UNIVERSITY OF CALIFORNIA, SAN DIEGO

Density estimation of delphinids using passive acoustics: A case study in the Gulf of Mexico

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Oceanography

by

Kaitlin Emily Frasier

Committee in charge:

John A. Hildebrand, Chair
Jay Barlow
Bruce D. Cornuelle
Yoav Freund
William S. Hodgkiss
Marie A. Roch

2015

The dissertation of Kaitlin Emily Frasier is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

_____

Chair

University of California, San Diego

2015

DEDICATION

In memory of Elsie Parker.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

Thanks first and foremost to my advisor, John Hildebrand, for his unfailing support, mentorship and humor, from beginning to end.

Thanks to Marie Roch, who pushed me to do my best work and taught me new things every day. Thanks also to my committee members Jay Barlow, Bill Hodgkiss, Yoav Freund and Bruce Cornuelle for their helpful and guidance.

Karlina Merkens was my inspiration, mentor, and cheerleader throughout my years at Scripps. Even after she'd left for other things, she never forgot me. I hope someday to be half as amazing as Karli. Thanks also to Rosa Leon and Scarla, the truest friends a girl could have.

Former students who paved the way before me were always willing to lend a helping hand: Liz Henderson, Melissa Soldevilla, Megan McKenna, Lisa Munger, and Ana Širović. My fellow grad students in the lab, Liz Vu, Anne Simonis, Martin Gassman, and Regina Guazzo were an excellent support group.

None of this work would have been possible without the many past and present lab members who built and fixed equipment, processed data, wrote permits, and so much more: Sean Wiggins, Simone Baumann-Pickering, Chris Garsha, Brent Hurley, Bruce Thayre, John Hurwitz, Ryan Griswold, Josh Jones, Erin O'Neill, Greg Campbell, Hannah Bassett, Amanda Debbich, Sara Kerosky, Lauren Roche, Rachel Gottlieb, Sarah Johnson, Elioth Gonzalez, Zoe Gentes, Leah Varga and Frank Chang.

I'm particularly grateful to Beve Kennedy for the many times she helped me organize travel, purchase equipment, attend meetings, and for her cheerful updates about treats in the kitchen. A big thank you also to the staff in the SIO grad office including Denise Darling, Josh Reeves, Satomi Saito, Cerise Maue, Adam Petersen, and Gilbert Bretado, and Almira Henson.

Many non-Scripps folks were indispensable along the way. The crew at NOAA

SEFSC, Lance Garrison, Tony Martinez, Jesse Wicker and Pascagoula lab's Keith Mullin helped us get out to sea and service our instruments many times. Len Thomas, Tiago Marques and Danielle Harris met with us weekly across eight time zones to dig us out of the many pitfalls of density estimation.

My parents, Richard and Robin Frasier, were always ready with emotional support, words of wisdom and unconditional love. I can't thank them enough for their patience and kindness. Thanks also to my sister and fellow scientist, Becca Frasier.

Above all, I want to thank my partner and best friend, Tyler Olmstead, for finding, supporting and putting up with me through this process. He deserves an honorary PhD and all the motorcycle racing he wants for his patience. Tyler helped me think through research challenges, provided technical support, and generally kept me moving forward.

This material is based upon work supported in large part by the NRDA partners, BP and the National Oceanic and Atmospheric Administration under Award Number 20105138. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of BP and/or any State or Federal Natural Resource Trustee. Financial support also came from C-IMAGE and C-IMAGE2 under a grant from BP/The Gulf of Mexico Research Initiative Contract #SA12-10/GoMRI-007 and from the Marine Mammal Commission with special acknowledgement of Tim Ragen. Additional Financial support was provided by the National Science Foundation's Graduate Research Fellowship Program, the University of California San Diego, and the Scripps Institution of Oceanography, Graduate Student Office.

VITA

2015	Ph.D., Oceanography Scripps Institution of Oceanography, University of California, San Diego
2012	M.S., Oceanography Scripps Institution of Oceanography, University of California, San Diego
2009-2015	Graduate Student Researcher Marine Physical Laboratory, University of California, San Diego
2006	B.S., Honors Biology & B. A. International Studies Oregon State University

PUBLICATIONS

Journal Articles

1. Frasier, K. E., Henderson, E. E., Bassett, H. R. & Roch, M. A. Automated identification and clustering of subunits within delphinid vocalizations. *Marine Mammal Science*, *In Review*.
2. Wiggins, S. M., Frasier, K. E., Henderson, E. E. & Hildebrand, J. A. (2013). Tracking dolphin whistles using an autonomous acoustic recorder array. *The Journal of the Acoustical Society of America*, 133(6), 3813-3818.
3. Lundsten, L., Schlining, K. L., Frasier, K., Johnson, S. B., Kuhnz, L. A., Harvey, J. B., & Vrijenhoek, R. C. (2010). Time-series analysis of six whale-fall communities in Monterey Canyon, California, USA. Deep Sea Research Part I: Oceanographic Research Papers, 57(12), 1573-1584.
4. Granek, E. F., & Frasier, K. (2007). The impacts of red mangrove (*Rhizophora mangle*) deforestation on zooplankton communities in Bocas del Toro, Panama. *Bulletin of Marine Science*, 80(3), 905-914.
5. Frasier, K. E. (2006). Impacts of mangrove conditions on zooplankton communities (Undergraduate dissertation). Oregon State University Honors College.

Conference Presentations

1. Frasier, K. E., Wiggins, S. M., Harris, D., Marques, T. A., Thomas, L., Merkens, K. P. & Hildebrand, J. A. (2015). Gulf of Mexico Oil Spill & Ecosystem Science Conference Special Session. Houston, TX. Oral Presentation.

2. Frasier, K. E., Wiggins, S. M.,Marques, T. A., Merkens, K. P., Harris, D., Thomas, L., & Hildebrand, J. A. (2014). Passive acoustic monitoring of sperm whales (*Physeter macrocephalus*) in the Gulf of Mexico: From detection to density estimation. Gulf of Mexico Oil Spill & Ecosystem Science Conference. Mobile, AL. Poster.

3. Frasier, K. E., Merkens, K. P., McDonald, M. A., Wiggins, S. M., Baumann-Pickering, S., Roch, M. A. & Hildebrand, J. A. (2013). Long-term monitoring of dolphins in the Gulf of Mexico. Gulf of Mexico Oil Spill & Ecosystem Science Conference. New Orleans, LA.Oral Presentation.

4. Frasier, K. E., Roch, M. A., Henderson E. E., Alldredge, A, Baumann-Pickering, S., Rankin S., Oleson, E. M., Kirker, A, & Hildebrand, J. A. (2013). A comparison of approaches for clustering tonal vocalizations of free ranging delphinids. Sixth International Workshop on Detection, Classification, Localization, and Density Estimation of Marine Mammals using Passive Acoustics. St. Andrews, Scotland. Oral Presentation.

5. Frasier, K. E., Merkens, K. P., Wiggins, S. M.,Marques, T. A., Harris, D., Thomas, L., McDonald, M. & Hildebrand, J. A. (2013). Effects of acoustic propagation on echolocation click detectability in the marine environment. Sixth International Workshop on Detection, Classification, Localization, and Density Estimation of Marine Mammals using Passive Acoustics. St. Andrews, Scotland. Poster.

6. Frasier, K. E., Soldevilla, M. S., McDonald, M. A., Merkens, K. P., Wiggins, S. M., Hildebrand, J. A., & Roch, M. A. (2011). Acoustic monitoring of dolphin populations in the Gulf of Mexico, San Diego, CA. Acoustical Society of America Meeting, Oral Presentation.

7. Frasier, K. E., Henderson, E. E., Roch, M. A., Bassett, H. R., & Hildebrand, J. A. (2011). Odontocete species identification by analysis of whistle component shape and sequence. Fifth International Workshop on Detection, Classification, Localization, and Density Estimation of Marine Mammals using Passive Acoustics. Mount Hood, OR. Oral Presentation.

8. Frasier, K. E., Wiggins, S. M., & Hildebrand, J. A. (2009). Acoustic tracking of whistling dolphins offshore of Southern California. Acoustical Society of America Meeting. Portland, OR. Oral Presentation.

ABSTRACT OF THE DISSERTATION

Density estimation of delphinids using passive acoustics:
A case study in the Gulf of Mexico

by

Kaitlin Emily Frasier

Doctor of Philosophy in Oceanography

University of California, San Diego, 2015

John A. Hildebrand, Chair

Dolphin populations are often considered an indicator of ocean health, yet they have historically been difficult to monitor. These animals live in remote, variable environments, and spend much of their time out of sight, below the sea surface. Passive acoustic monitoring (PAM) has become a viable method for gathering data on marine mammals. Instruments can be placed on the seafloor for long periods, recording animal sounds in the environment, regardless of oceanographic conditions, time of day and site accessibility.

The lingering challenge is in translating acoustic detections into quantitative

population density estimates. Density estimation techniques using PAM have been developed for and applied to marine mammals, but they have rarely been used for long term studies using single sensors, or applied to dolphins.

The Deepwater Horizon (DWH) oil spill event and its unknown impacts on offshore marine mammals provided an impetus for collecting a long term PAM dataset aimed at monitoring offshore marine mammals in the Gulf of Mexico (GOM). In this work I develop a framework for long-term, high resolution, quantitative monitoring of dolphin populations using this dataset, which stretches over three years at five sites.

Delphinid density estimation involves a series of steps: First, the probability of detecting delphinid cues using PAM is estimated, accounting for variability by site, species and season. Cues are then detected in the acoustic recordings, using methods consistent with the constraints of the detection probability estimation process, and classified to species. Finally, density estimates are produced by bringing together the detection counts, probabilities, and species-specific behavioral parameters.

We find distinct annual cycles in animal density in the northern GOM with peaks for most species in spring and summer months. Long-term increases in local densities were seen for Stenellid dolphins and pilot whales at sites east of the DWH site, which are not seen at sites to the south and east.

This work represents significant progress toward the goal of monitoring dolphin populations with minimal impact, high temporal resolution, and improved accuracy. These methods are broadly applicable to PAM efforts. As peripheral data are improved and expanded, estimates can be refined using this framework.

Chapter 1

Introduction

1.1 Passive acoustic monitoring of marine mammals

Underwater acoustic recording packages are widely used to passively monitor oceanic dolphins (delphinids) and other cetaceans by recording the sounds that they produce. These sounds can be interpreted as an indicator of animal presence. Passive acoustic monitoring (PAM) devices are useful for long-term, non-invasive monitoring and can operate continuously regardless of weather, visibility or site accessibility (Mellinger et al., 2007). These methods can be used as a complement to visual surveys to understand patterns of habitat use and population trends in marine mammals. Short term PAM technologies include sonobuoys and towed arrays, while long term monitoring typically involves seafloor-mounted recording devices or cabled arrays (e.g. Tiemann et al., 2004; Nosal and Frazer, 2007).

Technological limitations on non-permanent PAM devices constrain bandwidth and storage capacity (Diercks et al., 1973; Madsen and Wahlberg, 2007). Until recently, this limited the types of marine mammal signals that could be monitored over long time periods to the low frequency vocalizations of mysticetes (baleen whales) and

sperm whales (e.g. McDonald and Fox, 1999; Širović et al., 2004; Stafford et al., 2007). Sampling rates above 50 kHz (sometimes up to 500 kHz), and ample data storage are required to record the high-frequency vocalizations of smaller odontocetes (toothed whales, including dolphins). Recent technological advances have made this increasingly feasible (Mellinger and Heimlich, 2013). This study uses High-Frequency Acoustic Recording Packages (HARPs), which have the sampling capabilities and storage capacity to continuously record high frequency delphinid echolocation clicks for extended periods of time (many months to more than a year) (Wiggins and Hildebrand, 2007).

While increasing the number of questions that can be addressed with passive acoustics, advances in underwater recording technologies have also led to a dramatic increase in the quantity of acoustic data being collected. Historically marine mammal vocalizations are manually detected and classified by trained analysts. Increasingly, however, acoustic data sets are becoming too large for pure manual analysis to be practical. For instance, a standard 6-month HARP deployment currently yields 2 terabytes of data, which might require multiple weeks to analyze manually for presence of a single species, and much longer for more detailed and/or multispecies analyses. A major focus in the PAM community is therefore on the development of efficient, robust, automated tools capable of detection, classification and localization of recorded signals (Mellinger and Heimlich, 2013).

In order to monitor marine mammal populations quantitatively using PAM devices, detections of the species of interest need to be converted into a measure of local density or abundance. Density estimation is an active area of research in PAM, because many factors influence the relationship between animal densities, and the number of signals or cues detected. In order to quantitatively monitor populations, to compare monitoring locations, or to estimate trends over time, these variables need to be accounted for.

Factors that influence the relationship between detection events and actual animal

densities include cue rate, animal behavior, acoustic characteristics of the propagation in the environment, and monitoring effort (Helble et al., 2013a; Marques et al., 2009; McDonald and Fox, 1999). Other factors specific to the processes of detection and classification must also be considered, including false detections, missed detections, and detection range limitations (Buckland et al., 2001, 2007). As a result, quantitative passive acoustic population monitoring requires a combination of advanced marine technology, robust data analysis, and specialized statistical methods.

1.2 Echolocation Clicks: Structure and Function

Delphinids produce two main types of sound: (1) Echolocation clicks, which are short duration, broadband, directional pulses, and (2) whistles, which are longer duration, frequency-modulated, omnidirectional tonal signals. Pulsed calls, consisting of many clicks in rapid succession, are also produced.

Echolocation clicks are produced by all known delphinids. These signals are produced by a structure adjacent to the blowhole known as the monkeylips (Cranford et al., 1996; Madsen et al., 2003). The signal passes through a fatty body in the animal's head, known as the melon, which focuses the sound into a narrow beam (Aroyan et al., 1992; Norris, 1968). This forward-directed beam of sound leaves the animal, bounces off a target, and returns to the animal, which obtains information about the target from the reflected signal (Au, 1993).

The directional nature of echolocation clicks gives them unique acoustic properties (Au, 1993). When a dolphin orients its beam of sound directly at a target sensor, the received signal amplitude at the sensor is high. In this scenario, a click is termed "on-axis" with respect to the sensor. However, if the same animal, at the same location points its transmitting beam elsewhere, the received click amplitude at the

sensor will be low. In this case, the echolocation click is "off-axis" with respect to the sensor. Specifics of delphinid transmission beams, including beam width, source level, and amplitude difference between on and off-axis clicks, vary between species (Au and Benoit-Bird, 2003; Au et al., 2009; Roch et al., 2011; Soldevilla et al., 2008) and may be related to animal morphology, behavior, prey preferences and/or habitat (Akamatsu et al., 2007; Henderson et al., 2011; Jensen et al., 2009, 2013; Madsen et al., 2004a).

Echolocation click parameters, and directionality in particular, impact density estimation efforts because signal detectabilty depends on animal orientation relative to the recording device (Hildebrand et al., 2015; Küsel et al., 2011; Marques et al., 2009). At close range, an animal's echolocation clicks may be detectable even if it's beam is off-axis relative to the sensor. As range increases, however, only on-axis clicks will be detectable. Further, given a low amplitude echolocation click, it is currently not possible to determine whether the signal was produced close to the sensor but off-axis, or at a distance but on-axis. This leads to added complexity in click-based density estimation efforts, compared with estimates based on omnidirectional signals.

1.3 Delphinids in the Gulf of Mexico

The Gulf of Mexico (GOM) is home to at least fourteen species of delphinids (Fritts et al., 1983; Mullin and Fulling, 2004; Mullin, 2007). Twelve of these are primarily pelagic, living entirely in deep waters beyond the continental shelf, and are the focus of the research presented here.

Listed in order of expected abundance these species are:

1. Pantropical spotted dolphin (*Stenella attenuata*),

2. Spinner dolphin (*Stenella longirostris*),

3. Pilot whale (*Globicephala macrorhynchus*),
4. Melon-headed whale (*Peponocephala electra*),
5. Risso's dolphin (*Grampus griseus*),
6. Striped dolphin (*Stenella coeruleoalba*),
7. False killer whale (*Pseudorca crassidens*),
8. Rough-toothed dolphin (*Steno bredanensis*),
9. Clymene dolphin (*Stenella clymene*),
10. Pygmy killer whale (*Feresa attenuata*),
11. Killer whale (*Orcinus Orca*).
12. Fraser's dolphin (*Lagenodelphis hosei*),

These pelagic species are typically found beyond the shelf break in waters more than 200 meters deep (Mullin and Fulling, 2004). Abundances are computed by aggregating data from periodic summer visual surveys by NMFS NOAA and can vary considerably between estimates, due to survey limitations. Pantropical spotted dolphin is consistently reported as the most abundant delphinid in the Northern GOM, with most recent population size estimates around 50,000 animals in 2009, CV = 0.27 (but as many as 90,000, CV = 0.16, estimated in 2001). Large fluctuations in population size estimates ($\hat{E}$) are also seen for species like Clymene dolphin ($\hat{E}$ = 17,000, CV = 0.65, in 2001 vs. $\hat{E}$ = 129, CV = 1, in 2009), and Fraser's dolphin ($\hat{E}$ = 723, CV = 0.70, in 2001 vs. $\hat{E}$ = 0 in 2009 ; Mullin and Fulling 2004; Mullin 2007). Population estimates are more consistent for other species including, melon-headed whale and Risso's dolphin at roughly 2000-3000 animals, and killer whales between 20 and 100 animals (Mullin and

Fulling, 2004; Mullin, 2007). This variability highlights a need for alternative strategies to complement summer visual survey efforts in order to improve population estimates and to resolve long term trends.

Pelagic delphinid home ranges in the GOM overlap spatially, but several species' distributions are linked to mobile oceanographic features including sea surface height anomalies, eddies and temperature gradients (Davis et al., 2002). Pilot whales and Risso's dolphins are deep divers (Baird, 2002; Heide-Jørgensen et al., 2002; Wells et al., 2009), feeding primarily on squid (Clarke and Pascoe, 1985; Würtz et al., 1992). Both species are large, with body lengths up to four meters in Risso's dolphin and up to six meters for adult pilot whales (Baird, 2002; Olson and Reilly, 2002). Risso's dolphins are often found along strong, productive thermal gradients, and both species are associated with steep bathymetry where vertical mixing supports elevated primary productivity and prey aggregation (Davis et al., 1998; Baumgartner, 1997).

Most of the pelagic dolphins of the genus *Stenella* (Stenellid dolphins include pantropical, Clymene, spinner and striped dolphins and the more coastal Atlantic spotted dolphin) are found beyond the continental shelf (Davis et al., 2002). Although they live in deep water, these relatively small animals (approx. 2m in length; Archer and Perrin 1999; Perrin and Hohn 1994), are not considered deep divers, and use only the top few hundred meters of the water column (Hastie et al., 2006). Evidence suggests that offshore Stenellid dolphins feed nocturnally on myctophid fish and small squid that migrate toward the sea surface as part of the deep scattering layer (Baird et al., 2001; Fitch and Brownell Jr, 1968; Miyazaki et al., 1973).

Pelagic Stenellid dolphins are often associated with cyclonic cold core eddies, which aggregate their prey (Davis et al., 2002). However some habitat differences have been reported between the various Stenellid species. Pantropical spotted dolphins are generally found in warmer waters (Perrin, 2001), and have been shown to vary their

range in order to follow preferred habitat as it shifts geographically over seasons or years (Fiedler and Reilly, 1994; Reilly, 1990).

Fraser's dolphin is similar in size to the Stenellid dolphins, and associated with deep tropical and subtropical oceans worldwide (Perrin and Hohn, 1994). This is thought to be a deep diving species that may be more oceanic, and prey on slightly larger items than the Stenellid dolphins (Dolar, 2002).

Rough-toothed dolphins have been reported in both shelf and slope waters in the GOM (Fulling et al., 2003). These unique dolphins forage on surface-dwelling prey including flying fish, mahimahi and needlefish (Baird et al., 2008; Pitman and Stinchcomb, 2002).

The remaining pelagic GOM delphinid species, including melon-headed whale, false killer whale and pygmy killer whale, are infrequently sighted or known mainly from stranded specimens. Little is known about their distribution and foraging habits in the region (Jefferson and Schiro, 1997). This may be partially due to the difficulty in distinguishing between the blackfish species during visual surveys. A number of these have only been recognized in the region within the last 30 years (Barron and Jefferson, 1993; Mullin et al., 1994; O'Sullivan and Mullin, 1997).

Mixed species groups of delphinids are common in the GOM. Melon-headed whales have been seen mingling with pods of Fraser's dolphin (Mullin et al., 1994), and schools containing multiple species of pelagic Stenellid dolphins have also been reported (Fertl et al., 2003).

Only two species of delphinid are regularly found in shelf and coastal habitats in the GOM: Bottlenose dolphins (*Tursiops truncatus*) and Atlantic spotted dolphins (*Stenella frontalis*) (Fulling et al., 2003; Jefferson and Schiro, 1997). Atlantic spotted and striped dolphin are typically found in cooler waters with mild sea surface temperature gradients (Davis et al., 1998). Some have suggested that the genus *Stenella* may be

non-monophyletic (Perrin, 2001), and that Atlantic spotted dolphins may be more closely related to bottlenose dolphins than to other Stenellid dolphin species (LeDuc et al., 1999) . Atlantic spotted dolphins and bottlenose dolphins are often found in mixed groups (Herzing and Johnson, 1997).

Bottlenose dolphins typically outnumber Atlantic spotted dolphins in areas shallower than 20 meters (Griffin and Griffin, 2003), and can be found in extremely shallow bays and estuaries, where they appear to be resident (Wilson et al., 2013). Mean dive durations are very short for these animals (less than a minute), and foraging seems to peak in the morning and evening, rather than at night (Mate et al., 1995). Both bottlenose and Atlantic spotted dolphins are found on the continental shelf and out to the shelf break (Davis et al., 2002). Their fine-scale distributions are thought to be linked to oceanographic conditions, but little is known about seasonal distribution trends for these shallow water species (Griffin and Griffin, 2004).

1.4 The Gulf of Mexico Ecosystem

The GOM covers an area of roughly 1.5 million square kilometers, nearly 60% of which consists of shallow zones (less than 180 m deep) above a broad continental shelf (1.1; Gore, 1992). The continental shelf is more than 80 km wide at its narrowest point, and extends up to 250 km from the coast in some places (Carsey, 1950). Beyond the shelf, the continental slope represents about 20% of the GOM's surface area, and abyssal planes up to four kilometers deep make up the remaining 20% (Gore, 1992).

This unique system is caught between the North American continent and the Caribbean Sea, both of which heavily influence its oceanographic features. From the south, a warm flow known as the Loop Current travels up from the Caribbean, passing between Cuba and the Yucatan peninsula and into the GOM, where it loops around before

exiting past the tip of Florida to join the North Atlantic Gulf Stream (Murphy et al., 1999).

Anticyclonic, oligotrophic warm core rings up to 300km wide periodically detach from the Loop Current and move westward until they dissipate along the coast of Mexico and southern Texas (Biggs, 1992). Slightly smaller cyclonic eddies known as cold cores are generated at the edges of the warm cores. These cold cores are nutrient rich and productive, in contrast with their warm counterparts (Biggs and Müller-Karger, 1994). The Loop Current and the cores associated with it drive upwelling in along their edges. This is thought to be a major and dynamic source of nutrient inputs to offshore waters (Wiseman et al., 1999). Planktonic biomass is generally higher in cyclonic cold core rings which drive upwelling at their centers, and lower in anticyclonic warm core eddies (Wormuth et al., 2000). This in turn affects distribution of marine mammals looking for prey (Biggs et al., 2000). A prominent deep scattering layer lies around 450-550 meters during the day, but may shoal in areas around the Mississippi river outflow where high particulate levels reduce light penetration (Kaltenberg et al., 2007).

The GOM is routinely hit by tropical storms and hurricanes in summer months, and battered by northern cold fronts throughout the winter. Sea surface temperatures are markedly higher in late spring and summer than in fall and winter (Chang and Oey, 2010; Etter, 1983). This shift, along with seasonal weather patterns leads to seasonal fluctuations in the mixed layer depth, which varies between 20m in summer, to 125m in winter (Müller-Karger et al., 1991, 2014).

On its northern edge, the GOM receives an average of 580 km^3 of fresh water annually from the Mississippi River (Meade, 1996). The magnitude of the flow varies seasonally, with the largest inputs in spring. These waters drain from a basin that includes more than 40% of the continental US (Berner and Berner, 1987). The Mississippi river outflow brings levels of dissolved nitrogen, phosphates, and organic particulates to the

nutrient-depleted subtropical waters (Dagg and Breed, 2003). The combination of the nutrient-rich Mississippi outflow and strong stratification of the water column in summer months causes eutrophication, leading to the annual formation of the world's second largest hypoxic region along the Northern GOM continental shelf during summer months (Rabalais et al., 2002). Sediment cores show that this hypoxic zone first appeared around the turn of the 20th century, and increased dramatically around the 1950s (Turner and Rabalais, 1994), as nutrient inputs from human activities increased.

1.5 Human impacts in the Gulf of Mexico

The Northern GOM is a chronically impacted system that has been the site of extensive development and exploitation for over a century, with pressures increasing in recent decades (Birkett and Rapport, 1999; Austin et al., 2004). Shoreline development has contributed to extensive wetland and coastal ecosystem loss, leading to increased erosion, pollution, and lack of nurseries for nekton, including larval fish (Peterson and Lowe, 2009; Turner, 1990, 1997). Pollutant discharge and nutrient loading from the Mississippi river outflow are also ongoing sources of environmental stress (Birkett and Rapport, 1999).

Despite environmental heavy impacts, the GOM is home to productive commercial and recreational fisheries, with over 60 managed stocks. A number are identified as over-fished, and many more are of unknown status due to a lack of historical records (Karnauskas et al., 2013). Numerous studies have found evidence of bioaccumulation of heavy metals, petroleum residues and Polychlorinated biphenyls (PCBs) in fish and invertebrates in the region (Cai et al., 2007; Neff et al., 2011; Peterson et al., 1996).

The oil and gas industry is probably the most well known source of human impact

in the GOM. Early extraction efforts began onshore, and subsequently moved through wetlands and into nearshore sites by the 1930s (Austin et al., 2004). Today Deepwater (water depth >1000m), and Ultra Deepwater (water depth >5000m) drilling are common. Well density exceeds three rigs per km^2 in some areas along the Texas shelf, while deepwater densities were between 0.23 and one wells per km^2 in 2009 (Nicot, 2009). Drill sites are routinely plugged and abandoned (Kaiser and Dodson, 2008), causing leakage concerns, but also reportedly functioning as an artificial reef system (Stanley and Wilson, 2000).

Oil exploration using airguns is a major source of ambient noise in the GOM, and this maybe particularly problematic for cetaceans populations. An analysis of global seismic surveys activity from 1994 to 2005 indicated that the GOM had an average of 25 active oil explorations per month, the highest of any region globally (Hildebrand, 2009). Additional noise sources include commercial shipping, fishing, and recreational vessel activity.

1.5.1 The Deepwater Horizon oil spill

The Deepwater Horizon (DWH) oil spill began with a well explosion on April 20th, 2010, followed by oil flow into the GOM for approximately three months. This was the largest oil spill in US waters to date. Total flow has been estimated at seven hundred thousand m^3 of crude oil (Camilli et al., 2012; Crone and Tolstoy, 2010), with over two million gallons of dispersant chemicals applied (1.4 million gallons at surface, 0.7 million gallons at depth, Kujawinski et al. 2011).

This spill was unique in that it occurred at the seafloor, in deep water (Reddy et al., 2012). A large surface oil slick formed, however much of the oil released never rose to the sea surface, remaining instead in a large and diffuse plume at depth (Kessler et al., 2011). The deep plume sat at a depth of approximately 1100 m and had a southward

footprint which differed from that of the surface slick (Camilli et al., 2010; Hildebrand et al., 2012). Dispersants were applied both at the surface, and at the well outflow, which may have contributed to the formation of the deep plume (Kujawinski et al., 2011). Over time, the oil in this plume appeared to settle to the seafloor over an area of approximately 3,200 km^2 around the well site (Spier et al., 2013; Valentine et al., 2014).

Surface oil was deposited in sediments by a large marine snow event (Passow et al., 2012), in which mats of oil, particulate matter and phytoplankton sank to the seafloor. Oil-derived compounds were released into the atmosphere by evaporation and managed surface burns (Ryerson et al., 2011). DWH oil was also documented along 1,773 km of shoreline on the Northern Gulf coast (Michel et al., 2013).

Long term effects of the DWH oil spill on pelagic GOM biota have been difficult to quantify due to the size and depth of affected areas, a lack of baseline data, and the GOM's complex history of chronic impacts. Of particular interest from a marine mammal perspective is an unusual mortality event involving primarily coastal bottlenose dolphins, which began in February 2010 and was ongoing as of publication of this manuscript (Litz et al., 2014). Evidence suggests that the magnitude and duration of this event may have been linked at least in part to the DWH oil spill (Carmichael et al., 2012; Venn-Watson et al., 2015). Pelagic species have been a minority of reported deaths, but are likely under-reported (Williams et al., 2011).

1.6 This study

The goal of this study is to develop a framework for long-term, high resolution, quantitative monitoring of dolphin populations in remote locations. This is accomplished by estimating delphinid densities at sites of interest on a weekly timescale using PAM devices. The DWH oil spill event and its unknown impacts on offshore marine mammals

provided the impetus for collecting the long term passive acoustic dataset used here as a case study.

The acoustic dataset underlying this study was obtained from HARPs deployed at five sites of interest in the GOM: Two shallow sites on the continental shelf, and three deep sites on the continental slope (Figure 1.1 and Table 1.1). Sites were chosen to capture variation across a range of marine habitats. Instruments were serviced with new batteries and data storage on a 2-5 month cycle. These instruments recorded continuously at 200kHz throughout each deployment. Deployments occurred between between May 2010, and September 2013 (Table 1.2).

HARPs are bottom-mounted acoustic recorders containing a hydrophone, data logger, battery, power supply, ballast weights, acoustic release system, and flotation (Wiggins and Hildebrand, 2007). The hydrophone is tethered to the instrument and buoyed approximately 10 m above the seafloor. It consists of a high frequency stage and a low frequency stage. The high frequency stage consists of a spherical, omni-directional transducer (ITC-1042, www.channeltechgroup.com), with an approximately flat (+/- 2 dB) sensitivity response of about -200 dB re 1Vrms / ÂţPa between 1 Hz and 100 kHz. The a low frequency stage is made up of six cylindrical transducers (Benthos AQ1, www.teledynebenthos.com), with a flat response (+/- 1.5 dB) of about -187 dB re 1Vrms / ÂţPa from 1 Hz to 10 kHz. . All acoustic data were converted to sound pressure levels based on hydrophone calibrations performed at Scripps Institution of Oceanography and at the U.S. NavyâĂŹs Transducer Evaluation Center facility in San Diego, California.

In order to arrive at delphinid density estimates, a number of challenges need to be addressed. Each subsequent chapter of this document addresses one of these challenges. First, the probability of detecting delphinids at monitored sites must be estimated (Chapter 2) and verified (Chapter 3). This probability may vary by site, species and season. Next, cues need to be detected in the extensive acoustic recordings using a

detection method consistent with the constraints of the detection probability estimation process (Chapter 4). Once detections are made, they need to be categorized by type (Chapter 5) and attributed to particular species (Chapter 6). Finally, density estimates are achieved by bringing together the detection counts, detection probabilities, and species specific behavioral parameters, to generate site specific estimates of animals per unit area over time (Chapter 7).

The complexity of this undertaking exceeds what can be resolved here. Lingering refinements needed to solidify density estimates are outlined in the final chapter. A lack of pre-spill data limits the inferences that can be drawn with respect to the effects of the DWH event. Nonetheless, this work represents significant progress toward the goal of monitoring dolphin populations with minimal impact, high temporal resolution, and improved accuracy. These methods are broadly applicable to monitoring scenarios in which permanent sensors are unavailable and human effort is constrained. As peripheral data are improved and expanded, estimates can be refined using the framework developed here.

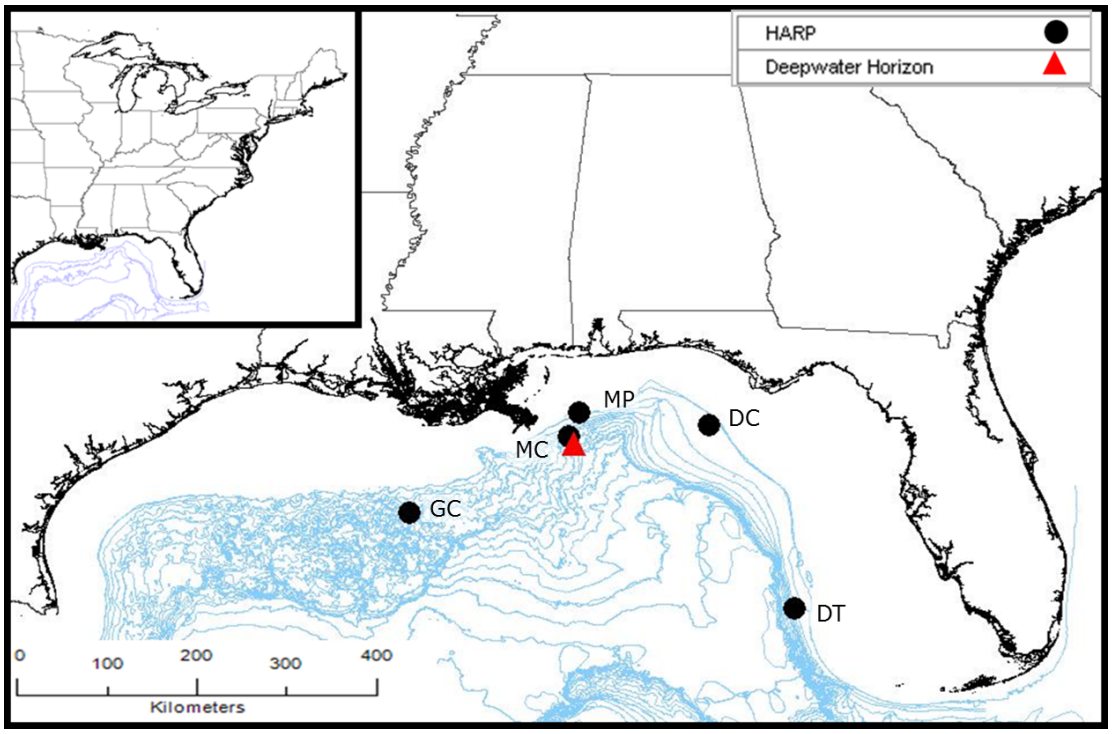


Figure 1.1: Map of the five monitoring sites (black circles) in the Gulf of Mexico used for this study, and the location of the 2010 DWH oil rig blowout (red triangle). Note two letter codes used to identify each site. Blue lines indicate 100m depth contours.

Table 1.1: Location of monitoring sites in the Gulf of Mexico. Sites are named after the federal lease block in which they are located.

Site name	Site Code	Depth (m)	Bathymetry Type	Dist. from spill site (km)	Oil coverage
Desoto Canyon	DC	270	Shelf	225	No
Dry Tortugas	DT	1320	Slope	484	No
Green Canyon	GC	1115	Slope	304	No
Main Pass	MP	80	Shelf	56	Yes
Mississipi Canyon	MC	980	Slope	14	Yes

Table 1.2: Dates, durations and locations of HARP deployments spanning 3 years at five sites. Site codes are based on the name of the federal lease block in which the site is located.

Data ID	Data Start	Data End	Duration (days)	Long. W	Lat. N	Depth (m)
DC02	10/21/2010	2/6/2011	108	86-05.773	29-03.134	268
DC03	3/21/2011	7/6/2011	107	86-05.800	29-03.210	260
DC04	10/26/2011	3/2/2012	144	86-05.899	29-02.886	260
DC05	3/3/2012	12/9/2012	282	86-05.839	29-02.822	260
DC06	12/9/2012	9/25/2013	289	86-05.873	29-02.902	260
DT01	8/9/2010	10/26/2010	78	84-38.251	25-31.911	1320
DT02	3/3/2011	7/12/2011	129	84-38.251	25-31.911	1320
DT03	7/13/2011	11/14/2011	124	84-38.262	25-31.859	1300
DT04	12/13/2011	1/9/2012	26	84-38.265	25-31.867	1300
DT05	5/27/2012	12/07/2012	195	84-38.041	25-31.938	1200
DT06	12/7/2012	8/18/2013	253	84-38.046	25-31.941	1200
GC01	7/15/2010	10/11/2010	88	91-10.010	27-33.470	1115
GC02	11/8/2010	2/2/2011	86	91-10.014	27-33.466	1160
GC03	3/23/3022	8/8/2011	138	91-10.073	27-33.424	1100
GC04	9/23/2011	2/17/2012	118	91-10.060	27-33.426	1100
GC05	2/28/2012	12/12/2012	289	91-10.562	27-33.440	1100
GC06	12/13/2012	9/10/2013	271	91-10.092	27-33.347	1100
MC01	5/16/2010	8/28/2010	104	88-27.927	28-50.746	980
MC02	9/7/2010	12/19/2010	103	88-27.907	28-50.771	980
MC03	12/20/2010	3/21/2011	91	88-27.909	28-50.775	980

Continued on next page

Table 1.2 – continued from previous page

Data ID	Data Start	Data End	Duration	Long. W	Lat. N	Depth
MC04	3/22/2011	8/15/2011	146	88-27.946	28-50.775	980
MC05	9/22/2011	2/21/2012	152	88-27.991	28-50.797	980
MC06	2/28/2012	12/11/2012	288	88-28.041	28-50.853	980
MC07	12/11/2013	8/3/2013	235	88-28.059	28-50.781	900
MP01	7/4/2010	9/25/2010	83	88-17.530	29-15.204	86
MP02	11/07/2010	2/19/2011	100	88-17.808	29-15.318	93
MP03	3/23/2011	9/6/2011	167	88-17.808	29-15.318	93
MP04	9/22/2011	3/1/2012	161	88-17.702	29-15.354	93
MP05	3/1/2012	11/24/2012	270	88-17.597	29-15.368	90
MP06	12/10/2012	9/25/2013	289	88-17.514	29-15.379	90

Chapter 2

Modeling detectability of delphinid echolocation clicks on seafloor sensors

2.1 Abstract

Passive acoustic monitoring devices can record high frequency, directional delphinid echolocation clicks in the environment. To translate click detections into a measure of local delphinid density, we need to know the likelihood that a dolphin in the vicinity of a sensor will be detected. Efforts to determine this probability are complicated by the limitations of single-sensor monitoring devices, the variability of dolphin behaviors and of the marine environment, and by the unique characteristics of delphinid clicks.

In this chapter, I describe two simulation-based methods designed to estimate the probability of detecting echolocating delphinids in the vicinity of a seafloor sensor. One method estimates the probability of detecting a single click (cue counting), while the other estimates the probability of detecting a group of delphinids (group counting). Simulations predict that groups are up to an order of magnitude more likely to be detected

than individual clicks. Species-specific dive behaviors and click characteristics, including frequency content and source level, are predicted to significantly affect detectability. Seasonal sound speed profile changes may affect the probability of detecting groups at some sites. This work suggests that species-specific, site-specific detection probabilities are necessary for delphinid density estimation based on passive acoustic methods.

2.2 Background: Detection Probability Estimation

A promising application of marine mammal passive acoustics is in developing the ability to estimate and monitor animal densities at field sites over time based exclusively on recording data. The problem of density estimation using passive acoustics has been largely solved for cases in which an animal produces a detectable acoustic cue at a known rate, and the distance between the animal and the sensor can be estimated (Thomas et al., 2010; Marques et al., 2009, 2013). Knowing the distribution of distances at which detections are made is critical for density estimation because the probability of detecting an animal is typically a function its distance from the sensor (Buckland et al., 2001).

Distance to an acoustic source has been estimated from recordings in optimal scenarios. Triangulation is possible when the receiver is an array of simultaneously recording sensors (e.g. Wiggins et al., 2013). Triangulation has also been applied to single sensor cases involving few animals producing mid-frequency, high amplitude sounds, when clear surface and/or seafloor reflections are available (Nosal and Frazer, 2006). Alternatively, the distance to an omnidirectional source with a known source level and/or initial frequency content can be estimated based on properties of the received signal (e.g. Marques et al., 2011). However in many cases, none of these ideal scenarios applies. This is particularly true for delphinids, which tend to travel and forage in large groups and produce large numbers of high-frequency, directional echolocation clicks

with poorly understood frequency content and variable source levels.

2.2.1 Detection Probability Simulation

An alternative method of determining detection probability is through simulation. Detection probability simulation is useful in cases where detection probabilities cannot be estimated from experimental data. In many cases, the data do not exist, or are too sparse to represent the full variability across a range of environmental parameters, species-specific behaviors, cue characteristics, and temporal variability. The field work and instrumentation necessary to characterize detectability experimentally can be prohibitively expensive, and the results do not translate well between studies. Models provide a viable alternative. A model framework can be used to combine parameter estimates across existing studies and incorporate the reported variability into an estimate of detection probability through iterative simulation.

Detection probability simulation has been used for fin, humpback, and beaked whale detection probability estimation (Harris et al., 2013; Helble et al., 2013a; Küsel et al., 2011). Using a Monte Carlo simulation approach (Metropolis and Ulam, 1949), sources are placed programmatically at random locations within a specified area around a receiver. An algorithm then determines whether or not a signal produced by that source would be received and detected, based on signal characteristics, acoustic propagation models, and detector parameters. This sequence is repeated millions of times to generate a map of detection probabilities in the vicinity of the sensor for the signal of interest.

Modeling does have considerable drawbacks: It is necessarily based on a simplified understanding of the acoustic environment, initial signal, animal behaviors, and detector properties. However, if input parameters are carefully verified and sufficient variability is incorporated, a simulation can provide a good first order estimate of detectability.

Modeling is somewhat simplified in the case of delphinids because the animals are typically located within 200 m of the sea surface, even when diving (see for example: Benoit-Bird and Au, 2003; Hastie et al., 2006; Minamikawa et al., 2013; Wells et al., 2009). Tagging studies of various delphinid species indicate that these animals often spend more than half of their time within 20m of the sea surface even during foraging periods associated with deeper dives (Baird et al., 2001; Scott and Chivers, 2009). The high frequency content of delphinid echolocation clicks is associated with rapid amplitude attenuation with distance, generally resulting in small detection ranges. This simplifies the acoustic propagation component of the model, but limits the area that can be monitored acoustically. Furthermore, current models of echolocation click signals are very primitive, therefore only simple propagation effects on cues can be simulated.

2.2.2 Cue-Based and Group-Based Methods

Two basic modeling strategies are implemented in this chapter:

1. Modeling the probability of detecting a single cue (cue counting).

2. Modeling the probability of detecting a group of animals, over a small window of time (group counting).

These methods are based on influential work by Buckland et al. (2001, 2007).

Cue counting methods use individual detection events (cues), as the basic unit for density estimation. The number of cues detected over a period of time is converted into an estimate of animal density. Group counting methods use a small window of time, or snapshot as the basic unit for density estimation. For each time window, animals are determined to be either detected or not, and the number of positive detection time windows is converted into an estimate of animal density, by assuming that a positive time window equates to presence of a group.

Both cue and group counting simulation methods require knowledge of cue properties, animal distributions in space, and local oceanographic conditions. An advantage of cue counting is that the only behavioral parameters required are animal depth and vertical orientation in the water column (assuming that animal orientation in the horizontal plane is random relative to the sensor). Disadvantages of the cue counting approach appear in other density estimation steps. These include sensitivity to false positives, risk of detector saturation, and the fact that cue rates may not be proportional to animal density (e.g. Götz et al., 2006).

In contrast, the group counting approach is relatively insensitive to false positives, detector saturation, and cue rate variation. However, simulation of group detectability requires a group behavior model, as well as knowledge about group orientation and vocalization synchrony (Marques et al., 2013). This method also assumes that average group size is based on data that is relevant for the time of the acoustic survey.

Regardless of the simulation method used, different delphinid species require species-specific parameter choices within the model framework. In this study, we are using echolocation clicks as our cue. (Delphinids also produce omnidirectional tonal calls, however these are not frequently seen in the Gulf of Mexico recording data). Click characteristics including frequency content, beam width, and source level vary between delphinid species (e.g. Au et al., 1986, 1995; Au and Herzing, 2003; Fish and Turl, 1976; Madsen et al., 2004a). Behaviors including depth distributions, group size, and dive synchrony are also species-dependent (e.g. Baird et al., 2001; Heide-Jørgensen et al., 2002; Scott and Chivers, 2009; Wells et al., 2009, 2008). Because true parameter distributions are generally not known or poorly constrained, variability for all of these parameters must be built into the model framework.

2.2.3 Parameterizing the Detector

An essential part of estimating detectability is accurate characterization of the detector's ability to find cues. In some cases, models have used simulated calls embedded in noise (Helble et al., 2013a), or signals with known source levels recorded at known distances (Ward et al., 2011) to empirically determine a detector's performance. However, in the case of delphinid echolocation, click production models are not yet able to predict frequency content as a function of animal orientation relative to a sensor (off-axis angle), therefore full cue simulation is not a viable option.

The solution applied here is to use a simple energy-based detector, which functions predictably based on a constant amplitude threshold. It is important to note that this threshold is an absolute measure of amplitude (peak-to-peak dB re: ÂţPa) from calibrated hydrophones, thus allowing ambient noise to be ignored in the model framework. This amplitude threshold can be implemented simply in the Monte Carlo simulation (see chapter 4 for a description and discussion of the detector that the model described here simulates).

2.2.4 Goals of this Chapter

This chapter details a model-based methodology for estimating the probability of detecting delphinid echolocation signals on seafloor sensors. Two modeling approaches, cue counting and group counting, are implemented and compared. Effects of diel and seasonal parameter shifts on detectability are explored. Site-specific detection probabilities for the Gulf of Mexico HARP dataset are computed separately for deep diving delphinids (incl. Risso's dolphin (*Grampus griseus*) and pilot whale (*Globicephala* sp.)), and shallow diving pelagic species (incl. Stenellid dolphins (*Stenella* spp.) and rough-toothed dolphin (*Steno brenanensis*)).

The model results are analyzed to address three main hypotheses:

- H1: For a given species type and site, predicted detection probability is expected to differ between night and day.
- H2: For a given species type and site, predicted detection probability is expected to differ between winter and summer.
- H3: At a given site deep divers are expected to have higher predicted detection probability than shallow divers.

2.3 Methods

Both models consisted of two loops, one nested within the other, in a Monte Carlo framework. One iteration of the outer loop is considered a simulation, and in this work, 500 simulations were run for each detection probability scenario of interest. For each simulation, a mean and standard deviation for each input parameter were drawn from a random uniform distribution between the ranges associated with that parameter (Tables 2.1 and 2.2. Using this mean and standard deviation, a second random distribution of the appropriate type was then generated for each input parameter. In the inner model loop, each source was assigned parameter values drawn from this second set of distributions.

2.3.1 Model Design: Cue Counting Method

Click detection probability was modeled using a Monte Carlo framework in which animal distributions and echolocation parameters were simulated to estimate the likelihood of detecting clicks produced in the area around each sensor. Within a single iteration of the model, 10^4 source positions were randomly selected in the horizontal

plane around the sensor within a maximum radius of five kilometers, beyond which it was assumed that clicks were not detectable (e.g. Zimmer et al., 2008, for beaked whales). Each source was assigned a depth, drawn from a probability distribution based on species dive behavior data in the literature (Tables 2.1 and 2.2) rendering the model three dimensional. Sources were then assigned body angle in the vertical (pitch), and orientation in the horizontal plane (yaw). Yaw was chosen from a uniform distribution in which all orientations relative to the sensor are equally likely (0° to 359°). Elevation angle distributions may be derived from tag data in the future, but are not currently available, therefore elevations were drawn from a normal distribution with a mean of 0° (body is parallel to the seafloor) and a large standard deviation selected from a uniform distribution between 20° and 40° upon each iteration of the model. Finally, each source was assigned source level and beam directivity based on distributions reported in the literature (Au et al., 1986, 1995; Au and Herzing, 2003; Au et al., 2012a; Fish and Turl, 1976; Madsen et al., 2004a,b; Rasmussen et al., 2002, 2004), and data described in chapter 4 .

Predicted received levels at the HARP were computed for a click produced at each modeled source position, using the parameters assigned to each source, and model-based transmission loss estimates at the click peak frequency (see Section 2.3.4). Clicks with received levels greater than a static amplitude threshold (based on the detector described in Chapter 4) were considered detectable. This is a necessary oversimplification given current click simulation capabilities. Ramifications are further described in the discussion section and in Chapter 4. Detection probability for each model iteration was given by the ratio of detectable clicks to total simulated clicks. The model iterated 500 times for a total of five million sources simulated per detection probability estimate.

2.3.2 Model Design: Group Counting Method

The group probability of detection model functioned similarly to the cue-counting model in overall structure, but assumptions at the level of a detection differed. This model estimated the probability of detecting a group of animals in a five minute window. Since most delphinids' mean dive durations are shorter than this, the model incorporated a range of behavioral possibilities within this time window, including heading, pitch and depth changes (Tables 2.1 and 2.2). Each source location in the xy plane was assigned a maximum depth selected from a literature-based distribution.

In the group model, rather than being assigned a specific heading and pitch, each source was assigned a main heading, and allowed to rotate vertically and horizontally by a certain amount about that heading. For group detection purposes, only the most detectable cue in a time window matters, therefore only the maximum possible received level for each source, given the allowed rotation and depth distribution, was retained.

This model design simulated the detectability of a group of animals that could change direction and dive during the detection time window. Although position could conceivably change during this time window, this behavior was not included in the model, because distance sampling assumptions require that nothing move in or out of the detection area in during the time window (Buckland et al., 2007).

2.3.3 Model Parameterization

Detection probabilities were computed separately for each site (MC, GC, DT, DC, and MP) and delphinid group (deep and shallow divers), using model parameters specific to each. Values of all parameters were randomly and repeatedly selected from expected distributions, which were constructed based on a literature review (Tables 2.1 and 2.2).

Risso's dolphin and pilot whales are generally deep divers, with relatively large

body sizes. Literature values suggest similar dive depth distributions and source levels for these species therefore a generalized parameter set was used to model their detectability (Table 2.1; Southall in prep; Chapter 3 data; Au et al., 1986; Au and Herzing, 2003; Fish and Turl, 1976; Heide-Jørgensen et al., 2002; Madsen et al., 2004a; Wells et al., 2009).

In contrast, Stenellid dolphins are smaller animals, with shallower dive profiles and lower source levels. Literature values for dive profiles and click parameters of rough-toothed dolphins and false killer whales suggest that they are more similar in these categories to Stenellid dolphins than to the deeper divers, therefore they are included in the shallow diver category (Table 2.2; Au and Herzing, 2003; Au et al., 1995; Baird et al., 2001; Madsen et al., 2004a; Scott and Chivers, 2009; Wells et al., 2008). Separate day and night models were created for both deep and shallow divers, because these pelagic species typically dive deeper at night than during the day (Baird, 2002; Scott and Chivers, 2009).

Due to computation limitations, beam directivity was drawn from a uniform distribution and kept constant within each model iteration, but varied between iterations. Beam directivity was used to estimate a simulated beam pattern using a piston model (Zimmer et al., 2005). This model is an approximation at best, and does not realistically predict amplitudes at large off-axis angles. In order to compensate for this, mean values for beam amplitudes at 90° and 180° off-axis angles were drawn from a uniform distribution based on beam patterns in the literature (Au et al., 1986, 2012a). A linear interpolation was used to complete the off-axis beam shape, by interpolating between the 90° and 180° amplitude values.

Signal peak frequency, which primarily affects model results by influencing absorption predictions, was also generalized for computational reasons. Because echolocation click models are not yet able to simulate click frequency content as a function of off-axis angle, single frequency value was used to approximate transmission

loss. For modeling purposes, the critical value is the mean peak frequency of received clicks recorded for a species of interest. This approximate frequency was identified for each species category based on a manual review of the HARP detection data (see Chapter 5).

Parameters differed slightly between slope and shelf site simulations (Table 2.3). Shelf sites were noisier in general, therefore a higher minimum received level threshold was used, in order to simulate the detector (see Chapter 4 for details).

Table 2.1: Literature-based parameter ranges from which inputs were drawn for the Monte Carlo simulation of deep diver (Risso's dolphin and pilot whale) detectability. Model type indicates click (C) or group (G) method. TL is transmission loss. For each outer loop of the simulation, a mean and standard deviation for each parameter were drawn from a random uniform distribution between the listed ranges associated with that parameter. A second random distribution of the type listed under "Distribution" was then generated for each parameter using the selected mean and standard deviation. Note that log-normal distributions were used to dive depths. In the inner model loop, each source was assigned parameter values drawn from this second set of distributions.

Parameter	**Model**	**Mean**	**Std. Dev.**	**Distribution**	**Ref.**
Dive depth (day)	C	1.5 to 2.5 m	0.5 - 1 m	Log-normal	(1)
Dive depth (night)	C	2 - 3.5 m	1 - 1.5 m	Log-normal	(1)
Dive depth (day)	G	2 - 2.5 m	0.5 - 1 m	Log-normal	(1)
Dive depth (night)	G	2.5 - 3.5 m	1 - 1.5 m	Log-normal	(1)
Directivity	C	23 - 27 dB_{pp}	n/a	Uniform	(2)
Min off-axis TL	G	23 - 27 dB_{pp}	2 - 5 dB_{pp}	Normal	(2)
90° off-axis TL	C	23 - 27 dB_{pp}	n/a	Uniform	(2)
180° off-axis TL	C	33 - 37 dB_{pp}	n/a	Uniform	(2)
Peak Frequency	C, G	34kHz	n/a	none	(3)
Source level	C, G	215-225 dB_{pp}	2 - 4 dB_{pp}	Normal	(3)
Orientation: Elevation	C	±0°	20 - 40°	Normal	
Orientation: Azimuth	C	±0-359°	n/a	Uniform	(4)
Rotation: Elevation	G	±55 - 65°	10-15°	Normal	
Rotation: Azimuth	G	±140 - 160°	10 - 20°	Normal	

(1) Wells et al., 2009; Heide-Jørgensen et al., 2002, Southall in prep,
(2) Au et al., 1986, 2012a
(3) Au and Herzing, 2003; Fish and Turl, 1976; Madsen et al., 2004a, Chapter 3 data,
(4) Required by distance sampling assumptions.

Table 2.2: Literature-based parameter ranges from which inputs were drawn for the Monte Carlo simulation of shallow diver (Stenellid dolphin, rough-toothed dolphin and false killer whale) detectability. Model type indicates click (C) or group (G) method. TL is transmission loss. For each outer loop of the simulation, a mean and standard deviation for each parameter were drawn from a random uniform distribution between the listed ranges associated with that parameter. A second random distribution of the type listed under "Distribution" was then generated for each parameter using the selected mean and standard deviation. Note that log-normal distributions were used to dive depths. predict In the inner model loop, each source was assigned parameter values drawn from this second set of distributions.

Parameter	**Model**	**Mean**	**Std. Dev.**	**Distribution**	**Ref.**
Dive depth (day)	C	0.5 - 1.5 m	0.5 - 1 m	Log-normal	(1)
Dive depth (night)	C	1.5 - 3 m	0.5 -1 m	Log-normal	(1)
Dive depth (day)	G	1 - 1.5 m	.5 - 1 m	Log-normal	(1)
Dive depth (night)	G	2 - 3 m	.5 - 1 m	Log-normal	(1)
Directivity	C	23 - 27 dB_{pp}	n/a	Uniform	(2)
Min off-axis TL	G	23 - 27 dB_{pp}	2 - 5 dB_{pp}	Normal	(2)
90° off-axis TL	C	23 - 27 dB_{pp}	n/a	Uniform	(3)
180° off-axis TL	C	33 - 37 dB_{pp}	n/a	Uniform	(3)
Peak Frequency	C, G	40kHz	n/a	none	(2)
Source level	C, G	205 - 215 dB_{pp}	2 - 4 dB_{pp}	Normal	(2)
Orientation: Elevation	C	±0°	20 - 40°	Normal	
Orientation: Azimuth	C	±0-359°	n/a	Uniform	(4)
Rotation: Elevation	G	±55 - 65°	10-15°	Normal	
Rotation: Azimuth	G	±140 - 160°	10 - 20°	Normal	

(1) Baird et al., 2001; Scott and Chivers, 2009; Wells et al., 2008
(2) Au and Herzing, 2003; Au et al., 1995; Madsen et al., 2004a,b; Rasmussen et al., 2002, 2004, Chapter 3 data,
(3) Au et al., 1986, 2012a
(4) Required by distance sampling assumptions.

2.3.4 Acoustic Propagation Modeling

Transmission loss as a function of distance and elevation relative to each sensor was simulated using the ray-tracing algorithm Bellhop (Porter and Bucker, 1987), with site-specific environmental and physical parameters drawn from the Oceanographic and Atmospheric Master Library (OAML). The ray tracing technique used by the Bellhop algorithm models acoustic wave motion by breaking the signal into a large number of narrow beams, radiating out from a source (Porter and Bucker, 1987). Each ray is made up of very short straight lines, whose directions change gradually as a function of the properties of the modeled medium. Together they form long curving paths. This approach relies on the assumption that the acoustic wavelength is very small compared to physical features in the propagation environment. The algorithm is therefore well-suited for high frequency modeling, where other models based on normal modes or parabolic equations, are inefficient.

Using the theorem of acoustic reciprocity (Rayleigh, 1887), incoherent transmission loss in a gridded volume around each site was simulated by defining the sensor location as an omnidirectional sound source. Three thousand rays were projected from the sensor position, at launch angles ranging from -90° to 90°, where 0° is directly above the instrument.

Transmission loss at each frequency was computed along 64 equally spaced radials around each HARP site to render the model three dimensional. Each radial calculation resulted in a grid covering the full depth of the water column, and extending 20 km horizontally from the sensor. Vertical grid resolution was one meter and horizontal resolution was ten meters. Vertical grid resolution was kept fine to limit the influence of the model's surface boundary layer to a one meter bin. Large radials were computed to check for non-monotonic detectability changes and to allow for development of models for other species with higher amplitude signals, such as sperm whales.

Environmental and physical parameters were extracted from OAML using ESME Workbench (Mountain et al., 2013). Bottom bathymetry was obtained from the global Digital Bathymetry Database (DBDB, version 5.4, 1' resolution) available from OAML. The bathymetry for each radial was computed as a linear, piece-wise interpolation between the grid points extracted from the DBDB. Bottom composition was clay at all sites according to OAML's Bottom Sediment Type database (BST, version 2.0, 2' resolution). The bottom boundary was modeled as an acousto-elastic half-space (Porter, 1992). Sound speed profiles were range dependent, such that the profile applied to each analysis position was dependent on geographical location, rather than applying a single profile to the entire volume. All sound speed profiles were draw from OAML's Generalized Digital Environment Model (GDEM, version 3.0, 15' resolution).

Potential seasonal environmental changes on detectability were tested for by computing and comparing detection probabilities based on mean monthly modeled sound speed profiles from January (winter) and July (summer), as obtained from OAML databases, which are in turn based on historical values. The sound speed profiles used are averages, and do not reflect particular hydrographic events. This is in keeping with the aim of the model to provide an average probability of detection over a large number of encounters and a wide variety of conditions.

2.3.5 Hypothesis Tests

Hypotheses were tested using a two-sample Kolmogorov-Smirnov test (Massey Jr, 1951), to compare distributions of mean detection probability estimates between iteration simulations. In all cases, the null hypothesis was that the two distributions were drawn from the same continuous distribution, and the alternative hypothesis was that they were drawn from different continuous distributions. A p-value less than a significance level of 5% significance level was required to reject the null hypothesis.

Table 2.3: Site-specific model input parameters. Note that a higher received level threshold (RL) was used for shelf sites MP and DC due to noise conditions at these locations.

Site	DC	DT	GC	MC	MP
Latitude (N)	29.047	25.531	27.557	28.846	29.256
Longitude (W)	86.097	84.638	91.168	88.465	88.295
Sensor Depth (m)	272	1158	1336	990	97
RL Threshold ($dB_{pp)}$	117	115	115	115	117
Model Radius (km)	5	5	5	5	5

2.4 Results

For each species group, deep divers and shallow divers, detection probability was computed using the both the click and group modeling methods (Tables 2.4 and 2.8). Detection probability decreased as a function of range in all simulations (Figures 2.1 to 2.4), with highest detection probabilities when an animal or group was located within 100m of the sensor (horizontal range).

Differences between model predictions, as well as results of the hypothesis tests were as follows:

- H1: Diel differences in detection probability

  Click model: For a given species group (deep or shallow divers) at slope sites the distributions of modeled click detection probabilities were not significantly different between day and night models (Table 2.5). At shelf sites, the only significant difference was seen for the deep diver model in summer at site MP, with higher average detectability at night.

  Group model: The probability of detecting groups was significantly different

between day and night models at shelf sites for both deep and shallow divers in both summer and winter (Table 2.9), with higher average detection probabilities at night. No significant differences were found between day and night group detection probability distributions at slope sites for either species category.

- H2: Seasonal differences in detection probability

  Click model: At shelf site MP, there was a significant seasonal change in detectability for shallow divers in both day and night scenarios (Table 2.6). Deep divers only showed a seasonal detection probability distribution difference at MP in night scenarios. At shelf site DC, there was no significant difference between night model predictions for the deep diver category as a function of season. A significant difference was found between shallow diver night models in Janauary compared to July models at site DC, but no difference was found for day models.

  Significant differences between summer and winter model prediction distributions were not seen at slope sites for either species category or time of day.

  Group model: The distribution of detection probabilities for deep diver groups was significantly different in summer than in winter models for northern slope sites MC and GC, with higher mean detection probabilities predicted in summer (Table 2.10). There was no significant difference in shallow diver models, and no significant difference at the southern slope site DT. At shelf sites, both deep and shallow diver detection probability distributions were significantly different between summer and winter models. Both were predicted to be less detectable on average in summer than in winter.

- H3: Inter-species differences in detection probability

  Click model: Deep diver and shallow diver models predicted significantly different click detection probability distributions in all scenarios (Table 2.7). Deep diver

clicks were predicted to be roughly five times more likely than shallow diver clicks to be detected at slope sites, and three times more detectable at shelf sites.

Group model: Deep diver and shallow diver models predicted significantly different group detection probabilities in all scenarios (Table 2.11). Deep diving groups were predicted to be two to three times more detectable than shallow diving groups on average, at all sites.

Table 2.4: Model-based estimates of the site specific probability (as a %) of detecting an individual echolocation click, plus or minus one standard deviation (σ), accounting for time of day and season, within a 5 km radius. Models were used to generate models associated with night and day dive behaviors using both summer (July) and winter (January) mean sound speed profiles. Separate estimates are presented for shallow diving delphinids (including Stenellid dolphins, Rough-toothed dolphins, and false killer whales), and deep diving delphinids (including Risso's dolphins and pilot whales). Clicks are assumed to be undetectable at ranges larger than 5 km.

		January		July	
Site	day/night	shallow	deep	shallow	deep
MC	day	1.4 ±0.8 σ	7.5 ±2.3 σ	1.4 ±0.8 σ	7.5 ±2.3 σ
	night	1.4 ±0.7 σ	7.7 ±2.4 σ	1.4 ±0.7 σ	7.7 ±2.3 σ
DT	day	1.1 ±0.7 σ	7.4 ±2.3 σ	1.1 ±0.7 σ	7.2 ±2.2 σ
	night	1.2 ±0.7 σ	7.4 ±2.3 σ	1.1 ±0.7 σ	7.3 ±2.3 σ
GC	day	1.1 ±0.7 σ	7.3 ±2.4 σ	1.1 ±0.7 σ	7.2 ±2.3 σ
	night	1.1 ±0.7 σ	7.4 ±2.4 σ	1.2 ±0.7 σ	7.3 ±2.4 σ
MP	day	2.1 ±0.6 σ	6.8 ±1.5 σ	2.0 ±0.6 σ	5.0 ±0.8 σ
	night	2.2 ±0.7 σ	6.7 ±1.6 σ	2.1 ±0.6 σ	5.6 ±1.0 σ
DC	day	2.2 ±0.7 σ	7.7 ±2.0 σ	2.2 ±0.8 σ	8.1 ±1.9 σ
	night	2.2 ±0.8 σ	8.0 ±2.1 σ	2.2 ±0.8 σ	8.1 ±2.1 σ

Table 2.5: Click Method: P-values from two-sample Kolmogorov-Smirnov goodness-of-fit hypothesis tests, comparing mean detection probability distributions obtained from day and night simulations, for a given diver type and season (January and July indicate winter and summer models respectively). A p-value less than 0.05, denoted with * is taken to indicate a significant difference between distributions.

	January		July	
Site	Shallow	Deep	Shallow	Deep
MC	0.90	0.40	0.36	0.25
DT	0.17	0.86	0.86	0.60
GC	0.76	0.28	0.07	0.32
MP	0.25	0.66	0.06	*0.00
DC	0.19	0.08	0.32	0.12

Table 2.6: Click Method: P-values from two-sample Kolmogorov-Smirnov goodness-of-fit hypothesis tests, comparing mean detection probability distributions obtained from January (winter) and July (summer) simulations, for a given diver type and time of day. A p-value less than 0.05, denoted with *, is taken to indicate a significant difference between distributions.

	Shallow divers		Deep divers	
Site	Day	Night	Day	Night
MC	0.66	0.71	0.98	0.93
DT	0.96	0.19	0.50	0.45
GC	0.36	0.86	0.81	0.66
MP	*0.03	*0.00	0.09	*0.00
DC	0.32	*0.02	0.19	0.22

Table 2.7: Click Method: P-values from two-sample Kolmogorov-Smirnov goodness-of-fit hypothesis tests, comparing mean detection probability distributions obtained for deep and shallow diver simulations, for a given month and time of day. January and July indicate winter and summer models respectively. A p-value less than 0.05, denoted with *, is taken to indicate a significant difference between distributions.

	January		July	
Site	Day	Night	Day	Night
MC	0.00*	0.00*	0.00*	0.00*
DT	0.00*	0.00*	0.00*	0.00*
GC	0.00*	0.00*	0.00*	0.00*
MP	0.00*	0.00*	0.00*	0.00*
DC	0.00*	0.00*	0.00*	0.00*

Table 2.8: Model-based estimates of the site specific probability (as a %) of detecting a group of delphinids, plus or minus one standard deviation (σ), accounting for time of day and season, within a 5 km radius. Models were used to generate models associated with night and day dive behaviors using both summer (July) and winter (January) mean sound speed profiles. Separate estimates are presented for shallow diving delphinids (including Stenellid dolphins, Rough-toothed dolphins, and false killer whales), and deep diving delphinids (including Risso's dolphins and pilot whales). Groups are assumed to be undetectable at ranges larger than 5 km.

		January		July	
Site	day/night	shallow	deep	shallow	deep
MC	day	15.5 ±2.9 σ	40.8 ±6.0 σ	15.8 ±3.1 σ	43.1 ±6.5 σ
	night	15.6 ±2.9 σ	41.1 ±6.0 σ	15.7 ±3.0 σ	43.4 ±6.6 σ
DT	day	15.5 ±3.2 σ	41.7 ±6.3 σ	15.3 ±3.2 σ	42.1 ±6.2 σ
	night	15.1 ±3.1 σ	41.8 ±6.5 σ	15.5 ±3.2 σ	42.1 ±6.5 σ
GC	day	15.0 ±3.2 σ	41.0 ±6.1 σ	15.2 ±3.1 σ	42.6 ±6.5 σ
	night	15.1 ±3.0 σ	41.3 ±6.1 σ	15.1 ±3.1 σ	42.7 ±6.4 σ
MP	day	12.2 ±1.7 σ	27.5 ±3.6 σ	7.8 ±0.9 σ	18.8 ±2.7 σ
	night	12.8 ±2.0 σ	30.9 ±4.2 σ	9.0 ±1.4 σ	24.3 ±3.7 σ
DC	day	14.3 ±2.0 σ	28.2 ±3.0 σ	13.7 ±1.6 σ	25.9 ±2.7 σ
	night	15.4 ±2.1 σ	30.3 ±2.9 σ	14.3 ±1.7 σ	27.8 ±2.9 σ

Table 2.9: Group Method: P-values from two-sample Kolmogorov-Smirnov goodness-of-fit hypothesis tests, comparing mean detection probability distributions obtained from day and night simulations, for a given diver type and season (January and July indicate winter and summer models respectively). A p-value less than 0.05, denoted with *, is taken to indicate a significant difference between distributions.

	January		July	
Site	Shallow	Deep	Shallow	Deep
MC	0.81	0.93	0.40	0.60
DT	0.12	0.90	0.32	0.86
GC	0.45	0.66	0.93	0.93
MP	*0.00	*0.00	*0.00	*0.00
DC	*0.00	*0.00	*0.00	*0.00

Table 2.10: Group Method: P-values from two-sample Kolmogorov-Smirnov goodness-of-fit hypothesis tests, comparing mean detection probability distributions obtained from January (winter) and July (summer) simulations, for a given diver type and time of day. A p-value less than 0.05, denoted with *, is taken to indicate a significant difference between distributions.

	Shallow divers		Deep divers	
Site	Day	Night	Day	Night
MC	0.11	*0.00	0.99	*0.00
DT	0.28	0.66	0.07	0.36
GC	0.40	*0.01	0.71	*0.04
MP	*0.00	*0.00	*0.00	*0.00
DC	*0.00	*0.00	*0.00	*0.00

Table 2.11: Group Method: P-values from two-sample Kolmogorov-Smirnov goodness-of-fit hypothesis tests, comparing mean detection probability distributions obtained for deep and shallow diver simulations, for a given month and time of day. January and July indicate winter and summer models respectively. A p-value less than 0.05, denoted with *, is taken to indicate a significant difference between distributions.

	January		July	
Site	Day	Night	Day	Night
MC	0.00*	0.00*	0.00*	0.00*
DT	0.00*	0.00*	0.00*	0.00*
GC	0.00*	0.00*	0.00*	0.00*
MP	0.00*	0.00*	0.00*	0.00*
DC	0.00*	0.00*	0.00*	0.00*

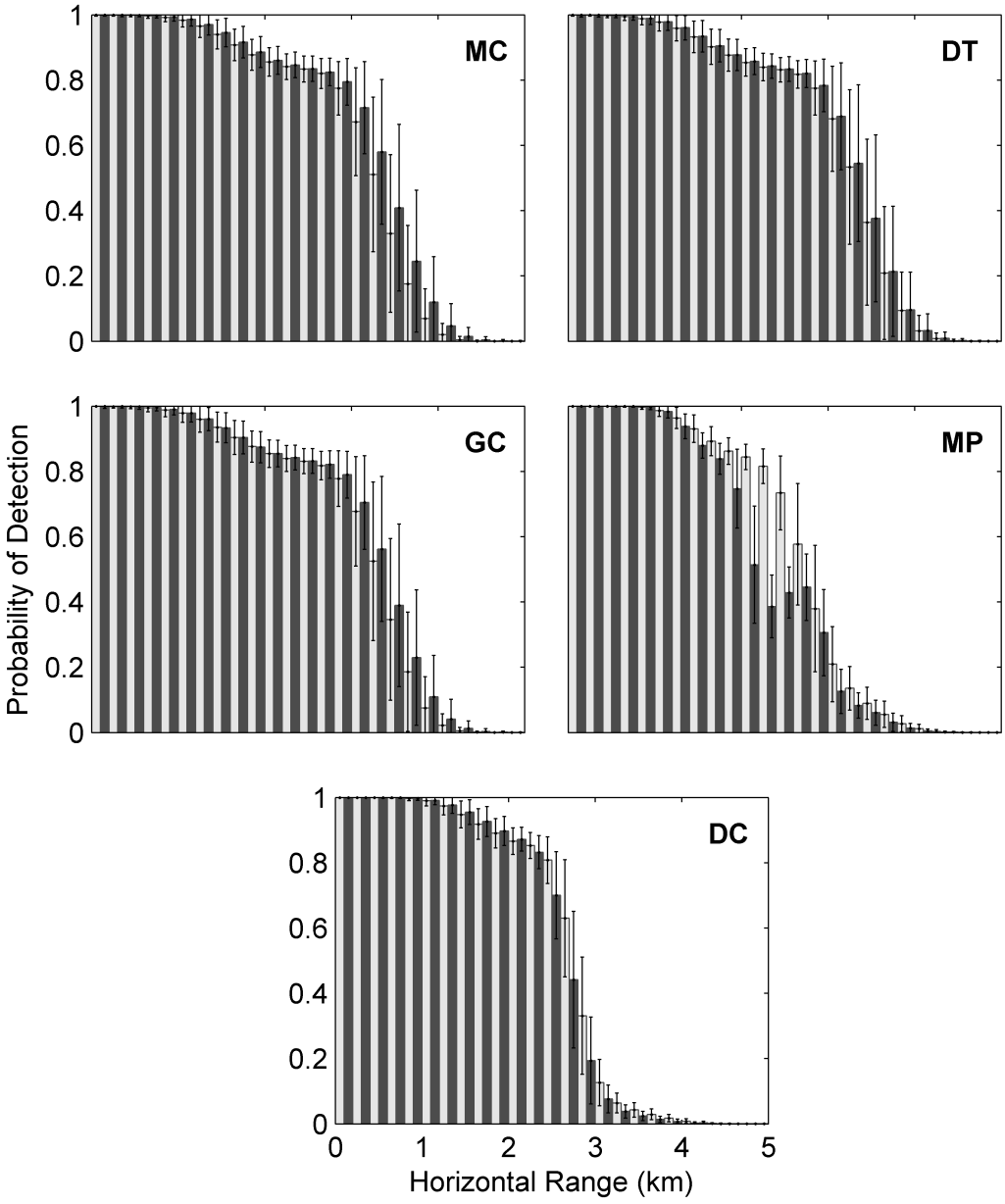


Figure 2.1: Group Model: Nighttime model-based probability of detecting a deep diving group, as a function of range, at all sites. Light gray bars: January (winter); Dark gray bars: July (summer).

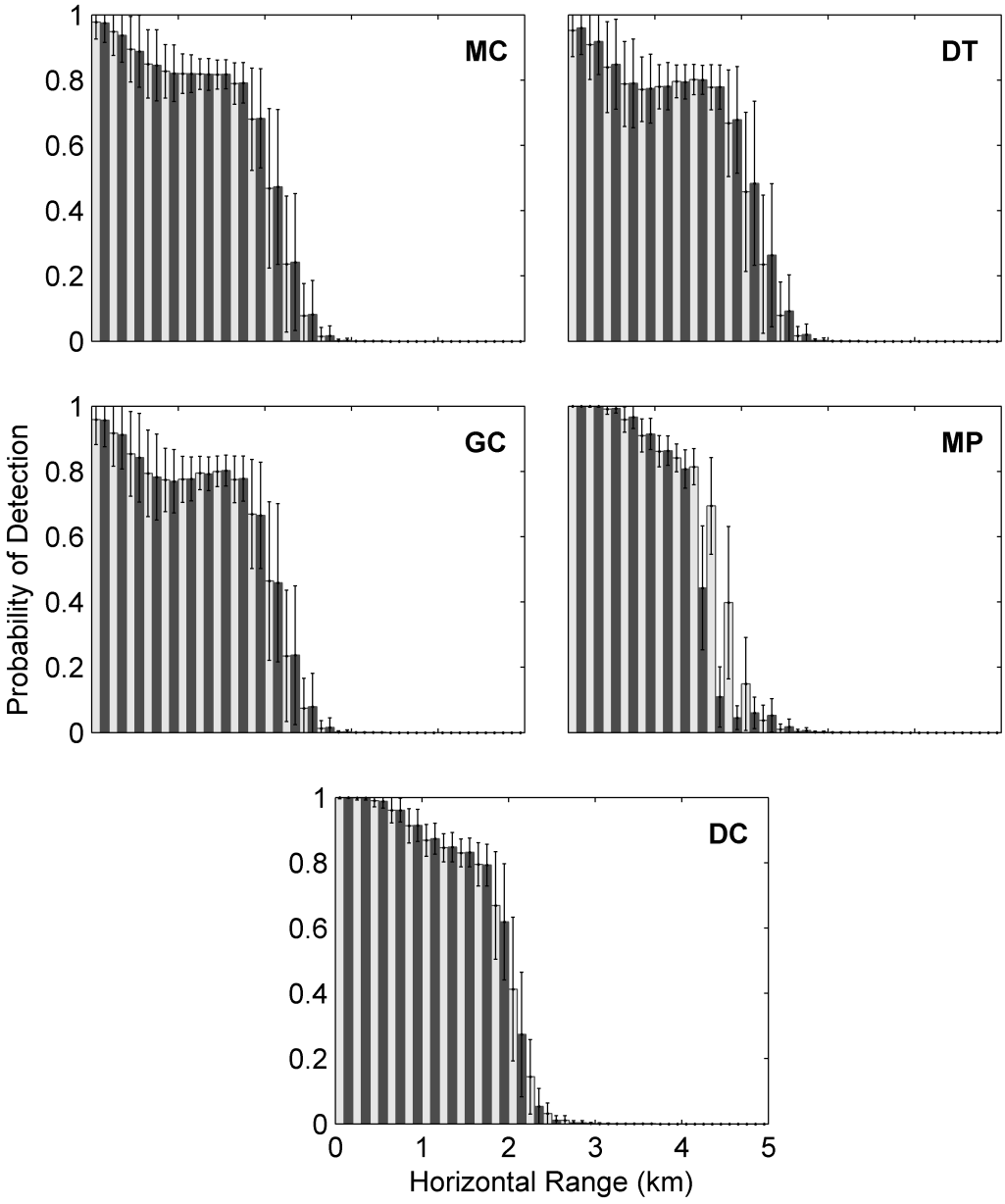


Figure 2.2: Group Model: Nighttime model-based probability of detecting a shallow diving group, as a function of range, at all sites. Light gray bars: January (winter); Dark gray bars: July (summer).

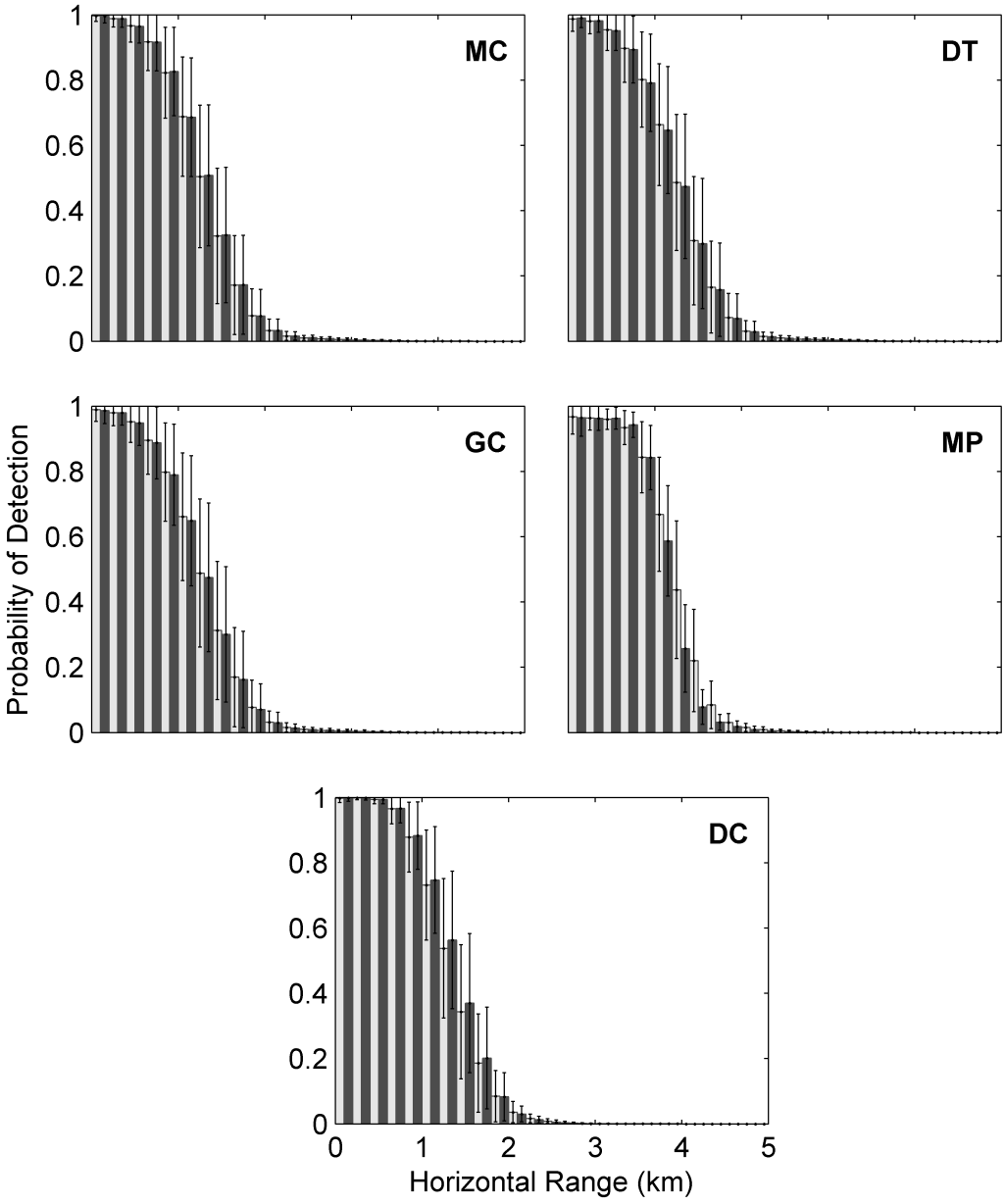


Figure 2.3: Click Model: Nighttime model-based probability of detecting a deep diver click, as a function of range, at all sites. Light gray bars: January (winter); Dark gray bars: July (summer).

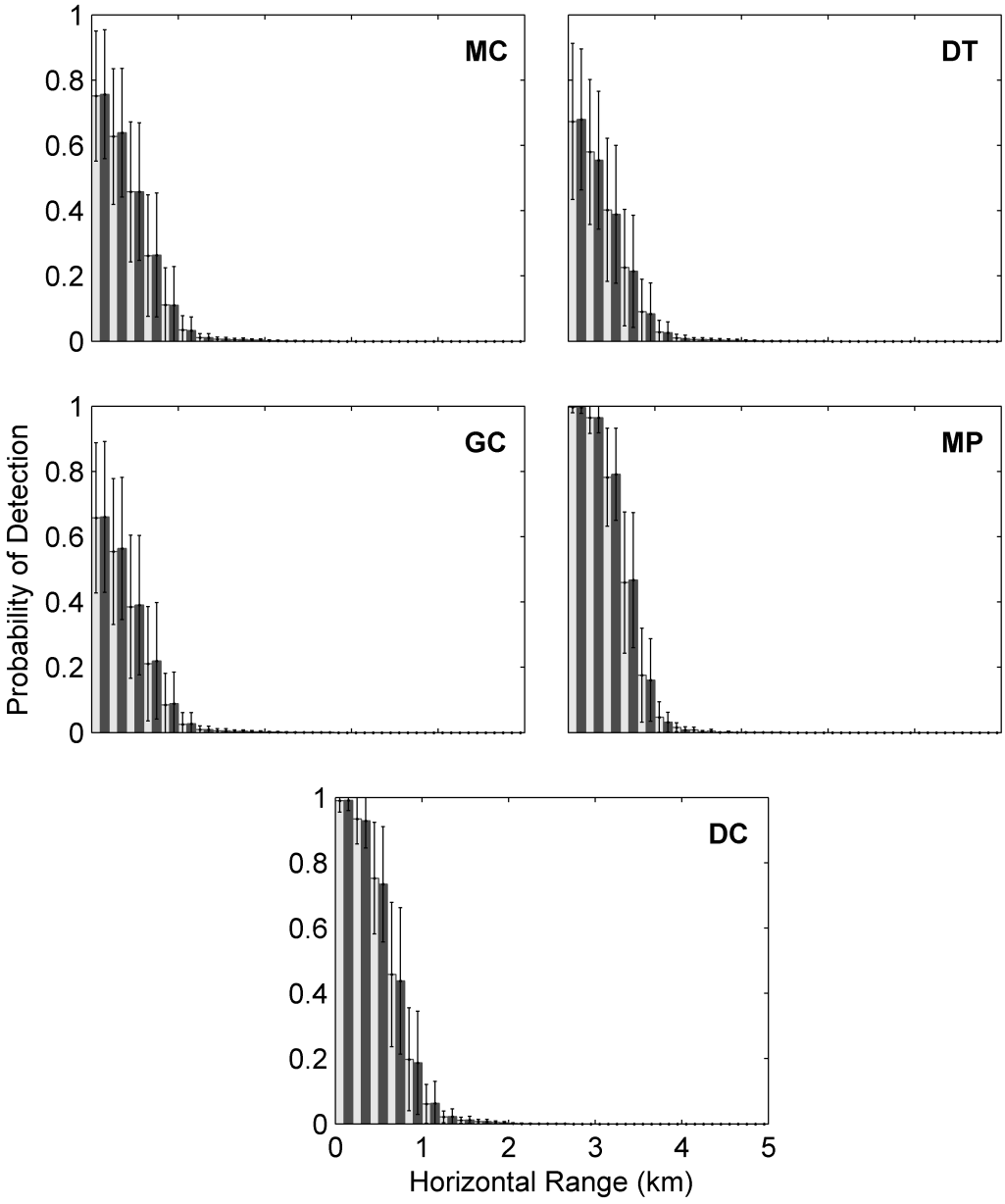


Figure 2.4: Click Model: Nighttime model-based probability of detecting a shallow diver click, as a function of range, at all sites. Light gray bars: January (winter); Dark gray bars: July (summer).

