overlapping and matching of codas in vocal interactions between sperm whales: insights into communication function," *Anim. Behav.* **76**, 1977–1988.
- Schulz, T. M., Whitehead, H., Gero, S., and Rendell, L. (2011). "Individual vocal production in a sperm whale (*Physeter macrocephalus*) social unit," *Mar. Mammal Sci.* **27**, 149–166.
- Schwarz, C. J., and Seber, G. A. (1999). "Estimating animal abundance: Review III," *Stat. Sci.* **14**, 427–456.
- Schwenker, F., and Trentin, E. (2014). "Pattern classification and clustering: A review of partially supervised learning approaches," *Pattern Recogn. Lett.* **37**, 4–14.

- Stafford, K. M., Mellinger, D. K., Moore, S. E., and Fox, C. G. (2007). "Seasonal variability and detection range modeling of baleen whale calls in the Gulf of Alaska, 1999–2002," *J. Acoust. Soc. Am.* **122**, 3378–3390.
- Teloni, V., Zimmer, W. M. X., Wahlberg, M., and Madsen, P. (2007). "Consistent acoustic size estimation of sperm whales using clicks recorded from unknown aspects," *J. Cetacean Res. Manage.* **9**, 127–136.
- The MathWorks, Inc. (2015). "Prominence," in *Signal Processing Toolbox: User's Guide (R2015a)* (The MathWorks, Inc., Natick, Massachusetts, USA), pp. 16-120–16-122.
- Thode, A., Mellinger, D. K., Stienessen, S., Martinez, A., and Mullin, K. (2002). "Depth-dependent acoustic features of diving sperm whales (*Physeter macrocephalus*) in the Gulf of Mexico," *J. Acoust. Soc. Am.* **112**, 308–321.
- Thomas, J. A., Fisher, S. R., and Awbrey, F. A. (1986a). "Use of acoustic techniques in studying whale behavior," *Rep. Int. Whal. Commn. (Special Issue)* **8**, 121–138.
- Thomas, J. A., Fisher, S. R., and Ferm, L. M. (1986b). "Acoustic detection of cetaceans using a towed array of hydrophones," *Rep. Int. Whal. Commn. (Special Issue)* **8**, pp. 139–148.
- Verfuß, U. K., Honnef, C. G., Meding, A., Dähne, M., Mundry, R., and Benke, H. (2007). "Geographical and seasonal variation of harbour porpoise (*Phocoena phocoena*) presence in the German Baltic Sea revealed by passive acoustic monitoring," *J. Mar. Biol. Ass. U.K.* **87**, 165–176.
- Waters, S., and Whitehead, H. (1990). "Population and growth parameters of Galápagos sperm whales estimated from length distributions," *Rep. Int. Whal. Commn.* **40**, 225–235.
- Watkins, W. A., Daher, M., Fristrup, K. M., and Howald, T. J. (1993). "Sperm whales tagged with transponders and tracked underwater by sonar," *Mar. Mammal Sci.* **9**, 55–67.
- Watkins, W. A., and Schevill, W. E. (1977). "Sperm whale codas," *J. Acoust. Soc. Am.* **62**, 1485-1490.
- Watwood, S. L., Miller P. J. O., Johnson, M., Madsen P. T., and Tyack, P. L. (2006). "Deep-diving foraging behavior of sperm whales (*Physeter macrocephalus*)," *J. Anim. Ecol.* **75**, 814–825.
- Weilgart, L. S., and Whitehead, H. (1988). "Distinctive vocalizations from mature male sperm whales (*Physeter microcephalus*)," *Can. J. Zool.* **66**, 1931–1937.

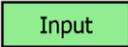
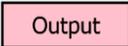
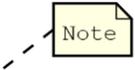
- Weilgart, L., and Whitehead, H. (1993). "Coda communication by sperm whales (*Physeter macrocephalus*) off the Galápagos Islands," *Can. J. Zool.* **71**, 744-752.
- Whitehead, H. (1996). "Babysitting, dive synchrony, and indications of alloparental care in sperm whales," *Behav. Ecol. Sociobiol.* **38**, 237-244.
- Whitehead, H. (2001). "Analysis of animal movement using opportunistic individual identifications: Application to sperm whales," *Ecology* **82**, 1417-1432.
- Whitehead, H. (2008). *Analyzing animal societies: quantitative methods for vertebrate social analysis* (Chicago University Press, Chicago)
- Whitehead, H., Antunes, R., Gero, S., Wong, S. N. P., Engelhaupt, D., and Rendell, L. (2012). "Multilevel Societies of Female Sperm Whales (*Physeter macrocephalus*) in the Atlantic and Pacific: Why Are They So Different?," *Int. J. Primatol.* **33**, 1142-1164.
- Whitehead, H., Coakes, A., Jaquet, N., and Lusseau, S. (2008). "Movements of sperm whales in the tropical Pacific," *Mar. Ecol. Prog. Ser.* **361**, 291-300.
- Whitehead, H., and Gero, S. (2015). "Conflicting rates of increase in the sperm whale population of the eastern Caribbean: positive observed rates do not reflect a healthy population," *Endang. Species Res.* **27**, 207-218.
- Whitehead, H., Waters, S., and Lyrholm, T. (1991). "Social organization of female sperm whales and their offspring: Constant companions and casual acquaintances," *Behav. Ecol. Sociobiol.* **29**, 385-389.
- Whitehead, H., and Weilgart, W. (1990). "Click rates from sperm whales," *J. Acoust. Soc. Am.* **87**, 1798-1806.
- Whitehead, H., and Weilgart, L. (1991). "Patterns of visually observable behaviour and vocalizations in groups of female sperm whales," *Behaviour* **118**, 275-296.
- Zimmer, W. M. X. (2011). *Passive Acoustic Monitoring of Cetaceans* (Cambridge University Press, Cambridge)
- Zimmer, W. M. X., Johnson, M. P., D'Amico, A. and Tyack, P. (2003). "Combining data from a multisensor tag and passive sonar to determine the diving behavior of a sperm whale (*Physeter macrocephalus*)," *IEEE J. Ocean. Engr.* **28**, 13-28.
- Zimmer, W. M. X., Madsen, P. T., Teloni, V., Johnson, M. P., and Tyack, P. L. (2005a). "Off-axis effects on the multipulse structure of sperm whale usual clicks with implications for sound production," *J. Acoust. Soc. Am.* **118**, 3337-3345.

Zimmer, W. M. X., Tyack, P. L., Johnson, M. P., and Madsent, P. T. (2005b). "Three-dimensional beam pattern of regular sperm whale clicks confirms bent-horn hypothesis," *J. Acoust. Soc. Am.* **117**, 1473–1485.

## **APPENDIX A - ACTIVITY DIAGRAMS FOR AUTOMATIC IPI COMPILATION ALGORITHM**

This section contains activity diagrams illustrating the flow of the complete inter-pulse interval (IPI) compilation algorithm. The diagrams are based roughly on the Unified Modelling Language (UML) 2.5 standard (Object Management Group Inc., 2015). Table A.1 describes the symbols used in each diagram.

**Table A.1** Summary of activity diagram symbols

| Symbol  | Description  |
|---|--|
|    | Initial Node. Marks the beginning of an activity.  |
|    | Activity Final Node. Marks the end of an activity.   |
|    | Action Node. Denotes a process in an activity.   |
|    | Call Activity Node. This is a special Action Node that indicates an action described in more detail by a separate activity diagram. "AX" is the corresponding figure number.   |
|    | Object Node. Used to model data flow between processes. To avoid clutter, not all variables are explicitly represented as Object Nodes (loop variables and global parameters are typically implicit).  |
|    | Input Parameter Node. Represents an input variable to an activity.   |
|   | Output Parameter Node. Represents an activity's output.  |
|  | Decision Node. Marks a point where the activity flow is dictated by a condition (e.g. "if" statements and loops). Conditions are denoted by statements in square brackets next to their corresponding outflows. They may also act as Merge Nodes. Objects associated with decision criteria can link directly to a Decision Node; these are indicated by inflows with the statement <<decisionInputFlow>>. |
|  | Merge Node. Indicates a point where multiple flows converge into one.  |
|  | Note Symbol. These are used as comments. They are often used to list important loop variables and their initial conditions.  |
|  | Connector. These merely indicate the continuation of a flow. They always come in pairs. Their main purpose is to avoid overlapping edges and to link distant nodes together.   |