2.5 Discussion

2.5.1 Click Detection Probabilities

The click-based probability of detection model aims to estimate the likelihood of detecting a single delphinid click given that it was produced within a 5 km radius of the sensor. In general, this probability is very low, when averaged across the entire 79 km^2 monitoring area (Table 2.4). However, when detectability is considered as a function of range (Figures 2.3 and 2.4), the reason behind the low probabilities becomes clear: Detection probabilities are high in the small region immediately around the sensor, but are outweighed by low detection probabilities at large ranges. Since the area monitored increases with the square of distance, the overall probability of detection is low.

Click detectability is most affected by two main factors. The first is source level. The main difference between the deep diver and shallow diver models is the estimated source level distribution. Deep divers' source levels are typically higher, making them detectable at larger ranges than the shallow divers, under otherwise similar model conditions (Table 2.4). Preliminary work during model development also suggested that the click model is sensitive to assumptions about beam pattern. Since most of the clicks are expected to be received off-axis, assumptions about beam width and off-axis amplitudes can have relatively large effects on model predictions. This sensitivity is addressed within the model by sampling from a range of probable beam directivities and off-axis amplitudes, and is therefore expected to be included in the error estimates. Further research on click directivity and off-axis amplitudes would improve the reliability of model predictions.

The second factor influencing click detectability within the model is site depth. Detection ranges are slightly shorter at shelf sites due to the higher detection thresholds used there relative to slope sites. However despite the received level threshold differences,

animals are more likely to be detected at close horizontal ranges because the water depth is much less than at slope sites (Table 2.3). For the same horizontal range, the animals are actually much closer to the sensor at shelf sites than slope sites. In general, when animal dive depths are large relative to the site depth, predicted detectability increases, therefore deep divers are more likely to be detected at deep sites than shallow divers.

Both sets of species are thought to dive deeper at night than during the day (Baird et al., 2001; Heide-Jørgensen et al., 2002; Scott and Chivers, 2009; Wells et al., 2008, 2009), but the effect of this difference on detectability is predicted to be very small. False killer whales may exhibit the reverse behavior, diving deeper during the day than at night (Minamikawa et al., 2013), but unless the depth distributions are vastly different, this would not be expected to have a large effect on detectability.

Overall, although there are some small, significant differences in the model predictions of mean click detection probability as a function of season and time of day, these differences are not expected to have a large effect on density estimates. The estimated detectability differences are below a few percentage points on average. When combined with other factors influencing density estimates and the associated variances, these temporal effects are expected to be minimal. The modeled sound propagation environment is generalized, and may not account for the effects of specific oceanographic events. In general, the smaller the predicted detection range, of the cue of interest, the smaller the impact of sound speed profiles on detectability.

The click detection model predicts that individual delphinid clicks are only detectable when an animal is on or nearly on-axis relative to the instrument, or the animal is close to the sensor. A shallow-diver click produced at the sea surface directly above a deep (>1000m) seafloor sensor may not be detected if the animal's beam is not pointed downward. On the other extreme, a deep diver click produced over three km away from the sensor, may be detected if the animal's beam is aimed toward the sensor.

The use of received level as the the only metric of detectability is an oversimplification. In reality echolocation click detectability is also influenced in part by frequency content and energy distribution in the time-series, both of which can vary as a function of off-axis angle. Since we currently have no methods for modeling the effects of off-axis angle on received click structure, an alternative approach is used, in which click detection is simplified to rely as much as possible on received level (Chapter 4), which can be modeled. Future work will likely seek to improve echolocation click simulation in order to allow for more sophisticated detection methods and more refined estimates of click detectability.

2.5.2 Group Detection Probabilities

The probability of detecting a group is related to the click detection probability in many ways, but it attempts to mimic a snapshot approach over a short time window, and to estimate the detection probability during that period. Integrating over a small time window is expected to increase detection probabilities relative to cue counting methods. This is borne out by the group model predictions, which estimate that the probability of detecting a shallow-diving group in a five minute window will be an order of magnitude higher than the probability of detecting a single click from that group. Likewise, the probability of detecting a deep-diving group in a five minute window is predicted to be nearly five times higher than the probability of detecting a single click from that group.

Because it integrates detectability across a time window, the group model predicts larger maximum detection ranges than the click model, given the same conditions. This is because an on-axis click is theoretically detectable at relatively large ranges, but the probability of an individual click being on-axis is very low. In the group model, the probability of a click being on-axis is much higher because time and multiple animals make it more likely that at least one animal in the group will orient toward the sensor at

least once during the window. While this is probably true in practice, in the model realm, the distribution of rotational angles and depths expected of an average group is poorly constrained, if at all. This is a considerable weakness in the group model because group rotation assumptions have a large effect on detectability predictions.

The high detection probabilities at horizontal ranges over 2 km predicted by the group model make detectability predictions more sensitive to seasonal changes in sound-speed profiles than the click model predictions. The sound propagation models predict that the strong summer thermocline will have two main effects (Urick, 1967). First it will trap sound energy produced above the thermocline in the shallow mixed layer at the sea surface, reducing the amplitude of the signal that reaches the seafloor sensor below. Second, the thermocline will strongly refract sound energy that does escape the surface layer toward the seafloor, reducing the distance that the signal will travel before it bounces off the seafloor (Figure 2.5 A). In contrast, the weak winter thermocline and deep mixed layer is predicted to allow more near-surface acoustic energy to reach the seafloor, and to cause weaker downward signal diffraction (Figure 2.5 B). For shallow diver click detectability, the probability of detecting an individual click at the ranges at which these effects would be significant is small. However the estimated probability of detecting a group is greater at these ranges, therefore the effects of seasonal thermocline shifts are predicted to be larger. This effect is seen at shelf sites MP and DC. Groups of both deep and shallow divers are predicted to be less detectable in summer than in winter at horizontal ranges over 1.5 km. Groups are also predicted to be more detectable during summer nights when they spend a larger portion of the time below the thermocline, than summer days. This effect is stronger for the deep-diving group.

A different seasonal effect is seen at the two northern slope sites, MC and GC. The group detectability model predicts slightly higher detection probabilities in the summer, but only at night. At night, the deep diver model puts more animals just below the

thermocline, which has the reverse effect of trapping energy at depth, pushing detection probabilities slightly higher (Figure 2.6). This effect is not seen at the shallow sites because the signal would bounce off the seafloor prior to arriving at the sensor at the ranges at which this effect is seen. The effect is not seen at site DT because the summer thermocline there is not typically as strong there as it is at the more northern sites.

Modeling group behavior over a short time window also requires assumptions about how a group's orientation changes over time. We know from manual analysis of recording data that the period of time during which a group of animals appears to be approaching the sensor (amplitudes steadily increasing over time) is generally much longer than the period during which the group appears to be leaving it (amplitudes steadily decreasing). This suggests that the animals in the group are more likely to be clicking in their direction of travel, and may not turn back 180° to click on-axis after the point of closest approach. In the case where the animals are foraging, full 360° rotation in the time window may be more likely than in a traveling mode. These different behavioral modes are not currently captured by the group model, but their incorporation, along with the availability of further data, could improve the reliability of the group model approach in the future.

Group spread is currently not incorporated into the model. Group spread is a complex variable to predict, because it is expected to vary as a function of group size and behavior, which are poorly characterized. Data on group spread could be extracted from tracking data in the future, but is not currently available. The effect of modeling a group as a single point is mainly seen at short ranges in figure 2.2, for slope sites GC and DT, where a slight dip in detectability is seen at approximately 1km. Incorporating spread would likely remove the predicted dip in detectability because group spread would be larger than the affected area. This would have the effect of smoothing out dips in detectability as a function of range. Average predicted group detectability is not

expected to change dramatically if group spread is incorporated, because if the animals are randomly distributed in space according to a uniform distribution, the distribution of distances between the sensor and the most detectable animal in a group should also be random and uniform.

The MP site is characterized by asymmetry in detection probabilities in summer months (Figure 2.7 A). Northward, up-slope radials show increased detectability at horizontal ranges over two kilometers, relative to down-slope radials, due to a higher likelihood of indirect arrival paths from sources near the surface. Detection probability drops off more quickly along southward, down-slope radials at this site. The model asymmetry is primarily seen summer months. This effect is less dramatic for the lower amplitude clicks of the Stenellid dolphins, which are not predicted to be highly detectable at the ranges for which the effect is seen. Models for all other sites predict largely symmetric detection probabilities about the sensor (Figure 2.7 B). Asymmetric detection probabilities are problematic if animals are not uniformly distributed in the area around the sensor over time. Since uniform distribution is one of the primary assumptions of point transect density estimates, it is not a problem here. However, if that assumption is violated at site MP (for example, dolphins are only present on the shallow side leading to an underestimation of detection probability) models would need to be adjusted to incorporate the non-uniformity. Data to verify this assumption are not currently available.

It is clear that detectabilty models must be tailored to each species' signal characteristics and behavior. Uncertainty in modeled detection probabilities could be significantly reduced given more specific characterizations of group and individual behaviors, as well as a more accurate acoustic model of echolocation clicks. Large-scale tagging and tracking efforts, which would provide much of the information needed to improve the behavioral components of the simulations, are beginning to target delphinids, but data is still sparse. Although these more detailed inputs are currently lacking, the

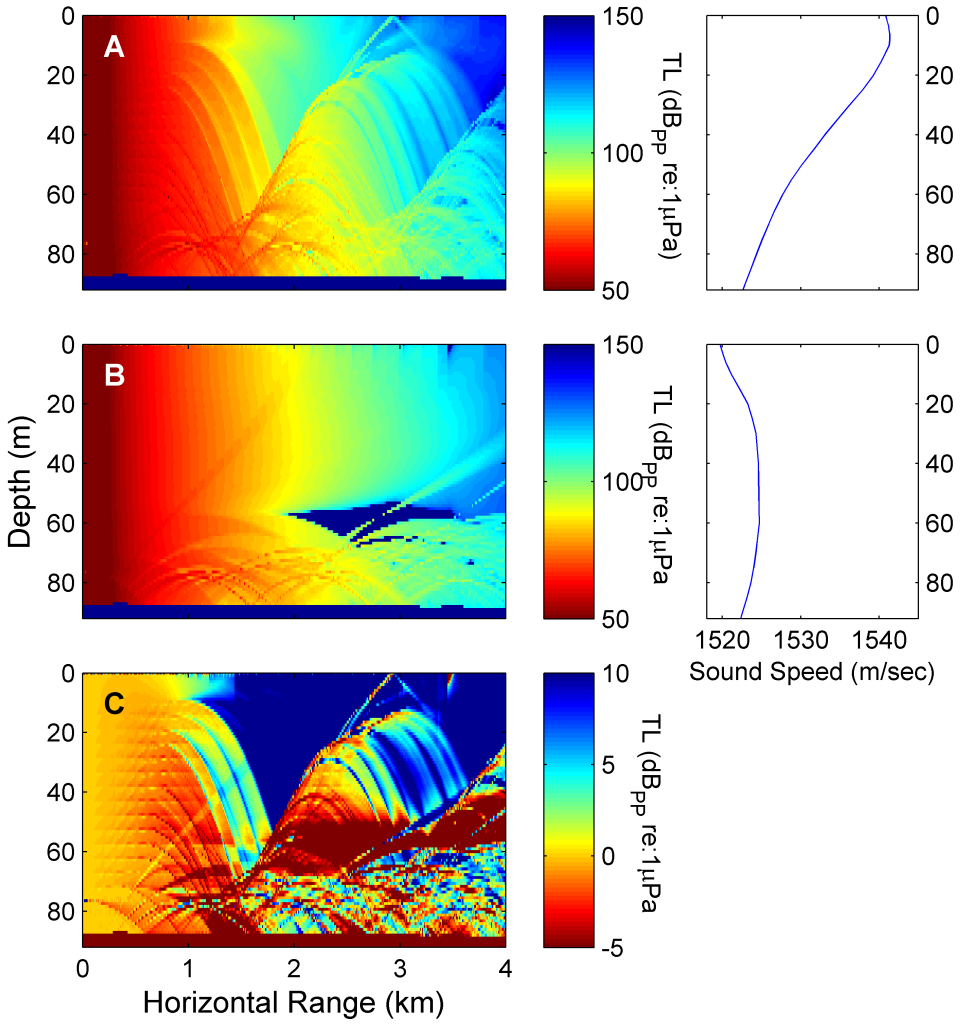


Figure 2.5: Comparison of modeled transmission loss (TL) at site MP. A) Average summer TL and sound speed profile; B) Average winter TL and sound speed profile. In A and B, red indicates areas of low TL, and blue indicates high TL. These plots can be interpreted using the theorem of reciprocity (Rayleigh, 1887), by imagining the sensor as located at the seafloor where horizontal range = 0 km. Estimated TL between the sensor and a source is then given by the color of the plot at the depth and range of the source.

C) Difference between summer and winter TL. Note that color scale is different from the top two panes. Dark blue areas at shallow depths indicate regions of the water column where sources are predicted to be less detectable in summer than in winter conditions. Note that these are the near surface areas where dolphins spend most of their time. Red areas indicate zones where sources are predicted to be more detectable in summer than in winter. Yellow areas indicate zones where detectability is expected to remain roughly constant across seasons.

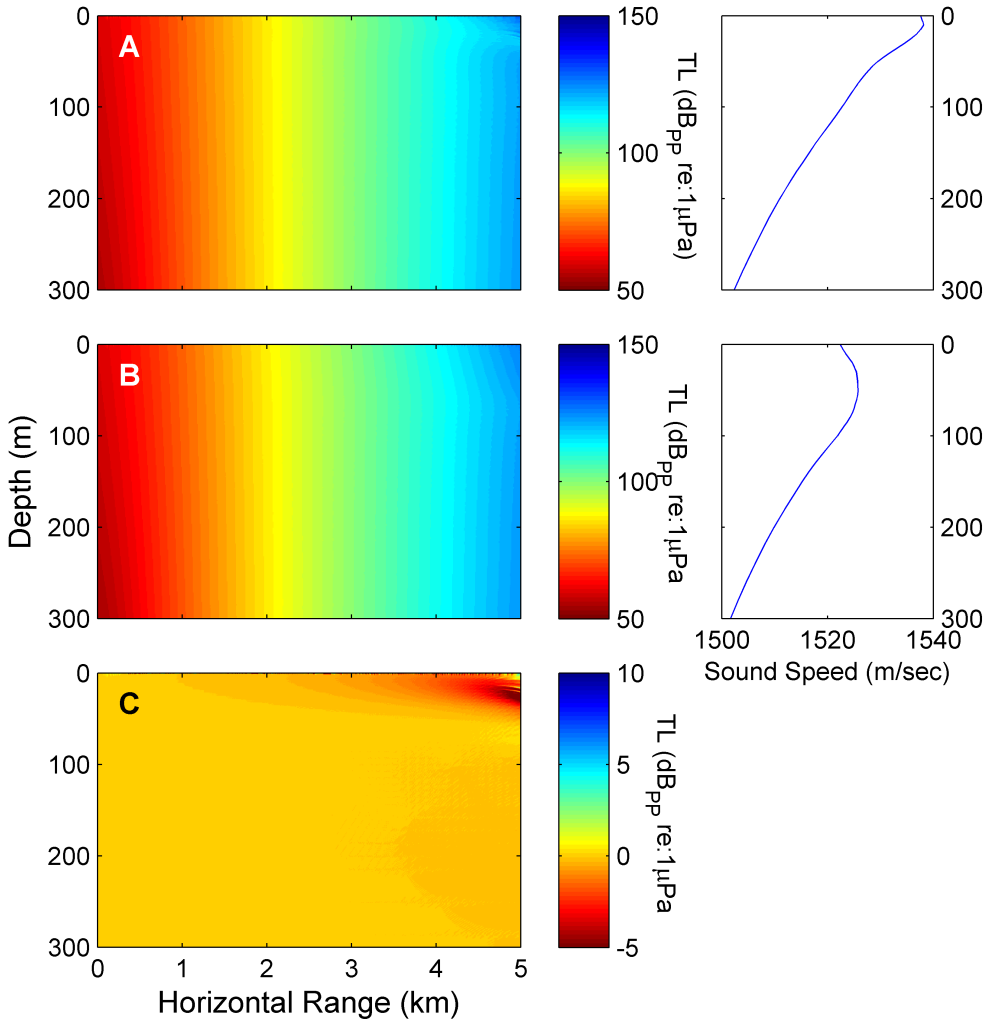
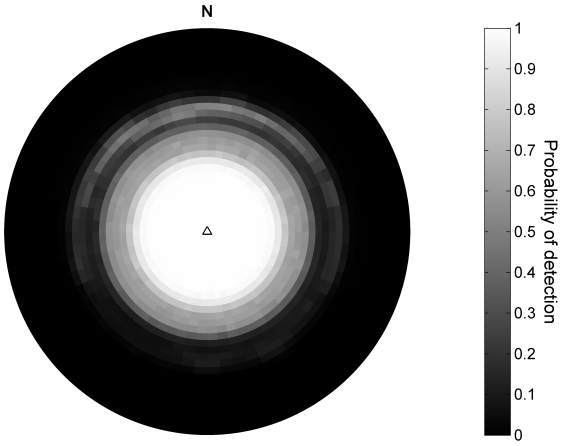
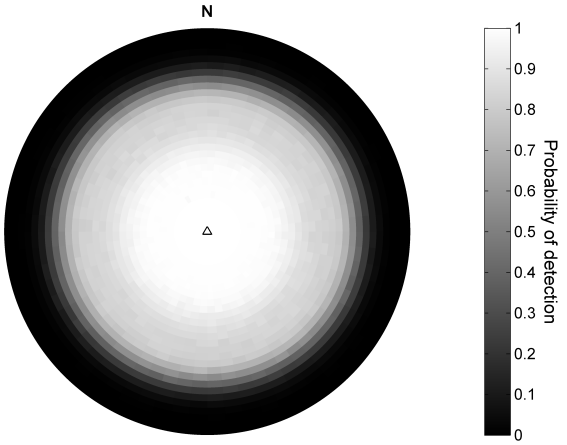


Figure 2.6: Comparison of modeled transmission loss (TL) at site MC. A) Average summer TL and sound speed profile; B) Average winter TL and sound speed profile. In A and B, red indicates areas of low TL, and blue indicates high TL. These plots can be interpreted using the theorem of reciprocity (Rayleigh, 1887), by imagining the sensor as located at the seafloor where horizontal range = 0 km. Estimated TL between the sensor and a source is then given by the color of the plot at the depth and range of the source.

C) Difference between summer and winter TL. Note that color scale is different from the top two panes. Dark blue areas at shallow depths indicate regions of the water column where sources are predicted to be less detectable in summer than in winter conditions. Red areas indicate zones where sources are predicted to be more detectable in summer than in winter. Yellow areas indicate zones where detectability is expected to remain roughly constant across seasons. A small increase in detectability at the most extreme ranges is seen in summer.



(A) MP



(B) GC

Figure 2.7: Examples of asymmetric (A) and symmetric (B) detectability patterns at two sites for deep diver group models in summer months. Asymmetric detection probabilities occur at site MP under certain conditions, such as strong thermocline. This can lead to biased density estimates if animals are not uniformly distributed around the sensor. Symmetric detectability patterns, such as that shown for GC, are common for models at deeper sites.

broad variability incorporated into the model framework enables us to produce an estimate of detectability with the understanding that further refinements will be incorporated into the existing model frameworks as data become available.

2.5.3 Propagation Model

There are a number of caveats to consider when interpreting the results of the ray-tracing model. First, ray tracing doesn't take phase into account, so it does not predict interference or diffraction. The number of rays that can be traced is also limited. More rays require more computation time, but reduce the likelihood of artifacts in the model due to gaps between calculated rays. In general, the three thousand rays used for each radial in these models provided good coverage of the water column. The shallowest slope site, MP, appeared to be the most vulnerable to artificial blind spots created by ray gaps Figure 2.5. This issue is expected to have a minor effect on propagation predictions, and therefore minor effects on detection probabilities and density estimates.

Bellhop also has known issues with modeling sea surface boundaries (Porter and Liu, 1994). Although sea surface roughness is included in the model, this captures only small scale features, not the constantly changing swells and surf. The effect of this weakness is unclear. In general, we do not expect indirect clicks to be detectable at deep sites, and this is supported by the fact that we rarely see multiple arrivals of the same click. However multiple arrivals are common at the slope sites, therefore poor handling of the surface boundary may have a larger effect there. Multiple arrivals are problematic because they can inflate detection counts is each arrival is counted as a separate detection. This could lead to an overestimate of animal density. Handling of multiple arrivals at slope sites is explored further in Chapter 4.

2.6 Conclusion

The probability of detecting delphinid clicks was computed for shelf and slope sites using both cue and group counting methods. Simulations predict that groups are between four and ten times more detectable than individual clicks, depending on the dive behavior and echolocation signals of the species of interest. Large, deep-diving delphinids including Risso's dolphin and pilot whales are expected to be detectable at greater ranges than smaller, shallow-diving species including Stenellid dolphins. Diel differences in dive depths within species are not predicted to have significant effects on detectability. Seasonal effects on detectability were not significant for the cue-counting method, but are predicted to significantly affect group detection probabilities at shallow sites. In general detectability is more dynamic at shelf sites and more consistent at slope sites. Modeling provides a good first order estimate of click and group detectability at all sites, however more detailed descriptions delphinid of echolocation click parameters and behaviors are needed to improve model reliability. Inputs to the model framework described can be updated as further data become available.

Chapter 3

Ground-truth of detectability models using acoustic localization

3.1 Abstract

A series of delphinid encounters were recorded during a short term deployment of a HARP equipped with multiple hydrophones in a three dimensional configuration. One encounter consisted of a small group of Risso's dolphins (*Grampus griseus*), and the others were groups of Stenellid dolphins (*Stenella* sp.). After verifying that the multi-channel HARP data could be used to accurately track a vessel of known location, the delphinid encounters were localized based on the time difference of arrivals (TDOAs) of individual delphinid clicks across each pair of hydrophones. Horizontal detection range (HR), received level (RL) and estimated source level (SL) distributions from localized encounters were compared with predicted distributions obtained from species-specific Monte Carlo simulations. Probability of vocalization was examined for both species by comparing expected click counts to detected counts. Agreement between the localization-based results and model predictions suggests that the simulation described in Chapter 2

can be used to estimate detection probabilities when localization is not an option.

3.2 Background

The models presented in Chapter 2 provide estimates of species detectability at each of the HARP sites monitored in the Gulf of Mexico (GOM), in the absence of experimental data. However, the models are only as good as their inputs and our understanding of the parameters affecting detectability, which are fairly simplified. Our assumptions and predictions need to be verified by comparing the model predictions to data.

Here I present a ground truth case in which model predictions are compared with experimental results. This was was accomplished using a short-term deployment of acoustic sensors with the ability to localize echolocating animals. Localizations were used to compare actual distributions of signal parameters to those predicted by the models. Ideally, all passive acoustic monitoring (PAM) devices would be made up of multiple hydrophones in a three-dimensional configuration, to enable localization of all detections. This would render detectability modeling unnecessary. However data storage and battery-life limitations make this currently unfeasible at the bandwidths needed for odontocete clicks.

3.2.1 Acoustic Localization

Acoustically-based marine mammal tracking efforts began in the 1970s with the development of a three-dimensional hydrophone array to record cetacean vocalizations by Watkins and Schevill (1971), who applied the technique to spinner dolphins (Watkins and Schevill, 1974). Since then, similar tracking techniques have successfully located and tracked large cetaceans, using far-ranging, stereotyped, low-frequency blue, fin and

humpback whale calls (Clark, 1995; McDonald et al., 1995; Tiemann et al., 2004). Semi-omnidirectional, broadband sperm whale clicks have also been used for localization (e.g. Giraudet and Glotin, 2006). More recently, beaked whale and delphinid echolocation clicks have been localized using closely spaced, time-synchronized hydrophones (Gassmann et al., 2013, 2011; Wiggins et al., 2012).

Well-established acoustic localization methods use the time difference of arrival (TDOA) of a signal on multiple hydrophones in a known spatial configuration to estimate a source position (e.g. Watkins and Schevill, 1971). If sensors are positioned in a three dimensional configuration, then both azimuth and elevation between the sensor and source can be determined trigonometrically. Determining source depth typically requires at least two multi-channel instruments. For localization using echolocation clicks, hydrophones generally need to be closely spaced so that the same click can be received and identified on all channels. This is primarily because delphinids often click rapidly and simultaneously. Click association across sensors becomes difficult if TDOAs can be larger than the click time between successive clicks. (A probability-based method for associating clicks in such cases is described by Baggenstoss (2011)for a small group of four sperm whales, but it remains to be seen whether the method can be applied to larger groups of dolphins).

Although closely-placed sensors are deemed necessary for click localization, there are disadvantages to this kind of small aperture array. These include high sensitivity to time resolution and sensor depth errors, as well as compromised accuracy at large horizontal ranges.

3.2.2 Vocalization Rate Estimates

A key variable in density estimates using passive acoustics is the likelihood that an animal is actively vocalizing, and therefore available for detection (see Chapter 7 for

further discussion). Probability of vocalization is a difficult value to pin down, because it is likely influenced by many variables, including group size, behavior and environment, which are highly dynamic. Data in which animals can be tracked provides an opportunity to examine vocalization probability in a small set of cases. If the species-specific click rate while actively vocal is known, and the number of animals present can be estimated, then the number of detections obtained when the animals are within detection range can be compared to the number of detections expected if all animals were actively echolocating at all times (e.g. Van Parijs et al., 2002). This ratio provides a preliminary estimate of probability of vocalization.

3.2.3 Goals of this Chapter

In this groundruth example, two different species were encountered and localized. The first was Risso's dolphin (*Grampus griseus*). This species has a unique click type (Soldevilla et al., 2008) which is easily recognized and distinguished from other Gulf of Mexico delphinids based on frequency content. The second localized species was *Stenella* sp., identified from inter-click interval and click frequency content (see Chapters 5 and 6 for details).

The primary goal of this chapter is to confirm that model predictions of detectability are in agreement with in situ results. This is accomplished by comparing predicted and actual distributions of received level (RL), source level (SL), and horizontal range (HR) for detected clicks. Because the models are simplified relative to reality, and are designed to simulate average detectability over a large number of detections and conditions, we do not expect a perfect match between the two. Rather we are looking for general agreement as an indication that model predictions are aligned with reality. A secondary goal is to obtain a preliminary estimate of probability of vocalization.

3.3 Methods

3.3.1 Data Collection

A four-channel HARP was deployed near the Dry Tortugas site (DT) for approximately 18 hours in November 2013 (Figure 1.1). The instrument was equipped with four hydrophones in a tetrahedral configuration, each sampled at 100 kHz (Figure 3.1; Wiggins and Hildebrand, 2007; Wiggins et al., 2012). Each hydrophone was recorded on a separate channel. Hydrophones were affixed to a low-reflectivity plastic frame to minimize signal reflections. Maximum distance between hydrophones was 104 cm when measured prior to deployment. Floats attached above the hydrophone frame held it vertically above the seafloor. The frame was able to rotate freely in the horizontal plane. A test deployment in southern California indicated that the frame rotated 180° around its vertical axes on a tidal cycle, but was generally stable over shorter periods of time. Due to tilt sensor failure at the DT site, frame rotation during the Gulf of Mexico deployment is not known deployment.

Instrument position after settling on the seafloor was computed by circumnavigating the instrument while sending 12 kHz pings to it from a towed transducer (Wiggins et al., 2012). The instrument responded with pings from an on-board transducer. The two-way travel times (interrogation and response), assuming a mean sound speed of 1490 m/s, were used to solve for instrument position and depth to within 8 m (e.g. Creager and Dorman, 1982).

3.3.2 TDOA Calculation

A hemispherical grid of expected signal time TDOAs was computed as a function of azimuth and elevation relative to hydrophone 1 (one of the bottom phones on the tetrahedral frame, Figure 3.1). This resulted in a set of six theoretical TDOAs for each grid

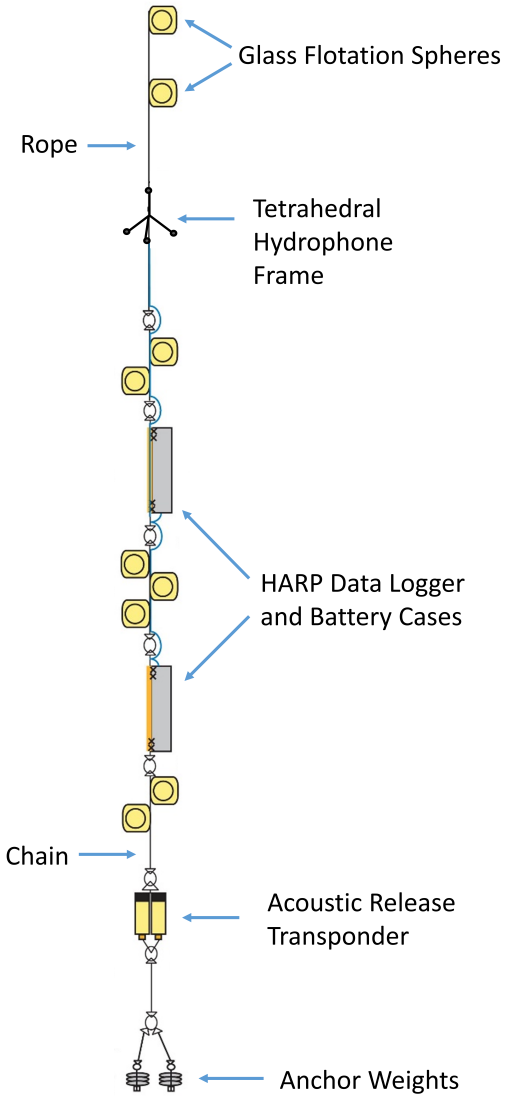


Figure 3.1: Diagram of four-channel HARP setup and tetrahedral hydrophone configuration. Image and instrumentation by Sean Wiggins.

position on the hemisphere. Grid spacing was in one degree increments, with elevation ranging from zero to 90° (directly above the instrument) and azimuth ranging from zero to 359°. A sound speed of 1500 m/s was assumed between phones, based on typical sound speeds for that site and depth (Oceanographic and Atmospheric Master Library, OAML, US Navy, 2009). To localize a source given a set of experimentally-derived TDOAs, the total difference between theoretical and actual TDOAs was computed for all grid locations. The elevation and azimuth of the position with the minimum difference was used to generate an estimated source position, assuming that the source was located near the sea surface.

3.3.3 Ship Localization

Ship position was computed by localizing the HARP interrogation pings originating from the vessel. Pings were transmitted at 11 kHz. A second order infinite impulse response resonator filter with peak frequency of 11 kHz and a 250 Hz bandwidth was applied to the recording data (MATLAB DSP Systems Toolbox, Version R2012b). The onset of the ping was defined as the moment the filtered signal rose above 50 counts. TDOAs were obtained by differencing the onset times across the four channels. The elevations obtained from the minimization process were translated into horizontal range from the sensor using the depth of the instrument, as calculated from the two-way travel times. The ship's transducer was assumed to be at the sea surface for localization purposes.

Azimuth obtained from the minimization process provided a relative heading for the ship over the tracking period. A constant rotation was applied to align the relative heading with the actual position of the ship, known from GPS data. The magnetometer aboard the instruments failed at depth, therefore true azimuthal orientation of the frame was unknown. It was assumed that the instrument was stationary and did not rotate during

the tracking period. Ship tracks obtained from the localization process were compared with those from the GPS receiver located at mid-ship.

3.3.4 Dolphin Localization

During the deployment, one 60-minute Risso's dolphin (*Grampus griseus*) encounter and four consecutive *Stenella* sp. encounters over a 3.5 hour period were recorded and tracked. Delphinid echolocation clicks were automatically detected on all channels (see Chapter 4 for detector details). The small array aperture made it possible to identify individual clicks across all four channels because the travel time between sensors (less than one millisecond) was less than the dolphin's interclick interval (>50 ms).

Clicks were matched automatically across channels by associating those which were temporally closest to one another. A TDOA set was computed for each detected click by computing the TDOA of the click amplitude maximum between sensor pairs. Elevation angles for the animals were translated into horizontal ranges by assuming that the animals were located at the sea surface. A second set of horizontal ranges were computed for Risso's dolphins assuming that animals were diving to 200 m, which is an intermediate dive depth for these animals (Wells et al., 2009).

3.3.5 Comparison with Model Predictions

Risso's dolphin and *Stenella* sp. SLs were computed from localized clicks by adding estimated transmission loss to RLs using the sonar equation (Urick, 1967):

$$SL = RL \cdot 20\log(r) + \alpha \cdot r^2 \tag{3.1}$$

where r is slant range in meters, and α is attenuation at peak frequency in dB/m. Estimated SLs were compared with literature values for consistency. Delphinid detection range

and RL distributions were compared to model predictions. Transmission loss within the simulation was computed for both sensors using typical environmental parameters for the site in month of November (OAML, US Navy, 2009).

3.3.6 Vocalization Rate Estimates

Vocalization rate estimates $\hat{P}_v$ were computed by estimating the number of animals n in a tracked group and comparing the expected number of clicks $\hat{c}_{\text{exp}}$ to the actual number of clicks detected c_{act}, such that

$$\hat{P}_v = \frac{c_{\text{act}}}{\hat{c}_{\text{exp}}} \tag{3.2}$$

The expected number of clicks is the product of the mean probability of detecting a click P_{det}, the mean species-specific click rate r (chapter 5), and n:

$$\hat{c}_{\text{exp}} = P_{\text{det}}\, r\, n \tag{3.3}$$

This calculation was done for a series of five minute time bins. For Risso's dolphin, the number of animals present in a five minute bin could be determined based on tracks, therefore n was adjusted to reflect this number for each bin. For Stenellid dolphins, animals could not be individually counted, therefore an average group size was used (chapter 7; Mullin and Fulling, 2004) to estimate expected click counts, and then an average $\hat{P}_v$ was computed across all bins.

Variances associated with P_v were computed using the delta method (Seber, 1982; Powell, 2007) :

$$\text{var}(P_v) = \text{var}(N_{act}) \cdot \left(\frac{1}{N_{exp}}\right)^2 + \text{var}(N_{exp}) \cdot \left(\frac{1}{N_{act}}\right)^2 \tag{3.4}$$

where

$$\mathrm{var}(N_{act}) = \mathrm{var}(D) \cdot (\hat{c}_k)^2 + \mathrm{var}(\hat{c}_k) \cdot \mathrm{mean}(D)^2 \tag{3.5}$$

and

$$\mathrm{var}(N_{act}) = \mathrm{var}(\hat{g}) \cdot (c_r P_k)^2 + \mathrm{var}(c_r) \cdot (\hat{g} P_k)^2 + \mathrm{var}(P_k) \cdot (\hat{g} c_r)^2 \tag{3.6}$$

3.4 Results

3.4.1 Harp and Ship Localization

Instrument latitude, longitude and depth were determined from the two-way travel times of interrogation pings sent from the ship (Table 3.1). The instrument drifted with the current between its surface release location and landing site on the seafloor. Azimuth and elevation estimates from the TDOA minimization scheme used to localize the ship's transducer generally agreed with the theoretical values expected based on the vessel's GPS coordinates (Figures 3.3 and 3.4). Acoustically-derived elevation estimates closely matched expected values. Azimuthal estimates deviated from expected values based on actual vessel position during the first part of the circumnavigation period, but matched closely for the second half of that period. This indicates that the tetrahedral frame was rotating, during the first part of the tracking period contrary to our expectations.

Calculated horizontal distances between the HARP and ship are comparable with expected ranges based on GPS positions. Timing differences between the GPS and HARP localizations are due to instrument rotation (Figure 3.4). The tracklines obtained from horizontal range estimates are similar to actual tracklines.

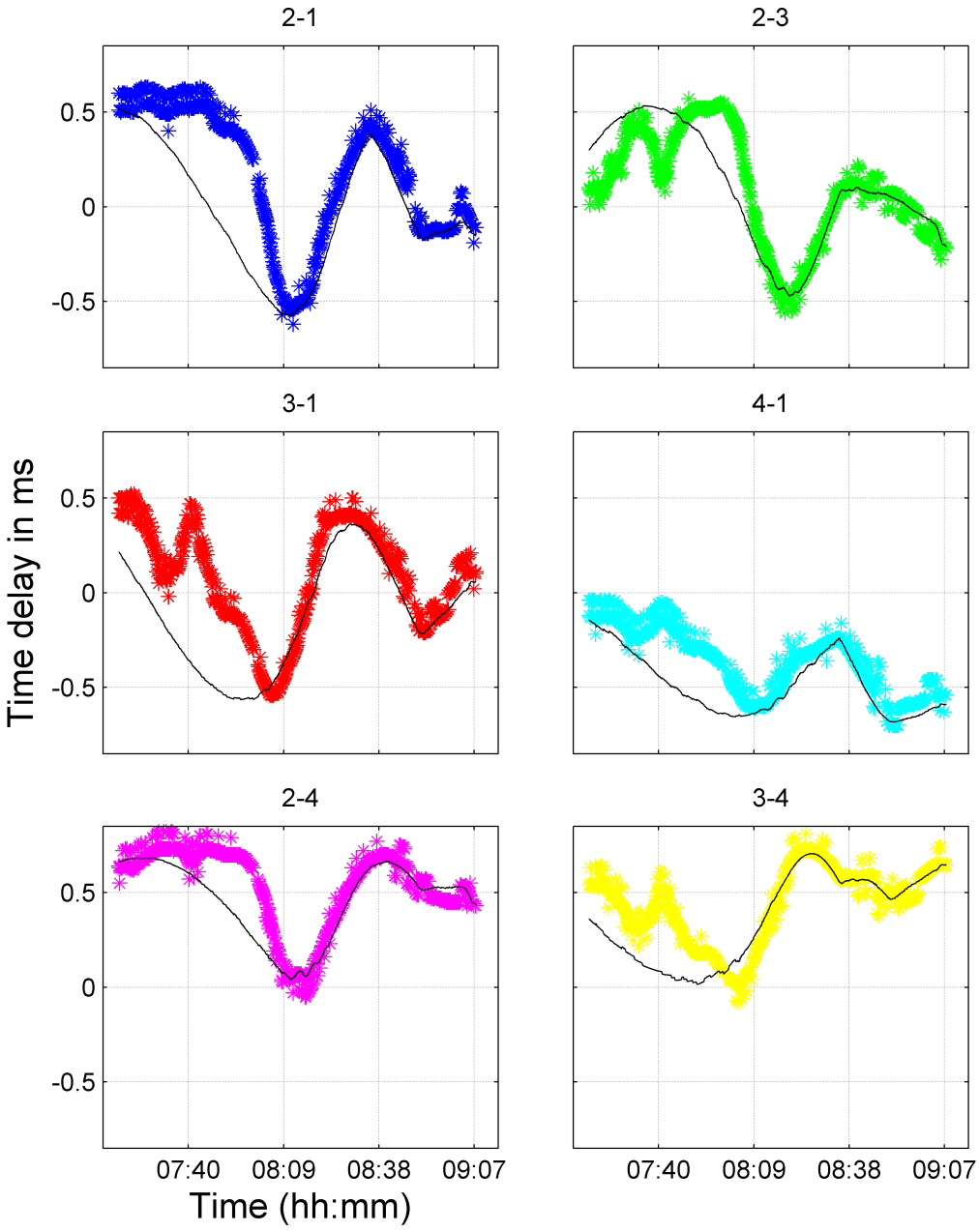


Figure 3.2: Ping onset TDOAs across sensor pairs on multi-channel sensor. Colored points indicate actual TDOAs. Black lines indicate expected TDOAs based on GPS position of ship and known array configuration. All times are in GMT. Local time is GMT -05:00.

Table 3.1: Position and depth of sensor on seafloor based on a minimization of two-way interrogation ping travel times.

Site	Latitude	Longitude	Depth (m)
HARP B	25.5366N	84.6320W	1220

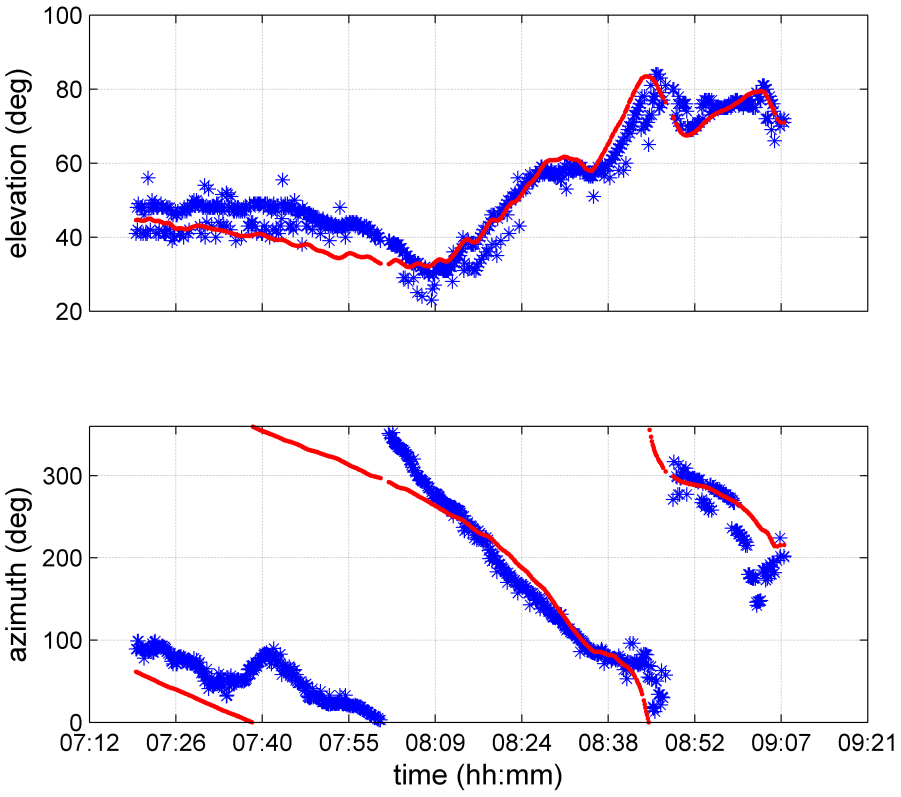


Figure 3.3: Theoretical (red) and actual (blue) elevation and azimuth of ship relative to sensor. Azimuths are relative. Elevation of 90° indicates that ship is directly above the sensor. All times are in GMT. Local time is GMT - 05:00.

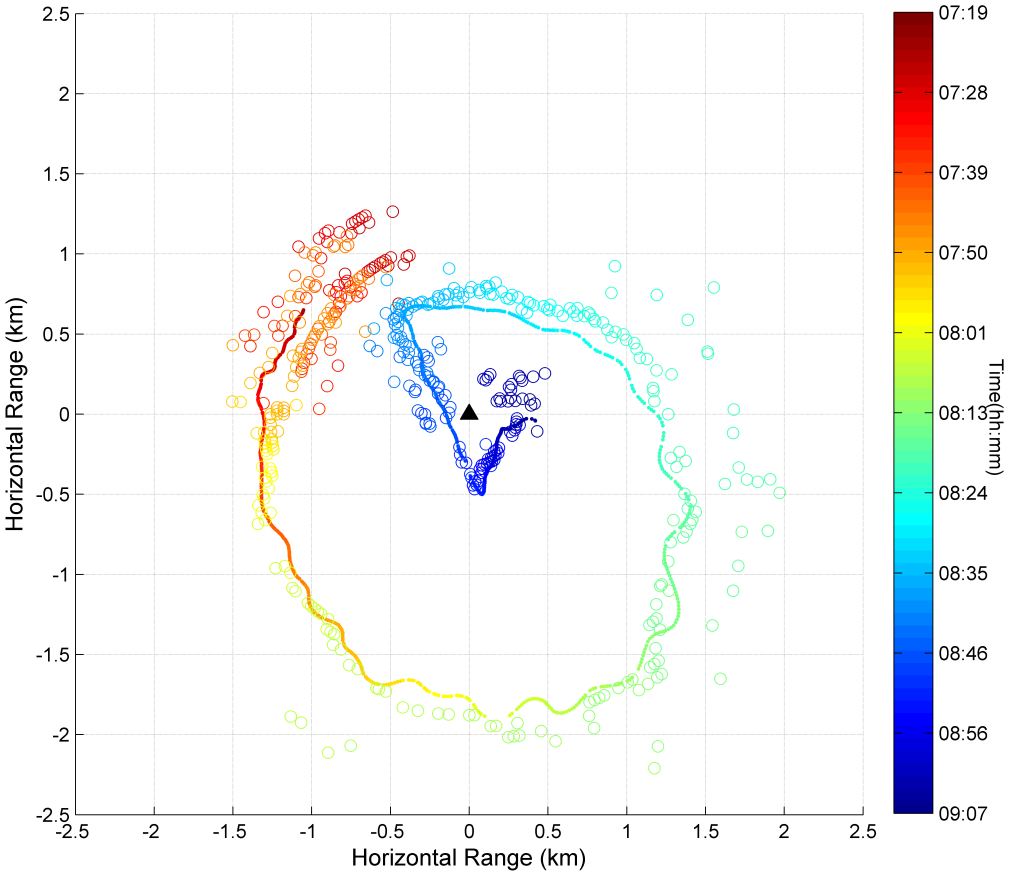


Figure 3.4: Ship position estimates (colored circles) based on localization using time difference of arrivals (TDOAs) of transducer pings from the ship on a multi-channel HARP. Colored line indicates actual ship position based on GPS. Color indicates the time associated with each position, in GMT, with red indicating the beginning of the vessel tracking period, and blue indicating the end of the tracking period. Local time is GMT - 05:00. The HARP position on the seafloor is indicated by the black triangle at (0,0).

3.4.2 Dolphin Localization

Risso's Dolphin Encounter

A total of 509 Risso's dolphin echolocation clicks were localized from the HARP data (Figures 3.5 and 3.7) at a maximum horizontal range of 3.0 km (Table 3.2). High SLs allowed the animals to be detected at elevation angles as low as 20° (where 90° is directly above the instrument). Localization range may have been limited by vertical array aperture.

The maximum detection range predicted by the model was 4.6 km, however large ranges would be rare in practice, therefore it is not unexpected that the theoretical maximum detection range is not seen in this small dataset. RL and horizontal detection range distributions match fairly well between the modeled and in situ data when animals are assumed to be located at the sea surface (Figure 3.8), however the match improves if animals are assumed to be diving to 200 m (Figure 3.9). In situ elevation angle distributions also match modeled data more closely if the animals are assumed to be at 200m (Figures 3.10 and 3.11). In situ RLs reached a maximum of 138 $dB_{pp}re : 1\mu Pa$ with most received clicks much lower in amplitude (Table 3.2). The modeled maximum RL was much higher, at 163 $dB_{pp}re : 1\mu Pa$, but again, it is not unexpected that this value is not seen in the in situ data. Mean in situ RLs were close to the modeled mean, around 122 $dB_{pp}re : 1\mu Pa$, with similar standard deviations.

SL distributions from the modeled data accurately predicted the in situ SL distribution back-calculated from the localizations (Table 3.2). In situ SL estimates ranged from 184 to 217 $dB_{pp}re : 1\mu Pa$. Maximum in situ SL estimates were lower than the theoretical maximum predicted by the model. The range of SLs estimated from the localizations was contained within the modeled maximum and minimum SLs (176 to 234 $dB_{pp}re : 1\mu Pa$). Mean SL was higher in the in situ data ($201 \pm 4\sigma$ $dB_{pp}re : 1\mu Pa$)

than in the modeled data ($196 \pm 4\sigma$ $dB_{pp}re : 1\mu Pa$) .

Multiple track lines are visible in the position data (Figure 3.7), suggesting a small group of five animals moving through the area (Figure 3.6). Larger numbers of clicks are detected in the first half of the encounter, when animals are moving toward the sensor. In the second half of the encounter the animals have passed their point of nearest approach to the sensor and are oriented away. Mean vocalization probability is estimated at 0.13 (CV = 0.37; Table 3.3).

Table 3.2: Comparison of modeled and in situ parameter distributions for Risso's dolphin encounters. The model distributions are based on 500 model iterations. HR indicates horizontal range. Subscripts indicate HARP data (H), or model value (M). All amplitudes are reported in $dB_{pp}re : 1\mu Pa$.

Parameter	$\mu \pm \sigma$	**Max**	**Min**
RL_H (dB)	$122 \pm 5\sigma$	138	115
RL_M (dB)	$122 \pm 5\sigma$	163	115
HR_H (km)	$1.3 \pm 0.6\sigma$	3.0	0.1
HR_M (km)	$1.1 \pm 0.5\sigma$	4.6	0.0
SL_H (dB)	$201 \pm 6\sigma$	217	184
SL_M (dB)	$196 \pm 6\sigma$	234	176

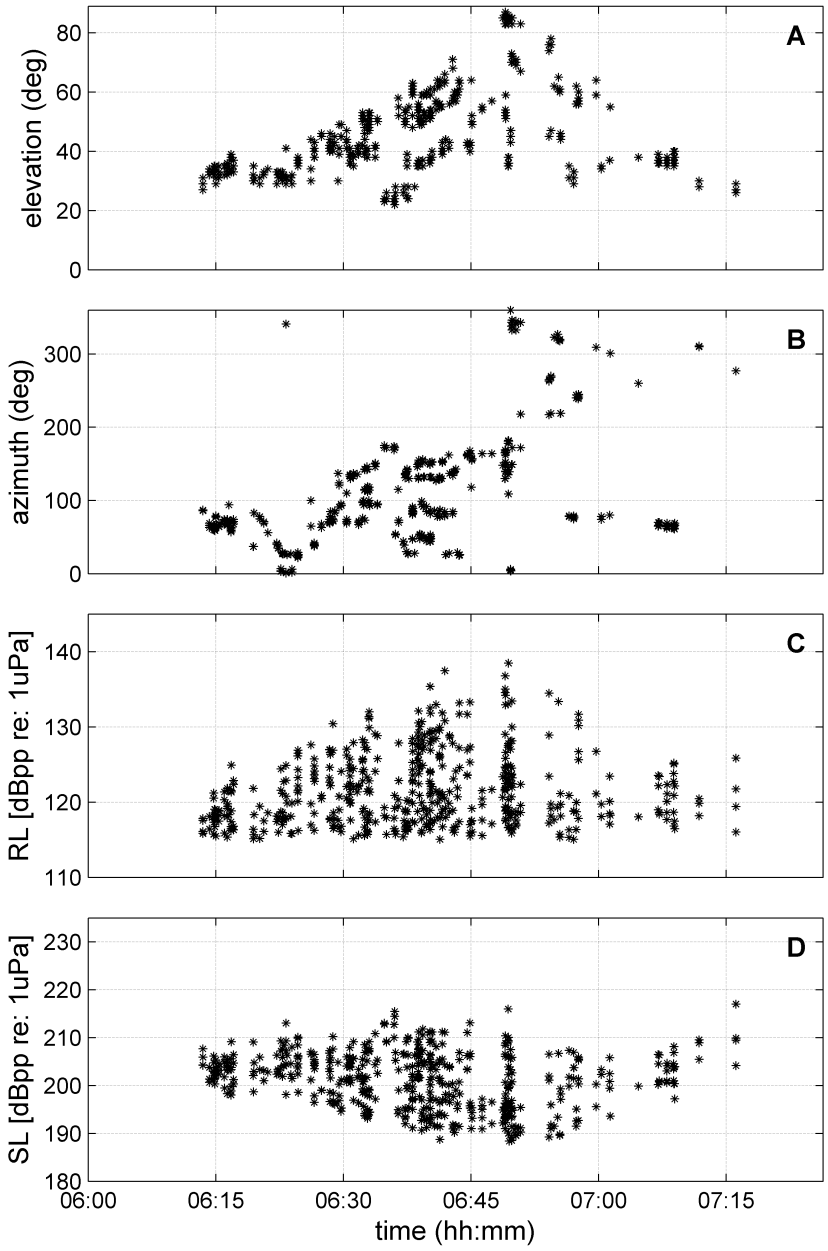


Figure 3.5: Risso's dolphin encounter localized by seafloor sensor. (A) Elevation relative to sensor. (B) Azimuth relative to sensor. (C) Received levels of localized clicks over time. (D) Estimated SLs of localized clicks based on received levels and position estimates. All times are in GMT. Local time is GMT - 05:00.

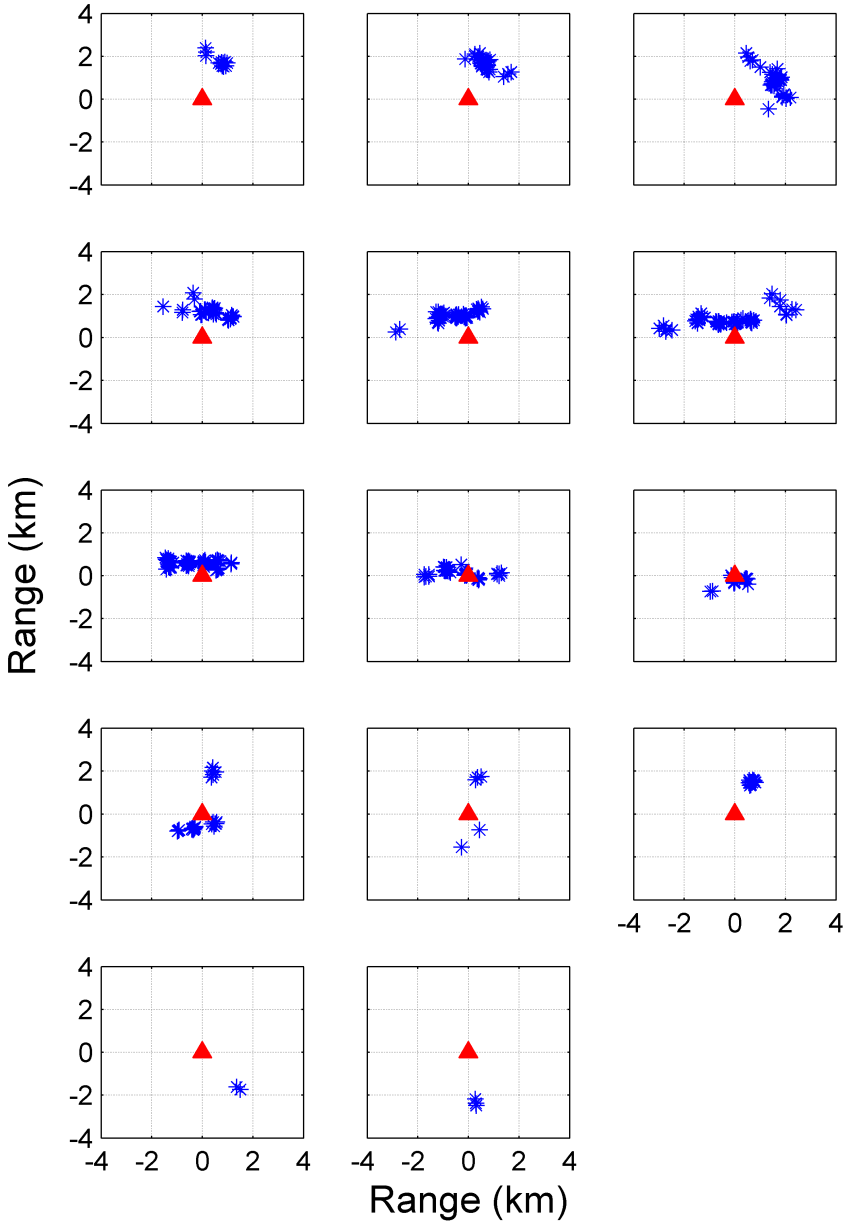


Figure 3.6: Risso's dolphin encounter localized by five minute bin. Blue stars indicate click localizations, red triangles indicate HARP position. Total group size is estimated at 5 animals, but all five are not within detection range at all times. Animals are assumed to be at the sea surface for localization purposes.

Table 3.3: Estimate of Risso's dolphin vocalization probability by five minute bin during a tracked encounter. The estimated number of animals (n) present during each time bin is used to estimate the expected number of clicks ($\hat{c}_{exp}$) that would be detected if all animals were actively vocal. This is compared to the actual number of clicks detected (c_{act}), to estimate the probability P_v that an individual animal is vocalizing.

bin #	**n**	$\mathbf{c_{act}}$	$\mathbf{\hat{c}_{exp}}$	$\mathbf{P_v}$
1	1	19	117	0.16
2	2	37	235	0.16
3	2	39	235	0.17
4	2	37	235	0.16
5	4	66	469	0.14
6	5	75	587	0.13
7	5	85	587	0.14
8	5	74	587	0.13
9	2	16	235	0.07
10	3	25	352	0.07
11	2	8	235	0.03
12	1	21	117	0.18
mean				0.13

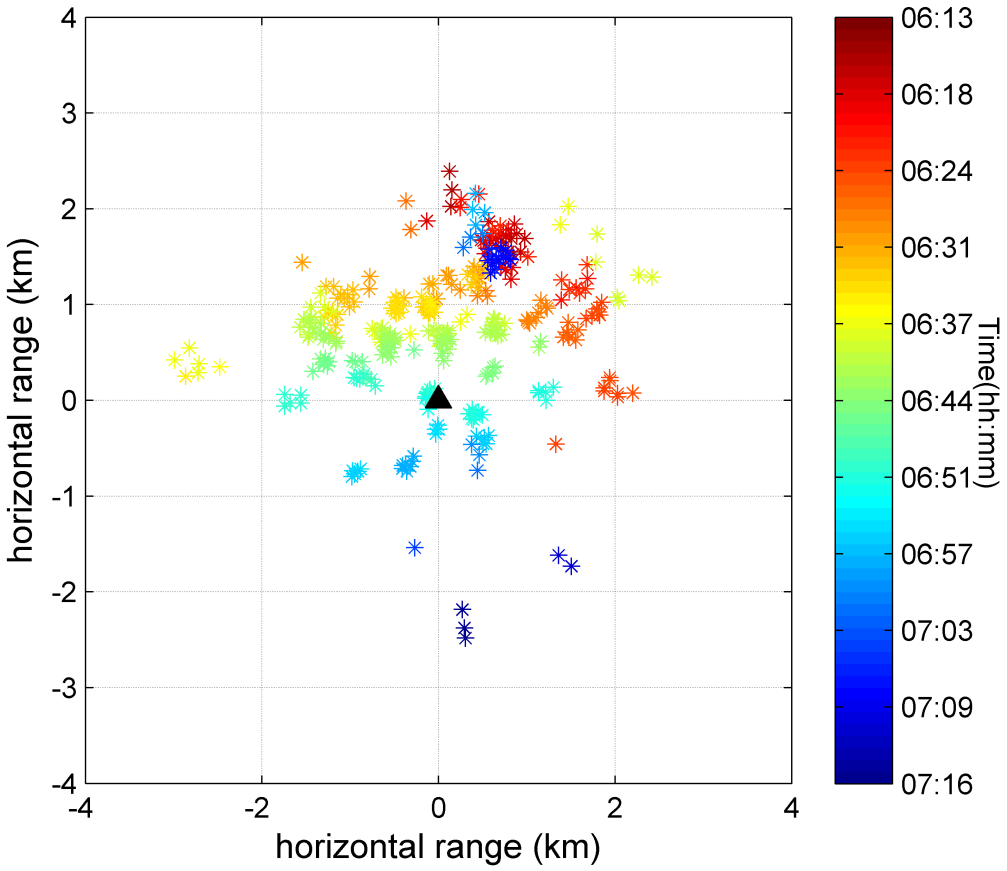


Figure 3.7: Bird's eye view of Risso's dolphin encounter as localized by seafloor sensor. Black triangle indicates tracking HARP position. Color indicates time in GMT.

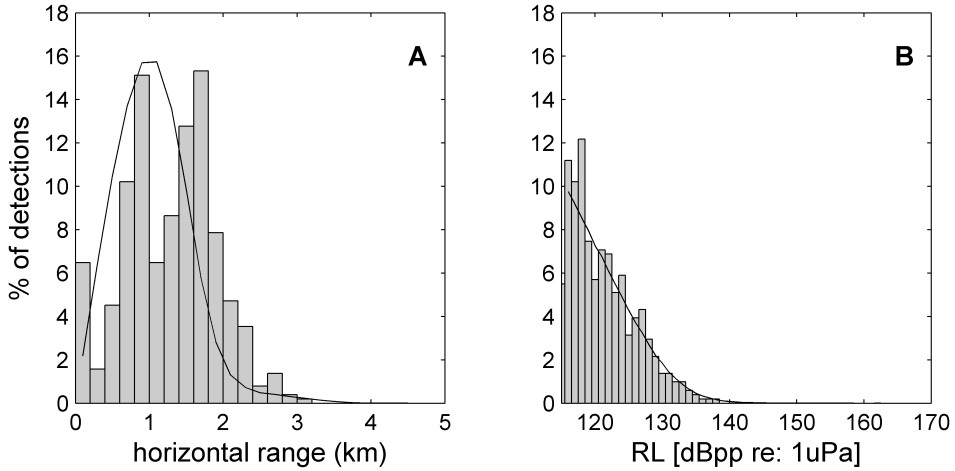


Figure 3.8: Risso's horizontal detection ranges (**A**) and received level distributions (**B**), assuming the animals are located at the sea surface. Grey bars indicate results from in situ data. Overlaid black line indicates model-based distribution.

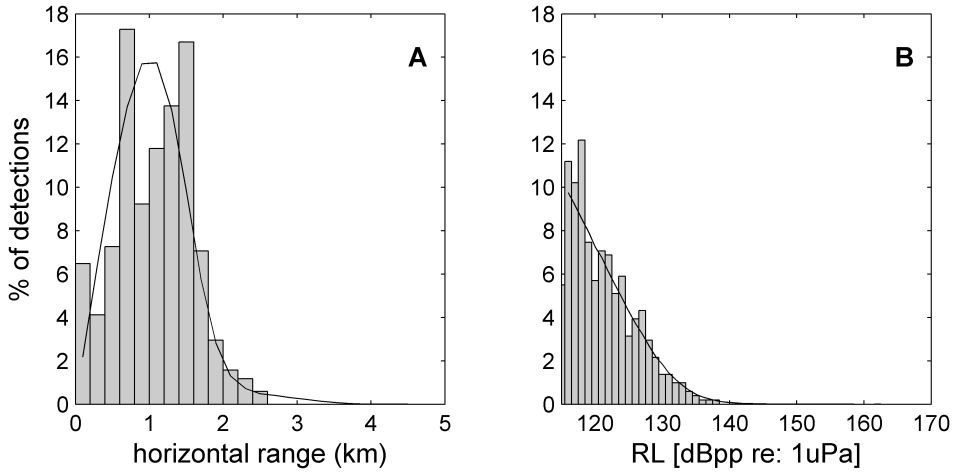


Figure 3.9: Risso's horizontal detection ranges (**A**) and received level distributions (**B**), assuming the animals are diving to 200 m. Grey bars indicate results from in situ data. Overlaid black line indicates model-based distribution. Agreement between in situ and model data improves if animals are assumed to be at depth.

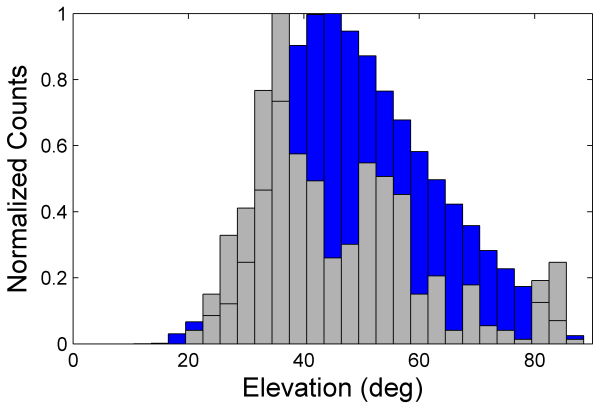


Figure 3.10: Blue bars indicate model-predicted distribution of elevation angles at which Risso's dolphin should be detected if randomly distributed in space, and animals are located at the sea surface. Grey bars indicate actual distribution of elevation angles at which Risso's dolphin were detected during this encounter. An elevation angle of 90° indicates that the animal was directly above the receiver.

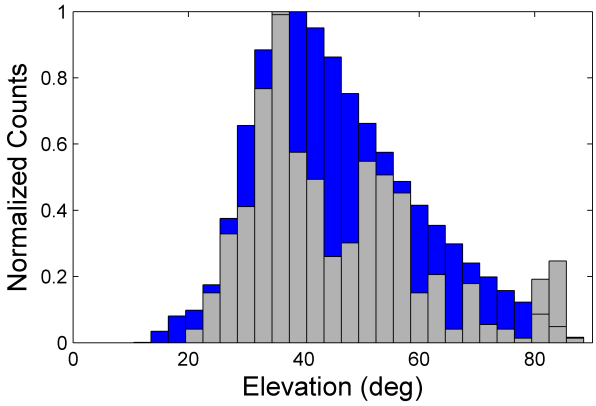


Figure 3.11: Blue bars indicate model-predicted distribution of elevation angles at which Risso's dolphin should be detected if randomly distributed in space, and diving to a depth of 200 m. Grey bars indicate actual distribution of elevation angles at which Risso's dolphin were detected during this encounter. Agreement between in situ and model data improves if animals are assumed to be at depth. An elevation angle of 90° indicates that the animal was directly above the receiver.

Stenellid Dolphin Encounters

Four successive Stenellid dolphin encounters were tracked. Encounter #1 was a small, tight group of animals traveling rapidly at approximately 6 km/hr (Figures 3.12 and 3.14) . Encounter #2 consisted of three small groups, one of which remained at the limit of the detection range. At least two of these groups were also traveling, but some milling behavior and looser association is seen in one group (Figures 3.15 and 3.17). Encounter #3 consisted of multiple large groups which appeared to merge and split over time (Figures 3.18 and 3.20) . Detection rates during this encounter were as high as 3,500 clicks per five minute bin. Encounter #4 was a small group of animals traveling through the area at roughly 4 km/hr (Figures 3.21 and 3.23).

The number of animals in each encounter could not be determined from the tracking data. If an average Stenellid dolphin group size of 60.7 animals based on visual data is assumed (Mullin and Fulling, 2004), P_v is estimated at 0.19 (CV = 1.6).

Stenella sp. echolocation clicks were localized at a maximum horizontal range of 2.3 km, slightly less than the 2.7 km maximum predicted by the model. Mean horizontal detection range and standard deviation was very similar to the model prediction. Most of the nearly 43,000 localizations associated with these encounters were within two kilometers of the sensor (Figure 3.24). The SL distribution from the ground truth data was had a higher mean and smaller standard deviation than predicted by the model results (Table 3.4). However, the distribution of SLs seen in the data was within in the range predicted by the model, suggesting that the difference may be attributable to a non-random uniform distribution of the animals in the area around the sensor.

Figure 3.25 shows that the distribution of elevation angles of localized clicks during this encounter does not perfectly match the modeled data for a uniform randomly distributed set of sources, although the two distributions have similar overall shapes. The majority of in situ Stenellid dolphin localizations were at an elevation of 50° or greater,

where 90° indicates that the animal is directly above the sensor. The number of counts above 70° begins to decline because area decreases. A tail of low number of detections extends out to low elevation angles, down to approximately 30°. At these angles, animals are not detected unless their beam is pointed at the hydrophone.

Received levels reached a maximum of 132 dB_{pp} re:1 μPa with most received clicks much lower in amplitude (Table 3.4). The model predicted a maximum RL of 150 dB_{pp} re:1 μPa, but this extreme value was not seen in the data. Mean RL was 118 ± 4σ dB_{pp} re:1 μPa, which was very similar to the model-predicted mean RL.

Table 3.4: Comparison of model and ground truth parameter distributions for *Stenella* sp. dolphin encounters. Model distributions are based on 500 model iterations. HR indicates horizontal range. Subscripts indicate HARP data (H), or model value (M). All amplitudes are reported in $dB_{pp}re : 1\mu Pa$.

Parameter	$\mu \pm \sigma$	**Max**	**Min**
RL_H (dB)	118 ±4σ	132	115
RL_M (dB)	119 ±3σ	150	115
HR_H (km)	0.6 ±0.3σ	2.3	0.0
HR_M (km)	0.6 ±0.3σ	2.7	0.0
SL_H (dB)	196 ±4σ	211	189
SL_M (dB)	188 ±6σ	221	176

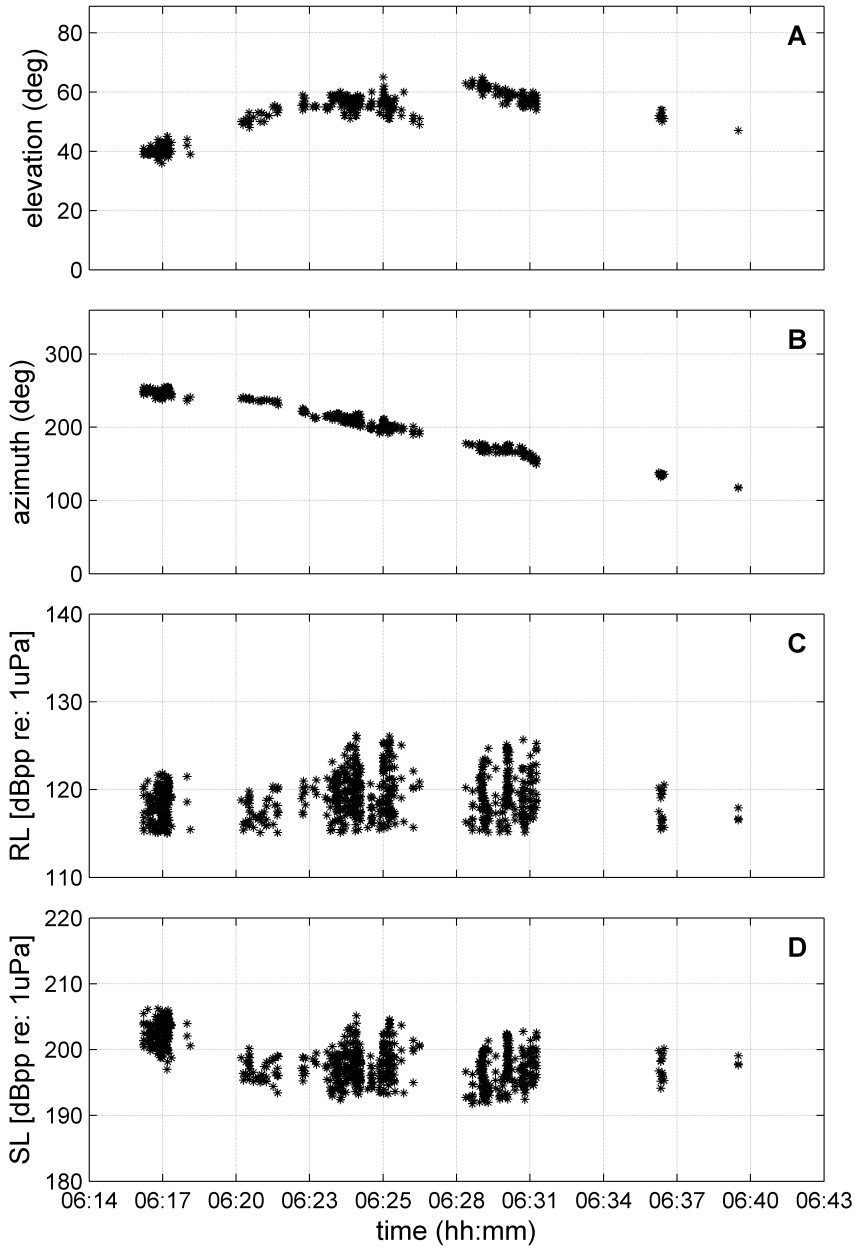


Figure 3.12: Stenellid dolphin encounter #1 localized by tracking HARP. (A) Elevation relative to sensor. (B) Azimuth relative to sensor. (C) RLs of localized clicks over time. (D) Estimated SLs of localized clicks based on received levels and position estimates. All times are in GMT. Local time is GMT - 05:00

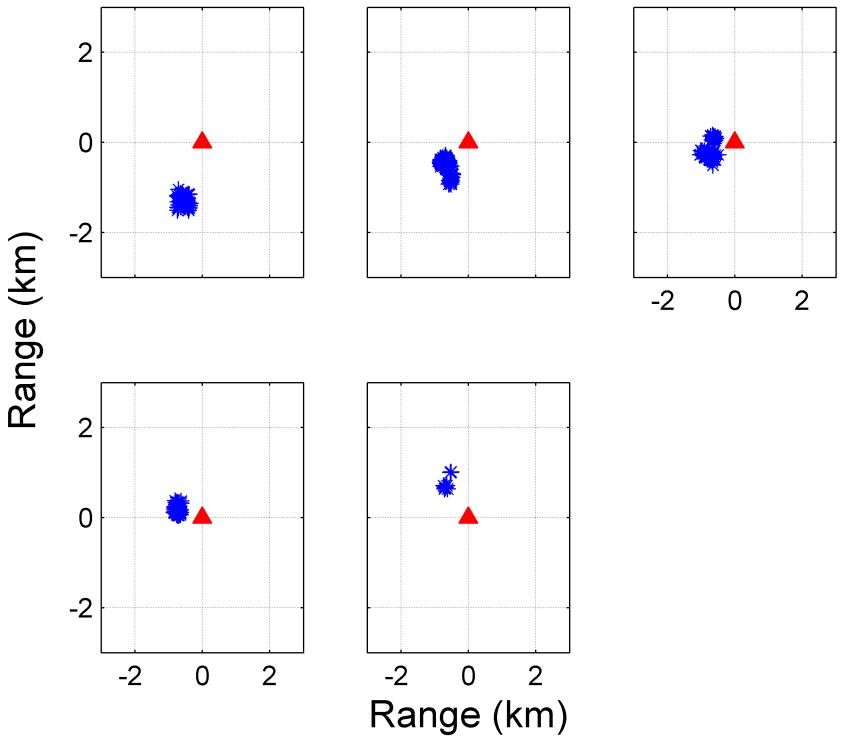


Figure 3.13: Stenellid dolphin encounter #1 localizations by five minute bin. Blue stars indicate click localizations, red triangles indicate HARP position. Animals are assumed to be at the sea surface for localization purposes.

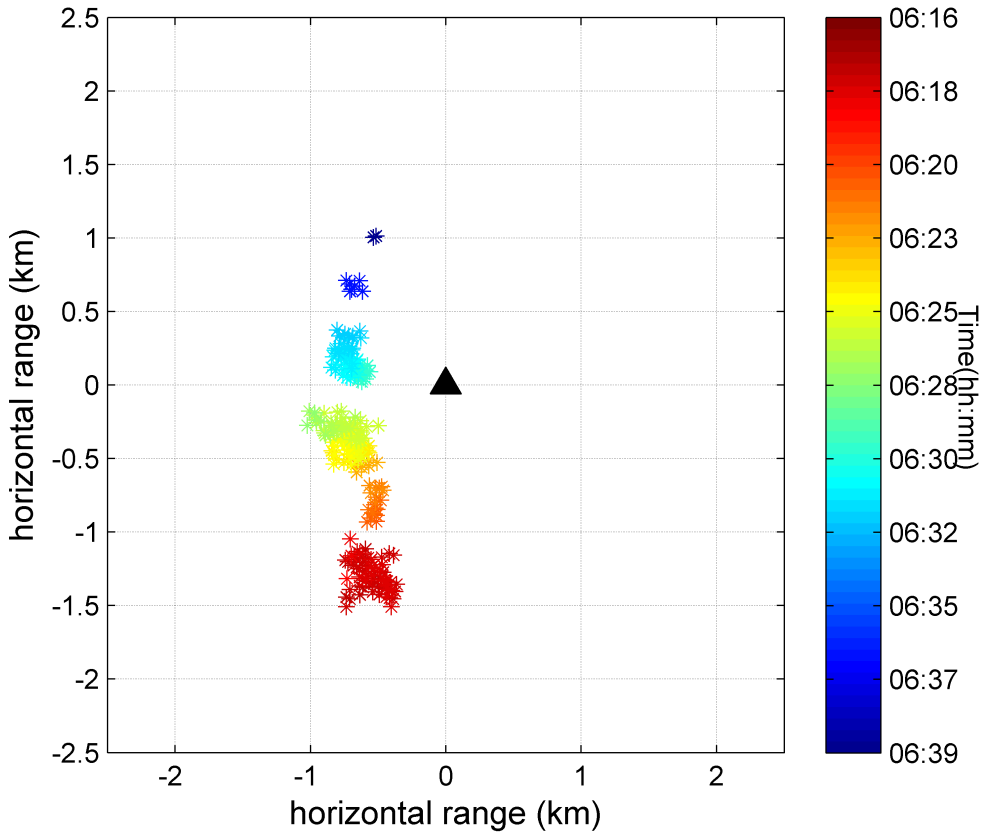


Figure 3.14: Bird's eye view of Stenellid dolphin encounter #1 as localized by seafloor sensor. Black triangle indicates the position of the tracking HARP. Color indicates time (GMT).

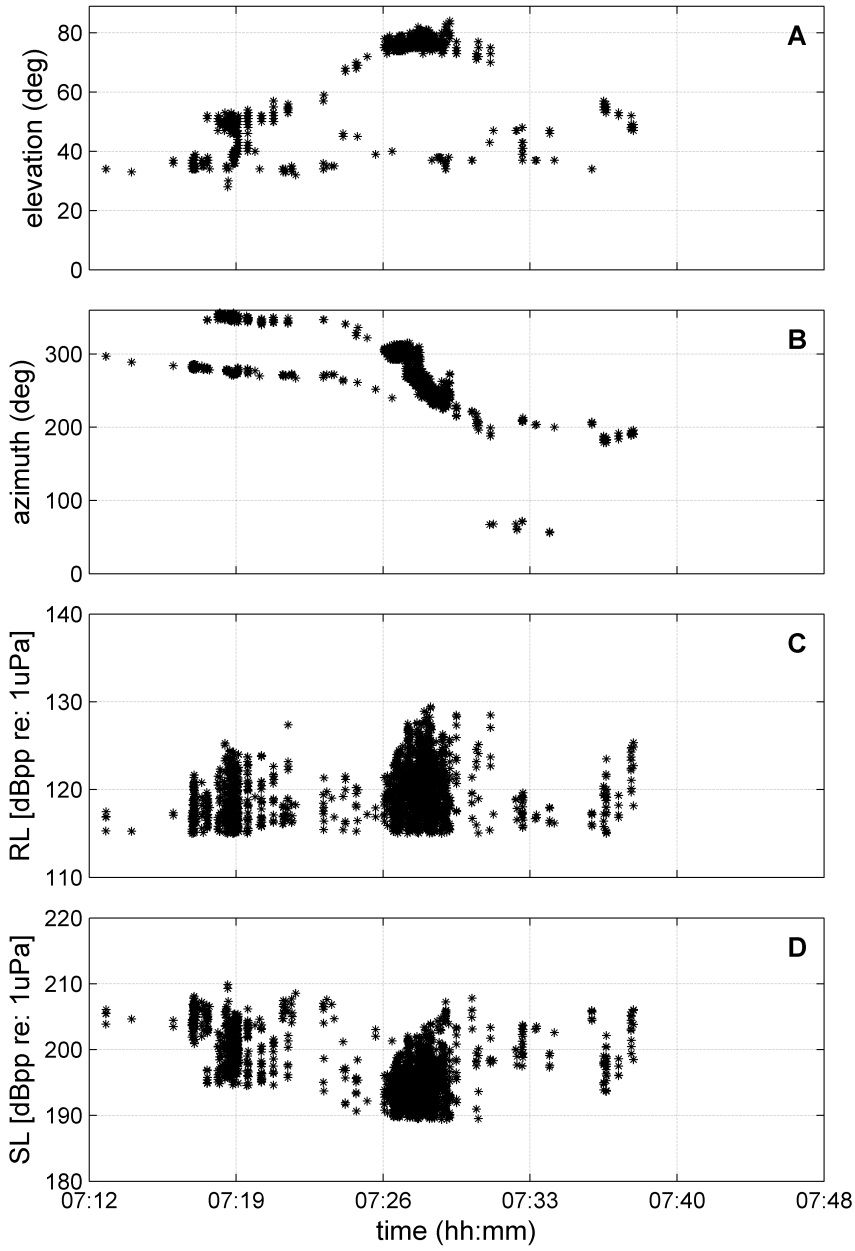


Figure 3.15: Stenellid dolphin encounter #2 localized by tracking HARP. (A) Elevation relative to sensor. (B) Azimuth relative to sensor. (C) RLs of localized clicks over time. (D) Estimated SLs of localized clicks based on received levels and position estimates. All times are in GMT. Local time is GMT - 05:00

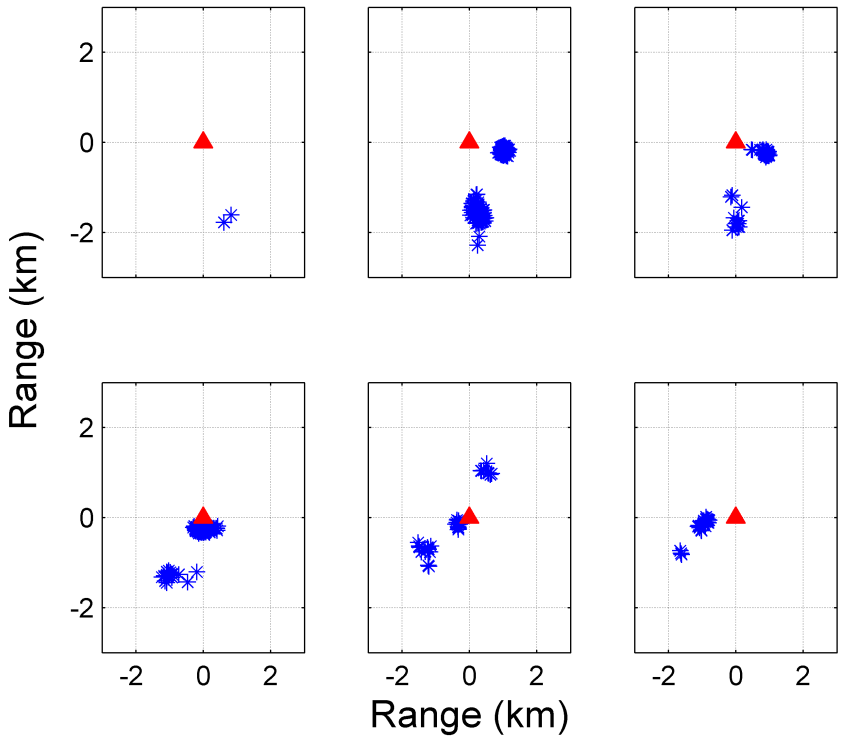


Figure 3.16: Stenellid dolphin encounter #2 localizations by five minute bin. Blue stars indicate click localizations, red triangles indicate HARP position. Animals are assumed to be at the sea surface for localization purposes.