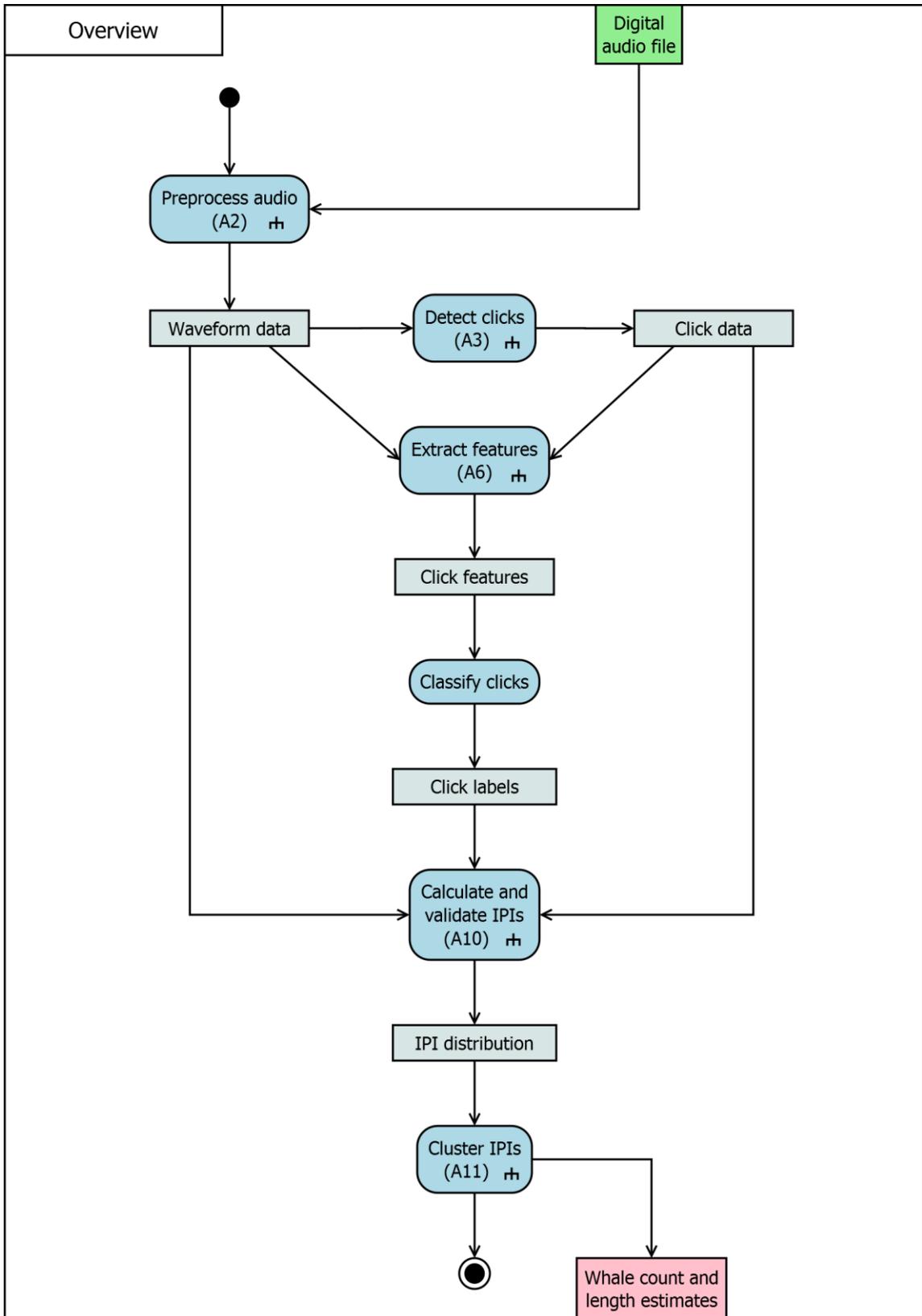


Figure A.1 Overview of complete IPI compilation algorithm

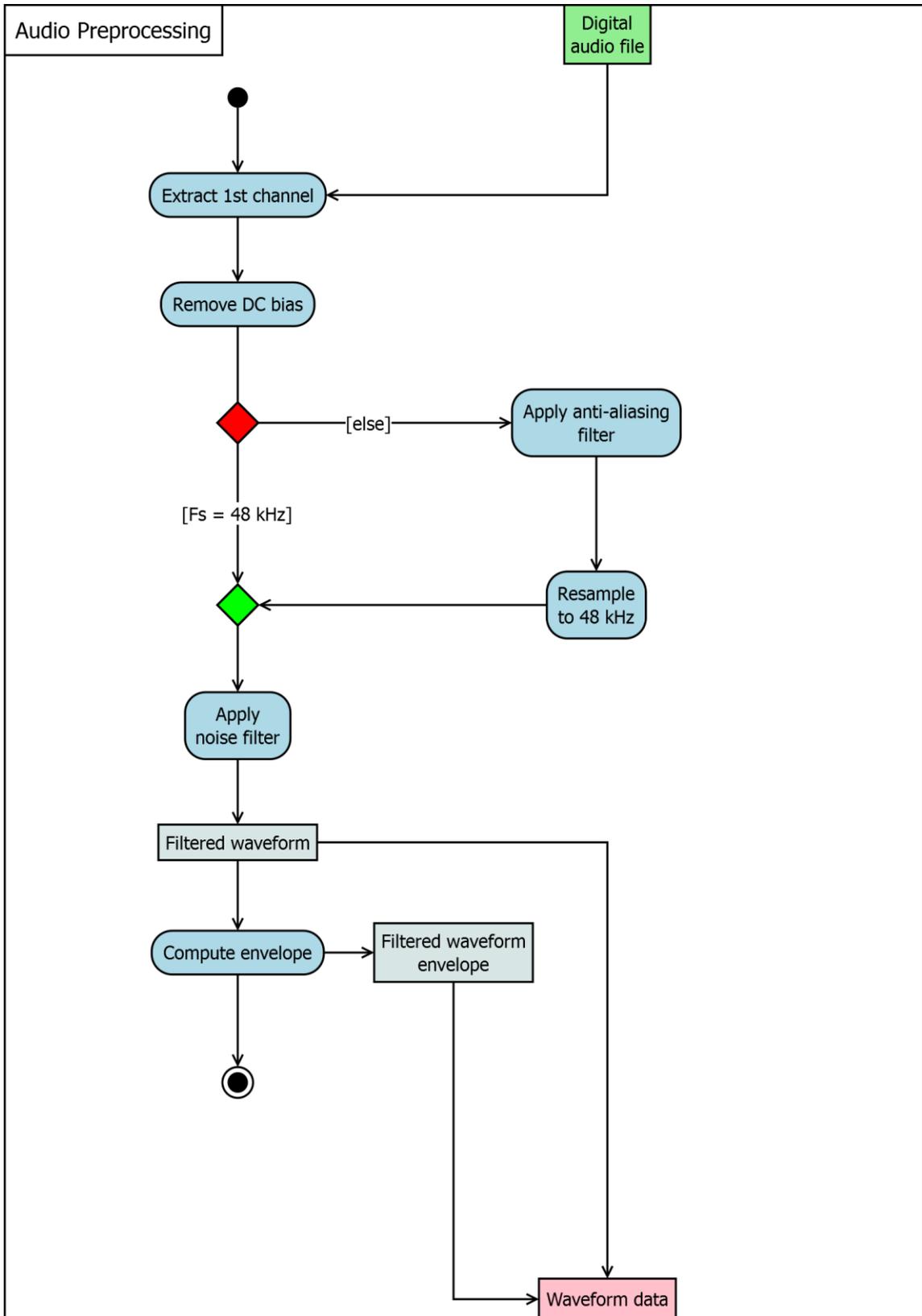


Figure A.2 Preprocessing of raw audio file

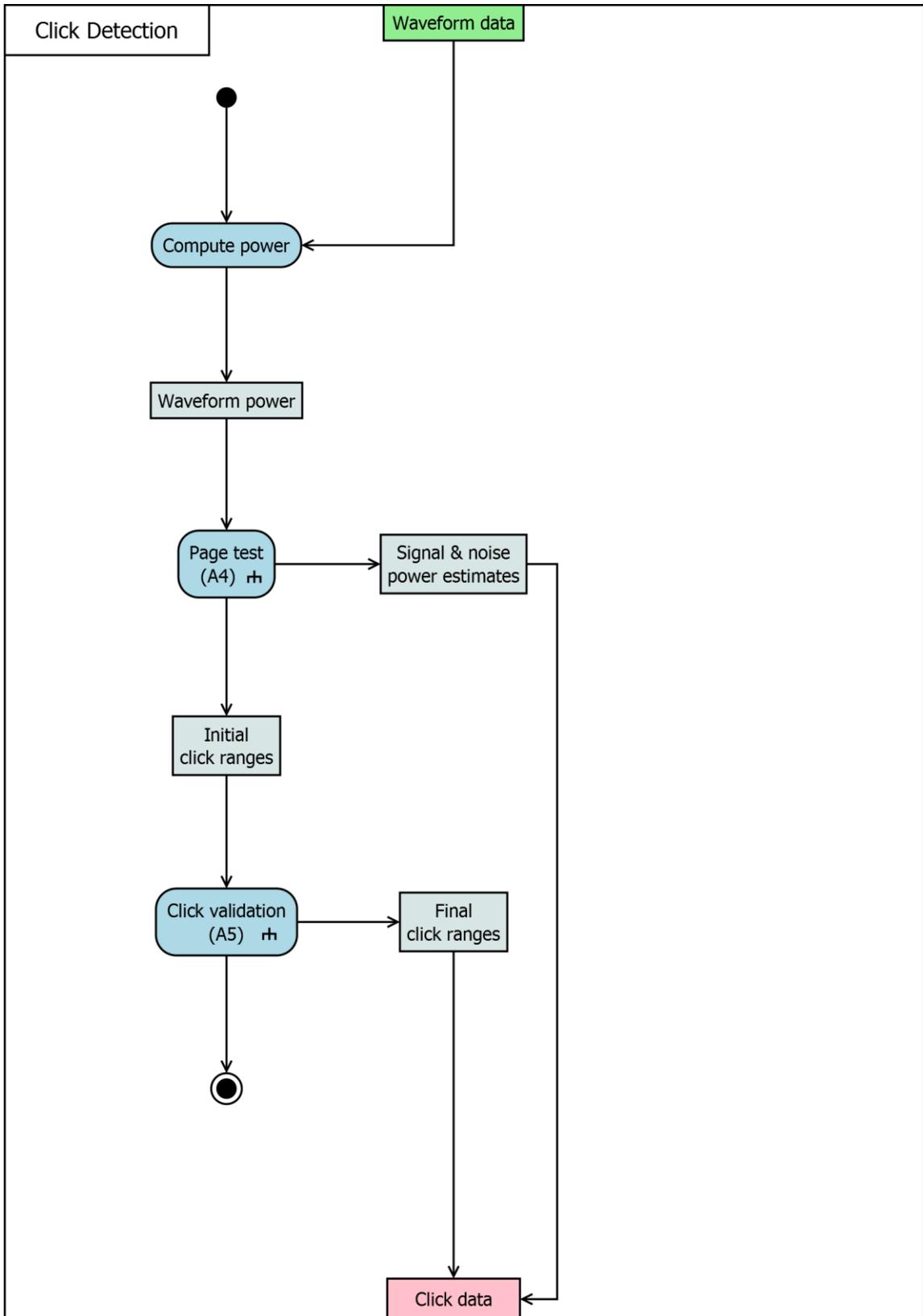


Figure A.3 Overview of click detection procedure

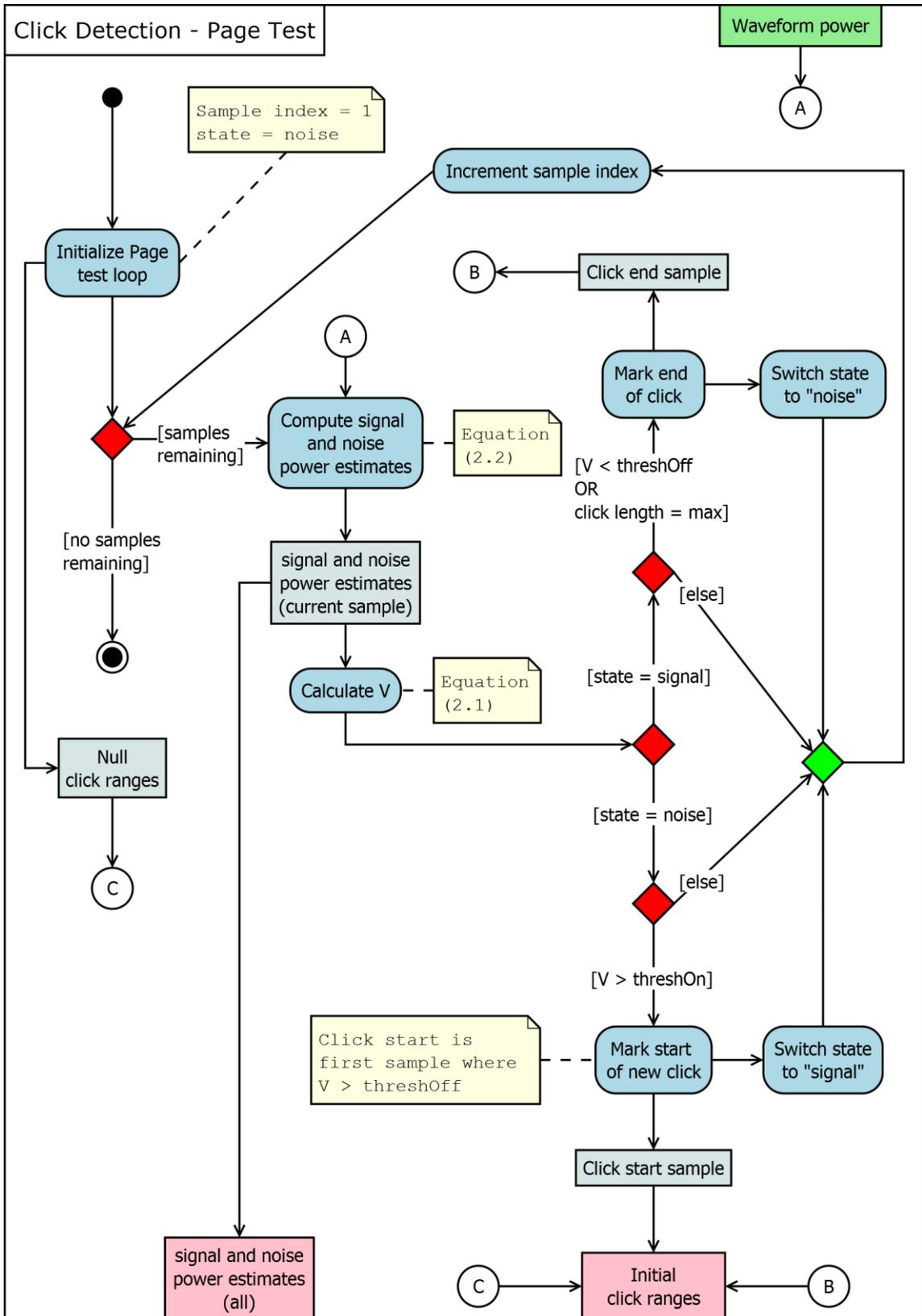
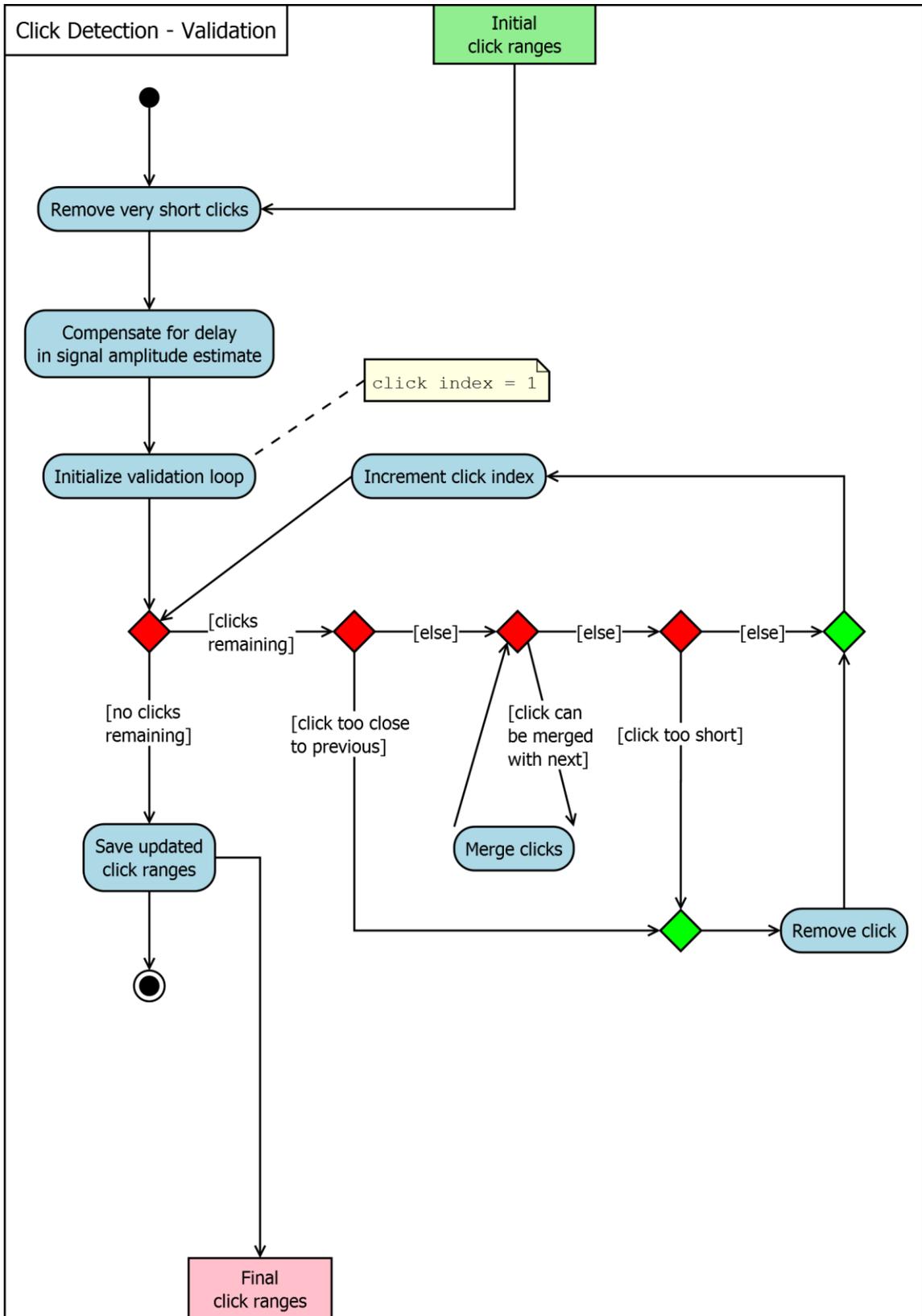


Figure A.4 Page test component of click detection



**Figure A.5** Validation component of click detection. Validates and refines output of Page test.

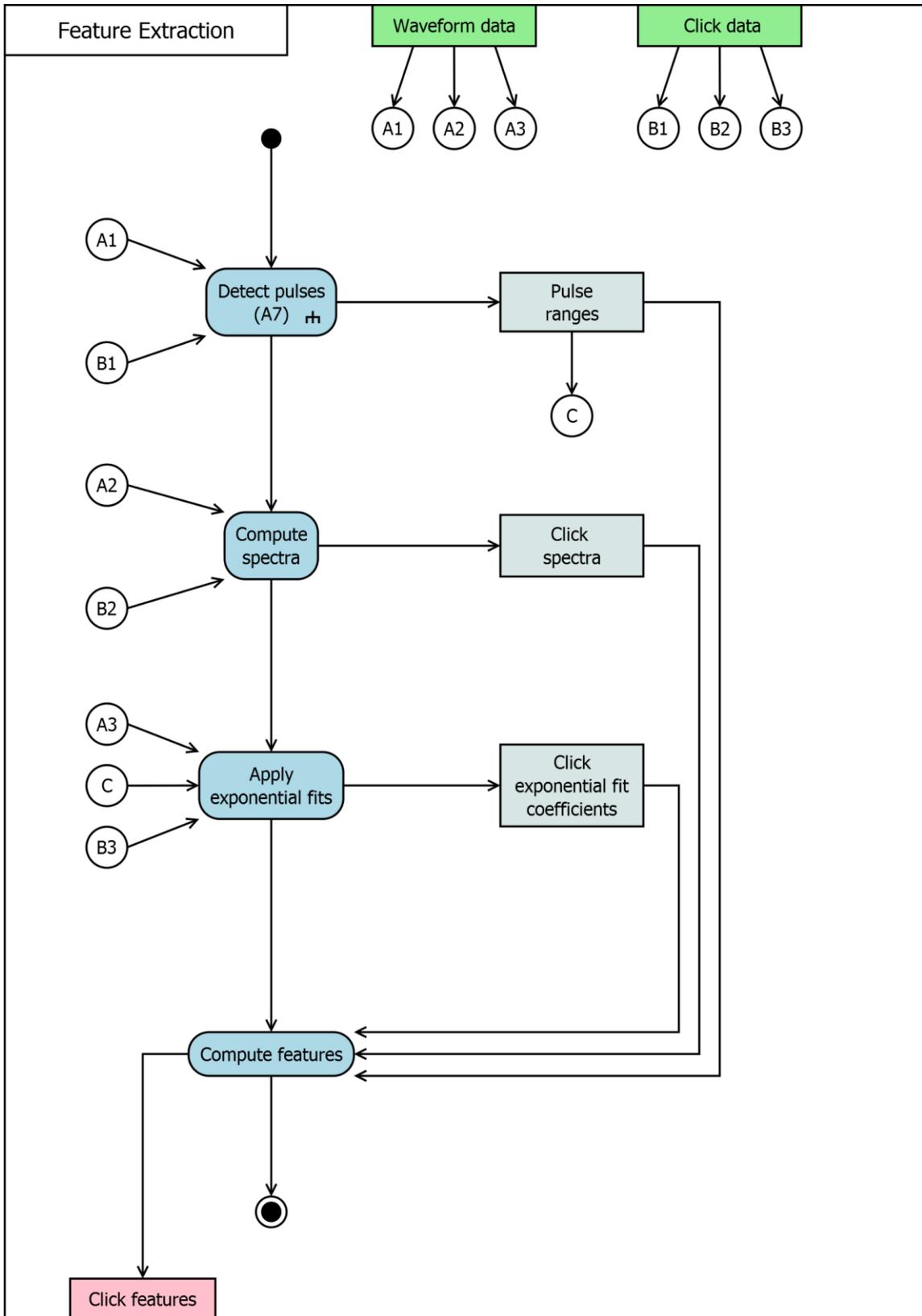


Figure A.6 Extraction of features for click classification

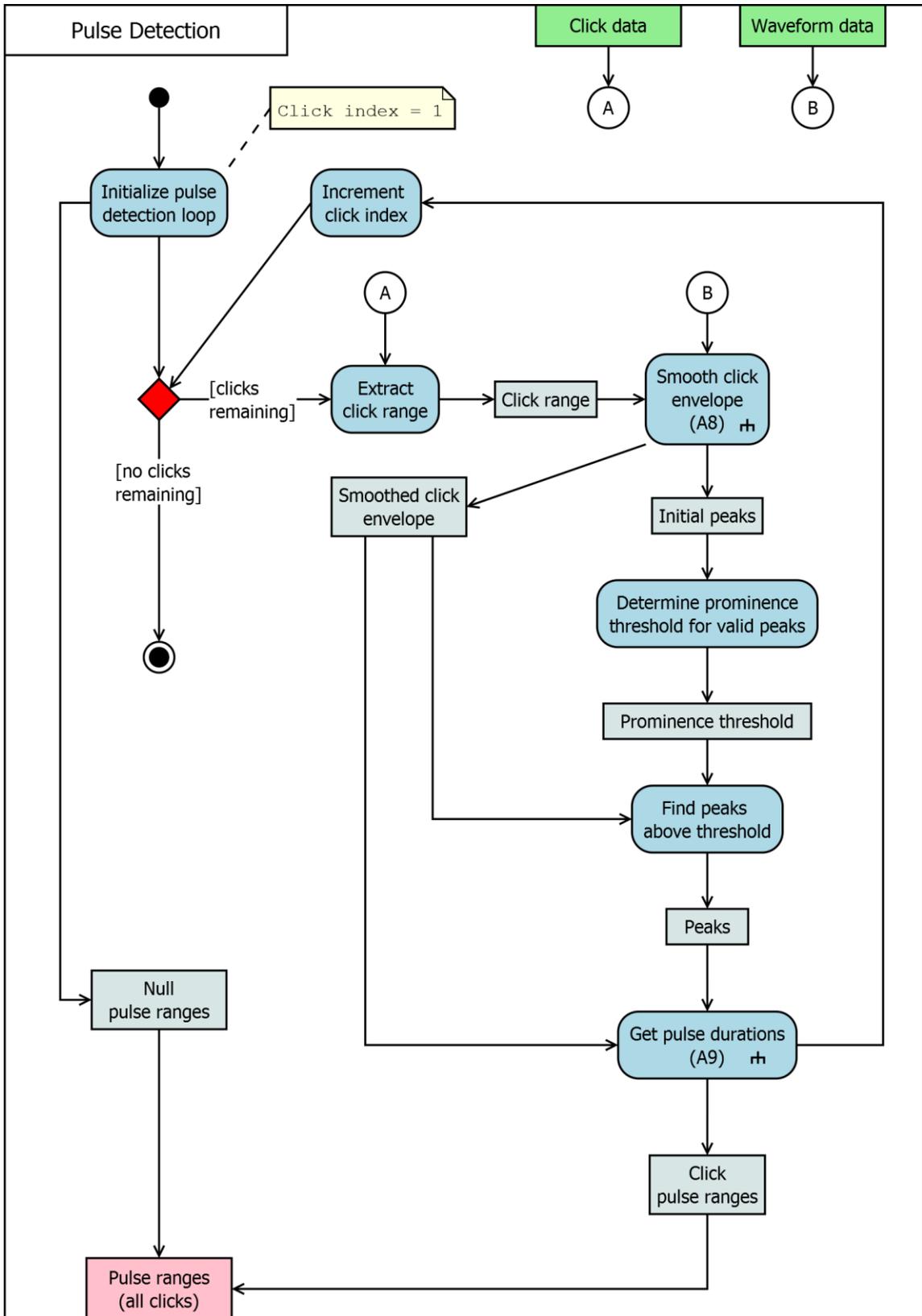


Figure A.7 Overview of pulse detection procedure

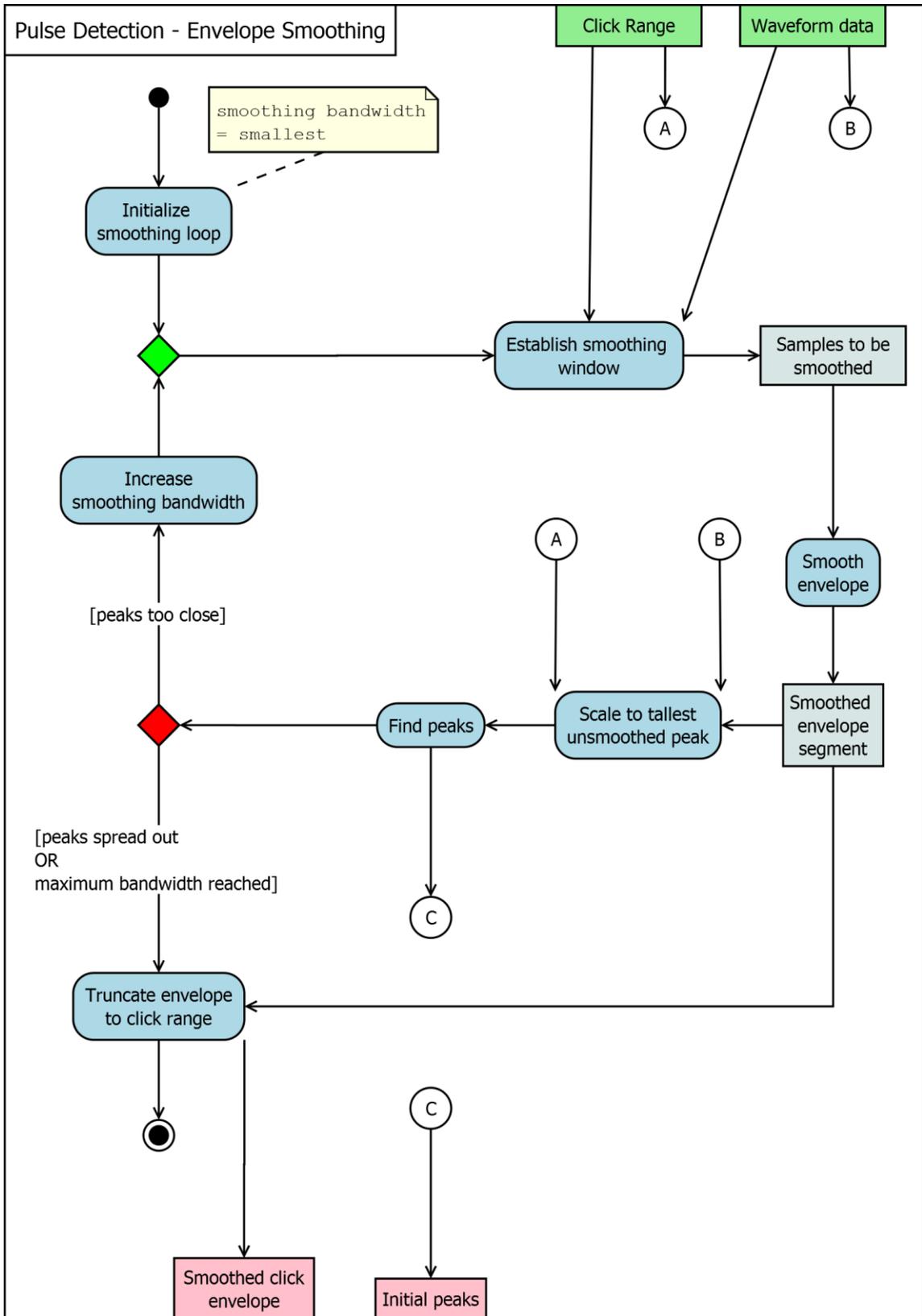


Figure A.8 Click envelope smoothing for pulse detection

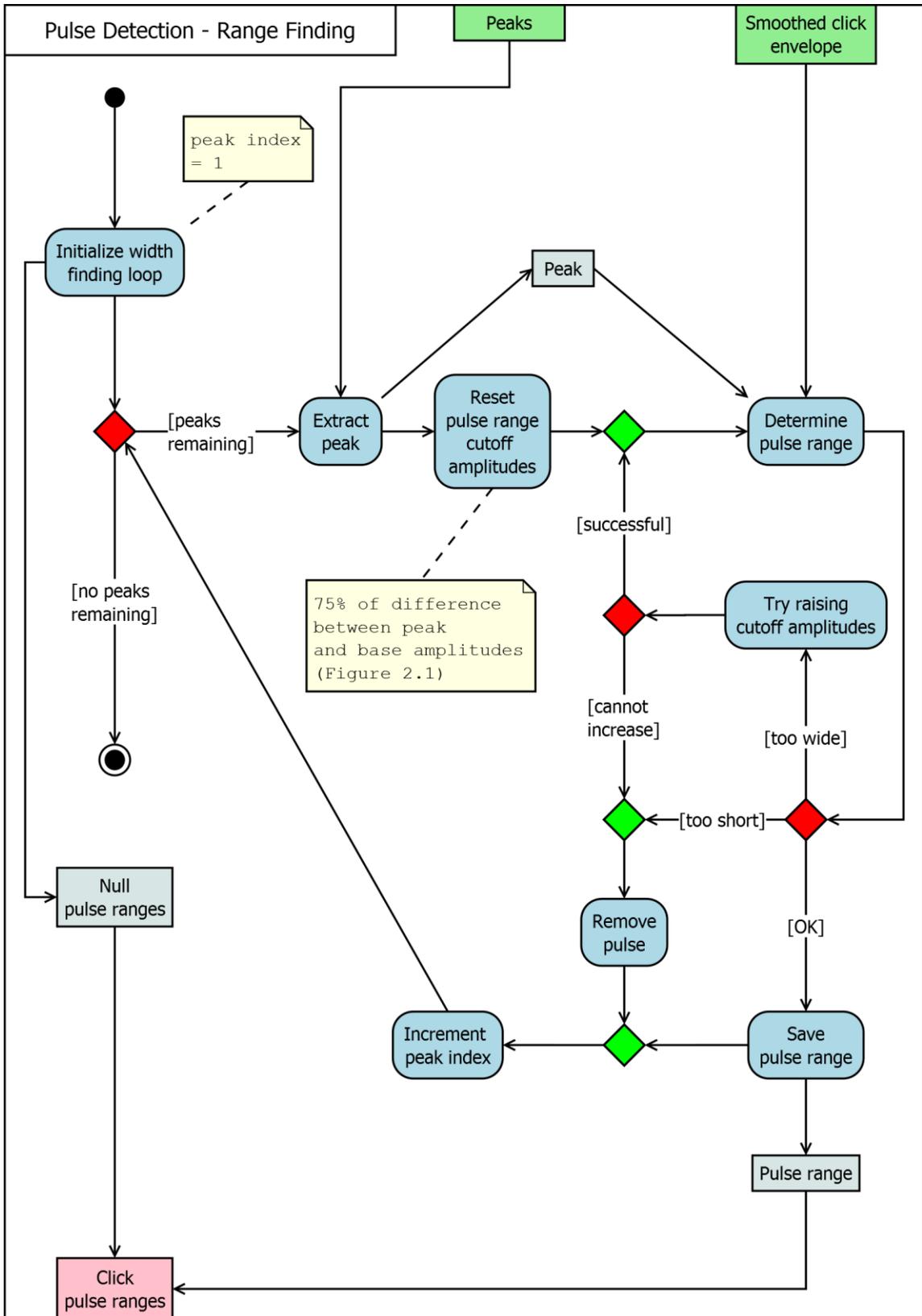


Figure A.9 Determination of pulse ranges



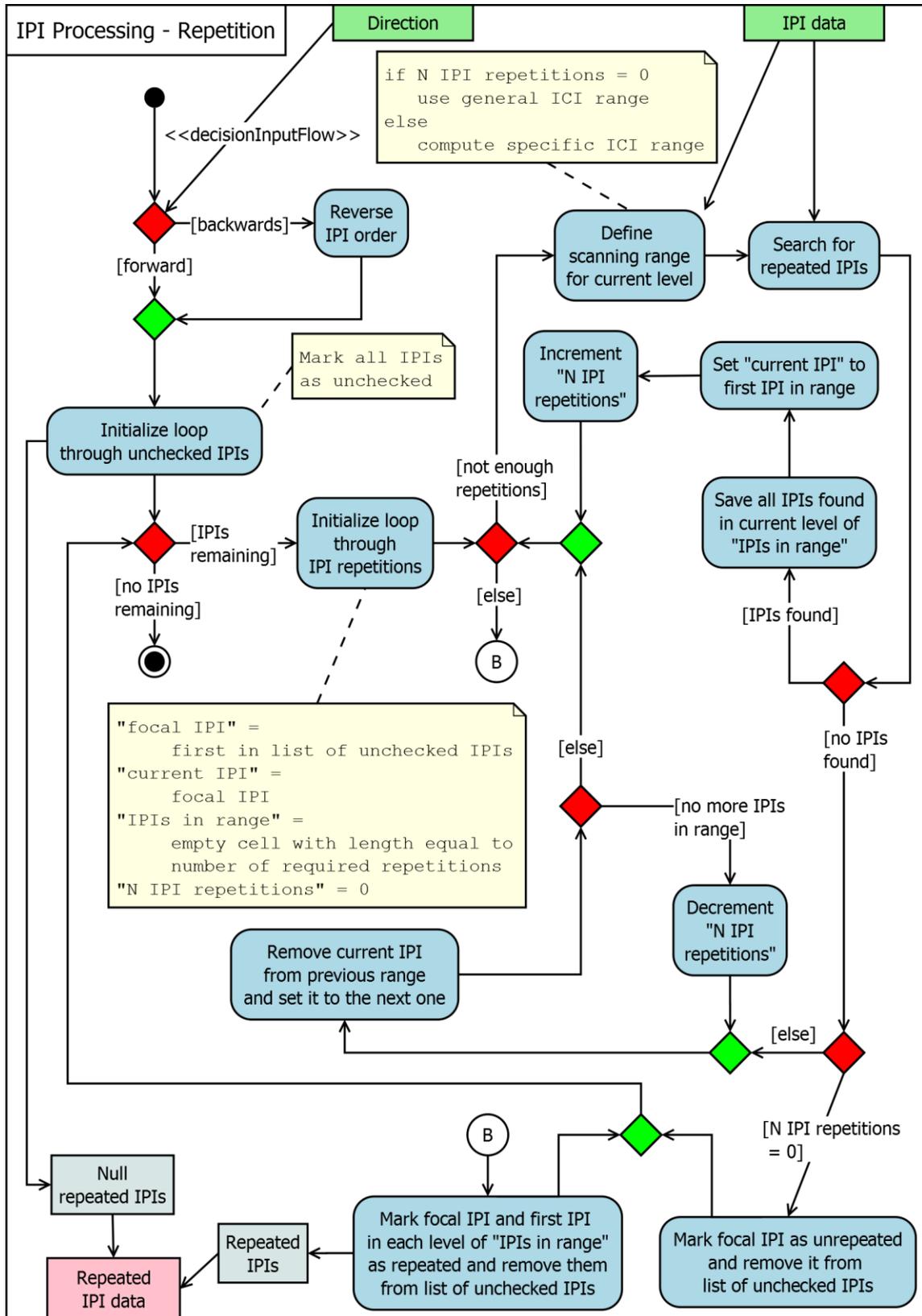


Figure A.11 IPI repetition check

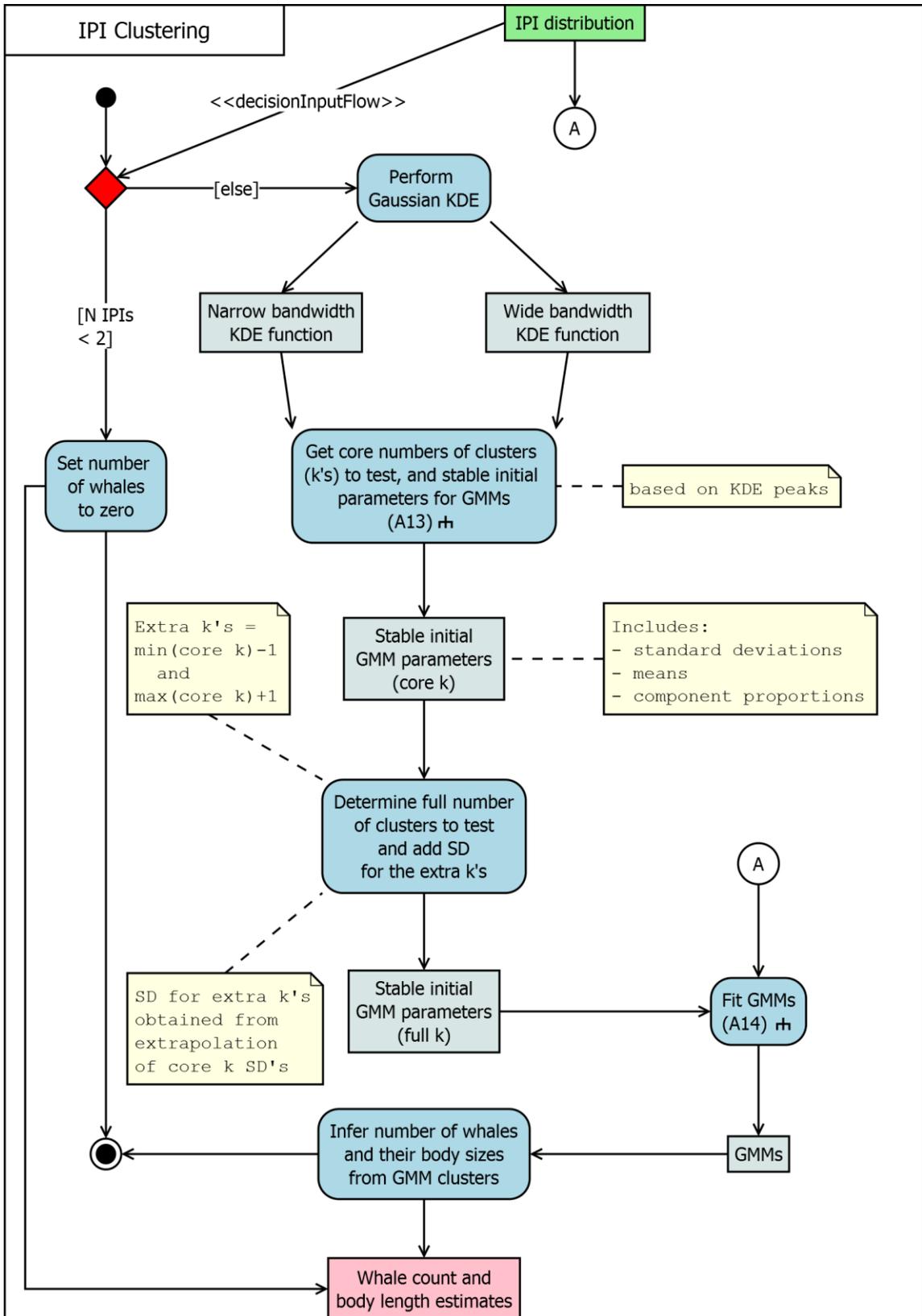
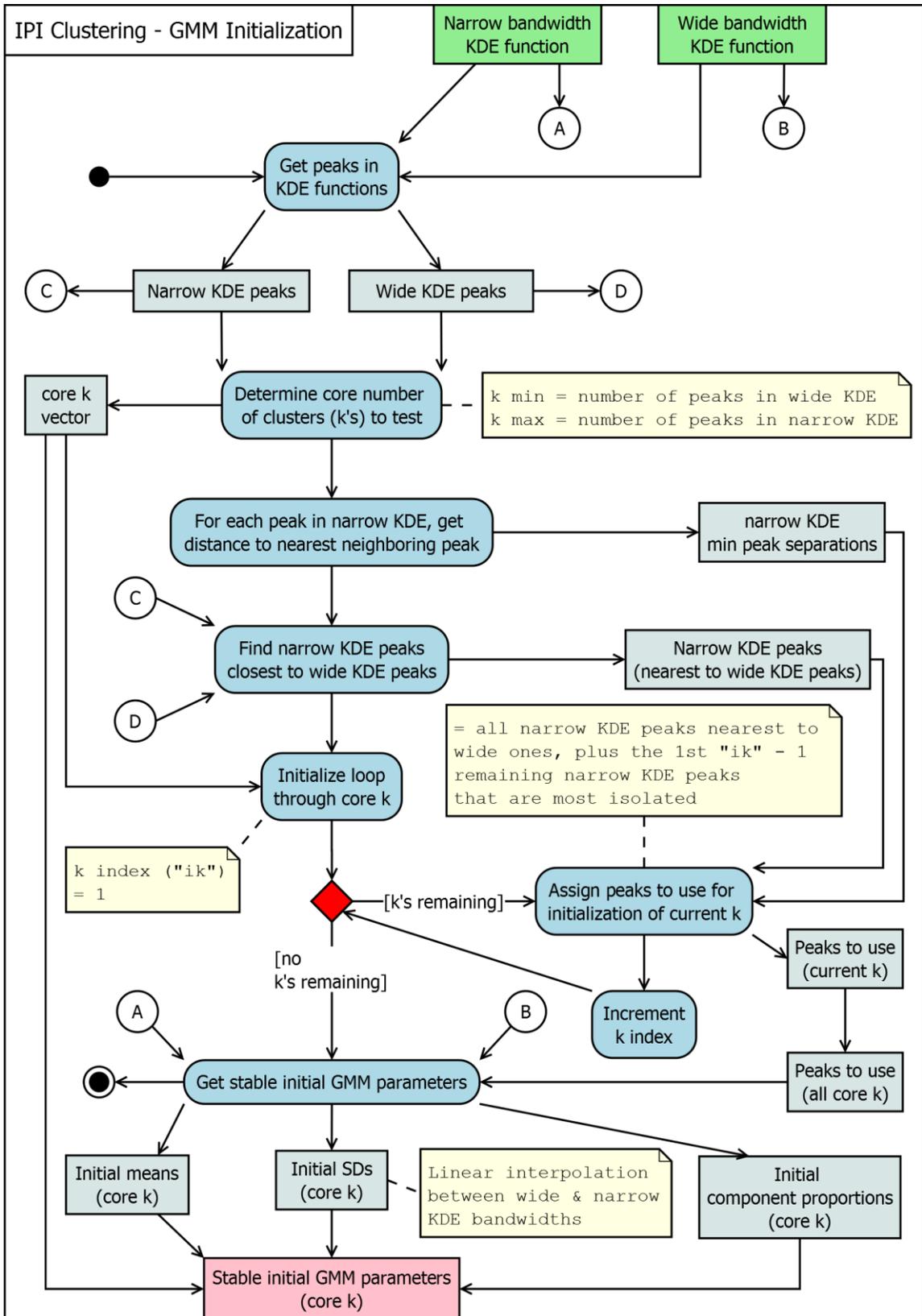


Figure A.12 Overview of IPI clustering



**Figure A.13** Determination of stable (non-random) parameters for GMM initialization

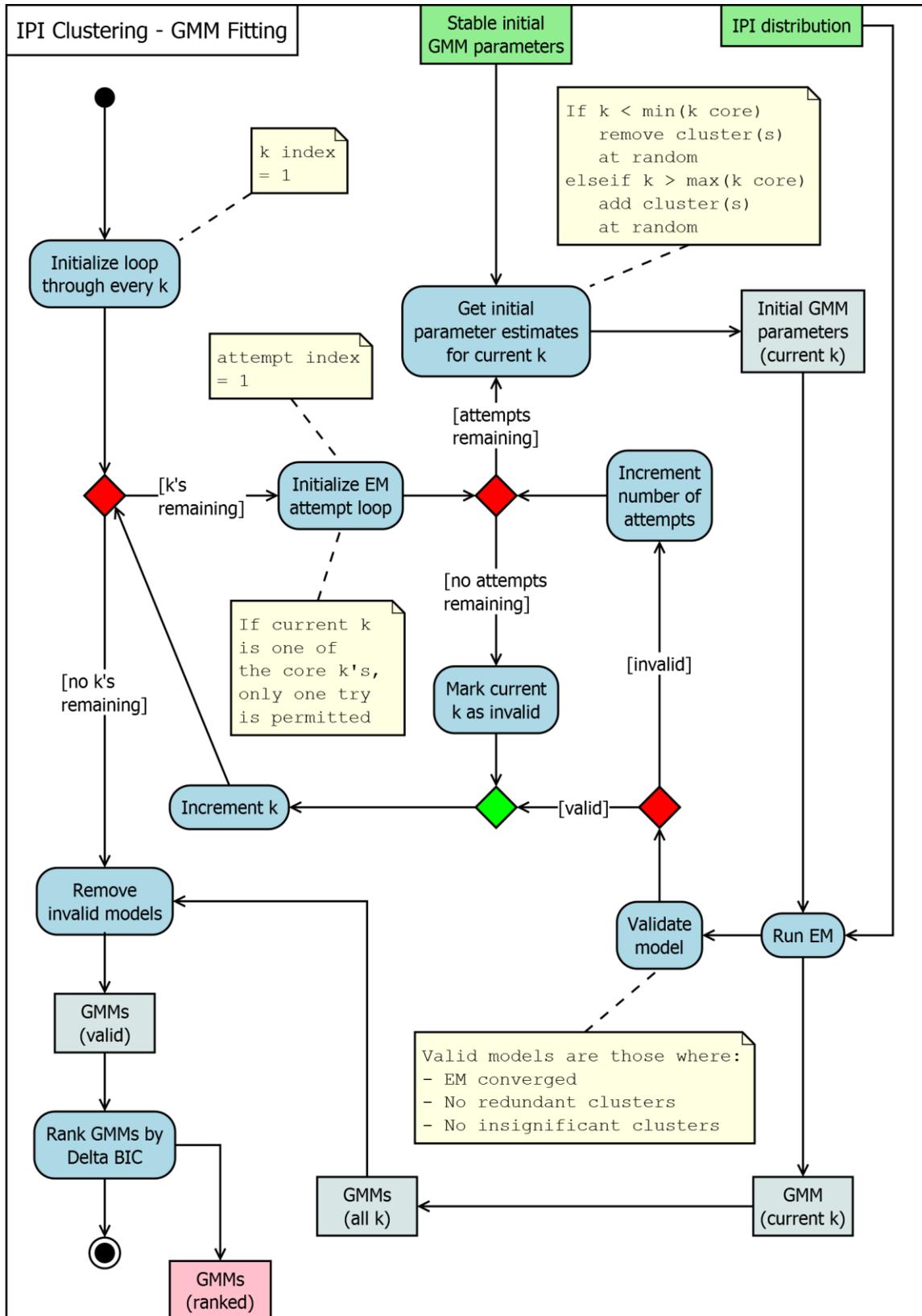


Figure A.14 Procedure for fitting GMMs to IPI distributions

## APPENDIX B - LIST OF PARAMETERS FOR AUTOMATIC IPI COMPILATION ALGORITHM

The following tables describe all parameters used by the IPI compilation routine that may be modified. Default values for each parameter have been selected such that they result in adequate performance under a wide range of recording qualities. With the exception of two parameters (specifically, *nIPIReps* and *minGoodProb*), these are essentially “advanced” parameters: default values are sufficient for casual use and do not need to be modified.

Square brackets around default values indicate vector inputs (i.e. multi-variable parameters).