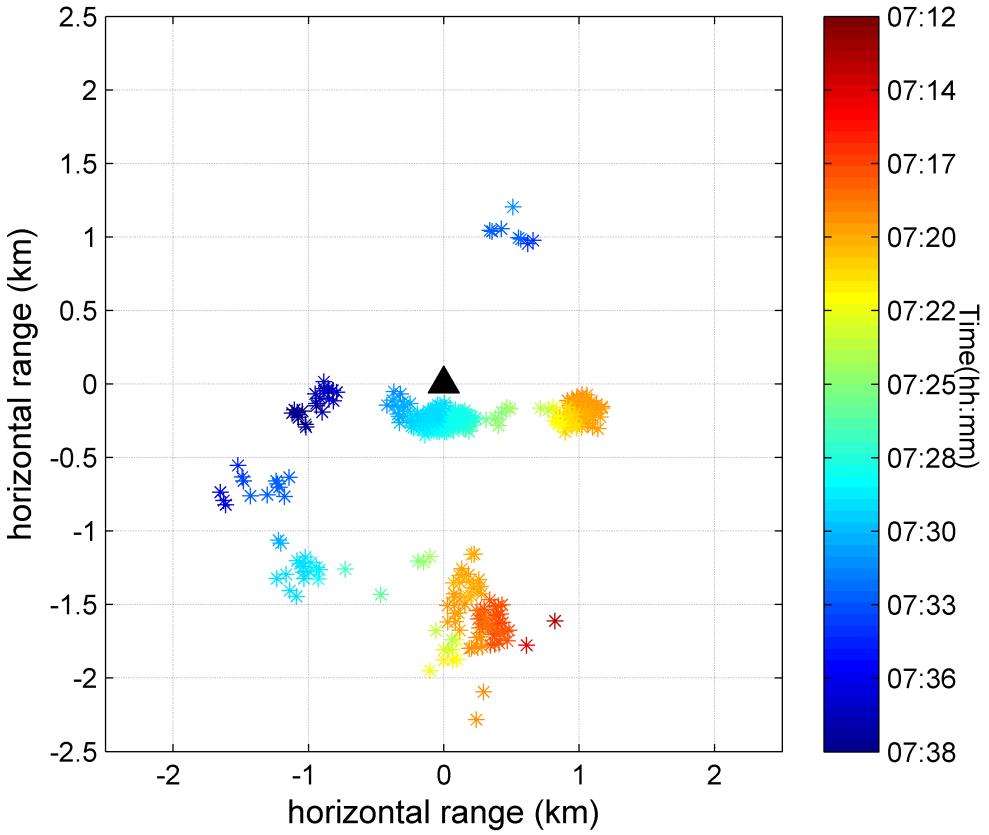


Figure 3.17: Bird's eye view of Stenellid dolphin encounter #2 as localized by seafloor sensor. Black triangle indicates tracking HARP location. Color indicates time (GMT).

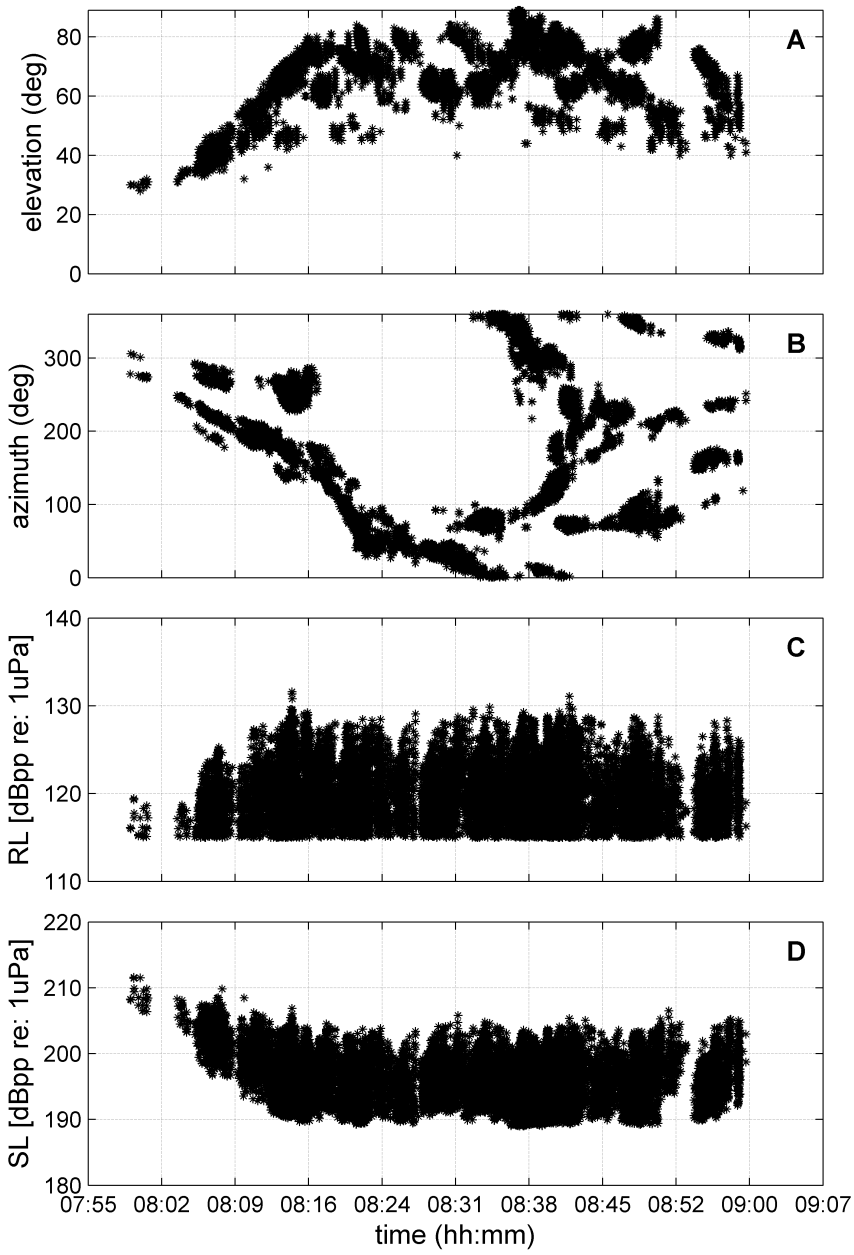


Figure 3.18: Stenellid dolphin encounter #3 localized by tracking HARP. (A) Elevation relative to sensor. (B) Azimuth relative to sensor. (C) RLs of localized clicks over time. (D) Estimated SLs of localized clicks based on received levels and position estimates. All times are in GMT.

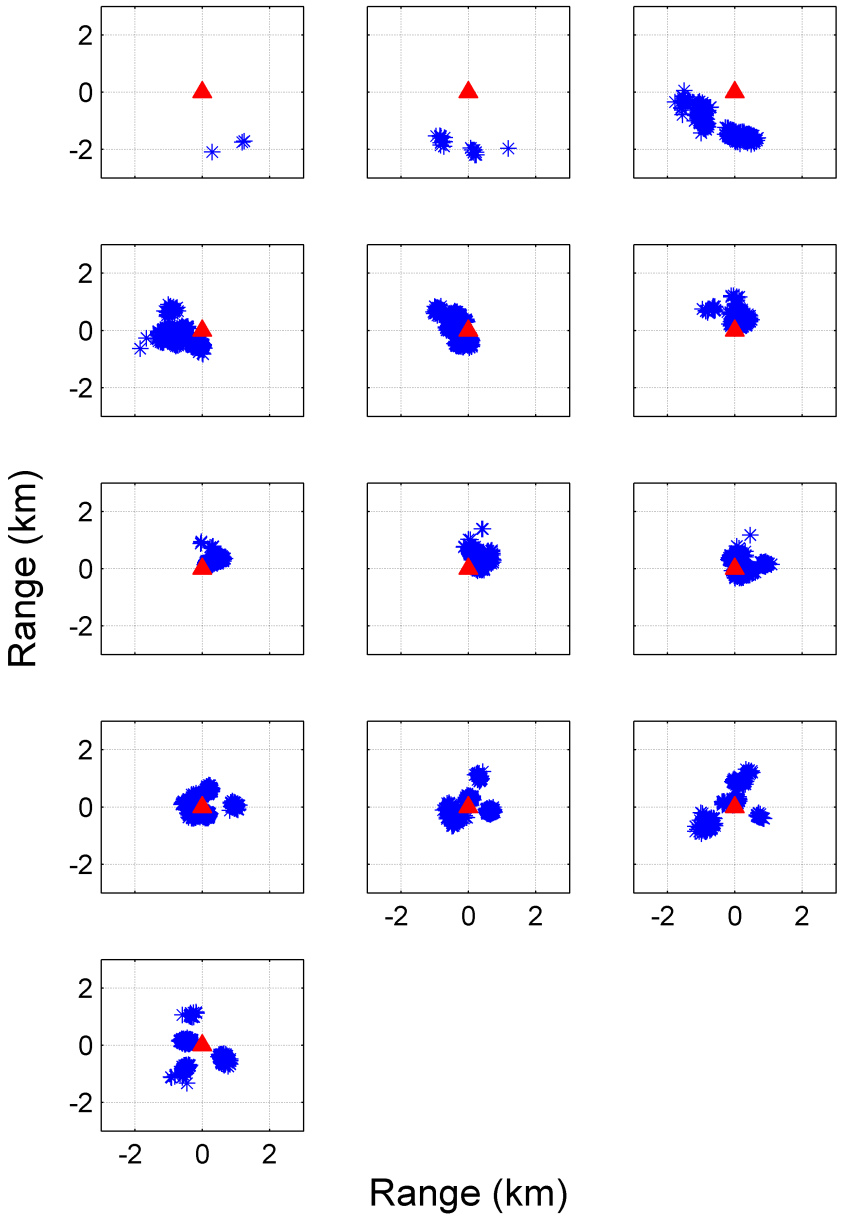


Figure 3.19: Stenellid dolphin encounter #3 localizations by five minute bin. Blue stars indicate click localizations, red triangles indicate HARP position. Animals are assumed to be at the sea surface for localization purposes.

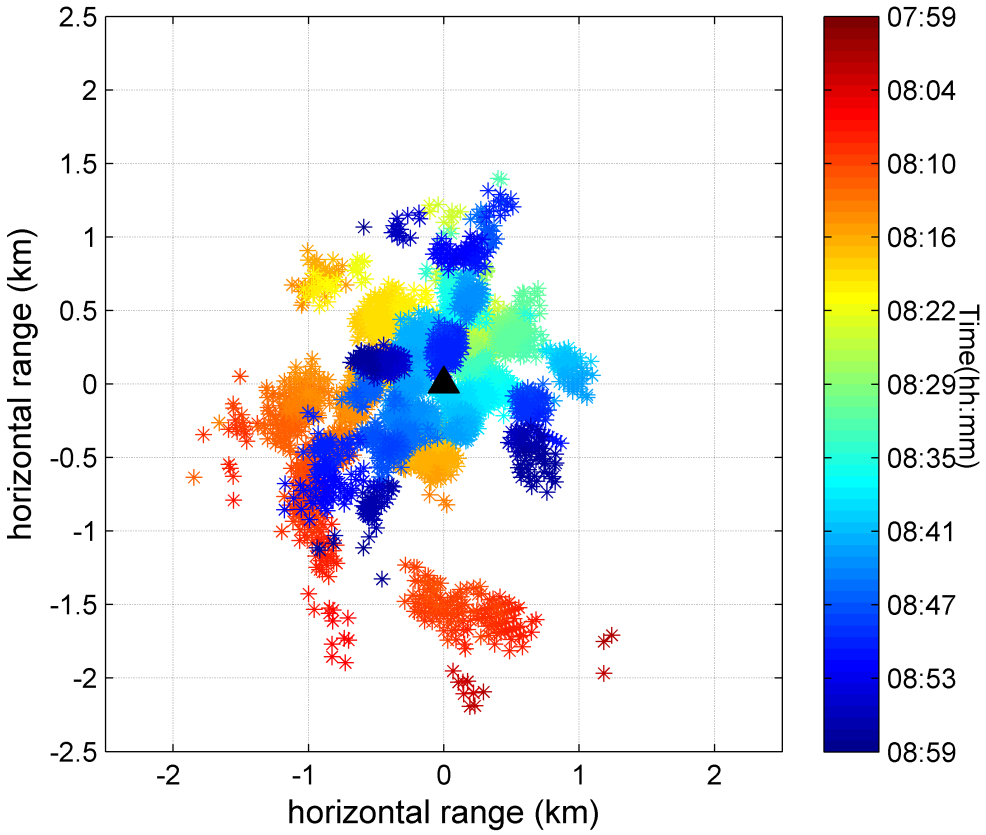


Figure 3.20: Bird's eye view of Stenellid dolphin encounter #3 as localized by seafloor sensor. Black triangle indicates tracking HARP location. Color indicates time (GMT).

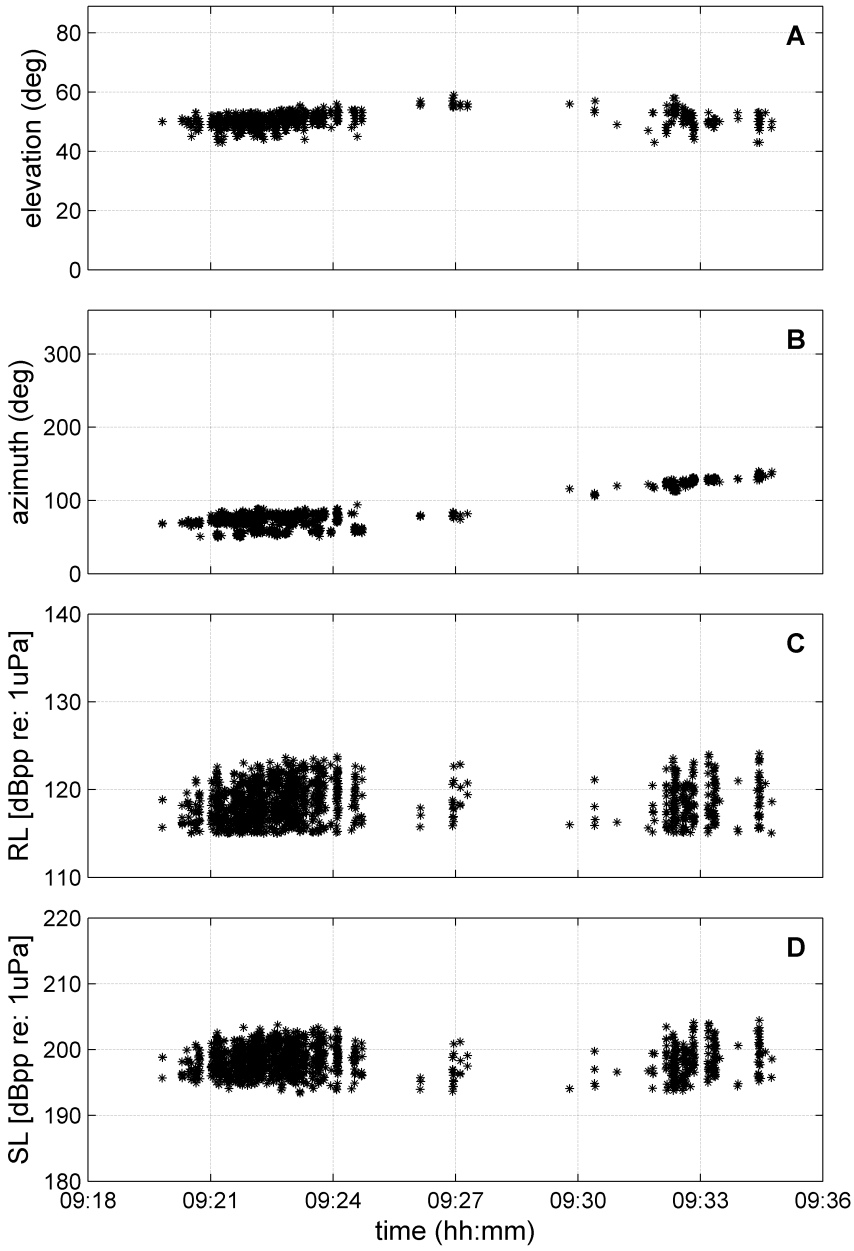


Figure 3.21: Stenellid dolphin encounter #4 localized by tracking HARP. (A) Elevation relative to sensor. (B) Azimuth relative to sensor. (C) RLs of localized clicks over time. (D) Estimated SLs of localized clicks based on received levels and position estimates. All times are in GMT.

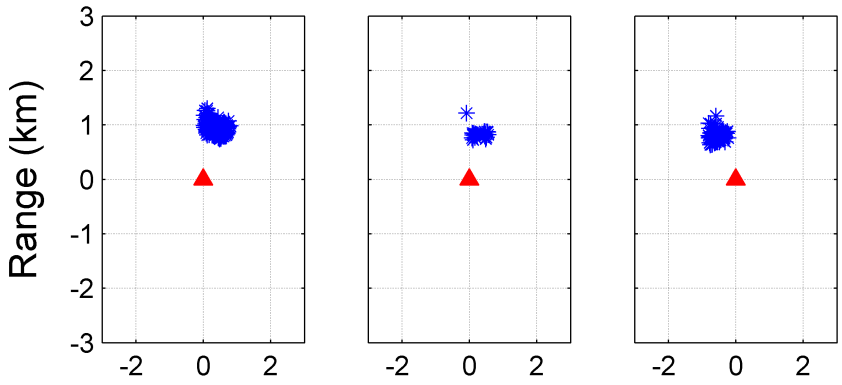


Figure 3.22: Stenellid dolphin encounter #4 localizations by five minute bin. Blue stars indicate click localizations, red triangles indicate HARP position. Animals are assumed to be at the sea surface for localization purposes.

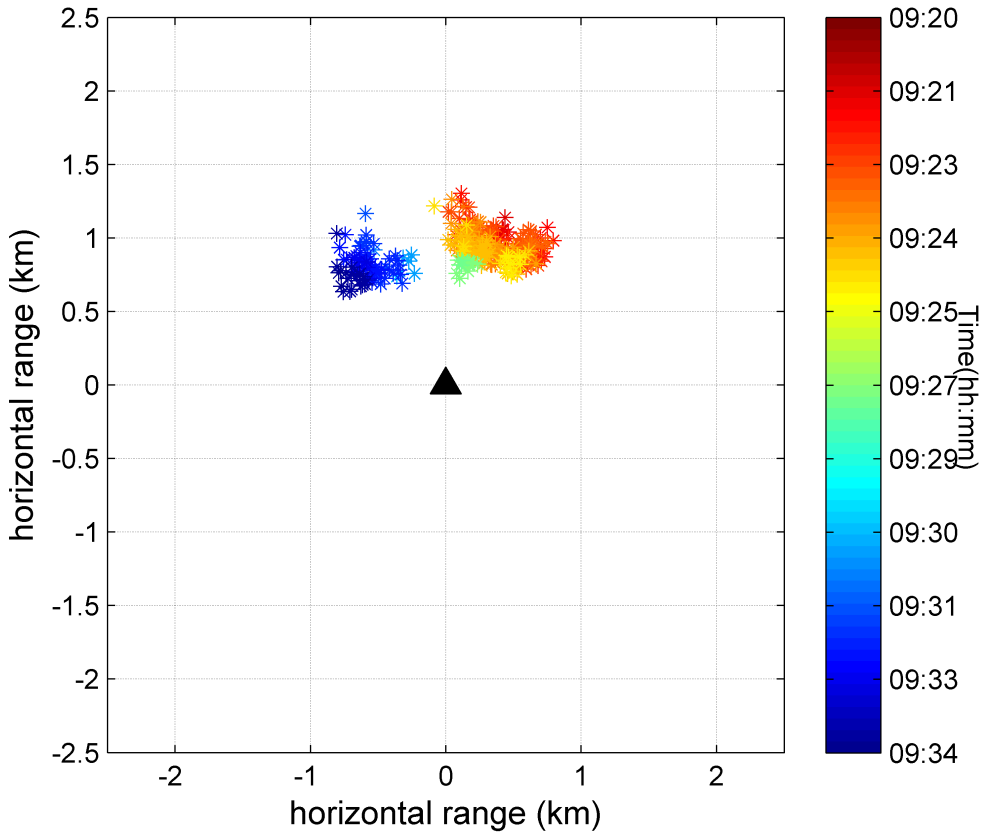


Figure 3.23: Bird's eye view of Stenellid dolphin encounter #4 as localized by seafloor sensor. Black triangle indicates position of tracking HARP. Color indicates time (GMT).

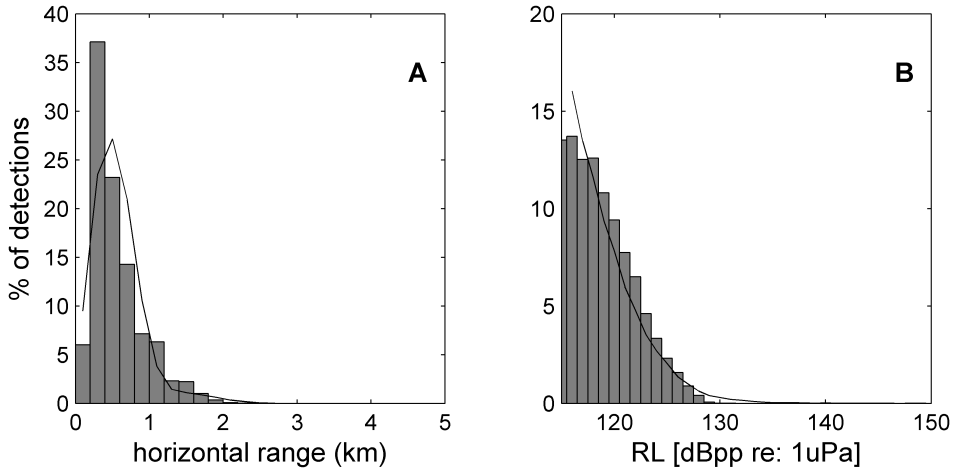


Figure 3.24: *Stenella* sp. in situ horizontal detection ranges (**A**) and received level distribution (**B**). Grey bars indicate results from in situ data. Overlaid black line indicates model-based distribution.

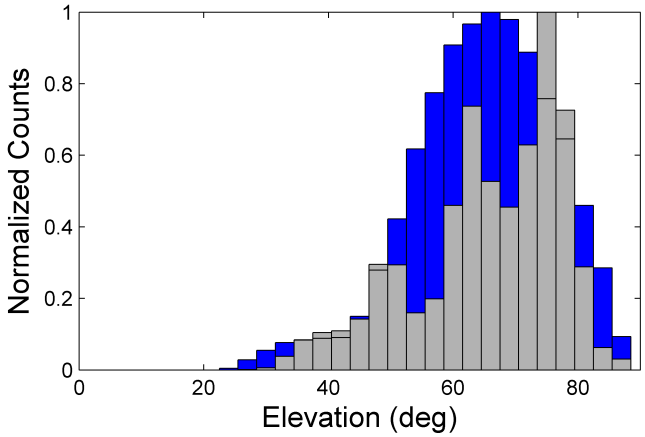


Figure 3.25: Blue bars show model-predicted distribution of elevation angles at which *Stenella* sp. should be detected if randomly distributed in space. Grey bars indicate actual distribution of elevation angles at which *Stenella* sp. were detected during all encounters. An elevation of 90° indicates that the source is directly above the sensor.

3.5 Discussion

3.5.1 Ship Localization

The in situ data demonstrates agreement between model predictions of detection range and detectability at site DT and real values from tracked encounters. Acoustically-based position estimates for the ship, a sound source of known location, closely matched the actual vessel positions in most cases. The fact that estimated elevations were very close to expected values, while azimuthal estimates deviated, suggests that the sensors might have been rotating slightly around the vertical axis during the tracking periods. This type of rotation was seen in test deployments, albeit on a tidal cycle much longer than these tracking periods.

A second line of elevation estimates appears to shadow the main trajectory in Figure 3.3 (top panel). This is attributable to errors associated with signal onset identification when detecting interrogation pings. When the signal onset is identified one cycle earlier on some sensors than others, the result is a 0.1 ms over or under-estimate of the TDOA for some sensor pairs (Figure 3.2), which translates into a localization error. This might be improved by fine-tuning the interrogation ping detection process. Previous ship localization efforts have used propeller noise as the signal of interest for localization, rather than interrogation pings (Wiggins et al., 2012). However, this vessel's propeller was small, low power, and did not generate enough noise for successful localization using that method.

Minor inaccuracies in sensor depth and position estimates may introduce further discrepancies between the position estimates and true locations. The algorithm used to estimate sensor position from the interrogation ping two-way travel times assumed that the ship's GPS and the transducer were at the same location. In reality, the transducer was towed alongside the vessel, and may have been as much as 20 m away from the GPS

receiver.

The ship tracking exercise demonstrates that the elevation estimates from which horizontal ranges are computed are generally reliable. Horizontal range estimation to echolocating animals for ground truth purposes is therefore feasible. Although deviations were seen in azimuthal estimates, azimuth appears to be fairly reliable over small time windows.

3.5.2 Dolphin Localization

One Risso's dolphin and four *Stenella* sp. encounters were successfully localized. Both encounters were nocturnal (local time is GMT - 05:00).

The Risso's dolphin tracks over time indicate a small group of animals moving through the area during a one hour period, with a possible single animal following behind in the same direction, toward the end of that period. Animals were initially localized at a horizontal range of 3 km. Four waves of *Stenella* sp. transits were identified in the area over a 3.5 hour period. Possible rotation of the sensor during this period prevents us from identifying the absolute direction of travel.

Animals were assumed to be at the surface, for localization purposes, however some clicks may have been produced at depth. Multiple seafloor packages would be necessary to determine depth. Clicks produced at depth would be closer to the sensor than estimated assuming a surface location, which would reduce some of the horizontal ranges for these encounters. In the case of the Risso's dolphin encounter, agreement between modeled and in situ data improves if the animals are assumed to be at depth. Here we assumed that all clicks were produced at 200 m below the sea surface, for comparison with the assumption that the animals were at the sea surface. This is an over-simplification aimed at examining the effects of depth assumptions. The comparison indicates that the Risso's dolphins were likely diving to a variety of depths during the

tracking period.

The Stenellid dolphins were also likely diving during the tracking period, rather than staying at the sea surface. Because these animals do not typically dive as deep as Risso's dolphins (Baird et al., 2001; Scott and Chivers, 2009), the effect on localizations is expected to be smaller. However some of the in situ horizontal range estimates obtained by assuming the animals were at the sea surface are likely overestimates.

The method for estimating source positions assumes that the sound travels from source to receiver along a linear path. In reality the path is not truly linear, due to the bending effects of the thermocline (Urick, 1967). At horizontal ranges less than or close to the water depth, the effect of this bending on position estimates is small, as can be seen in the ship localization case: when the ship is close to the sensor, errors decrease. However, as the horizontal distance increases, the error resulting from the linear path assumption also increases. Low elevation angles associated with large horizontal ranges are also problematic, as range estimates become sensitive to small errors in angle estimates. In the case of dolphins, however, detection ranges are relatively short, and errors due to signal path curvature are estimated to be less than 10 meters.

Tracks over time appear patchy. This is primarily attributed to a 'flashlight' effect, in which many clicks are detected when an animal or cluster of animals turns toward the sensor, and few clicks are detected when they turn away, due to the directional nature of echolocation clicks. An alternative explanation is that the animals fall silent during these periods. Although the latter may be true part of the time, one clue that this is not the primary explanation lies in the relationship between animal direction of travel and number of detections. Localized clicks were recorded more often during the approach phase, when the animals were traveling toward the sensor, than the departure phase. Assuming that probability of vocalization is independent of whether the animals are approaching or departing, we can infer from this that the animals are probably off-axis,

rather than silent, during gaps in the tracks. As a result, both approach and departure bins need to be taken into account when computing the vocalization probability P_v because the model-based estimate of detection probability is an average across all orientations of animals relative to the sensor.

The method used here to estimate P_v is not ideal, because it assumes that we can deconvolve this parameter from group size, click rate, and detection range. These parameters are highly intertwined. A better approach would be to use acoustic tag data to come up with an independent estimate of clicking activity across a variety of behaviors. Acoustic delphinid tag data is currently rare, but is expected to become increasingly available over the next few years.

The density of clicks and localizations during the Stenellid dolphin encounters suggests more animals than seen in the Risso's encounter, however the click rate of Stenellid dolphins is expected to be two to three times that of Risso's dolphin, which partially explains the larger numbers of clicks. Clicks were localized at a maximum horizontal range of roughly 2.5 km. This smaller detection range relative to Risso's dolphin is expected due to lower SLs and higher peak frequencies in Stenellid dolphins. Individual animals are not readily distinguishable in the Stenellid tracking data. This is partially due to lower click detectability, but also is likely due to behavioral differences including tighter groupings of individuals, and coordinated vocalization behaviors such as 'eavesdropping' in order to avoid sonar jamming (Götz et al., 2006; Gregg et al., 2007). Computing P_v from these data using an average group size is the best we can do given the current data, but further refinements are needed. Our estimates for Risso's dolphin are in line with $\hat{P}_v$ estimates for beaked whales (Hildebrand et al., 2015), which are also deep divers. The Stenellid $\hat{P}_v$ computed here is higher than the Risso's dolphin estimate. Further data are needed refine probability of vocalization estimates for Stenellids and other delphinid species.

Back-calculated SLs from horizontal range and RL yields peak SL estimates below the maximum predicted for both species. Values are lower than expected based on the literature because the in situ clicks are not strictly on-axis. It is possible that none of the received clicks from the Risso's dolphin encounter were perfectly on-axis with respect to the sensor given the small sample size, and the narrow aperture of delphinid beam patterns (Au et al., 2012a). Experimentally-derived SLs increased slightly with increasing range, particularly when the animals were approaching the instrument. This relationship is also predicted by the model results, and is attributed to the fact that animals are more likely to click on-axis relative the seafloor instrument when they are farther away. As animals approach the instrument, a steeper body inclination is required to obtain an on-axis click. Further refining SL estimates will likely require an array of multi-channel sensors capable of resolving source depth. Model-predicted and in situ elevations are right-skewed for the Stenellid dolphin encounters, and left-skewed for the Risso's dolphin encounter. This is primarily attributed to source level differences.

Differences between model and experimental results for both encounters are attributable in large part to non-uniform distributions of the animals in space. The model is designed to simulate average detectability over a large number of encounters and conditions. Overall, the tracking data from this ground truth experiment are consistent with the model predictions and support the use of the model-based detection probability in generating preliminary delphinid density estimates.

3.6 Conclusion

By localizing a ship transducer with known coordinates, we demonstrated that the tetrahedral HARP configuration could be used to accurately localize a sound source at the sea surface within a range of a few kilometers. Five delphinid encounters were

successfully tracked. Predicted distributions of RL, SLs and horizontal detection ranges are consistent with distributions computed from these encounters. The consistency of the results supports the use of the model to estimate detection probabilities, when multi-channel data are not available.

Chapter 4

Designing a delphinid echolocation click detector for density estimation in the Gulf of Mexico

4.1 Abstract

A delphinid echolocation click detector was designed to find clicks in passive acoustic HARP recordings from five sites in the Gulf of Mexico in a predictable way, such that detection probabilities could be simulated in a model framework. Unlike other detectors, this click detector is based on an absolute amplitude ($dB_{p-p}re : 1\mu Pa$) threshold. This detection scheme was coupled with a manual review phase aimed at removing large batches of false positives where ambient noise was high. A false positive rate was then determined for each site, across the entire monitoring period. Time series of detection results are presented for each site both in terms of weekly average click counts and as weekly mean percentage of bins containing detections. By satisfying the density estimation requirements that detection probabilities be predictable, and that false

positives be evenly distributed throughout each dataset, we produced a detection time series specifically tailored for use in delphinid density estimation.

4.2 Background

Recording efforts in the Gulf of Mexico yielded a large amount of acoustic data in which delphinid signals needed to be detected for monitoring purposes. A review of the HARP data as well as array data collected Gulf indicated that delphinids in this region whistle infrequently, therefore echolocation clicks were a better option for documenting delphinid presence.

4.2.1 Developing a Predictable Detector

Detection and classification are closely intertwined. Detections are not typically useful unless there is some indication of what was detected. Many so-called detectors incorporate multiple layers of filters, transformations, and thresholds designed to exclude undesired signals from the final set of detections. These layers are actually preliminary classification steps, in which the classifier attempts to decide whether the detected signal is of the kind the user is interested in.

This approach becomes problematic with respect to density estimation because the extra layers of complexity make it difficult to predict the behavior of the detector. When classification steps are added, there is no longer a simple way to determine whether a signal will be detected under a given set of conditions.

If synthetic signals can be generated and fed to a detection algorithm, then these can be used to characterize the detector's performance (Helble et al., 2013b). Alternatively, if recordings are available which are coincident with high resolution tracking data capable of pinpointing the source location, such recordings can also be used

to characterize the detector (Marques et al., 2009). However, in the case of delphinid clicks, no such data are currently available, and click models (e.g. Zimmer, 2011) are not developed enough to simulate the complex interference patterns associated with off-axis clicks (Au et al., 2012a) in a model framework.

An alternative approach is to build a predictable click detector, with behavior that can be incorporated into a model such as that described in Chapter 2. The detector must rely on clearly identifiable thresholds in order to be accurately modeled in simulations. Extra filters can result in unpredictable detection behavior, opening up the possibility that the detector will behave differently in different noise conditions for instance, unaccounted for by the simulation. This can lead to potential biases and erroneous results (Helble et al., 2013a). At the same time, the detector also needs to minimize the number of false positives returned, and to be efficient due to the large quantity of data to be analyzed.

A number of choices were made in the design of this click detector to maintain predictability. A peak-to-peak signal amplitude threshold ($dB_{p-p}re : 1\mu Pa$) was used, rather than a signal to noise ratio, because the impacts of different types of noise of click detectability were too complex to simulate, without a sophisticated click model. Once candidate detections were identified, the detector relied on three variables: duration, energy distribution, and peak frequency, to determine whether or not a signal was a delphinid click.

The trade-off for detector simplicity is a higher false positive rate in complex or noisy environments. In a density estimation framework, the solution to this is to characterize the detector's false positive rate and factor it in to the density estimates (Buckland et al., 2001). However, this approach is only reliable if false positives can be assumed to be uniformly distributed throughout the time series. If false positives occur instead in concentrated episodes, they can lead to biased density estimates.

4.2.2 Bin vs. Cue Counting Detection Rates

For density estimation, we are considering two possible approaches: Cue counting and group counting. Each method has its pros and cons, as discussed in chapter 2. Group counting is less sensitive to false positives than cue counting, however the process of binning detections for this method may obscure detail in the time series. Part of the detection process therefore is to understand the relationship between these two methods across sites, in order to recognize potential biases in the results associated with each.

4.2.3 Goals of this Chapter

The goals of this chapter are as follows:

- Describe the detector implemented.
- Describe the false positive rate estimation process and present resulting estimates.
- Provide time series of delphinid click detections at all sites, adjusted for false positives, for use in click counting and group counting density estimation methods respectively.
- Account for the effects of detector dead time on detection counts.
- Investigate the relationship between the click and group counting metrics.

4.3 Methods

4.3.1 Detection

The goal in creating this detector was to build an algorithm with behavior that could be fully modeled in the detection probability simulations described in chapter

2. Peak-to-peak (PP) amplitude in decibels ($dB_{p-p}re : 1\mu Pa$) was selected rather than signal to noise ratio (SNR) as the primary detection threshold because SNR is much more difficult to model accurately. Issues include colored noise unequally affecting different frequencies, and temporal noise patterns biasing detectability over time. Using peak to peak amplitude ensures that detector performance will remain consistent across noise conditions, however false positives may increase during noisy periods.

A two-step detection process modeled after work by Soldevilla et al. (2008) was implemented to find click-like signals in the dataset exceeding a minimum peak-to-peak amplitude threshold. Both detection steps analyzed a band-passed version of the time series data (10 to 85 kHz). Band passing was used to filter out low frequency, high amplitude signals which can otherwise overpower delphinid clicks in the time series. In the first step, a low resolution detector flagged 30 second segments of band passed time series data containing energy peaks greater than 5000 counts (counts are defined as the squared amplitude of the band passed signal).

In the second detection step, a high-resolution detector returned to flagged time periods. All candidate energy peaks within each time period were identified using the same energy threshold approach as in the first step, and then further pruned to retain only clicks with a received level (RL) of at least 115 $dB_{p-p}re : 1\mu Pa$ at peak frequency (clicks with RL >=112 $dB_{p-p}re : 1\mu Pa$ were retained for classification in chapter 5). Start and end times (defined as the moment when energy rose above and fell below the 60th percentile from the mean, respectively) were found for each energy peak. Overlapping clicks, and clicks within 50 μs of each other were merged. Clipped signals, defined as those with amplitudes greater than 90% of the maximum recordable by the instrument A/D were ignored. The time series of each candidate click was extracted, with an added 250μs buffer on either end, for analysis by a simplified classification scheme designed to exclude impulses that did not have the basic expected features of delphinid clicks.

Classification and Post-processing

A rudimentary, preliminary classification step followed the detection process, in order to discard clearly non-delphinid signals. The main categories of such signals at most sites were beaked whale clicks, sperm whale clicks, and cavitation noise from ship propellers. Classification heuristics were therefore based on manual analysis of the classifier's ability to reject these unwanted detections on a subset of data.

In the classification step, the amplitude envelope of each click was computed using a Hilbert transform, and normalized on a scale from zero to one. The duration of the portion of the envelope above 0.5, as well as the ratio of the envelope amplitude in the first half vs. the second half of the click envelope were computed (method based on work by SBP, pers. com.). A 240 point-fft with 50% overlap of Hanning windowed data was used to estimate the frequency content and the peak frequency of the signal. Signals with an envelope duration between 10 and 70 μs, total click duration less than 1 ms, peak frequency between 20 and 80 kHz, and a positive envelope energy distribution ratio were classified as delphinid clicks. Each threshold choice was linked to a specific anticipated false positive source: Beaked whale clicks have longer envelope and overall durations than delphinid clicks, sperm whale clicks have lower peak frequencies, and propeller cavitation impulses typically have irregular envelope shapes and energy distributions.

Lone clicks or pairs of clicks separated from a neighboring signal by more than 0.5 seconds both before and after were discarded, because manual analysis revealed that such detections were likely to be false positives. Click bouts, defined as periods of clicking with more than 15 minutes without detections both before and after, were identified automatically. Bouts shorter than 75 seconds and bouts containing fewer than 25 clicks were excluded from further analyses.

After the detection process was complete, a long term spectral average (LTSA, Wiggins and Hildebrand, 2007) of each bout was reviewed visually by an analyst (KF),

who flagged concentrated bouts of false detections for removal. If false positives are uniformly distributed throughout the dataset, they can be accounted for later on using a false positive rate, however some false positives (particularly ship noise) occur in concentrated events. The manual review process was aimed at minimizing the inclusion of such events in the detection dataset.

Time series were computed for click counts, in preparation for cue-based density estimation methods. Time series of positive (containing clicks) and negative (no clicks) five minute bins, were generated in preparation for group-counting density estimation methods.

Shallow water detections

A number of challenges appear when attempting to detect clicks in data recorded at shallow sites (<300 m deep). In particular, the clicks often arrive via an indirect path, bouncing off the sea surface and/or seafloor before arriving at the sensor. Multiple arrivals are common for this reason. There is also increased noise from anthropogenic, biological and oceanographic sources relative to deep sites. Shipping activity is high at some coastal sites. Sensors are close to the surface and record more weather-related noise. Snapping shrimp are a particular problem at site MP, where their signals are difficult to distinguish automatically from delphinid clicks. Conversely, beaked and sperm whales are much less prevalent at these sites.

In order to address these challenges, several parameter changes were made when running the detector on site MP and DC recordings. The low pass end of the band pass filter was increased slightly for shallow site detection at sites MP and DC, to 90 kHz, to account for the fact that animals were closer to the sensors at these sites, and therefore high frequencies would not be as strongly attenuated. The RL minimum detection threshold was also raised to 117 dB_{PP} to reduce the number of indirect arrivals and poor

quality signals detected, and therefore to improve the performance of the preliminary classifier at shallow sites. (Note that this decreases the maximum detection range, and is accounted for in simulations in chapter 2). Allowed click envelope duration was lengthened to 20 to 80 μs to allow for more complex arrivals, and a 1.5 ms lockout was imposed after every detection to avoid counting echos as individual clicks. Lastly, the maximum allowed time between neighboring clicks was reduced to 0.25 seconds. Batches of up to 5 detections isolated from neighboring detections by this amount were discarded, due to the prevalence of snapping shrimp impulse signals, which often occur as one or a small number of impulses, rather than a longer click train.

4.3.2 Estimation of False Positive Rates

Error rates after manual false positive removal were estimated by reviewing a subset of individual clicks. For each deployment, a starting value between 1 and 1000 was chosen to indicate the click number of the first click to be inspected . Using that value, the time series, spectrum and spectrogram of every 5000^{th} click was manually reviewed and labeled as a correct or false detection. This sampling method ensured that the final error rate would represent the entire monitoring period at each site. Once the analysis was complete, a bootstrap approach was used to estimate the mean and standard deviation of the false positive rate. For each site, 500 detections were sampled without replacement from the set of manually reviewed clicks (Table 4.2). The proportion of false positives in this subset was computed, and the process was then repeated 100 times. The mean and standard deviation of the false positive rates by site were estimated from the sampled subsets.

Due to the simplicity of the energy detector it was assumed that no signals exceeding the minimum amplitude threshold were missed, therefore the false negative rate is negligible. Manual removal of large bouts of false positives ensured that no five

minute intervals were entirely composed of false positives, therefore the error rate on the group method is negligible.

4.3.3 Accounting for Detector Dead Time

Counts were adjusted by accounting for detector saturation or dead time (Lucke, 1976). For each time bin T, a true number of counts N was estimated as a function of the detector dead time τ in seconds, and the actual number of counts N_T in the bin, as:

$$\hat{N} = \frac{N_T}{1 - \frac{N_T \tau}{T}} \tag{4.1}$$

In this case, T equals five minutes, or 300 seconds. Detector dead time was higher at shelf sites because of the longer lockout time implemented to avoid counting echos as new detections (τ_{slope} = 50μs; τ_{shelf} = 1500μs).

4.4 Results

Delphinid clicks were detected at all sites (Table 4.1). Detection results are presented here both as raw counts of detections for use in eventual cue-based density estimates, and as positive (containing detections) and negative (no detections) five minute bins for use with group-based density estimates. In general, the two detection indices are correlated, however Figures 4.1 and 4.2 show that they are not strictly proportional to one another, and that the relationship between the two metrics varies across sites. At the upper end of this variation are sites DC and MC, where the ratio of number of clicks detected to the percentage of positive bins per day tends to be higher than at other sites. Site DT is at the lower end. This may have to do with regional differences in species composition and group sizes.

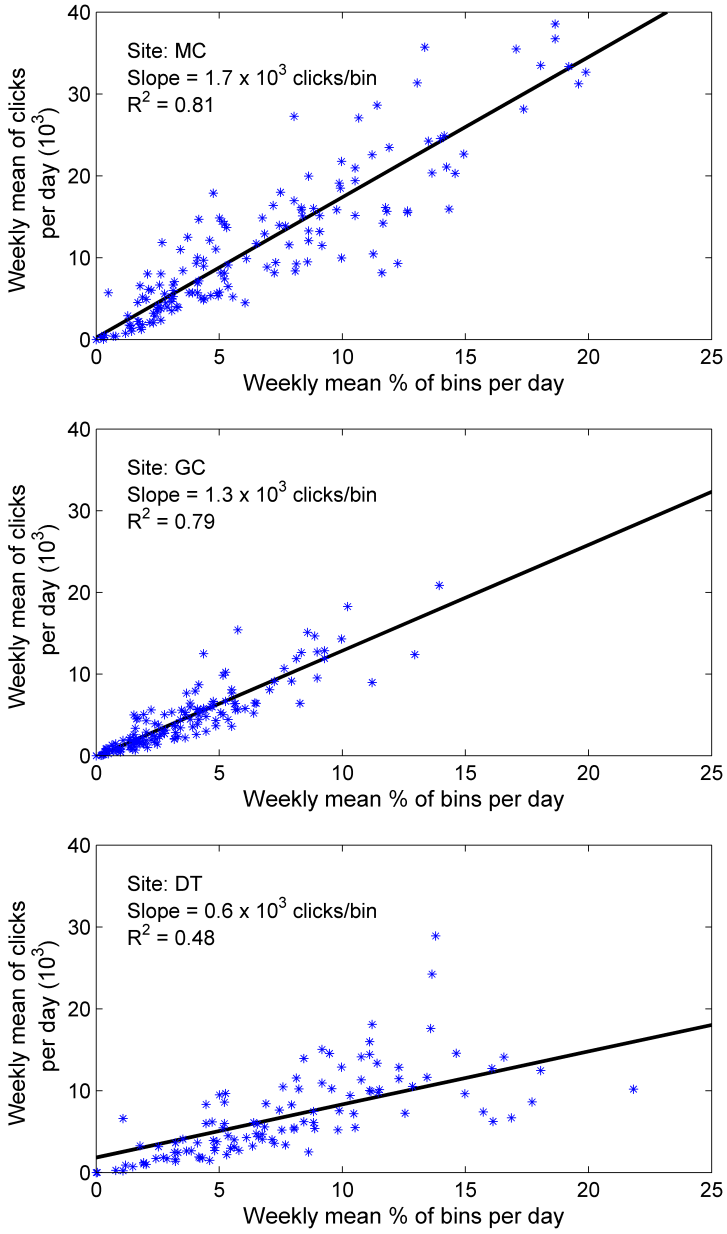


Figure 4.1: Slope Sites: Relationship between weekly average number of detections per day and weekly average number of positive bins per day by site. A first order polynomial fit to each site's dataset (black line), with slope indicating the approximate relationship between the two metrics. R^2 values provide a measure of goodness of fit.

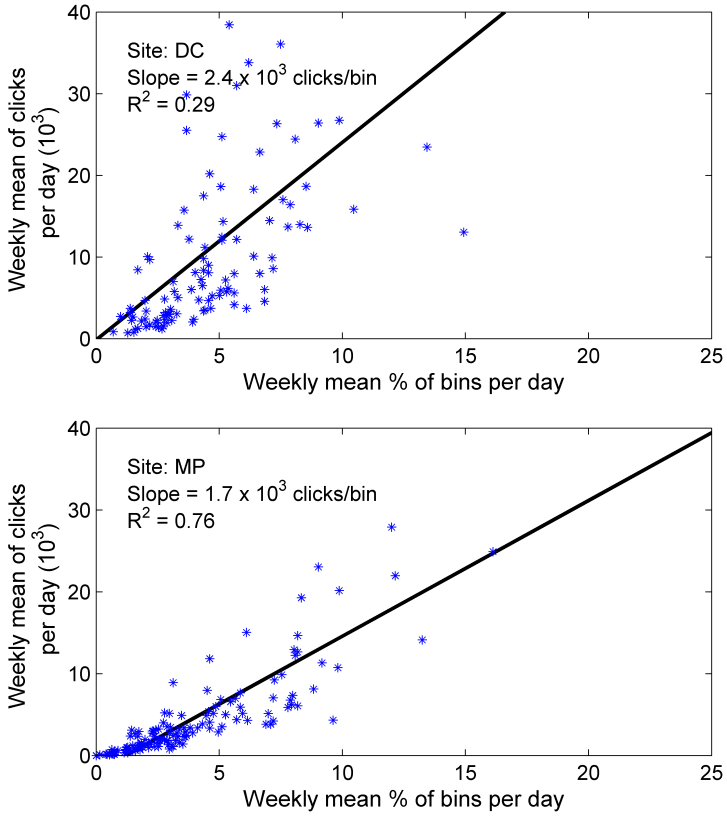


Figure 4.2: Shelf Sites: Relationship between weekly average number of detections per day and weekly average number of positive bins per day by site. A first order polynomial fit to each site's dataset (black line), with slope indicating the approximate relationship between the two metrics. R^2 values provide a measure of goodness of fit.

Table 4.1: Detection summary statistics by site. Total detections indicate the number of clicks detected at each site, while positive bins are the total number of five minute bins containing detections at each site. Total possible five minute bins per day is 288.

Site	Total Detections	Positive Bins	Effort (days)	Mean Clicks per Day (10^3)	Mean % Bins per Day
MC	13,089,754	20,949	1089	10.6 ±16.2 σ	5.9 ±6.8σ
GC	4,842,472	10,809	1001	3.9 ±7.0 σ	3.0 ±4.2σ
DT	5,924,725	18,911	791	4.8 ±8.5 σ	5.3 ±7.4σ
DC	9,296,068	10,463	749	7.5 ±24.0σ	2.9 ±4.5σ
MP	5,711,326	11,594	1020	4.6 ±13.2σ	3.2 ±5.2σ

4.4.1 Detections by Site

Detection counts at all sites increase exponentially with decreasing RL, until RLs fall below the site-specific detection threshold (Figure 4.3). This suggests that detections above the RL threshold are not being systematically missed, which is important for density estimation using model-based detectability estimates.

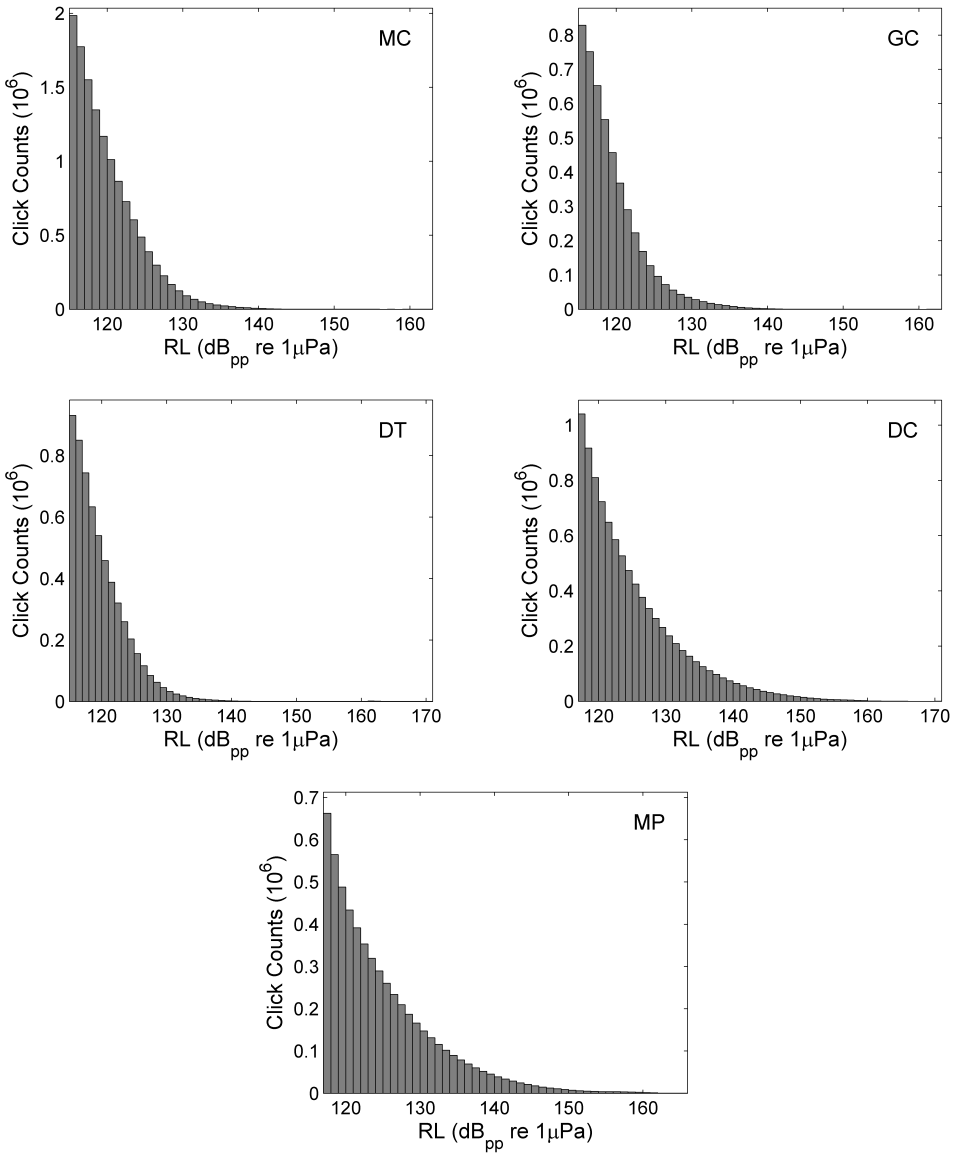


Figure 4.3: Total click counts are shown in 1 dB_{pp} bins as a function of received level (RL) by site. This analysis is performed as a check to ensure that the detector does not start to miss detections as RL decreases. Missed detections would result in a leveling off of counts as RLs approached the minimum amplitude threshold (115 dB_{pp} for slope sites, 117 dB_{pp} for shallow sites). The exponential increase in counts as RL approaches the threshold, seen here for all sites, indicates that clicks above the RL threshold are not being systematically missed, and that the detector is performing predictably.

Missisippi Canyon

The largest dataset in terms of days of recording effort and total clicks detected came from MC (Table 4.1). Mean numbers of clicks per week at site MC were variable, with higher averages in summer months (Figure 4.4). Weekly averages of percent of five-minute bins containing detections follow a similar pattern. The shapes of the two time series (click and group) are similar, suggesting overall agreement between the two methods. The diel view shows a clear nocturnal pattern, with most clicks detected at night (Figure 4.14). Detections tend to increase after dusk and decrease before dawn. Ephemeris in Figures 4.5, 4.7, 4.9, 4.11 and 4.13 were obtained from NASA Jet Propulsion Laboratory's Horizons ephemeris service (Giorgini et al., 1996) via the Tethys workbench (Roch et al., 2013).

Click counts were often high at site MC, with well over ten thousand detections in a five minute period in some cases. The mean false positive rate at this site was the lowest of all sites (Table 4.2).

Table 4.2: False positive detection rates by site.

Site	Mean false positive rate (%)	Number of clicks analyzed
MC	1.4 $\pm 0.55\sigma$	3950
GC	2.0 $\pm 0.51\sigma$	1942
DT	2.2 $\pm 0.55\sigma$	1784
DC	6.4 $\pm 0.96\sigma$	1864
MP	20 $\pm 1.1\sigma$	898

Green Canyon

Mean numbers of clicks per week at site GC were generally lower than at site MC, however the time series retains a similar character, with possibly higher averages in summer months (Figure 4.6). Fewer than half as many clicks were detected overall at this site compared to site MC. Weekly averages of percent of five-minute bins containing detections are also lower overall than at site MC. A pattern of predominantly nocturnal click detection is evident at this site (Figure 4.14). The false positive rate at this site was low and comparable to that at site MC (Table 4.2).

Dry Tortugas

Mean numbers of clicks per week at site DT are comparable to those at site GC, while the proportion of five minute bins containing clicks is generally larger (Figure 4.8 and Table 4.1). Unlike sites MC and GC, no clear seasonal pattern is visible in the time series of delphinid detections at this site, however the same nocturnal increase in detections is seen here (Figure 4.14). As at the other two slope sites, the mean false positive rate was low at site DT. Instrument failure in the first half of 2012 resulted in reduced total recording effort at this location compared to the other slope sites.

Desoto Canyon

Mean numbers of clicks per week at site DC are comparable to numbers of detections at the slope sites, however it is important to remember that the amplitude detection threshold was higher at this site and at site MP, therefore the effective detectability is smaller. No clear seasonal pattern appears in the weekly averaged time series (Figure 4.10). A possible increase in weekly mean of number of click detections may be occurring at site DC during the monitoring period. This increase is not discernible in the binned time series. The hourly view indicates that any diel pattern is weak at this

site (Figure 4.14). Although there appear to be slightly more click detections at night than during the day, daytime detections appear to be more common at this location than at any of the slope sites. The false positive rate at this site was roughly three times higher than at the slope sites, due in large part to snapping shrimp noise.

Main Pass

The number of clicks detected was smallest at this site, despite a high level of recording effort. This is partially attributable to the higher RL threshold used, as well as the relatively low detection probabilities associated with this site (chapter 2). Mean numbers of clicks per week at site MP are generally lower than at site DC, with a few exceptional spikes (Figure 4.12). The time series of proportion of bins containing clicks is similar in overall shape to the counts time series. Maximum click counts per 5 minute bin are lowest at this site, with numbers of detections rising above 10 thousand on only a few occasions. The nocturnal increase in detections at this site appears to be stronger than at site DC, however daytime detections are common (Figure 4.13). The false positive rate was highest at this site, mainly due to snapping shrimp which generate a constant and confounding noise source in these shallow coastal waters.

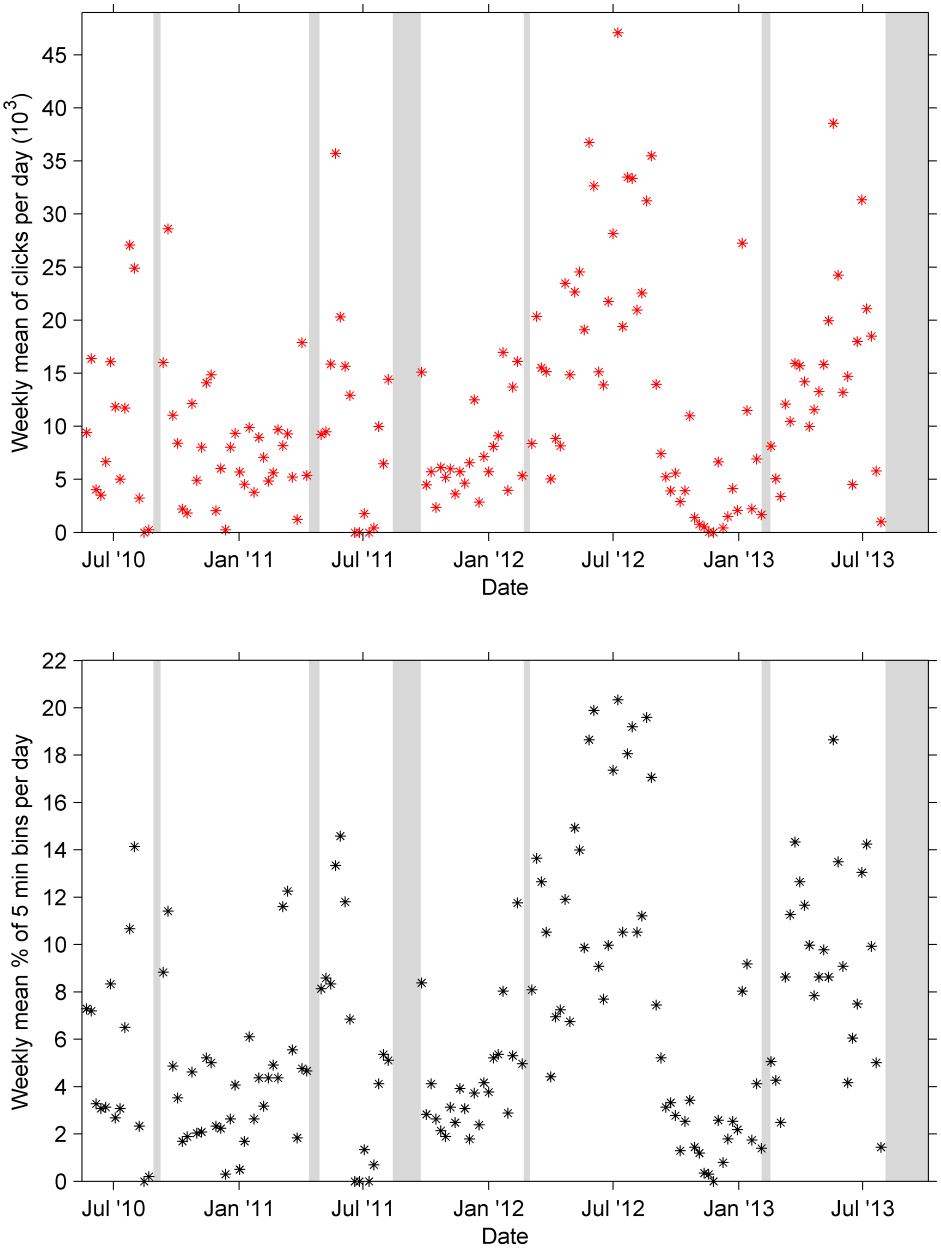


Figure 4.4: Site MC detections. *Top*: Weekly mean of daily number of click detections. Counts are adjusted by mean false positive rate. *Bottom*: Weekly mean of daily percentage of five minute intervals containing delphinid clicks. Gray bands indicate gaps in time series.

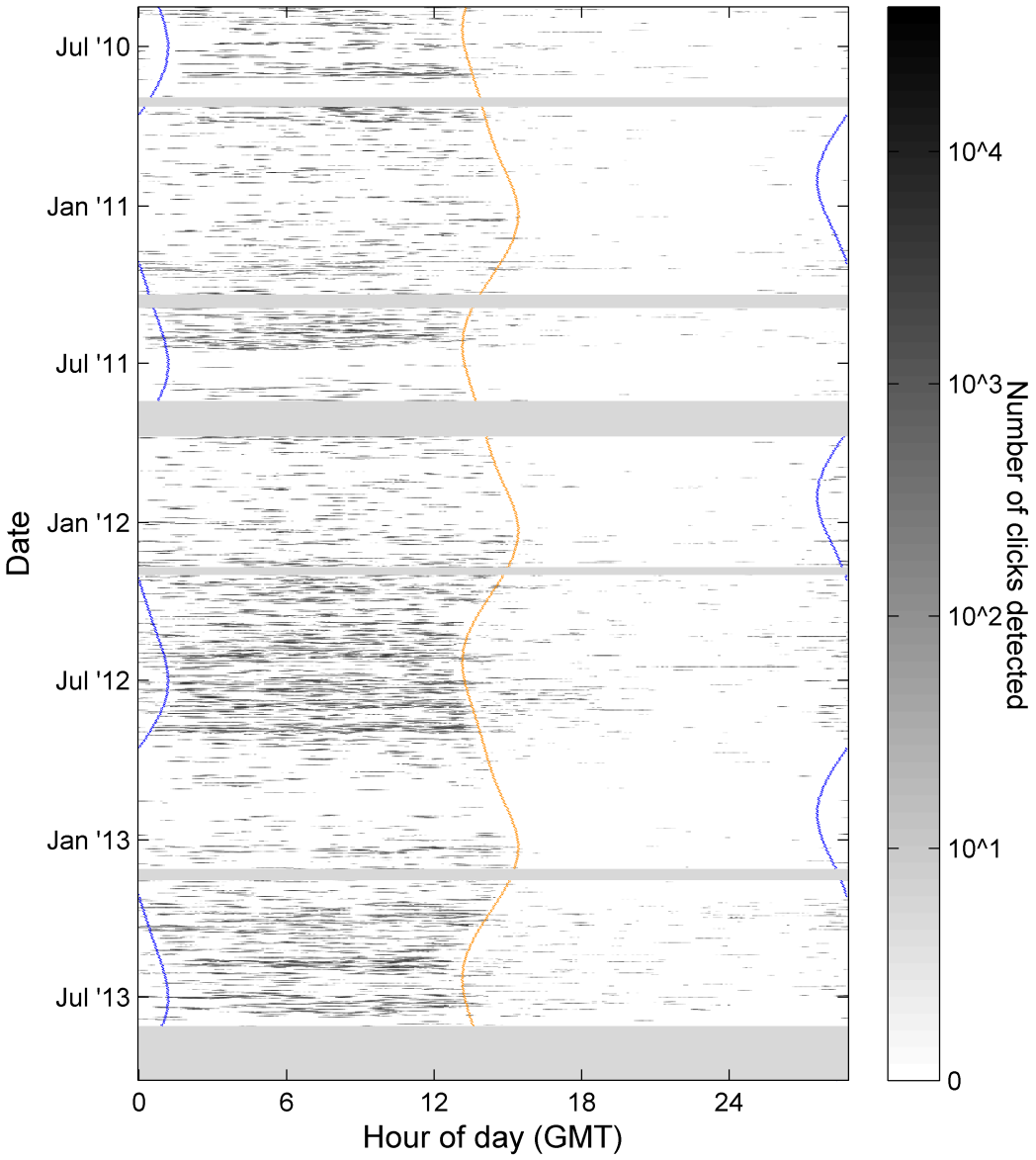


Figure 4.5: Click detections at site MC by time of day for the duration of monitoring period. Grayscale indicates number of clicks detected. Resolution is one day in the vertical, and one minute in the horizontal. Gray bands indicate gaps in time series. Orange curve indicates time of sunrise, blue indicates time of sunset.

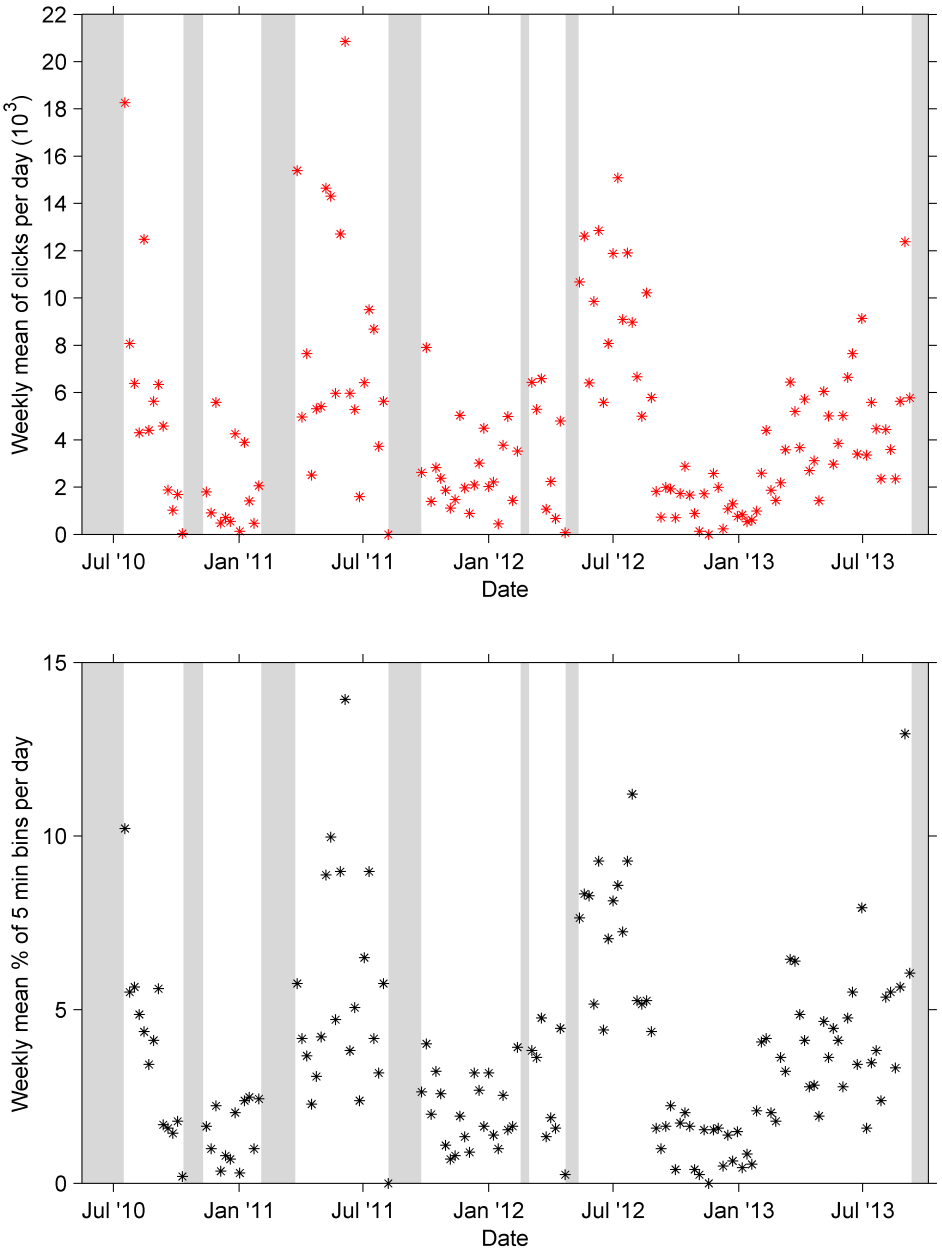


Figure 4.6: Site GC detections. *Top*: Weekly mean of daily number of click detections. Counts are adjusted by mean false positive rate. *Bottom*: Weekly mean of daily percentage of five minute intervals containing delphinid clicks. Gray bands indicate gaps in time series.

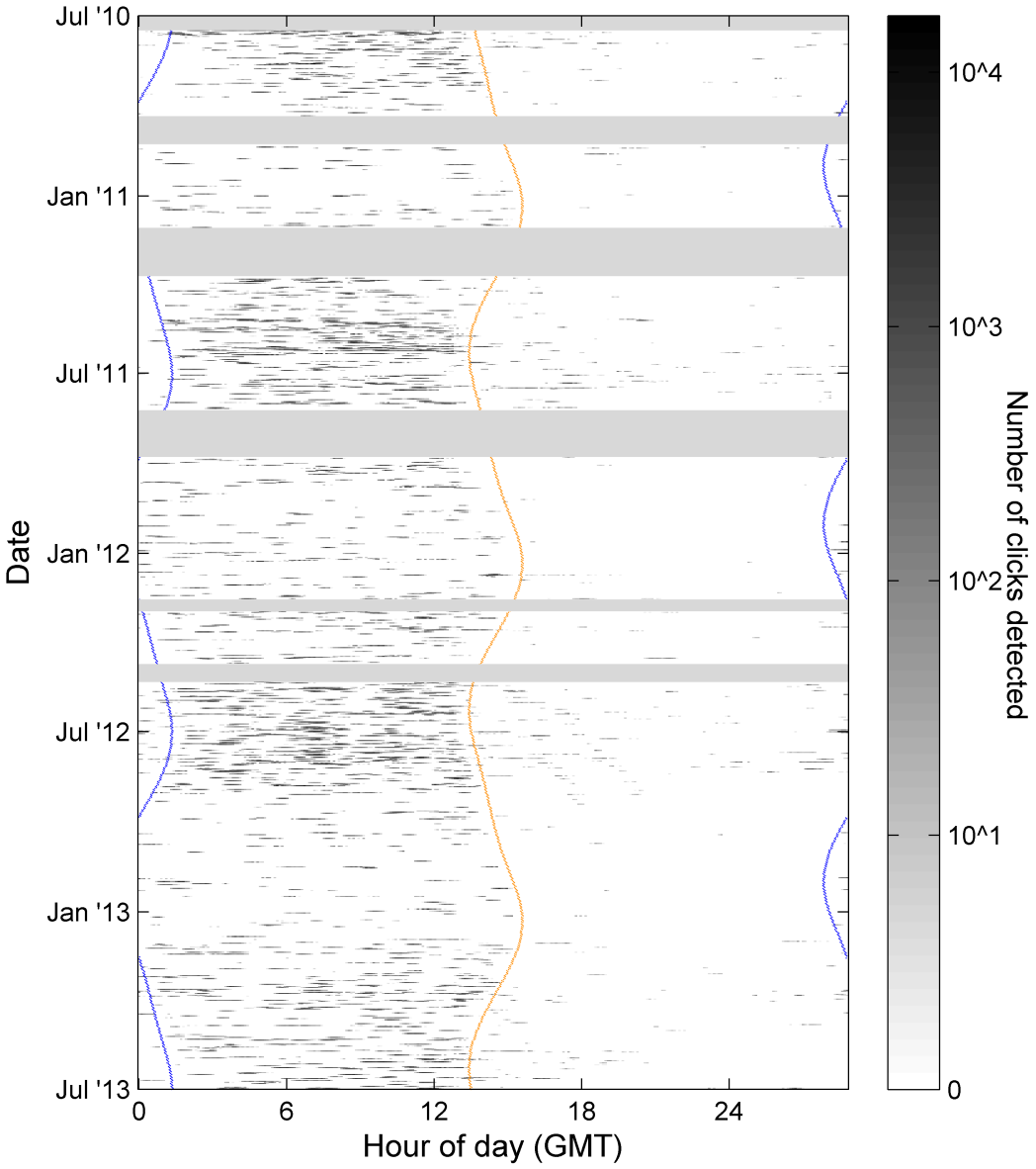


Figure 4.7: Click detections at site GC by time of day for the duration of monitoring period. Grayscale indicates number of clicks detected. Resolution is one day in the vertical, and one minute in the horizontal. Gray bands indicate gaps in time series. Orange curve indicates time of sunrise, blue indicates time of sunset.

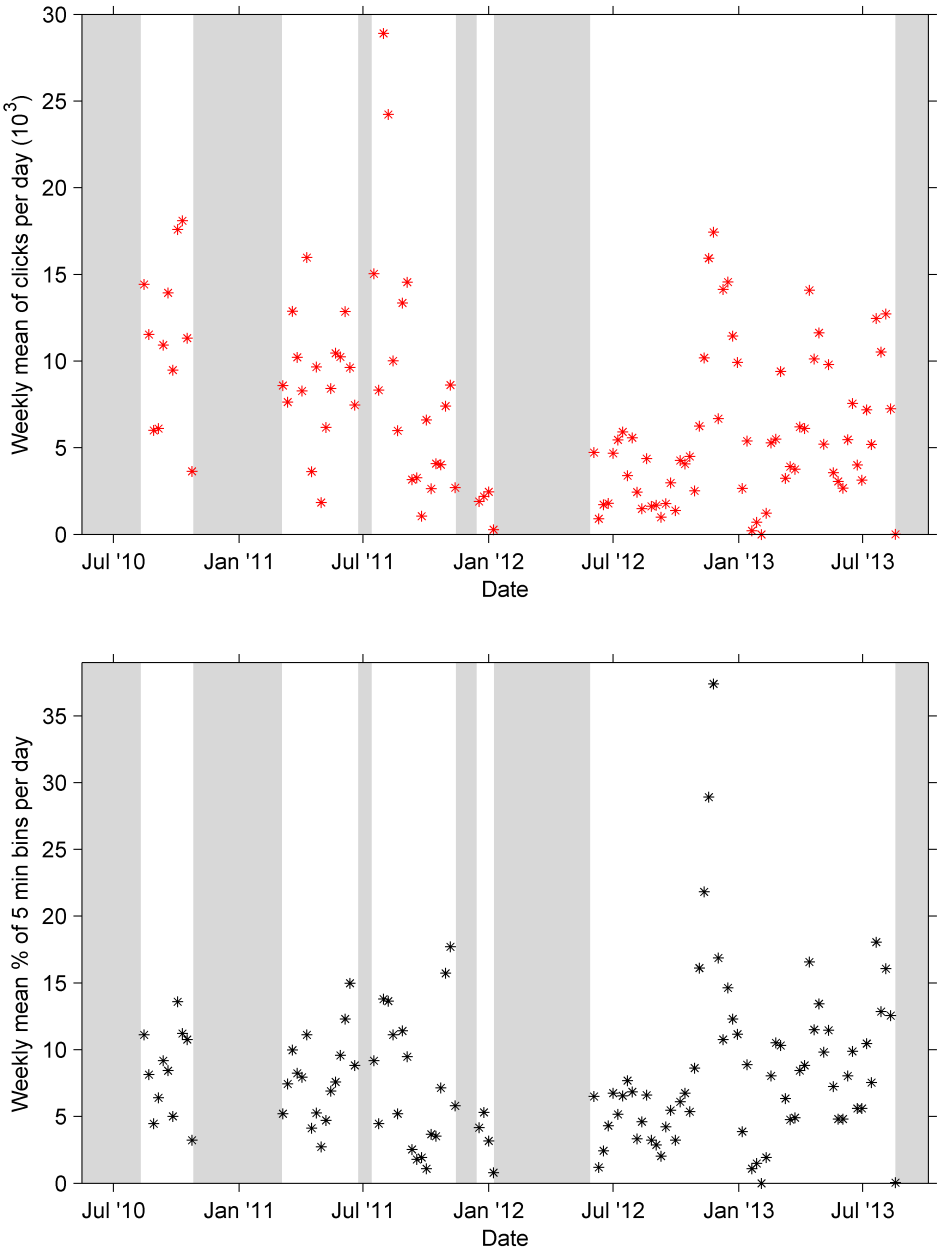


Figure 4.8: Site DT detections. *Top*: Weekly mean of daily number of click detections. Counts are adjusted by mean false positive rate. *Bottom*: Weekly mean of daily percentage of five minute intervals containing delphinid clicks. Gray bands indicate gaps in time series.

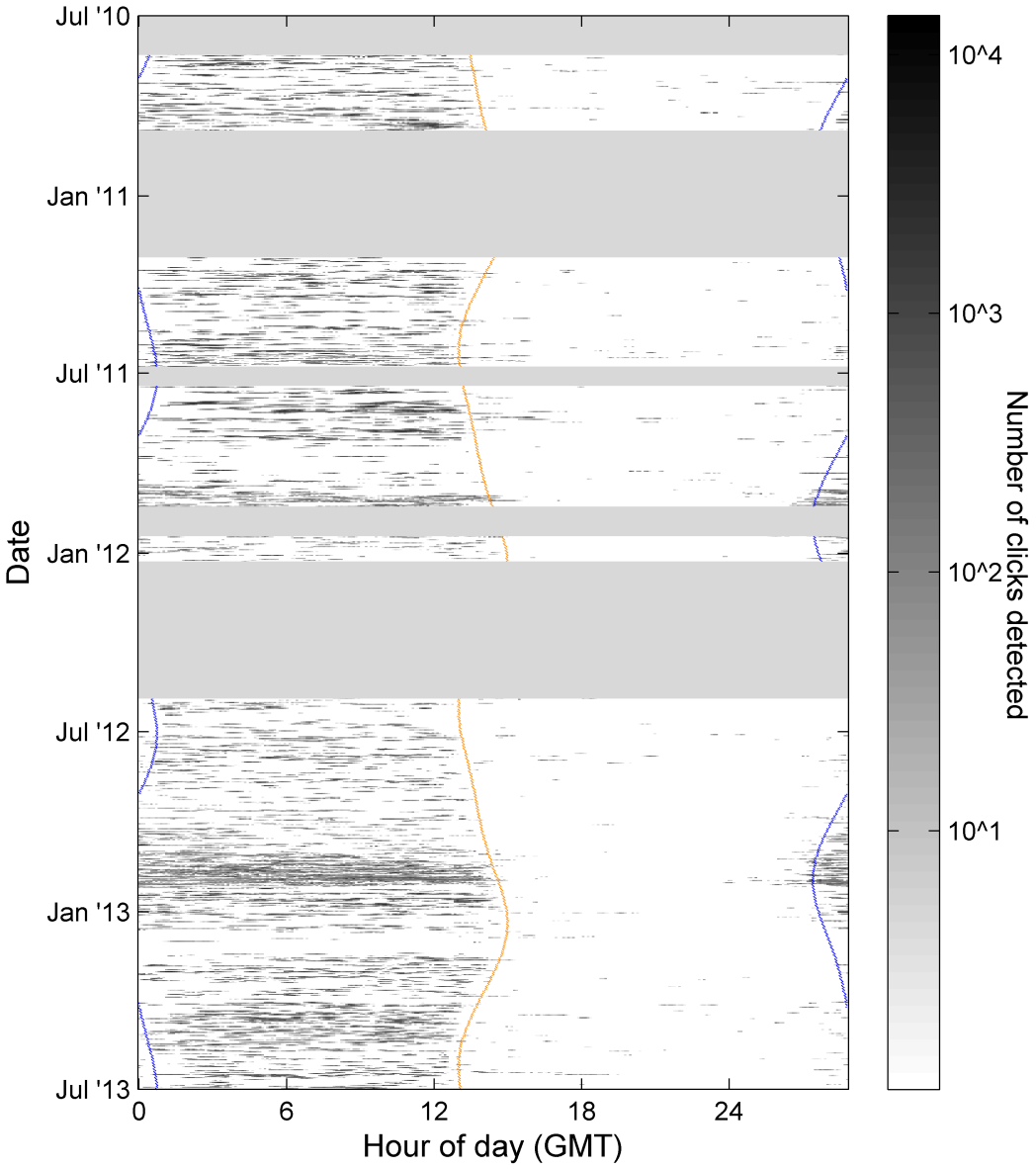


Figure 4.9: Click detections at site DT by time of day for the duration of monitoring period. Grayscale indicates number of clicks detected. Resolution is one day in the vertical, and one minute in the horizontal. Gray bands indicate gaps in time series. Orange curve indicates time of sunrise, blue indicates time of sunset.

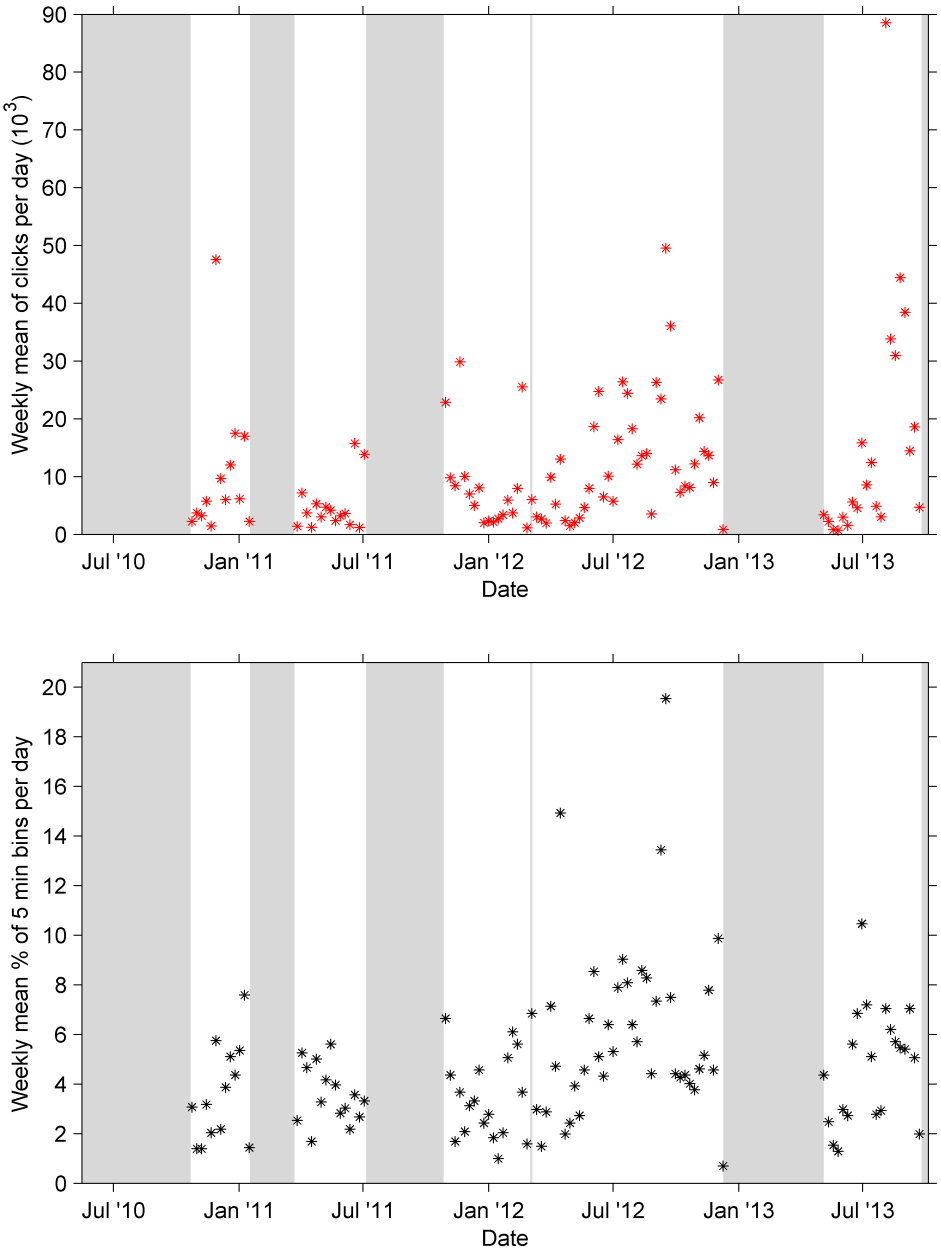


Figure 4.10: Site DC detections. *Top*: Weekly mean of daily number of click detections. Counts are adjusted by mean false positive rate. *Bottom*: Weekly mean of daily percentage of five minute intervals containing delphinid clicks. Gray bands indicate gaps in time series.