**Table B.1** List of key parameters used by the IPI compilation routine. Sensitivity analysis for these parameters is presented in Chapter 2.

| Parameter   | Default | Description  |
|-------------|---------|--|
| nIPIReps    | 1       | Number of times that an IPI must be repeated in succession to be considered valid. This parameter has a very large impact on the output.<br><br>Default: Based on performance evaluation for 174 recordings lasting approximately 4 minutes. See Chapter 2.  |
| minGoodProb | 0.7     | Probability threshold that determines if a click is “Good” or not. The automatic classifier assigns probabilities to each click based on how well it fits in the “Good” class. Clicks with a “Goodness” probability greater than the threshold are then labelled as “Good”.<br><br>Default: Based on performance evaluation for 174 recordings lasting approximately 4 minutes. See Chapter 2. |

**Table B.2** List of advanced parameters used in multiple steps of the IPI compilation routine

| Parameter            | Default   | Description  |
|----------------------|-----------|--|
| IPIRange             | [2, 9]    | <p>Full possible range of IPIs, in milliseconds. It dictates the lower and upper limits within which IPIs are calculated for both autocorrelation and cepstrum methods. In click detection, the maximum also limits the distance within which detections can be merged.</p> <p>Default: 2-9 milliseconds was the range used by Marcoux <i>et al.</i> (2006). According to them, values below 2 ms may result in erroneous IPIs due to long first pulses. 9 ms is an acceptable upper limit because it is longer than that of the mature males measured by Rhinelander and Dawson (2004).</p> |
| pulseDurationRange   | [0.05, 1] | <p>Minimum and maximum expected pulse durations, in milliseconds. In click detection, the minimum is a lower limit for click duration: clicks shorter than this are immediately removed. In pulse detection, the minimum is also a threshold below which detections are removed, and the maximum is a threshold above which pulses are trimmed, or rejected if trimming is not possible.</p> <p>Default: Based on visual observation of pulses within several sperm whale clicks. These values should tolerate extreme but plausible cases.</p>  |
| tukeyFalloffDuration | 5         | <p>The extent of each falloff region in a Tukey window, in milliseconds. Tukey windows are used prior to <i>FFT</i> to minimize ringing in the spectra of clicks. Small values result in greater ringing, whereas large values include more of the surrounding waveform, potentially contaminating the spectrum. Setting this to 0 is equivalent to using a rectangular window.</p> <p>Default: Reasonable tradeoff between amount of ringing and inclusion of non-click components.</p>   |

**Table B.3** List of advanced parameters for click detection

| Parameter     | Default  | Description   |
|---------------|----------|---|
| threshOn      | 10       | <p>Linear SNR threshold that defines click detection events during Page test. When <math>V \geq threshOn</math>, a new click is detected.</p> <p>Default: Heuristically determined to be effective and robust against a wide range of SNRs</p>  |
| threshOff     | 1        | <p>Linear SNR threshold that marks the limits of a detected click during Page test. After a detection is triggered by <i>threshOn</i>, its range in the time series is recorded based on the surrounding region where <math>V &gt; threshOff</math>.</p> <p>Default: Heuristically determined to be effective and robust against a wide range of SNRs</p> |
| alphaSignal   | 0.2      | <p>Smoothing factor for the signal power estimation filter.</p> <p>Default: Heuristically determined to be effective and robust against a wide range of SNRs, for <math>F_s = 48</math> kHz.</p>  |
| alphaNoiseOn  | 0.000002 | <p>Smoothing factor for the noise power estimation filter during the "signal" state.</p> <p>Default: Heuristically determined to be effective and robust against a wide range of SNRs, for <math>F_s = 48</math> kHz.</p>   |
| alphaNoiseOff | 0.0002   | <p>Smoothing factor for the noise power estimation filter during the "noise" state.</p> <p>Default: Heuristically determined to be effective and robust against a wide range of SNRs, for <math>F_s = 48</math> kHz.</p>  |

**Table B.3 (Continued)**

| Parameter        | Default | Description   |
|------------------|---------|---|
| minEchoProp      | 0.6     | <p>Minimum proportion of a click envelope peak beyond which the next click will be considered an echo. Setting this too low may break up multi-pulsed clicks, while setting it too high may cause clicks to merge with their reflections.</p> <p>Default: Based on a rough observation of average difference between the maximum amplitude of clicks and their surface reflections.</p>   |
| maxClickDuration | 40      | <p>Maximum expected click duration, in milliseconds. Forces termination of the "signal" state during the Page test.</p> <p>Default: A value that is larger than the duration of virtually all sperm whale clicks. This value was also used by Miller (2010).</p>  |
| minClickSep      | 0       | <p>Minimum expected click separation, in milliseconds. During the Page test, when the state switches from "signal" to "noise", it will be forced to remain in "noise" for the duration of this value. This helps prevent the detection of echoes, but also risks rejecting some direct-path clicks when multiple whales are clicking simultaneously.</p> <p>Default: With a value of 0, this parameter is effectively disabled by default. It is only useful if one whale is present, or if the ICI is otherwise very predictable, which rarely occurs when recording female sperm whales. Thus, this parameter was ultimately never used for analysis.</p> |

**Table B.4** List of advanced parameters for pulse detection

| Parameter          | Default          | Description   |
|--------------------|------------------|---|
| smoothBandwidths   | [0.5, 1, 1.5, 2] | <p>Vector of candidate bandwidths for smoothing click waveform envelopes using locally weighted linear regression (LOWESS). Bandwidth correspond the (temporal) range of data used to compute the regression for a given sample. Larger values include more samples, resulting in greater smoothing. They are measured in milliseconds.</p> <p>Default: Heuristically determined to perform well, based on visual observation of smoothed envelopes for several clicks.</p> |
| nSmoothRuns        | 2                | <p>Number of times to apply envelope smoothing.</p> <p>Default: Heuristically determined to perform well, based on visual observation of smoothed envelopes for several clicks.</p>   |
| peakBaseHeightProp | 0.75             | <p>Proportion of the difference between the amplitude of a peak and one of its secondary bases, which determines the amplitude cutoff that defines the endpoints of a pulse. In other words, this parameter controls how “wide” a pulse is.</p> <p>Default: Heuristically determined to perform well, based on visual observation of automatically-detected pulse ranges for several clicks.</p>  |

**Table B.4 (Continued)**

| <b>Parameter</b>   | <b>Default</b> | <b>Description</b>   |
|--------------------|----------------|--|
| promThreshProp     | 0.05           | <p>Proportion of the difference between the prominences of the most prominent and least prominent peaks in a smoothed click envelope, use to establish a threshold for “significant” prominence. If a peak has a prominence greater than the threshold, it is considered to mark the location of a pulse.</p> <p>Default: Heuristically determined to perform well, based on visual observation of automatically-detected pulse ranges for several clicks.</p> |
| minPromThreshScale | 2              | <p>Scale factor relative to the RMS of the absolute value of the difference of the unsmoothed envelope. This is meant to provide a measure of peak prominences for peaks which arise only from noise, and is thus the minimum allowable prominence threshold.</p> <p>Default: Heuristically determined to perform well, based on visual observation of automatically-detected pulse ranges for several clicks.</p>   |

**Table B.5** List of advanced parameters for IPI calculation and validation

| Parameter                         | Default | Description   |
|-----------------------------------|---------|---|
| doMethod_<br>Autocorrelation      | true    | <p>Determines if IPIs should be computed using the autocorrelation method.</p> <p>Default: IPIs are more precise when both autocorrelation and cepstrum methods are enabled.</p>  |
| doMethod_<br>Cepstrum             | true    | <p>Determines if IPIs should be computed using the cepstrum method.</p> <p>Default: IPIs are more precise when both autocorrelation and cepstrum methods are enabled.</p>   |
| useChiSquared_<br>Autocorrelation | false   | <p>Determines if <math>\chi^2</math> type windows should be used when computing IPIs through autocorrelation. <math>\chi^2</math> windowing was proposed by Goold (1996) as a means of amplifying the signal of successive pulses within a click.</p> <p>Default: <math>\chi^2</math> windows were used by Goold (1996) for cepstral analysis only. The application of <math>\chi^2</math> windows to autocorrelation was experimented with for this thesis, but this was found to have minimal impact.</p> |
| useChiSquared_<br>Cepstrum        | true    | <p>Determines if <math>\chi^2</math> type windows should be used when computing IPIs through cepstral analysis. <math>\chi^2</math> windowing was proposed by Goold (1996) as a means of amplifying the signal of successive pulses within a click.</p> <p>Default: As in Goold (1996), <math>\chi^2</math> windows were deemed to be appropriate for cepstral IPI calculation.</p>   |

**Table B.5 (Continued)**

| Parameter       | Default     | Description   |
|-----------------|-------------|---|
| maxIPIDeviation | 0.05        | <p>The maximum allowable difference between each IPI estimate of a given click (i.e. autocorrelation- and cepstrum-derived IPIs), and the final mean IPI of that click. This is measured in milliseconds. This parameter determines if a click's IPI is "precise" or not. Clicks that do not meet this precision criterion are rejected.</p> <p>Default: Determined based on IPI precision of clicks that were manually classified as on-axis, and with consideration of IPI precision reported by Schulz <i>et al.</i> (2011).</p> |
| ICIRange        | [0.25, 1.5] | <p>Minimum and maximum expected ICI, in seconds. This is the range within which potentially repeated clicks are searched for.</p> <p>Default: Based on the observed ICI range of sperm whale usual clicks (female and male).</p>  |
| ICITol          | 0.2         | <p>The maximum allowable deviation from the expected ICI of a click train, in seconds. This parameter becomes relevant only if <math>nIPIReps &gt; 1</math>. Once a first repetition is found within <i>ICIRange</i>, its ICI is remembered, and the next repetitions are searched within that <math>ICI \pm ICITol</math>.</p> <p>Default: Heuristically determined to perform well, with consideration of observed stability in the ICIs of several echolocation click trains.</p>  |
| IPITol          | 0.05        | <p>The maximum allowable difference between the IPI estimates of two successive clicks in a click train, in milliseconds. This parameter is used when checking for IPI repetition.</p> <p>Default: Heuristically determined to perform well, with consideration of observed differences in IPI between successive on-axis clicks, and IPI precision reported by Schulz <i>et al.</i> (2011).</p>  |

**Table B.6** List of advanced parameters for IPI clustering.

| Parameter     | Default        | Description   |
|---------------|----------------|---|
| KDEBandwidths | [0.1, 0.05]/3  | <p>Bandwidth values for Gaussian kernel density estimation. Narrow and wide-bandwidth KDE is used to initialize the Gaussian mixture modelling procedure.</p> <p>Default: Based on the possible range of variability in variance of an individual whale's observed IPI.</p>   |
| nkExtra       | 1              | <p>The number of extra clusters to test for during mixture modelling, beyond the estimates provided by KDE. Larger values mean a greater number of GMMs are tested.</p> <p>Default: Heuristically determined to perform well. Based on several observed model selection outcomes, there is no need for this value to be any larger.</p>                       |
| shareSigma    | true           | <p>Determines if all clusters in a GMM should have the same variance or not.</p> <p>Default: Based on the expectation that the variance in observed IPIs of individual whales should not differ greatly within a given recording. Also, when variance is unconstrained, mixture modelling has a tendency to include dense groups of peaks as one cluster.</p> |
| sigma2RegVal  | $((1/Fs)/4)^2$ | <p>Regularization value for variance in GMMs. This is a small number systematically added to the variance of each cluster during EM fitting.</p> <p>Default: Designed to account for IPI quantization due to limited sampling resolution.</p>   |

**Table B.6 (Continued)**

| <b>Parameter</b> | <b>Default</b> | <b>Description</b>  |
|------------------|----------------|---|
| EMTol            | 1e-9           | <p>Tolerance value that determines when maximum likelihood has converged during EM fitting. Large values result in faster convergence and thus less accurate models, whereas small values require more iterations and thus more computation time.</p> <p>Default: Somewhat arbitrary, but with priority to accuracy over computation speed.</p>                               |
| maxEMIterations  | 1000           | <p>Maximum number of EM iterations. Small values increase the risk of GMMs not converging, while large values can increase computation time for models that are difficult to fit.</p> <p>Default: Somewhat arbitrary, but with priority to accuracy over computation speed. It is rare for plausible models to require many iterations, so there is no need to go higher.</p> |
| maxEMTries       | 3              | <p>In some cases, the initial conditions of an EM run may be ill-suited and cause the run to fail. This parameter is the maximum allowed number of EM attempts for a particular model.</p> <p>Default: Somewhat arbitrary. Since models that fail the first time are often not good ones, the impact of this parameter is rather minimal.</p>                                 |

## APPENDIX C - DATASET FOR TESTING AUTOMATIC UNIT DETECTION

The following tables describe parts of the dataset used for testing the automatic unit detection algorithm presented in Chapter 3, and how it was organized and processed. This dataset consists of photo-identifications of whales and audio recordings taken off the coast of Dominica from February-April 2015.

**Table C.1** Individual sperm whales encountered off the west coast of Dominica from February-April 2015 and their unit designations. Key for sex/age class is as follows:

A = adult female or juvenile male

C = calf

M = mature male

Note that unit *FU* has historically been treated as two separate units in other studies (*F* and *U*; see Gero *et al.*, 2014). However, these units have been associating together more frequently in recent years, enough that they may now be considered to be one unit (Konrad, 2017). In 2015, units *F* and *U* were always seen within the study area together on the same days.

| Individual         | Sex/Age Class | Unit/Grouping |
|--------------------|---------------|---------------|
| 5586 (Atwood)      | A             | A             |
| 5712 (Lady Oracle) | A             | A             |
| 5714 (Rounder)     | A             | A             |
| 5719 (Soursop)     | A             | A             |
| 5720 (Fruit Salad) | A             | A             |
| 6088 (Allan)       | A             | A             |
| 6196 (Snowman)     | A             | A             |
| 5151 (Fork)        | A             | FU            |
| 5560 (Pinchy)      | A             | FU            |
| 5562 (Knife)       | A             | FU            |
| 5722 (Fingers)     | A             | FU            |
| 6058 (Canopener)   | A             | FU            |
| 6070 (Tweak)       | A             | FU            |

**Table C.1 (Continued)**

| <b>Individual</b> | <b>Sex/Age Class</b> | <b>Unit/Grouping</b> |
|-------------------|----------------------|----------------------|
| 6219 (Digit)      | C                    | FU                   |
| 5979 (Sophocles)  | A                    | J                    |
| 5981 (Laius)      | A                    | J                    |
| 5987 (Jocasta)    | A                    | J                    |
| 5133 (Nalgene)    | A                    | N                    |
| 5148 (Nicki)      | A                    | N                    |
| 5877 (Nanni)      | A                    | N                    |
| 5729 (Mrs. Right) | A                    | R                    |
| 5730 (Roger)      | A                    | R                    |
| 5731 (Rip)        | A                    | R                    |
| 5732 (Raucous)    | A                    | R                    |
| 5733 (Rita)       | A                    | R                    |
| 6208 (Rap)        | C                    | R                    |
| 57321 (Riot)      | C                    | R                    |
| 57332             | C                    | R                    |
| 5726 (Sam)        | A                    | S                    |
| 5759 (TBB)        | A                    | S                    |
| 6052 (Sally)      | A                    | S                    |
| 5144              | A                    | UNK                  |
| 6197              | A                    | UNK                  |
| 6199              | A                    | UNK                  |
| 6201              | A                    | UNK                  |
| 6202              | A                    | UNK                  |
| 6203              | A                    | UNK                  |
| 6204              | A                    | UNK                  |
| 6205              | A                    | UNK                  |
| 6206              | A                    | UNK                  |
| 6207              | A                    | UNK                  |

**Table C.1** (Continued)

| Individual | Sex/Age Class | Unit/Grouping |
|------------|---------------|---------------|
| 6212       | A             | UNK           |
| 6213       | A             | UNK           |
| 6214       | A             | UNK           |
| 6215       | A             | UNK           |
| 6216       | A             | UNK           |
| 6200       | M             | MALE          |
| 6209       | M             | MALE          |
| 6210       | M             | MALE          |
| 6211       | M             | MALE          |
| 6217       | M             | MALE          |
| 6218       | M             | MALE          |

**Table C.2** Audio recording groups used for the construction of unit IPI profiles. Units present were determined based on all whales photo-identified within 2 hours prior to the start of a recording, and 2 hours after the end of the recording. Unidentified whales were assigned to the *UNK* grouping. IPIs were extracted from each recording using the method described in Chapter 2, and were compiled together based on group. There were no recordings where unit *N* was alone.

| Group number | Units present | Total duration (hh:mm:ss) | Number of extracted IPIs |
|--------------|---------------|---------------------------|--------------------------|
| 1            | A             | 01:19:28                  | 144                      |
| 2            | FU            | 04:05:25                  | 1017                     |
| 3            | J             | 01:38:26                  | 961                      |
| 4            | R             | 03:34:21                  | 1511                     |
| 5            | S             | 01:20:14                  | 504                      |

**Table C.3** Audio recording groups used for testing automatic unit detection. Units present were determined based on all whales photo-identified within 2 hours prior to the start of a recording, and 2 hours after the end of the recording. Unidentified whales were assigned to the *UNK* grouping. IPIs were extracted from each recording using the method described in Chapter 2, and were compiled together based on group. Whales from units for which IPI profiles were not available are considered to be in the *UNK* grouping. This list considers unit *A* to be a known unit. For the case where *A* was treated as unknown, replace all instances of "A" in "Units present" to "UNK", and change the corresponding scenario to "Unknowns".

| Group number | Day    | Units present | Scenario  | Total duration (hh:mm:ss) | Number of extracted IPIs |
|--------------|--------|---------------|-----------|---------------------------|--------------------------|
| 1            | 09-Feb | A             | Singles   | 00:05:08                  | 20                       |
| 2            | 12-Feb | S             | Singles   | 00:04:29                  | 69                       |
| 3            | 12-Feb | R+S           | Multiples | 00:15:19                  | 85                       |
| 4            | 12-Feb | R+S+UNK       | Unknowns  | 00:16:11                  | 51                       |
| 5            | 13-Feb | UNK           | Unknowns  | 00:07:33                  | 27                       |
| 6            | 25-Feb | R             | Singles   | 00:12:26                  | 8                        |
| 7            | 25-Feb | R+UNK         | Unknowns  | 00:05_47                  | 11                       |
| 8            | 26-Feb | R+UNK         | Unknowns  | 00:15:52                  | 70                       |
| 9            | 26-Feb | R+UNK+MALE    | Unknowns  | 00:13:30                  | 164                      |
| 10           | 01-Mar | A             | Singles   | 00:17:00                  | 11                       |
| 11           | 02-Mar | A             | Singles   | 00:14:41                  | 38                       |
| 12           | 02-Mar | A+FU          | Multiples | 00:27:16                  | 125                      |
| 13           | 03-Mar | A+J+UNK       | Unknowns  | 00:04:18                  | 72                       |
| 14           | 04-Mar | FU            | Singles   | 00:19:36                  | 77                       |
| 15           | 08-Mar | A+J           | Multiples | 00:29:10                  | 286                      |
| 16           | 09-Mar | A+J+MALE      | Unknowns  | 00:04:01                  | 102                      |
| 17           | 12-Mar | A             | Singles   | 00:25:14                  | 67                       |
| 18           | 16-Mar | A+J           | Multiples | 00:12:17                  | 17                       |
| 19           | 19-Mar | FU+UNK        | Unknowns  | 00:44:06                  | 203                      |
| 20           | 20-Mar | S             | Singles   | 00:11:43                  | 206                      |
| 21           | 20-Mar | J+R+UNK+MALE  | Unknowns  | 00:09:43                  | 88                       |

**Table C.3 (Continued)**

| <b>Group number</b> | <b>Day</b> | <b>Units present</b> | <b>Scenario</b> | <b>Total duration (hh:mm:ss)</b> | <b>Number of extracted IPIs</b> |
|---------------------|------------|----------------------|-----------------|----------------------------------|---------------------------------|
| 22                  | 21-Mar     | R                    | Singles         | 00:09:18                         | 12                              |
| 23                  | 21-Mar     | FU+J+MALE            | Unknowns        | 00:12:45                         | 49                              |
| 24                  | 21-Mar     | R+MALE               | Unknowns        | 00:09:18                         | 71                              |
| 25                  | 21-Mar     | J+UNK+MALE           | Unknowns        | 00:04:03                         | 63                              |
| 26                  | 22-Mar     | R                    | Singles         | 01:19:39                         | 1082                            |
| 27                  | 23-Mar     | S                    | Singles         | 00:08:24                         | 84                              |
| 28                  | 23-Mar     | R+S                  | Multiples       | 00:20:22                         | 216                             |
| 29                  | 24-Mar     | R                    | Singles         | 00:09:33                         | 177                             |
| 30                  | 24-Mar     | R+S                  | Multiples       | 00:13:00                         | 5                               |
| 31                  | 25-Mar     | J                    | Singles         | 00:24:04                         | 335                             |
| 32                  | 25-Mar     | J+UNK                | Unknowns        | 00:39:15                         | 151                             |
| 33                  | 26-Mar     | FU                   | Singles         | 01:50:32                         | 266                             |
| 34                  | 30-Mar     | R                    | Singles         | 00:41:56                         | 42                              |
| 35                  | 31-Mar     | FU                   | Singles         | 01:29:32                         | 635                             |
| 36                  | 01-Apr     | S                    | Singles         | 00:23:18                         | 57                              |
| 37                  | 02-Apr     | J                    | Singles         | 00:08:06                         | 137                             |
| 38                  | 02-Apr     | J+UNK                | Unknowns        | 00:14:05                         | 45                              |
| 39                  | 02-Apr     | J+R+S+UNK            | Unknowns        | 00:13:23                         | 21                              |
| 40                  | 06-Apr     | R                    | Singles         | 00:35:43                         | 168                             |
| 41                  | 07-Apr     | S                    | Singles         | 00:32:19                         | 84                              |
| 42                  | 08-Apr     | FU                   | Singles         | 00:20:34                         | 23                              |
| 43                  | 08-Apr     | FU+UNK               | Unknowns        | 00:30:27                         | 42                              |
| 44                  | 10-Apr     | J                    | Singles         | 00:57:28                         | 486                             |
| 45                  | 10-Apr     | J+S                  | Multiples       | 00:31:10                         | 138                             |
| 46                  | 11-Apr     | J+S                  | Multiples       | 00:39:17                         | 124                             |

## APPENDIX D - RESULTS OF AUTOMATIC UNIT DETECTION FOR ALL RECORDING GROUPS

The following figures illustrate the results of the “All-All” unit detection test for all groups of recordings. There are 46 groups in total. The “All-All” test involved inferring which units were present from all recordings, by evaluating the likelihood of models for all possible combinations of units. Recordings were grouped by day and by units present. The figures show the distribution of IPIs extracted from each recording as a histogram (bin width = 1/48), with the probability density functions of the best supported model and the “true” model overlaid.