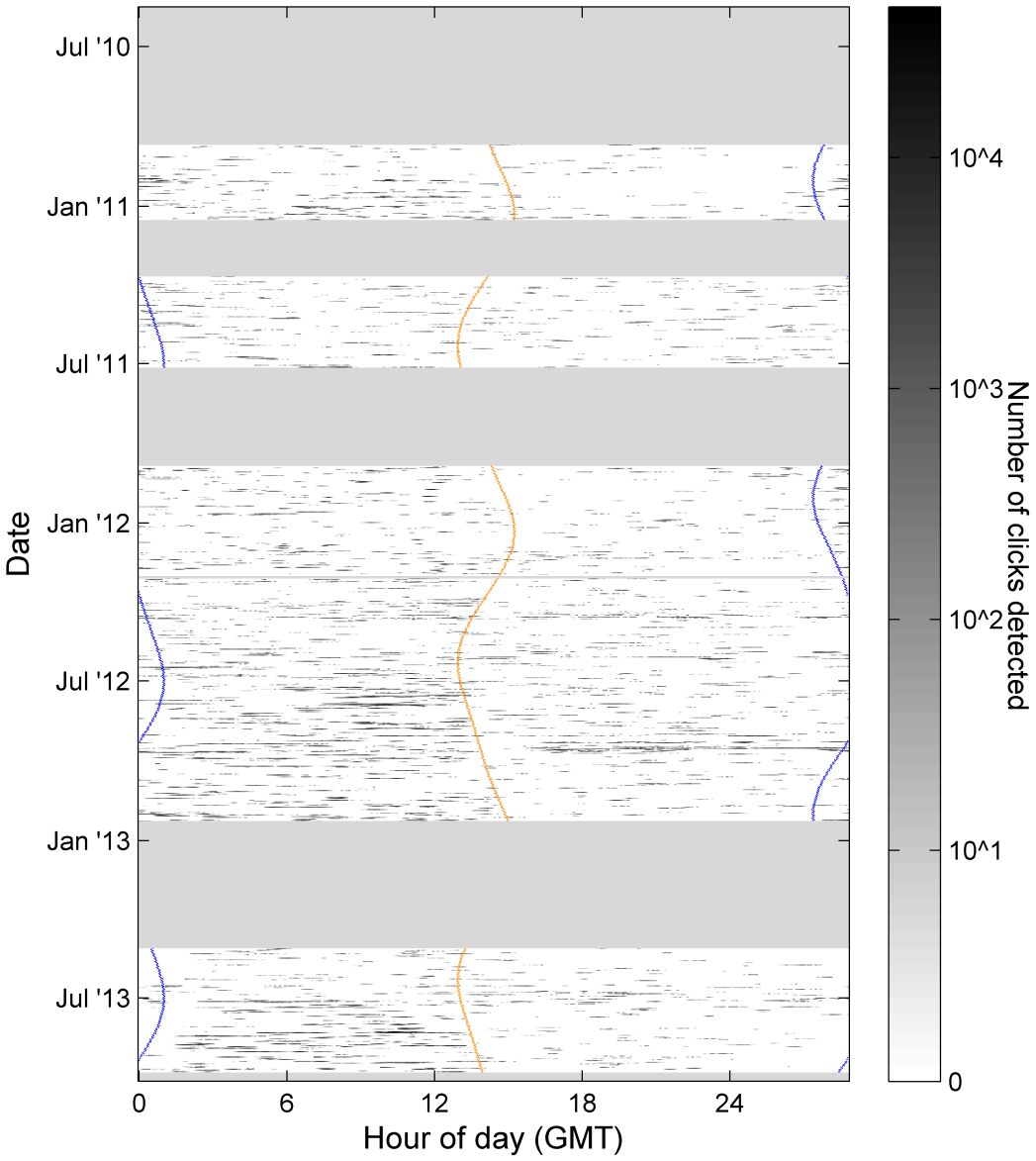


Figure 4.11: Click detections at site DC by time of day for the duration of monitoring period. Grayscale indicates number of clicks detected. Resolution is one day in the vertical, and one minute in the horizontal. Gray bands indicate gaps in time series. Orange curve indicates time of sunrise, blue indicates time of sunset.

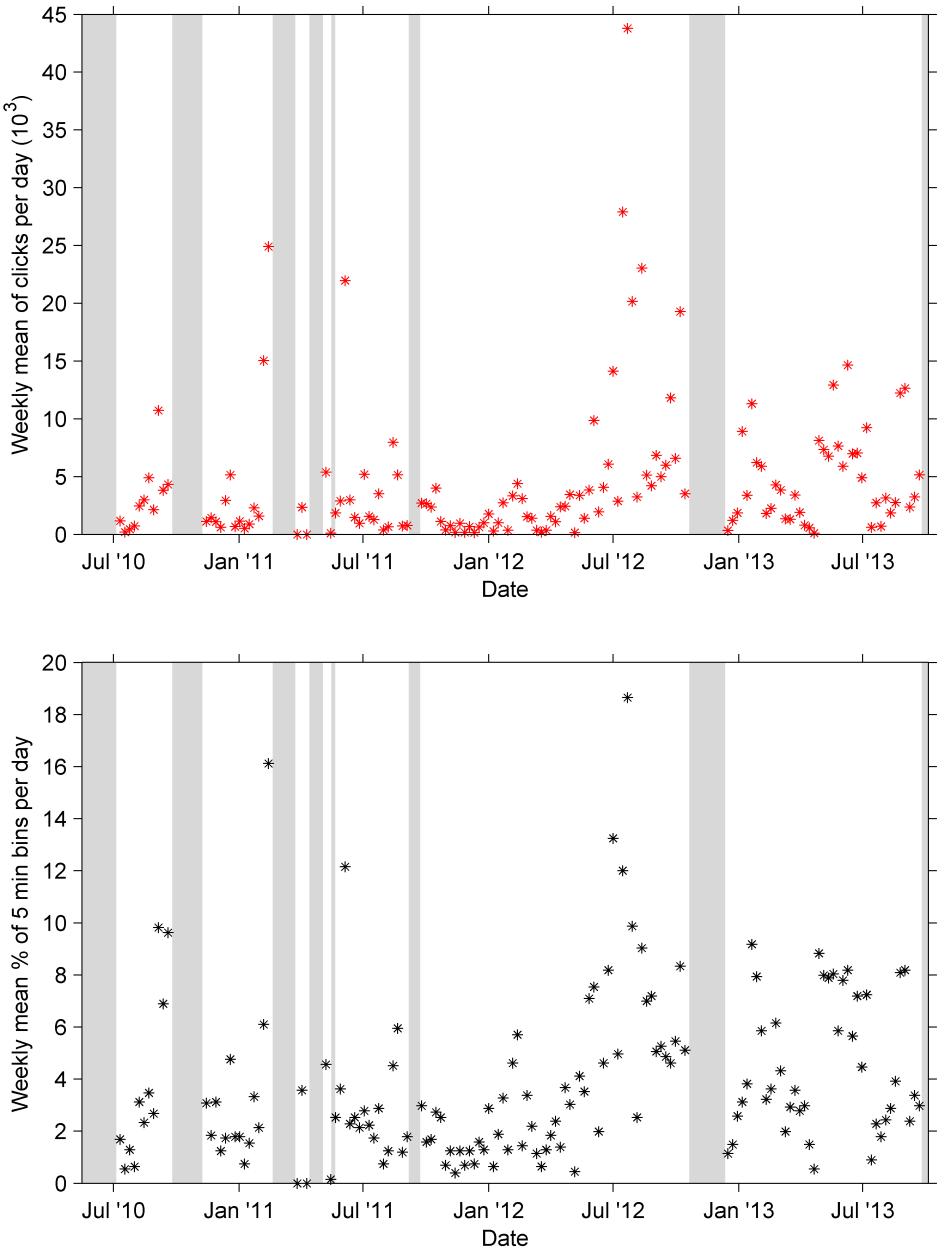


Figure 4.12: Site MP detections. *Top*: Weekly mean of daily number of click detections. Counts are adjusted by mean false positive rate. *Bottom*: Weekly mean of daily percentage of five minute intervals containing delphinid clicks. Gray bands indicate gaps in time series.

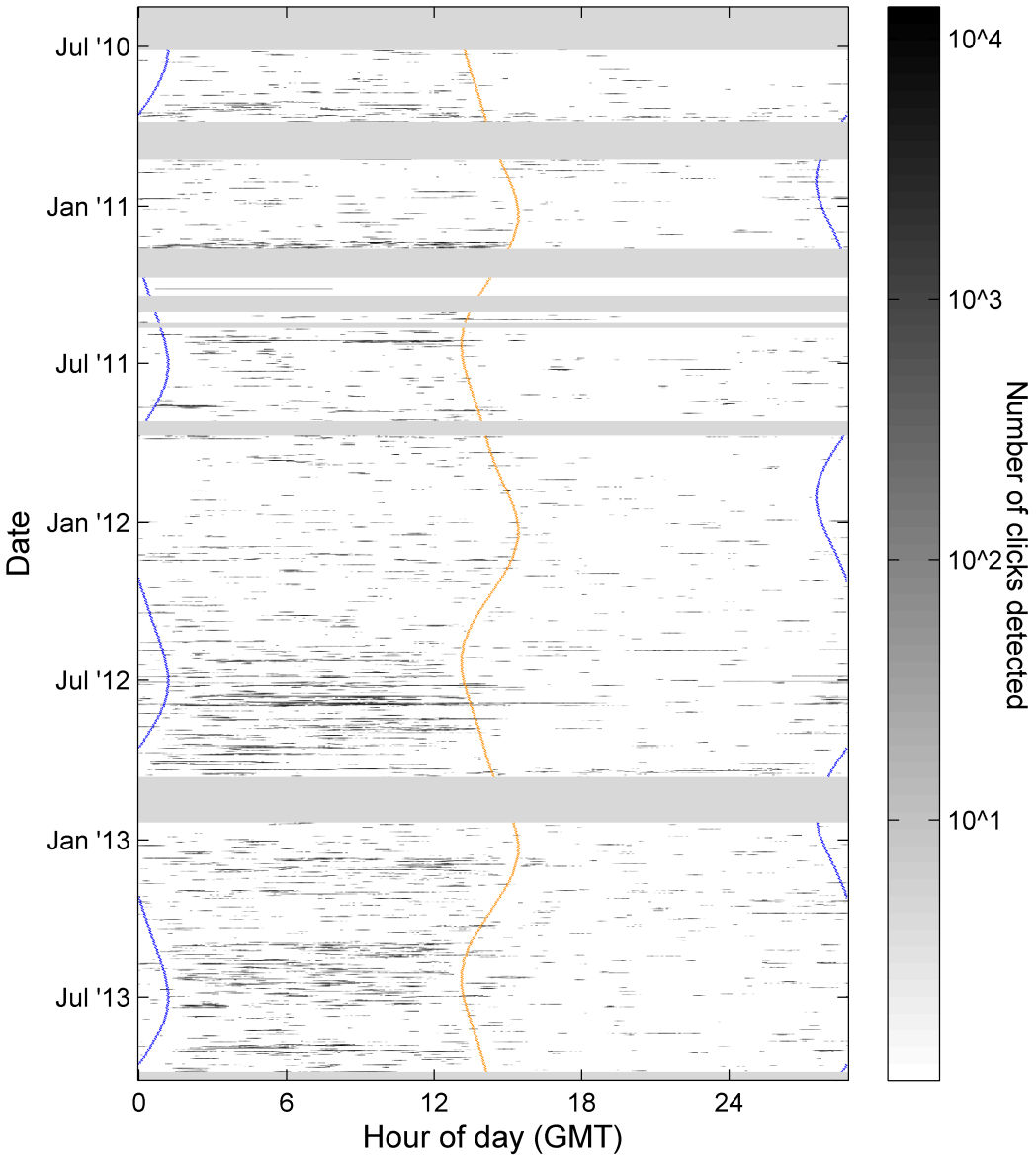
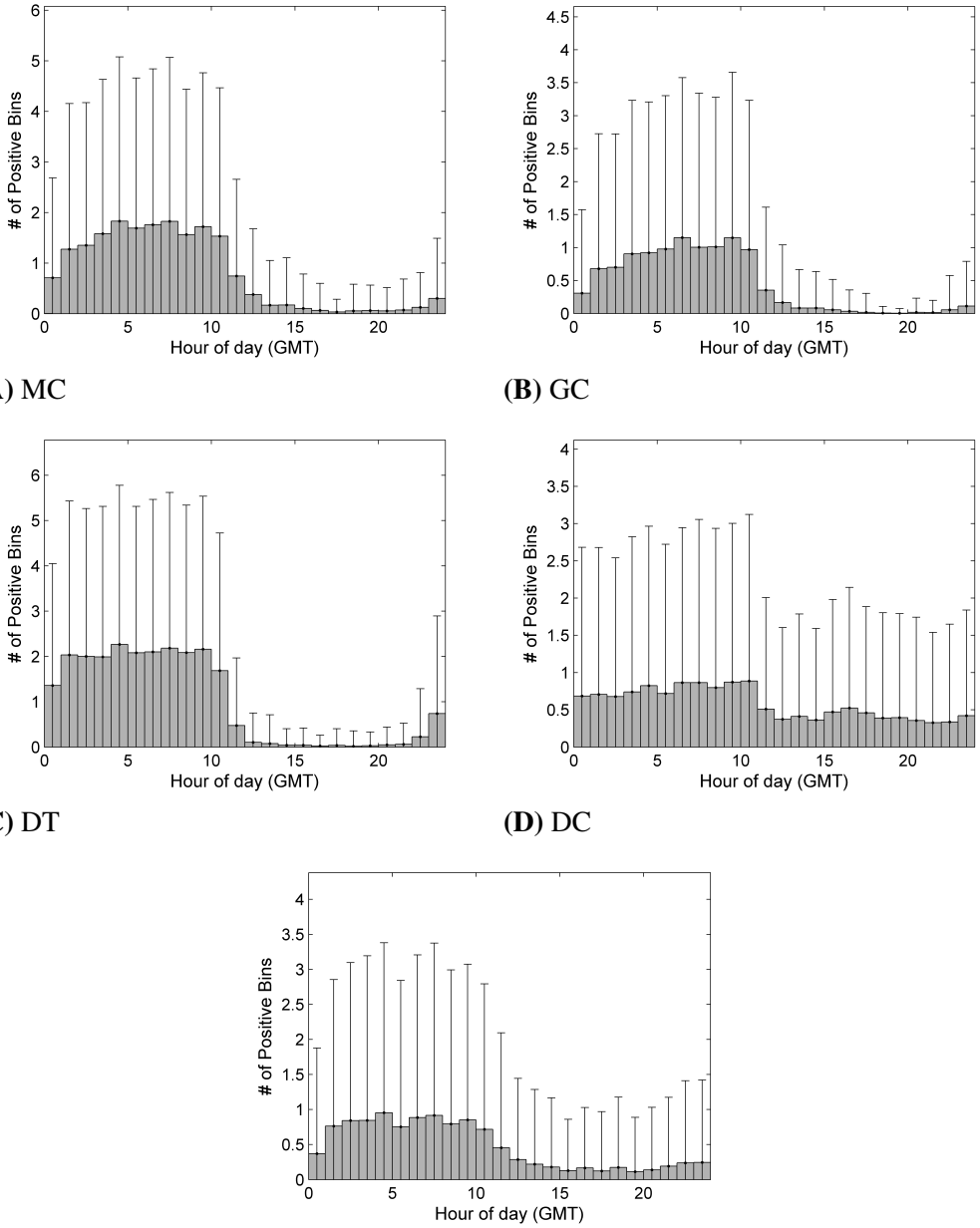


Figure 4.13: Click detections at site MP by time of day for the duration of monitoring period. Grayscale indicates number of clicks detected. Resolution is one day in the vertical, and one minute in the horizontal. Gray bands indicate gaps in time series. Orange curve indicates time of sunrise, blue indicates time of sunset.



(A) MC **(B)** GC

(C) DT **(D)** DC

(E) MP

Figure 4.14: Mean number of positive five minute bins per hour by time of day. Error bars indicate one standard deviation from the mean. Times are in GMT. Local time is GMT -05:00.

4.4.2 Detector Dead Time

The use of a detector lockout period, in which a short period of time after each detection was skipped to avoid multiple arrivals, likely has a significant effect on total click counts at shelf sites, where the lockout period was longest, during periods of high clicking activity (Figure 4.15). As a result, detection counts from the shelf sites need to be adjusted based on an estimate of missed rate, prior to use in cue-based density estimates. Detector saturation effects are estimated to be negligible at slope sites, due to the shorter lockout period.

4.5 Discussion

4.5.1 Detector Performance

These results indicate that delphinids are detected year round at each of the five sites monitored. The detector design emphasized predictability while effectively excluding non-delphinid clicks using a few targeted thresholds. Efficient manual removal of large bouts of false positives from ships made it possible to assume a uniform distribution of remaining false positives throughout the dataset. The detector performed best at the deep slope sites MC, GC and DT, where potential sources of click-like false positives were few. False positive rates were lowest at site MC, due to high levels overall of delphinid vocal activity at that site. Click counts increased exponentially as RLs approached the minimum amplitude threshold, indicating that the detector was not missing low amplitude clicks. This is critical for density estimation, and confirms that this detector is performing predictably.

The simplicity of the detector had to be compromised somewhat for shallow water detections. One way to reduce false positives at these sites without adding unwanted

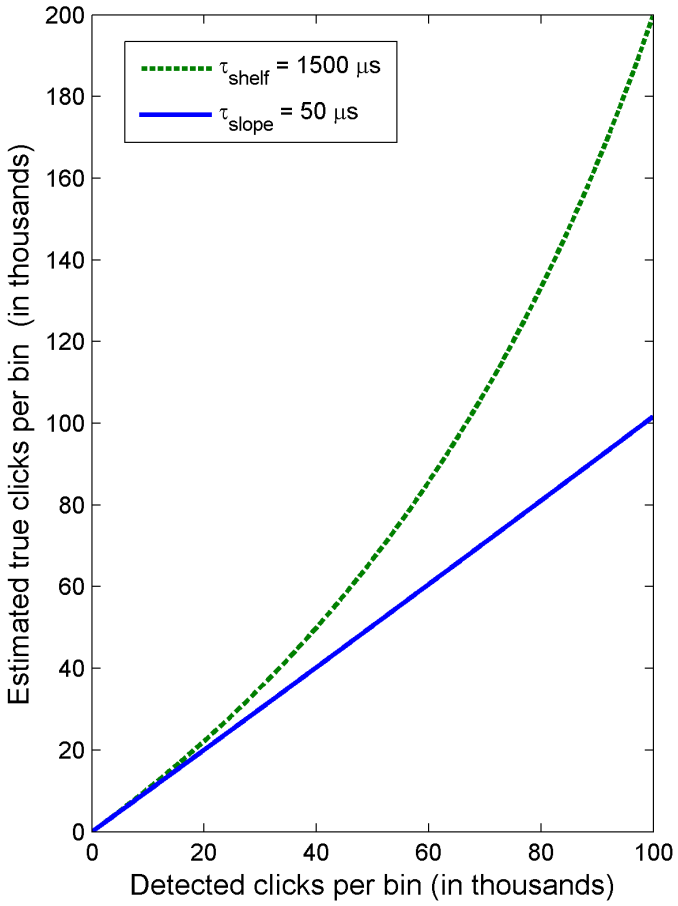


Figure 4.15: Estimated true click count after adjusting for clicks missed due to detector saturation, as a function of number of clicks detected per bin, using slope (solid blue line) and shelf (dashed green line) detection parameter settings. Bin size is five minutes. τ represents the duration of the lockout period imposed for detection purposes at shelf and slope sites, in microseconds.

thresholds and filters was to raise the minimum RL threshold. In general, higher amplitude clicks were found to be of better quality and to conform as expected to the filters already in place. However, this did not completely resolve false positive issues, particularly because higher amplitude signals are more likely to be followed by multiple arrivals.

In order to handle multiple arrivals at shelf sites, a detector lockout period was implemented after each click, before another click could be identified. This was deemed necessary to avoid detecting multiple arrivals of the same click, however it leads to a potential underestimation bias, because true detections may be missed during the lockout period. In both shallow and deep scenarios, the detector has a theoretical saturation point, which it is expected to approach asymptotically: As numbers of detections increase, so does the likelihood that the detection window around two signals will overlap. In the intermediately bad case, only one detection is made, when there should be two or more. In the worst case, no detections are made, because the detector doesn't recognize the interfering signals.

It is likely that some clicks were missed due to overlap, but the number is presumably quite small, particularly at slope sites. At the shelf sites it is much more likely that click detections obscure one another occasionally. In fact at shallow sites, once the click rate exceeds 30 clicks/sec, there is approximately a 50% chance that at least two click windows will overlap each second. In a few cases at site MP and more often at site DC, click counts exceeds 10,000 in a five minute bin, or 33 clicks/second, and the potential for interference becomes a reality. Click saturation is only an issue for cue-counting applications, but must be accounted for if a lockout period is being implemented as part of the detection strategy.

Group counting methods based on time bins have the advantage of being insensitive to saturation issues. The binning approach has a different sensitivity issue: It is theoretically possible that a single click could be detected in a five-minute window,

and this would be considered a positive bin. However in reality neither the detector nor an analyst could reliably find that single click, and attempting to do so would make the output very sensitive to false positives. A single false positive could add an extra five minutes of false detection time. For this reason, minimum click counts are required in this model before a bin is considered positive for delphinid clicks. Unfortunately, this leads to a subset of very sparsely populated bins being ignored, effectively reducing probability of detecting groups at large ranges. Ideally, a bin-based detection probability model would simulate the probability of detecting N clicks, in a five-minute time window, where N is the minimum number of clicks required. However, given the current state of the behavioral models, this is not a viable option.

Future detection improvements are possible. One promising approach is to use training data to develop a scheme capable of estimating a probability that a detection of interest was a click, rather than applying independent thresholds. This approach would improve the validity of the assumption that the model and detector are consistent, and could provide a method of estimating a false negative rate. Another technique is to identify click trains. Trained analysts rely heavily on temporal regularity as an identifying feature in noisy environments to distinguish clicks from other sounds with similar appearance in the time domain, such as snapping shrimp and cavitation. Currently, this detector makes minimal use of the temporal relationships between clicks. Incorporating such information could be particularly useful at shallow shelf sites. The use of lockout to avoid detecting echos at shallow sites could be replaced by using cross-correlation and/or a click model to determine whether a detected signal is as a second arrival or a new detection. Future analyses may look for cyclical trends in snapping shrimp noise, and use only times when snapping shrimp noise is low for density estimation.

4.5.2 Detection Time Series

Diel patterns were seen at all sites, with greater numbers of detections occurring at night, but the pattern was most prominent at slope sites. The difference between slope and shelf in terms of diel pattern may be due to species differences and/or differing foraging strategies. Pelagic animals are expected to dive at night to forage on prey in the deep scattering layer at slope sites. This may not be the primary foraging pattern at shallower sites. The predominance of clicking at night suggests that echolocation may be more typical of nocturnal behaviors at most of the monitoring sites, and may not be a good cue for evaluating density during the day at these sites. Alternatively, the animals may simply not be present at these sites during daylight hours. Acoustic propagation is not expected to differ between night and day.

The possibility that species composition differs between sites is further suggested by the differing ratios of click counts to bins per day among sites. The presence of animals that click more slowly on average is one possible explanation for the lower ratios at site DT compared to MC and GC (Figure 4.1). The poor fit of the first order polynomial to the DC dataset (Figure 4.2) may be due to the presence of two distinct count:bin relationships, which may indicate two different species groups.

Seasonal shifts in detection rates are seen seen at the two northern slope sites, MC and GC. These and other trends will be explored further in subsequent chapters, but these preliminary results indicate that findings will likely differ between northern (MC, GC, DC and MP) sites and southern (DT) sites. Before long term trends in these time series can be definitively identified, and inter-site comparisons can be made, these detection rates need to be adjusted for detectability and cue rates, when translated into density estimates.

4.6 Conclusion

A fundamental part of density estimation is understanding the likelihood that a cue will be detected given the methodology used. Here, a simplified, targeted delphinid echolocation click detector was designed to maintain predictabilty so that detector output would be appropriate for use in density estimation. This approach was explored as an alternative to characterizing an existing detector, which would require specialized tracking or localization data. Trade-offs in this case included high false positive rates under noisy conditions, and the need for a manual phase to remove concentrated instances of false positives associated with ship passages.

Clicks were detected at all five monitoring locations in the Gulf of Mexico, with highest detection rates at site MC. Time series depicting weekly averages of percentage of bins containing clicks per day were similar in shape to average click count time series. However, the relationship between these two metrics was found to differ between sites. Click detections followed a strongly diel cycle at slope sites with higher detection rates at night. This trend was weaker at shelf sites. These results provide a first glimpse at delphinid presence in offshore areas of the Gulf of Mexico over a multi-year period, on a weekly time scale.

Chapter 5

Classification of delphinid echolocation clicks in the Gulf of Mexico

5.1 Abstract

Bouts of delphinid echolocation clicks were manually classified by type according to inter-click interval (ICI) and frequency content. At least seven distinct click types were seen throughout the recording period across all slope sites. Two of these types were also seen at shelf sites, along with one additional click type not found at slope sites. An automated classification scheme was trained using the manually identified click types, and used to label click types in five minute bins throughout the time series. Automated and manual click methods are compared. These distinct, recognizable click types likely represent different species of delphinids known to be present in the monitored areas, and may be useful for species specific density estimation.

5.2 Background

5.2.1 Echolocation Click Characteristics

Dolphins produce echolocation clicks to find prey and interpret their environment (Au, 1993). Characteristics of echolocation clicks, such as frequency content, duration, and inter-click interval (ICI), are known to differ between species (e.g. Madsen et al., 2004a,b; Baumann-Pickering et al., 2010). Animal size and head morphology, as well as habitat and target prey are among the factors thought to influence the acoustic properties of clicks (Litchfield et al., 1979; Au, 2004).

In addition to species-specific click differences, captive studies have shown that individual delphinids can vary the structure of their clicks (Moore et al., 2008; Au and Benoit-Bird, 2003; Kloepper et al., 2012). Source level and beam width are among the parameters that appear to be variable both between individuals and by individuals. The highly directional nature of echolocation clicks also adds a layer of complexity to click characterization, because the structure of received signals depends on the orientation of the animal producing the sound relative to the recording sensor (Au et al., 2012a).

Despite the many sources of click variation, click-based classification of odontocetes is an active area of research (see Bittle and Duncan, 2013, for a review). Studies have shown that porpoise and delphinid echolocation clicks differ in frequency and duration between species (Kamminga et al., 1996; Akamatsu et al., 1998; Baumann-Pickering et al., 2010). Classification efforts have used these discriminating features, as well as transformations thereof, with some success (Gillespie and Caillat, 2008; Roch et al., 2007, 2011, 2015; Soldevilla et al., 2010, 2008). In general, these methods work well for a subset of species with very distinct click spectra, but struggle in cases where differences are more subtle or variable.

ICI is also used to identify delphinid click trains to species. Most notably, a

proprietary algorithm implemented in C-POD sensors (Chelonia Ltd., Cornwall, U.K.) is able identify certain species using this approach (e.g. Roberts and Read, 2015; Castellote et al., 2013), however no data is available on the algorithm's ability to distinguish between these species.

The ICI of coastal bottlenose dolphins has been shown to vary as a function of depth (Simard et al., 2010). This is consistent with evidence that captive delphinids adjust the timing of their clicks to allow for the two way travel time of the signal between the source a target (Au, 1993). However, typical ICI ranges have been shown to differ between species (Baumann-Pickering et al., 2010). ICI variability in other species, as well as across behaviors and environments, remains to be quantified.

5.2.2 Network Analysis and Clustering

The classification effort described here is aimed at using network analysis to find a subset of high quality, recurring click spectra amid a larger body of poor quality, highly variable click spectra. This subset is then used in combination with ICI as the basis for classification. Using network analysis, the relationships between numerous spectra are summarized as a map in which nodes representing clicks are linked to each other based on their similarity. Two nodes which have high similarity are expected to be neighbors in a network, with a strong connection (referred to as an edge) between them. Two nodes which are very different should be situated far apart, with a weak edge between them. Visualizing a network representation of a relational dataset facilitates the identification of unique nodes with little in common, as well as clusters of closely-related nodes. In the case of this application, the ability to identify subsets of highly similar click types within a larger body of clicks may be useful for species classification.

Many methods have been developed for partitioning networks into clusters of highly related nodes (Newman, 2004). One of these methods is to search for a network

partition that optimizes a function known as modularity (Newman, 2006). The modularity calculation is based on the idea that a good partitioning of a network should be defined by clusters with strong edges connecting nodes within clusters, and weak edges linking outward to other clusters. Accordingly, the modularity Q of a partition is a value between -1 and 1 that represents the strength or weights of the edges within clusters compared to the weights of the edges between clusters. The best partition of a network is taken to be the one that maximizes Q.

A simple example case could involve a network containing n nodes being divided into two clusters. For a pair of nodes i and j, the weight of the edge between them is defined as A_{ij}. The sum of the weights of all edges attached to a node i, is defined as k_i, such that:

$$k_i = \sum_j A_{ij} \tag{5.1}$$

The expected weight E of the edge between i and j, if edge weights were evenly distributed throughout the graph, is given by the product the edge sums k_i and k_j, divided across the weights of all of edges in the network:

$$E_{ij} = \frac{k_i k_j}{\sum_{ij} A_{ij}} = \frac{k_i k_j}{2m} \tag{5.2}$$

where

$$m = \frac{1}{2} \sum_{ij} A_{ij} \tag{5.3}$$

The modularity Q of a network partition is the sum of $A_{ij} - E_{ij}$ across all pairs of nodes i and j that fall into the same cluster. It is divided by the sum of the weights of edges in the network to ensure that Q is a value between zero and 1. In order to include the caveat that nodes must be in the same cluster to increase Q, a delta function $\delta(c_i, c_j)$ is used, in which c_i and c_j represent the clusters to which nodes i and j are assigned. If

$c_i = c_j$, then $\delta(c_i, c_j) = 1$, and it is zero otherwise (Blondel et al., 2008). The statement for Q becomes:

$$Q = \frac{1}{2m} \sum_{ij} (A_{ij} - E_{ij}) \delta(c_i, c_j) \tag{5.4}$$

Algorithms for maximizing network modularity vary, but the one used in this study is described in greater detail in Blondel et al. (2008). It uses an inverted approach with two phases which together consitute one pass over the dataset. In the first phase, each node is initially assigned to its own cluster, and the algorithm looks for cases in which reassigning a node i to the cluster of one of its neighbors j will increase Q. Iterations cease when no further re-assignments increase Q. In the second phase, a new network is created in which the clusters found in the first phase are redefined as nodes, and the weights of the edges between them are given by the cumulative weights of the edges between them. This completes the first pass. A second pass begins, applying the same two phases to the new network. Passes continue until no more changes occur, and modularity is considered to be maximized. This algorithm is designed for computational speed.

One of the benefits of the use of this modularity approach over a simpler clustering algorithm such as k-means is that the number of clusters created does not have to be specified apriori. However, modularity algorithms are known to suffer from intrinsic resolution limits (Fortunato and Barthelemy, 2007). In particular, they are not good at finding modules of extremely different sizes within a network. Given the relatively small networks considered in this work, and the fact that we are not interested in small clusters, this is not a significant issue. However, it should be kept in mind if the approach is developed further.

The goals of this chapter are:

- to describe a manual echolocation click classification method based primarily on ICI.
- to describe an automated echolocation click classification method based on unsupervised clustering of manually labeled training data.
- to compare manual and automated classification results.
- to present type-specific time series for automatically classified and manually classified types

5.3 Methods

5.3.1 Manual Classification

Echolocation clicks were detected using the detector described in chapter 4. Clicks with an RL $\geq$ 112 dB_{pp} re:1μPa were used for classification. An analyst (KF) reviewed each bout of detections, manually classifying clicks based on a long term spectral average (LTSA) of the bout, time series of ICI and RL, and mean spectra (Figure 5.1, mean spectra not shown). RL was useful for getting a sense of group passages (approach, closest point of approach, and departure) over the instrument, which can be used to distinguish click types in overlapping encounters. A bout was defined as a period of detections separated from other periods by at least 15 minutes before and after. Bouts longer than six hours were subdivided into six hour segments for manual review purposes. Recurrent click types with typical ICIs and spectra were labeled with an ID number representing that type. Clicks that did not clearly conform to any recurrent type were left unclassified. Bouts less than 75 seconds long, and/or containing fewer than 25 clicks were discarded.

Click spectra were normalized to improve comparability. First, each click was truncated to focus on the frequency content between 12 and 60 kHz, where the bulk of the distinctive spectral features were found. The slope of each click spectrum between the two end frequencies was subtracted out of the overall spectrum (based on work by S. Baumann-Pickering, pers comm). This amplified structural differences within the frequency band of interest. Click spectral amplitudes were then further normalized by re-scaling them on a [0,1] scale. After normalization of all clicks, a mean spectrum was computed from the set of spectra within each bout. The use of ICI as an additional discriminating feature allowed multiple simultaneously occurring species to be distinguished despite spectral averaging.

Clicks were classified manually by batch selection in the time series space. The large number of clicks detected at each site made individual click examination unfeasible, therefore manual classifications were subject to possible cross-contamination in cases where multiple species were detected simultaneously.

5.3.2 Automated Classification

The goal of the the automated classification step was to refine and improve the consistency of the generalized manual classification results. The automated classification algorithm included three steps: Building templates from the manually labeled data, identifying major types in test data, and assigning labels to test types by comparing them to templates.

Building Templates

The labeled data from the manual classification step was used to build spectral templates and to determine characteristic ICIs for each type. For each manually labeled click type at a given site, all associated clicks were extracted from the acoustic record.

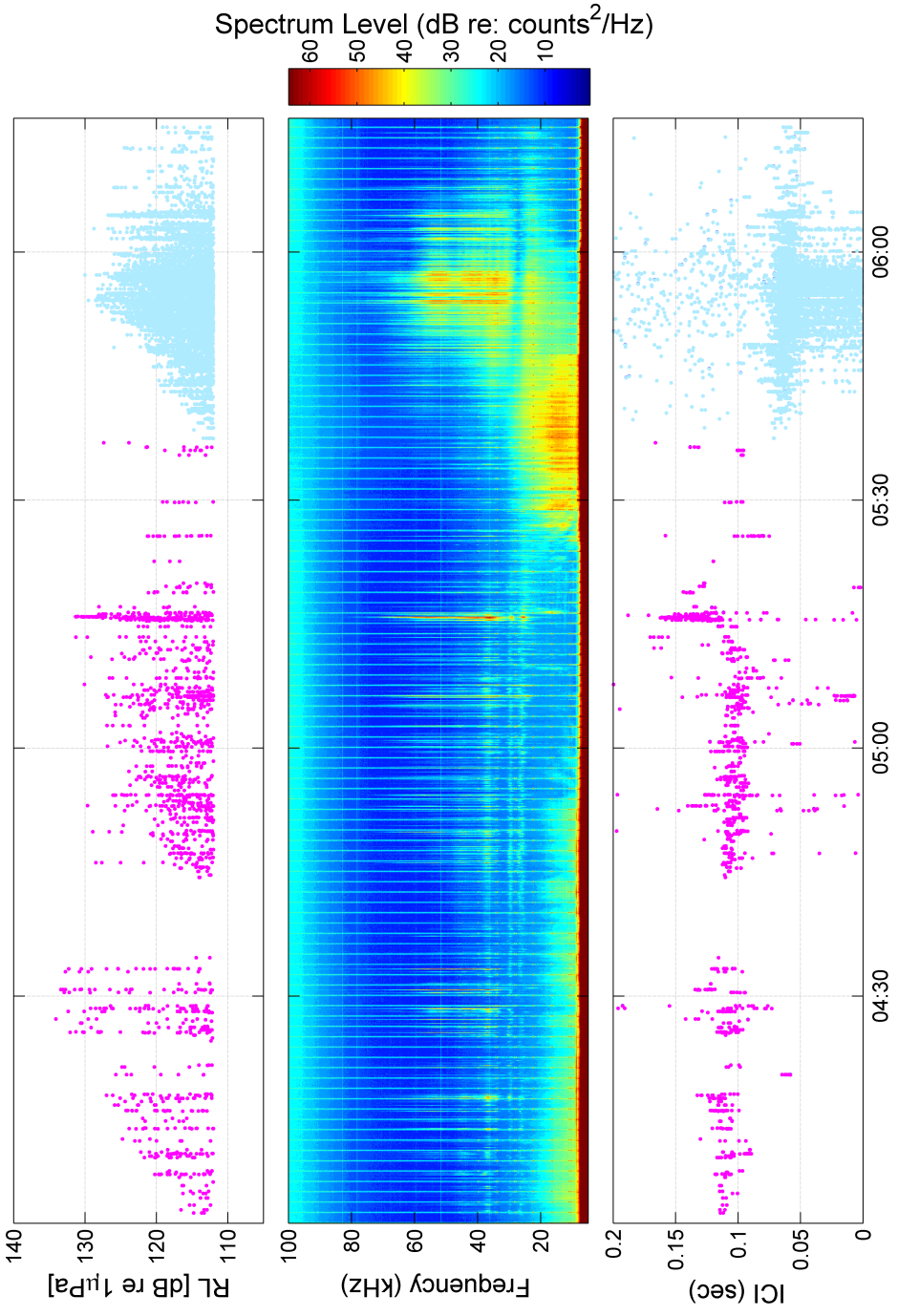


Figure 5.1: Example of manual classification step using ICI. The top panel shows received level over time, middle panel shown mean spectral energy in 5 second bins over the same period, and the lower panel shows the first difference between click times, or ICI. Pink and blue points represent two different manually identified signal types.

Clicks with received amplitudes below 118 $dB_{pp}re : 1\mu Pa$ were pruned out of the set in order to exclude poor quality clicks from the dataset used for template building.

ICI was computed for each set of high quality clicks by taking the first time difference between them. A gaussian kernel-smooth (Bowman and Azzalini, 1997) was used to fit a probability density distribution to the ICI data between 0 and 0.3 seconds. This range encompassed the variety of ICIs seen in the manual analysis step. The kernel smooth was used instead of a simpler probability distribution, because ICI distributions were complex, and in some cases multi-modal. The idea behind this approach is to estimate a density $\hat{f}(x,h)$ by centering a kernel function K over each observation x. The kernel function is used to average neighboring points. The formula for a Gaussian kernel is

$$K(x;\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} \tag{5.5}$$

where σ is the standard deviation of the input data.

The kernel probability density $\hat{f}_h(x)$ is estimated by applying K at each value of x as

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right) \tag{5.6}$$

where n is the sample size and h is the bandwidth of the kernel. For a Gaussian kernel, the optimal kernel bandwidth is computed as

$$h = \sigma\left(\frac{4}{3 * n}\right)^{1/5} \tag{5.7}$$

(Scott and Chivers, 2009; Silverman, 1986).

Characteristic spectra were computed for each high-quality click set by repeatedly selecting a randomized subset of 1,000 clicks from the labeled data. For each pair

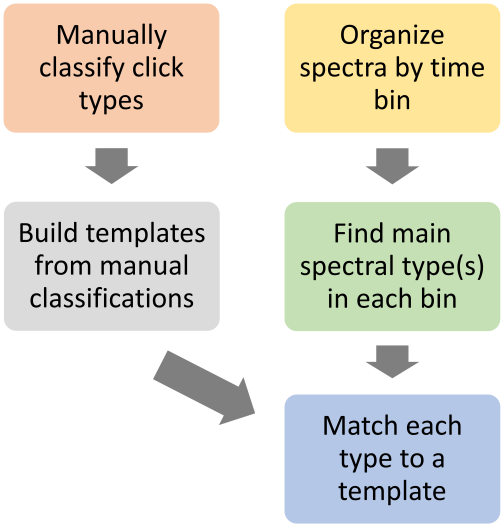


Figure 5.2: Flow chart of classification steps. A manual classification step is used to roughly assign clicks to categories. Detailed templates describing each type are then trained using the labeled data. These templates are then used automatically assign types to the detected clicks on a much finer scale.

of n-point spectra x_s and x_t, in the randomized set, a correlation distance D was computed as one minus the sample correlation between points of the normalized spectra (Equation (5.8); Pearson, 1920).

$$D = 1 - \frac{(x_s - \bar{x}_s)(x_t - \bar{x}_t)'}{\sqrt{(x_s - \bar{x}_s)(x_s - \bar{x}_s)'}\sqrt{(x_t - \bar{x}_t)(x_t - \bar{x}_t)'}} \tag{5.8}$$

where $\bar{x}_s$ and $\bar{x}_t$ are the averages of x_s and x_t respectively:

$$\bar{x}_s = \frac{1}{n}\sum_{j=1} x_{sj} \text{ and } \bar{x}_t = \frac{1}{n}\sum_{j=1} x_{tj}$$

The similarity S between two spectra was computed as exp(-D). This procedure resulted in a network in which each click was represented by a node, and S represented the strength of the connections or linkages between nodes. Values of S closer to one indicated

closely-related spectra with strong linkages, whereas smaller values indicated dissimilar spectra with weak linkages. A dynamic pruning threshold set at the 90th percentile of the distribution of network edge weights was used to trim the click networks.

Clusters were identified in the network using a modularity algorithm implemented in the network analysis tool *Gephi* (Blondel et al., 2008; Bastian et al., 2009). This algorithm does not require the user to specify a number of clusters to be identified, however a modularity coefficient can be adjusted to influence the number of clusters formed. A larger modularity coefficient (>1) favors the identification of few, larger clusters, while a smaller coefficient (<1) favors the identification of many, small clusters. A modularity coefficient of two was used in the template-building step, to favor the identification of larger, more generalized clusters. For each cluster containing 25 or more clicks, a mean spectrum was computed and retained. This process of selecting and clustering a randomized subset of clicks was repeated 50 times for each click type.

After the sets of mean spectra were generated for each click type, each set was manually reviewed. If multiple spectral variations were present in the same set, having been lumped together in the manual classification step, the types were split into subcategories for improved template matching. This was done by clustering the mean spectra once more using a pruning coefficient of 0.95 and a modularity coefficient of one (example Figure 5.3). Contaminating types were discarded.

A spectral template for each click type was created by taking the first derivative of each of the associated mean spectra, and determining the mean and standard deviation of the first derivatives at 58 evenly-spaced frequencies across the band of interest. The standard deviations were floored to a minimum of 0.005 and re-scaled to a maximum of 0.025 normalized amplitude units/kHz.

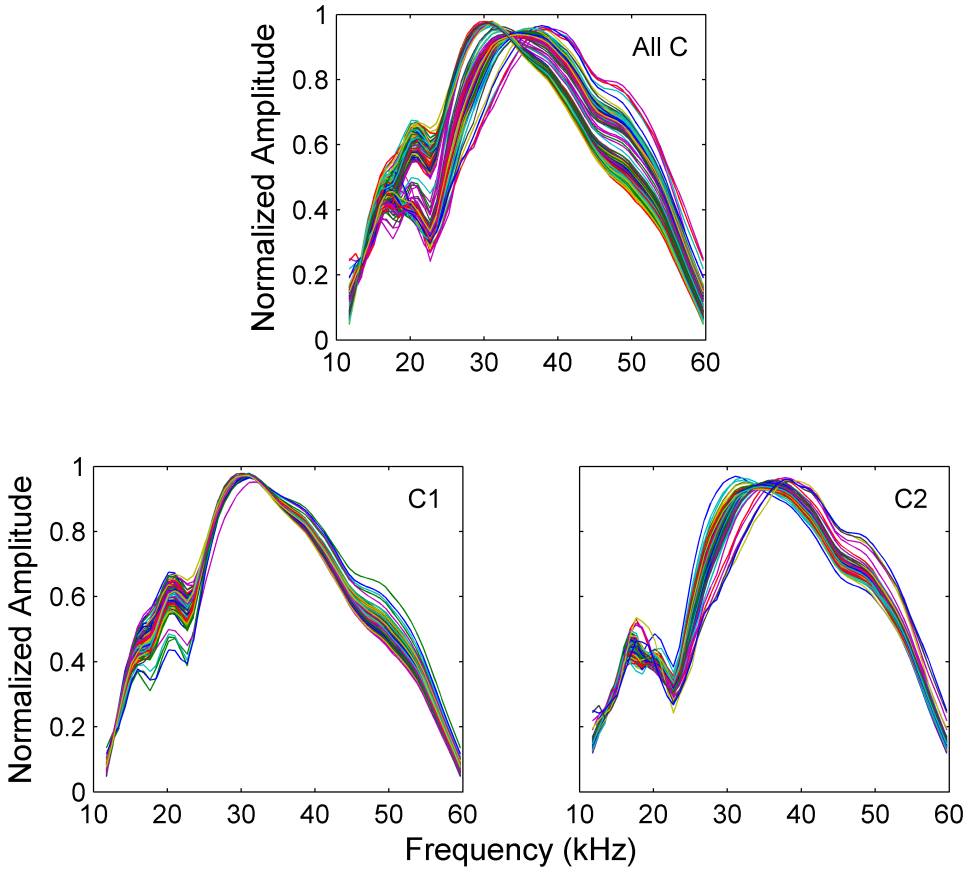


Figure 5.3: Example of mean spectra partitioning. **Top**: Set of mean spectra resulting from the template building process for click type C. **Bottom**: Two spectral variations found within the manually identified training data.

Test Data

Clicks were extracted from the recording data and batched according to consecutive five minute bins. Predominant types for each time bin were identified by clustering all clicks in the bin using the methods described above, with the same input parameters for the modularity algorithm. Mean spectra were computed and stored for each cluster with more than 100 high quality clicks. ICIs of the clicks in each cluster were also computed and retained.

The classifier did not assign a label to clicks that were not associated with a cluster of at least 100 nodes. If a five minute window contained more than 3000 high quality clicks, a randomized subset of 3000 click spectra was selected for clustering, due to computation limitations. If a window contained fewer than 100 clicks, no mean spectra were computed, and no further attempt at classification was made.

Assigning Labels

For automated classification, the first derivative of each mean test spectrum t was compared to each template T. The probability $P(T_n)$ that each point n along t came from a normal distribution defined by the template mean μ_n and standard deviation σ_n was computed along the test spectrum using a normal probability density function (f):

$$P_{T_n} = f(t_n|\mu_n, \sigma_n)$$

The resulting probability vectors were log-transformed and normalized by setting the maximum value across all vectors to zero, and subtracting that maximum from all other vectors. Each probability vector was then multiplied by a linear interpolation from one to zero (A) of the same length. This decreased the influence of the probability values as frequency increased, the rationale being that the most informative part of the spectrum

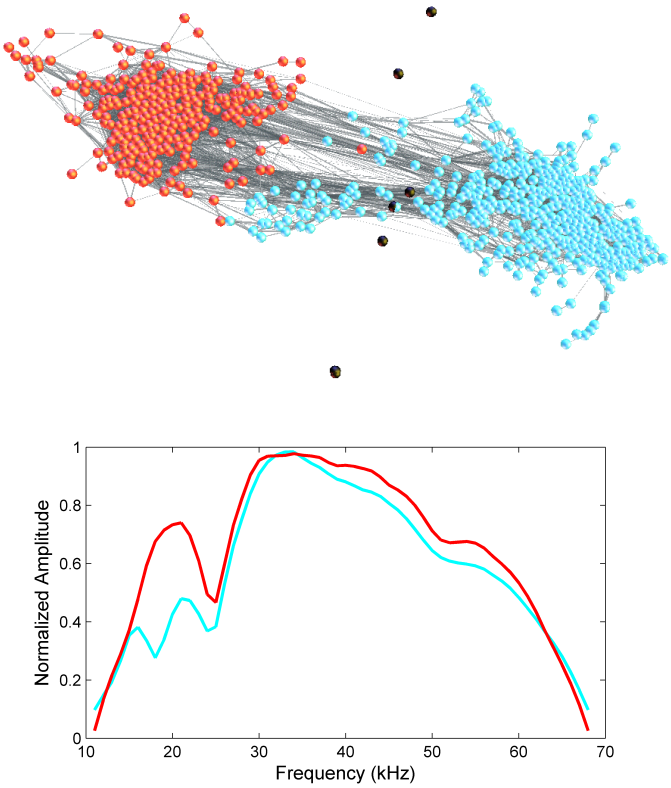


Figure 5.4: Top: Visualization of network of 736 clicks detected in a five minute bin. Each node (sphere) represents a single click in the set. Lines between nodes indicate connection strength with longer lines indicating weaker connections. Colors indicate two distinct click types identified by the clustering algorithm described in the text. Black nodes are not assigned to a cluster. Network image was generated using a force-directed layout routine Force Atlas 2 (Jacomy et al., 2014), of the graph visualization tool Gephi (Bastian et al., 2009).
Bottom: Two spectral types, B2 (red) and C1 (blue) identified in a the time bin associated with the network above.

tended to be in the lower frequencies, while variability was high and less meaningful at higher frequencies (Ibsen et al., 2007, 2009). The score S_T associated with the match between the test spectrum and each template was obtained by taking the mean of each vector.

$$S_T = \text{mean}(\log(P_{T_n}^2)) \cdot A \tag{5.9}$$

A second probability was computed using a similar strategy to estimate the likelihood P_{ICI} that the set of ICIs in the test sample came from the ICI distribution associated with each template. For each click type, the mean probability across all of the ICI values in the test bin was computed using the probability density function associated with that type. The mean ICI probabilities were log-transformed and normalized to a maximum of -1, so that ICI would have less influence than spectral shape on the goodness of fit calculated between templates and test data. This resulted in an ICI score, S_I , associated with each template:

$$S_I = log(P_{ICI}^{1/2}) - \max(\log(P_{ICI}^{1/2})) - 1; \tag{5.10}$$

The match score (M) between the the template and spectrum was computed as $M = S_T * S_I$. The template with the lowest score was the best match (M_{best}). If M_{best} was less than a threshold value of ten, the test cluster was assigned the template label, otherwise, the test cluster type was unknown. This threshold value was chosen based on manual review of a subset of automated click assignments to prevent poor matches from being retained.

5.3.3 Comparing Manual and Automated Results

Manual and automated classification results were compared by totaling numbers of click detections assigned to each click type, and by reviewing plots of click-type specific time series obtained by each method. Because subtypes were not recognized manually, comparisons were conducted only at the level of the more generalized manual click types. Confusion matrices were generated by comparing the manual classifications to the automated classifications by bin. Since the true click type is not known, neither method can be said to represent true classifications. Further field data is needed to assess the accuracy of both methods.

5.4 Results

5.4.1 Manual Classification

Manual review of the click detections from the three slope sites MC, GC and DT revealed approximately seven visually distinguishable click types (A through G) (Figures 5.6, 5.12 and 5.17). Examples of each type were seen at all three slope sites, with the exception of type F, which was not seen at site DT. Click types B, E and F, seen at the slope sites were also seen at the deepest shelf site DC (Figure 5.22). Type K was not equivalent to any of the types seen at slope sites. Only one distinguishable type, B2, was found at site MP. This type was comparable to type B2 from the other four sites (Figure 5.27). High false positive rates at site MP contaminated mean spectra and ICI distributions, limiting the feasibility of manual classification.

In the manual classification phase, click types were recognized primarily by ICI distribution, and secondarily by mean spectra. Mean spectra were not manually categorizable on their own with the visual method described, because averaging across

many clicks made spectral differences difficult to distinguish, particularly in mixed encounters. Click type A was distinguished by its large ICI, (mean near 0.2 seconds), and lower frequency spectra (Figures 5.9, 5.14 and 5.24). Click type B was the most common across all sites (Tables 5.3 to 5.5). Although type B appeared during manual analysis to include multiple spectral types with roughly the same ICI, the types were too intermingled to be consistently discerned. Type C was closely related to type B (similar ICI, and often co-occurring), but different enough to distinguish, due to two small energy peaks at approximately 16 and 22 kHz. Types D, F and G were distinguished primarily by ICI. Types E and K were recognized spectrally based on their distinct and consistent spectral shapes. Type E is associated with Risso's dolphin based on previous work by Soldevilla et al. (2008).

Overall, approximately 98% of clicks detected at the slope sites and 94% of click detected at site DC were assigned a manual identification (Table 5.1). Manual classification was less successful at site MP, where roughly half of the detected clicks were assigned a manual identification. The percentages of total bins classified by site using the manual method were similar to the click count percentages (Table 5.2)

All click types were detected primarily at night, with the exception of type K from site DC, which was detected at roughly the same rate regardless of time of day (Figure 5.5).

5.4.2 Automated Classification

Automated classification revealed that some of the visually identified click types were consistent spectrally, while others contained multiple subtypes. A single typical spectrum characterized types A, F, G and K. Click type B contained multiple distinct spectral types at all sites except for MP, therefore type B was split into three subtypes, B1, B2, and B3 for classification purposes (Figures 5.6, 5.12 and 5.12). All versions of

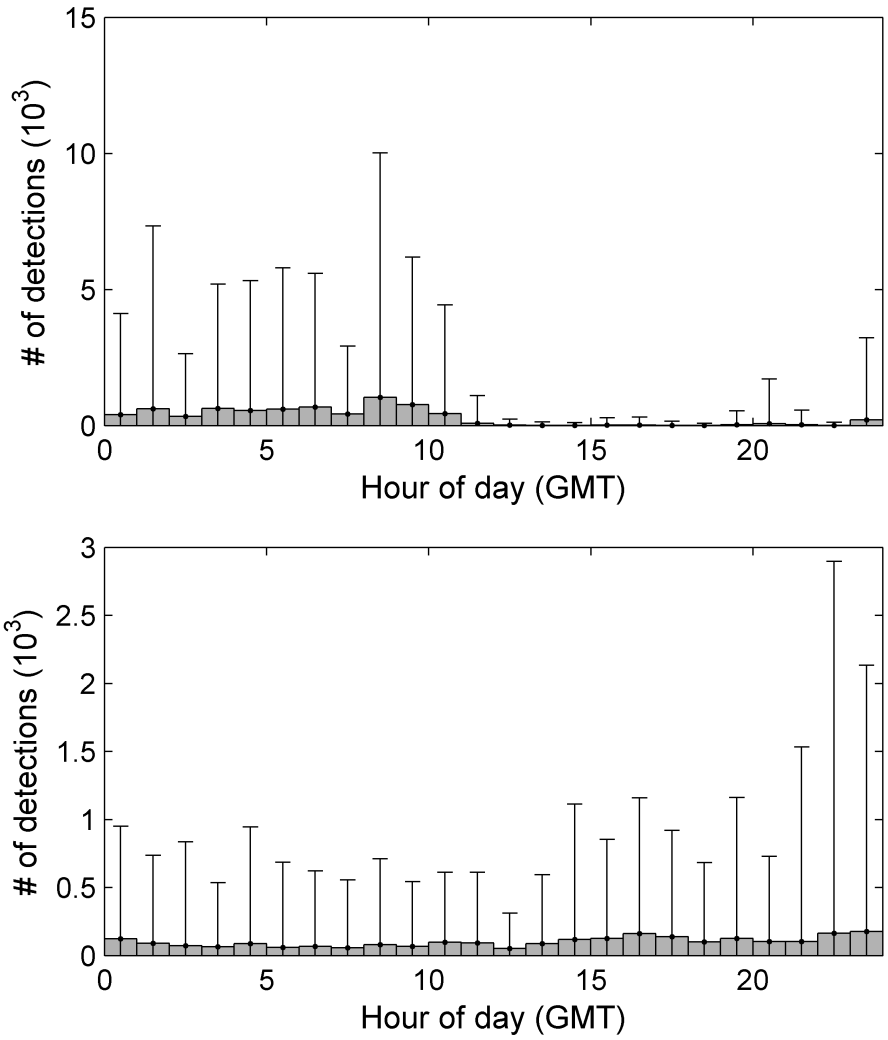


Figure 5.5: Top: Mean number of type B clicks detected by hour of the day at slope site DC. This nocturnal pattern was typical of most click types. **Bottom**: Mean number of type K clicks detected by hour of the day at slope site DC. This click type was unique in that it was found during all hours of the day. Error bars indicate one standard deviation above the mean.

Table 5.1: Total clicks with classified by site using manual and automated methods. Total classified clicks are provided both as counts and as percentages of all detected clicks.

Site	**Manually Classified**		**Auto Classified**		**Total Clicks**
	counts	% of total	counts	% of total	
MC	12,868,599	(98%)	12,796,841	(98%)	13,089,754
GC	4,750,237	(98%)	4,760,356	(98%)	4,842,472
DT	5,780,329	(98%)	5,727,053	(97%)	5,924,725
DC	8,692,519	(94%)	8,705,755	(94%)	9,225,644
MP	2,804,598	(49%)	4,997,696	(86%)	5,711,326

Table 5.2: Total bins classified by site using manual and automated methods. Total classified bins are provided both as counts and as percentages of all positive bins (bins containing clicks).

Site	**Manually Classified**		**Auto Classified**		**Total Positive**
	counts	% of total	counts	% of total	**Bins**
MC	20,129	(96%)	19,891	(95%)	20,949
GC	10,374	(96%)	10,316	(95%)	10,809
DT	18,490	(98%)	17,048	(90%)	18,911
DC	9,325	(89%)	7,292	(70%)	10,463
MP	5,514	(48%)	8,054	(69%)	11,594

click type B had very similar ICIs.

Click types C, D, and E were each split into two subtypes for most slope sites, to better capture the diversity of associated mean spectra, but it was not clear whether the subtypes were truly distinct, or if they represented a continuum of variability within a single type. Click types at the shelf sites were noisier and less consistent. Click type K was very common at site DC, however the type had high spectral variability, and no clear ICI mode, making it difficult to recognize automatically.

ICI probability distributions varied between types (Figures 5.10, 5.15, 5.20, 5.25

and 5.30). ICI variability was higher, reflected by broader ICI peaks, for types A, D and E, which also had larger ICIs on average than the other types. Types B, C and F had lower ICI variability, and lower, more distinct modes. These types also often had multiple smaller peaks associated with multiples of the main ICI peak, as reflected in the ICI probability distributions from the slope sites. Complexity in the ICI pdfs was captured by the Gaussian kernel smooths. Secondary ICI peaks at multiples of the main peak are likely associated with cases in which a click in a click train is missed, while the clicks on either side of it are detected. They were not visible in the ICI distributions at the shelf sites (Figures 5.25 and 5.30).

Slope site ICI distributions were also characterized by peaks at very low ICIs (0.02 seconds or less). These low ICI peaks are likely due to false positives, a higher probability of detecting off-axis clicks at these shallow sites, and a higher probability of detecting multiple animals simultaneously.

5.4.3 Counts and Time Series Comparison

Total manually classified click and bin counts were very similar to automatically classified counts (Tables 5.3 to 5.7). Confusion matrices indicate however that there was significant confusion between types, and that there is room for improvement. It is important to remember that interpretation of the confusion matrices is limited by the fact that they only compare the manual and automated classifications to each other, not to reality.

At slope sites, click counts assigned to types A through E were generally similar between manual and automated methods. Type B was so common at slope sites that it was confused more often between the two methods, in absolute terms, but on a relative scale it appears to have been strongly identifiable. Large differences between the two methods were seen for types F and G at slope sites. Bin counts were more variable

between the two methods than click counts. Automated bin counts were typically higher because the automated method allowed multiple types to be found in a single bin fairly often, while the manual method rarely did, due to resolution limitations. Interestingly, at slope sites there is little agreement in what falls into the unknown type categories suggesting that the two approaches have different weaknesses.

At site DC, click counts assigned to click types B and F were comparable between manual and automated methods. Type K was not matched as frequently as expected with the automated approach. Type E was identified more often by the automated method than the manual method. Bin followed the same trends as cue counts. At site MP, the automated classifier assigned more clicks to the type B category than the manual method did (86% vs. 49%).

The manual and automated classification time series were similar overall shapes for most click types (figs. 5.31 - 5.54). The automated classification results indicate that certain B subtypes occurred at different times over the course of the monitoring period at all three slope sites (Figures 5.32, 5.39 and 5.46). Subtypes of other click types do not appear to be temporally separated.

5.5 Discussion

5.5.1 Manual Classification

Manual classification proved essential for developing templates and understanding the relationships between click types found in the data. However, manual classifications may be inconsistent across large deployments, and the ability to distinguish between mingled types was limited. It was clear from the data that different types of clicks were present, and we expected that we would be able to distinguish them to some level using spectral shape and ICI information. Further refinements are needed to define some the

more similar types, while more unique types are fairly easily classifiable. Similar findings have been seen in other attempts to distinguish delphinid vocalizations across species (e.g. Baumann-Pickering et al., 2010; Roch et al., 2015; Soldevilla et al., 2008).

Further data, such as seafloor sensor recordings associated with visual species IDs, are needed to refine click types, and distinguish between within-type vs. between-type variation. This would improve the quality and consistency of training data, and likely lead to improved templates, and better classification results. Given further training data, some of these types might be merged, while others might be teased apart. Delphinid species diversity in the Gulf of Mexico is fairly high, and the number of click types documented manually is in keeping with the 10 to 14 species expected around these sensors. However, the manual classification process applied here is aimed at recognizing dominant click types in the data, and rare types associated with rare species were likely overlooked.

5.5.2 Automated Classification

The automated classification step provided a method for refining the manual classifications, teasing apart intermingled click types and operating consistently across datasets. The expectation was that an algorithm could be developed based on the analysts' manual classification process, and the automated method might be more consistent across the large volume of data. The findings in this chapter support that expectation, but also reflect the fact that in a supervised learning setup like this, the automated classifier is only as good as the training data provided by the manual process. For click types F and G the manual and automatic classification output did not agree well, because the two types were not clearly distinguished in the manual training step.

The use of two lines of evidence, ICI and spectrum, for classification led to better results than using either independently. Other choices within the automated classification method, such as weighting spectral fit at lower frequencies more than at

higher frequencies, and balancing the contribution of ICI and spectrum fits to overall template scores, significantly improved performance. The use of first derivative of frequency to capture shape significantly improved the matching ability of the algorithm. More complex approaches such as the use of click cepstrum and first derivative of cepstrum could further improve the quality of template matching.

Click classification suffers when the number of clicks in a bin is low. In these cases, clicks are often low amplitude, with noisy spectra, and ICI distributions are sparse. Handling of sparsely populated bins was improved by implementing a variable network pruning threshold. Smaller click networks were pruned less harshly than larger networks using an edge-weight distribution-based threshold rather than a static threshold. This increased the likelihood that a summary spectrum would be produced for sparsely populated bins. For large networks, a higher pruning threshold tends to result in clearer, more consistent summary spectra, by reducing the influence of outlier spectra and speeds up processing times. The drawback is the potential loss of minor types from the classification pool as more and more nodes become fully isolated (i.e. all linkages to the rest of the network are severed).

A number of refinements are possible and may be implemented in the future. In its current form, the template approach still becomes unreliable when operating on a small set of input clicks. In cases where a bin has no clear summary spectra, the classifier cannot reliably assign a type. The ICI data can also become unreliable in cases where there are very few data points. Those same bins might be classifiable by a human analyst through more sophisticated pattern recognition and incorporation of contextual information.

Contextual information could be incorporated into the automated classifier by considering classifications in neighboring time bins, such that the probability that a bin preceded by click type A and followed by type A, would have a higher than average

probability of containing type A itself. Expanding time bin durations would also likely improve classifications, however longer time bins increase the probability that animals either enter of leave the detection range during a bin, posing problems for density estimation later on.

Developing a method for adjusting the score contribution from ICI as a function of the number of datapoints available might also improve automatic classification performance. If an ICI distribution is robust, it could be given more weight than if only a few data points were available.

The template matching approach has an innate weakness in that it performs best if all possible click types are defined. If a click type is not represented in the body of templates, it is possible that clicks belonging to the missing type will be assigned to an alternate type. This appears to be the case at site DC, for instance, where a manual review of the template matches revealed that a type not represented in the template set was intermingling with type E inflating the counts associated with that type. Another issue is that weak templates, characterized by high variability, few examples, or very generic mean spectra and ICI distributions, may score more highly than strong, well-defined templates. This was partially addressed in the classification scheme, by normalizing the standard deviations of the spectral templates over a common interval. However, types F and G remained fairly generic. These types were weak due to confusion with type B during the manual classification process.

The ultimate solution to this is probably to collect visually verified, species-specific training data, and to build classification templates from that data.An alternative might be to use an unsupervised learning approach to identify distinct click types in the data without a manual input step.

5.5.3 Time Series

The click types described here are likely each associated with different species. An initial look at possible identifications is covered in the following chapter, although more research is required to refine the types and finalize species. The fact that these types occur at different times over the course of the monitoring period, are common across sites, and occur in different quantities, supports the idea that these are in fact different species.

Type B is probably associated with multiple species, given the variety of subtypes found. Unlike the subtypes of C, D and E, some of the subtypes of B are have distinct energy distributions from the others. A pulse of type B2 in the summer of 2012 at sites MC and GC may be indicative of a temporary species composition shift during that period. At site, DT a shift from predominantly type B2 in the first half of the time-series to predominantly type B3 in the second half is also noteworthy.

There is likely variability both spectrally and in terms of ICI within each click types which is not indicative of different species, but rather of other factors including behavior, oceanographic conditions, or individual differences. This is in keeping with established research on delphinid echolocation click properties and variability.

5.6 Conclusion

Manual and automated echolocation click classification methods identified click types with distinct spectra and ICI distributions. An automated time series generated based on both methods indicate that different types are encountered at different rates and times, but are common across sites. These click types may be associated with different species of delphinids frequently encountered in the monitored areas. Applications include species-specific density estimates based on passive acoustic recordings.

Table 5.3: Click and bin counts associated with each click identified at site MC. Manual and automated classification results are shown for each type. Subtypes were identified by the automated classification scheme which were not identified manually.

Type	Manual ID clicks	Auto ID clicks	Manual ID bins	Auto ID bins
A	55,836	43,583	242	261
B1	n/a	3,416,460	n/a	6,313
B2	n/a	1,704,513	n/a	3,991
B3	n/a	2,467,035	n/a	7,416
B (all)	6,671,061	7,588,008	11730	14,745*
C1	n/a	843,524	n/a	1,805
C2	n/a	251,517	n/a	764
C (all)	1,094,861	1,095,041	1,402	2,273*
D	764,387	672,015	742	1894
E1	n/a	245,425	n/a	1479
E2	n/a	120,767	n/a	721
E (all)	204,575	366,192	1,687	1,985*
F	2,729,558	805,372	1,712	1,312
G	1,348,321	2,226,630	2,563	4,479
unknown	221,155	292,913	1,251	1,058
TOTAL	**13,089,754**	**13,089,754**	**20,949***	**20,949***

*Bins do not sum to this value because some bins contain multiple click types.

Table 5.4: Click and bin counts associated with each click identified at site GC. Manual and automated classification results are shown for each type. Subtypes were identified by the automated classification scheme which were not identified manually.

Type	Manual ID clicks	Auto ID clicks	Manual ID bins	Auto ID bins
A	49,254	51,003	253	338
B1	n/a	950,244	n/a	3529
B2	n/a	844,188	n/a	2329
B3	n/a	785,880	n/a	3516
B (all)	2,751,348	2,580,312	6,844	7,351
C1	n/a	586,614	n/a	1,975
C2	n/a	308,787	n/a	826
C (all)	849,492	895,401	1,677	2,545*
D1	n/a	156,537	n/a	543
D2	n/a	447,966	n/a	1,575
D (all)	579,476	604,503	652	1,878*
E1	n/a	27,229	n/a	216
E2	n/a	18,491	n/a	88
E (all)	17,448	45,720	230	278*
F	99,740	292,093	81	924
G	403,479	290,724	722	802
unknown	92,235	82,716	435	493
TOTAL	**4,842,472**	**4,842,472**	**10,809***	**10,809***

*Bins may not sum to this value because some bins contain multiple click types.

Table 5.5: Click and bin counts associated with each click identified at site DT. Manual and automated classification results are shown for each type. Subtypes were identified by the automated classification scheme which were not identified manually.

Type	Manual ID Clicks	Auto ID clicks	Manual ID bins	Auto ID bins
A	37,156	21,944	253	301
B1	n/a	7,15,199	n/a	3150
B2	n/a	2,274,587	n/a	6747
B3	n/a	1,384,202	n/a	6531
B (all)	4,690,477	4,373,988	12,373	12,572*
C	140,613	69,064	496	537
D1		152,314		927
D2		131680		687
D (all)	130,479	283,994	97	1,502*
E1	n/a	187,440	n/a	1,885
E2	n/a	318,166	n/a	2,403
E (all)	554,847	505,606	5,192	3,803*
F	31,390	237,746	61	786
G	195,367	457,992	362	1,014
Unk	144,396	197,672	421	1,863
TOTAL	**5,924,725**	**5,924,725**	**18,911***	**18,911***

*Bins may not sum to this value because some bins contain multiple click types.

Table 5.6: Click and bin counts associated with each click identified at site DC. Manual and automated classification results are shown for each type. Subtypes were identified by the automated classification scheme which were not identified manually.

Type	Manual ID Clicks	Auto ID clicks	Manual ID bins	Auto ID bins
B1	n/a	542,866	n/a	841
B2	n/a	313,727	n/a	559
B3	n/a	5,472,880	n/a	5170
B (all)	5,414,470	6,329,473	3,749	5,597*
E	33,231	142,640	166	382
F	1,433,230	1,307,691	1,060	942
K	1,811,588	925,951	5,119	2609
Unk	533,125	519,889	881	2,197
TOTAL	**9,225,644**	**9,225,644**	**10,463***	**10,463***

*Bins may not sum to this value because some bins contain multiple click types.

Table 5.7: Click and bin counts associated with the single click type identified at site MP. Manual and automated classification results are shown for this type.

Type	Manual ID Clicks	Auto ID clicks	Manual ID bins	Auto ID bins
B	2,804,598	4,997,696	5,514	8,054
Unk	2,906,728	713,630	6,080	3,540
TOTAL	**5,711,326**	**5,711,326**	**11,594**	**11,594**

Table 5.8: Site MC: Confusion matrix comparing manual and automated classification results by five minute bin. Numbers on the diagonal (**bold**) indicate the total number of bins that were given the same classification by both methods. Values off of the diagonal indicate bins that were classified differently by the two methods.

		Manual								
		A	**B**	**C**	**D**	**E**	**F**	**G**	**Unk**	**Total**
Automatic	**A**	**108**	6	3	7	0	1	6	52	183
	B	11	**10,289**	194	30	19	308	393	2,219	13,463
	C	2	164	**886**	4	3	2	17	802	1,880
	D	10	113	3	**343**	5	50	37	741	1,302
	E	2	50	5	63	**1,171**	10	11	347	1,659
	F	2	40	13	0	7	**555**	4	457	1,078
	G	2	97	18	33	47	82	**1,705**	1,350	3,334
	Unk	64	610	74	79	408	45	231	**900**	2,411
	Total	201	11,369	1,196	559	1,660	1,053	2,404	6,868	25,310

Table 5.9: Site GC: Confusion matrix comparing manual and automated classification results by five minute bin. Numbers on the diagonal (**bold**) indicate the total number of bins that were given the same classification by both methods. Values off of the diagonal indicate bins that were classified differently by the two methods.

		Manual								
		A	**B**	**C**	**D**	**E**	**F**	**G**	**Unk**	**Total**
Automatic	**A**	**123**	15	1	2	0	1	1	100	243
	B	23	**5,727**	219	24	3	5	55	870	6,926
	C	2	275	**1,148**	5	1	2	13	744	2,190
	D	16	88	25	**518**	15	4	34	810	1,510
	E	2	15	11	4	**123**	0	1	72	228
	F	5	63	27	3	1	**48**	7	460	614
	G	2	38	6	2	0	1	462	**196**	707
	Unk	48	367	83	55	84	2	77	**186**	902
	Total	221	6,588	1,520	613	227	63	650	3,438	13,320

Table 5.10: Site DT: Confusion matrix comparing manual and automated classification results by five minute bin. Numbers on the diagonal (**bold**) indicate the total number of bins that were given the same classification by both methods. Values off of the diagonal indicate bins that were classified differently by the two methods.

		Manual								
		A	**B**	**C**	**D**	**E**	**F**	**G**	**Unk**	**Total**
Automatic	**A**	**108**	8	0	1	2	0	1	47	167
	B	24	**11,143**	212	6	93	30	70	740	12,318
	C	1	43	**150**	0	4	0	1	247	446
	D	22	157	10	**74**	79	3	8	909	1,262
	E	3	76	5	2	**3,245**	0	0	341	3,672
	F	0	61	4	0	38	**16**	14	459	592
	G	1	75	0	1	21	0	**187**	477	762
	Unk	78	560	39	4	1,641	3	46	**415**	2,786
	Total	237	12,123	420	88	5,123	52	327	3,635	22,005

Table 5.11: Site DC: Confusion matrix comparing manual and automated classification results by five minute bin. Numbers on the diagonal (**bold**) indicate the total number of bins that were given the same classification by both methods. Values off of the diagonal indicate bins that were classified differently by the two methods.

		Manual					
		B	**K**	**E**	**F**	**Unk**	**Total**
Automatic	**B**	**2,956**	295	6	367	1	3,625
	K	32	**2,410**	0	7	1	2,450
	E	35	17	**94**	6	1	153
	F	77	9	2	**260**	1	349
	Unk	1	1	1	1	**1**	5
	Total	3,101	2,732	103	641	5	13,164

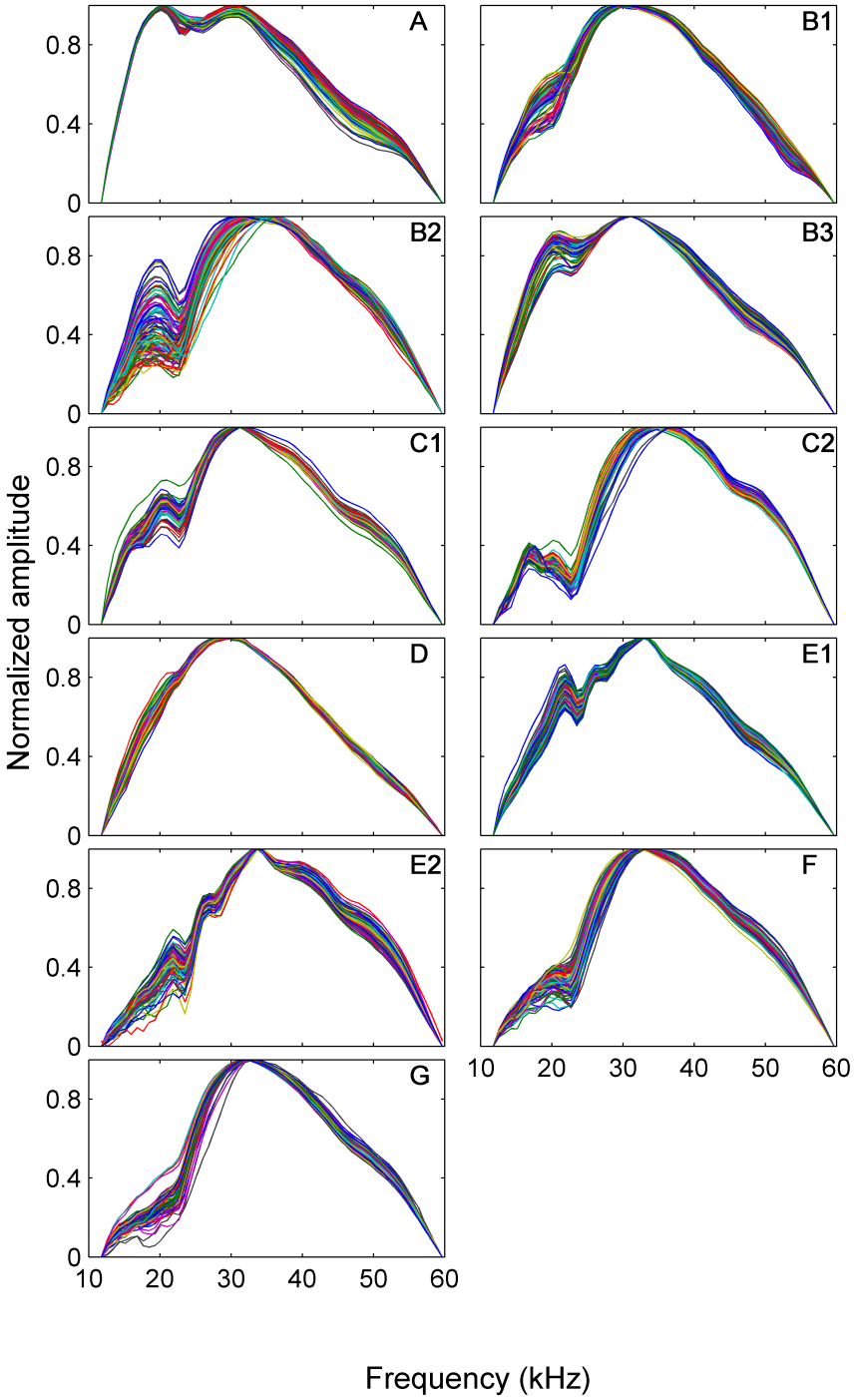


Figure 5.6: MC: Normalized mean click spectra extracted from manually labeled data using spectral clustering methods.

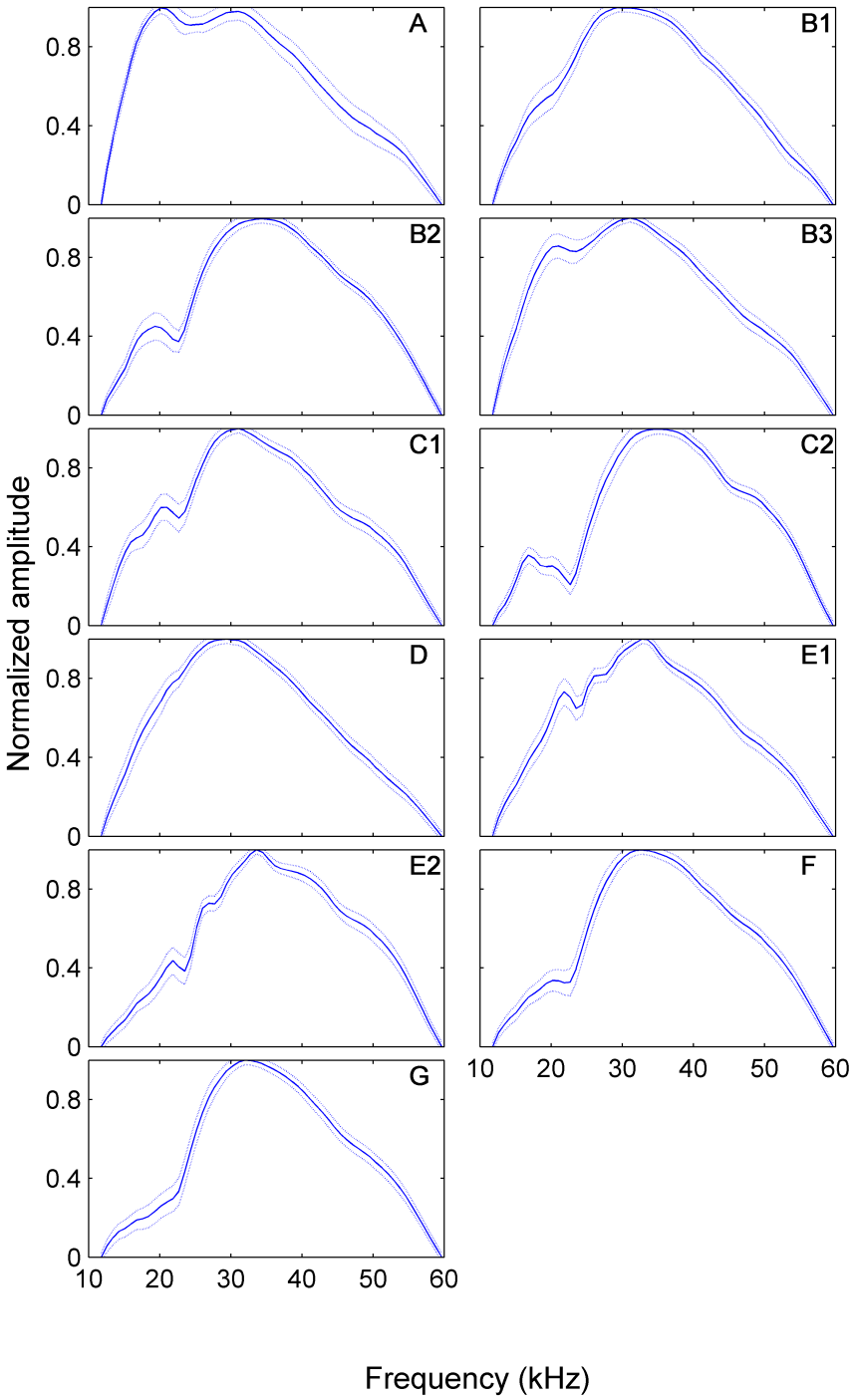


Figure 5.7: Site MC mean and standard deviation normalized amplitude of clustered mean click spectra by type.

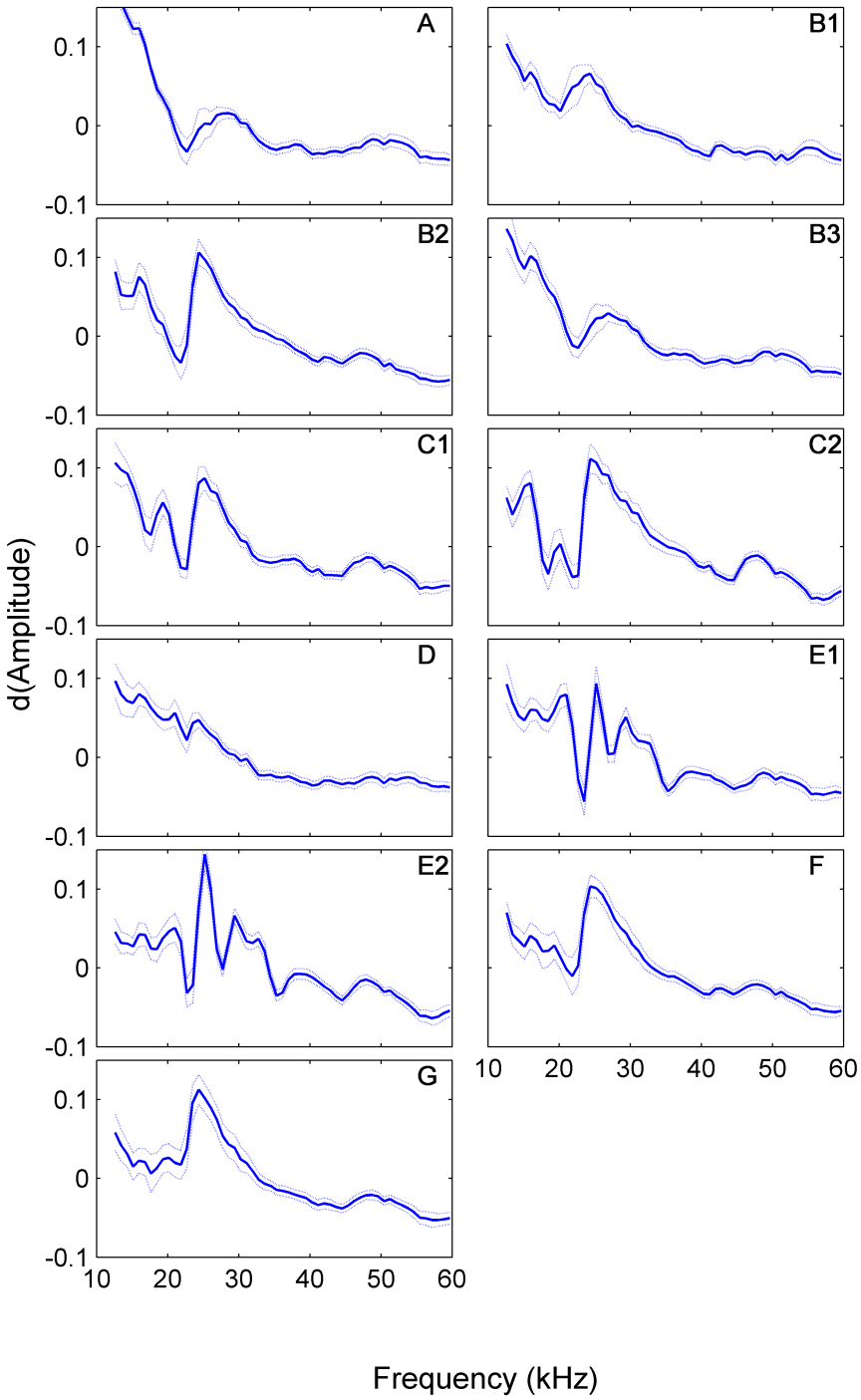


Figure 5.8: Site MC click templates based on mean and standard deviation of the first derivative of clustered mean spectra.