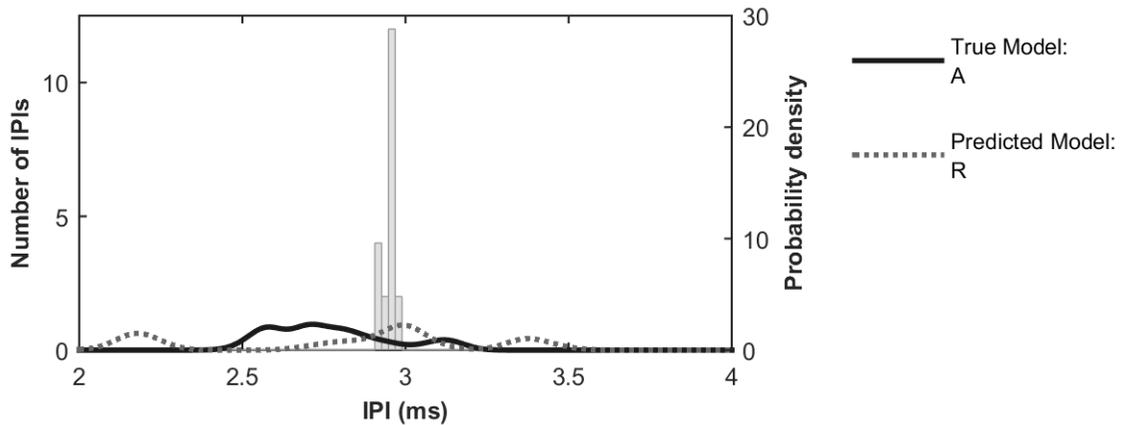


Figure D.1 Results for Group 1: Feb. 9, unit A

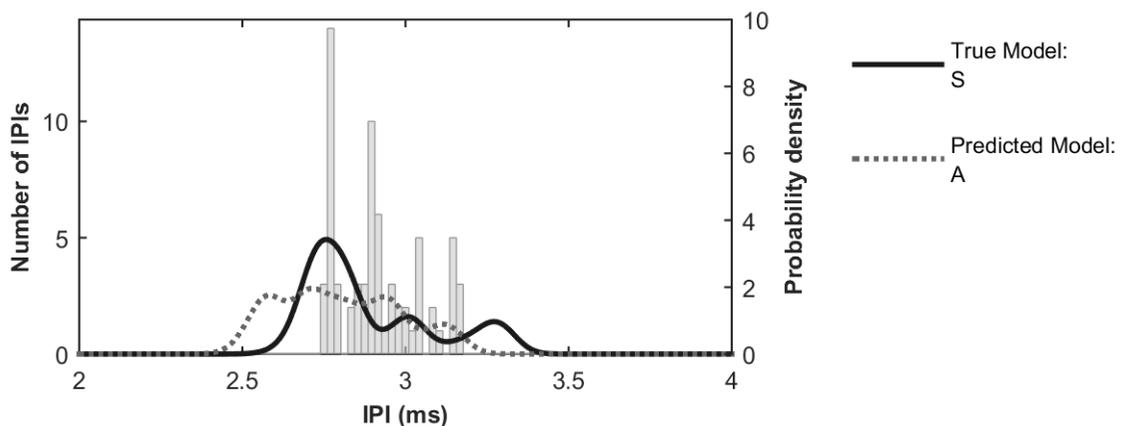
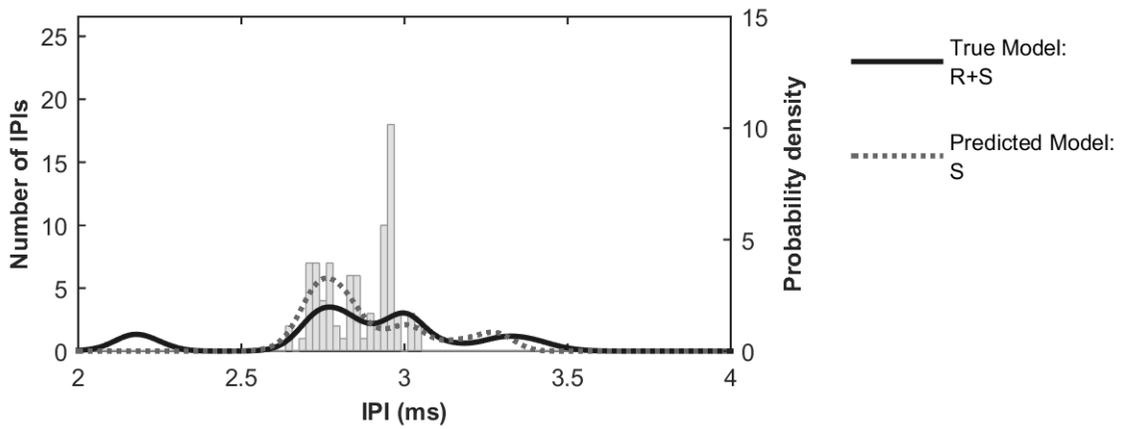
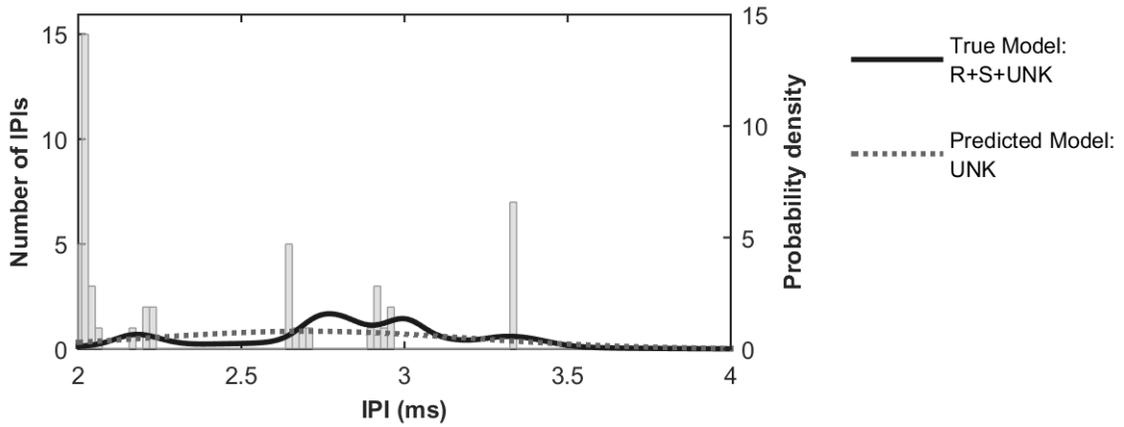


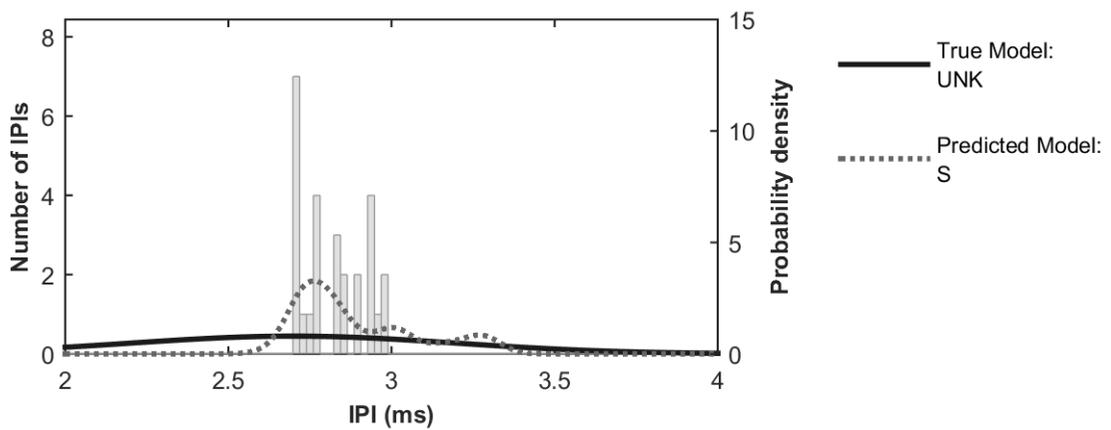
Figure D.2 Results for Group 2: Feb. 12, unit S



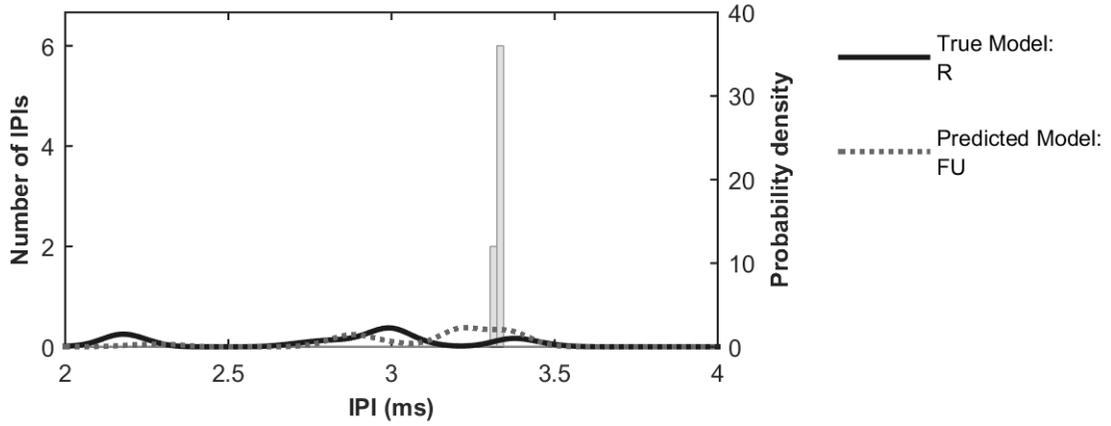
**Figure D.3** Results for Group 3: Feb. 12, units R + S



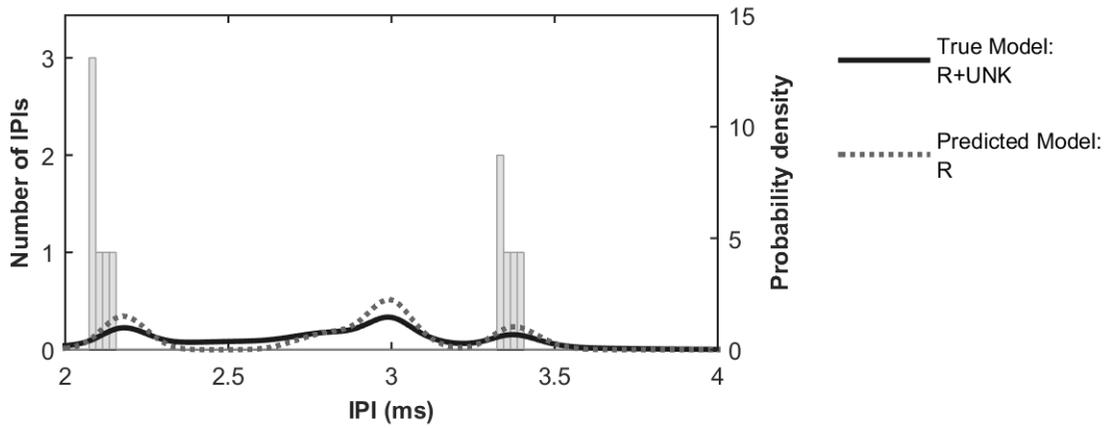
**Figure D.4** Results for Group 4: Feb. 12, units R + S + UNK



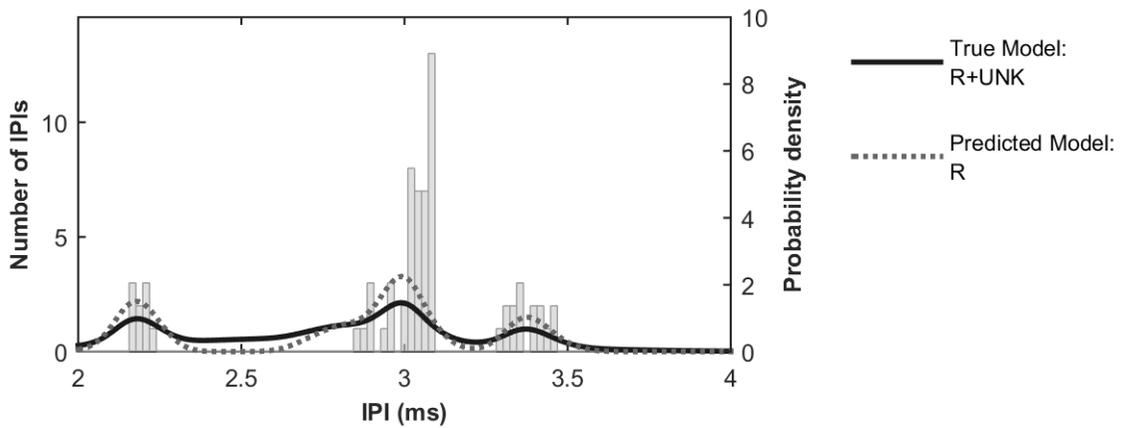
**Figure D.5** Results for Group 5: Feb. 13, UNK



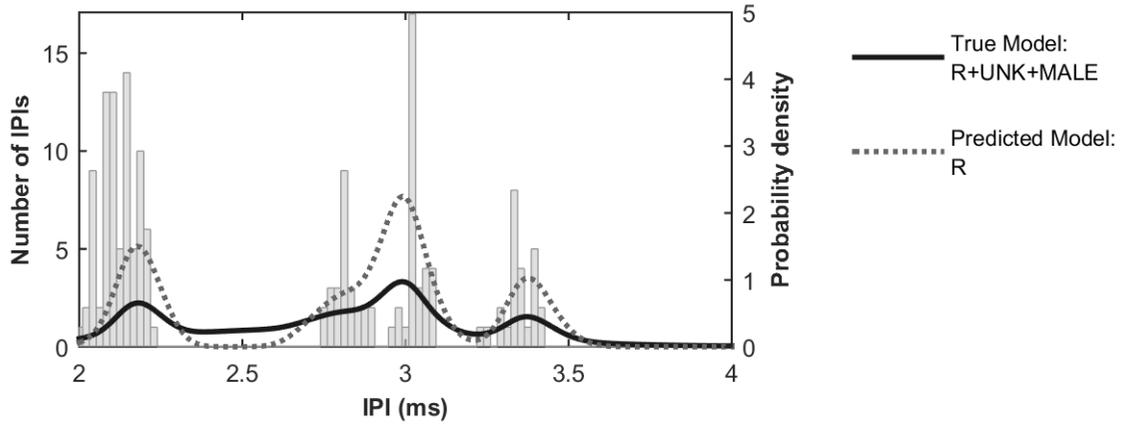
**Figure D.6** Results for Group 6: Feb. 25, unit R



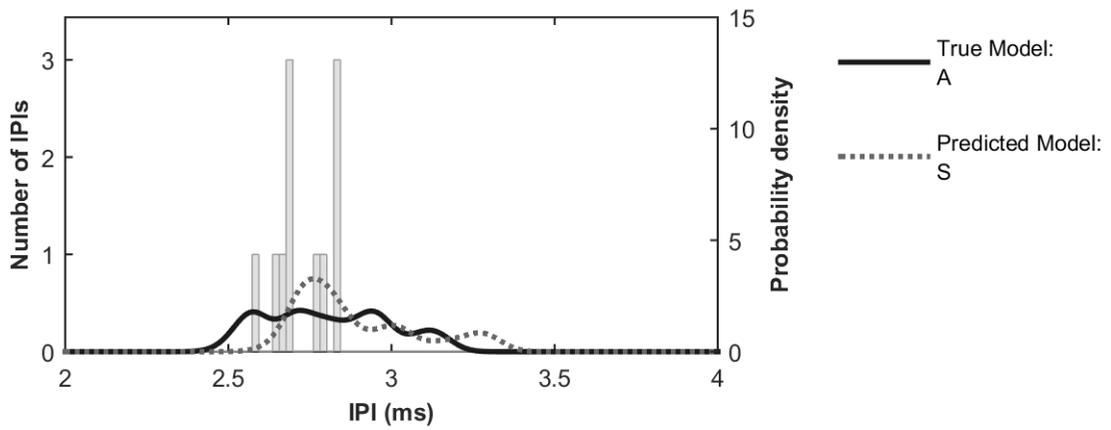
**Figure D.7** Results for Group 7: Feb. 25, unit R + UNK



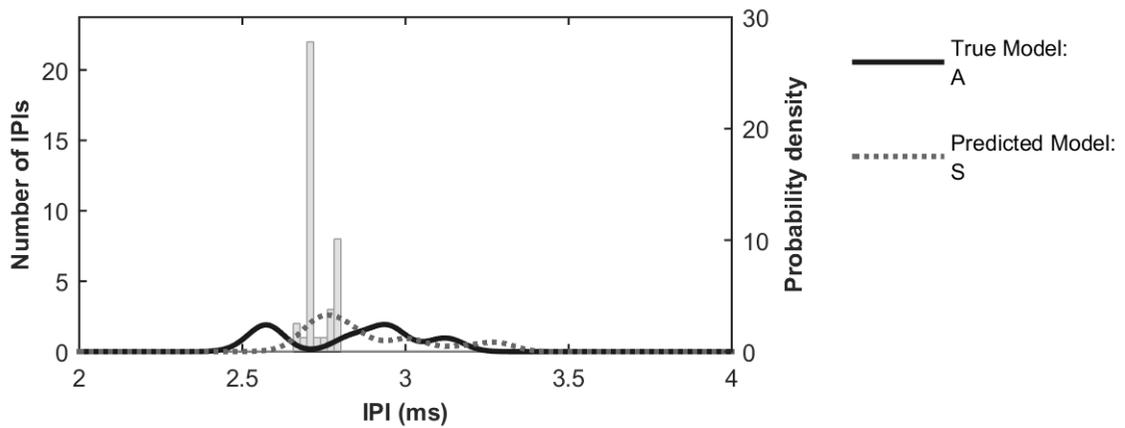
**Figure D.8** Results for Group 8: Feb. 26, unit R + UNK