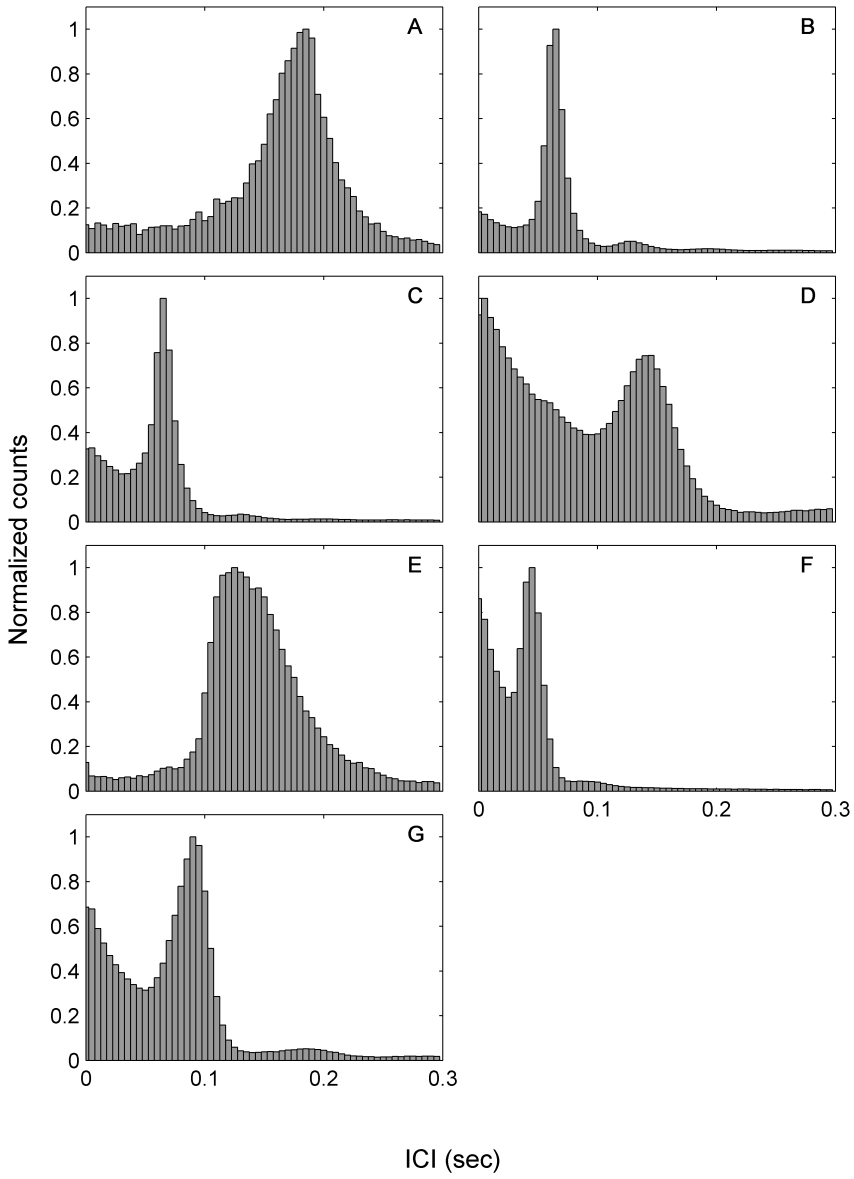


Figure 5.9: Site MC ICI distributions associated with each click type identified at this site.

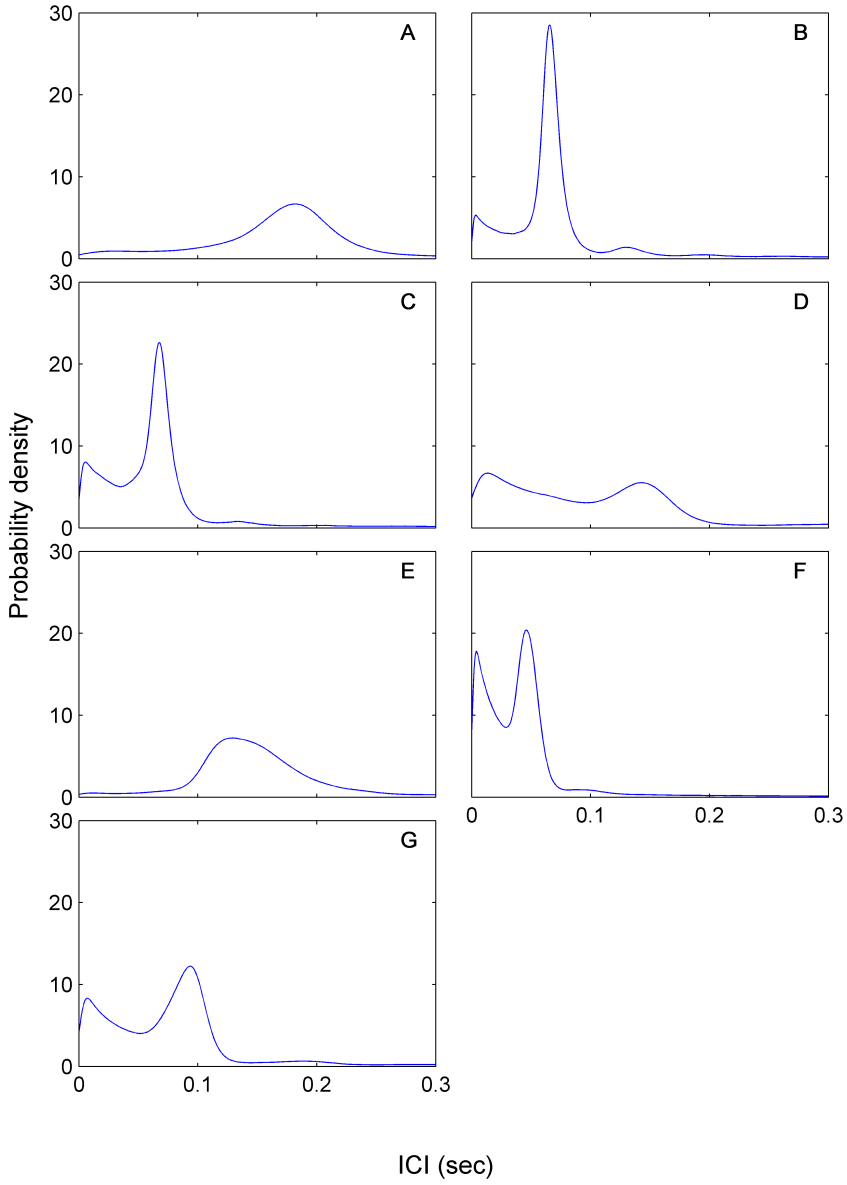


Figure 5.10: Site MC: ICI probability density distributions associated with each click type identified at this site.

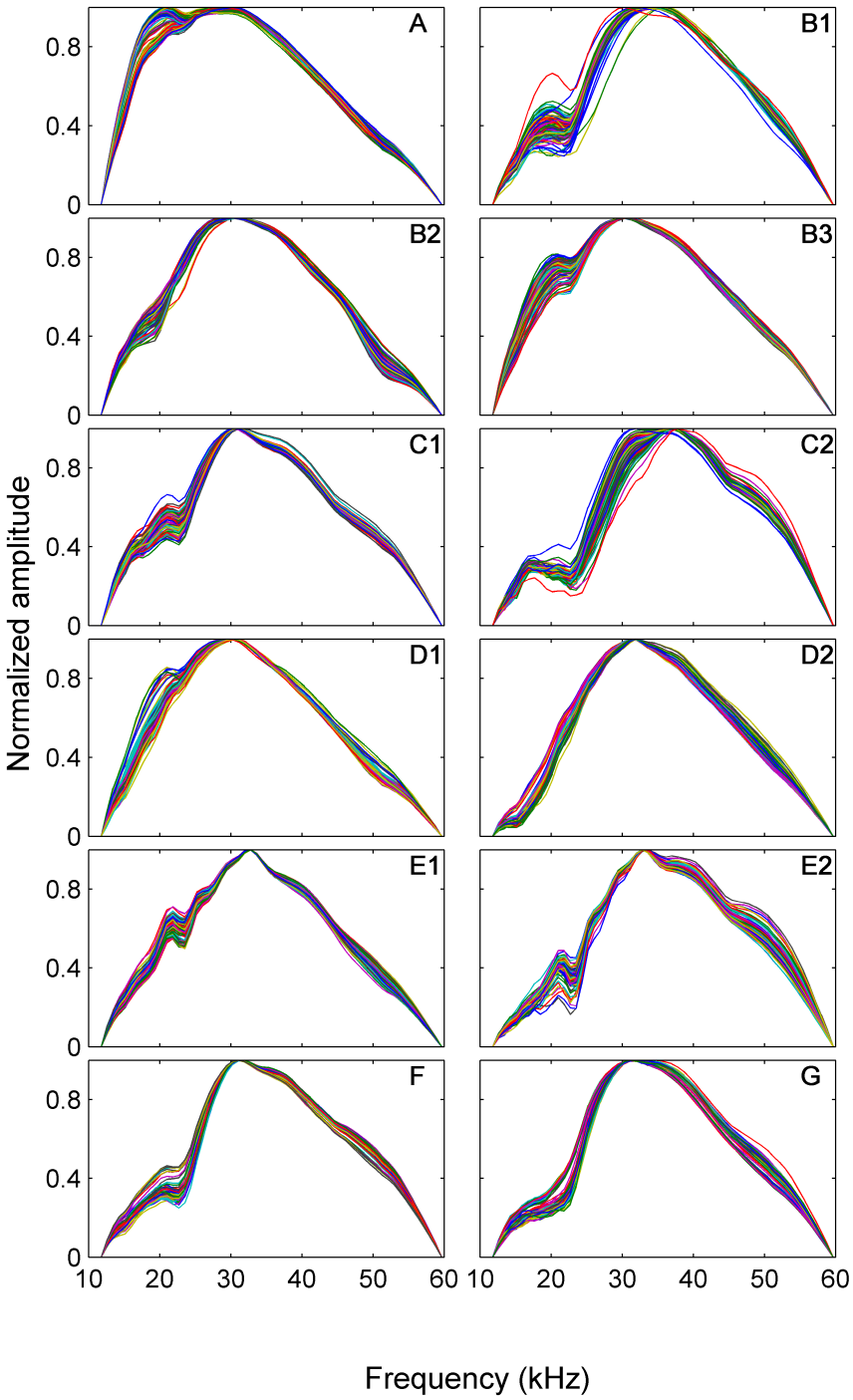


Figure 5.11: Site GC: Normalized mean click spectra extracted from manually labeled data using spectral clustering methods.

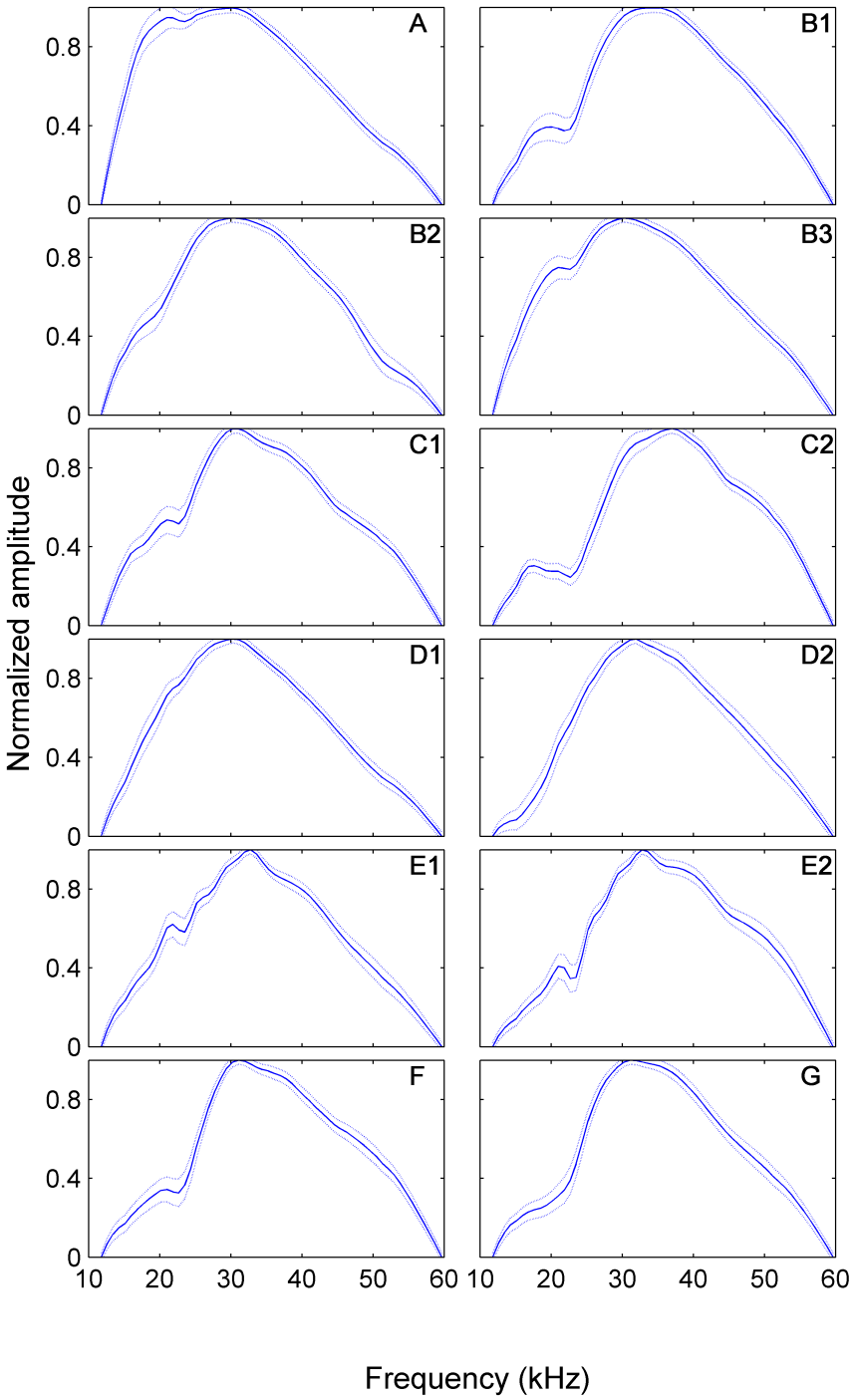


Figure 5.12: Site GC: Mean and standard deviation normalized amplitude of clustered mean click spectra by type.

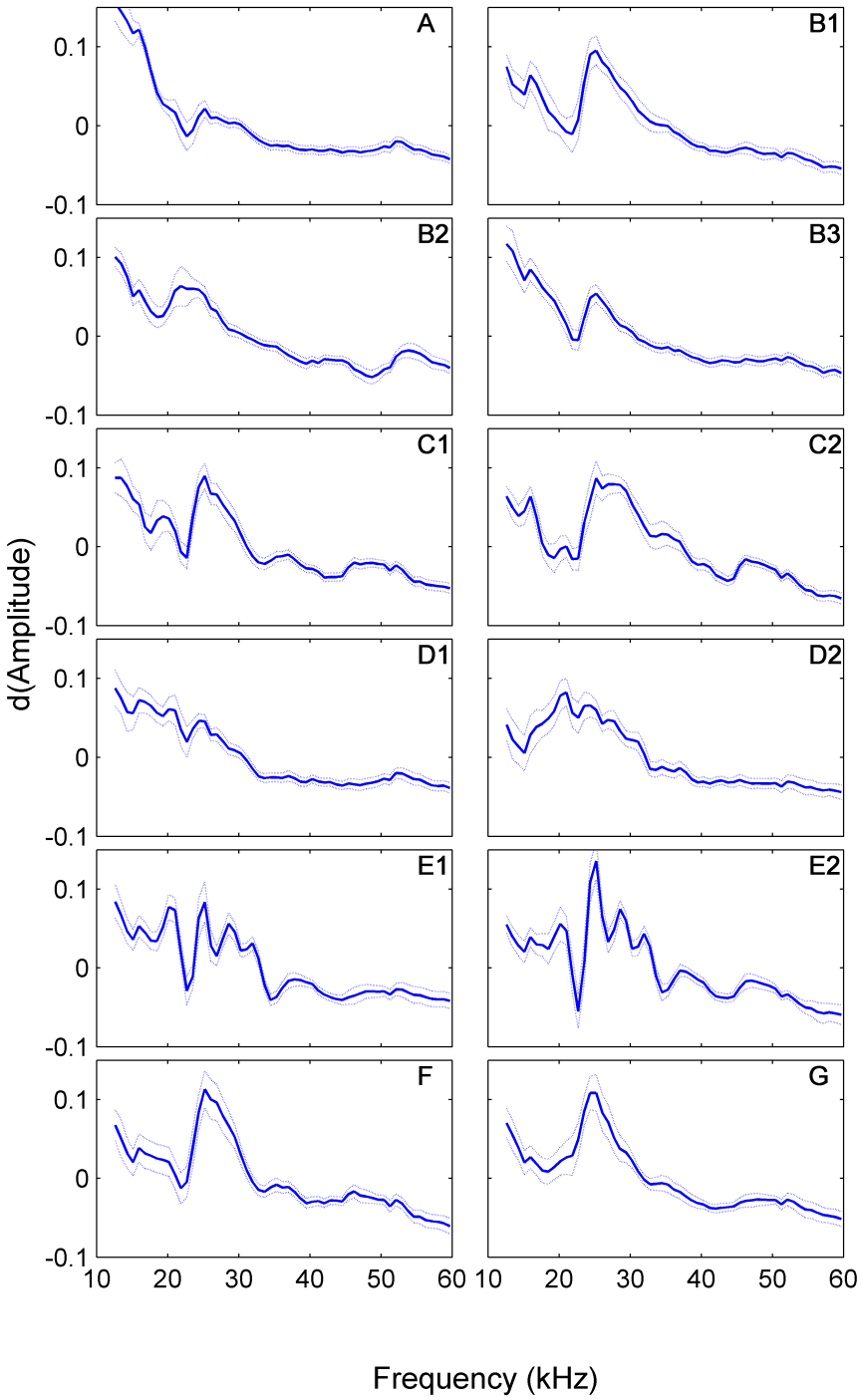


Figure 5.13: Site GC: Click templates based on mean and standard deviation of the first derivative of clustered mean spectra.

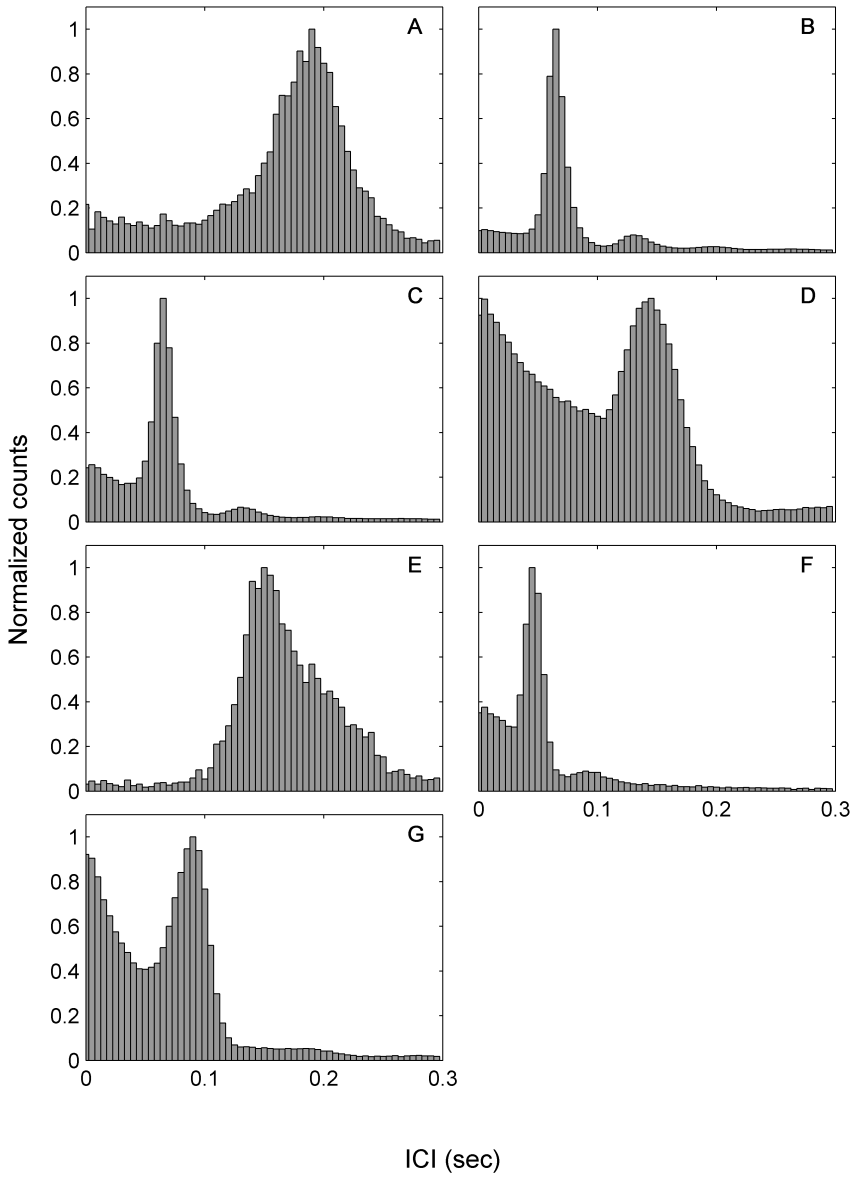


Figure 5.14: Site GC: ICI distributions associated with each click type identified at this site.

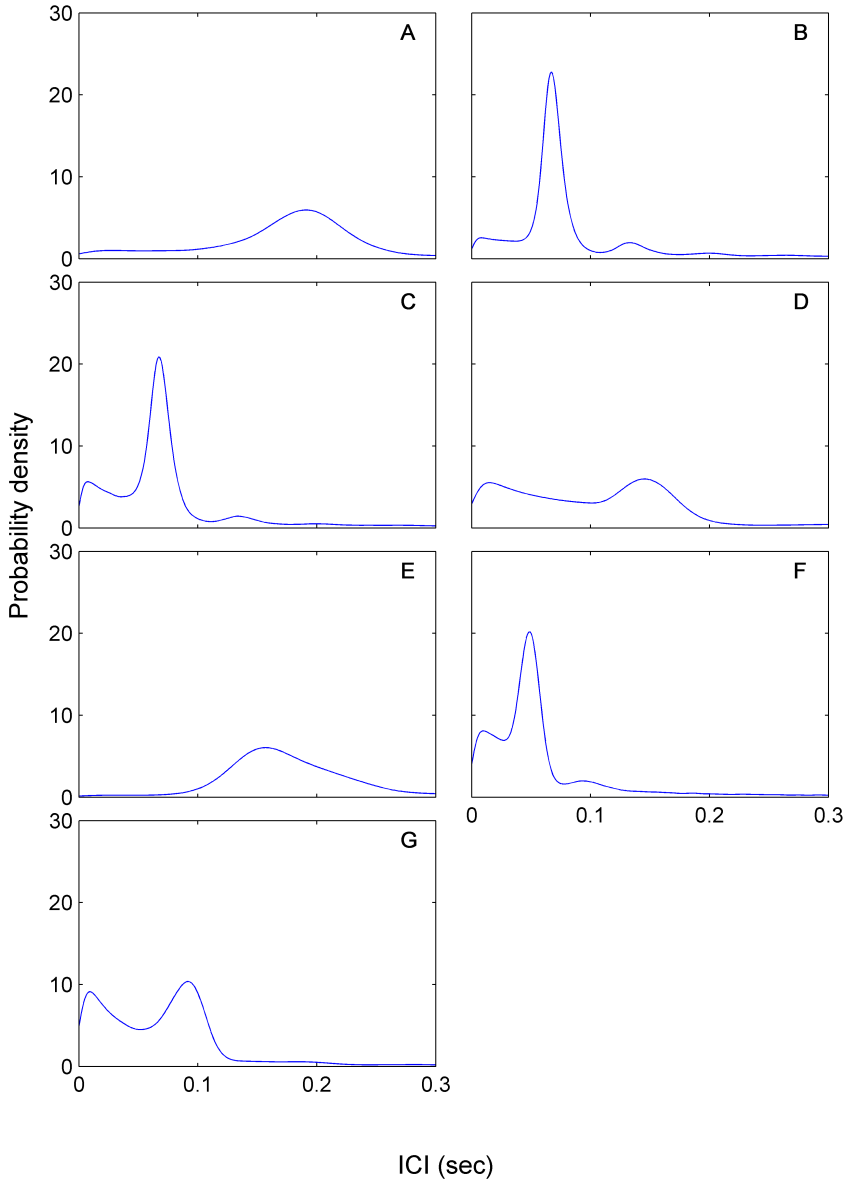


Figure 5.15: GC: ICI probability density distributions associated with each click type identified at this site.

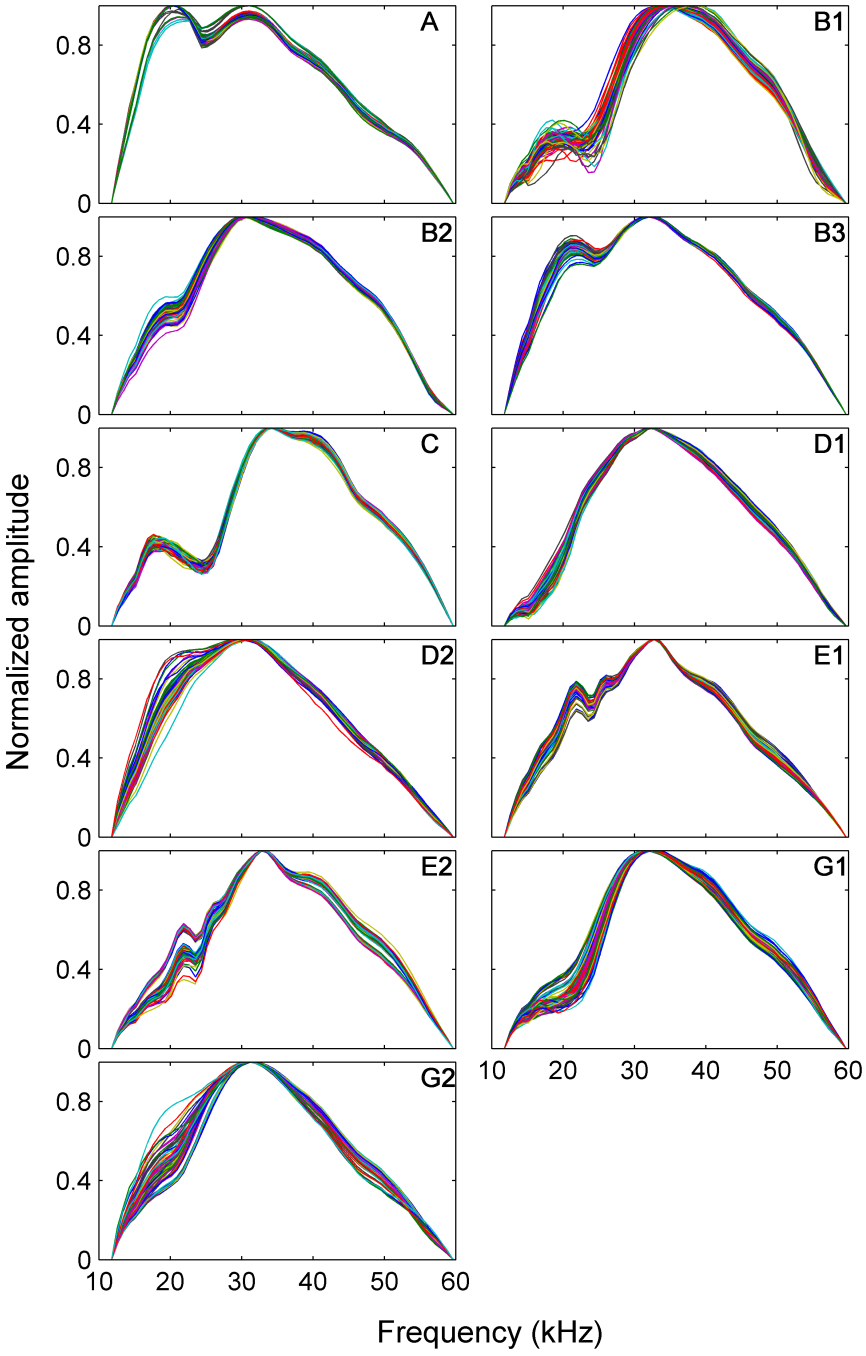


Figure 5.16: Site DT: Normalized mean click spectra extracted from manually labeled data using spectral clustering methods.

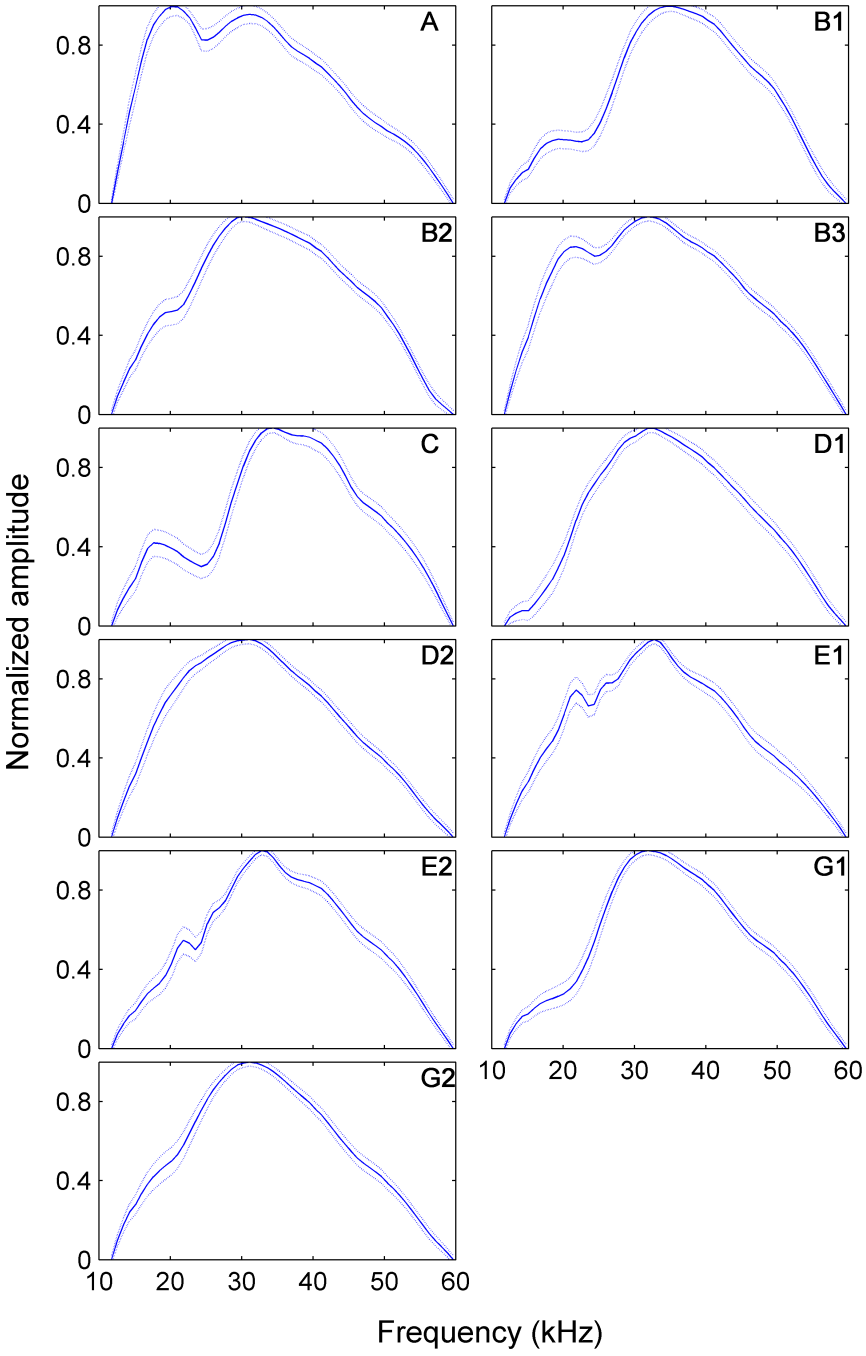


Figure 5.17: Site DT: Mean and standard deviation of normalized amplitude of clustered mean click spectra by type.

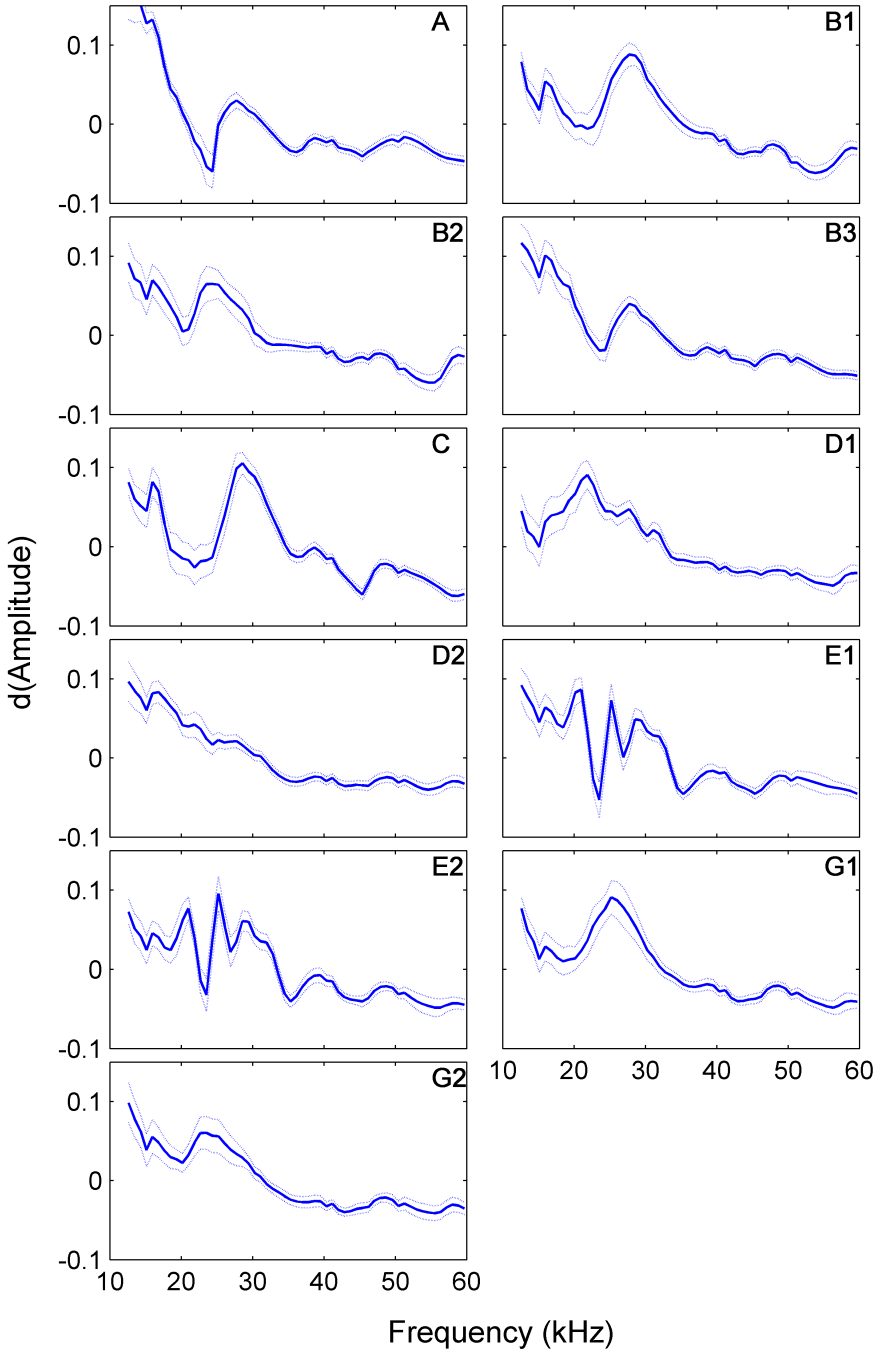


Figure 5.18: Site DT: Click templates based on mean and standard deviation of the first derivative of clustered mean spectra.

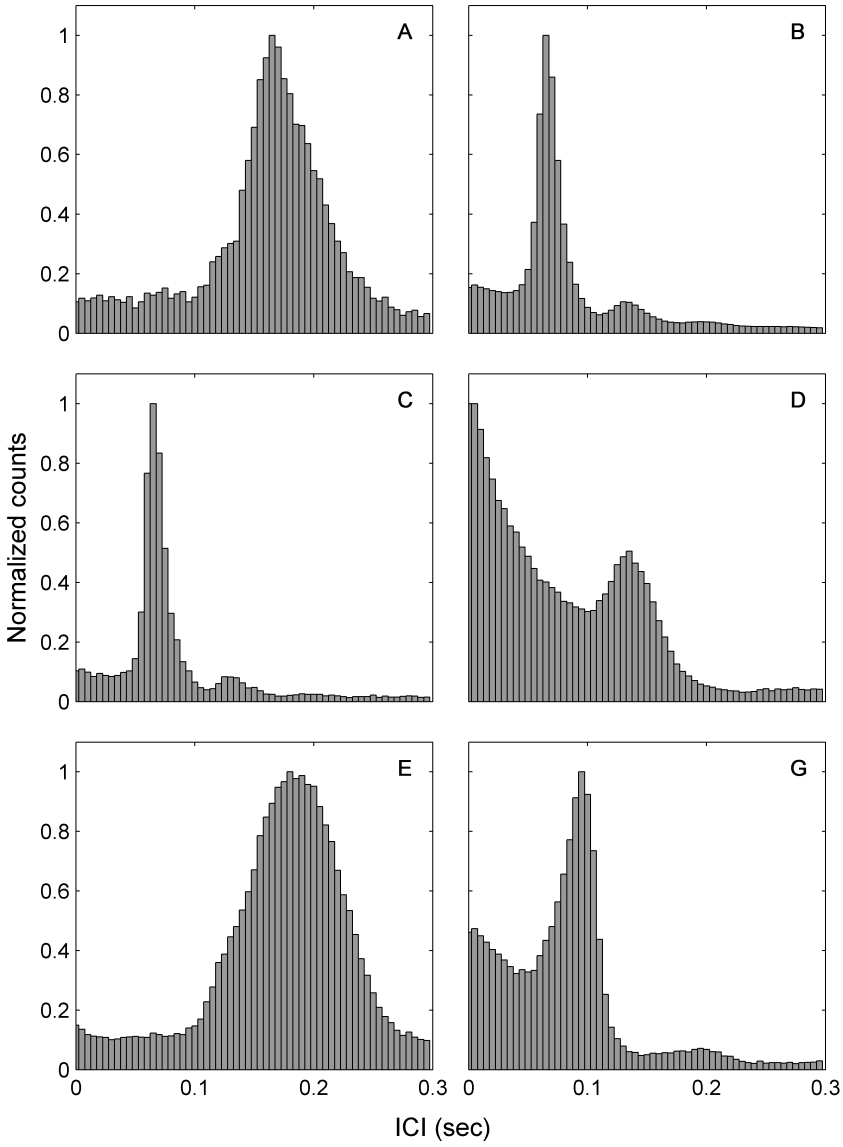


Figure 5.19: Site DT: ICI distributions associated with each click type identified at this site.

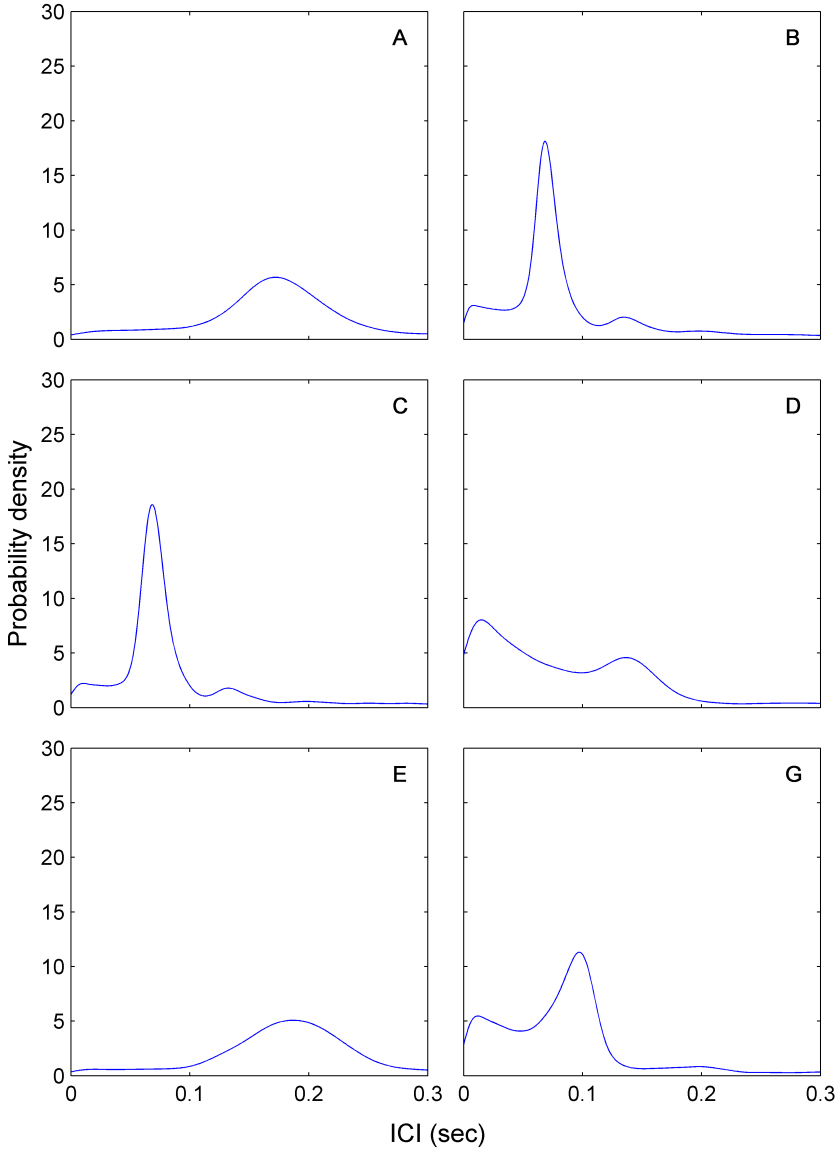


Figure 5.20: Site DT: ICI probability density distributions associated with each click type identified at this site.

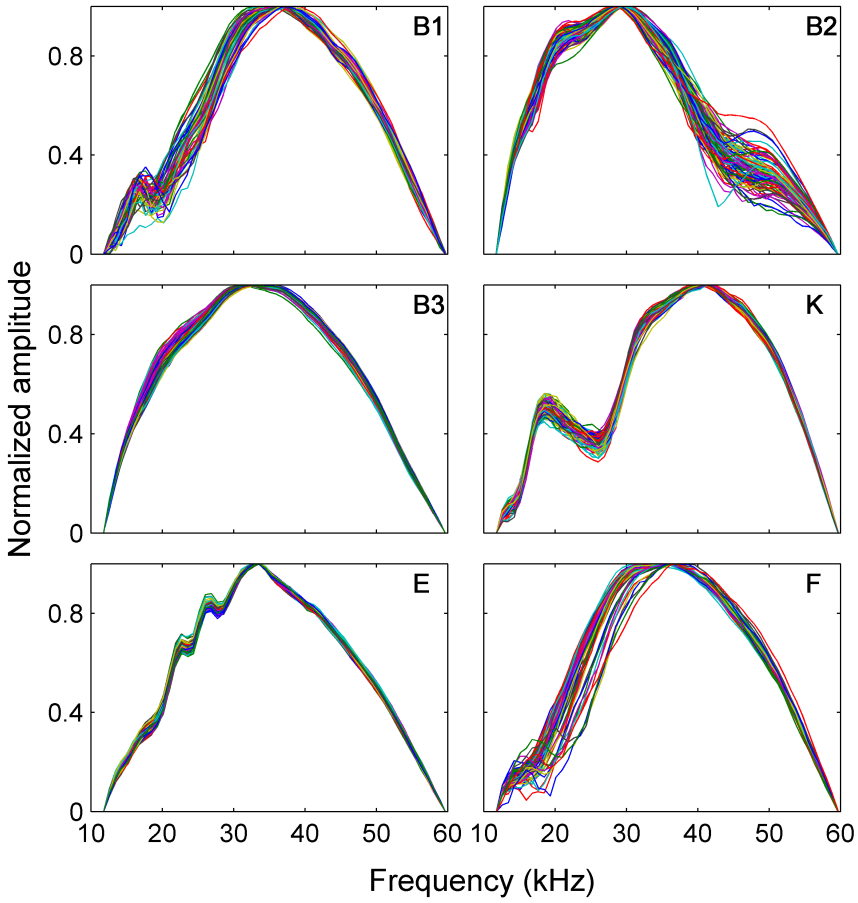


Figure 5.21: Site DC: Normalized mean click spectra extracted from manually labeled data using spectral clustering methods.

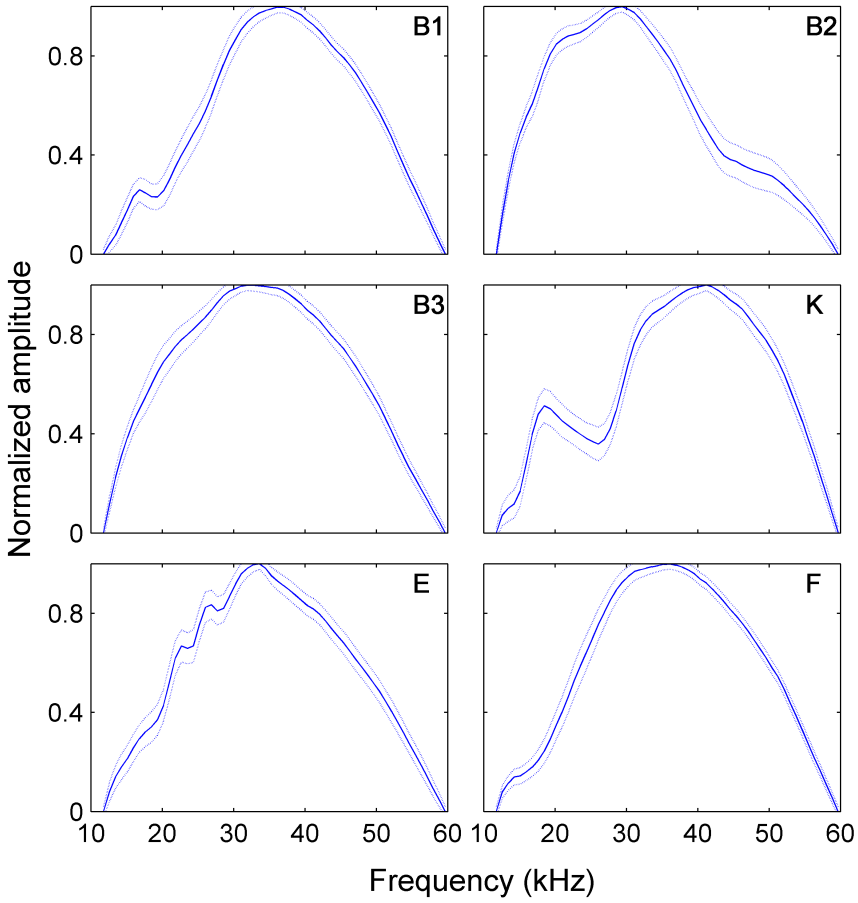


Figure 5.22: Site DC: Mean and standard deviation normalized amplitude of click spectra by type.

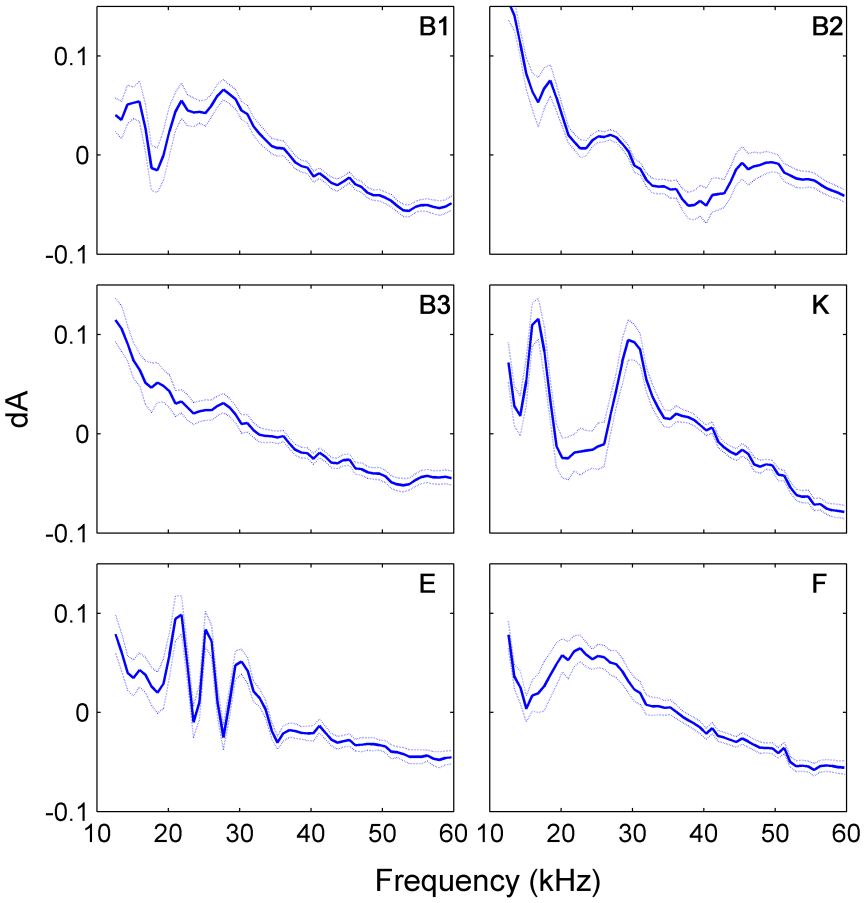


Figure 5.23: Site DC: Click templates computed from mean and standard deviation of the first derivative of clustered mean spectra.

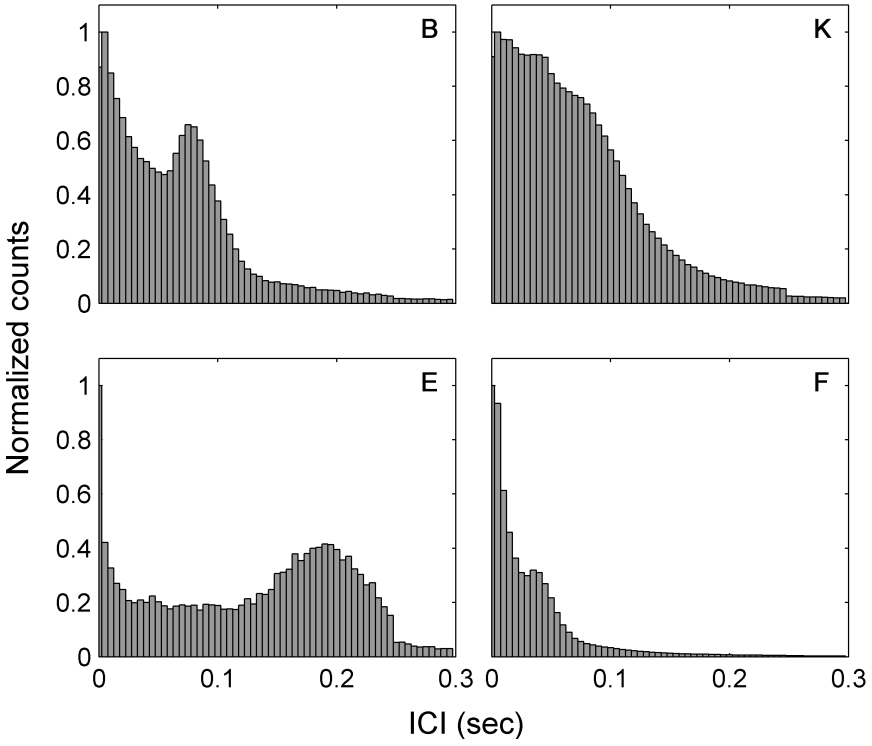


Figure 5.24: Site DC: ICI distributions associated with each click type identified at this site.

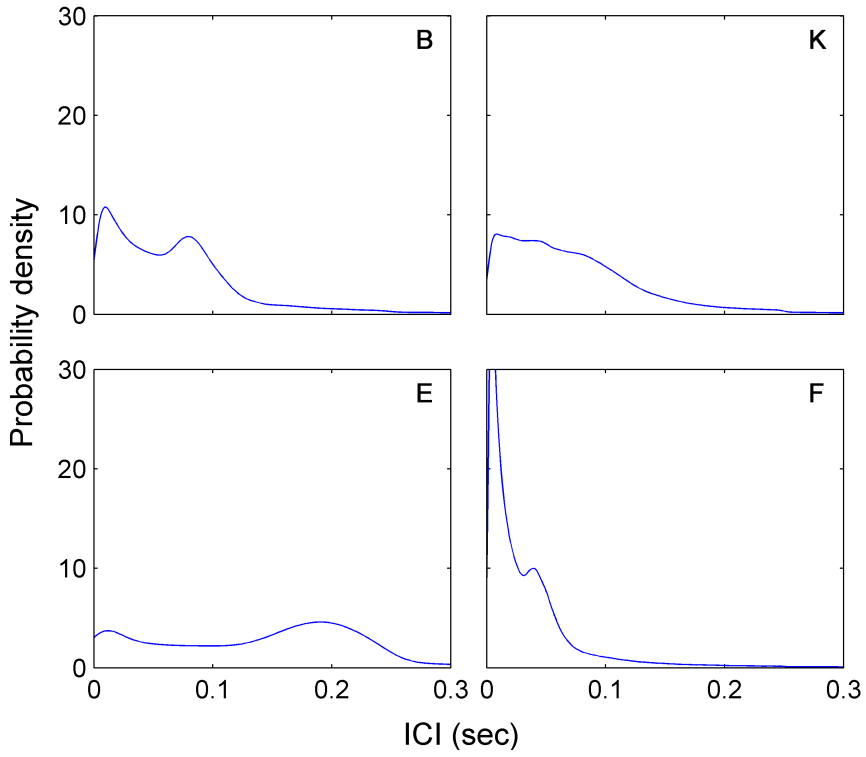


Figure 5.25: Site DC: ICI probability density distributions associated with each click type identified at this site.

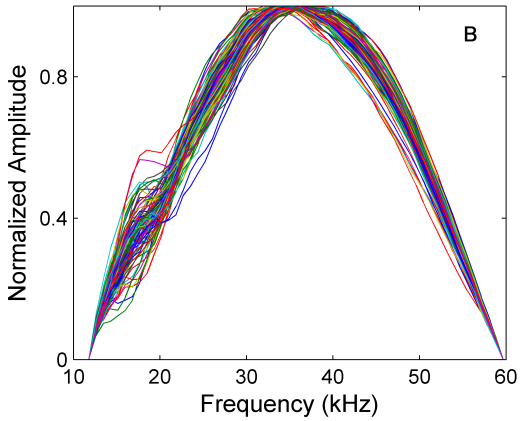


Figure 5.26: Site MP click type B normalized mean spectra extracted from manually labeled data.

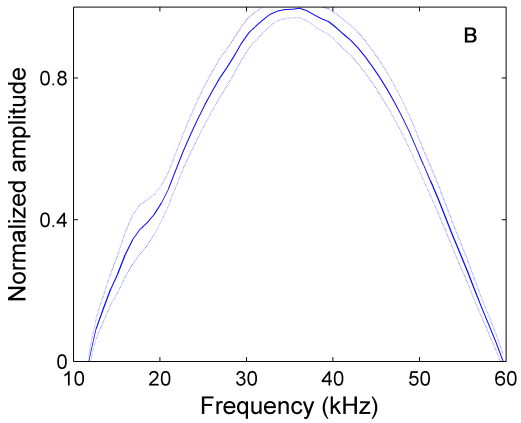


Figure 5.27: Site MP: Mean and standard deviation of normalized amplitude of mean click spectra for type B.

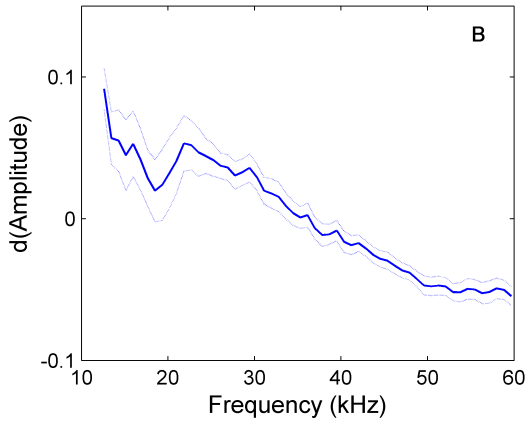


Figure 5.28: Site MP: Click template based on mean and standard deviation of the first derivative of the clustered mean spectra.

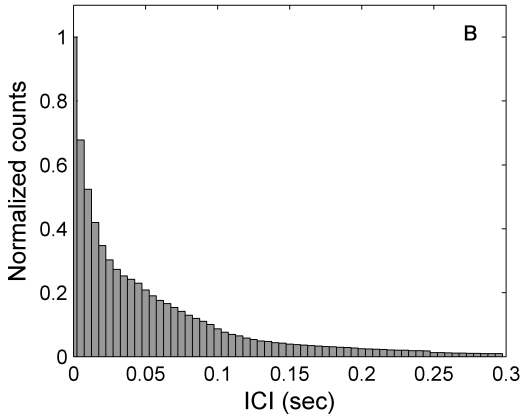


Figure 5.29: Site MP: Inter-click interval distribution associated with click type B.

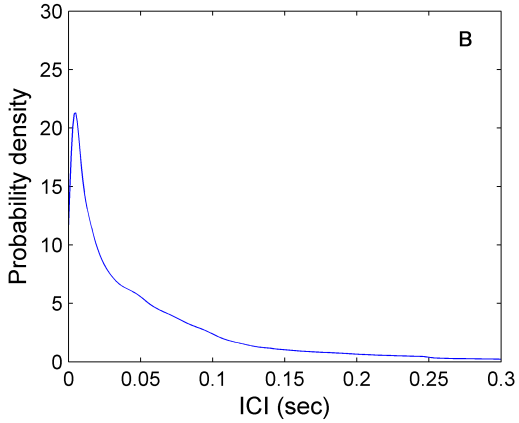


Figure 5.30: Site MP: ICI probability density distribution associated with click type B.

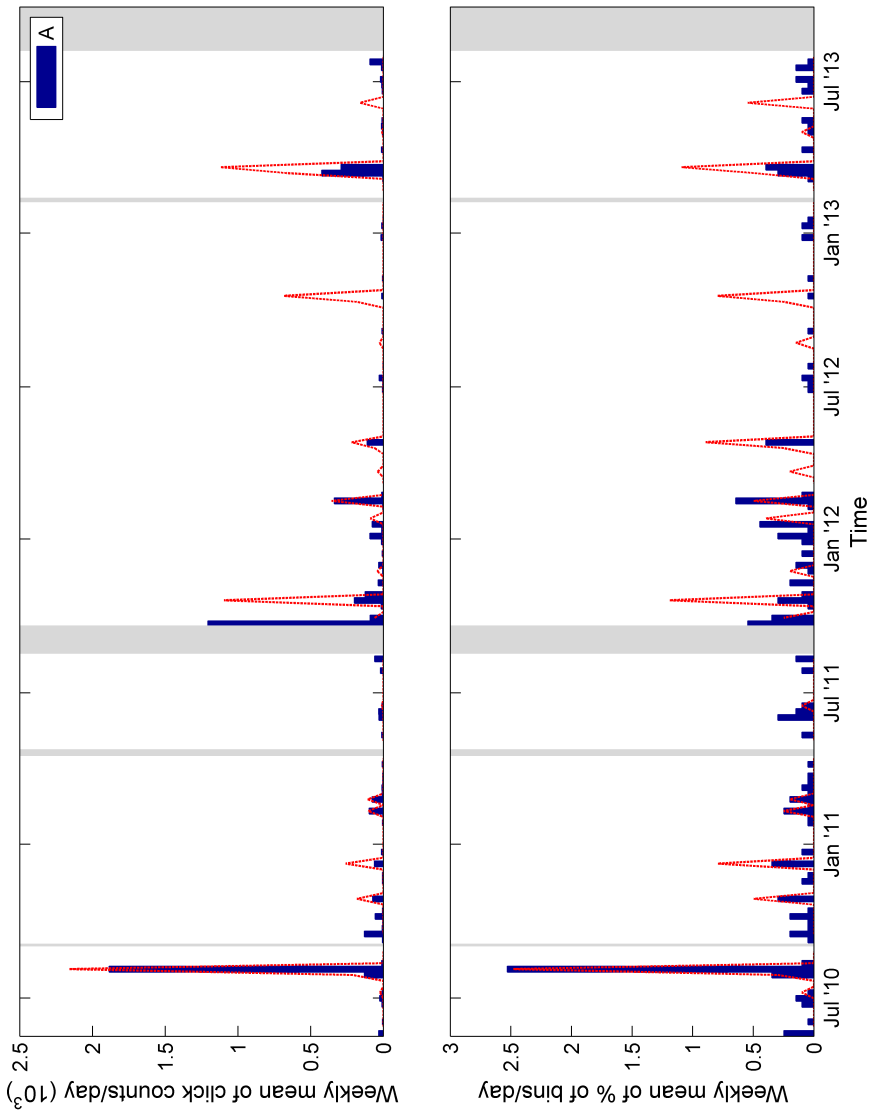


Figure 5.31: Occurrence of click type A at site site MC, using automated (histogram) and manual (dashed line). Top panel shows classification as a weekly mean of thousands click counts per day. Bottom panel shows the same data as a weekly mean percentage of total daily five-minute bins.

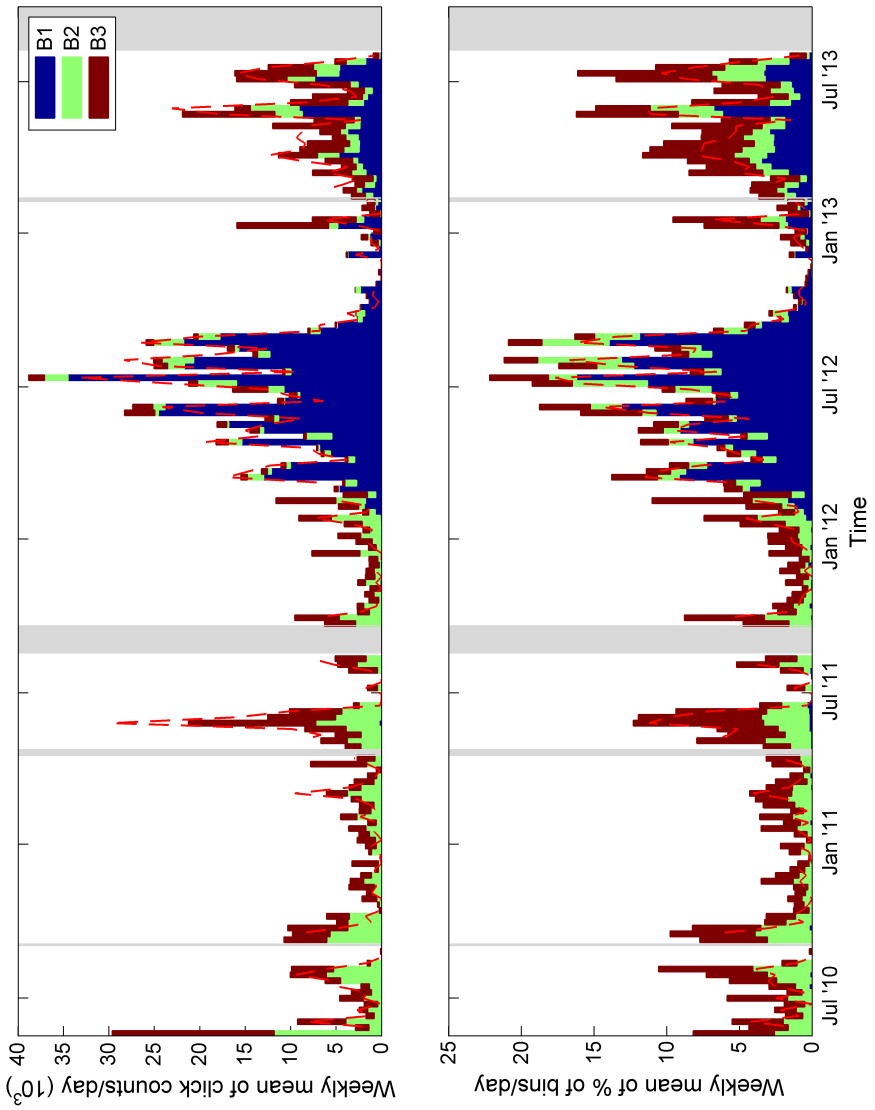


Figure 5.32: Occurrence of click B subtypes at site MC, using automated (histogram) classification methods. Manual classification (dashed red line) did not distinguish between subtypes, but indicates type B totals. Top panel shows classification as a weekly mean of thousands click counts per day. Bottom panel shows the same data as a weekly mean percentage of total daily five-minute bins.

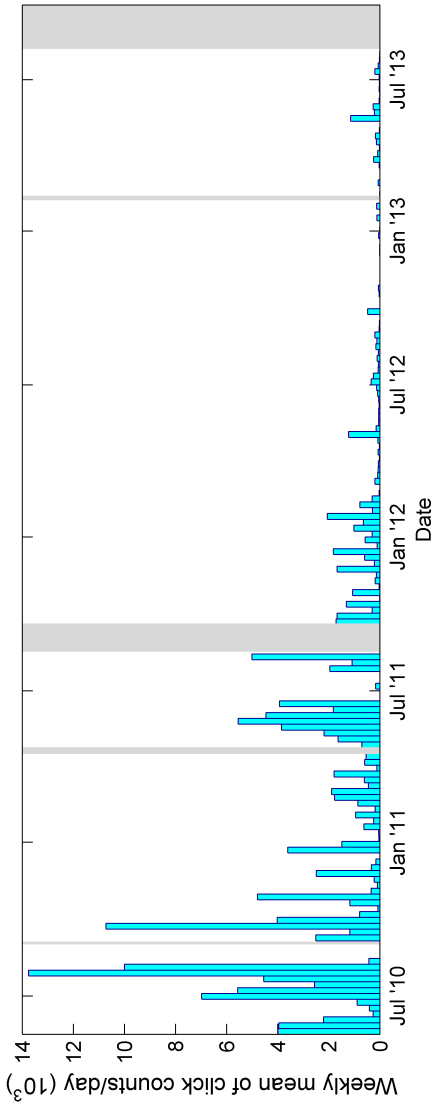


Figure 5.33: Occurrence of click C subtypes at site MC, using automated (histogram) classification methods. Manual classification (dashed red line) did not distinguish between subtypes, but indicates type C totals. Top panel shows classification as a weekly mean of thousands click counts per day. Bottom panel shows the same data as a weekly mean percentage of total daily five-minute bins.

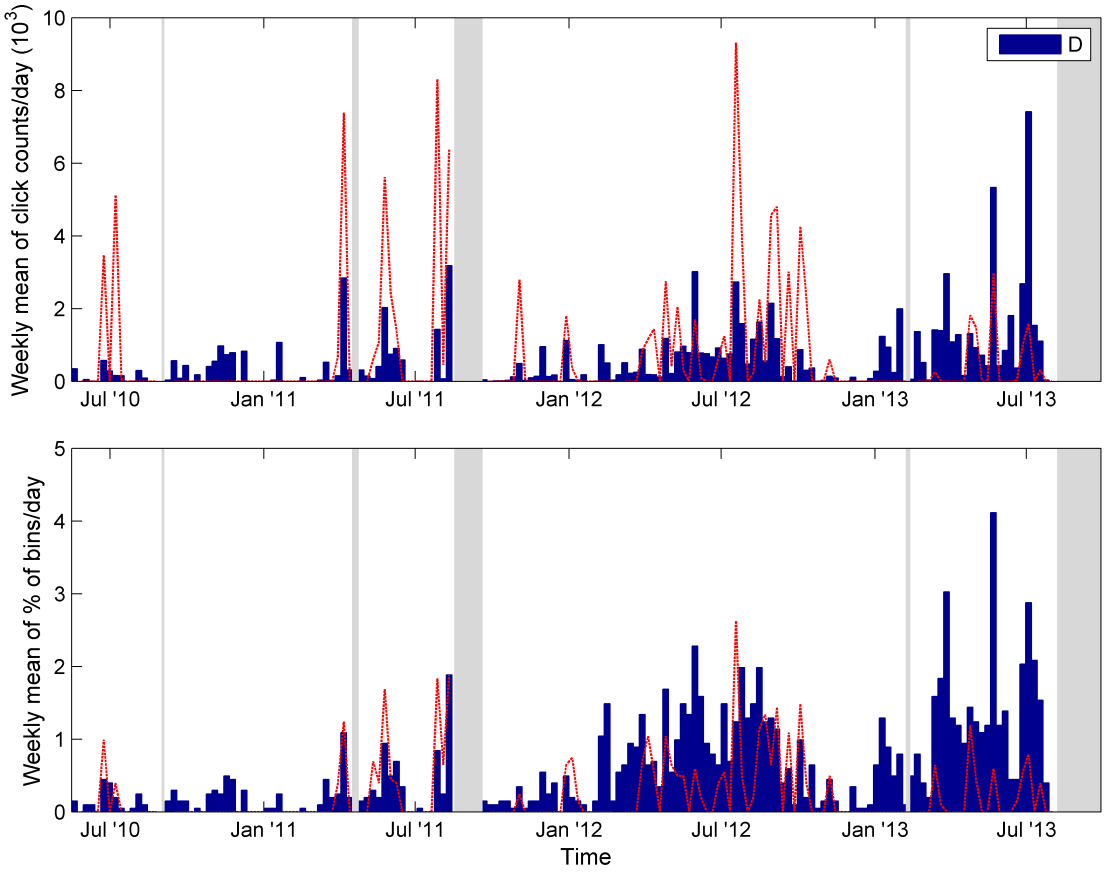


Figure 5.34: Occurrence of click type D at site MC, using automated (histogram) and manual (dashed red line). Top panel shows classification as a weekly mean of thousands click counts per day. Bottom panel shows the same data as a weekly mean percentage of total daily five-minute bins.

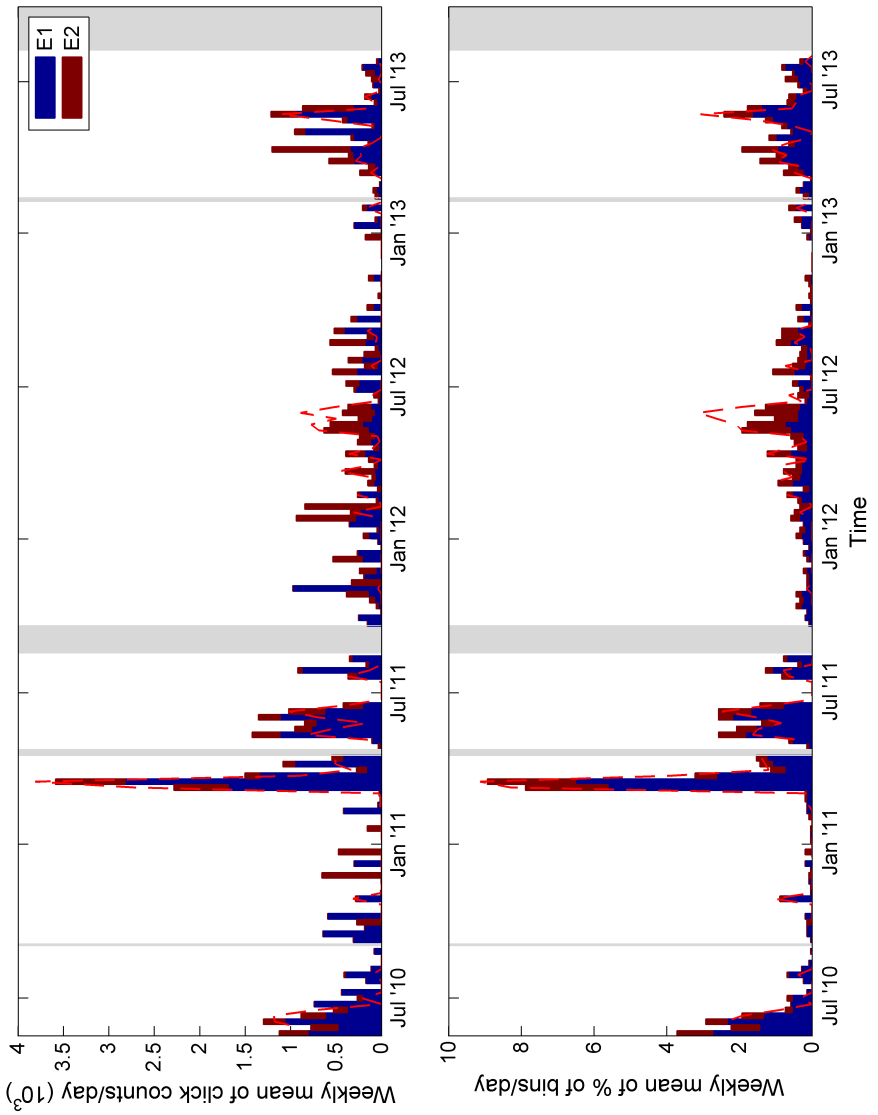


Figure 5.35: Occurrence of click E subtypes at site MC, using automated (histogram) classification methods. Manual classification (dashed red line) did not distinguish between subtypes, but indicates type E totals. Top panel shows classification as a weekly mean of thousands click counts per day. Bottom panel shows the same data as a weekly mean percentage of total daily five-minute bins.

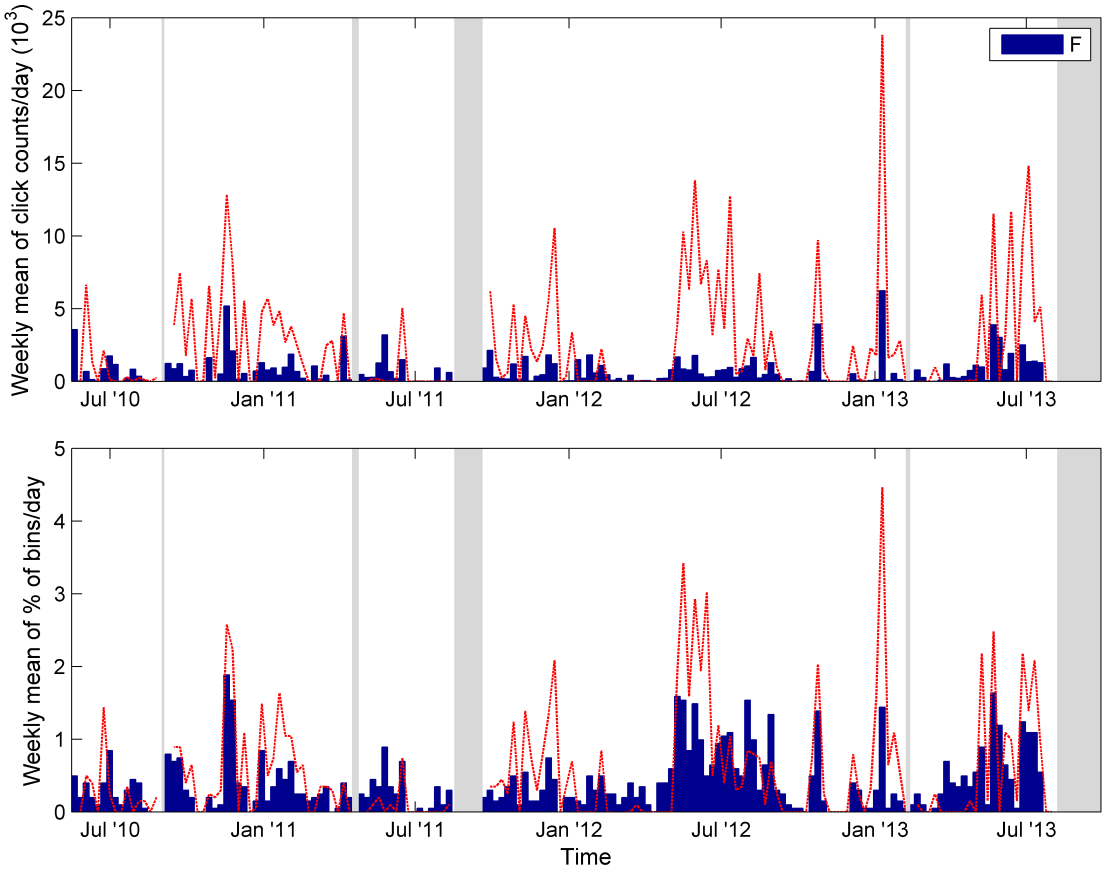


Figure 5.36: Occurrence of click type F at site MC, using automated (histogram) and manual (dashed red line). Top panel shows classification as a weekly mean of thousands click counts per day. Bottom panel shows the same data as a weekly mean percentage of total daily five-minute bins.

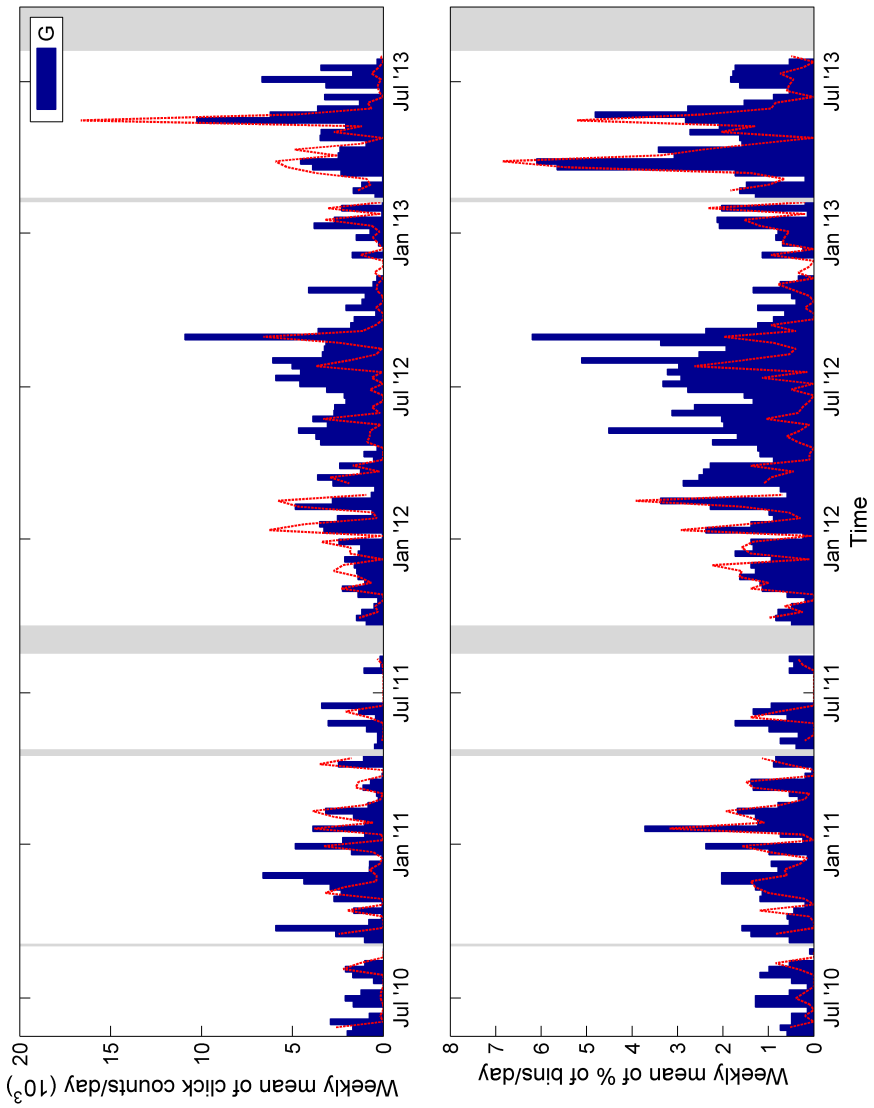


Figure 5.37: Occurrence of click type G at site MC, using automated (histogram) and manual (dashed line). Top panel shows classification as a weekly mean of thousands click counts per day. Bottom panel shows the same data as a weekly mean percentage of total daily five-minute bins.

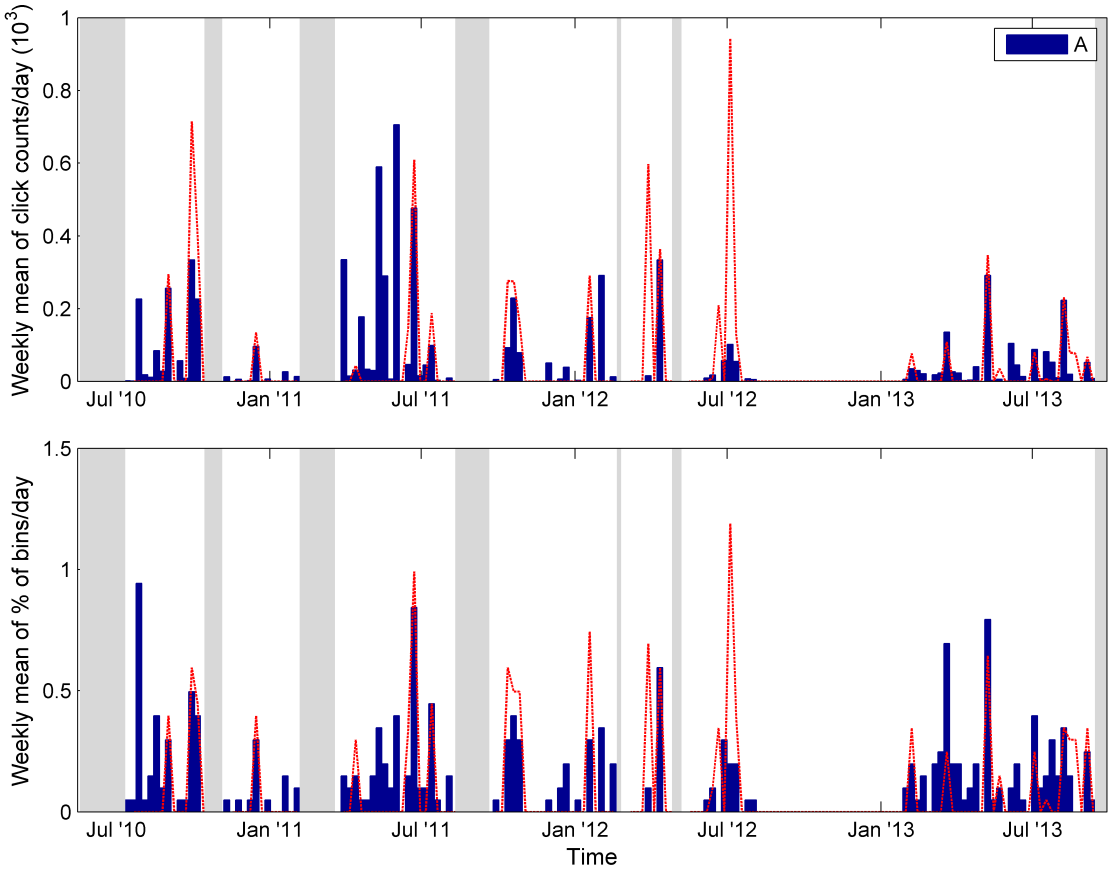


Figure 5.38: Occurrence of click type A at site GC, using automated (histogram) and manual (dashed line). Top panel shows classification as a weekly mean of thousands click counts per day. Bottom panel shows the same data as a weekly mean percentage of total daily five-minute bins.