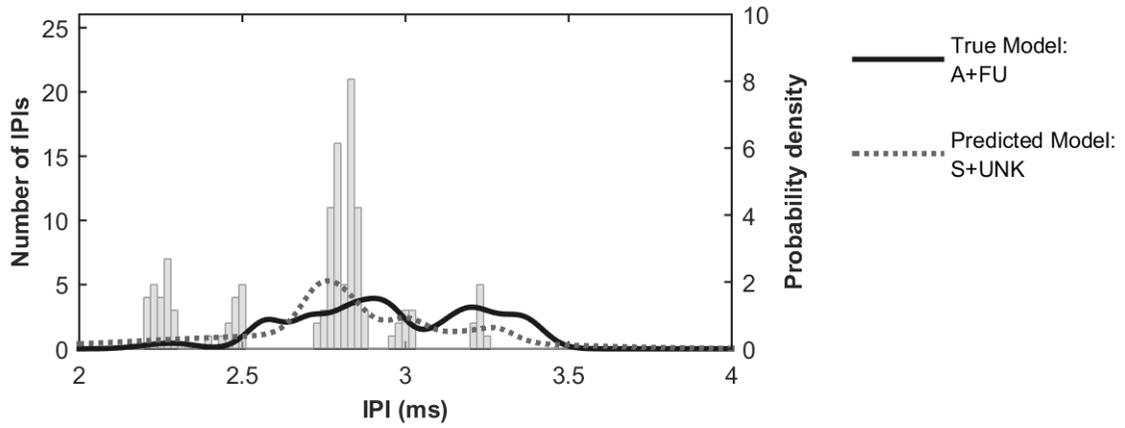
**Figure D.9** Results for Group 9: Feb. 26, unit R + UNK + MALE



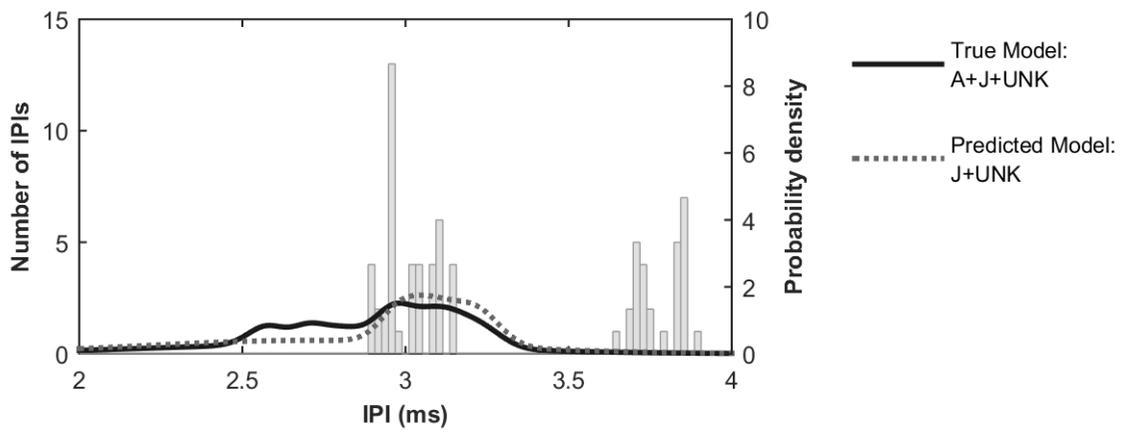
**Figure D.10** Group 10: Mar. 1, unit A



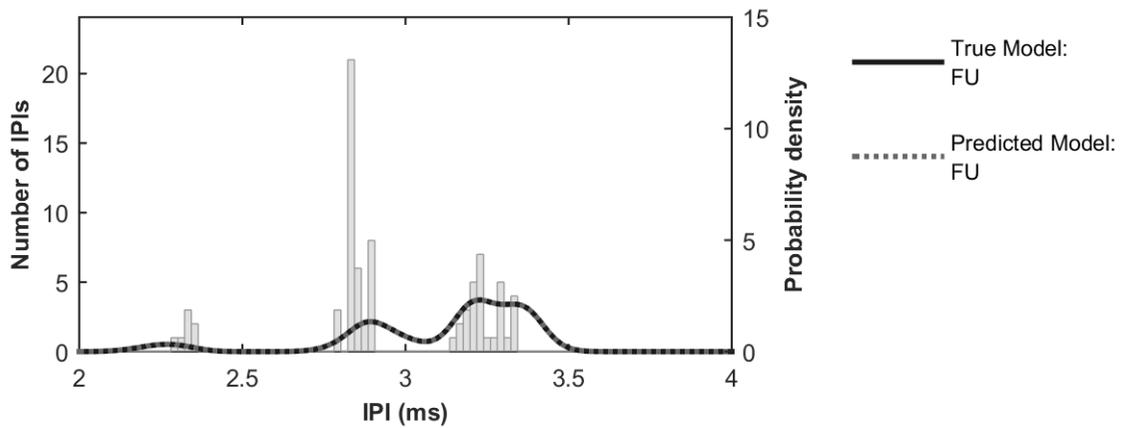
**Figure D.11** Results for Group 11: Mar. 2, unit A



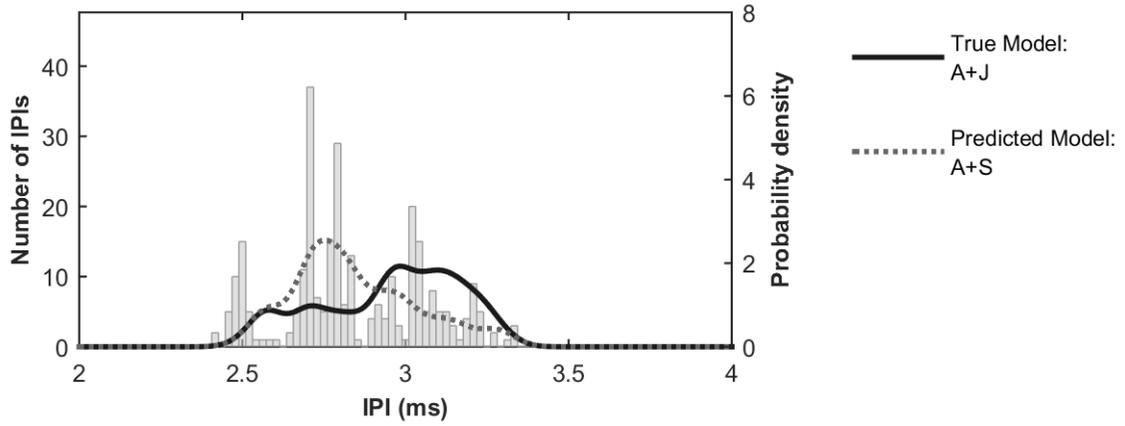
**Figure D.12** Results for Group 12: Mar. 2, units A + FU



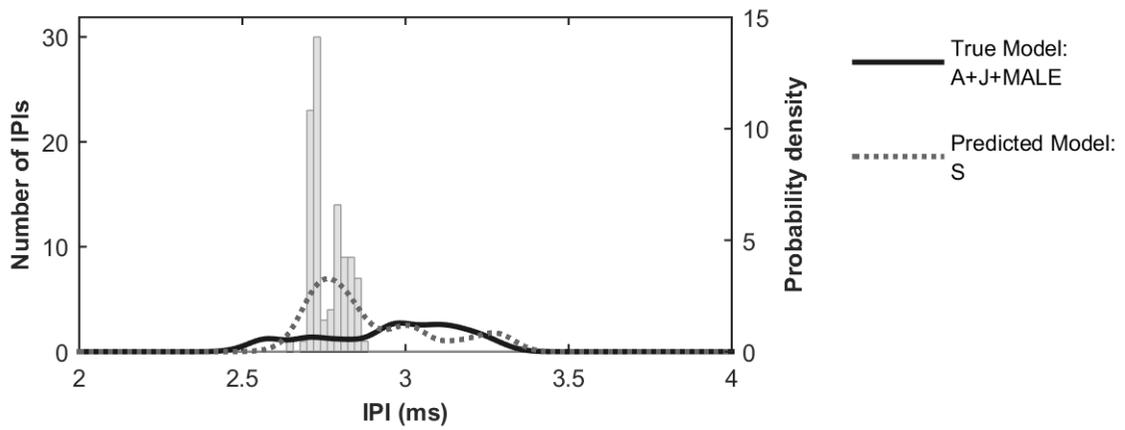
**Figure D.13** Results for Group 13: Mar. 3, units A + J + UNK



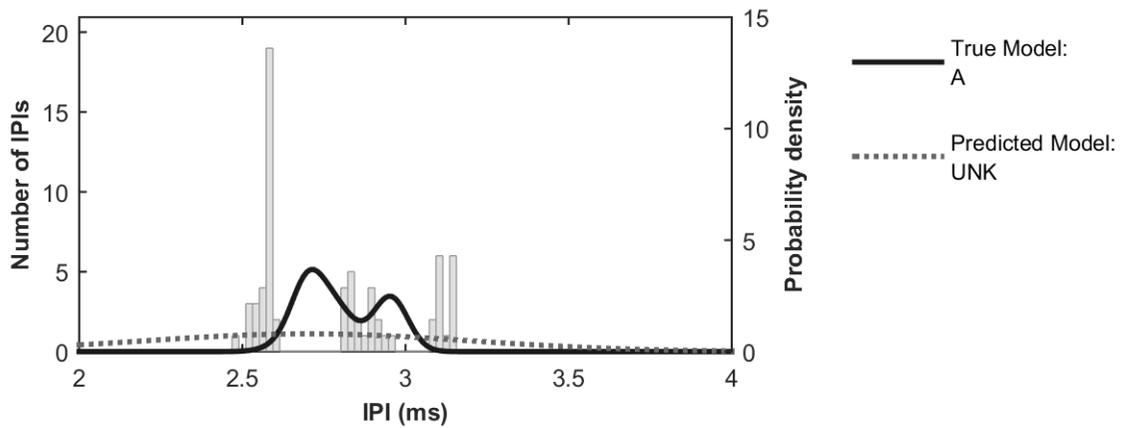
**Figure D.14** Results for Group 14: Mar. 4, unit FU



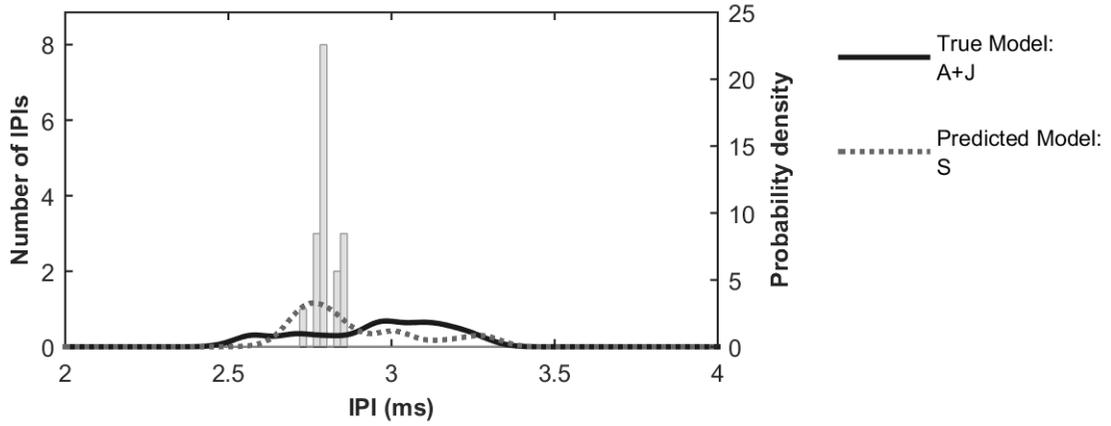
**Figure D.15** Results for Group 15: Mar. 8, units A + J



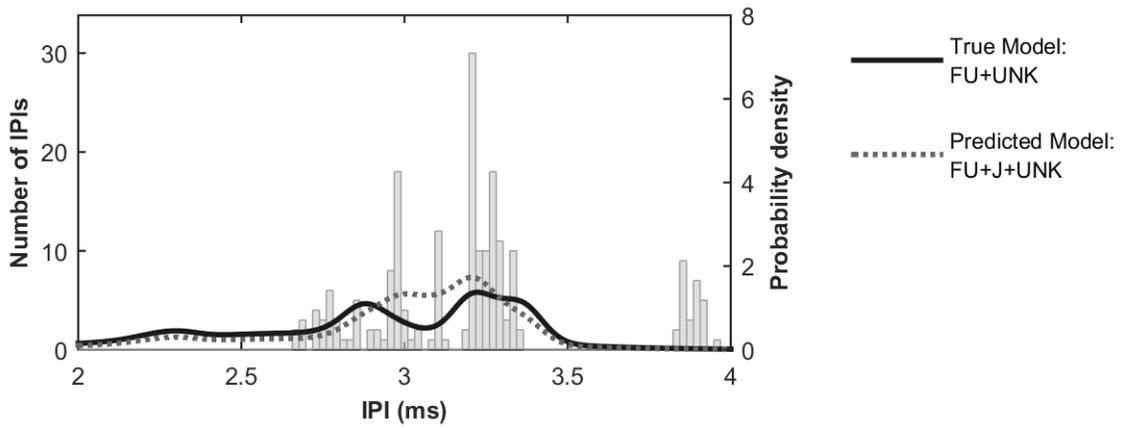
**Figure D.16** Results for Group 16: Mar. 9, units A + J + MALE



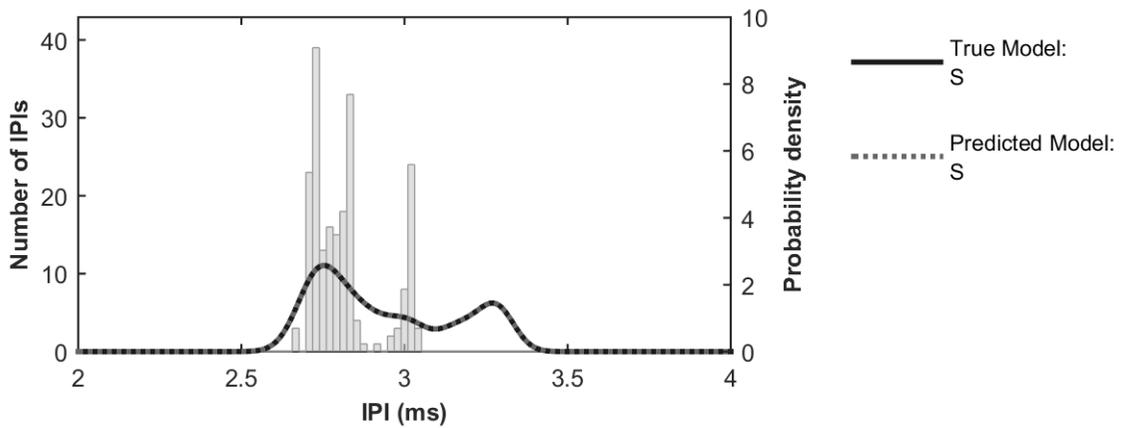
**Figure D.17** Results for Group 17: Mar. 12, unit A



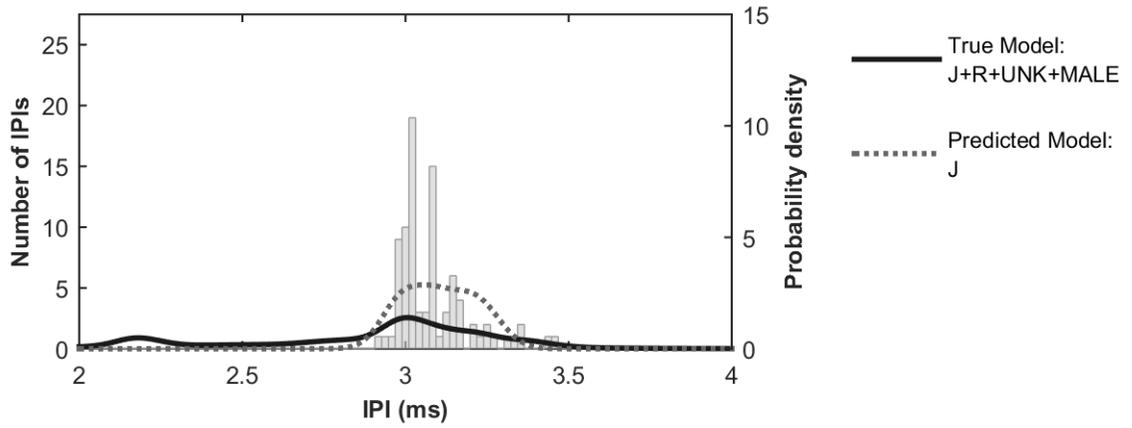
**Figure D.18** Results for Group 18: Mar. 16, units A + J



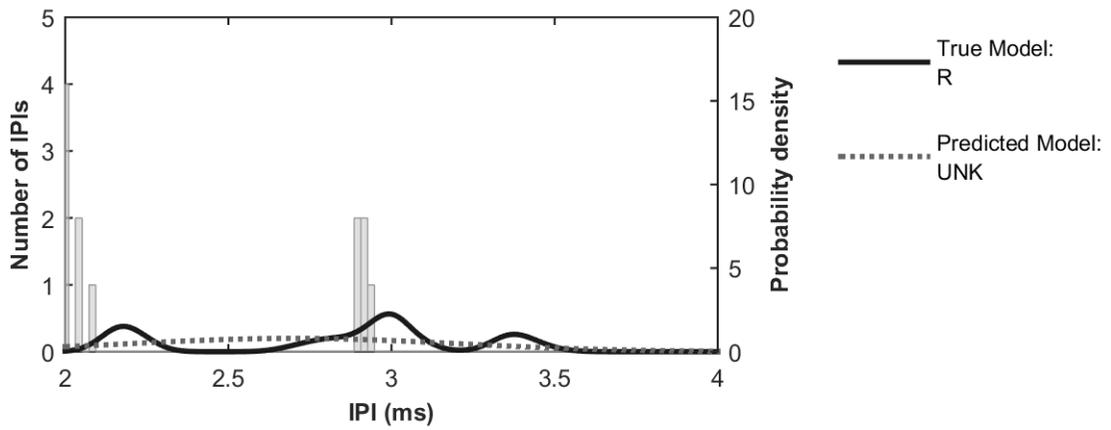
**Figure D.19** Results for Group 19: Mar. 19, unit FU + UNK



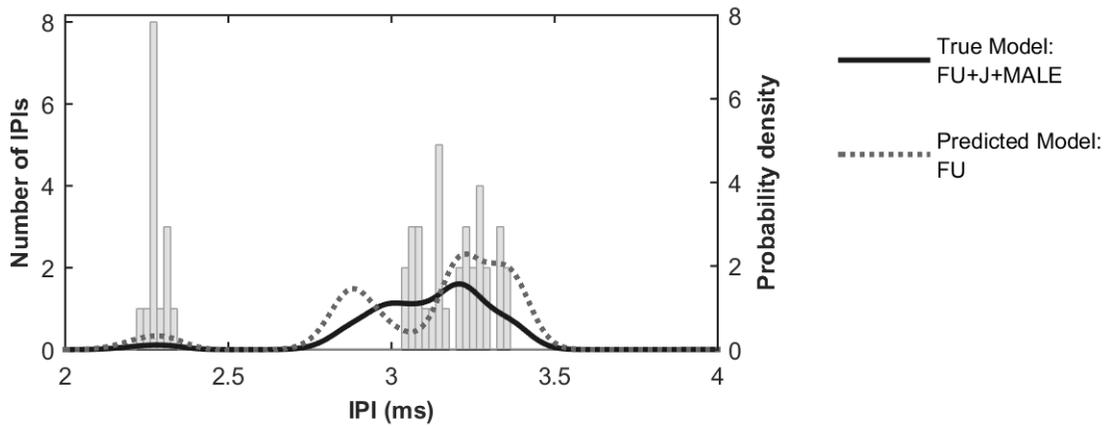
**Figure D.20** Results for Group 20: Mar. 20, unit S