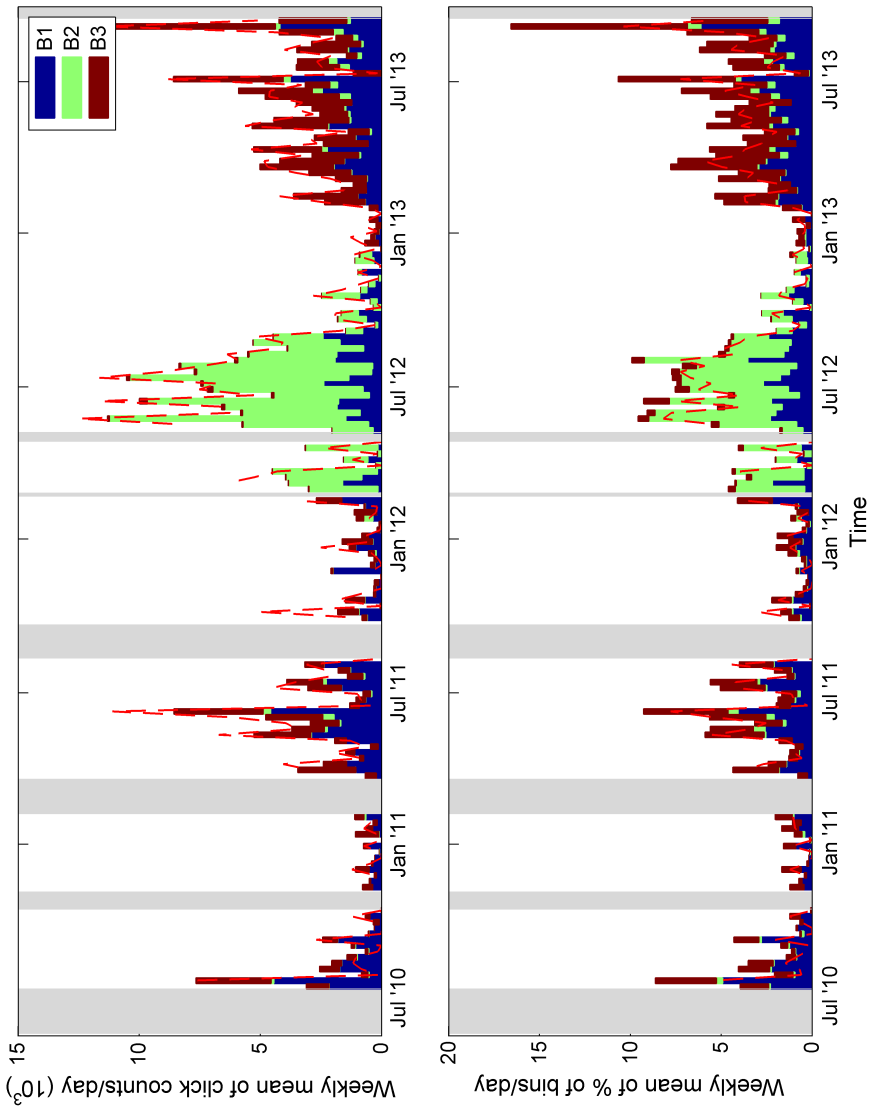


Figure 5.39: Occurrence of click B subtypes at site GC, using automated (histogram) classification methods. Manual classification (dashed red line) did not distinguish between subtypes, but indicates type B totals. Top panel shows classification as a weekly mean of thousands click counts per day. Bottom panel shows the same data as a weekly mean percentage of total daily five-minute bins.

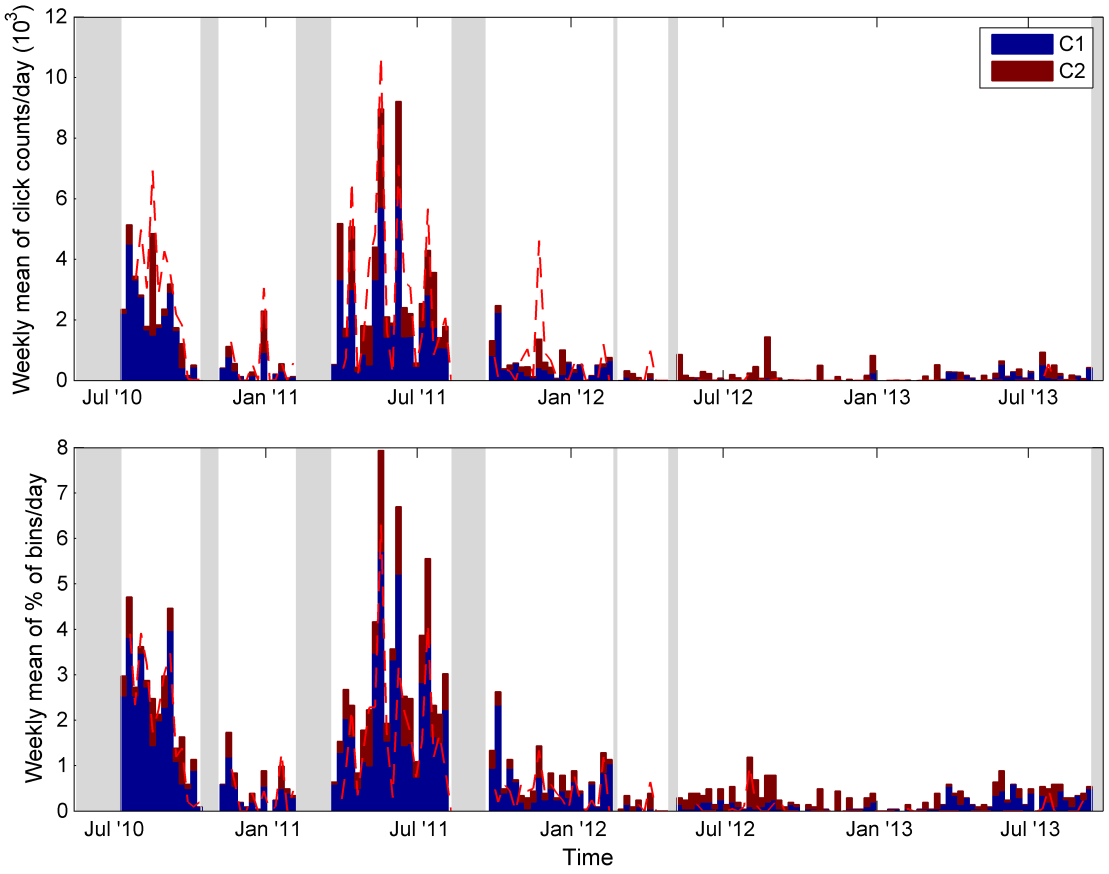


Figure 5.40: Occurrence of click C subtypes at site GC, using automated (histogram) classification methods. Manual classification (dashed red line) did not distinguish between subtypes, but indicates type C totals. Top panel shows classification as a weekly mean of thousands click counts per day. Bottom panel shows the same data as a weekly mean percentage of total daily five-minute bins.

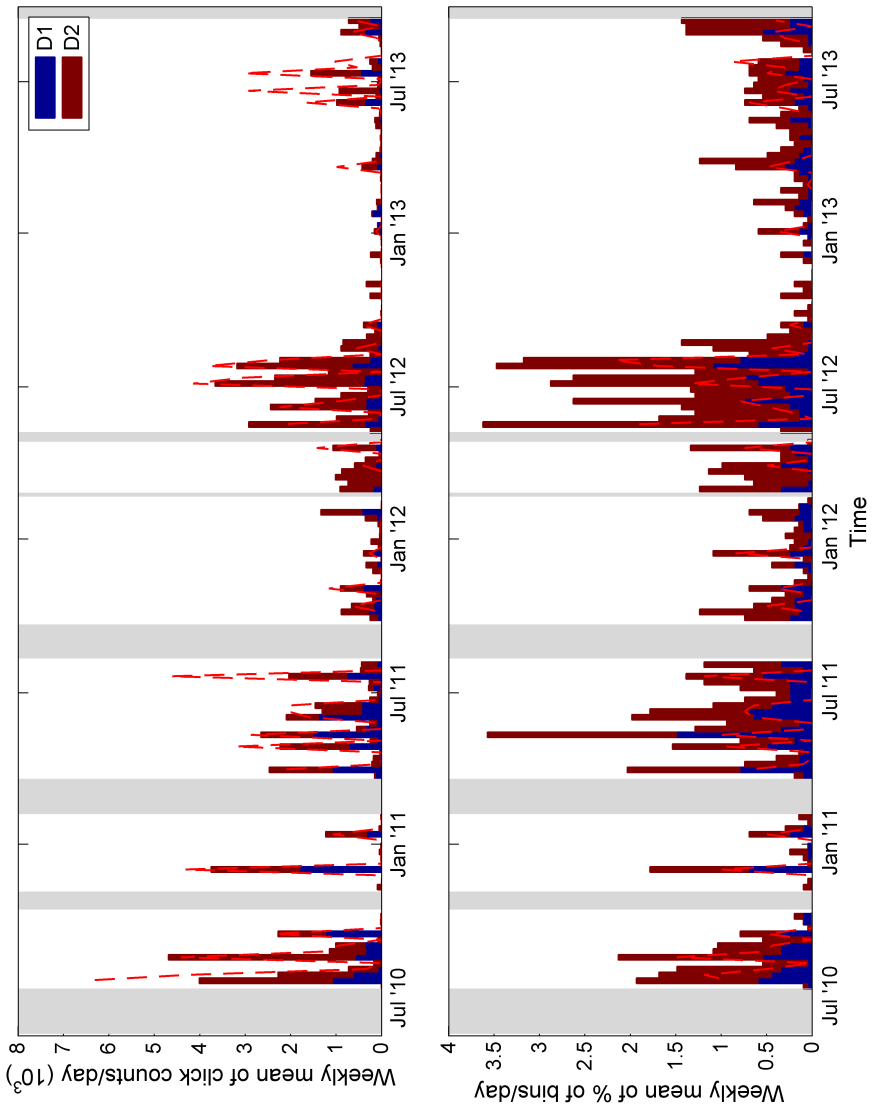


Figure 5.41: Occurrence of click D subtypes at site GC, using automated (histogram) classification methods. Manual classification (dashed line) did not distinguish between subtypes, but indicates type D totals. Top panel shows classification as a weekly mean of thousands click counts per day. Bottom panel shows the same data as a weekly mean percentage of total daily five-minute bins.

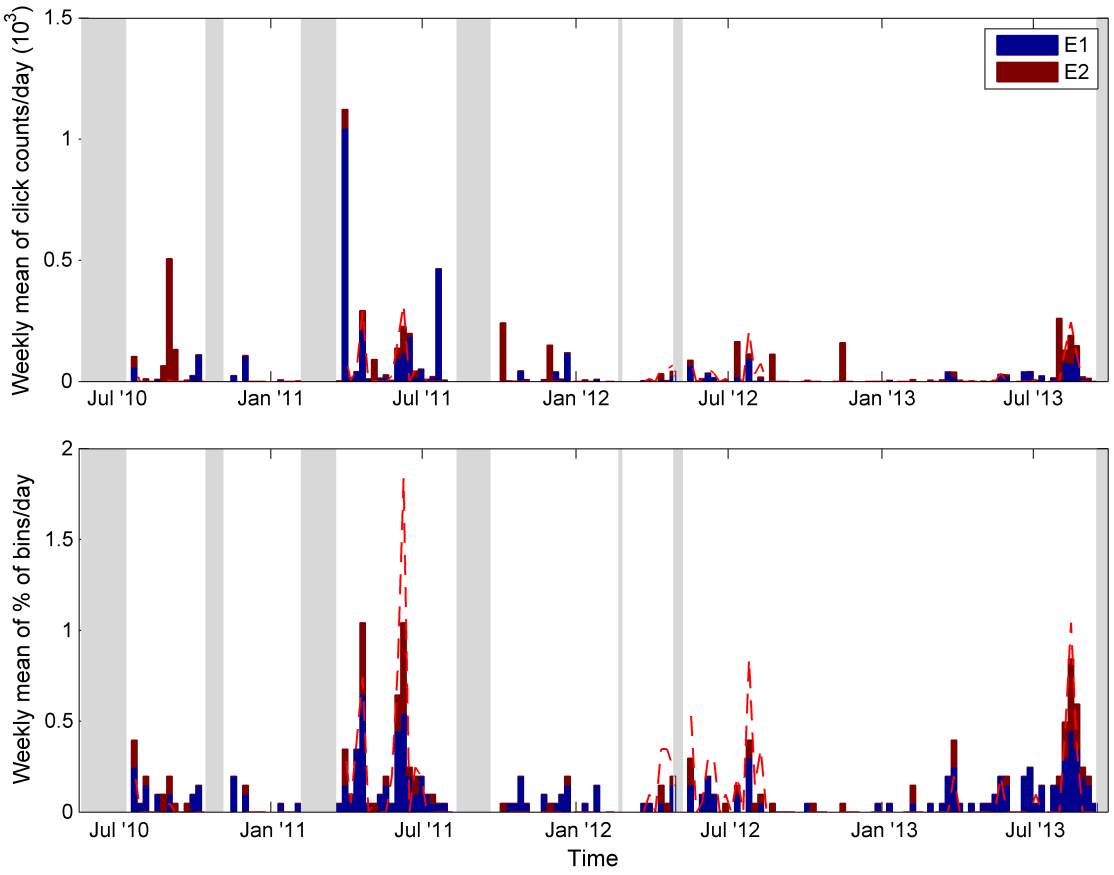


Figure 5.42: Occurrence of click E subtypes at site GC, using automated (histogram) classification methods. Manual classification (dashed line) did not distinguish between subtypes, but indicates type E totals. Top panel shows classification as a weekly mean of thousands click counts per day. Bottom panel shows the same data as a weekly mean percentage of total daily five-minute bins.

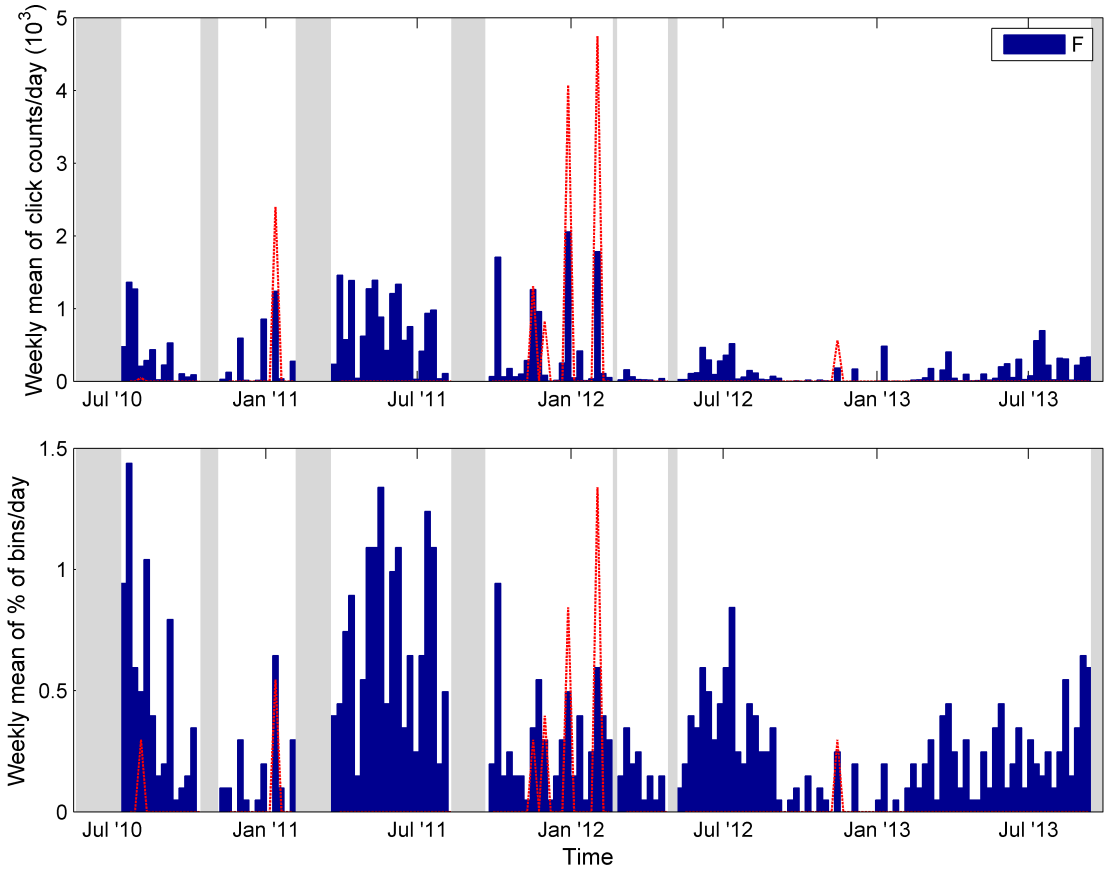


Figure 5.43: Occurrence of click type F at site GC, using automated (histogram) and manual (dashed line). Top panel shows classification as a weekly mean of thousands click counts per day. Bottom panel shows the same data as a weekly mean percentage of total daily five-minute bins.

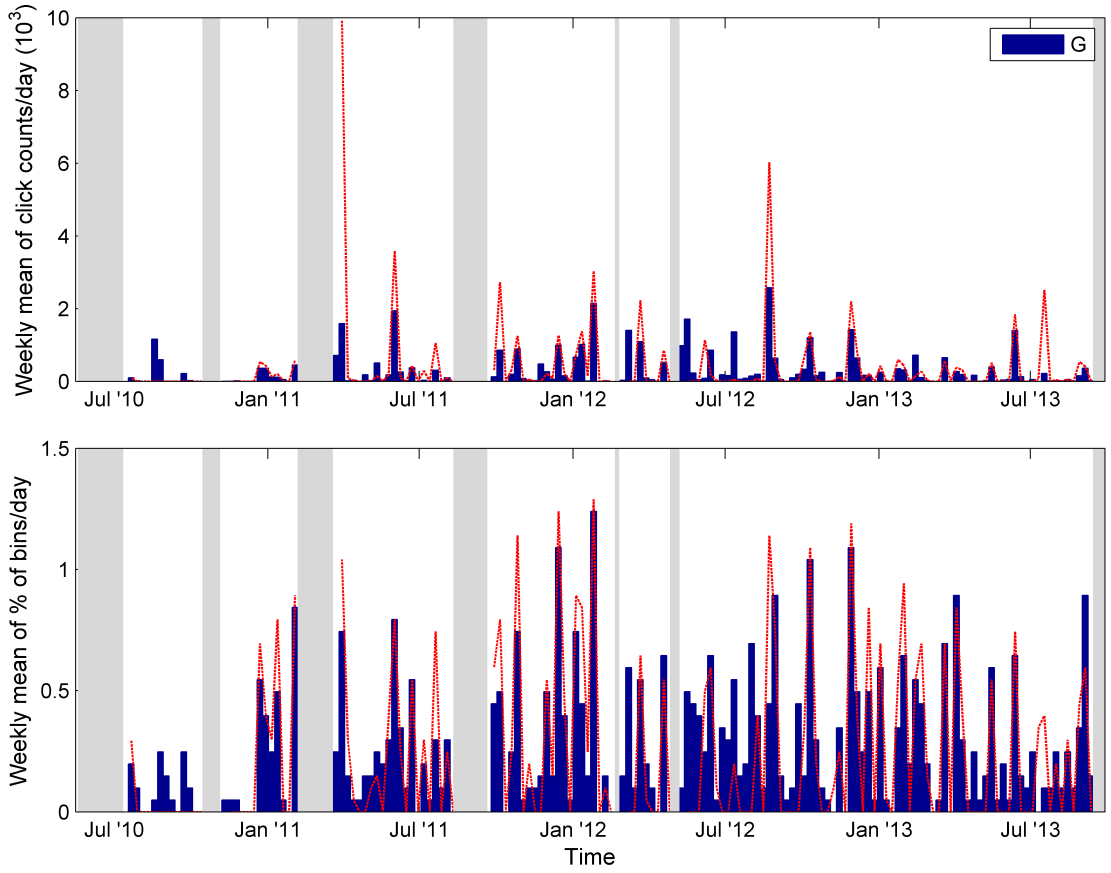


Figure 5.44: Occurrence of click type G at site GC, using automated (histogram) and manual (dashed line). Top panel shows classification as a weekly mean of thousands click counts per day. Bottom panel shows the same data as a weekly mean percentage of total daily five-minute bins.

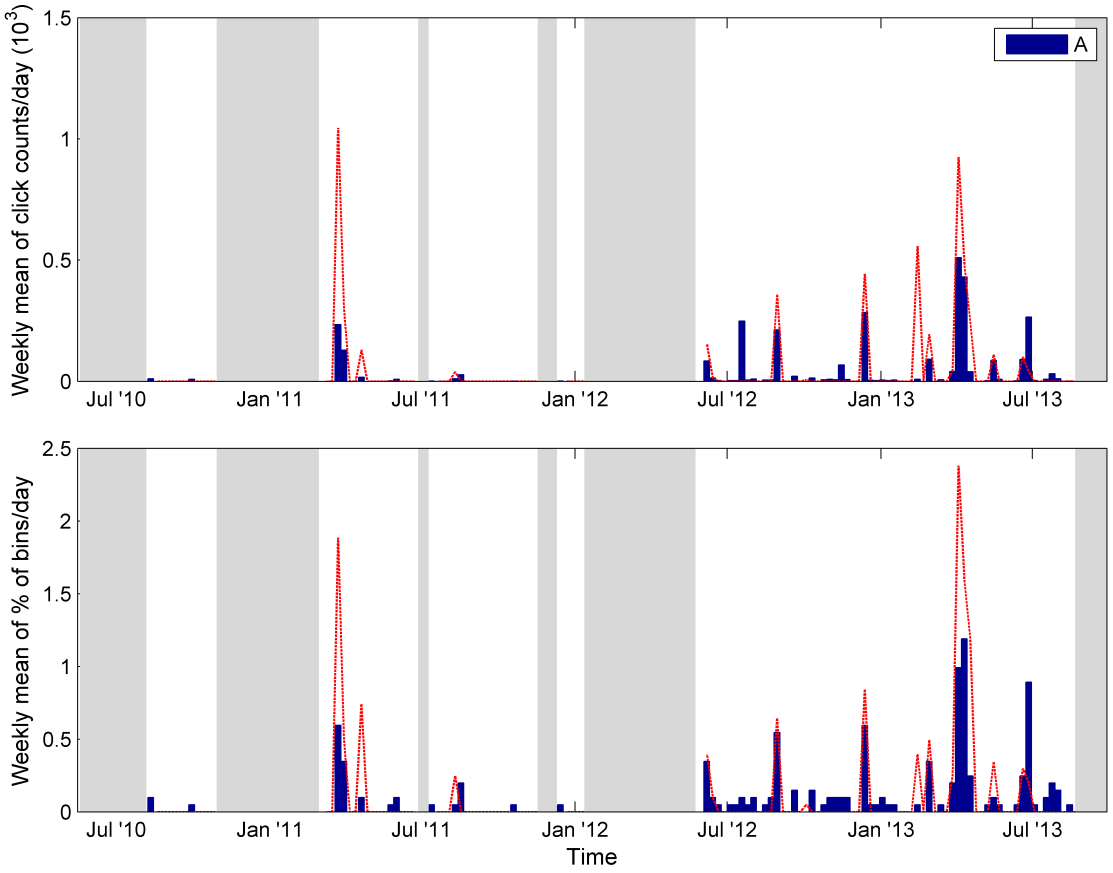


Figure 5.45: Occurrence of click type A at site DT, using automated (histogram) and manual (dashed line). Top panel shows classification as a weekly mean of thousands click counts per day. Bottom panel shows the same data as a weekly mean percentage of total daily five-minute bins.

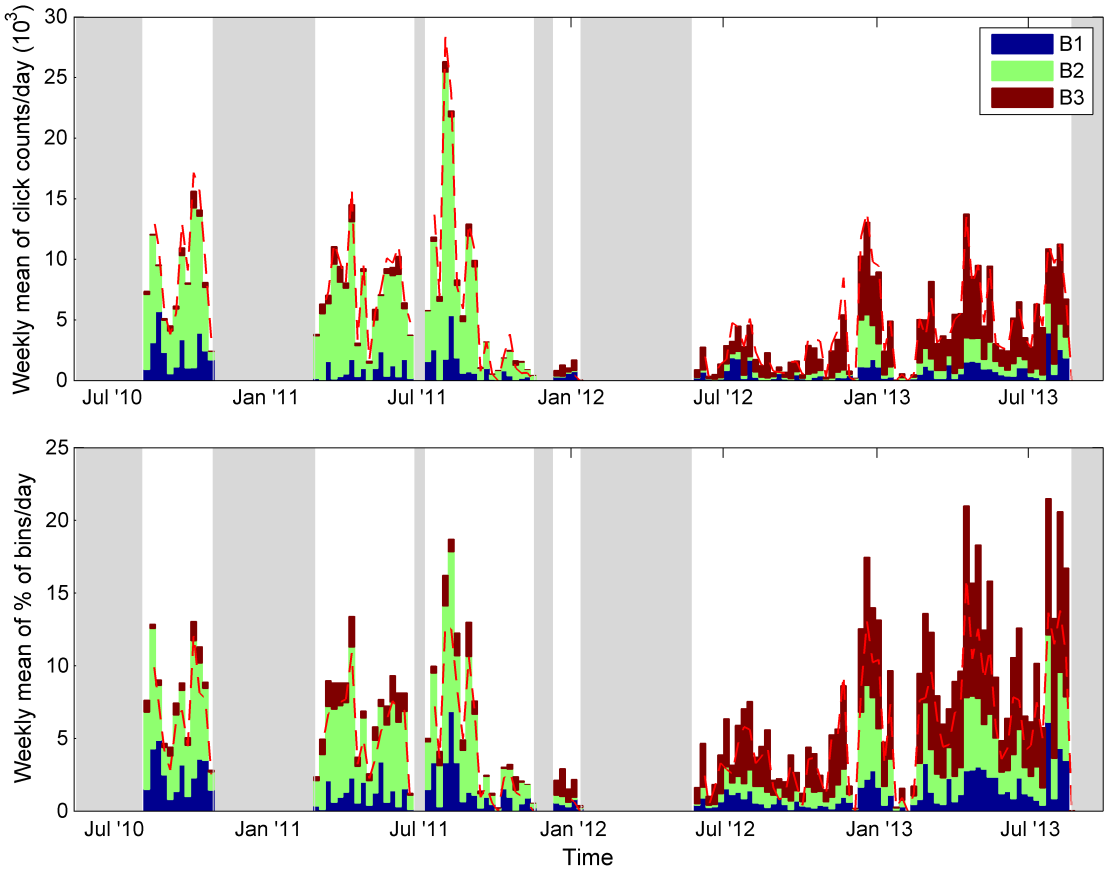


Figure 5.46: Occurrence of click B subtypes at site DT, using automated (histogram) classification methods. Manual classification (dashed line) did not distinguish between subtypes, but indicates type B totals. Top panel shows classification as a weekly mean of thousands click counts per day. Bottom panel shows the same data as a weekly mean percentage of total daily five-minute bins.

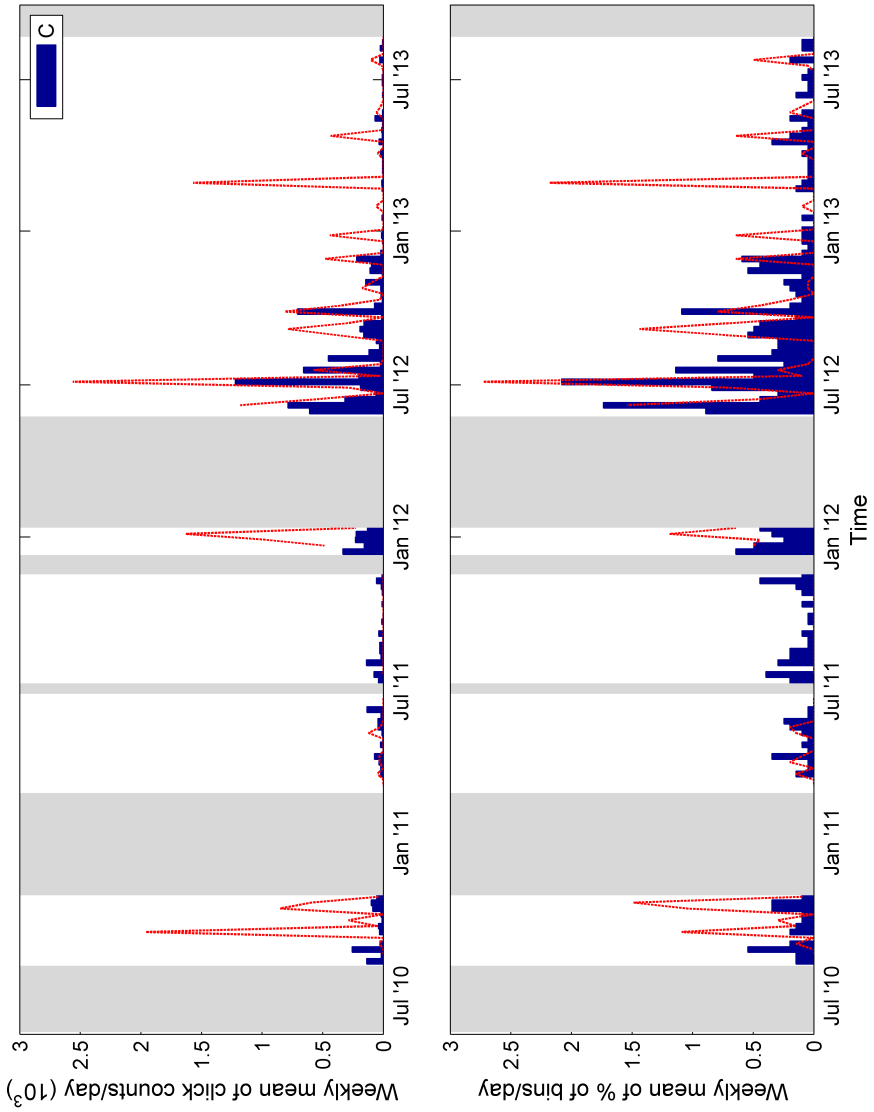


Figure 5.47: Occurrence of click type C at site DT, using automated (histogram) and manual (dashed line). Top panel shows classification as a weekly mean of thousands click counts per day. Bottom panel shows the same data as a weekly mean percentage of total daily five-minute bins.

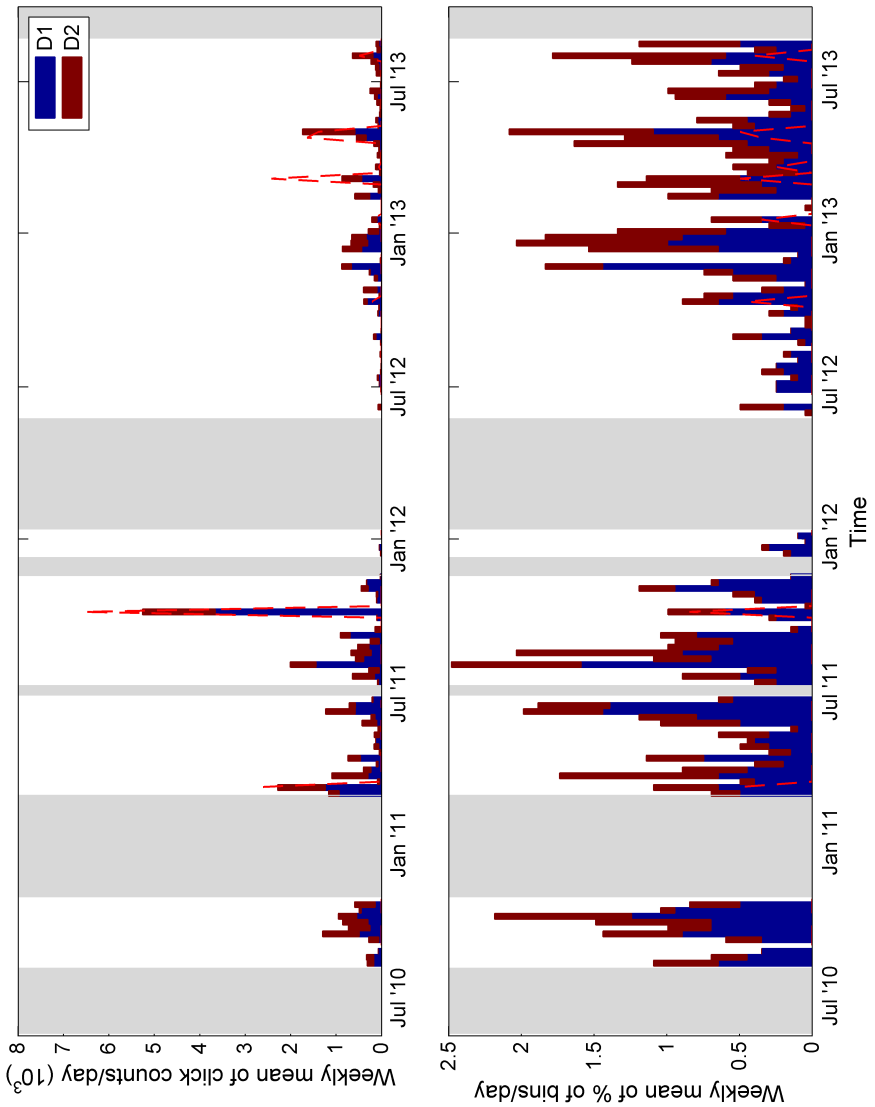


Figure 5.48: Occurrence of click D subtypes at site DT, using automated (histogram) classification methods. Manual classification (dashed line) did not distinguish between subtypes, but indicates type D totals. Top panel shows classification as a weekly mean of thousands click counts per day. Bottom panel shows the same data as a weekly mean percentage of total daily five-minute bins.

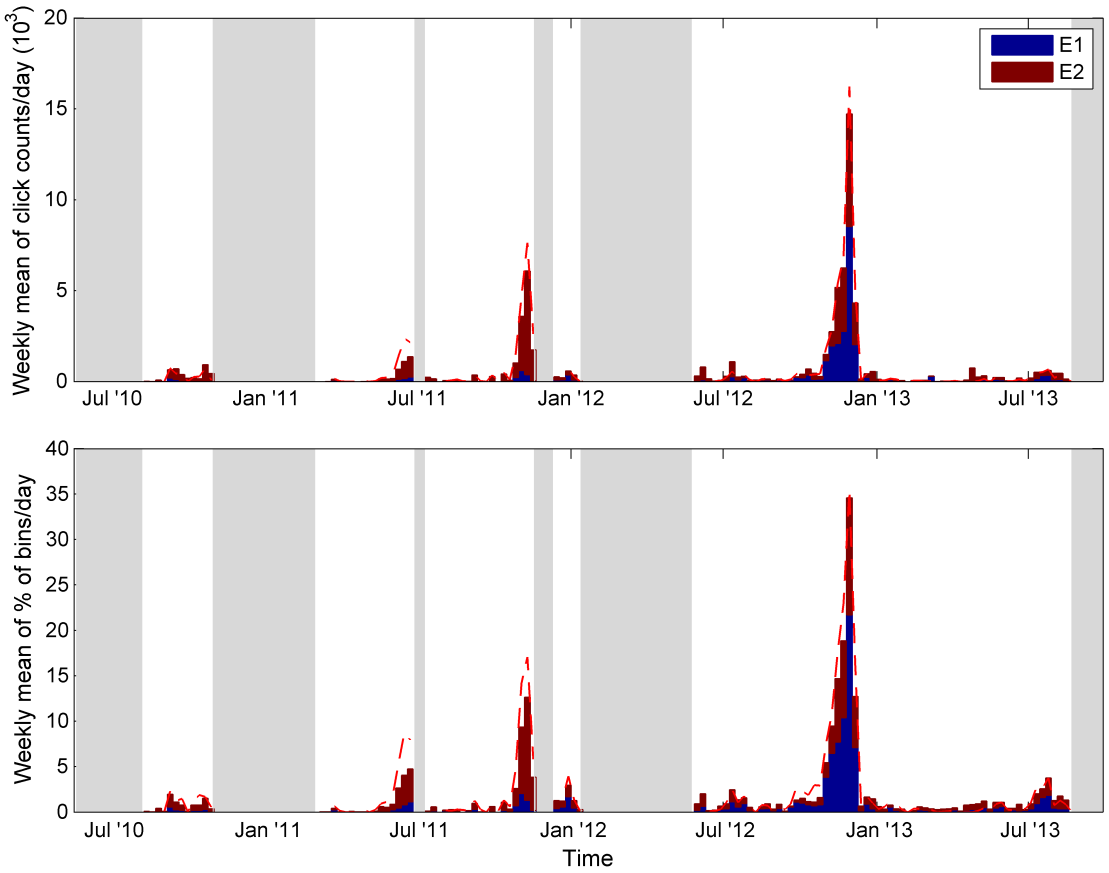


Figure 5.49: Occurrence of click E subtypes at site DT, using automated (histogram) classification methods. Manual classification (dashed line) did not distinguish between subtypes, but indicates type E totals. Top panel shows classification as a weekly mean of thousands click counts per day. Bottom panel shows the same data as a weekly mean percentage of total daily five-minute bins.

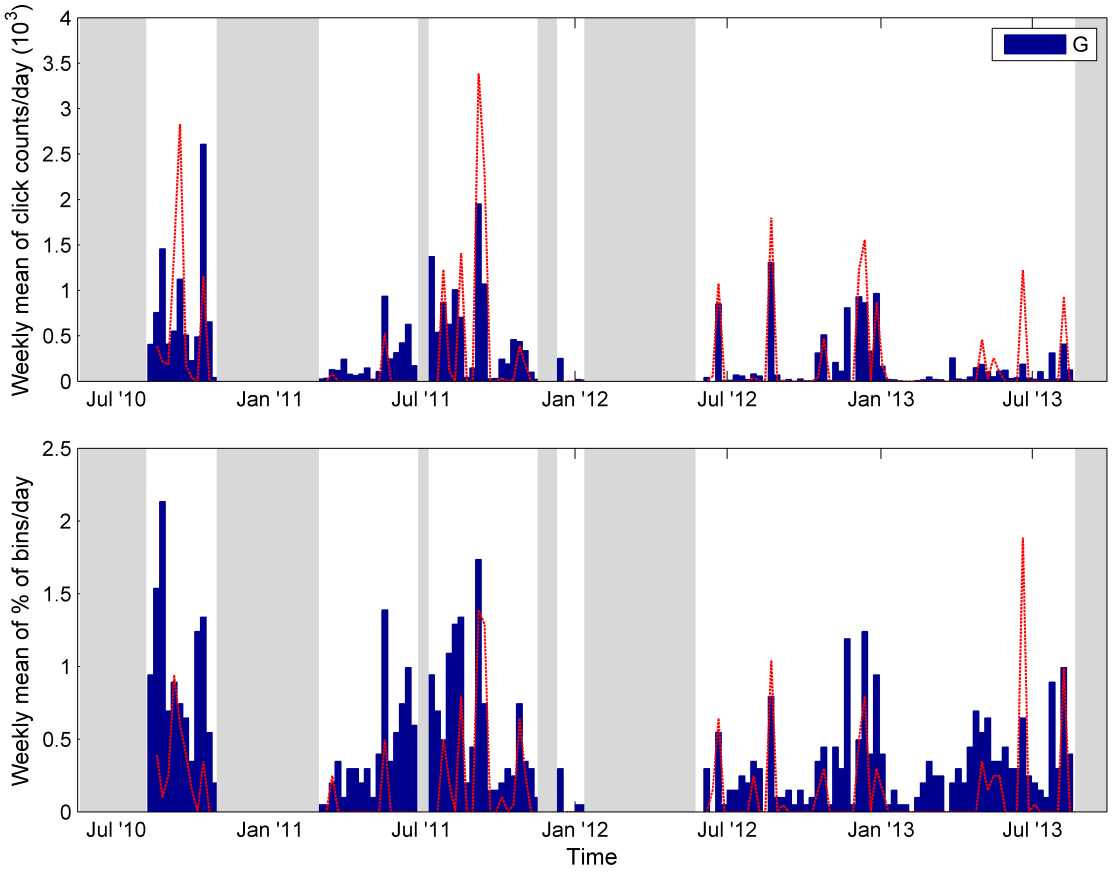


Figure 5.50: Occurrence of click type G at site DT, using automated (histogram) and manual (dashed line). Top panel shows classification as a weekly mean of thousands click counts per day. Bottom panel shows the same data as a weekly mean percentage of total daily five-minute bins.

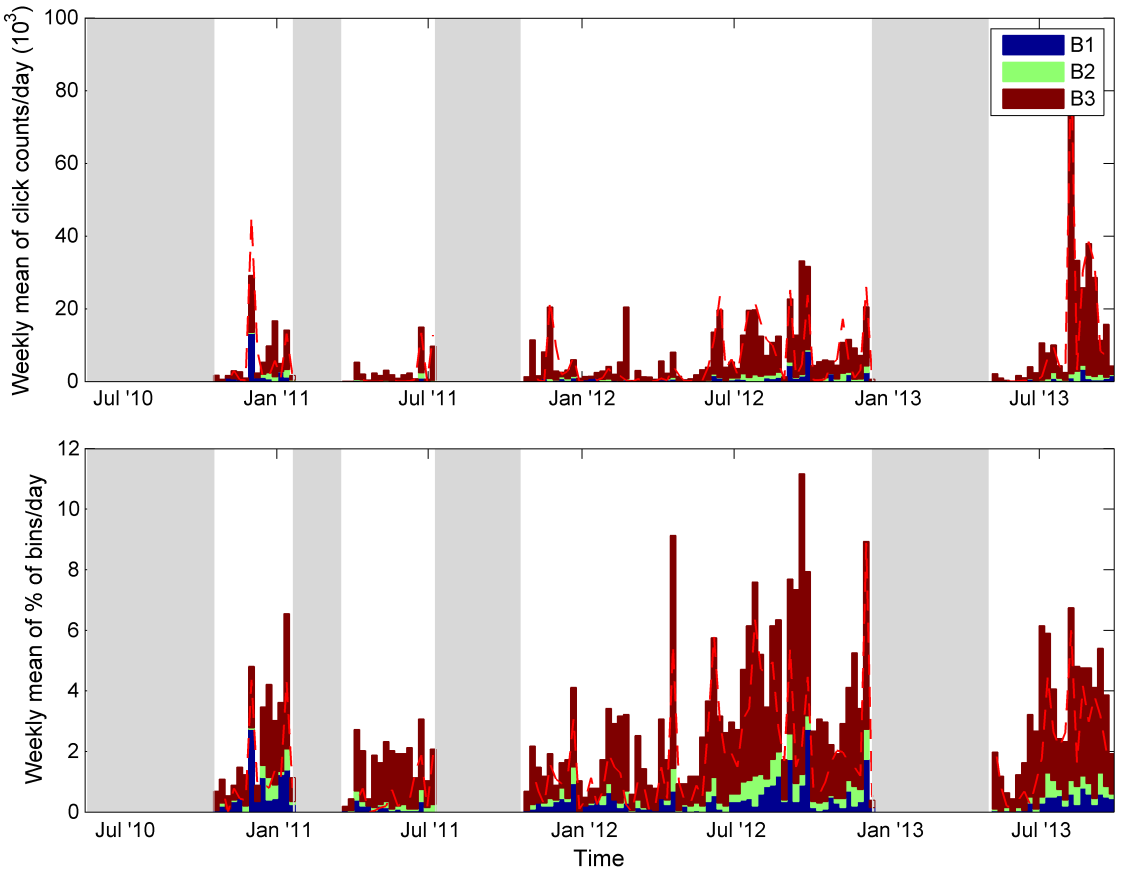


Figure 5.51: Occurrence of click B subtypes at site DC, using automated (histogram) classification methods. Manual classification (dashed line) did not distinguish between subtypes, but indicates type B totals. Top panel shows classification as a weekly mean of thousands click counts per day. Bottom panel shows the same data as a weekly mean percentage of total daily five-minute bins.

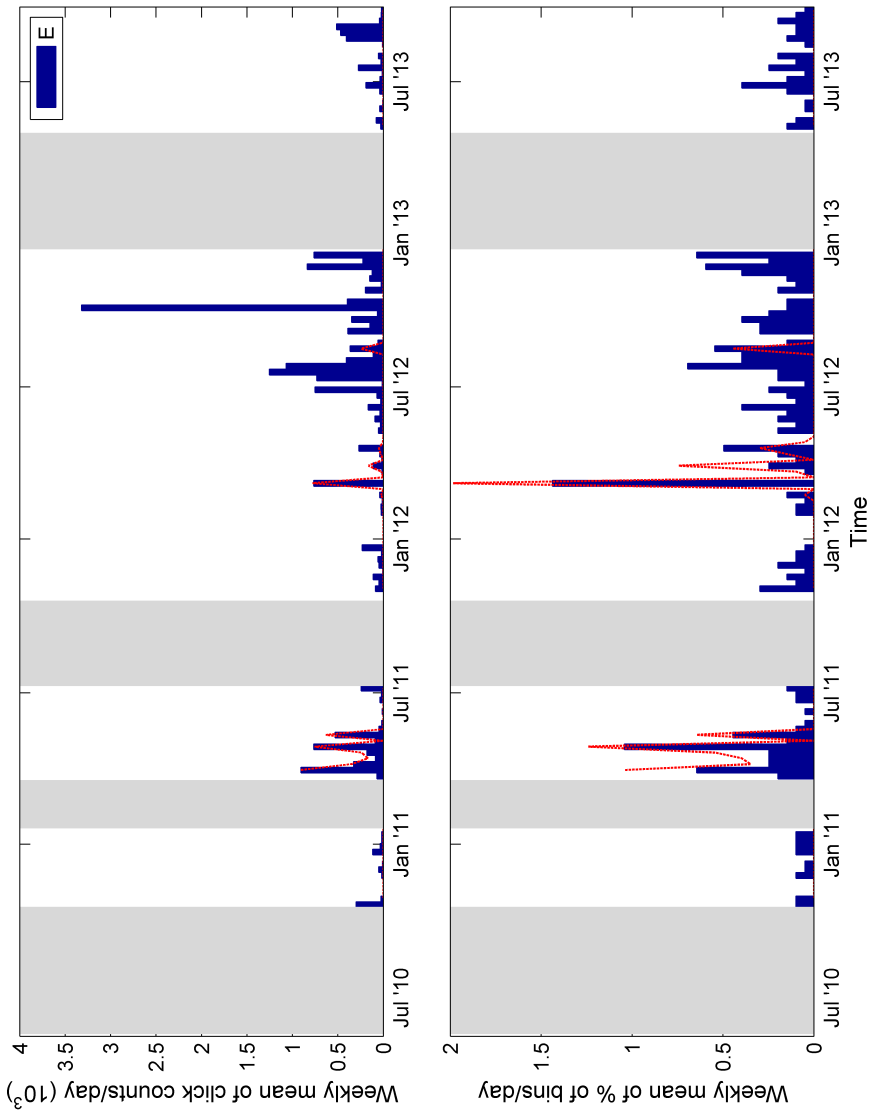


Figure 5.52: Occurrence of click type E at site DC, using automated (histogram) and manual (dashed line). Top panel shows classification as a weekly mean of thousands click counts per day. Bottom panel shows the same data as a weekly mean percentage of total daily five-minute bins.

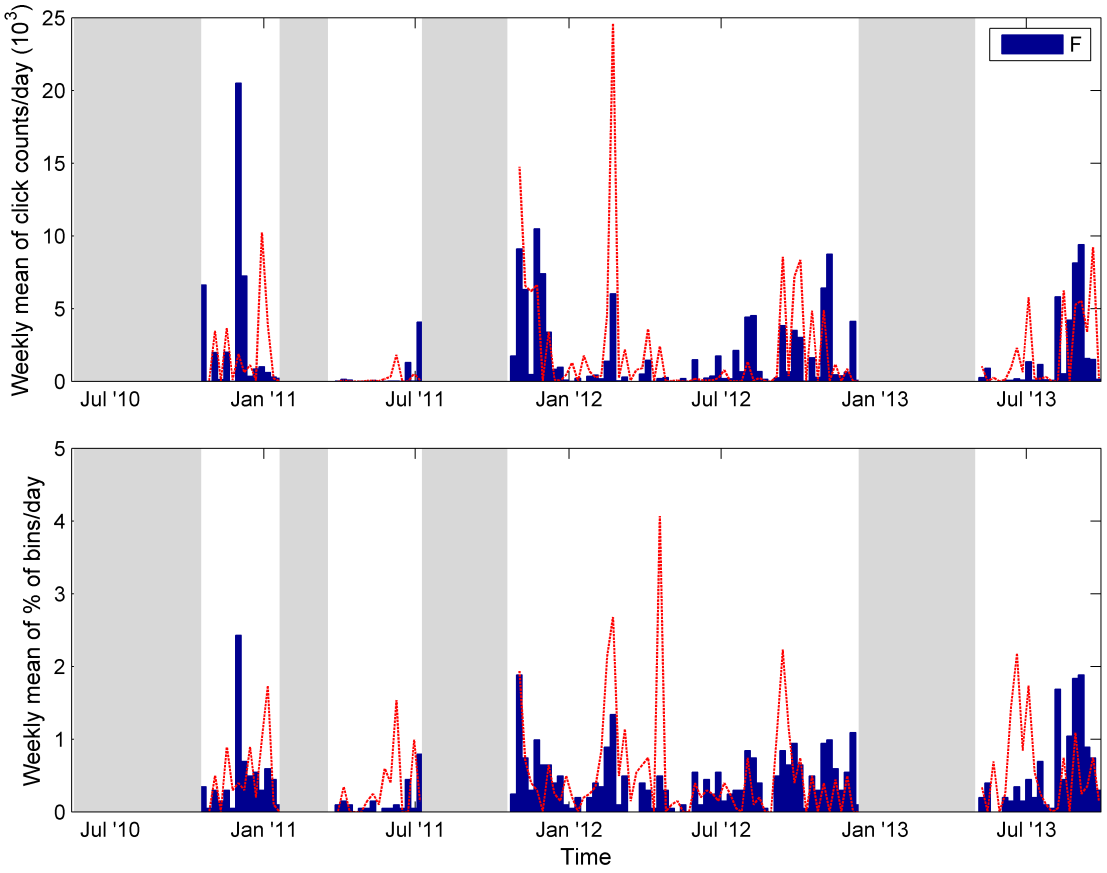


Figure 5.53: Occurrence of click type F at site DC, using automated (histogram) and manual (dashed line). Top panel shows classification as a weekly mean of thousands click counts per day. Bottom panel shows the same data as a weekly mean percentage of total daily five-minute bins.

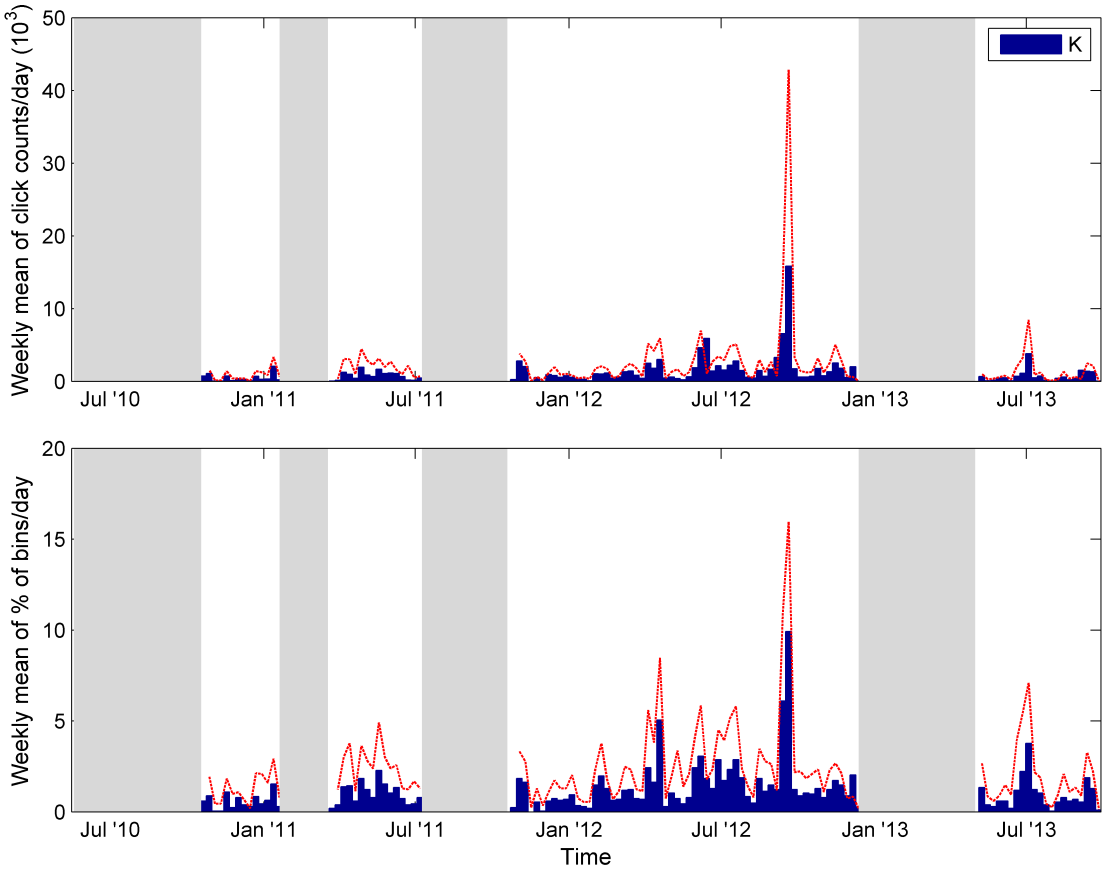


Figure 5.54: Occurrence of click type K at site DC, using automated (histogram) and manual (dashed line). Top panel shows classification as a weekly mean of thousands click counts per day. Bottom panel shows the same data as a weekly mean percentage of total daily five-minute bins.

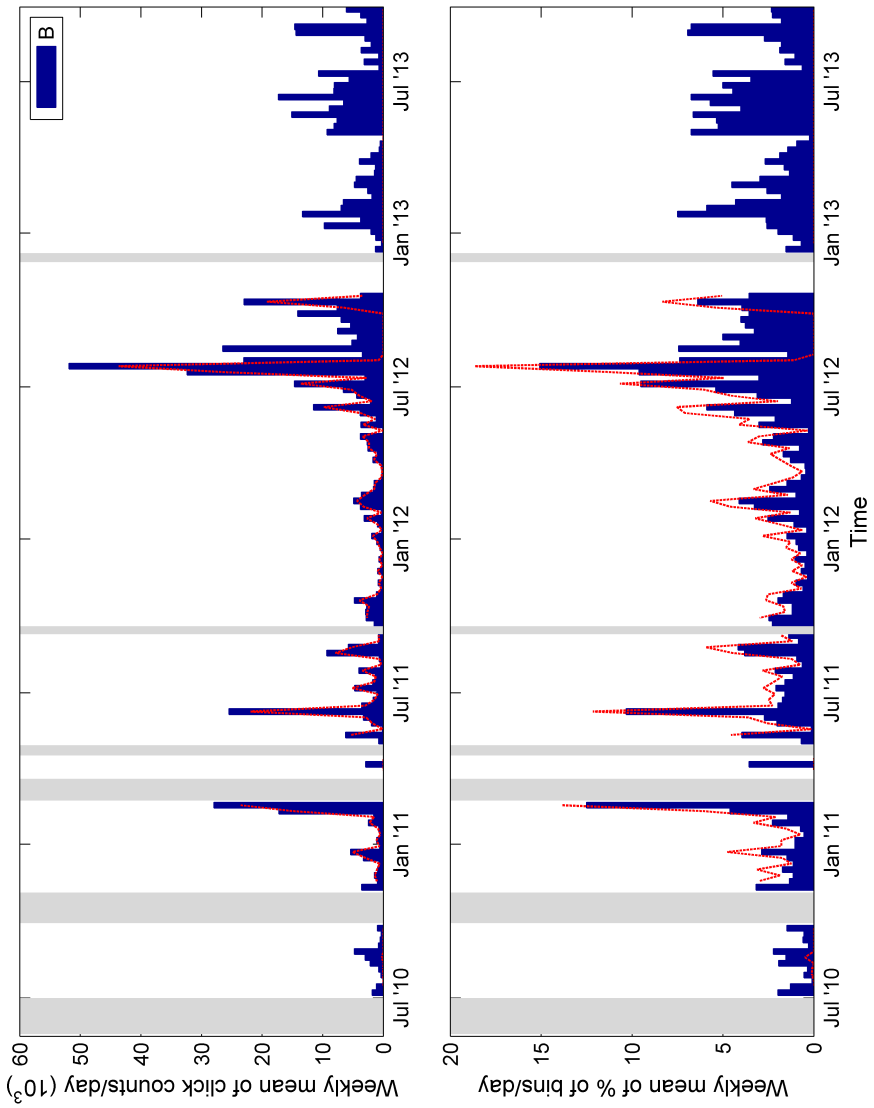


Figure 5.55: Occurrence of click type B at site MP, using automated (histogram) and manual (dashed line). Top panel shows classification as a weekly mean of thousands click counts per day. Bottom panel shows the same data as a weekly mean percentage of total daily five-minute bins

Chapter 6

Towed array recordings of Gulf of Mexico delphinid echolocation clicks

6.1 Abstract

Echolocation clicks recorded on seafloor sensors are useful for delphinid species identification, if they can be compared with recordings of known species. In order to obtain labeled recordings, towed array data were collected over three years in the Gulf of Mexico and along the southern US Atlantic coast. Visually verified, single species delphinid encounters were recorded. Spectral characteristics and inter-click interval (ICI) distributions of echolocation clicks from these recordings are presented here for ten species of delphinids commonly found in the Gulf of Mexico. These click types are compared with click types seen in Gulf HARP data for potential matches.

6.2 Introduction

HARPs and other deployable passive acoustic monitoring devices can record over long periods of time, in variable oceanographic conditions, around the clock (Mellinger et al., 2007). However, making sense of the data requires an intermediate step, in which simultaneous recording and visual efforts are used to associate species with characteristic signals. Conducting surveys with towed arrays and visual observation teams is a common strategy (e.g. Oswald et al., 2003; Rankin and Barlow, 2007; Baron et al., 2008; Soldevilla et al., 2008). Similar efforts were undertaken in the Gulf of Mexico to aid in acoustic species identification in the context of this study.

6.2.1 Comparing Array and HARP data

Although array recordings are a useful tool for interpreting unlabeled HARP data, there are considerable differences between recordings obtained via the two methods.

Noise

Towed array recordings are often noisier than seafloor sensor recordings. Surface sounds from wind and wave action are common. Vessel propeller, echosounder and electronic noise can be problematic. Noise is also associated with flow across the towed hydrophone. Because of the elevated noise levels, false positives rates are often higher in towed array detections than for seafloor sensors. Low amplitude clicks are likely to be masked.

Proximity

Animals are typically much closer to the sensor in the case of a towed array. When listening for near-surface dwelling odontocetes on a seafloor sensor, we can assume that

we mainly detect on-axis, or nearly on-axis click trains associated with one animal at a time. This is because off-axis clicks are not very detectable given the distance between the animals and the sensor, and the likelihood of each animal being on-axis is relatively small given their narrow transmitting beam.

As a simple illustration, consider a click produced on-axis relative to a sensor, with a source level, SL_{ON} of 210 $dB_{pp}re:1\mu Pa@1m$ (Au and Herzing, 2003; Au et al., 1995). If the received level (RL) threshold imposed by the detector is 115 $dB_{pp}re:1\mu Pa$, then maximum transmission loss (TL) is given by

$$TL = SL_{ON} - RL = 95dB_{pp}re:1\mu Pa \tag{6.1}$$

Assuming spherical spreading to approximate TL

$$TL = 20\log_{10}(r) + (r\alpha) \tag{6.2}$$

$$95dB_{pp}re:1\mu Pa = 20\log_{10}(r) + (r\alpha) \tag{6.3}$$

where r is the maximum detection range (slant range) for this signal, and α is frequency dependent absorption (Approximately 9 dB_{pp}/km at 35 kHz; Francois and Garrison, 1982). Solving for r in equation 6.3, yields a maximum detection slant range of roughly 3 km. If the receiver is a HARP sitting at a depth z of 1000 m below the sea surface, then the maximum horizontal detection range h = $\sqrt{r^2 - z^2}$ = 2.8 km.

In the off-axis case, SL_{OFF} is estimated to be at least 25 $dB_{pp}re:1\mu Pa$ lower than the SL_{ON}. Repeating the previous calculations:

$$SL_{OFF} = SL_{ON} - 25dB_{pp}re:1\mu Pa = 185dB_{pp}re:1\mu Pa@1m \tag{6.4}$$

$$TL = SL_{OFF} - RL = 70dB_{pp}re : 1\mu Pa \tag{6.5}$$

$$95dB_{pp}re : 1\mu Pa = 20\log_{10}(r) + (r\alpha) \tag{6.6}$$

Solving for r now yields a maximum detection range of approximately 1 km. Now if the receiver is a HARP sitting 1000 m below the sea surface, then the maximum horizontal detection range is zero meters, i.e. only off-axis clicks from animals directly above the instrument will be detected. An animal diving to a depth of 200m would be detectable off-axis at a maximum horizontal range of 0.6 km.

Based on these rough calculations it is clear that off-axis clicks are only detectable when the horizontal range between a clicking dolphin and a seafloor sensor is small. We can therefore assume that off-axis click detections are fairly rare. Since the odds that an animal will be on-axis with respect to a seafloor sensor are low, we are only likely to detect multiple animals simultaneously when a group is very close to the sensor.

This assumption does not hold for towed array data. In the towed array case, the animals and the sensor are both near the sea surface, at comparable depths, therefore horizontal and slant ranges are approximately equal. As a result, off-axis clicks produced within roughly 1km of the array will be detectable, and it is much more likely that many animals will be detected simultaneously, both on and off-axis.

Since sequential clicks are often from different animals in towed array recordings, the use of a simple first difference approach to capture ICI distributions (as described for HARP data in Chapter 5) is not as effective as it is for seafloor sensor data. An alternative is to identify click trains associated with individual animals. Once sets of sequential clicks from the single animals have been identified, ICI distributions can be computed for this subset of related clicks. One method for identifying click trains is cross-correlation

of click timeseries (e.g. *RainbowClick* click train identification, Gillespie and Leaper, 1996; Baggenstoss, 2011).

Proximity between animals and a towed array results in higher variability in received click structure and quality, relative to HARP recordings. This makes click characterization more challenging and complex. Animal proximity also changes the frequency content of received clicks. Higher frequencies attenuate more rapidly than lower frequencies with distance (Fisher and Simmons, 1977), therefore when the source animal is closer to the hydrophone, as in the array case, high frequency content will be more apparent (eg. Au, 2004; Au et al., 2012a,b).

Behavior

Although towed arrays can record at night, and can record animals that are not at the surface, visual identification typically requires daylight and surface behaviors. This means that identified array recordings are generally daytime recordings of animals engaged in surface behaviors such as bow riding, travel, or near-surface foraging. In contrast, daytime detections are rare in HARP data, where nighttime detections are more common (e.g. Chapter 4). At slope sites in particular, the vast majority of detections occur at night when animals are presumed to be foraging (Benoit-Bird, 2004; Klatsky et al., 2007; Herzing and Elliser, 2014). Click characteristics may differ between these behavioral modes.

6.2.2 Goals of this Chapter

Goals of this chapter are:

- to summarize a multi-year towed array dataset collected in the Gulf of Mexico and along the southern US Atlantic coast.

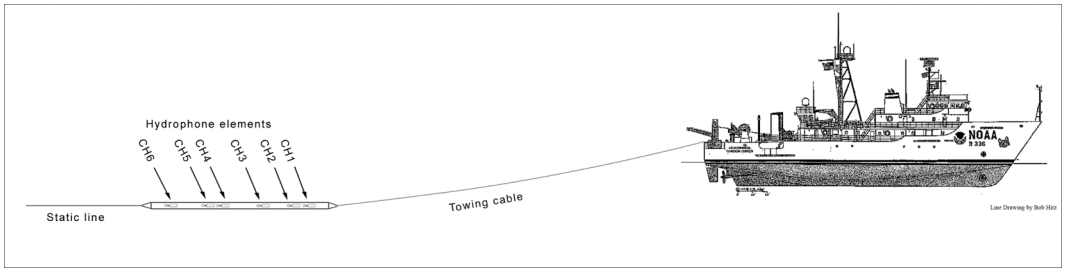


Figure 6.1: Schematic of R/V Gordon Gunter towing a six element hydrophone array used to record delphinid vocalizations during encounters with visually identified species.

- to generate inter-click interval (ICI) distributions for recorded species associated with confirmed visual species identifications.
- to extract click spectra for recorded species associated with confirmed visual identifications.
- to compare ICI and click spectra from array data to those from HARP data, and look for possible matches.

6.3 Methods

6.3.1 Data Collection: Array Recordings

Dolphins in the Gulf of Mexico and US Atlantic coast were recorded in 2010, 2011 and 2012 using a six-element towed hydrophone array on three summer cruises aboard the NOAA research vessel *R/V Gordon Gunter* (Figure 6.1). Cruises proceeded along pre-planned tracklines.

Delphinid groups were found by acoustic localization or visual observation. During daylight hours, localized groups were approached for species identification by professional marine mammal observers. When visual and acoustic detections were correlated in time and space, recordings were associated with the visually identified

species. Additional information was gathered based on visual observations, including estimated group size, presence of calves, and occurrence of mixed species groups.

Acoustic monitoring for marine mammals was conducted 24 hours a day on all cruises. Visual identification of recorded mammals was only possible during daylight hours. Acoustic signals from the array were recorded continuously, and monitored both visually and aurally for marine mammal vocalizations during all hours of acoustic survey effort. All detections, acoustic effort, deployments, environmental data, and recordings were logged to a *Microsoft Access* database using the *Logger 2000* software package (Gillespie, IFAW).

A six-channel hydrophone array was towed 274 m behind the ship, at an estimated depth of 15 to 18 meters at a minimum survey speed of 10 knots (Figure 6.2). Array depth was measured by attaching a dive watch to the head of the array during deployment. Measured depths were matched with vessel speed using time stamps associated with each measurement.

Each hydrophone element in the array consisted of a custom built preamplifier with a high pass filter set at 1 kHz and roll-off of either 100 or 200kHz connected to an SRD HS-150 hydrophone (Table 6.1). While the array was deployed, four channels were recorded continuously using a MOTU 896 HD digital audio interface, sampling at 192 kHz. The two remaining channels were recorded continuously at 500 kHz using a National Instruments (NI) USB 6251 data acquisition module (except in 2010, when NI recordings were sampled at 300 kHz). Recording was controlled through the *Logger 2000* software package.

Hydrophone channels 1 and 3 were monitored by trained acousticians using headphones connected to the MOTU to aid in aural detection of marine mammal calls. Channels 1, 3 and 4 were monitored visually using running, real-time spectrograms displayed by the recording program *Ishmael* (D. Mellinger, NOAA). Channels 1 and 3

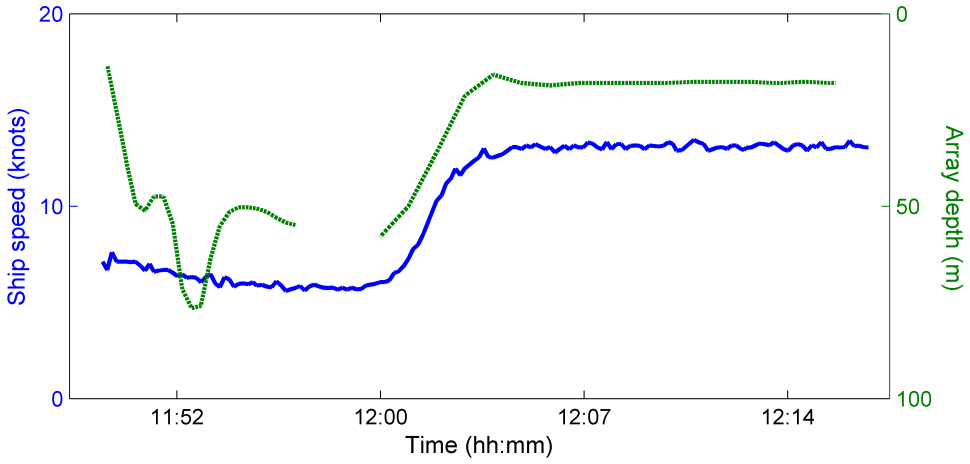


Figure 6.2: Relationship between vessel speed and array depth when towing.

were displayed across the full bandwidth (1 to 92 kHz) to monitor for echolocation clicks, while Channel 4 was displayed from 1 to 35 kHz to monitor for whistles.

Localization of signals of interest was performed by entering time difference of arrival (TDOA) data from Ishmael into *Whaletrak*, a localization and distance estimation program developed by J. Barlow (NOAA, SIO). TDOAs were computed between hydrophones with known spacing. Using localization, it was possible to detect odontocetes while underway and often to derive relative bearings to click or whistle producing animals when their received levels rose sufficiently above background noise. Localization also enabled acousticians to determine whether acoustic detections were spatially coincident with visual detections.

6.3.2 Click Detection and Characterization

Clicks were detected using the click detector described in Chapter 4. Only recordings associated with confirmed single species visual identifications were analyzed. The analyses reported here are based on NI recordings only, because they were found to be of higher quality than the MOTU recordings. Recordings containing few detectable clicks and those containing click types associated with other known species (e.g. Risso's

dolphin) were excluded from the analyses.

ICI Analysis

Click timeseries were extracted by finding the maximum amplitude of each detected click and extracting the time period between 30μs before to 40μs after this peak. The timeseries of each click was then cross-correlated with a set of subsequent candidate clicks, using an unbiased cross-covariance (mean-removed cross correlation) function evaluated at 41 sequential one-sample time lags ([-20:20]).

Candidate matching clicks were defined as clicks occurring within 0.5 seconds of the original click. The candidate with the highest maximum cross-covariance was considered the best match. If the maximum cross-covariance between a click and it's best matching candidate exceeded a threshold value of 10^5, the time interval between the two clicks was computed and retained. The threshold excluded poor matches from the dataset when no good match was found. ICIs were rounded to the nearest millisecond. ICI distributions were generated for each species by binning the time differences into 10ms bins between zero and 300 ms.

Echolocation Click Spectral Analysis

A subset of recorded echolocation clicks were digitally clipped, due to the broad variability of received click amplitudes and lack of adjustment of recording tools. Another subset were of poor quality due to low amplitudes. Detected clicks were therefore pruned to exclude both high and low amplitude clicks, retaining only mid-amplitude clicks for spectral clustering.

Click spectra were normalized as described in chapter 5, and then clustered using an iterative two-step process. For each species, click spectra from all detections were assembled. A subset of 1000 clicks (or 50% of total clicks, whichever was

smaller) was then selected at random from the overall set and clustered according to pairwise correlation distances between normalized spectra, using the previously described modularity algorithm (see Chapter 5). A high pruning threshold of 0.92 was used to trim the click networks. A modularity coefficient of 0.5 was used to facilitate identification of small clusters.

Mean spectra were computed and retained for clusters which contained at least two percent of the clicks in the clustered subset. This process of selecting a subset of clicks, clustering them, and computing spectral means of resulting clusters was repeated 50 times for each species. The resulting set of mean spectra for a given species was then re-clustered to obtain summary spectra, using the same algorithm, but with a modularity coefficient of 1.

Table 6.1: Towed array hydrophone specifications. Hydrophone array was towed approximately 284 feet behind the vessel.

Array Ch.	Rolloff	Dist. from Ch1	Sampling device	Sampling Rate
1 (front)	100 kHz	0 cm	MOTU Ch1	192 kHz
2	200 kHz	22 cm	NI Ch1	300 or 500 kHz
3	100 kHz	78 cm	MOTU Ch2	192 kHz
4	100 kHz	193 cm	MOTU Ch3	192 kHz
5	200 kHz	233 cm	NI Ch2	300 or 500 kHz
6 (rear)	100 kHz	333 cm	MOTU Ch4	192 kHz

6.4 Results

Visually verified single-species delphinid encounters were collected in the Gulf of Mexico and along the southern US Atlantic coast (Figure 6.3). Recordings of ten

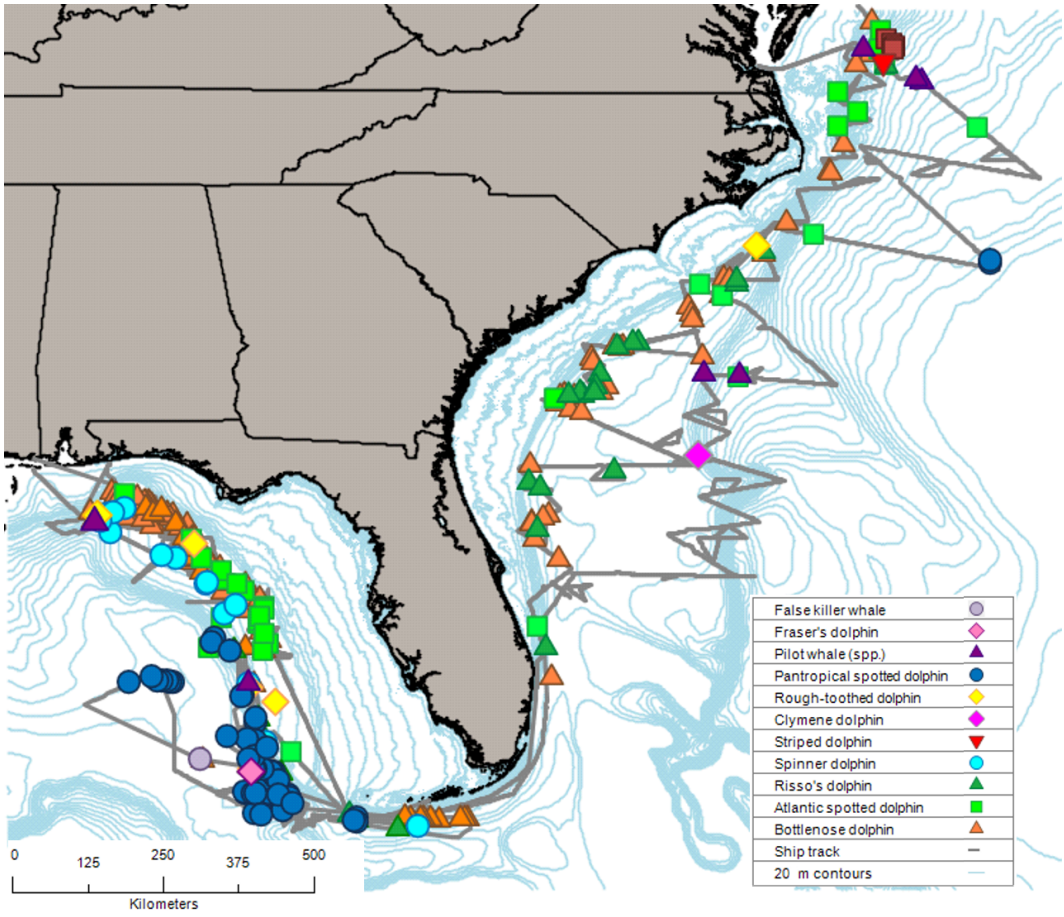


Figure 6.3: Map of all towed array recording effort conducted in tandem with NOAA Southeast Fisheries Science Center's ship-based marine mammal surveys in the Gulf of Mexico and along the Atlantic Coast between 2010 and 2012. Tracklines in gray. Colored markers indicate sighting locations of delphinid species listed in the legend.

Table 6.2: Click counts from visually confirmed array detections used for ICI and spectral analyses respectively. Modal ICI is reported for each species. ICI data were divided into 10ms bins.

Species	**Clicks for ICI Analysis**	**Clicks for Spectral Analysis**	**Modal ICI Bin (ms)**
Atlantic spotted dolphin	21,345	19,130	60
Pantropical spotted dolphin	26,924	23,873	70
Spinner dolphin	11,009	10,678	100
Striped dolphin	1,058	1,048	140
Bottlenose dolphin	45,742	20,574	50
False killer whale	638	608	140
Fraser's dolphin	1,768	1,690	200
Pilot whale spp.	5,000	4,647	160
Risso's dolphin	1,796	1,563	110
Rough-toothed dolphin	4,914	4,514	100

species contained sufficient numbers of detected clicks for analysis (Table 6.2). These included: pantropical spotted dolphin (*Stenella attenuatta*), Atlantic spotted dolphin (*Stenella frontalis*), spinner dolphin (*Stenella longirostris*), striped dolphin (*Stenella coeruleoalba*), Risso's dolphin (*Grampus griseus*), pilot whale (*Globicephala* spp.), rough-toothed dolphin (*Steno bredanensis*), false killer whale (*Pseudorca crassidens*), Fraser's dolphin (*Lagenodelphis hosei*), and bottlenose dolphin (*Tursiops truncatus*). One additional Stenellid dolphin species, Clymene dolphin (*Stenella clymene*), was encountered but too few high quality clicks were detected for further analysis.

6.4.1 ICI

ICI distributions were generated for ten delphinid species. Modal ICI values for the Atlantic and pantropical spotted dolphins were similar, with modes of 60 and 70 ms respectively (Figures 6.4A and 6.4B). The modal ICI for spinner dolphins was higher, at 100ms (Figure 6.4C). ICI distributions for the Stenellid dolphins were similar to one another, with Atlantic spotted dolphin ICIs skewed slightly lower, and spinner dolphin slightly higher than pantropical spotted dolphin. The striped dolphin dataset was very small, and the ICI histogram showed no clear peak.

The bottlenose dolphin ICI distribution was close to that of Atlantic spotted dolphin, with a modal ICI of 50 ms (Figure 6.5F), the lowest of the recorded species. Rough-toothed dolphin had an ICI distribution centered around a mode at 100 ms (Figure 6.5E). Modal ICIs for the blackfish species were larger in general than for the Stenellid dolphins. Risso's dolphin had the lowest modal ICI of the blackfish, at 110 ms. Pilot and false killer whale ICIs were higher, at 160 and 140 ms respectively. Fraser's dolphin ICI data were noisy, due in part to the small sample size (a single encounter). The main ICI mode was at 200 ms, but the distribution may be bi-modal, with a secondary peak at approximately 100 ms.

6.4.2 Click Spectra

For all species, fewer clicks were used for spectral analysis than for ICI analysis, because high amplitude, clipped signals were pruned out of the spectral set.

Risso's dolphin mean click spectra were the most distinct, with a characteristic three-peak pattern across most clusters (Figure 6.6). A subset of pantropical spotted and Atlantic spotted, spinner and striped dolphin mean click spectra were distinct, with variations of a bimodal-structure (Figures 6.7 to 6.10).

Pilot whale mean click spectra had some variation of an energy peak at roughly 15kHz (Figure 6.11). False killer whale and rough-toothed dolphin mean click spectra were skewed toward lower frequencies than the Stenellid dolphin spectra (Figures 6.12 and 6.13).

The remaining click types typically had generic unimodal shapes. Bottlenose dolphin click spectra had few distinguishing features and were generally variable (Figure 6.14). Fraser's dolphin mean click spectra did not appear to have distinguishing features (Figure 6.15) but sample size was small, and further recordings might yield more details.

6.5 Discussion

6.5.1 Towed Array Data Collection in the Gulf of Mexico

Detectability of echolocating dolphins in the Gulf of Mexico was found to be lower than expected based on array recording efforts in other areas. Animals were often only acoustically detected once they had approached the ship, and then only for a brief period of time. This is likely at least partially due to the warm summer surface layer in the Gulf during summer surveys. As discussed in Chapter 2, warm waters in the Gulf

cause sound energy produced near the surface to be strongly refracted down toward the seafloor, limiting detection ranges using a towed array (Figure 6.16).

Propeller noise was also constant challenge in these data, particularly since visual survey speeds were in excess of 10 knots, and the survey vessel was large and powerful. False positives due to cavitation noise are common across all detections, and vessel noise likely masked weak and distant cues. At the same time when animals were near the hydrophone array, digital signal clipping became an issue. This is particularly problematic for the analyses described here, and for comparison with HARP data, because the high amplitude, on-axis clicks recorded by seafloor sensors are likely distorted in the array recordings. Clipping is definitely an issue for spectral comparisons, but may also affect ICI distributions because cross-correlated time-series may have been distorted for some clicks of interest. Future attempts to compare seafloor sensor and towed array data may have greater success if detection of low amplitude clicks was sacrificed in the interest of ensuring that high amplitude clicks were recorded without clipping.

6.5.2 ICIs

ICI distributions differed between species, supporting the use of ICI as a classification feature, and confirming the significance of ICI differences observed in seafloor sensor recordings. The Stenellid dolphins all had similar ICI distributions to one another. Although the modal ICI for spinner dolphins was higher than for Atlantic and pantropical spotted dolphins, this is likely because the number of ICIs retained was small. The striped dolphin ICI was also small, and no clear modal peak was visible. The overall distributions for both species suggested that their typical ICIs were comparable to those of Atlantic and pantropical spotted dolphins.

Bottlenose dolphin ICI was highly variable, with a mode similar to that of Atlantic spotted dolphin. This indicates that the main species expected on the continental shelf

in the Gulf of Mexico (bottlenose and Atlantic spotted dolphins), cannot be easily distinguished using ICI.

In general, the blackfish species, including Risso's dolphin, pilot whales, and false killer whales, were found to have longer modal ICIs than the Stenellid and bottlenose dolphins. Differences may reflect differences in prey type and behavior. More data is needed for some of the rarer species, which were not recorded often enough to generate meaningful distributions.

It is important to note that these ICI distributions are based on daylight encounters at the surface and may differ significantly from the predominantly nocturnal encounters recorded by the Gulf of Mexico HARPs. Additionally, a subset of the recordings were collected in the Atlantic. Although the species are the same, acoustic characteristics may differ between populations. Little is known about genetic exchange between Gulf of Mexico delphinid populations and those in the greater Atlantic.

6.5.3 Click Spectra

Click spectra varied widely even within an encounter. The clicks that constitute the majority of signals detected on a seafloor instrument (on or nearly on-axis), are likely the minority in array recordings. Many clicks fell into a generic featureless category for all species. However a number of species had one or more subsets of more distinctive click types that may be characteristic, and therefore useful for classification.

Stenellid dolphin clicks often had a bimodal shape, similar to spectra seen on seafloor sensors, but with higher energy at the upper end of the normalized spectra (>40 kHz). This is expected because animals are closer to the towed array. Spectral details in the lower frequency end of the spectra (below 15kHz) were often obscured by noise. The normalization and truncation parameters used here are consistent with those used for HARP data, in order to facilitate comparison, however, there may be more distinct

features above 60 kHz. In general, these spectra are useful, but difficult to compare with those seen on seafloor sensors, because of the many differences in recording conditions.

Bottlenose dolphins were most often encountered in shallow waters (Figure 6.3), where ship noise was worsened by of reflections off the seafloor. False positives are likely more common under such conditions, and this may partially explain the lack of consistency in the bottlenose dolphin summary spectra. Pilot whale clicks had consistent small energy peaks at approximately 15 kHz, which may be useful for identification. Rough-toothed dolphin and false killer whale clicks seemed to have higher energy at lower amplitudes relative to a number of the other species. The Fraser's dolphin spectra set was small, and more data is needed before drawing conclusions about the species' average click characteristics.

A two step clustering process was used because of the high variablilty and number of clicks detected. This works well for the large datasets, but is probably unnecessary for small datasets of a few thousand clicks. In the interest of treating all sets the same, it was used here on all species' sets. A challenge with these data is that some encounters yielded more detections than others, and therefore a small number of encounters may dominate in some datasets. If those dominant encounters are not representative, or have very high false positive rates, they may obscure information contained in smaller encounters. In this analysis, all clicks associated with a given species are treated equally. However, some encounters may be higher quality than others, and a weighting strategy might balance information across detections in further analyses.