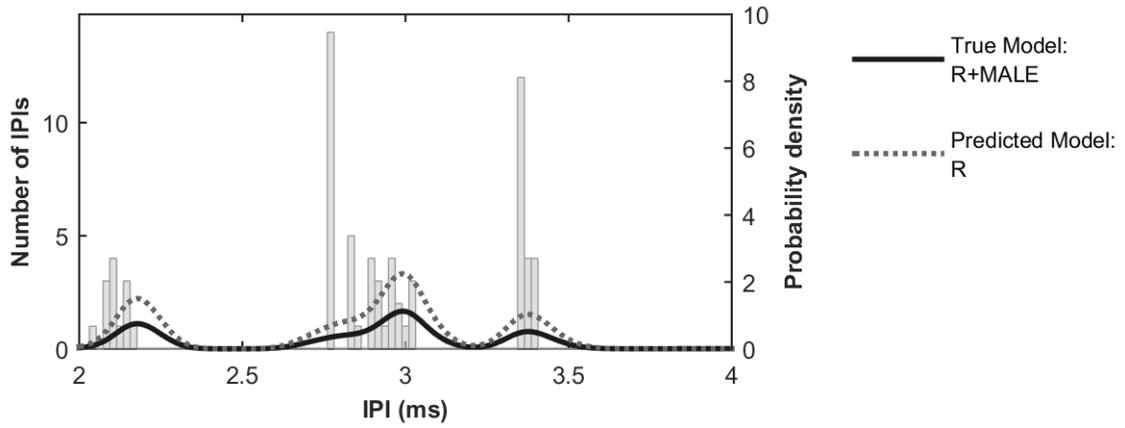
**Figure D.21** Results for Group 21: Mar. 20, units J + R + UNK + MALE



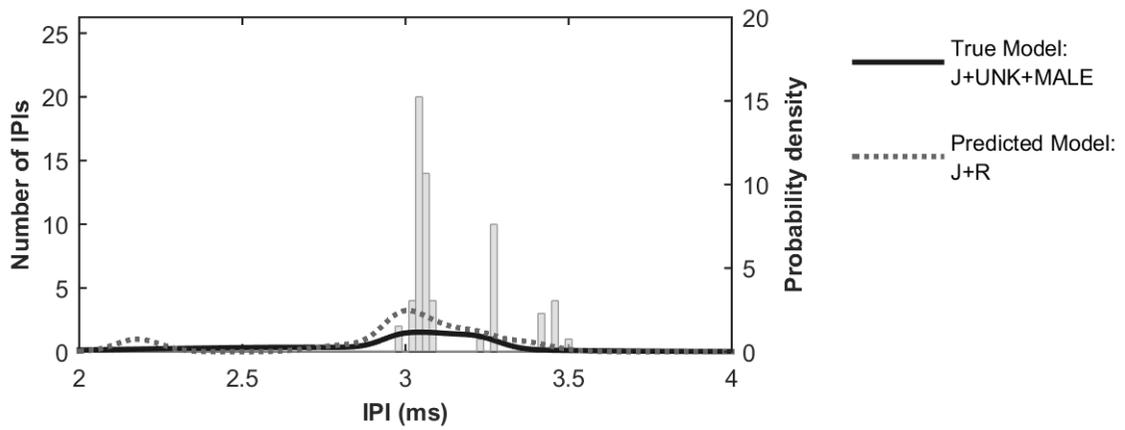
**Figure D.22** Results for Group 22: Mar. 21, unit R



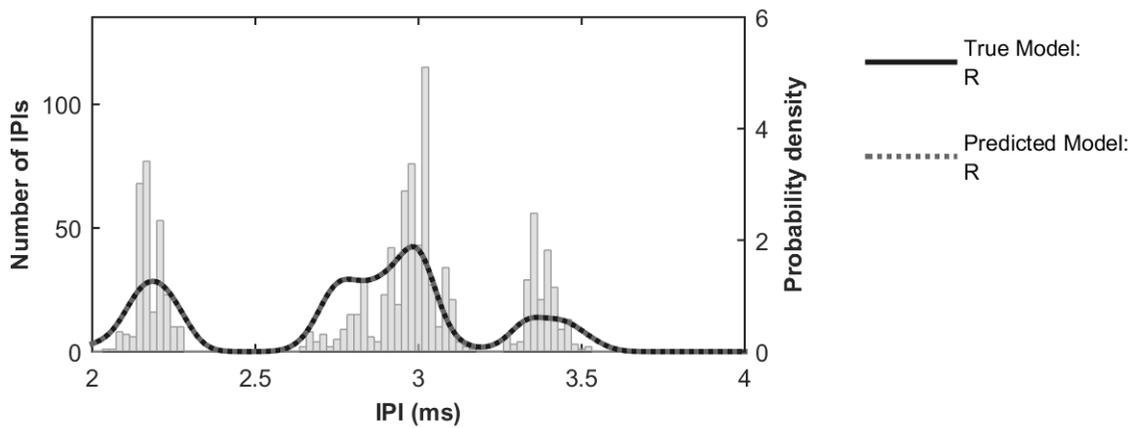
**Figure D.23** Results for Group 23: Mar. 21, units FU + J + MALE



**Figure D.24** Results for Group 24: Mar. 21, unit R + MALE



**Figure D.25** Results for Group 25: Mar. 21, unit J + UNK + MALE



**Figure D.26** Results for Group 26: Mar. 22, unit R

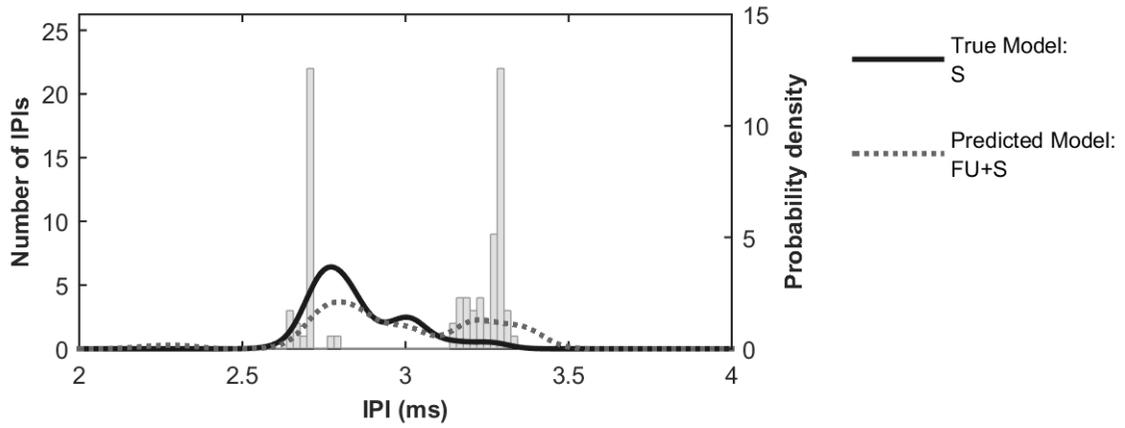


Figure D.27 Results for Group 27: Mar. 23, unit S

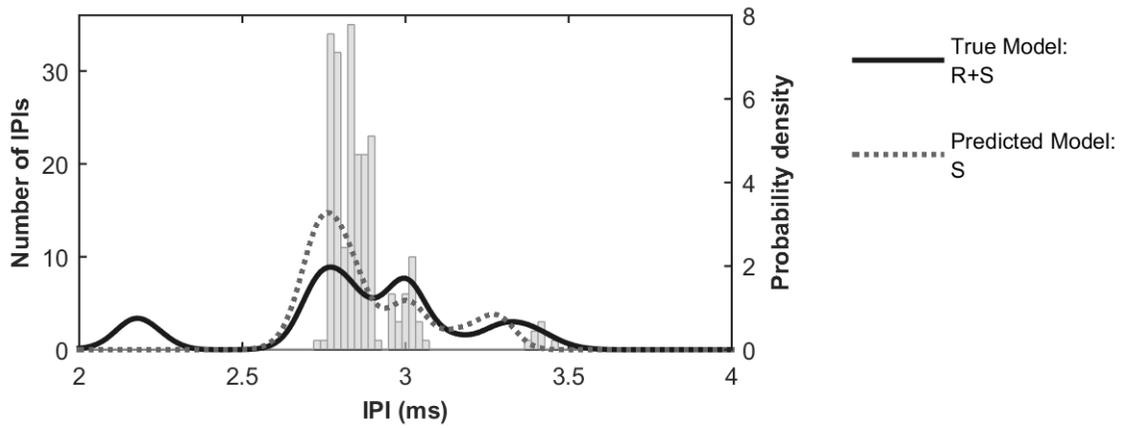


Figure D.28 Results for Group 28: Mar. 23, units R + S

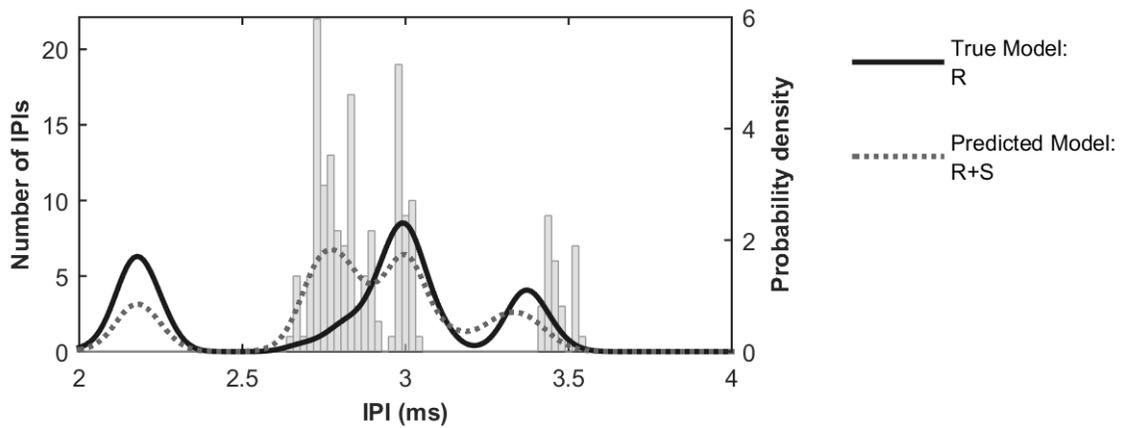
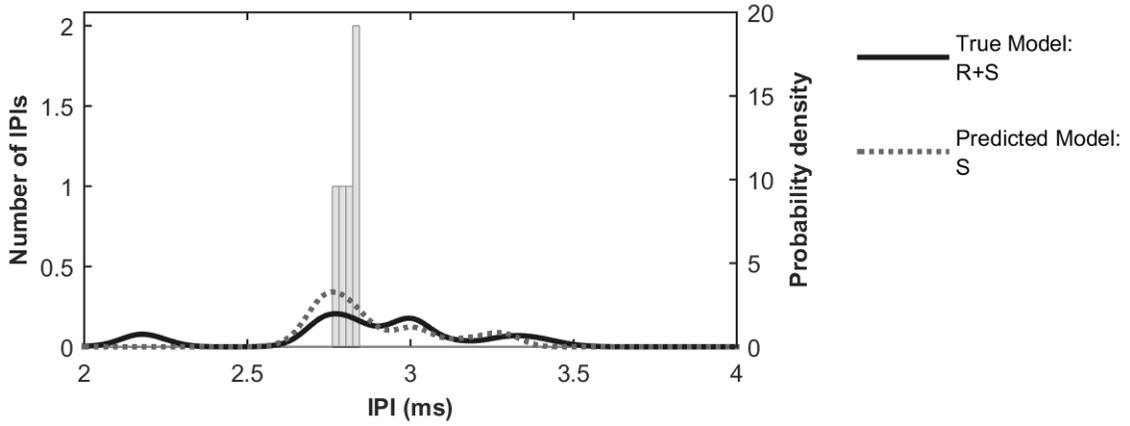
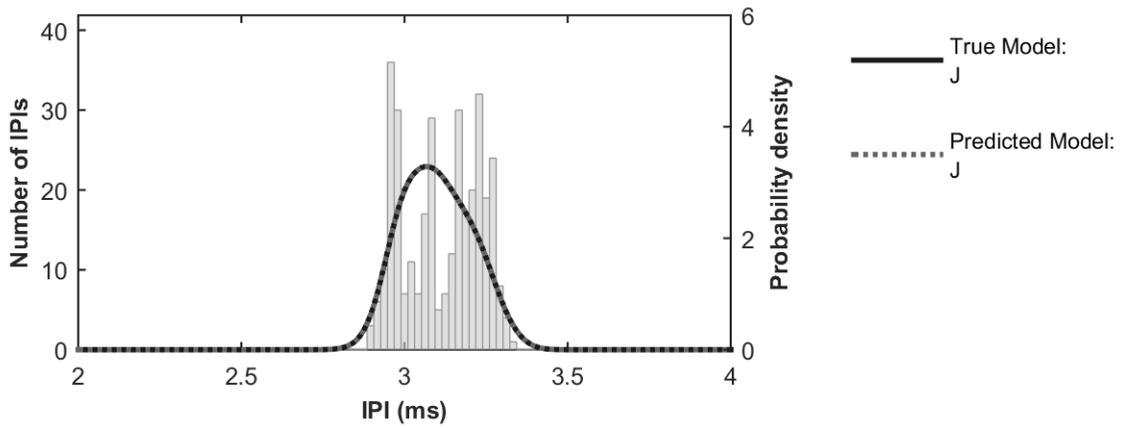


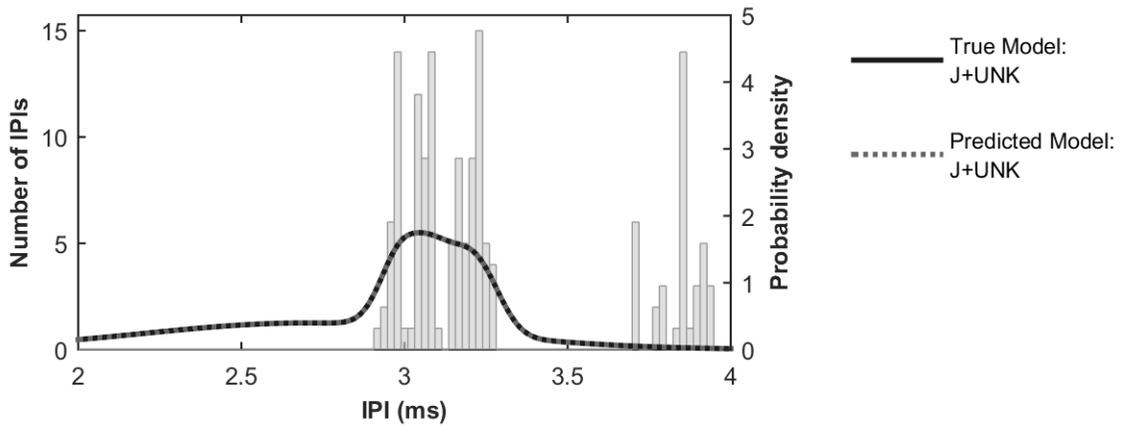
Figure D.29 Results for Group 29: Mar. 24, unit R



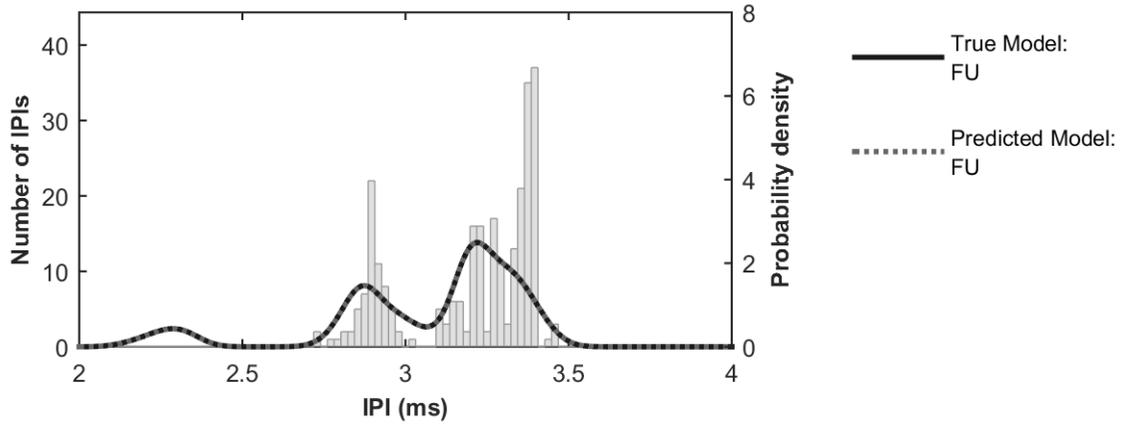
**Figure D.30** Results for Group 30: Mar. 24, units R + S



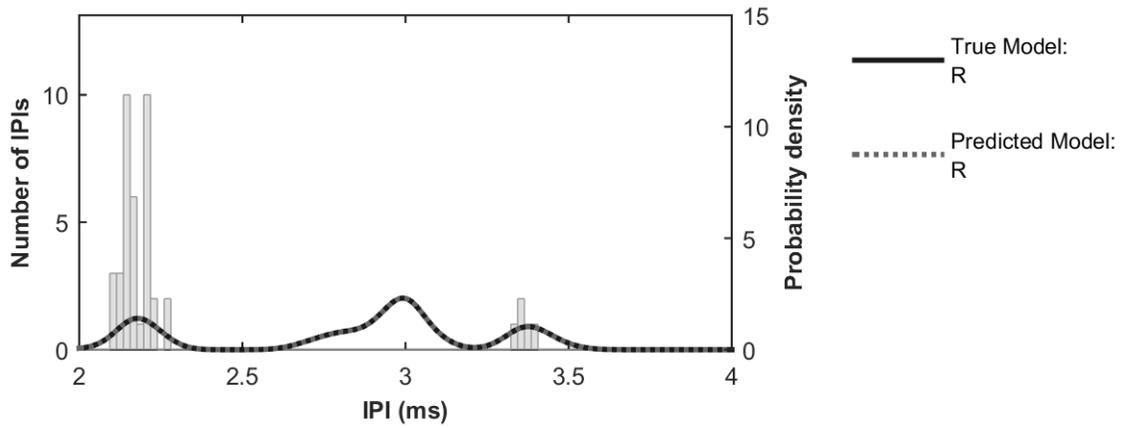
**Figure D.31** Results for Group 31: Mar. 25, unit J



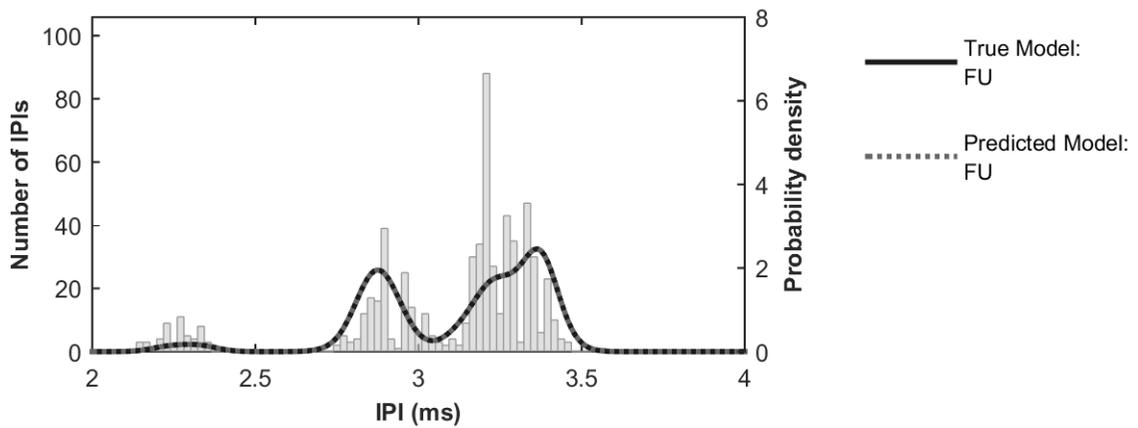
**Figure D.32** Results for Group 32: Mar. 25, unit J + UNK