6.5.4 Comparison to HARP types

Based on this work, click types B and C in the HARP data (see Chapter 5) are likely associated with Stenellid dolphins. Further labeled acoustic data might improve the ability to associate these species with specific subtypes. Bottlenose dolphin ICI was

in the same range, and might also be consistent with type F. Pilot whale ICIs and spectra were consistent with type D clicks from the HARP datasets.

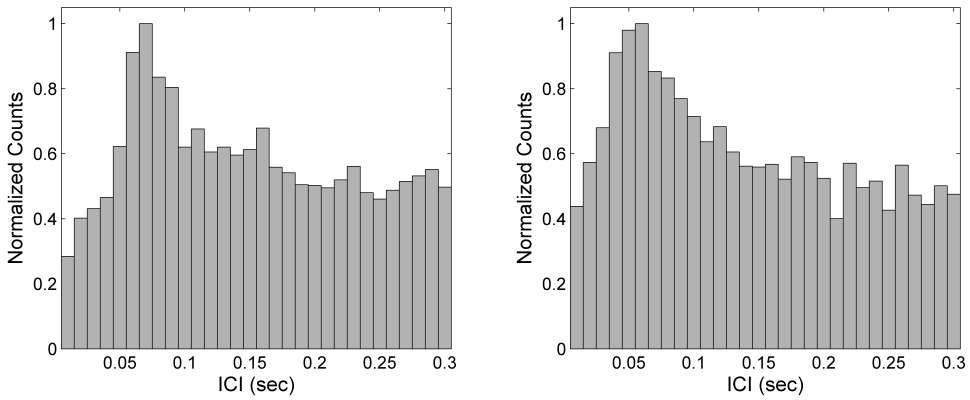
Rough-toothed dolphin ICI was similar to type G from the HARP data, and spectra were not inconsistent with the type. However Fraser's dolphin may click at similar rates, and more work is needed to develop methods for distinguishing the two.

Click type A in the HARP data is likely associated with one of the blackfish species, given it's long ICI. False killer whale is a likely match for that type, but melon-headed whale (*Peponocephala electra*) is another possible candidate which cannot be ruled out, because it was not encountered during these surveys. This may be because melon-headed whales are more often encountered in the western half of the northern Gulf, where little survey effort was applied. Melon-headed whales are likely among the species detected in the HARP datasets. Killer whale and pygmy killer whale have also been reported in the Gulf of Mexico, but they are uncommon (estimated population sizes 49, CV = 0.77, and 323, CV = 0.60, respectively; Mullin, 2007), and probably infrequently detected on seafloor sensors.

Datasets were strongest for pantropical and Atlantic spotted dolphins, bottlenose dolphins, and pilot whales, which are common species in the Gulf of Mexico, and were encountered numerous times. Risso's dolphin click spectra are well documented elsewhere (Soldevilla et al., 2008), and are clearly associated with click type E from the HARP data. Further data is needed to improve the click characterizations for the other three species, as well as the species not recorded here. For truly comparable data, an alternative to towed arrays may be necessary.

6.6 Conclusion

Echolocation click characteristics from towed array recordings of ten delphinids species commonly found in the Gulf of Mexico were examined for characteristic ICI distributions and spectral features. These characteristics were compared with click types extracted from seafloor sensor recordings. Likely matches were found for Stenellid dolphin and pilot whale click types. Other potential matches were found, but further data are needed to refine them. Array data is challenging to compare to HARP data, and further recording efforts would be necessary to fine-tune associations between click types and species. Possible improvements to the array data collection process, including gain and array depth adjustments and revised encounter protocols might facilitate future comparisons.



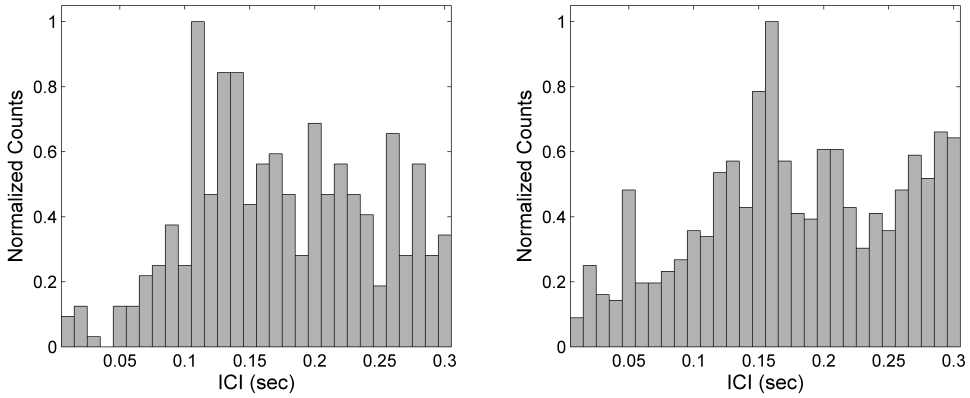
(A) Pantropical spotted dolphin

(B) Atlantic spotted dolphin

(C) Spinner dolphin

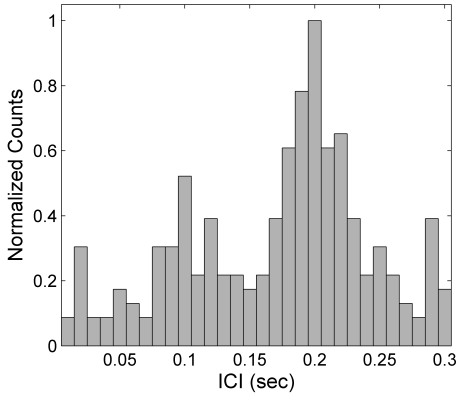
(D) Striped dolphin

Figure 6.4: ICI distributions for Stenellid dolphins based on visually identified, single species array recordings.

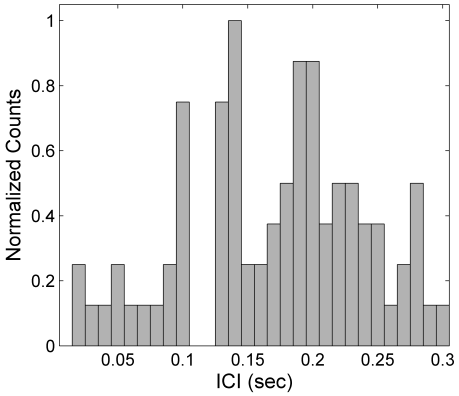


(A) Risso's dolphin

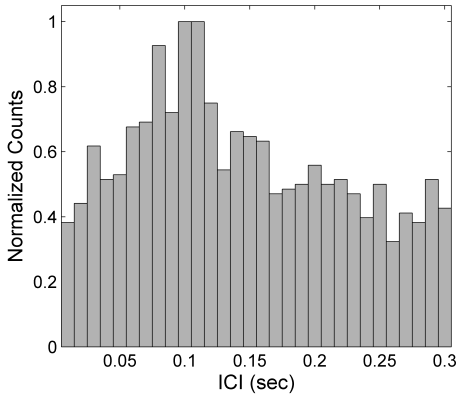
(B) Pilot whale (spp.)



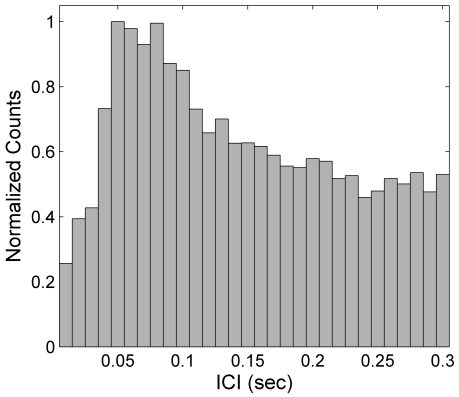
(C) Fraser's dolphin



(D) False killer whale



(E) Rough-toothed dolphin



(F) Bottlenose dolphin

Figure 6.5: ICI distributions for Gulf of Mexico delphinid species based on visually verified, single-species array recordings. Dark lines indicate spectral means, with light lines indicating one standard deviation above and below the mean.

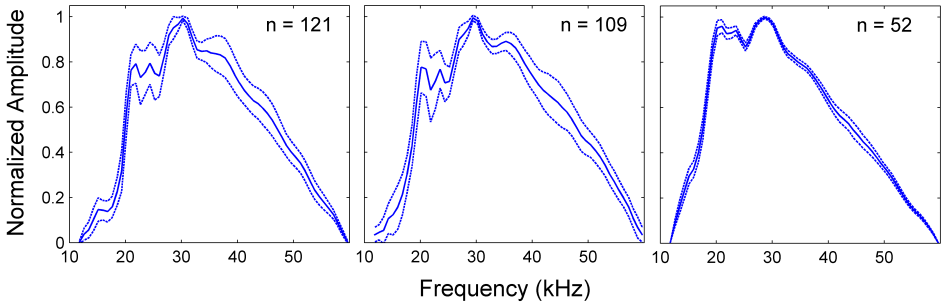


Figure 6.6: Representative Risso's dolphin (*Grampus griseus*) mean echolocation click summary spectra from towed array recordings of visually verified encounters. Dark lines indicate spectral means, with light lines indicating one standard deviation above and below the mean.

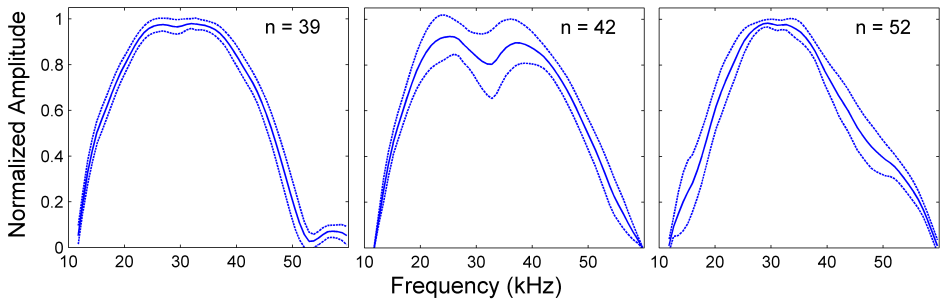


Figure 6.7: Representative pantropical spotted dolphin (*Stenella attenuatta*) mean echolocation click summary spectra from towed array recordings of visually verified encounters. Dark lines indicate spectral means, with light lines indicating one standard deviation above and below the mean.

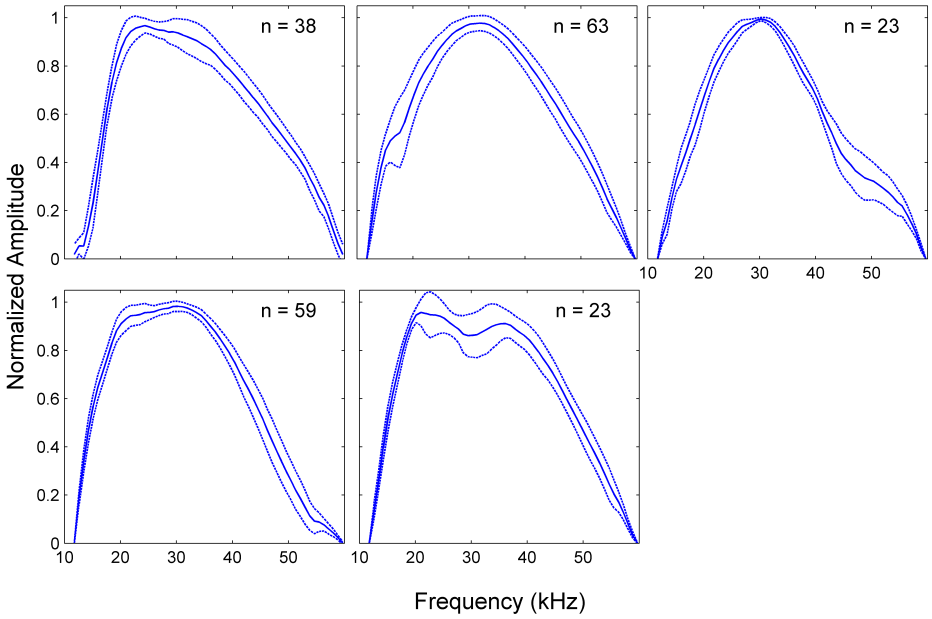


Figure 6.8: Representative Atlantic spotted dolphin (*Stenella frontalis*) mean echolocation click summary spectra from towed array recordings of visually verified encounters. Dark lines indicate spectral means, with light lines indicating one standard deviation above and below the mean.

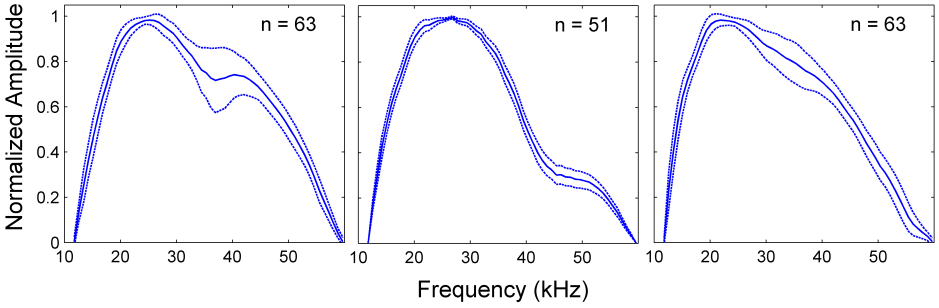


Figure 6.9: Representative spinner dolphin (*Stenella longirostris*) mean echolocation click summary spectra from towed array recordings of visually verified encounters. Dark lines indicate spectral means, with light lines indicating one standard deviation above and below the mean.

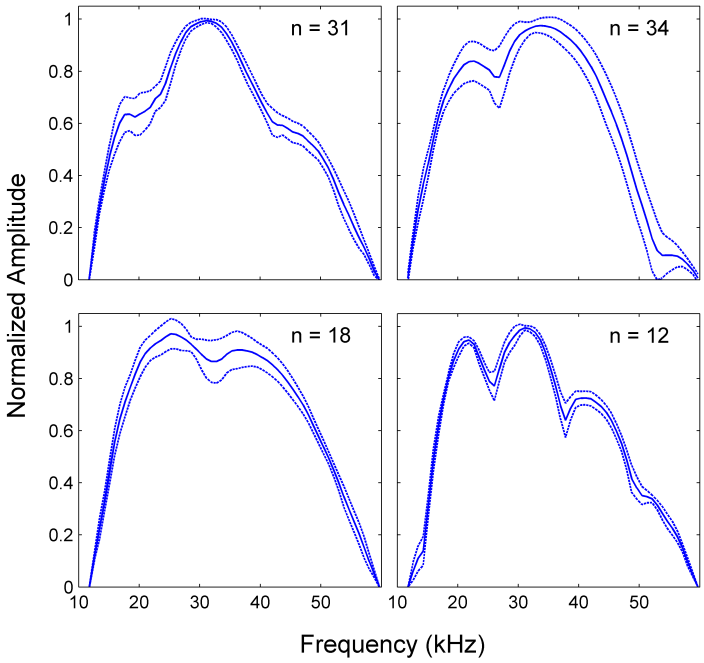


Figure 6.10: Representative striped dolphin (*Stenella coeruleoalba*) mean echolocation click summary spectra from towed array recordings of visually verified encounters. Dark lines indicate spectral means, with light lines indicating one standard deviation above and below the mean.

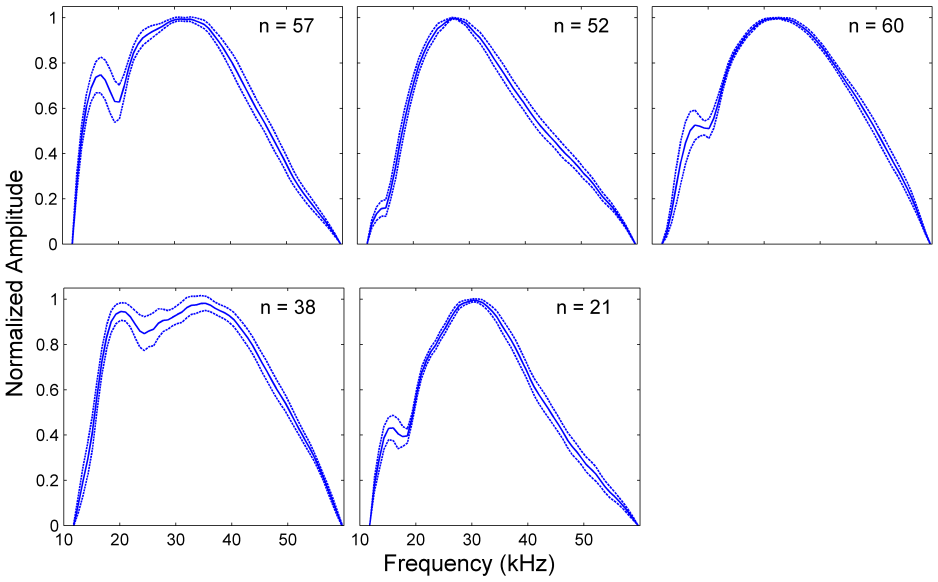


Figure 6.11: Representative pilot whale (*Globicephala* spp.) mean echolocation click summary spectra from towed array recordings of visually verified encounters. Dark lines indicate spectral means, with light lines indicating one standard deviation above and below the mean.

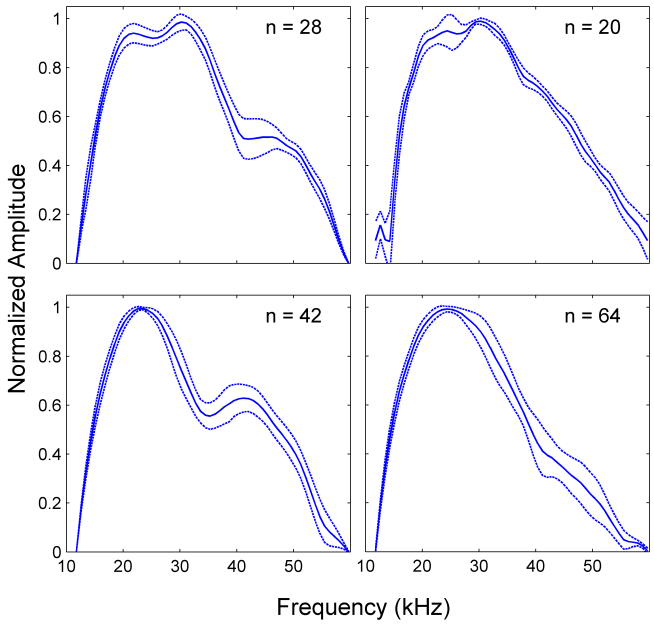


Figure 6.12: Representative false killer whale (*Pseudorca crassidens*) mean echolocation click summary spectra from towed array recordings of visually verified encounters. Dark lines indicate spectral means, with light lines indicating one standard deviation above and below the mean.

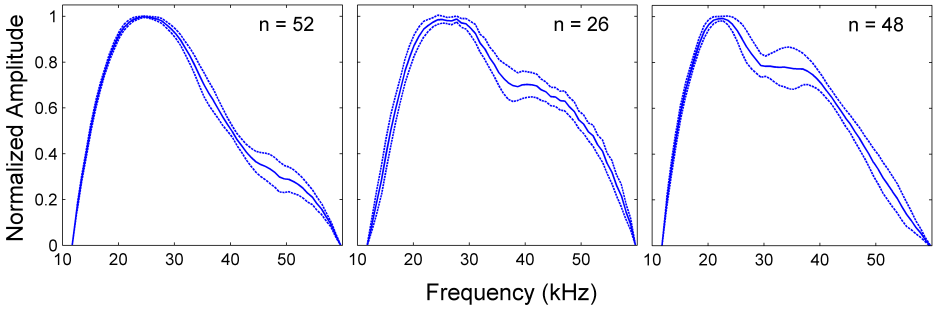


Figure 6.13: Representative rough-toothed dolphin (*Steno bredanensis*) mean echolocation click summary spectra from towed array recordings of visually verified encounters. Dark lines indicate spectral means, with light lines indicating one standard deviation above and below the mean.

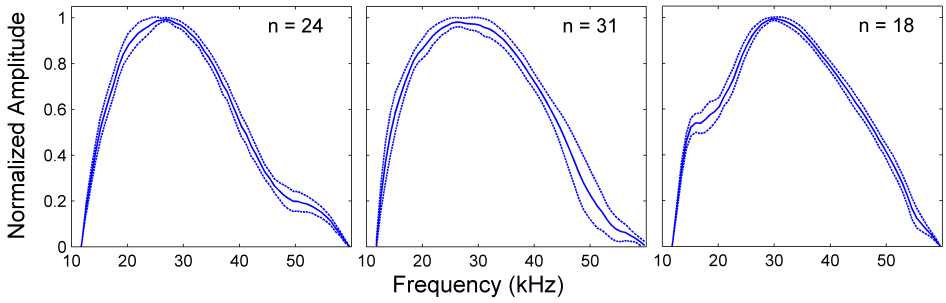


Figure 6.14: Representative bottlenose dolphin (*Tursiops truncatus*) mean echolocation click summary spectra from towed array recordings of visually verified encounters. Dark lines indicate spectral means, with light lines indicating one standard deviation above and below the mean.

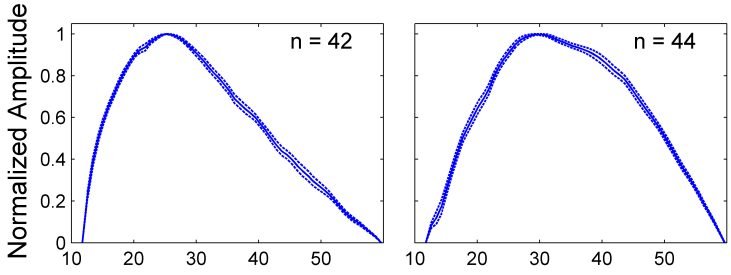


Figure 6.15: Representative Fraser's dolphin (*Lagenodelphis hosei*) mean echolocation click summary spectra from towed array recordings of visually verified encounters. Dark lines indicate spectral means, with light lines indicating one standard deviation above and below the mean.

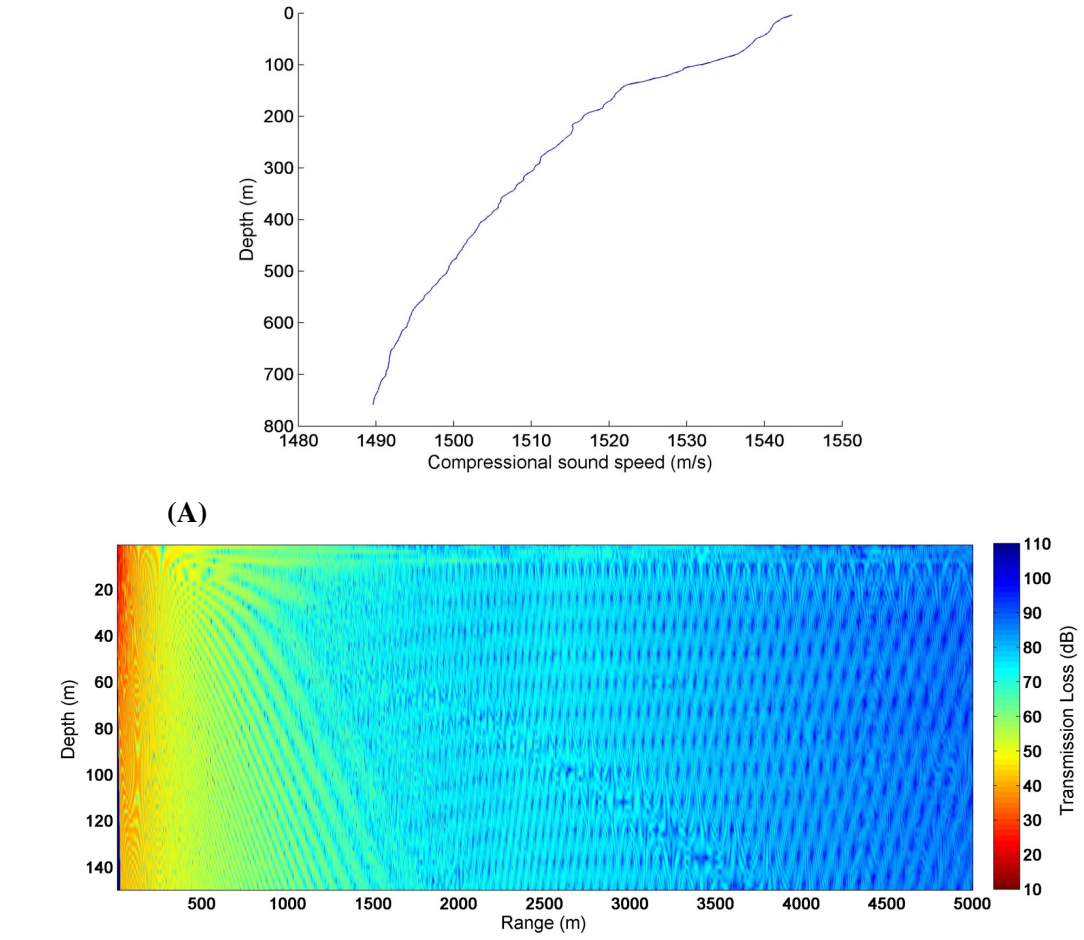


Figure 6.16: **A**. Typical Gulf of Mexico summer sound speed profile collected in July 2011. **B**: Propagation model for a 20 kHz signal produced at a depth of 5 m based on profile. The warm surface waters lead to strong downward refraction of sound energy toward the seafloor, limiting the ranges at which dolphins can be detected using a towed array.

Chapter 7

Estimating delphinid densities in the Gulf of Mexico

7.1 Abstract

Delphinid echolocation clicks were recorded at five sites in the Gulf of Mexico (GOM) over a three year period. Cues were translated into preliminary density estimates for three species groups including Risso's dolphin (*Grampus griseus*), short-finned pilot whale (*Globicephala macrorynchus*) and Stenellid dolphins (*Stenella* spp.)

Two different density estimation methods, click-counting and group-counting were used. Cue counts were converted into density estimates by factoring in species-specific cue rates, vocalization probabilities, and seasonally-adjusted click detection probabilities. Group detections, defined as five-minute time windows containing detections, were converted into density estimates by factoring in species-specific group vocalization probability, group sizes, and seasonally-adjusted group detection probabilities. Group-based density estimates are higher than click-based estimates, but trends are similar between the two. Further refinement of inputs including group size and

vocal activity is expected to increase agreement between the two methods.

Time series of weekly density estimates using both methods are provided for three classifiable species at the three slope sites and one shelf site. Seasonal cycles are apparent for some species, primarily at northern GOM sites. Longer term trends for deseasoned time series are evaluated using a Theil-Sen linear slope estimator.

7.2 Background

Cue counts alone are not a reliable tool for quantifying the presence of delphinids at acoustically monitored sites. Numerous factors influence the number of cues detected at a monitoring site and the relationship between that number and the estimated number of animals present (Buckland et al., 2001, 2007). Detection probabilities, cue rates, and sampling effort must be accounted for in order to convert cue counts into an estimate of local animal density (e.g. Helble et al., 2013a). By incorporating these types of information we can estimate local animal densities from passive acoustic data, while accounting for environmental and behavioral parameters and variability.

7.2.1 Density Estimation Through Distance Sampling

Stationary seafloor sensors observe a point in space and the area around that point. This is known as a point transect (Buckland et al., 2001, 2007). The monitored area extends out to a maximum detection radius ω, beyond which no cues are detected. The probability of detecting a cue typically decreases as the distance between the sensor and the sound source increases. In acoustic applications, this decrease in detection probability is primarily attributable to acoustic transmission loss whereby the amplitude of a signal decreases as it travels through the water column.

Density estimation methods for marine mammals using fixed passive acoustic

sensors have been described by (Marques et al., 2009). First, the number of cues detected are adjusted to a total estimated number of clicks within range ω using the principles of distance sampling for the point transect case. False-positive and false-negative rates are used to further adjust the cue counts.Counts are then converted into an estimate of the average number of animals present per unit area by factoring in variables such as cue rate and detection range.

Density estimation using distance sampling relies on a number of assumptions (Marques et al., 2013):

- On average, animals are uniformly distributed in the area around the sensor.
- Detections are statistically independent events.
- Detection ranges are measured without errors.
- Each detection is a snapshot, considered instantaneous in time.

Two main approaches have been explored for passive acoustic density estimation (Marques et al., 2013). One method uses each cue as the basis for density estimation (cue-counting), while the other looks at windows of time, for a group or snapshot method (group-counting). These methods rely on a different basic metric, but the goal of estimating local numbers of animals per unit area remains the same.

7.2.2 Cue-Counting Methods

Cue-based density estimation methods use individual click counts as the basic unit for estimating local density over time. These methods have been used for beaked whale density estimates (Hildebrand et al., 2015; Küsel et al., 2011; Marques et al., 2009). In those cases, as in the case described here, the cue is a single echolocation click.

For site k, and a unit of time t, the number of cues detected at that site, during that time is n_{kt}. This cue count is adjusted by incorporating $\hat{c}_k$, an estimate of detector's false positive rate. Site-specific density D can be estimated using the formula

$$\hat{D}_{kt} = \frac{n_{kt}\,(1-\hat{c}_k)}{\pi\,\omega^2\,\hat{P}_k\,T_{kt}\,r} \tag{7.1}$$

where T_{kt} is the number of time intervals sampled, P_k is the probability of a detecting a vocal cue produced within the radius ω from the site, and r is the cue production rate. Some of these variables, including false positive rate and detection probability, are unique to each site. Cue rate is expected vary between species of interest, and to depend on behavior. All of the input variables should be as specific to the sites, species, and time periods of interest as possible.

7.2.3 Group-Counting Methods

The group-based density estimation methods applied here use a unit of time in which cues were, or were not detected, as the basic unit for density estimation (Marques et al., 2013). This method has been used for beaked whale density estimation (Hildebrand et al., 2015, *submitted*). Rather than counting individual cues, the presence or absence of cues is used to determine whether or not a group of animals was present during a specific time bin. The assumption of instantaneity is more tenuous with this approach, because animals must be assumed not to move during the time window used. A short time window or bin is therefore preferable. The group-counting approach has the advantage of being more robust to cue rate variability than the click-counting approach, but relies on an estimate of group size.

The estimated density $\hat{D}_k$ of animals at site k during a period of time t is given by

$$\hat{D}_{kt} = \frac{n_{kt}(1-c_k)\,\hat{g_k}}{\pi\,\omega^2\,\hat{P}_k\,\hat{P}_{va}\,T_{kt}} \tag{7.2}$$

where n_{kt} is number of positive bins, c_k is the proportion of false positive bins. An estimated group size $\hat{g_k}$ is used to convert from positive bins to a number of animals, and the probability that a group is acoustically active $\hat{P}_{va}$ replaces the individual cue rate in equation 7.1. As for the cue counting method, ω is the maximum detection range of calls, $\hat{P}_k$ is the probability of a detecting a vocalizing group within radius ω, and T_{kt} is the number of time intervals sampled at site k during time period t.

7.2.4 Goals of this Chapter

The goal of this chapter is to generate preliminary estimates of site-specific densities for Risso's dolphin (*Grampus griseus*), Stenellid dolphins (*Stenella* spp.), and short-finned pilot whales (*Globicephala macrorynchus*) on a weekly basis, throughout the three-year monitoring period. Density time series are analyzed for seasonality and long term trends.

Two density estimation approaches, cue and group counting are implemented for comparison. Density estimates are based on cue and bin counts from Chapter 4, which are adjusted using detection probabilities and radii from Chapter 2, and with cue rates from Chapters 3 and 5, and group sizes from the literature. Click-based species classifications are based on Chapters 5 and 6. Cumulative error rates are estimated by assembling the known variances across the various inputs.

Preliminary local density estimates are reported, however further refinements are needed to reconcile estimates obtained using the two methods. Limitations, data gaps, and future improvements are discussed.

7.3 Methods

Only nocturnal periods between local sunset and local sunrise were used for density estimation, so as to simplify behavioral assumptions. The dominant behavior during nocturnal periods was assumed to be foraging. Three delphinid types could be identified from echolocation clicks in the passive acoustic recording data. These included Risso's dolphin (*Grampus griseus*), Stenellid dolphins (*Stenella* spp.), and short-finned pilot whales (*Globicephala macrorynchus*). Detections were translated into density estimates for the three slope sites MC, GC, and DT, and the deepest shelf site DC. Density estimates were not computed for the shallowest shelf site MP because the two species known to inhabit that area, bottlenose dolphins (*Tursiops truncatus*) and Atlantic spotted dolphins (*Stenella frontalis*) (Fulling et al., 2003), could not be reliably distinguished.

Click counts, false positive rates and dead time adjustments were obtained from the analyses described in Chapter 4. Click types associated with the three species of interest were determined based on the analyses described in Chapters 5 and 6.

7.3.1 Detection Probability Estimation

Detection probabilities $\hat{P}_k$ were obtained for each site from the Monte Carlo simulation described in Chapter 2. Both the probability of detecting a click, and the probability of detecting a group were computed (Tables 2.4 and 2.8). Deep diver model estimates were used for pilot whale and Risso's dolphin calculations. Shallow diver model estimates were used for Stenellid dolphins. January detection probability estimates were used for the months of November through March. July detection probability estimates were used for the months of April through October.

7.3.2 Cue Rate Estimates

Click Method

Individual cue rates r are a function of the probability that an animal is actively echolocating P_v, and animal's click rate when vocally active c_r. The mean and variance of c_r for each species and site was obtained from the inverse of the species and site-specific peak inter-click intervals (ICI) in Chapter 5. For each species, a mean and variance for I_p was computed by repeatedly selecting a subset of 5% of the ICIs obtained for that species' click type. The subset was binned into 50 millisecond intervals between 0.01 and 0.3 seconds, and the peak ICI for the subset was defined as the most populated bin. This process was repeated 100 times, and the mean and standard deviation were computed for the resulting set of peaks. The minimum allowed ICI was increased to 0.04 seconds at site DC due to false positives which interfered with ICI calculation.

The second input, P_v, was estimated in Chapter 3 by comparing the expected number of clicks per bin to the actual number of clicks per bin for localized encounters (similar to Van Parijs et al., 2002). Currently, there is little or no further vocalization probability data for delphinids in the literature.

Group Method

No data is currently available on group vocalization probability $\hat{P}_{va}$ for delphinids. Since only nocturnal detections are considered, the dominant behavior across all groups is assumed to be foraging, and therefore echolocation probability is expected to be high. In order to achieve a conservative density estimate from the group method, vocalization probability at night is assumed to be 100%, i.e. at least one animal in a group will echolocate at some point in a five minute interval.

7.3.3 Group Size Estimates

Mean species-specific group sizes (Table 7.1) were obtained from Mullin and Fulling (2004). Mullin and Fulling's estimates were based on visual survey data from the Gulf of Mexico (GOM; Mullin and Fulling, 2004).

Group size estimates ($\hat{g}$) for eastern and western slopes were combined, and combined variances were computed using the delta method:

$$var(\hat{g}) = \frac{1}{n^2} \sum_{i=1}^{n} var(\hat{g}_i) \tag{7.3}$$

where $\hat{g}_i$ is each individual estimate $i = 1...n$ and n is the total number of estimates being combined (Powell, 2007). For Stenellid dolphins, group size estimates for all GOM Stenellid species were combined using this method.

Table 7.1: Estimated group sizes from (Mullin and Fulling, 2004)

Species	**N**	**Mean**	**CV**
Stenella spp.	110	60.7	0.22
Globicephala spp.	6	34.2	0.32
Grampus griseus	27	7.0	0.14

7.3.4 Classification Errors

Classification error rates are not currently known. Labeled click detections, are needed to estimate these error rates. Classification error rates are not incorporated here, however understanding classification errors is critical for accurate estimation, particularly for rarer species (Caillat et al., 2013). Classification error rates likely differ between species, and between methods. See discussion for further analysis.

7.3.5 Cumulative Error Estimates

An approximation of the delta method (Seber, 1982; Powell, 2007) was used to estimate cumulative error for each density estimate (Marques et al., 2009):

$$\hat{var}(\hat{D}_{kt}) = \hat{D}_{kt}^2\{CV^2(\hat{n}_{kt}) + CV^2(\hat{c}_k) + CV^2(\hat{s}) + CV^2(\hat{P}_v) + CV^2(\hat{P}_k)\} \tag{7.4}$$

where $CV(x)$ is the coefficient of variation of the random quantity x, (i.e., the standard error of the estimate of x divided by the estimate itself).

7.3.6 Trend analysis

Trend analysis was done in R (R Core Team, 2015), using the 'openair' package for time series analysis (Carslaw and Ropkins, 2012, 2015). Each data point in the fitted time series is a monthly average of daily density estimates. Interpolated time series data for each species at each site were deseasonalized using a seasonal trend decomposition procedure (STL; Cleveland et al., 1990). First, monthly mean densities are computed, and linear interpolation is used to fill in any gaps in the time series. The time series (Y_v) is then decomposed into seasonal (S_v), trend (T_v), and remainder (R_v) such that

$$Y_v = T_v + S_v + R_v \tag{7.5}$$

where v = 1 to N.

The STL process involves repeatedly running the data through two recursive loops. In each pass of the inner loop, the seasonal and trend smooths are updated. In the outer loop, robustness weights are computed, after the inner loop has completed, and these weights are used to improve the seasonal and trend components in the next inner loop pass. The inner loop consists of six steps that run for k iterations:

1. A detrended time series is computed as $P_v = Y_v - T_v$, where T_v is initialized to zero on the first pass.

2. P_v reorganized as a cycle-subseries C_v on a 12-month scale, and smoothed using a one-dimensional locally weighted regression (loess) filter (Cleveland and Devlin, 1988).

3. A low pass filter is applied to C_v to generate the filtered cycle-subseries L_v.

4. The seasonal component S_v is computed as $C_v - Lv$, to remove low frequency signals from the seasonal component.

5. A deseasonalized time series T_v is computed as $Y_v - S_v$.

6. The deseasonalized time series is smoothed using a loess filter.

In the outer loop, the remainder R_v is computed using the inner-loop outputs as

$$R_v = Y_v - S_v - T_v$$

Robustness weights ρ_v are computed for each time point, such that weights are small if R_v is large. The weights are then factored in to the loess filter in steps 2 and 6 on the next iteration of the inner loop.

Theil-Sen slope estimates for linear trends in the time series (Sen, 1968; Thiel, 1950), were computed for the deseasoned data. The Theil-Sen method estimates the slope of a dataset (**x**,**y**) by computing the median slope $\frac{y_j - y_i}{x_j - x_i}$ between all pairs of points (i, j) in the set.

A 95% confidence interval for the slope is given by the interval containing the middle 95% of the pairwise slope estimates. The Theil-Sen approach was chosen because,

unlike simple linear regression, it is insensitive to outliers and is accurate for skewed data.

7.4 Results

7.4.1 Accounting for Detector Dead Time

Detector dead time adjustments had relatively large effects on counts for the shelf site DC. Total adjusted counts were over 4% higher than original counts at this site for Stenellid dolphins and over 2% higher Risso's (Table 7.2). This was due to the long lockout period implemented for detection purposes in shallow conditions (Figure 4.15). Slope site counts were negligibly affected, due to the much shorter lockout period used there, with increases on the order of a few hundredths of a percent for all slope sites and species.

7.4.2 Cue rate estimates

Mean active click rate estimates (r) based on ICI were consistent across sites for Stenellid dolphins and pilot whales, and more varied for Risso's dolphins (Tables 7.4, 7.6 and 7.8). Stenellid dolphins were estimated to produce approximately 15 clicks per second per individual when actively clicking, at all slope sites, and 14 clicks per second at site DC, based on classified clicks. Variability around the mean was low in all cases. Pilot whales were estimated to produce seven clicks per second on average, at slope sites. Risso's dolphin mean active click rates were highest at site MC, at eight clicks per animal per second and lowest at sites DC and DT, at roughly five clicks per second. Variance was low at all sites.

Table 7.2: Effect of detector dead time adjustment on total counts.

Site	Species	Total Counts	Adjusted Total Counts	% Increase
MC	Stenellid dolphin	7,241,631	7,244,434	0.04
	Pilot whale	654,037	654,229	0.03
	Risso's dolphin	341,860	341,920	0.02
GC	Stenellid dolphin	2,514,023	2,514,473	0.02
	Pilot whale	600,752	600,933	0.03
	Risso's dolphin	44,766	44,781	0.03
DT	Stenellid dolphin	4,284,210	4,284,963	0.02
	Pilot whale	278,658	278,739	0.03
	Risso's dolphin	486,500	486,529	0.01
DC	Stenellid dolphin	5,512,109	5,772,245	4.72
	Risso's dolphin	134,339	138,297	2.95

7.4.3 Group Size Estimates

Group sizes ($\hat{g}$) are predicted to be largest for Stenellid dolphins, with a mean of 60.7 animals per group, and lowest for Risso's dolphins at an expected mean of seven animals per group (Tables 7.5, 7.7 and 7.9). Group size variability is fairly high for Stenellid dolphins, reflecting what is typically seen in the field. Variability is also high for pilot whales due to at least in part to small sample size.

7.4.4 Density estimates

Stenellid dolphins

Mean group-based Stenellid dolphin density estimates across the entire period were between six and ten times larger at all sites than click-based density estimates

(Tables 7.4 and 7.5). CVs were approximately three times larger for the click-based method than for the group-based method.

Stenellid dolphin densities were higher on average at site MC than at the neighboring northern slope site GC. Daily and weekly density estimates from both methods increased in summer and decreased in winter (Figure 7.1), although this seasonal shift appeared to break down somewhat during the first half of the monitoring period. Stenellid densities were highest at site MC in summer between April and September (Figure 7.2). Densities declined rapidly in fall, remained low through December, and then began to ramp up beginning in January. Theil-Sen slope estimates indicated a significant increase in Stenellid dolphin densities at site MC over the three year period, with a rate of increase of 81 animals per 1000 km^2/year using the group method and seven animals per 1000 km^2/year according to the click method estimates (Figure 7.3).

Mean Stenellid dolphin densities were two to three times lower at site GC than at site MC depending on the density estimation method used (Tables 7.4 and 7.5). A regular, seasonal density shift was seen at site GC (Figure 7.10), similar to the pattern at site MC. Average densities were higher in the spring and summer months, between March and September, and lower in winter, between October and January (Figure 7.11). Theil-Sen slope estimates indicated no significant change in Stenellid densities at site GC over the three year monitoring period (Figure 7.12).

Mean Stenellid dolphin densities at site DT were comparable to densities at the northern slope site MC (Tables 7.4 and 7.5). A weak seasonal density shift was seen at this site, with the same overall pattern of higher densities in summer (March through August) as the other two slope sites (Figure 7.20). However, fewer total days were monitored at site DT than the other slope sites, due to longer servicing intervals. Only one full winter season was monitored at site DT without interruption, therefore seasonal trends may not be accurate. A reduction in estimated density at site DT is apparent in

the time series between the end of 2011 through 2012 (Figure 7.19). Theil-Sen slope estimates indicated no significant change in Stenellid densities at site DT over the three year monitoring period (Figure 7.21).

At the shelf site DC, mean Stenellid densities across the monitoring period were comparable with densities at slope site GC (Tables 7.4 and 7.5). Both density estimation methods suggest year-round presence with an increase in summer, between July and September (Figure 7.29). This seasonal pattern is weak in the first half of the time series, becoming more distinct in the second half (Figure 7.28). Theil-Sen slope estimates indicated a significant increase in Stenellid dolphin densities at site DC over the three year period, with rates of increase of 71 animals per 1000 km^2/year based on group method estimates, and nearly 10 animals per 1000 km^2/year according to the click method estimates (Figure 7.30).

Pilot whales

Mean pilot whale densities were similar across all three slope sites (Tables 7.6 and 7.7). Pilot whales were not detected at shelf site DC. The group model average predicted densities were five to eight times higher than the click-based density estimates. CVs are approximately two times larger for click-based estimates than group-based estimates. Both estimates place pilot whale densities far below Stenellid dolphin densities.

Mean pilot whale densities were very similar at sites MC and GC, with estimates between two and 10 animals per 1000 km^2 depending on the method used. A seasonal shift, with higher densities between March and August was seen at both sites (Figures 7.5 and 7.14). This seasonal shift grew more pronounced over the course of the monitoring period at site MC (Figure 7.4). At site GC, the seasonal increase was consistent during the first three summers monitored, but was relatively weak in the summer of 2013 (Figure 7.13).

Theil-Sen slope estimates indicated a significant increase in pilot whale densities at site MC over the three year period, with rates of increase between 1.5 and 6.4 animals per 1000 km^2/year depending on the density estimation method (Figure 7.6). In contrast, a small decrease in pilot whale densities, on the order of one to two animals per 1000 km^2/year was seen at site GC (Figure 7.15). However the decrease was only significant according to the click-based density estimates (Table 7.3).

Click-based estimates put mean pilot whale densities lower at site DT than at the two northern slope sites. This was difference was not seen in the group-based estimates (Tables 7.6 and 7.7). This is likely because pilot whale clicks were detected in low numbers throughout the year, and this low-level presence was scaled up by the group-based method more than the click-based method. No seasonal trend was seen in the group-based density estimates, while click based estimates predicted small peaks in March and October (Figure 7.23). As previously mentioned, gaps in the time series at this site limit the identification of seasonal trends.

No significant change in pilot whale densities was seen at site DT over the course of the monitoring period (Figure 7.24). Pilot whale densities decreased during 2012, relative to the periods before and after (Figure 7.22). (Recall that Stenellid dolphin densities decreased at this site during the same period). Daily and weekly density estimates from the click-based method at site DT are more variable and wider ranging than those obtained from the group method.

Risso's dolphins

Group and click-based density estimates of mean Rissos' dolphin density were similar within sites indicating overall agreement between the models (Tables 7.8 and 7.9). Risso's dolphin presence was seasonal at all sites. At site MC, they were predominantly detected between March and June, when densities peaked at approximately four animals

per 1000 km^2 (Figure 7.8). Theil-Sen slope estimates indicated a significant decrease in click-based Risso's dolphin density estimates of approximately -0.4 animals per 1000 km^2 per year, or -22%, over the three year monitoring period (Figure 7.9). No significant change was found in the group-based density estimates at this site.

Mean Risso's dolphin densities were lowest at site GC. Average densities rose to between 0.5 and 1 animal per km^2 between March and August, and then decreased to nearly zero in other months (Figure 7.17). The click-based method indicated a significant decrease in abundance over the monitoring period of -0.2 animals per year or approximately -29% (Figure 7.18). No significant change was found in the group-based density estimates at this site.

Mean Risso's dolphin densities were highest at site DT, but presence there was highly intermittent, and the average is not a good indicator of expected densities at any given time (Figure 7.25). Risso's dolphin densities spiked in November, at this site, briefly reaching between 30 and 60 animals per 1000 km^2 (Figure 7.26). A smaller influx of animals was seen in June and July. Outside these periods, Risso's dolphin densities were near zero. Both density estimation methods indicated a small increase in Risso's dolphin densities at this site over the course of the monitoring period, however it was only significant according to group-based estimates, which put the increase at roughly 0.9 animals per 1000 km^2/year (Figure 7.27). Due to gaps in the time series and high density variability at this site, the long term trend is not particularly robust.

Table 7.3: TheilSen slope estimates for trends in deseasoned time series by site and species. Slopes are listed as rate of change in number of animals per year ("slope"), and as percent change ("% change"). Significant rates of change are in **bold**. P-values and slopes are calculated through bootstrap simulations in the *R* package *openair* (Carslaw and Ropkins, 2012, 2015).

Site	**Method**	**Stenellid dolphin**			**Pilot whale**			**Risso's dolphin**		
		slope	% change	p-value	slope	% change	p-value	slope	% change	p-value
MC	group	**+81.8**	**40%**	**0.00**	**+6.4**	**68%**	**0.00**	-0.1	-3%	0.85
MC	click	**+7.1**	**23%**	**0.00**	**+1.5**	**42%**	**0.00**	**-0.4**	**-22%**	**0.00**
GC	group	+6.4	3%	0.42	**-2.4**	**-15%**	**0.03**	-0.1	-14%	0.20
GC	click	**+2.6**	**20%**	**0.00**	**-1.3**	**-26%**	**0.00**	**-0.2**	**-29%**	**0.00**
DT	group	-25.4	-5%	0.67	-2.7	-24%	0.13	**+0.9**	**22%**	0.00
DT	click	-17.3	-25%	0.07	**-0.5**	**-30%**	0.04	**+0.3**	**9%**	**0.03**
DC	group	**+71.0**	**79%**	0.00		n/a		-0.0	0%	0.89
DC	click	**+9.7**	**42%**	0.00		n/a		-0.1	-10%	0.54

7.5 Discussion

7.5.1 Density Estimation Input Variables

Six main variables were needed for each of the density estimation methods, and each fell along a spectrum of robustness. Click and bin counts are robust at the level of a detection, but lack of information on classification error rates is a serious issue. Classification errors can affect density estimates by assigning a portion of signals to the wrong species. The effect on more more abundant species is less drastic than that on rarer species, for which density estimates vary wildly in response to small changes in classification error rates. No classifier is perfect, therefore misclassifications are undoubtedly having an effect on the density estimates presented here. The magnitude of that effect is not yet determined. The collection of labeled seafloor sensor recording data on which to run the classifier, in order to evaluate error rates, remains a priority.

A second issue relating to classification is the question of how to deal with time bins in which multiple species have been identified. Here, if both click type B1 and B2 were identified in a bin, for example, the bin was counted only once, i.e. one group of type B was present. If both click type B1 and type E2 were identified, the bin was counted twice, i.e. one group of type B and one group of type E were present. However, since mixed species groups are common in the GOM, and some species are particularly well known for foraging together, further refinements might be necessary. Likewise, if type B1 and B2 are later determined to be associated two different species of Stenellid dolphin, the question of how to handle cases multi-species bins will need to be revisited.

Detection probability model predictions for the click method were ground-truthed with the available data (Chapter 3), although pilot whales were not encountered in the groundtruth recordings. More verification is needed to determine how well the model predicts group detectability. This would likely require a system of multiple,

simultaneously recording, multi-channel instruments, deployed for long enough to track multiple encounters with each species of interest. Multiple instruments are required in order to resolve the question of whether a group is undetectable at on a given sensor because too distant, or because no group is present in the area.

Group detectability likely varies as a function of group size, and group sizes may or may not be normally distributed around a mean. The group sizes used in these calculations are fairly robust given the amount of visual survey data that went into Mullin and Fulling's estimates (Mullin, 2007). One consideration is that all group size information was collected during summer cruises in daylight hours. If group size varies by time of day, or seasonally, this variabilty would not be reflected in the group size estimates.

A second group size issue is the possible difference between the visual and acoustic definition of a group. For acoustic purposes, a group is the total number of animals of one species within maximum detection radius ω at one time. The visual definition of a group from Mullin and Fulling may be integrated across more space and time than appropriate for an acoustic estimate. The main viewing platform from which group sizes were estimated visually in the Mullin and Fulling study had a height (c) of 14.5 m.

If the earth is assumed to be a perfect sphere with radius r, The maximum visual v range from this vantage point is computed as

$$v = \sqrt{(r+c)^2 - r^2} \tag{7.6}$$

Approximating the radius of the earth at 6378 km, v equals 13.6 km, for a total viewing area of 580 km^2. In contrast, our maximum click detection radius is estimated at 5 km, for a total monitoring area of 79 km^2. Because of the larger area monitored, visual group

size estimates may be high compared to acoustic group size estimates. Visual estimates may also be biased toward larger group sizes, because larger groups are more easily spotted than smaller groups. In contrast, some visual surveys are very granular in their definition of groups, and may identify multiple groups of the same species in the vicinity of a vessel.

At this time, there is does not seem to be an obvious solution to the acoustic group size issue. For species that typically travel in large groups, including Stenellid dolphins and other small delphinids, acoustic tracking of individual animals for counting purposes would require very high spatial resolution, because the animals seem to travel in tight clusters.

Group size is also difficult to deconvolve from vocal activity. Given robust cue rate data, it might be possible to estimate group size from click counts, however cue rates are currently tenuous. The active click rate portion of cue rate is fairly robust and based on a large amount of ICI data. The probability that an animal is actively vocal remains obscure. Individual cue rates may vary as a function of group size (e.g. Götz et al., 2006), behavior and/or environment.

The method used in Chapter 3 to estimate individual probability of vocal activity is not ideal, because it is dependent on estimates of other inputs such as click rate and detectability and therefore suffers from errors in those estimates. Risso's and Stenellid dolphin estimates of vocal activity using this method are comparable to estimates for beaked whales in the region (Hildebrand et al., 2015), however further refinements are necessary. Acoustic tag data are likely the best option for resolving the cue rate question (e.g Marques et al., 2009), however delphinids are difficult and expensive to tag, and few successful examples currently exist. A relatively large number of samples across behaviors, time periods, and locations will be needed to comprehensively address this question. Nonetheless, forthcoming work by other research groups may address this data

gap in the near future.

In these calculations groups were assumed to be vocally active 100% of the time. That is, at least one animal in the group is actively clicking during the period of a bin. Since only nocturnal data were used for density estimation purposes, and groups were likely foraging, this assumption is reasonable for now but not ideal.

7.5.2 Density Estimation Assumptions

The assumptions required for density estimation based on distance sampling methods were mentioned previously. The first was that animals are, on average, uniformly distributed around each sensor. Without extensive localization data, this assumption cannot be verified for this dataset. However, we currently have no reason to expect a non-uniform distribution of animals within the small area in which echolocation clicks were detected.

The second assumption was that detections are statistically independent events, i.e. given that one click was detected, we are no more or less likely to detect another. This assumption is violated in the group-based method because a minimum number of clicks are required in a time bin, before that click is deemed positive. This will lead to a small negative bias in our group-based density estimates. Given a perfect detector, this minimum number of detections would be unnecessary, however in practice, removing this threshold would make the group-based method highly sensitive to false positives in the click detection data. Similarly, in the click-based method, isolated clicks were pruned from the dataset because they could not be classified. This will also lead to a slight negative bias in density estimates, however the magnitude is expected to be very small, because individual clicks have much less significance than time bins.

The third assumption was that detection ranges are measured without errors. Because we used a simulation, rather than localization, to estimate distances in this case,

this assumption has more to do with the accuracy of our model in estimating detection probabilities than to actual distances. The detection probability estimates are not perfect, therefore this assumption is violated, but the direction of resulting bias is unknown. The model accuracy can likely be improved, as more and better input parameters become available, however, the ideal solution to this violation is to be able to localize each detection, rendering the model unnecessary. Current constraints make localization-ready sensors impractical in most cases, however future work may find ways to move away from detection probability simulation.

The final assumption is that each detection represents an instant in time. The point of this assumption is that we need to be able to assume that nothing moves in or out of our monitoring area during a detection. This is fine in the case of the click-based method because each click is so short in duration that animal movement is not a concern. In the case of the group-based method, however, it is possible that a group could enter or leave the monitored area during a five-minute window. In practice, we are assuming that each group was present for the entire five minute bin in which they were detected, which will lead to an overestimation of densities using the group-based method. Because the five minute time window is relatively small, and a group is relatively large, the amount of bias this violation introduces is expected to be minimized. An estimate of the magnitude of the bias might be possible given a better understanding of group travel speeds during detectable behaviors.

7.5.3 Density Estimates

The two methods used to estimate local delphinid densities, group-counting and cue-counting, have different strengths and weaknesses. The group counting method has the advantage of being insensitive to click rate fluctuations, and generally functions as an averaging window, compared to the peaky, highly variable click method predictions.

Predicting detection probabilities is more straightforward for the cue-counting method however.

Here, two methods predict similar trends in relative density for each site and species pair over time, and similar seasonal shifts. They differ however in the magnitude of the densities they predict for Stenellid dolphins and pilot whales. This is not surprising given refinements needed for some of the input parameters as listed above.

The Risso's dolphin density estimates agree fairly well between the two approaches, and gives us an indication of where the other estimates are falling short. Unlike pilot whales and Stenellid dolphins, Risso's dolphin group sizes are small, the animals are relatively independent and dispersed in the field, and we have a fairly good estimate of vocalization probability from our tracking data (Chapter 3). Stenellid dolphins and pilot whales are highly social and are found in large groups with a wide range of behaviors.

The takeaway message from this work is that density estimation for highly social delphinids will likely require a more complex behavioral model than the one implemented here. Rather than using a static value for vocalization probability in our density estimates, we would likely have better results if we incorporated vocalization probability into the Monte Carlo simulations. Within the model framework, estimates of group spread, group size, vocalization rates and rotation could be varied as a function of behavior. As more data on these variables becomes available, we expect the Stenellid dolphin and pilot whale density estimates from the two methods to converge.

Comparing the PAM-based density estimates to existing estimates based on visual surveys provides further insight on the accuracy of our numbers. In 2009, NOAA stock assessments estimated the number of pelagic Stenellid dolphins in the GOM at around 64,000 with a CV of 0.27, based on summer visual surveys (Waring et al., 2013). If animals were uniformly distributed in the survey area, this number would predict an

average of roughly 200 animals per 1000 km^2, in summer. However, our data suggest that animals are not uniformly distributed, and some areas, like Mississippi Canyon have higher relative Stenellid densities than others. The NOAA estimates provide the sense that the Stenellid dolphin densities predicted by the group method are probably a little high, but not unreasonable, and that the cue-based estimates are likely low.

Summer stock assessments for Risso's dolphin and pilot whales estimated their numbers at approximately 2,400 for both species in 2009, with CVs of 0.57 and 0.66 respectively (Waring et al., 2013). If animals were uniformly distributed in the US GOM, one would expect approximately 8 animals per 1000 km^2 in summer. The density estimates presented here support similar numbers, however because Risso's dolphin densities are strongly seasonal, average estimated densities over a year are much lower than the stock assessment estimates.

Each sensor in this study provided a long term, detailed look at a specific site. In exchange for temporal coverage, we sacrificed the kind of spatial coverage that a visual survey can achieve. As a result, we can draw definite conclusions about what happened at the sites we monitored, but cannot extrapolate what we see there to the larger GOM ecosystem. A future goal is to move from local animal densities, to estimates of population size and total abundance, while retaining temporal resolution. This will require more sensors, distributed across a wider variety of monitoring sites (e.g. Koblitz et al., 2014), and will increase the need for automated methods to make sense of the increasing amounts of data. Mobile passive acoustic sensors may also improve our ability to estimate population sizes (e.g. Bingham et al., 2012).

7.5.4 Population Density Trends

The fundamental question underlying this work is whether dolphin population sizes in the GOM were affected by the Deepwater Horizon Oil spill event. However,

comparable pre-spill or baseline density estimates do not exist. Nonetheless, we can draw a few conclusions from these data. First, it is clear that the spring and summer months, when the spill was occurring, are the times when dolphin densities are highest in the northern GOM. Stenellid dolphins, Risso's dolphins and pilot whales were all present at the site closest to the wellhead (site MC) while surface oil was present.

In the three years since the oil spill, Stenellid dolphin and pilot whale densities have been increasing at the site MC, and Risso's dolphin densities have stayed fairly constant. Stenellid dolphin densities also increased at site DC, to the east of the spill over the same period. The rates of increase at sites MC and DC are likely too large to be explained by reproduction rates, and may instead reflect a period of avoidance of the area, followed by a return.

Pilot whale densities decreased at site GC, while they were increasing at site MC, which is consistent with possible avoidance of site MC during the early part of the monitoring period. The click-based method indicated a significant increase in Stenellid dolphin populations at site GC, but no such increase was seen in the group-based estimates, so it is unclear if densities of Stenellid dolphins were increasing at all northern sites. The support for the density increase is more significant at sites MC and DC than at site GC, and no significant change was seen at site DT, which was furthest away from the spill site.

In the first half of the three year monitoring period, seasonal cycles for all three species were weak at sites MC and DC. This may have been due to habitat variability unrelated to oil, however seasonal cycles did not appear to be disrupted during the same period at site GC, to the west of the spill site, where oil was not observed. .

7.6 Conclusion

Density estimates for delphinid species provide insight into seasonal and long term trends at specific sites. Accounting for detectability and cue rate allows time series to be compared between sites. Group-based density estimates are slightly higher than, yet comparable to density estimates based on NOAA data. Click-based density estimates are comparable to group-based estimates for Risso's dolphin, but are likely too low for the highly social species including Stenellid dolphins and pilot whales. More complex behavioral models are needed to accurately estimate densities for these species using click-based methods.

The data indicate seasonal increases in delphinid densities at most sites in spring and summer months, between April and August. Significant increases in density are seen at sites north and east of the Deepwater Horizon spill site during the monitoring period, was well as a possible interruption of seasonal cycles in the first half of the time series. Density increases are absent or less significant at sites to the south and west of the spill site, and seasonal cycles appear to be more stable.

Table 7.4: *Stenella* sp. density estimation inputs for click method. Local animal density ($\hat{D}$) at each site is computed as a number of animals/1000 km^2. The number of click counts (N_{kt}) is adjusted by false positive rate (c_k), and then divided by the number of seconds of monitoring effort (T_{kt}), the active click rate (c_r), the probability of vocalization (P_v), the area monitored ($2\pi\ \omega^2$) and the probability of detecting a click within that area (P_k). Values are site and species-specific when possible.

Site	$\hat{\mathbf{D}}$		$\mathbf{N_{kt}}$	$\mathbf{T_{kt}}$	$\mathbf{c_k}$		$\mathbf{c_r}$		$\mathbf{P_v}$		ω	$\mathbf{P_k}$	
	mean	CV	click counts	sec	mean	CV	mean	CV	mean	CV	km	mean	CV
MC	48.13	0.64	7,244,434	46,126,500	0.01	0.39	15.40	0.00	0.19	1.6	5	0.014	0.50
GC	22.95	0.69	2,514,473	42,466,200	0.02	0.26	15.40	0.00	0.19	1.6	5	0.011	0.64
DT	45.94	0.63	4,284,963	33,076,500	0.02	0.25	15.40	0.00	0.19	1.6	5	0.012	0.58
DC	37.34	0.40	5,772,245	31,477,800	0.06	0.15	14.00	0.04	0.19	1.6	5	0.022	0.36

Table 7.5: *Stenella* sp. density estimation inputs for group method. Local animal density ($\hat{D}$) at each site is computed as a number of animals/1000 km^2. The number of positive five minute bins (N_{kt}) is adjusted by false positive bin rate (c_k), and then divided by the number of bins of monitoring effort (T_{kt}), the group size ($\hat{g}$), the probability of that the group is vocally active (P_{va}), the area monitored ($2\pi\ \omega^2$) and the probability of detecting a group within that area (P_k). Values are site and species-specific when possible.

Site	$\hat{\mathbf{D}}$		$\mathbf{N_{kt}}$	$\mathbf{T_{kt}}$	$\mathbf{c_k}$		$\hat{\mathbf{g}}$		$\mathbf{P_{va}}$		ω	$\mathbf{P_k}$	
	mean	CV	positive bins	total bins	mean	CV	mean	CV	mean	CV	km	mean	CV
MC	410.15	0.62	12,729	153,755	0	0	60.7	0.59	1	0	5	0.16	0.19
GC	238.06	0.62	6,584	141,554	0	0	60.7	0.59	1	0	5	0.15	0.20
DT	537.20	0.62	11,572	110,255	0	0	60.7	0.59	1	0	5	0.15	0.21
DC	193.18	0.61	4,039	104,926	0	0	60.7	0.59	1	0	5	0.15	0.14

Table 7.6: Pilot whale density estimation inputs for click method. Local animal density ($\hat{D}$) at each site is computed as a number of animals/1000 km^2. The number of click counts (N_{kt}) is adjusted by false positive rate (c_k), and then divided by the number of seconds of monitoring effort (T_{kt}), the active click rate (c_r), the probability of vocalization (P_v), the area monitored ($2\pi\ \omega^2$) and the probability of detecting a click produced within that area (P_k). Values are site and species-specific when possible.

Site	$\hat{\mathbf{D}}$		$\mathbf{N_{kt}}$	$\mathbf{T_{kt}}$	$\mathbf{c_k}$		$\mathbf{c_r}$		$\mathbf{P_v}$		ω	$\mathbf{P_k}$	
	mean	CV	click counts	sec	mean	CV	mean	CV	mean	CV	km	mean	CV
MC	2.5	0.62	654,229	46,126,500	0.01	0.39	7.10	0.03	0.13	0.37	5	0.08	0.31
GC	2.6	0.56	600,933	42,466,200	0.02	0.26	7.00	0.03	0.13	0.37	5	0.07	0.32
DT	1.4	0.55	278,739	33,076,500	0.02	0.25	7.60	0.05	0.13	0.37	5	0.07	0.31

Table 7.7: Pilot whale density estimation inputs for group method. Local animal density ($\hat{D}$) at each site is computed as a number of animals/1000 km^2. The number of positive five minute bins (N_{kt}) is adjusted by false positive bin rate (c_k), and then divided by the number of bins of monitoring effort (T_{kt}), the group size ($\hat{g}$), the probability of that the group is vocally active (P_{va}), the area monitored ($2\pi\ \omega^2$) and the probability of detecting a group within that area (P_k). Values are site and species-specific when possible.

Site	$\hat{\mathbf{D}}$		$\mathbf{N_{kt}}$	$\mathbf{T_{kt}}$	$\mathbf{c_k}$		$\hat{\mathbf{g}}$		$\mathbf{P_{va}}$		ω	$\mathbf{P_k}$	
	mean	CV	positive bins	total bins	mean	CV	mean	CV	mean	CV	km	mean	CV
MC	10.9	0.35	1,582	153,755	0	0	34.2	0.32	1	0	5	0.41	0.15
GC	11.8	0.36	1,586	141,554	0	0	34.2	0.32	1	0	5	0.41	0.16
DT	11.5	0.35	1,216	110,255	0	0	34.2	0.32	1	0	5	0.42	0.15

Table 7.8: Risso's dolphin density estimation inputs for click method. Local animal density ($\hat{D}$) at each site is computed as a number of animals/1000 km^2. The number of click counts (N_{kt}) is adjusted by false positive rate (c_k), and then divided by the number of seconds of monitoring effort (T_{kt}), the active click rate (c_r), the probability of vocalization (P_v), the area monitored ($2\pi\ \omega^2$) and the probability of detecting a click produced within that area (P_k). Values are site and species-specific when possible.

Site	$\hat{\mathbf{D}}$		$\mathbf{N_{kt}}$	$\mathbf{T_{kt}}$	$\mathbf{c_k}$		$\mathbf{c_r}$		$\mathbf{P_v}$		ω	$\mathbf{P_k}$	
	mean	CV	click counts	sec	mean	CV	mean	CV	mean	CV	km	mean	CV
MC	1.162	0.626	341,920	46,126,500	0.01	0.39	8.00	0.06	0.13	0.37	5	0.08	0.31
GC	0.204	0.556	44,781	42,466,200	0.02	0.26	6.70	0.05	0.13	0.37	5	0.07	0.32
DT	3.526	0.546	486,529	33,076,500	0.02	0.25	5.40	0.04	0.13	0.37	5	0.07	0.31
DC	0.950	0.482	138,297	31,477,800	0.06	0.15	5.30	0.06	0.13	0.37	5	0.08	0.26

Table 7.9: Risso's dolphin density estimation inputs for group method. Local animal density ($\hat{D}$) at each site is computed as a number of animals/1000 km^2. The number of positive five minute bins (N_{kt}) is adjusted by false positive bin rate (c_k), and then divided by the number of bins of monitoring effort (T_{kt}), the group size ($\hat{g}$), the probability of that the group is vocally active (P_{va}), the area monitored ($2\pi\ \omega^2$) and the probability of detecting a group within that area (P_k). Values are site and species-specific when possible.

Site	$\hat{\mathbf{D}}$		$\mathbf{N_{kt}}$	$\mathbf{T_{kt}}$	$\mathbf{c_k}$		$\hat{\mathbf{g}}$		$\mathbf{P_{va}}$		ω	$\mathbf{P_k}$	
	mean	CV	positive bins	total bins	mean	CV	mean	CV	mean	CV	km	mean	CV
MC	2.334	0.205	1,655	153,755	0	0	7.00	0.14	1	0	5	0.41	0.15
GC	0.349	0.214	229	141,554	0	0	7.00	0.14	1	0	5	0.41	0.16
DT	6.782	0.205	3,507	110,255	0	0	7.00	0.14	1	0	5	0.42	0.15
DC	0.903	0.173	322	104,926	0	0	7.00	0.14	1	0	5	0.30	0.10

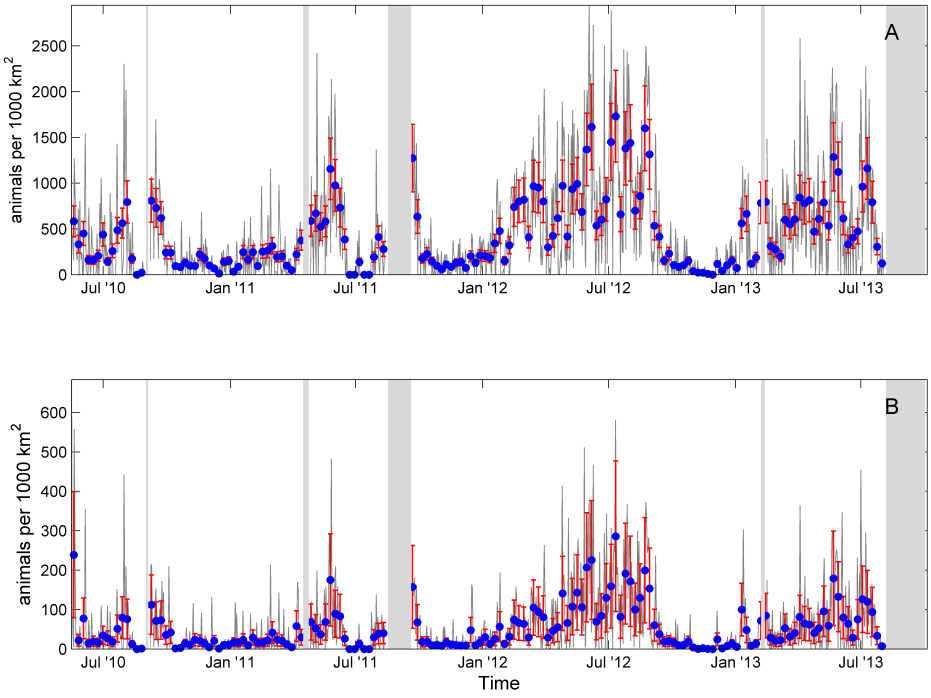


Figure 7.1: Density estimates for Stenellid dolphins at site MC. **A**: Group-method results. **B**: Click method results. Grey line indicates daily density estimates, blue squares are weekly density estimates with red error bars indicating standard error. Grey rectangles indicate gaps in recording effort.

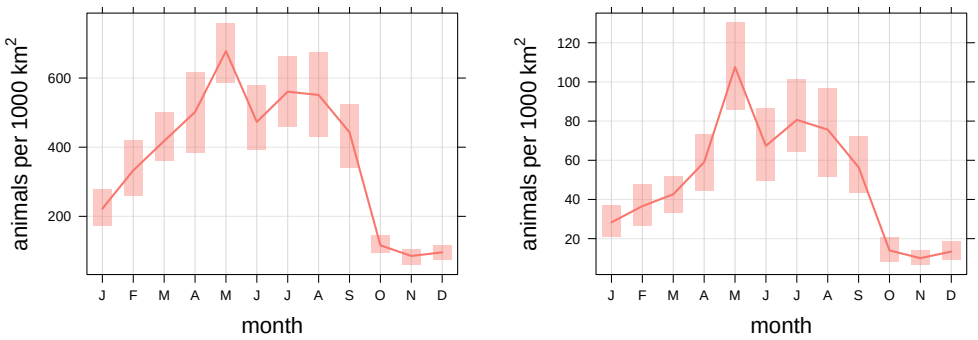


Figure 7.2: Mean density estimates and 95% confidence intervals of the mean by month for Stenellid dolphins at site MC. ***Left***: Group method estimates; ***Right***: Click method estimates.

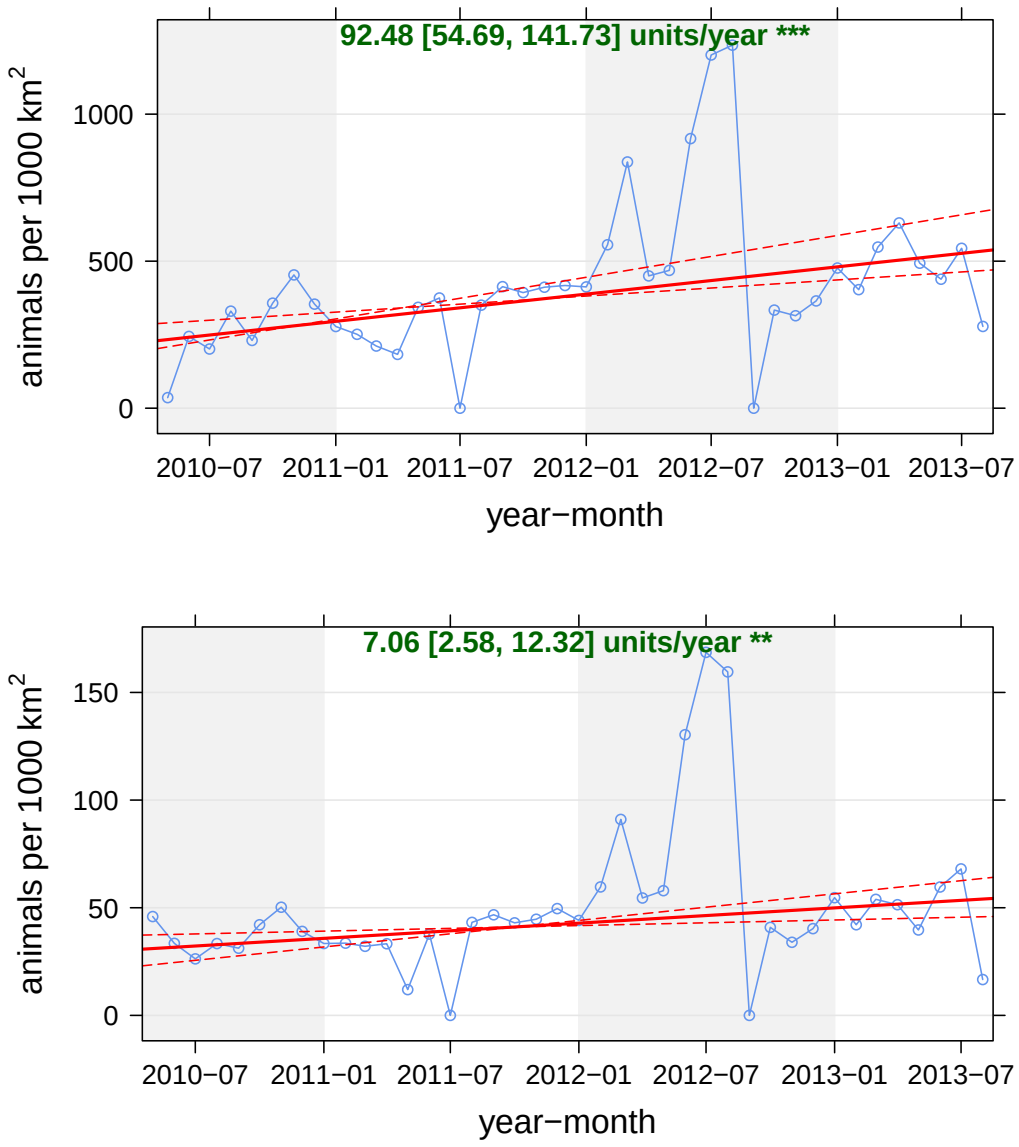


Figure 7.3: Theil-Sen slope estimates (solid red line) for deseasonalized Stenellid dolphin density time series at site MC, with 95% confidence intervals (dashed red lines). Blue dots are monthly averages of daily density estimates. ***Top***: Group method time series; ***Bottom***: Click method time series.

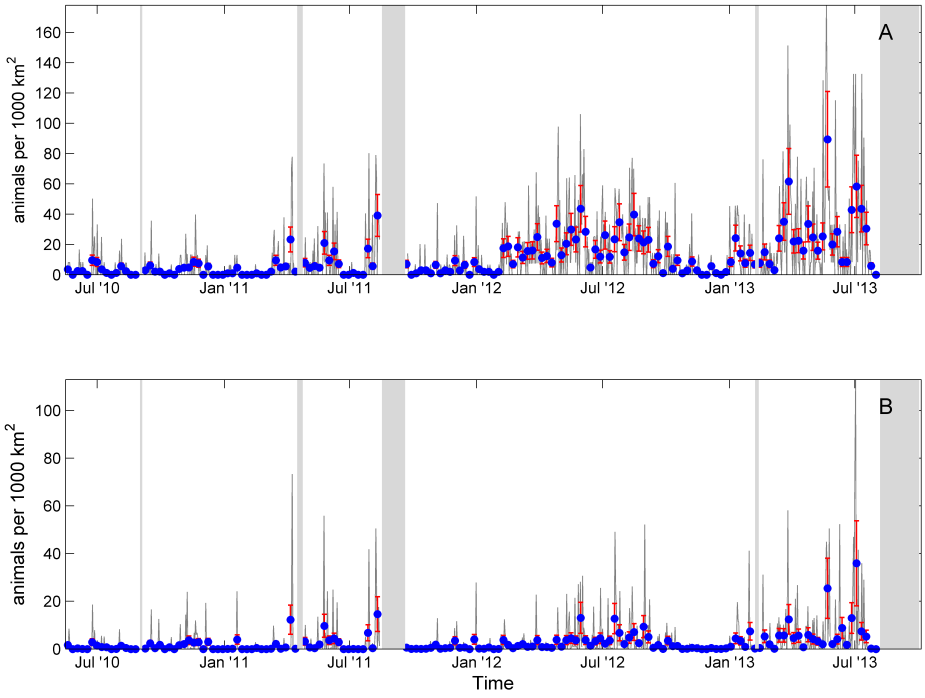


Figure 7.4: Density estimates for pilot whales at site MC. **A**: Group-method results. **B**: Click method results. Grey line indicates daily density estimates, blue squares are weekly density estimates with red error bars indicating standard error. Grey rectangles indicate gaps in recording effort.

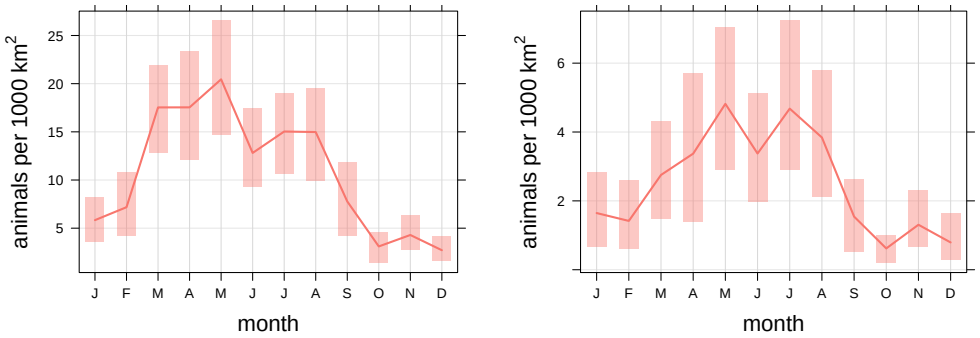


Figure 7.5: Mean density estimates and 95% confidence intervals of the mean by month for pilot whales dolphins at site MC. ***Left***: Group method estimates; ***Right***: Click method estimates.

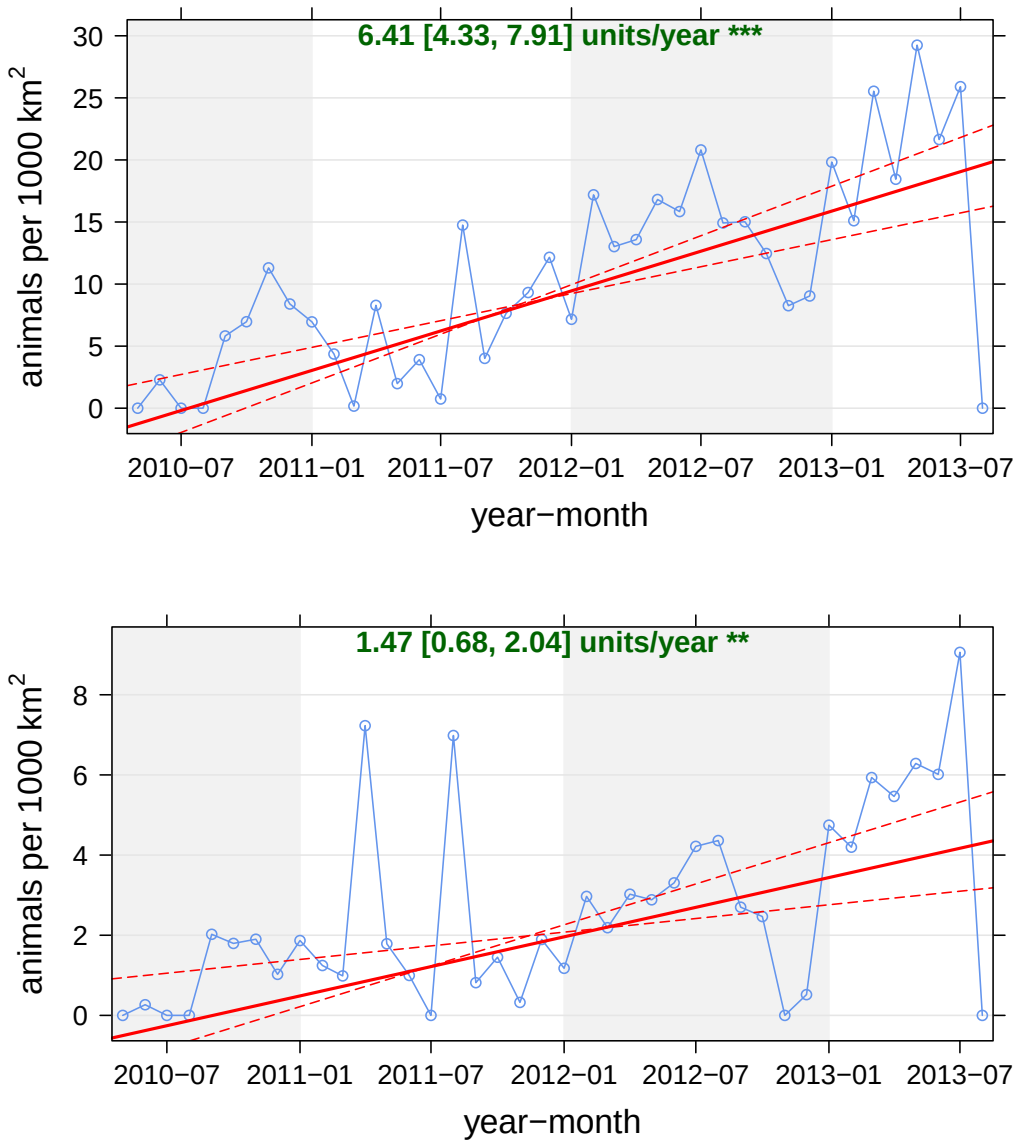


Figure 7.6: Theil-Sen slope estimates (solid red line) for deseasonalized pilot whale density time series at site MC, with 95% confidence intervals (dashed red lines). Blue dots are monthly averages of daily density estimates. ***Top***: Group method time series;***Bottom***: Click method time series.