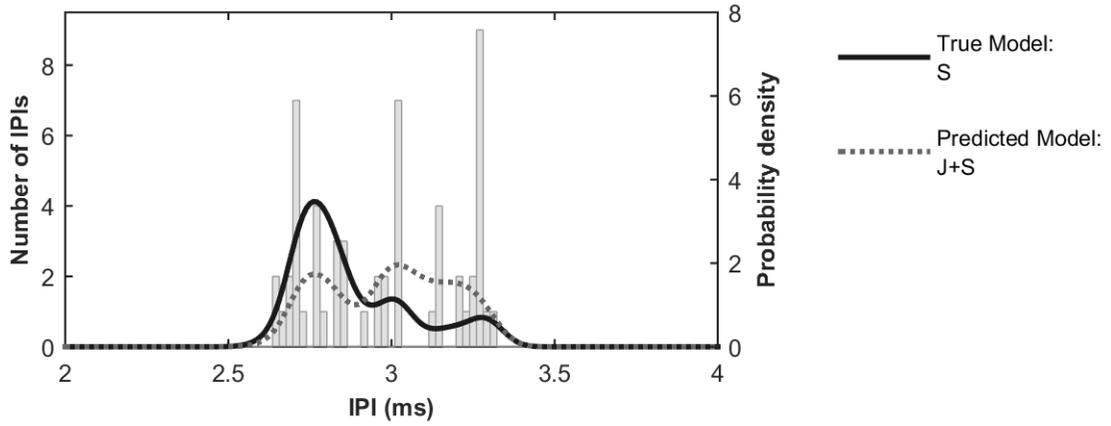
**Figure D.33** Results for Group 33: Mar. 26, unit FU



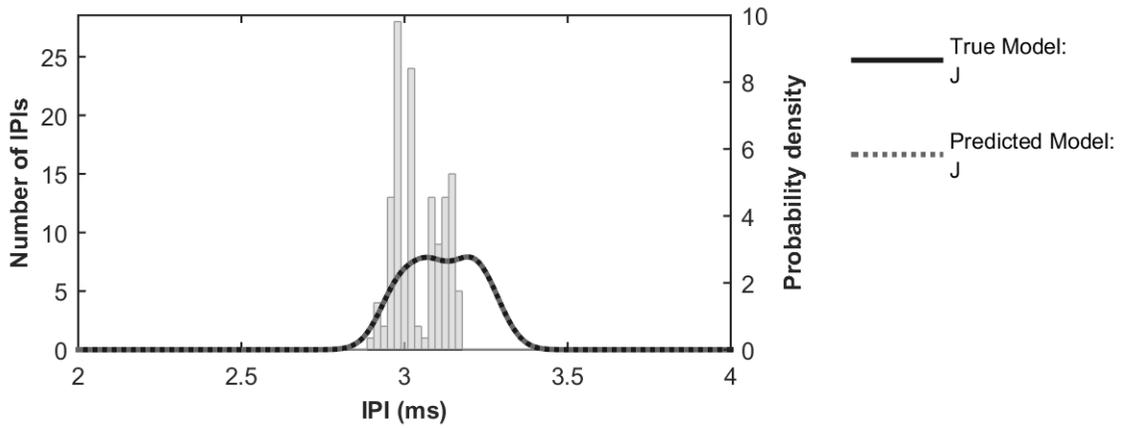
**Figure D.34** Results for Group 34: Mar. 30, unit R



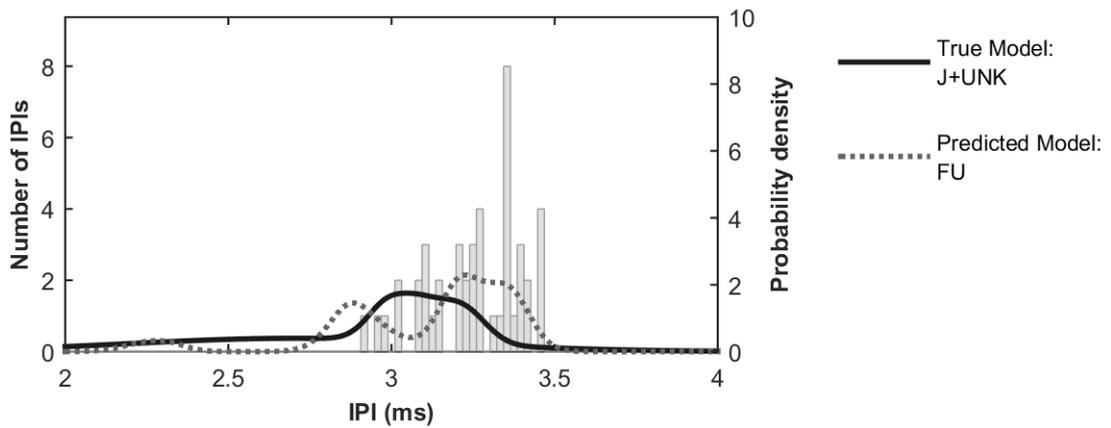
**Figure D.35** Results for Group 35: Mar. 31, unit FU



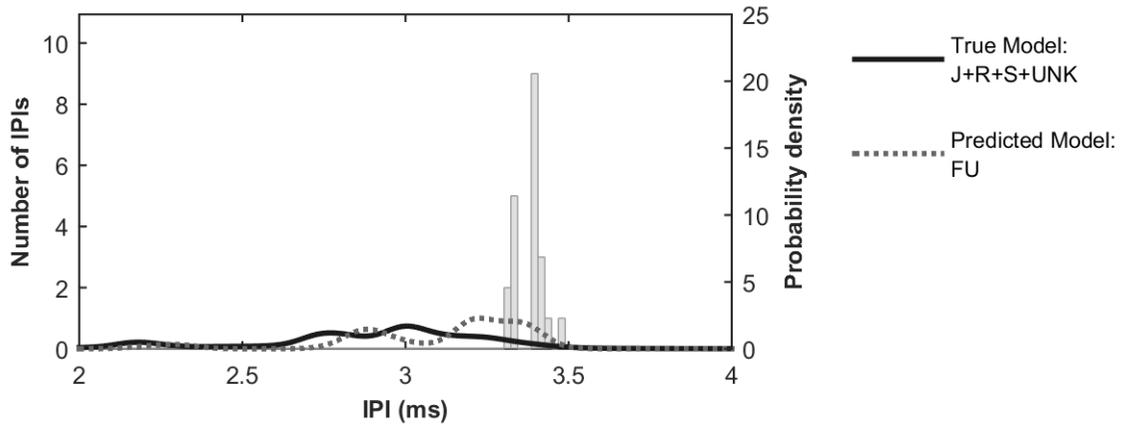
**Figure D.36** Results for Group 36: Apr. 1, unit S



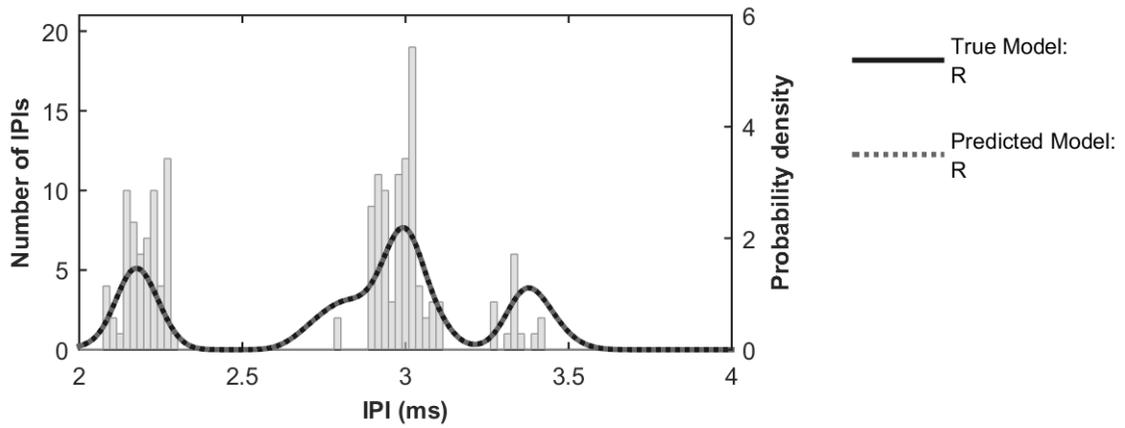
**Figure D.37** Results for Group 37: Apr. 2, unit J



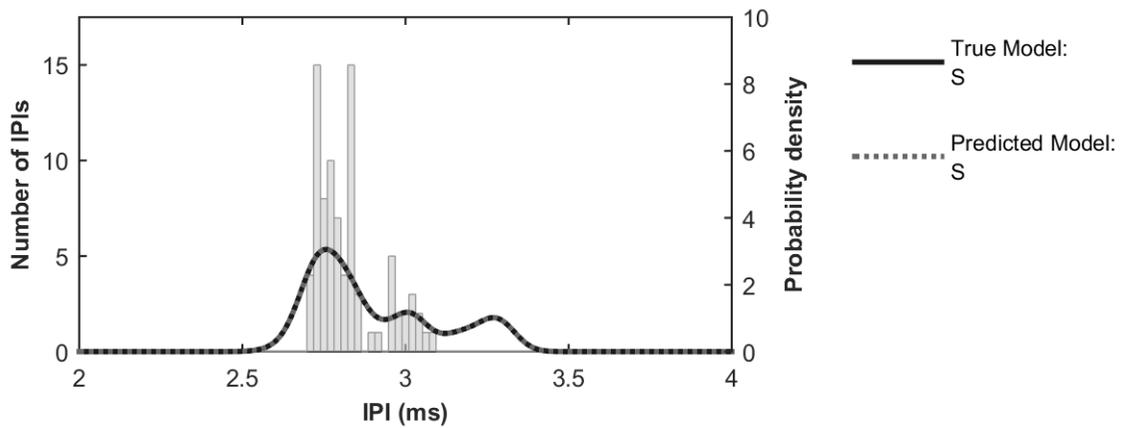
**Figure D.38** Results for Group 38: Apr. 2, unit J + UNK



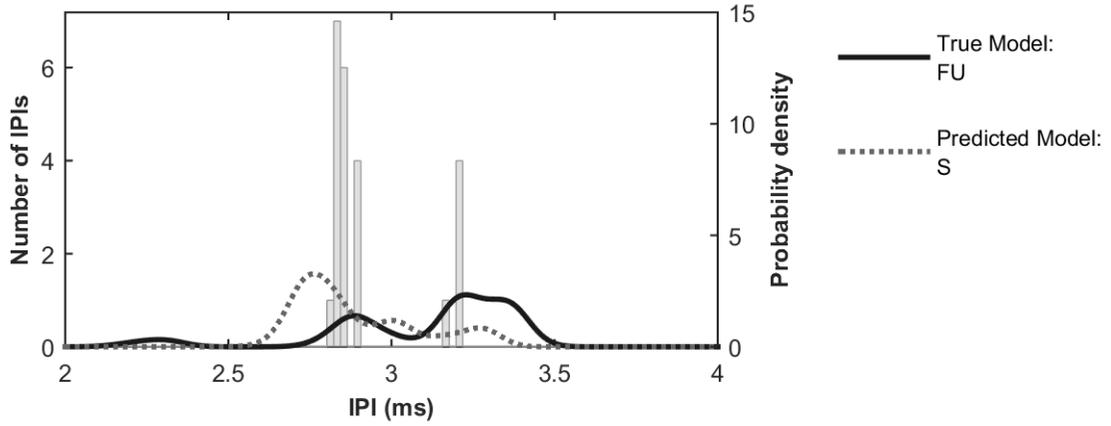
**Figure D.39** Results for Group 39: Apr. 2, units J + R + S + UNK



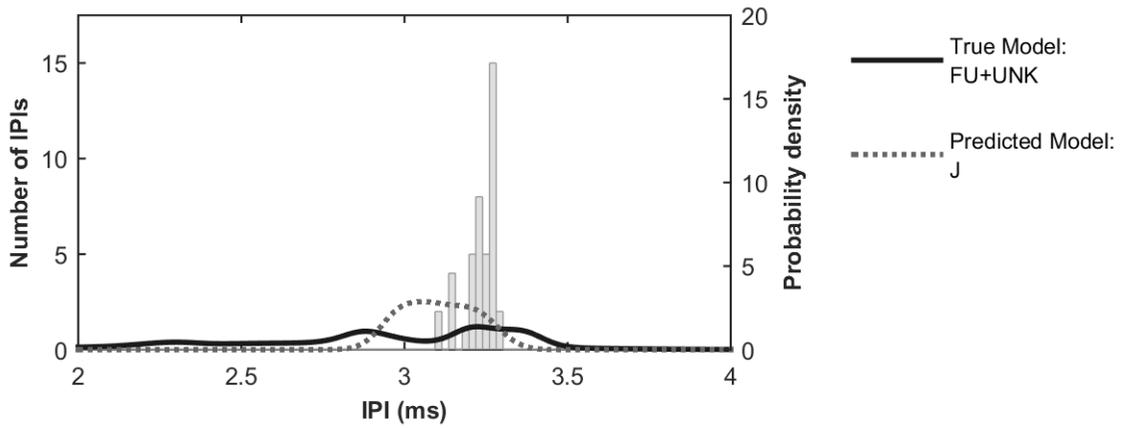
**Figure D.40** Results for Group 40: Apr. 6, unit R



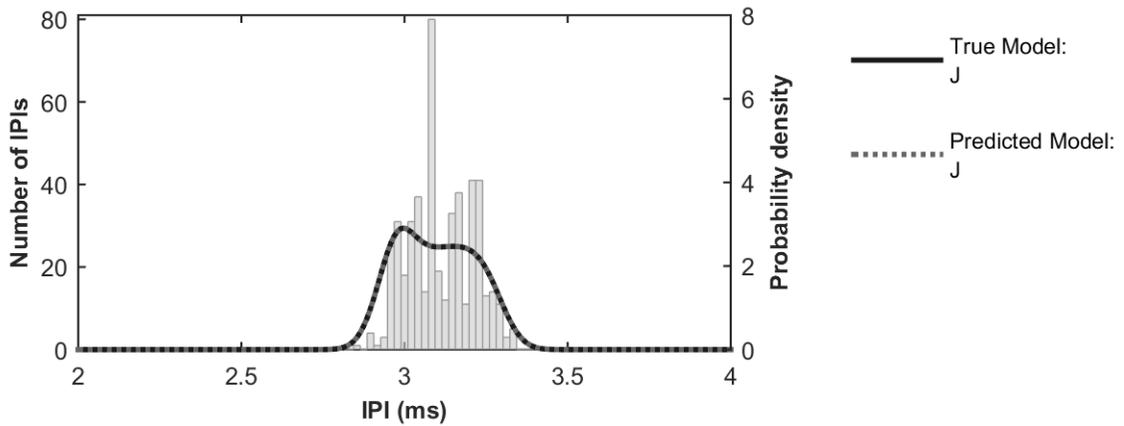
**Figure D.41** Results for Group 41: Apr. 7, unit S



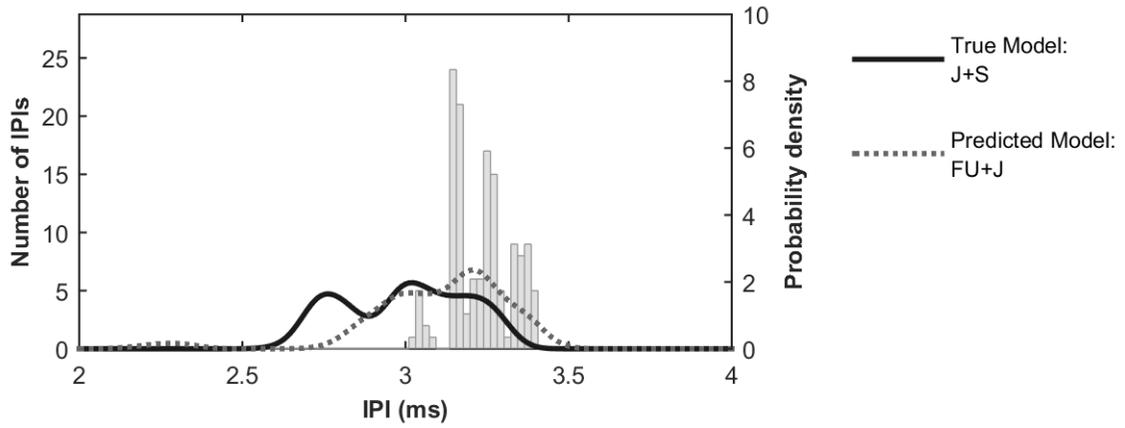
**Figure D.42** Results for Group 42: Apr. 8, unit FU



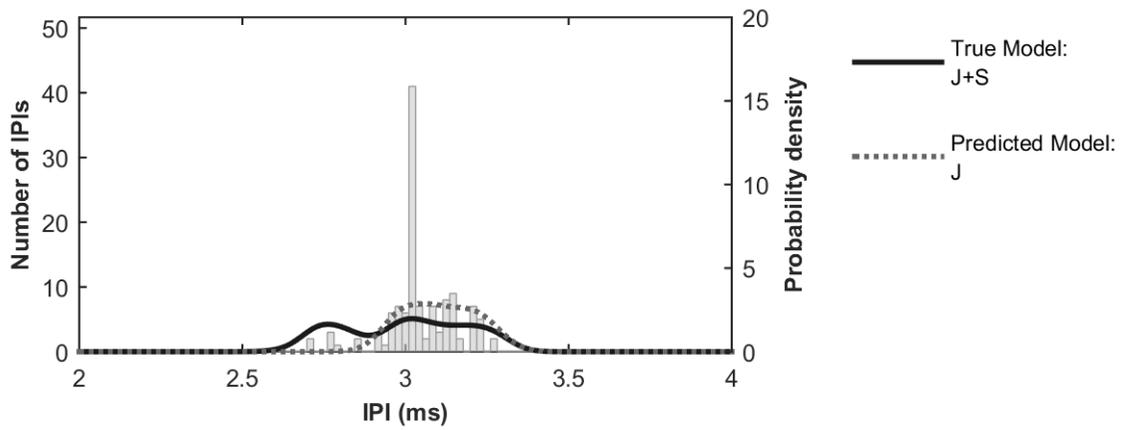
**Figure D.43** Results for Group 43: Apr. 8, unit FU + UNK



**Figure D.44** Results for Group 44: Apr. 10, unit J



**Figure D.45** Results for Group 45: Apr. 10, units J + S



**Figure D.46** Results for Group 46: Apr. 11, units J + S

## APPENDIX E – REFERENCES FOR APPENDICES

- Gero, S., Milligan, M., Rinaldi, C., Francis, P., Gordon, J., Carlson, C., Steffen, A., Tyack, P., Evans, P., and Whitehead, H. (2014). "Behavior and social structure of the sperm whales of Dominica, West Indies," *Mar. Mammal Sci.* **30**: 905–922.
- Goold, J. C. (1996). "Signal processing techniques for acoustic measurement of sperm whale body lengths," *J. Acoust. Soc. Am.* **100**, 3431–3441.
- Konrad, C. M. (2017). "Kinship in sperm whale society: Effects on association, alloparental care and vocalizations," M.Sc. thesis, Dalhousie University, Halifax, Nova Scotia, Canada.
- Marcoux, M., Whitehead, H., and Rendell, L. (2006). "Coda vocalizations recorded in breeding areas are almost entirely produced by mature female sperm whales (*Physeter macrocephalus*)," *Can. J. Zool.* **84**, 609–614.
- Miller, B. S. (2010). "Acoustically derived growth rates and three-dimensional localisation of sperm whales (*Physeter macrocephalus*) in Kaikoura, New Zealand," Ph.D. thesis, University of Otago, New Zealand.
- Object Management Group Inc. (2015). OMG Unified Modelling Language (OMG UML): Version 2.5 (Object Management Group Inc., Needham, Massachusetts, USA). Retrieved on October 31, 2016, from <http://www.omg.org/spec/UML/2.5/PDF>
- Rhineland, M. Q., and Dawson, S. M. (2004). "Measuring sperm whales from their clicks: Stability of interpulse intervals and validation that they indicate whale length," *J. Acoust. Soc. Am.* **115**, 1826–1831.
- Schulz, T. M., Whitehead, H., Gero, S., and Rendell, L. (2011). "Individual vocal production in a sperm whale (*Physeter macrocephalus*) social unit," *Mar. Mammal Sci.* **27**, 149–166.