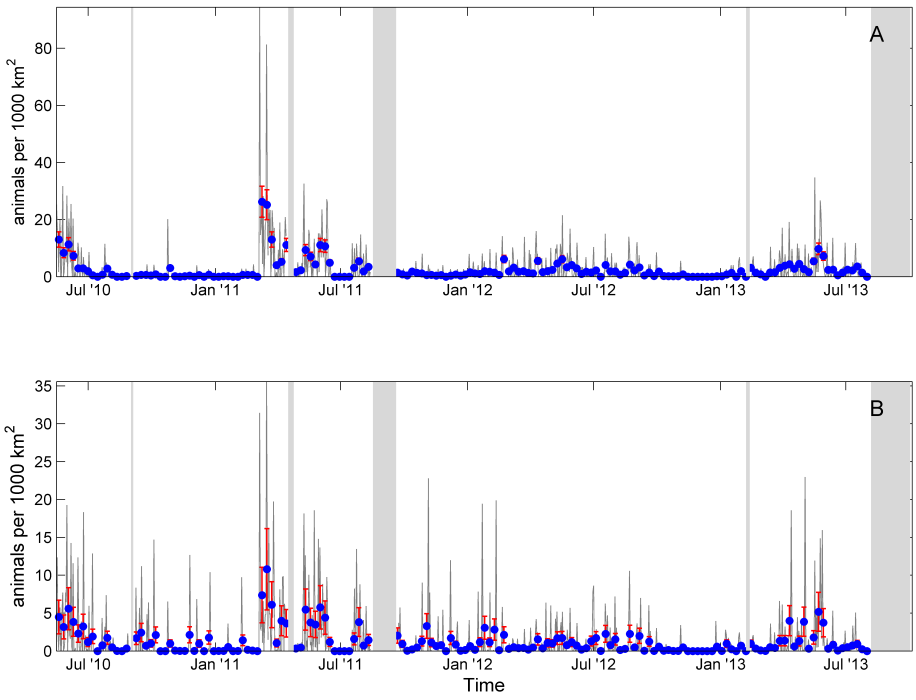


Figure 7.7: Density estimates for Risso's dolphins at site MC. **A**: Group-method results. **B**: Click method results. Grey line indicates daily density estimates, blue squares are weekly density estimates with red error bars indicating standard error. Grey rectangles indicate gaps in recording effort.

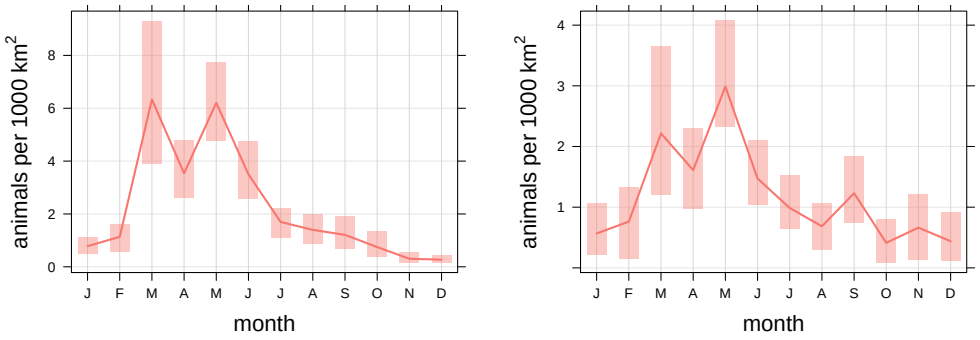


Figure 7.8: Mean density estimates and 95% confidence intervals of the mean by month for Risso's dolphins at site MC. ***Left***: Group method estimates; ***Right***: Click method estimates.

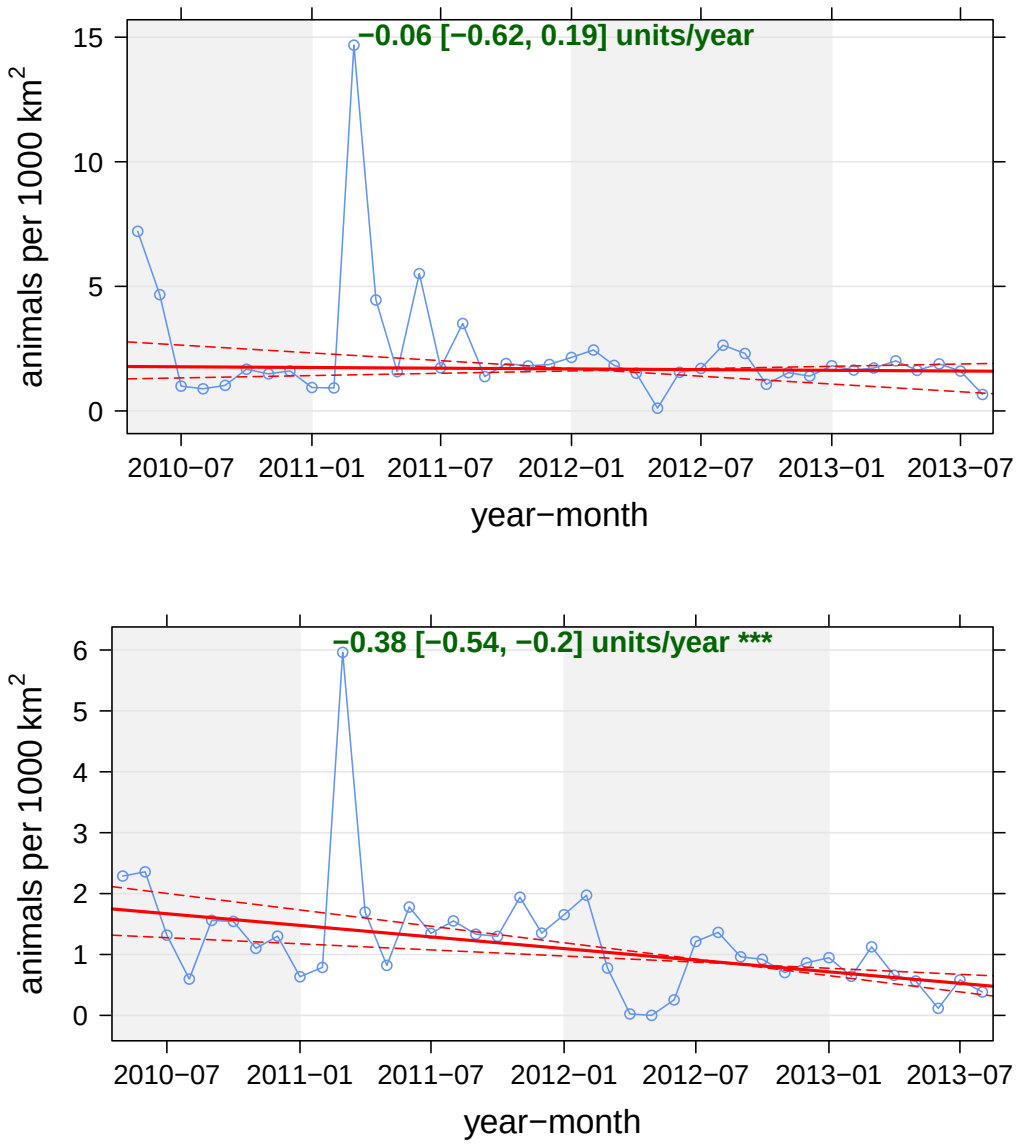


Figure 7.9: Theil-Sen slope estimates (solid red line) for deseasonalized Risso's dolphin density time series at site MC, with 95% confidence intervals (dashed red lines). Blue dots are monthly averages of daily density estimates. ***Top***: Group method time series; ***Bottom***: Click method time series.

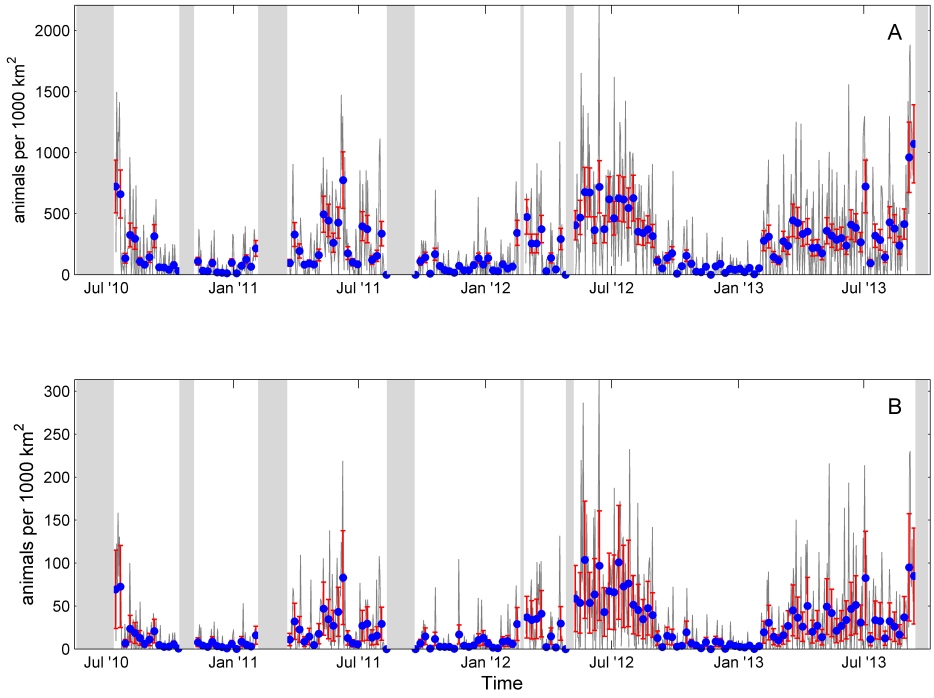


Figure 7.10: Density estimates for Stenellid dolphins at site GC. **A**: Group-method results. **B**: Click method results. Grey line indicates daily density estimates, blue squares are weekly density estimates with red error bars indicating standard error. Grey rectangles indicate gaps in recording effort.

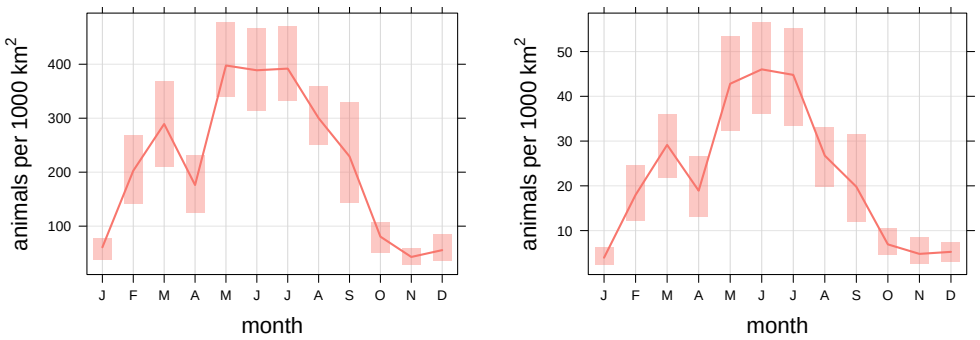


Figure 7.11: Mean density estimates and 95% confidence intervals of the mean by month for Stenellid dolphins at site GC. ***Left***: Group method estimates; ***Right***: Click method estimates.

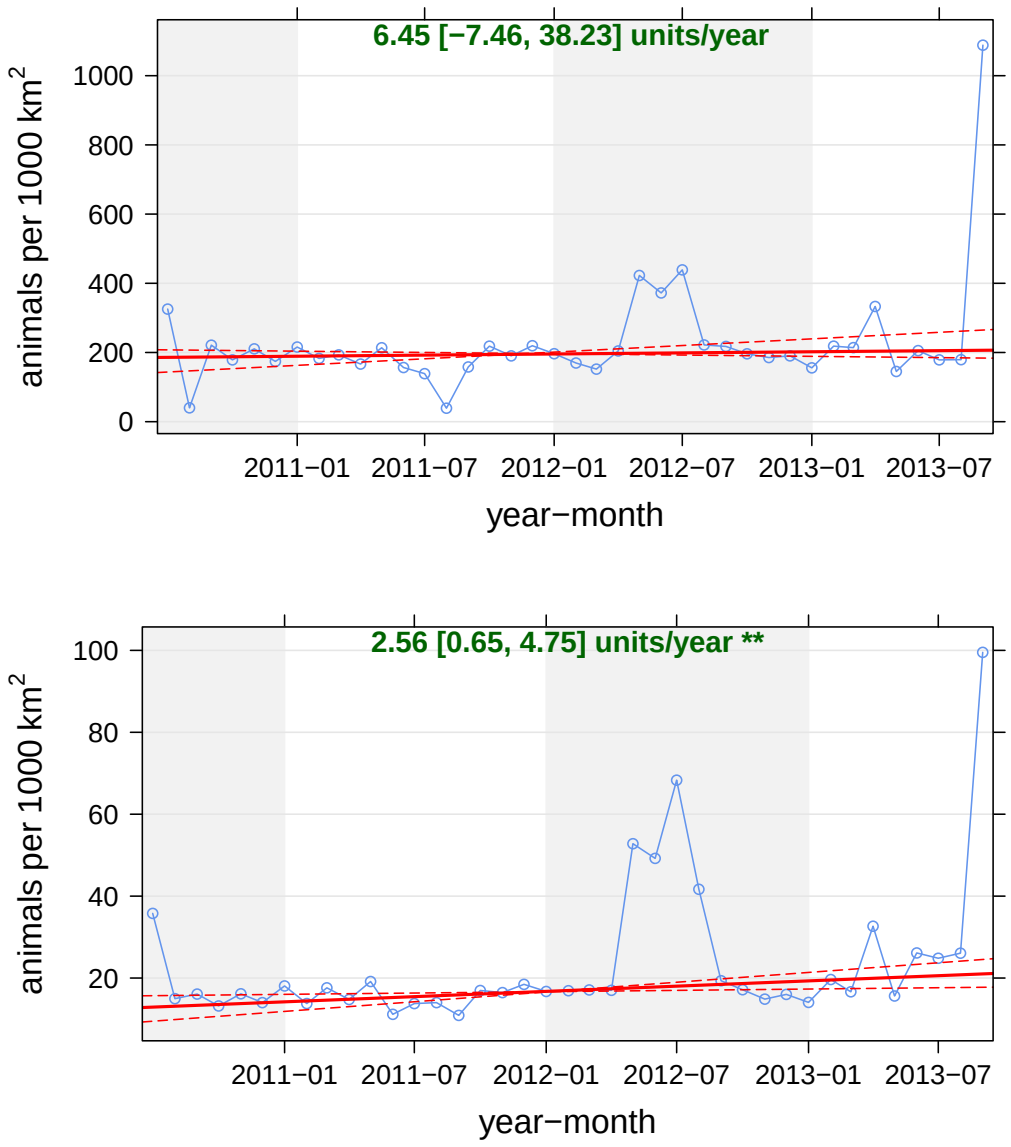


Figure 7.12: Theil-Sen slope estimates (solid red line) for deseasonalized Stenellid dolphin density time series at site GC, with 95% confidence intervals (dashed red lines). Blue dots are monthly averages of daily density estimates. ***Top***: Group method time series; ***Bottom***: Click method time series.

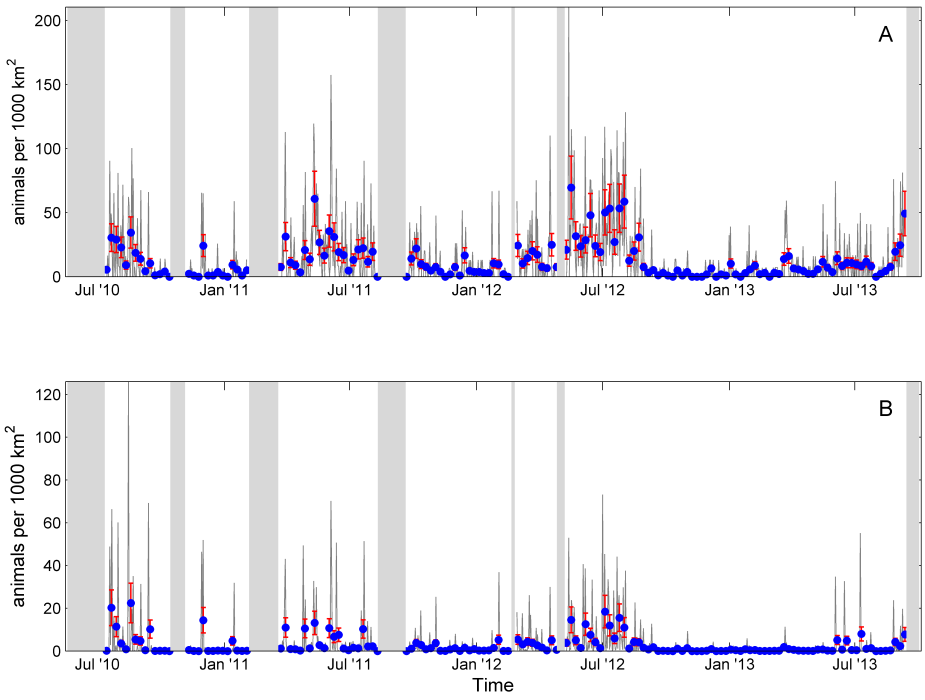


Figure 7.13: Density estimates for pilot whales at site GC. **A**: Group-method results. **B**: Click method results. Grey line indicates daily density estimates, blue squares are weekly density estimates with red error bars indicating standard error. Grey rectangles indicate gaps in recording effort.

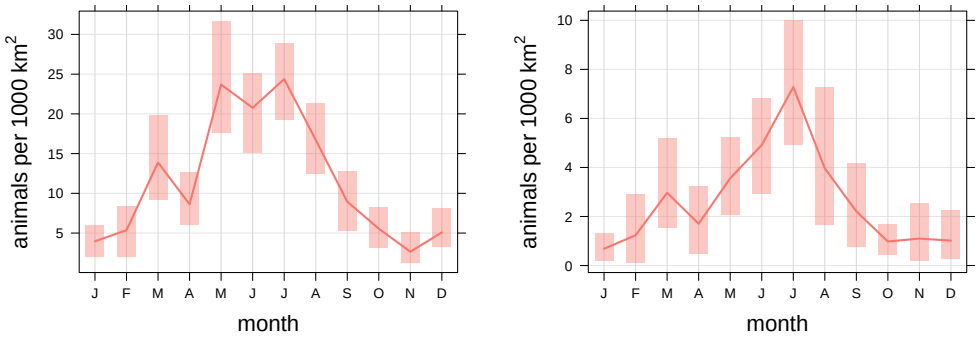


Figure 7.14: Mean density estimates and 95% confidence intervals of the mean by month for pilot whales dolphins at site GC. ***Left***: Group method estimates; ***Right***: Click method estimates.

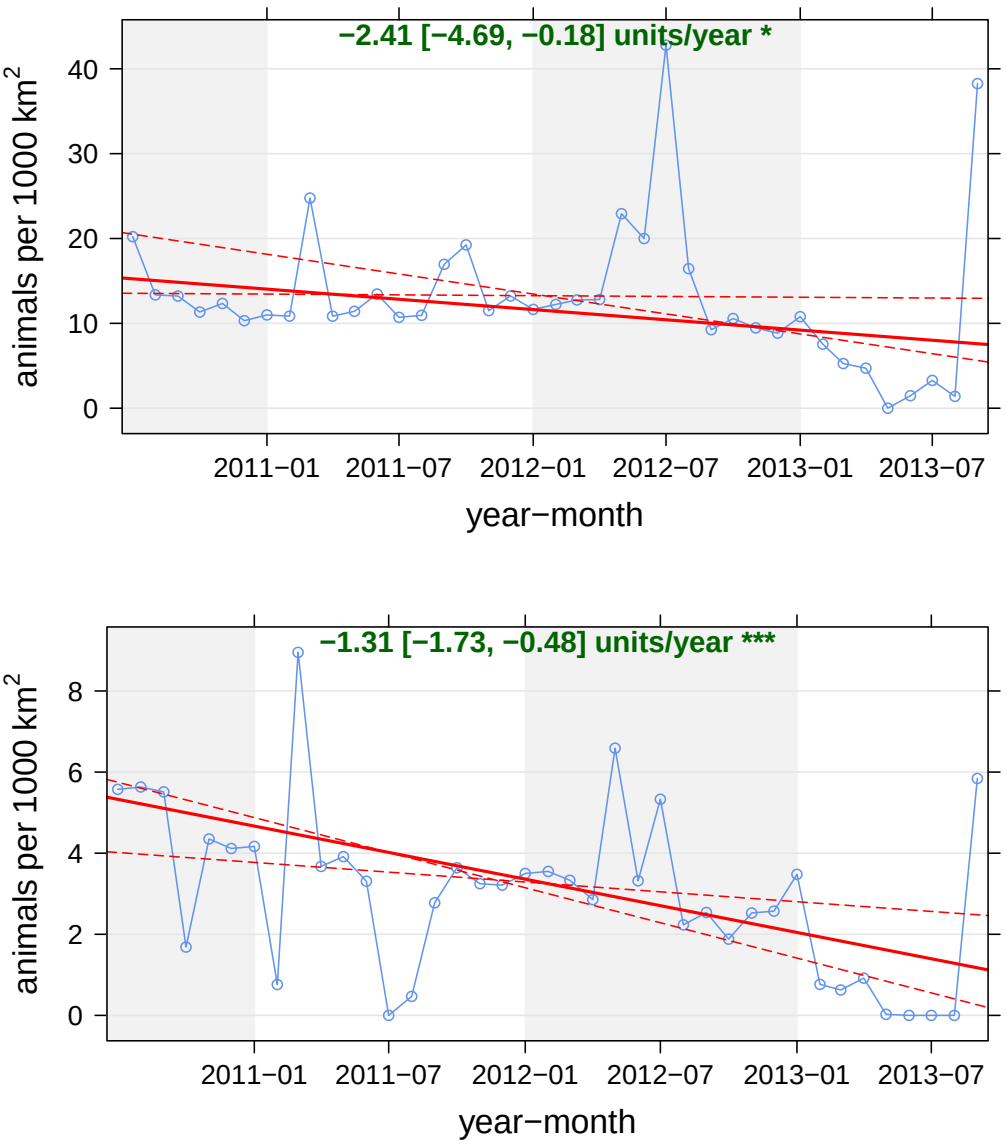


Figure 7.15: Theil-Sen slope estimates (solid red line) for deseasonalized pilot whale density time series at site GC, with 95% confidence intervals (dashed red lines). Blue dots are monthly averages of daily density estimates. ***Top***: Group method time series; ***Bottom***: Click method time series.

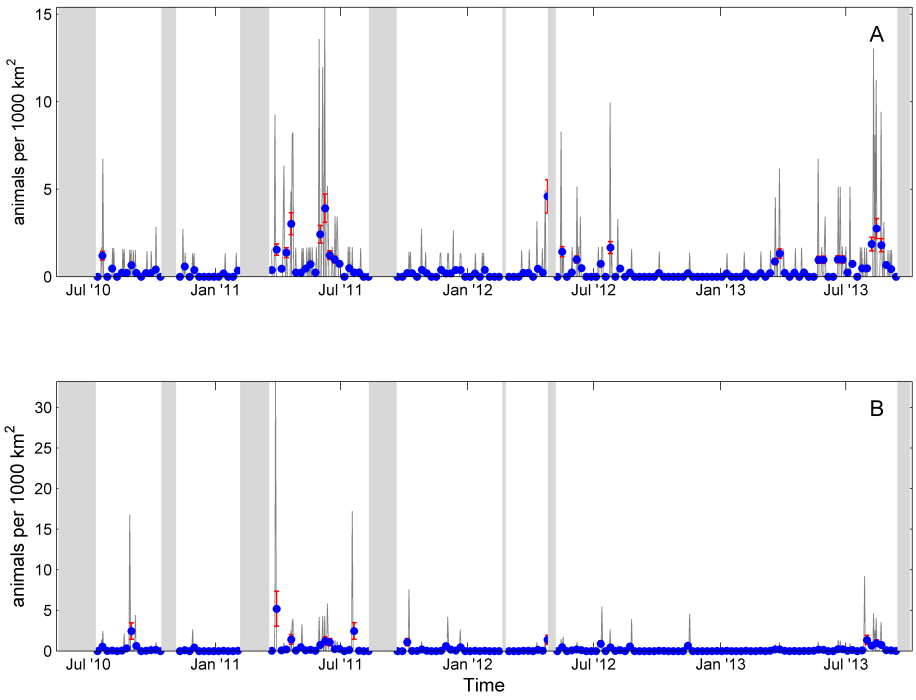


Figure 7.16: Density estimates for Risso's dolphins at site GC. **A**: Group-method results. **B**: Click method results. Grey line indicates daily density estimates, blue squares are weekly density estimates with red error bars indicating standard error. Grey rectangles indicate gaps in recording effort.

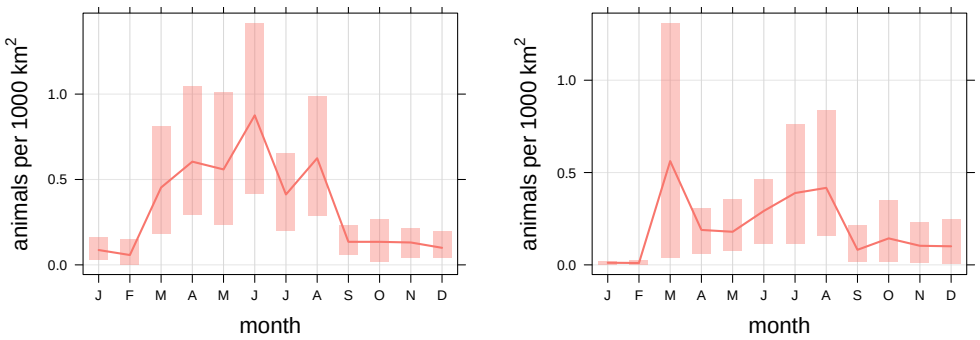


Figure 7.17: Mean density estimates and 95% confidence intervals of the mean by month for Risso's dolphins at site GC. ***Left***: Group method estimates; ***Right***: Click method estimates.

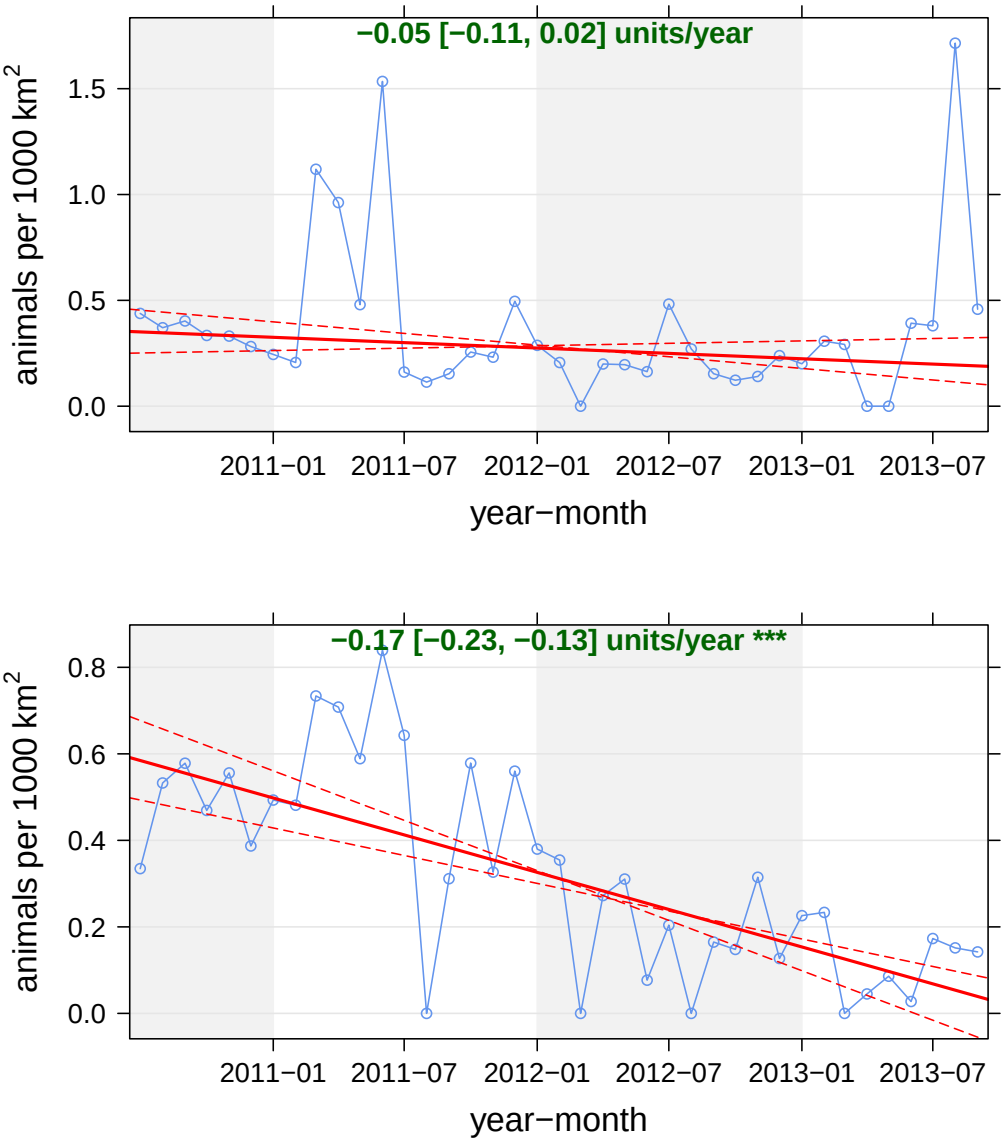


Figure 7.18: Theil-Sen slope estimates (solid red line) for deseasonalized Risso's dolphin density time series at site GC, with 95% confidence intervals (dashed red lines). Blue dots are monthly averages of daily density estimates. ***Top***: Group method time series; ***Bottom***: Click method time series.

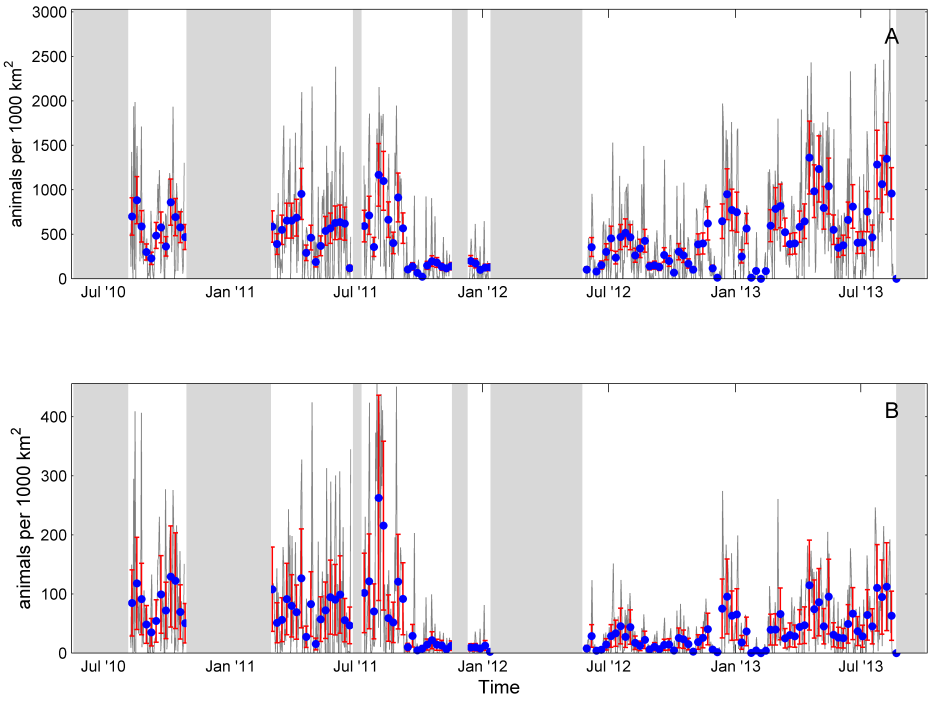


Figure 7.19: Density estimates for Stenellid dolphins at site DT. **A**: Group-method results. **B**: Click method results. Grey line indicates daily density estimates, blue squares are weekly density estimates with red error bars indicating standard error. Grey rectangles indicate gaps in recording effort.

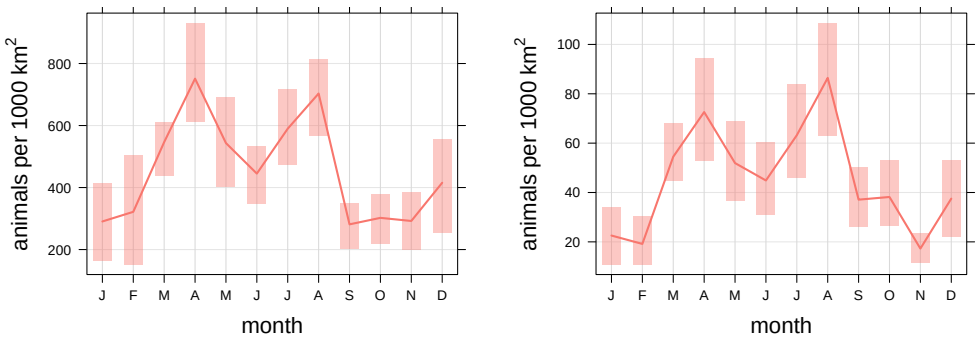


Figure 7.20: Mean density estimates and 95% confidence intervals of the mean by month for Stenellid dolphins at site DT. ***Left***: Group method estimates; ***Right***: Click method estimates.

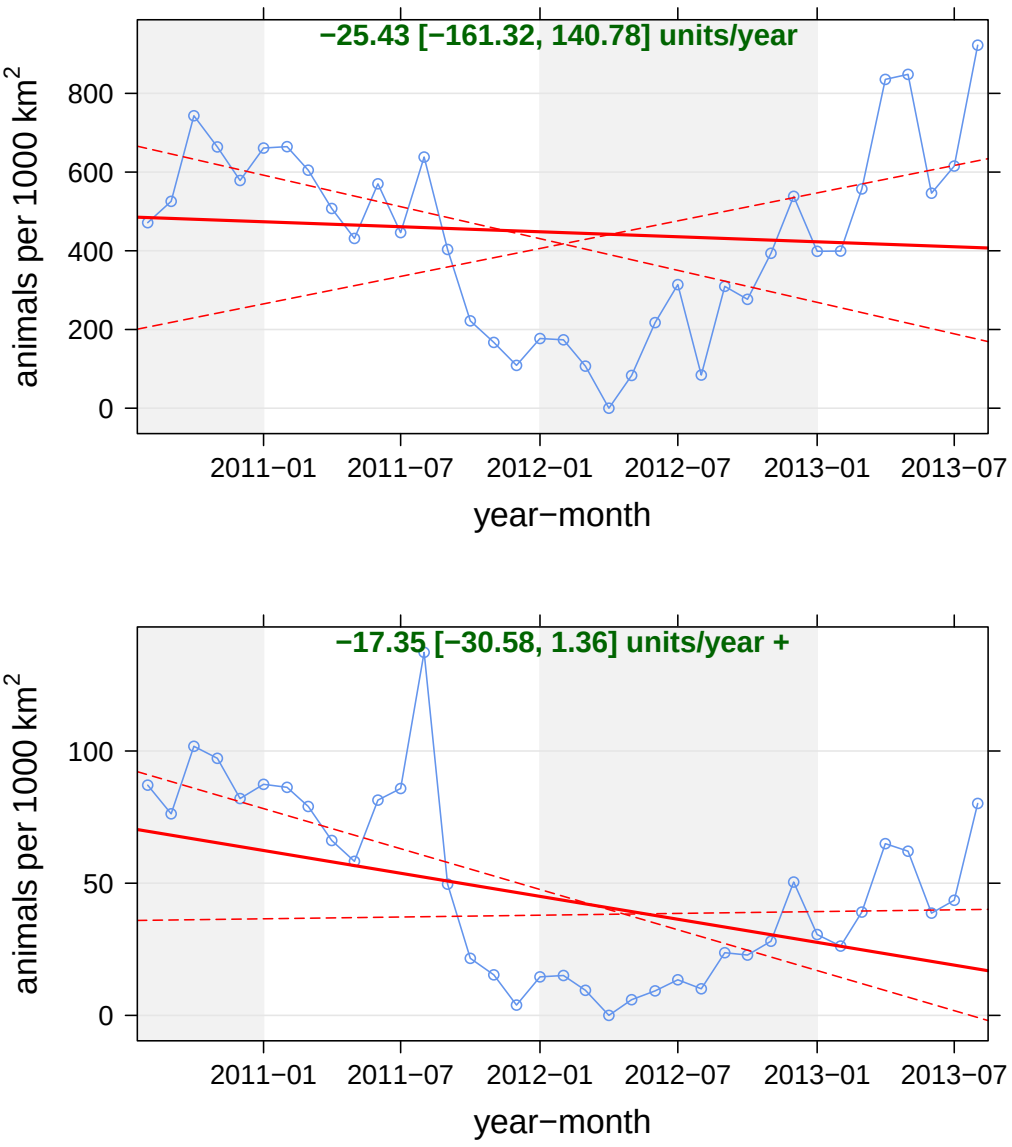


Figure 7.21: Theil-Sen slope estimates (solid red line) for deseasonalized Stenellid dolphin density time series at site DT, with 95% confidence intervals (dashed red lines). Blue dots are monthly averages of daily density estimates. ***Top***: Group method time series; ***Bottom***: Click method time series.

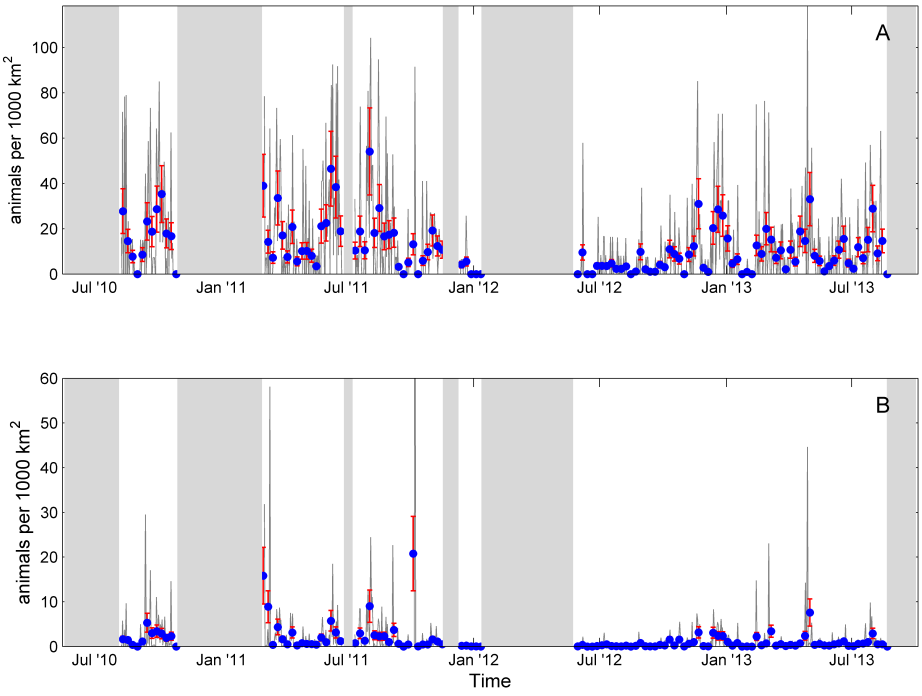


Figure 7.22: Density estimates for pilot whales at site DT. **A**: Group-method results. **B**: Click method results. Grey line indicates daily density estimates, blue squares are weekly density estimates with red error bars indicating standard error. Grey rectangles indicate gaps in recording effort.

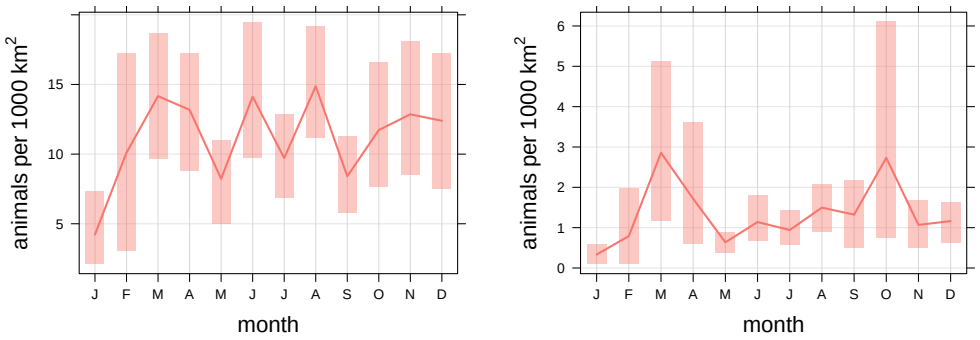


Figure 7.23: Mean density estimates and 95% confidence intervals of the mean by month for pilot whales dolphins at site DT. ***Left***: Group method estimates; ***Right***: Click method estimates.

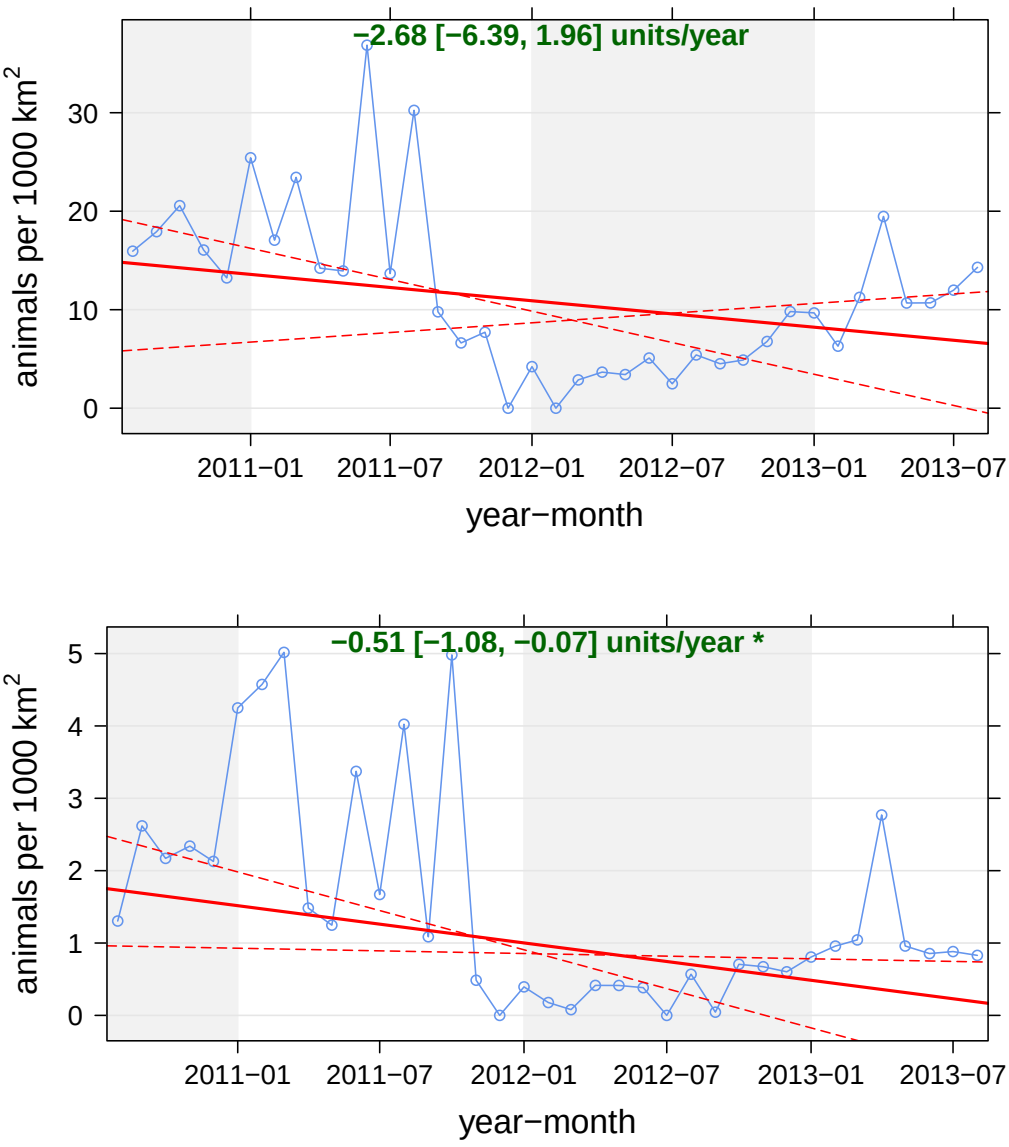


Figure 7.24: Theil-Sen slope estimates (solid red line) for deseasonalized pilot whale density time series at site DT, with 95% confidence intervals (dashed red lines). Blue dots are monthly averages of daily density estimates. ***Top***: Group method time series; ***Bottom***: Click method time series.

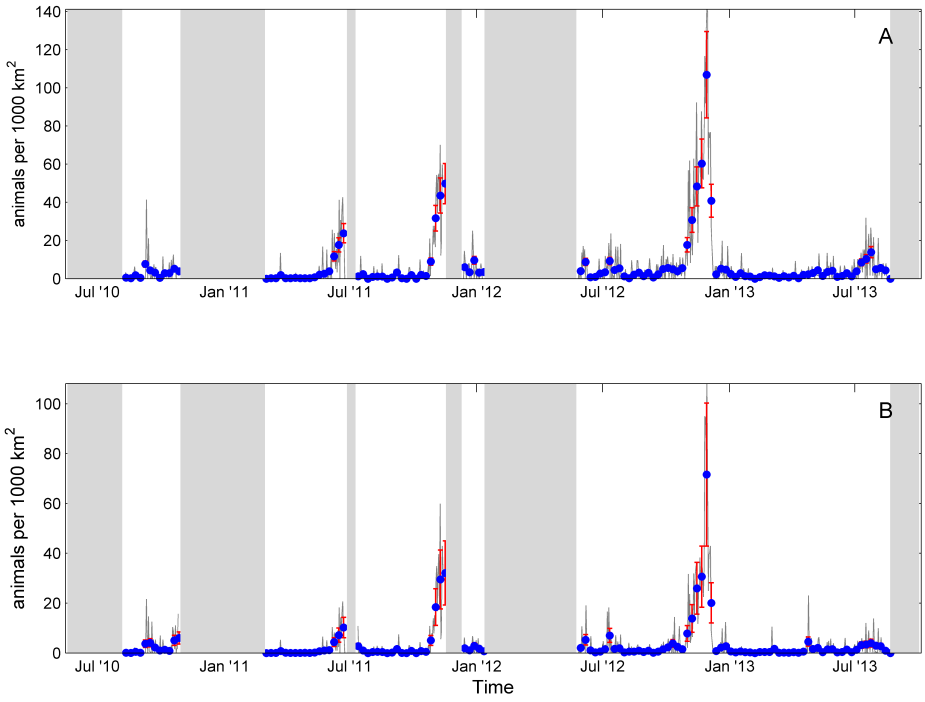


Figure 7.25: Density estimates for Risso's dolphins at site DT. **A**: Group-method results. **B**: Click method results. Grey line indicates daily density estimates, blue squares are weekly density estimates with red error bars indicating standard error. Grey rectangles indicate gaps in recording effort.

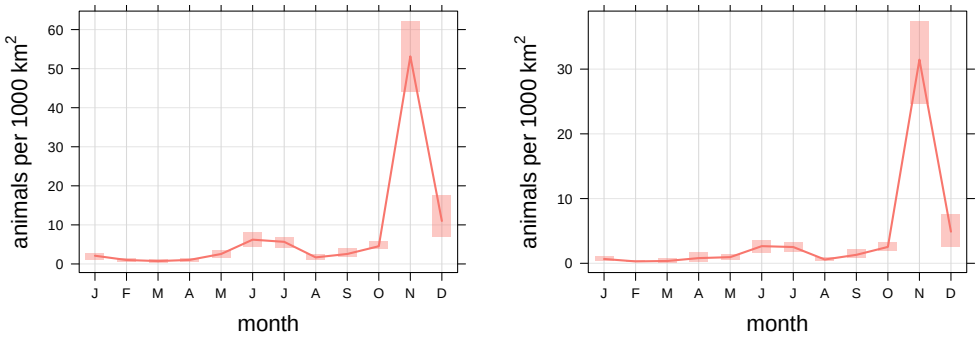


Figure 7.26: Mean density estimates and 95% confidence intervals of the mean by month for Risso's dolphins at site DT. ***Left***: Group method estimates; ***Right***: Click method estimates.

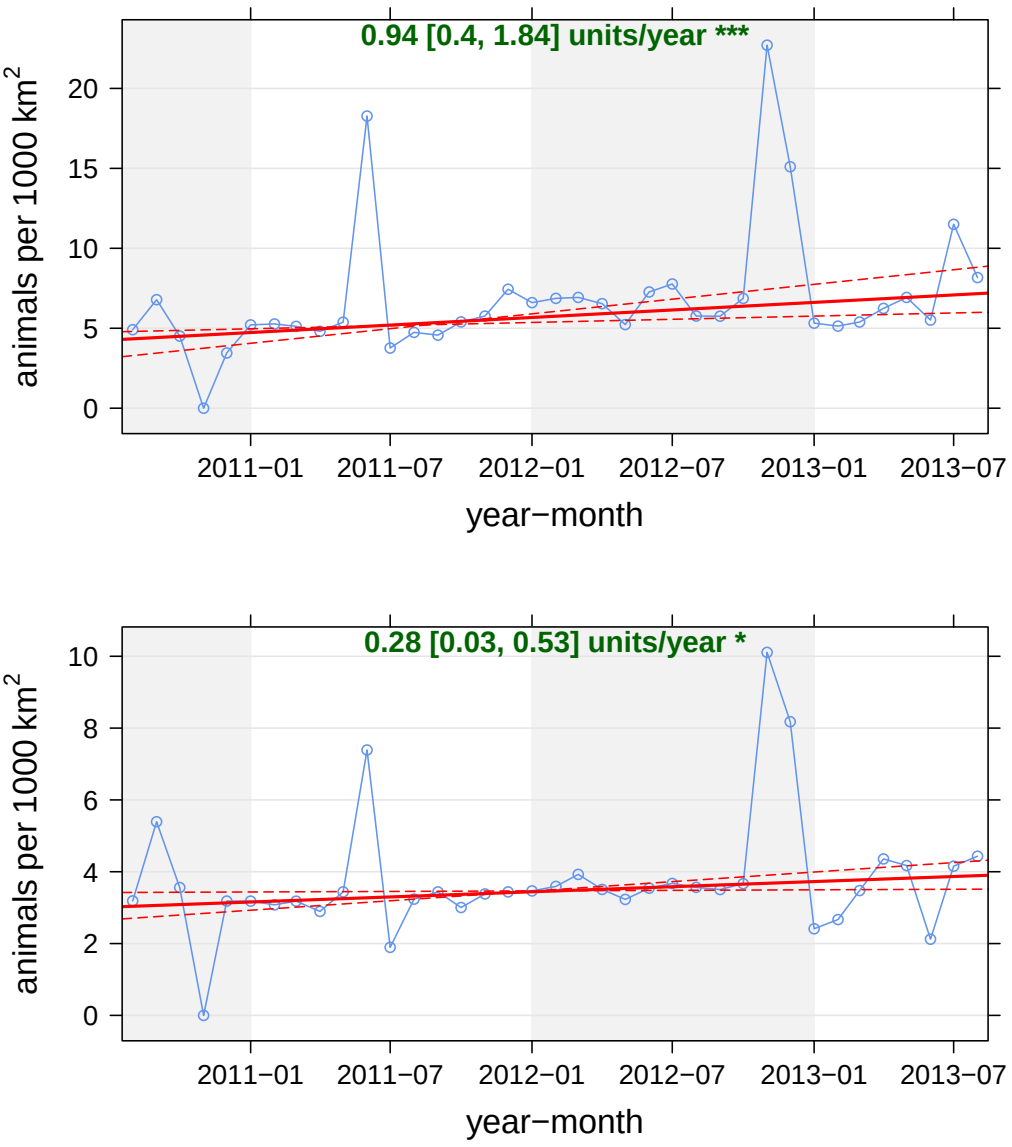


Figure 7.27: Theil-Sen slope estimates (solid red line) for deseasonalized Risso's dolphin density time series at site DT, with 95% confidence intervals (dashed red lines). Blue dots are monthly averages of daily density estimates. ***Top***: Group method time series; ***Bottom***: Click method time series.

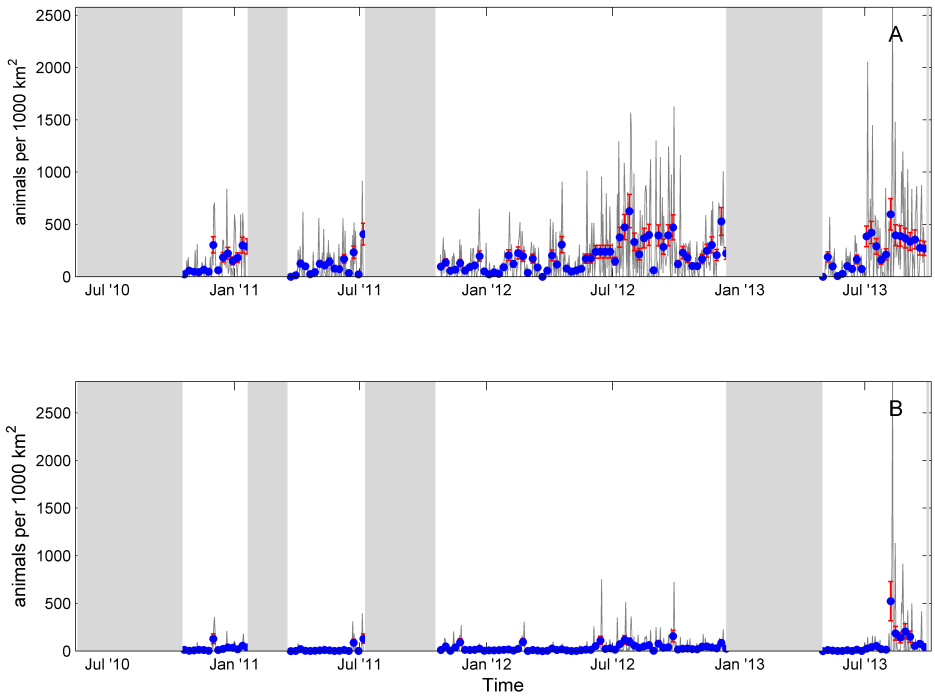


Figure 7.28: Density estimates for Stenellid dolphins at site DC. **A**: Group-method results. **B**: Click method results. Grey line indicates daily density estimates, blue squares are weekly density estimates with red error bars indicating standard error. Grey rectangles indicate gaps in recording effort.

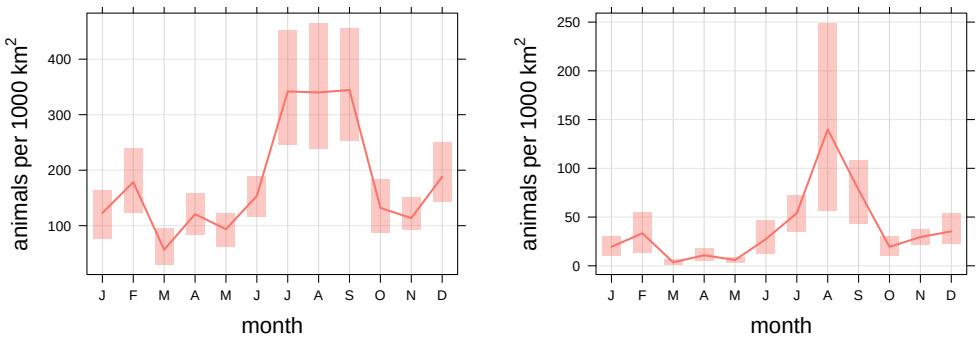


Figure 7.29: Mean density estimates and 95% confidence intervals of the mean by month for Stenellid dolphins at site DC. ***Left***: Group method estimates; ***Right***: Click method estimates.

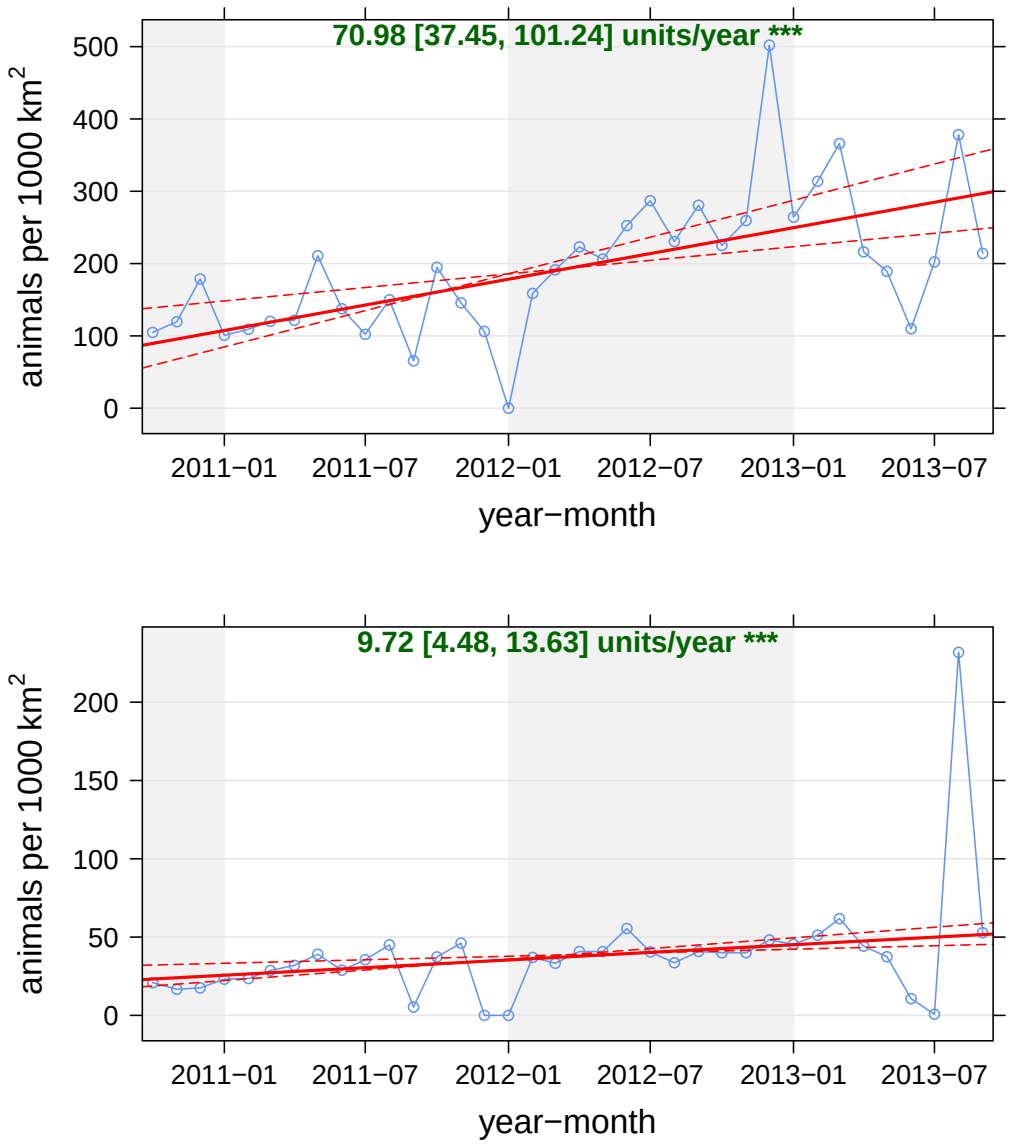


Figure 7.30: Theil-Sen slope estimates (solid red line) for deseasonalized Stenellid dolphin density time series at site DC, with 95% confidence intervals (dashed red lines). Blue dots are monthly averages of daily density estimates. ***Top***: Group method time series; ***Bottom***: Click method time series.

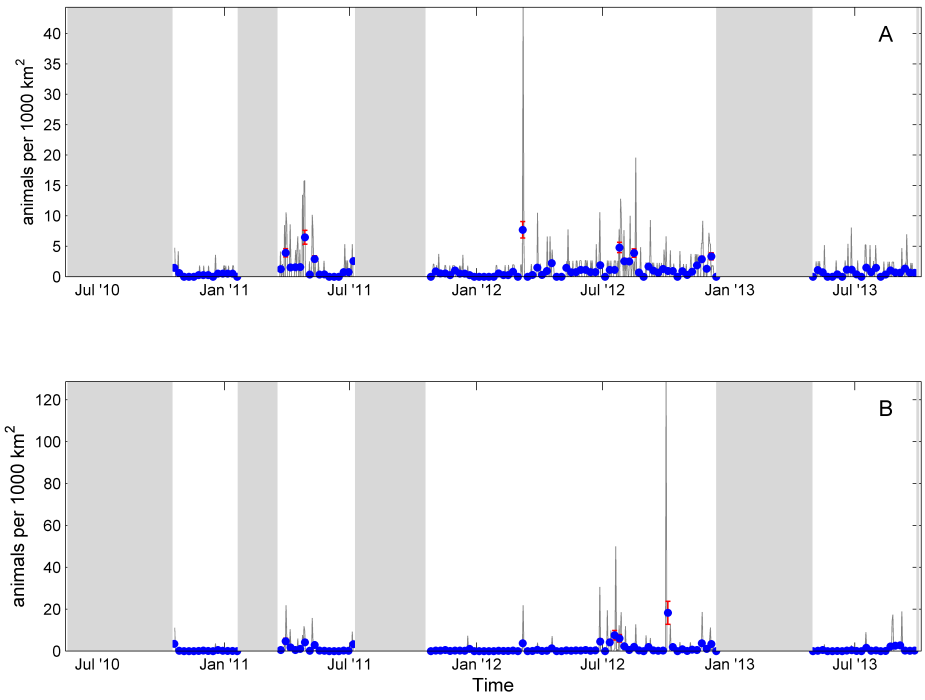


Figure 7.31: Density estimates for Risso's dolphins at site DC. **A**: Group-method results. **B**: Click method results. Grey line indicates daily density estimates, blue squares are weekly density estimates with red error bars indicating standard error. Grey rectangles indicate gaps in recording effort.

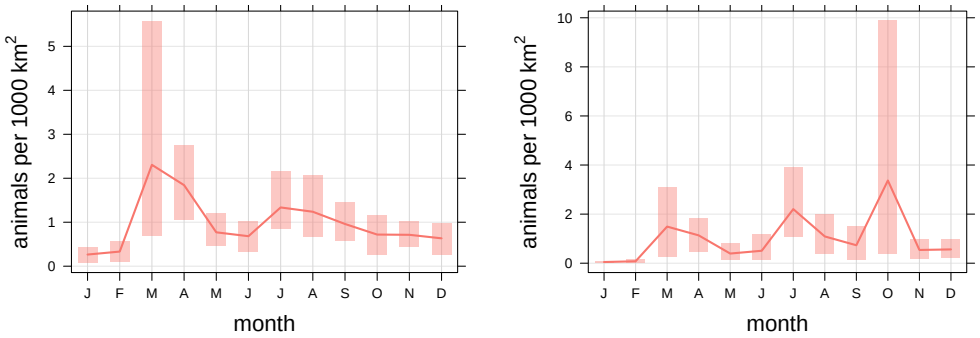


Figure 7.32: Mean density estimates and 95% confidence intervals of the mean by month for Risso's dolphins at site DC. ***Left***: Group method estimates; ***Right***: Click method estimates.

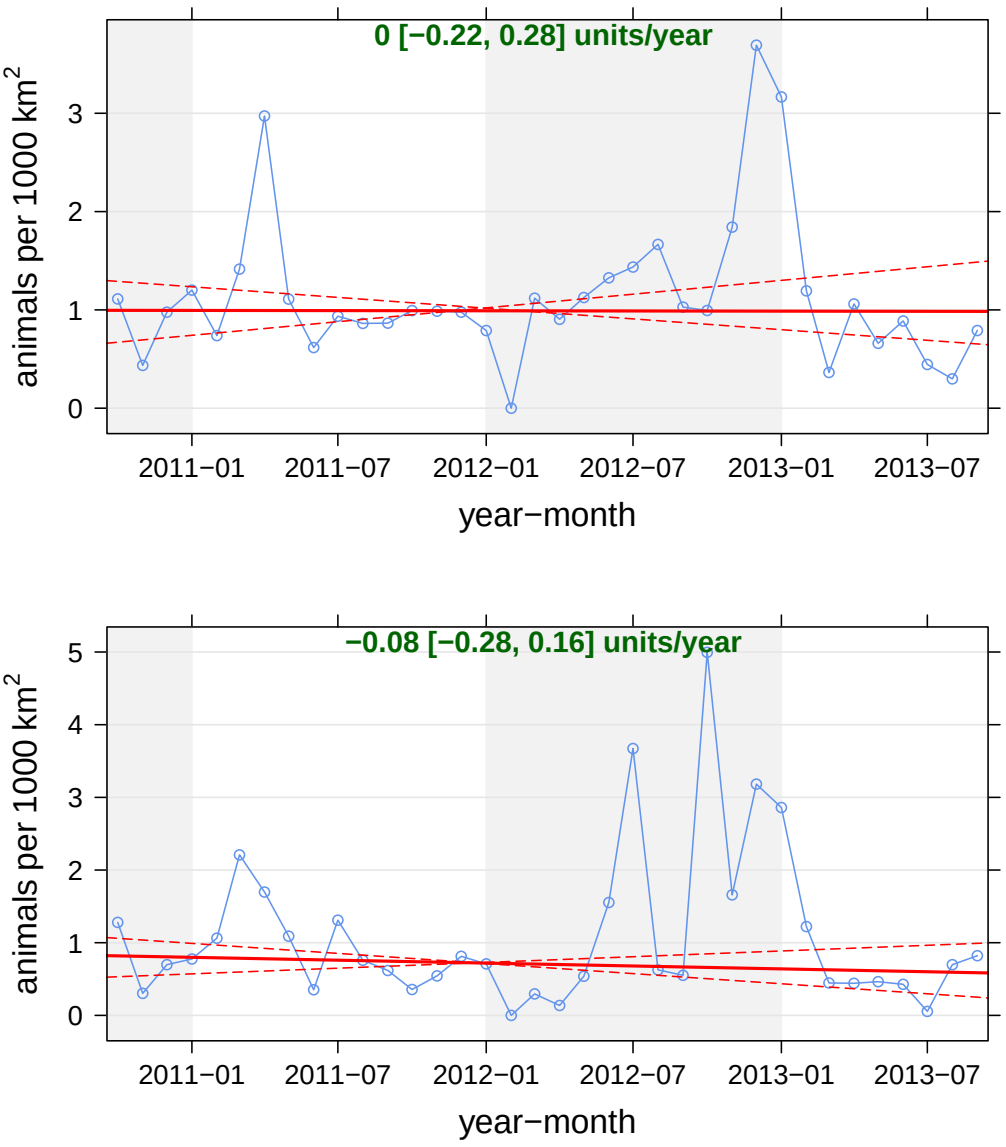


Figure 7.33: Theil-Sen slope estimates (solid red line) for deseasonalized Risso's dolphin density time series at site DC, with 95% confidence intervals (dashed red lines). Blue dots are monthly averages of daily density estimates. ***Top***: Group method time series; ***Bottom***: Click method time series.

Bibliography

Akamatsu, T., Teilmann, J., Miller, L. A., Tougaard, J., Dietz, R., Wang, D., Wang, K., Siebert, U., and Naito, Y. (2007). Comparison of echolocation behaviour between coastal and riverine porpoises. *Deep Sea Research Part II: Topical studies in Oceanography*, 54(3):290–297.

Akamatsu, T., Wang, D., Nakamura, K., and Wang, K. (1998). Echolocation range of captive and free-ranging baiji (*Lipotes vexillifer*), finless porpoise (*Neophocaena phocaenoides*), and bottlenose dolphin (*Tursiops truncatus*). *Journal of the Acoustical Society of America*, 104(4):2511–2516.

Archer, F. I. and Perrin, W. F. (1999). *Stenella coeruleoalba*. *Mammalian Species*, pages 1–9.

Aroyan, J. L., Cranford, T. W., Kent, J., and Norris, K. S. (1992). Computer modeling of acoustic beam formation in *Delphinus delphis*. *The Journal of the Acoustical Society of America*, 92(5):2539–2545.

Au, W. (2004). Echolocation signals of wild dolphins. *Acoustical Physics*, 50(4):454–462.

Au, W. and Benoit-Bird, K. (2003). Automatic gain control in the echolocation system of dolphins. *Nature*, 423(6942):861–863.

Au, W., Branstetter, B., Benoit-Bird, K., and Kastelein, R. (2009). Acoustic basis for fish prey discrimination by echolocating dolphins and porpoises. *Journal of the Acoustical Society of America*, 126(1):460–467.

Au, W., Branstetter, B., Moore, P., and Finneran, J. (2012a). Dolphin biosonar signals measured at extreme off-axis angles: Insights to sound propagation in the head. *Journal of the Acoustical Society of America*, 132(2):1199–1206.

Au, W., Branstetter, B., Moore, P., and Finneran, J. (2012b). The biosonar field around an Atlantic bottlenose dolphin (*Tursiops truncatus*). *Journal of the Acoustical Society of America*, 131(1):569–576.

Au, W. and Herzing, D. (2003). Echolocation signals of wild Atlantic spotted dolphin (*Stenella frontalis*). *Journal of the Acoustical Society of America*, 113(1):598–604.

Au, W. W. (1993). *The sonar of dolphins*. Springer.

Au, W. W. L., Moore, P. W. B., and Pawloski, D. (1986). Echolocation transmitting beam of the Atlantic bottlenose dolphin. *The Journal of the Acoustical Society of America*, 80(2):688–691.

Au, W. W. L., Pawloski, J. L., Nachtigall, P. E., Blonz, M., and Gisner, R. C. (1995). Echolocation signals and transmission beam pattern of a false killer whale (*Pseudorca crassidens*). *The Journal of the Acoustical Society of America*, 98(1):51–59.

Austin, D., Carriker, B., McGuire, T., Pratt, J., Priest, T., and Pulsipher, A. (2004). History of the Offshore Oil and Gas Industry in Southern Louisiana. Interim Report. Volume 1: Papers on the Evolving Offshore Industry. Technical report, U.S. Dept. of the Interior, Minerals Management Service, Gulf of Mexico OCS Region, New Orleans, LA. OCS Study MMS 2008-042.

Baggenstoss, P. (2011). An algorithm for the localization of multiple interfering sperm whales using multi-sensor time difference of arrival. *Journal of the Acoustical Society of America*, 130(1):102–112.

Baird, R., Ligon, A., Hooker, S., and Gorgone, A. (2001). Subsurface and nighttime behaviour of pantropical spotted dolphins in Hawai'i. *Canadian Journal of Zoology-Revue Canadienne De Zoologie*, 79(6):988–996.

Baird, R. W. (2002). Risso's dolphin. In Perrin, W., Würsig, B., and Thewissen, J., editors, *Encyclopedia of marine mammals*, pages 1037–1039. Academic Press, San Diego, California.

Baird, R. W., Webster, D. L., Mahaffy, S. D., McSweeney, D. J., Schorr, G. S., and Ligon, A. D. (2008). Site fidelity and association patterns in a deep-water dolphin: Rough-toothed dolphins (*Steno bredanensis*) in the Hawaiian Archipelago. *Marine Mammal Science*, 24(3):535–553.

Baron, S. C., Martinez, A., Garrison, L. P., and Keith, E. O. (2008). Differences in acoustic signals from Delphinids in the western North Atlantic and northern Gulf of Mexico. *Marine Mammal Science*, 24(1):42–56.

Barron, G. L. and Jefferson, T. A. (1993). First records of the melon-headed whale (*Peponocephala electra*) from the Gulf of Mexico. *The Southwestern Naturalist*, 38(1):82–85.

Bastian, M., Heymann, S., and Jacomy, M. (2009). *Gephi*: An open source software for exploring and manipulating networks. *ICWSM*, 8:361–362.

Baumann-Pickering, S., Wiggins, S., Hildebrand, J., Roch, M., and Schnitzler, H. (2010). Discriminating features of echolocation clicks of melon-headed whales

(*Peponocephala electra*), bottlenose dolphins (*Tursiops truncatus*), and Gray's spinner dolphins (*Stenella longirostris longirostris*). *Journal of the Acoustical Society of America*, 128(4):2212–2224.

Baumgartner, M. F. (1997). The distribution of Risso's dolphin (*Grampus griseus*) with respect to the physiography of the northern Gulf of Mexico. *Marine Mammal Science*, 13(4):614–638.

Benoit-Bird, K. (2004). Prey caloric value and predator energy needs: foraging predictions for wild spinner dolphins. *Marine Biology*, 145(3):435–444.

Benoit-Bird, K. J. and Au, W. W. (2003). Prey dynamics affect foraging by a pelagic predator (*Stenella longirostris*) over a range of spatial and temporal scales. *Behavioral Ecology and Sociobiology*, 53(6):364–373.

Berner, E. K. and Berner, R. A. (1987). *Global water cycle: geochemistry and environment*. Prentice-Hall.

Biggs, D. (1992). Nutrients, plankton, and productivity in a warm-core ring in the western Gulf of Mexico. *Journal of Geophysical Research: Oceans (1978-2012)*, 97(C2):2143–2154.

Biggs, D. C., Leben, R. R., and Ortega-Ortiz, J. (2000). Ship and satellite studies of mesoscale circulation and sperm whale habitats in the northeast Gulf of Mexico during GulfCet II. *Gulf of Mexico Science*, 18(1):15–22.

Biggs, D. C. and Müller-Karger, F. E. (1994). Ship and satellite observations of chlorophyll stocks in interacting cyclone-anticyclone eddy pairs in the western Gulf of Mexico. *Journal of Geophysical Research: Oceans (1978-2012)*, 99(C4):7371–7384.

Bingham, B., Kraus, N., Howe, B., Freitag, L., Ball, K., Koski, P., and Gallimore, E. (2012). Passive and active acoustics using an autonomous wave glider. *Journal of Field Robotics*, 29(6):911–923.

Birkett, S. and Rapport, D. (1999). A stress-response assessment of the North-western Gulf of Mexico ecosystem. In Kempf, H., Steidinger, K., and Sherman, K., editors, *The Gulf of Mexico Large Marine Ecosystem: Assessment, Sustainability, and Management*, pages 438–458. Blackwell Science, Oxford.

Bittle, M. and Duncan, A. (2013). A review of current marine mammal detection and classification algorithms for use in automated passive acoustic monitoring. In McMinn, T., editor, *Proceedings of Acoustics: Science, Technology and Amenity*, pages 1–8, Victor Harbour, South Australia. Australian Acoustical Society.

Blondel, V., Guillaume, J., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics-Theory and Experiment*, page 12.

Bowman, A. W. and Azzalini, A. (1997). *Applied Smoothing Techniques for Data Analysis: The Kernel Approach with S-Plus Illustrations*. Oxford University Press.

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.

Buckland, S. T., Anderson, D. R., Burnham, K. P., Laake, J. L., Borchers, D. L., and Thomas, L. (2007). *Advanced Distance Sampling: Estimating abundance of biological populations*. Oxford.

Cai, Y., Rooker, J. R., Gill, G. A., and Turner, J. P. (2007). Bioaccumulation of mercury in pelagic fishes from the northern Gulf of Mexico. *Canadian Journal of Fisheries and Aquatic Sciences*, 64(3):458–469.

Caillat, M., Thomas, L., and Gillespie, D. (2013). The effects of acoustic misclassification on cetacean species abundance estimation. *The Journal of the Acoustical Society of America*, 134(3):2469–2476.

Camilli, R., Di Iorio, D., Bowen, A., Reddy, C., Techet, A., Yoerger, D., Whitcomb, L., Seewald, J., Sylva, S., and Fenwick, J. (2012). Acoustic measurement of the Deepwater Horizon Macondo well flow rate. *Proceedings of the National Academy of Sciences of the United States of America*, 109(50):20235–20239.

Camilli, R., Reddy, C. M., Yoerger, D. R., Van Mooy, B. A., Jakuba, M. V., Kinsey, J. C., McIntyre, C. P., Sylva, S. P., and Maloney, J. V. (2010). Tracking hydrocarbon plume transport and biodegradation at Deepwater Horizon. *Science*, 330(6001):201–204.

Carmichael, R., Graham, W., Aven, A., Worthy, G., and Howden, S. (2012). Were Multiple Stressors a 'Perfect Storm' for Northern Gulf of Mexico Bottlenose Dolphins (*Tursiops truncatus*) in 2011? *Plos One*, 7(7).

Carsey, J. B. (1950). Geology of Gulf coastal area and continental shelf. *AAPG Bulletin*, 34(3):361–385.

Carslaw, D. and Ropkins, K. (2012). openair — an R package for air quality data analysis. *Environmental Modelling & Software*, 27-28(52-61).

Carslaw, D. and Ropkins, K. (2015). *openair: Open-source tools for the analysis of air pollution data.*, R package version 1.1-5. edition.

Castellote, M., Leeney, R. H., O'Corry-Crowe, G., Lauhakangas, R., Kovacs, K. M., Lucey, W., Krasnova, V., Lydersen, C., Stafford, K. M., and Belikov, R. (2013). Monitoring white whales (*Delphinapterus leucas*) with echolocation loggers. *Polar Biology*, 36(4):493–509.

Chang, Y. and Oey, L. (2010). Eddy and wind-forced heat transports in the Gulf of Mexico. *Journal of Physical Oceanography*, 40(12):2728–2742.

Clark, C. W. (1995). Application of U.S. Navy underwater hydrophone arrays for scientific research on whales. Report, Scientific Report International Whaling Commission.

Clarke, M. R. and Pascoe, P. (1985). The stomach contents of a Risso's dolphin (*Grampus griseus*) stranded at Thurlestone, South Devon. *Journal of the Marine Biological Association of the United Kingdom*, 65(03):663–665.

Cleveland, R. B., Cleveland, W. S., McRae, J. E., and Terpenning, I. (1990). STL: A seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics*, 6(1):3–33.

Cleveland, W. S. and Devlin, S. J. (1988). Locally weighted regression: an approach to regression analysis by local fitting. *Journal of the American Statistical Association*, 83(403):596–610.

Cranford, T. W., Amundin, M., and Norris, K. S. (1996). Functional morphology and homology in the odontocete nasal complex: implications for sound generation. *Journal of Morphology*, 228(3):223–285.

Creager, K. C. and Dorman, L. M. (1982). Location of instruments on the seafloor by joint adjustment of instrument and ship positions. *Journal of Geophysical Research: Solid Earth*, 87(B10):8379–8388.

Crone, T. J. and Tolstoy, M. (2010). Magnitude of the 2010 Gulf of Mexico oil leak. *Science*, 330(6004):634–634.

Dagg, M. J. and Breed, G. A. (2003). Biological effects of Mississippi River nitrogen on the northern Gulf of Mexico: A review and synthesis. *Journal of Marine Systems*, 43(3-4):133–152.

Davis, R., Fargion, G., May, N., Leming, T., Baumgartner, M., Evans, W., Hansen, L., and Mullin, K. (1998). Physical habitat of cetaceans along the continental slope in the north-central and western Gulf of Mexico. *Marine Mammal Science*, 14(3):490–507.

Davis, R., Ortega-Ortiz, J., Ribic, C., Evans, W., Biggs, D., Ressler, P., Cady, R., Leben, R., Mullin, K., and Würsig, B. (2002). Cetacean habitat in the northern oceanic Gulf of Mexico. *Deep-Sea Research Part I-Oceanographic Research Papers*, 49(1):121–142.

Diercks, K., Trochta, R., and Evans, W. (1973). Delphinid sonar: measurement and analysis. *The Journal of the Acoustical Society of America*, 54(1):200–204.

Dolar, M. (2002). Fraser's dolphin *Lagenodelphis hosei*. In WF Perrin, B. W. and Thewissen, J., editors, *Encyclopedia of Marine Mammals*, pages 485–487. Academic Press, San Diego, California.

Etter, P. C. (1983). Heat and freshwater budgets of the Gulf of Mexico. *Journal of physical oceanography*, 13(11):2058–2069.

Fertl, D., Jefferson, T. A., Moreno, I. B., Zerbini, A. N., and Mullin, K. D. (2003). Distribution of the Clymene dolphin *Stenella clymene*. *Mammal Review*, 33(3-4):253–271.

Fiedler, P. C. and Reilly, S. B. (1994). Interannual variability of dolphin habitats in the eastern tropical Pacific. II: Effects on abundances estimated from tuna vessel sightings, 1975-1990. *Fishery Bulletin - NOAA*, 92:451–451.

Fish, J. F. and Turl, C. W. (1976). Acoustic source levels of four species of small whales. Technical report, DTIC Document.

Fisher, F. and Simmons, V. (1977). Sound absorption in sea water. *The Journal of the Acoustical Society of America*, 62(3):558–564.

Fitch, J. E. and Brownell Jr, R. L. (1968). Fish otoliths in cetacean stomachs and their importance in interpreting feeding habits. *Journal of the Fisheries Board of Canada*, 25(12):2561–2574.

Fortunato, S. and Barthelemy, M. (2007). Resolution limit in community detection. *Proceedings of the National Academy of Sciences*, 104(1):36–41.

Francois, R. and Garrison, G. (1982). Sound absorption based on ocean measurements. Part II: Boric acid contribution and equation for total absorption. *The Journal of the Acoustical Society of America*, 72(6):1879–1890.

Fritts, T. H., Irvine, A. B., Jennings, R. D., Collum, L. A., Hoffman, W., and McGehee, M. A. (1983). Turtles, birds, and mammals in the northern Gulf of Mexico and nearby Atlantic waters. An overview based on aerial surveys of OCS areas, with emphasis on oil and gas effects. Technical report, New Mexico Univ., Albuquerque (USA). Museum of Southwestern Biology.

Fulling, G. L., Mullin, K. D., and Hubard, C. W. (2003). Abundance and distribution of cetaceans in outer continental shelf waters of the US Gulf of Mexico. *Fishery Bulletin*, 101(4):923–932.

Gassmann, M., Henderson, E. E., Roch, M. A., Wiggins, S. M., and Hildebrand, J. A. (2011). Tracking dolphins with hydrophone arrays deployed from the floating instrument platform R/P FLIP in the Southern California Bight. *The Journal of the Acoustical Society of America*, 129(4):2574–2574.

Gassmann, M., Henderson, E. E., Wiggins, S. M., Roch, M. A., and Hildebrand, J. A. (2013). Offshore killer whale tracking using multiple hydrophone arrays. *The Journal of the Acoustical Society of America*, 134(5):3513–3521.

Gillespie, D. and Caillat, M. (2008). Statistical classification of odontocete clicks. *Canadian Acoustics*, 36(1):20–26.

Gillespie, D. and Leaper, R. (1996). Detection of sperm whale (*Physeter macrocephalus*) clicks and discrimination of individual vocalisations. *European Research on Cetaceans*, 10:87–91.

Giorgini, J., Yeomans, D., Chamberlin, A., Chodas, P., Jacobson, R., Keesey, M., Lieske, J., Ostro, S., Standish, E., and Wimberly, R. (1996). JPL's on-line solar system data service. In *Bulletin of the American Astronomical Society*, volume 28, page 1158.

Giraudet, P. and Glotin, H. (2006). Real-time 3D tracking of whales by echo-robust precise TDOA estimates with a widely-spaced hydrophone array. *Applied Acoustics*, 67(11-12):1106–1117.

Gore, R. H. (1992). The Gulf of Mexico. *Pineapple, Sarasota, Florida*.

Götz, T., Verfuß, U. K., and Schnitzler, H.-U. (2006). "Eavesdropping" in wild rough-toothed dolphins (*Steno bredanensis*)? *Biology letters*, 2(1):5–7.

Gregg, J. D., Dudzinski, K. M., and Smith, H. V. (2007). Do dolphins eavesdrop on the echolocation signals of conspecifics? *International Journal of Comparative Psychology*, 20(1).

Griffin, R. B. and Griffin, N. J. (2003). Distribution, habitat partitioning, and abundance of Atlantic spotted dolphins, bottlenose dolphins, and loggerhead sea turtles on the eastern Gulf of Mexico continental shelf. *Gulf of Mexico Science*, 21(1):23–34.

Griffin, R. B. and Griffin, N. J. (2004). Temporal variation in Atlantic spotted dolphin (*Stenella frontalis*) and bottlenose dolphin (*Tursiops truncatus*) densities on the west Florida continental shelf. *Aquatic Mammals*, 30(3):380–390.

Harris, D., Matias, L., Thomas, L., Harwood, J., and Geissler, W. H. (2013). Applying distance sampling to fin whale calls recorded by single seismic instruments in the northeast Atlantic. *The Journal of the Acoustical Society of America*, 134(5):3522–3535.

Hastie, G., Wilson, B., and Thompson, P. (2006). Diving deep in a foraging hotspot: acoustic insights into bottlenose dolphin dive depths and feeding behaviour. *Marine Biology*, 148(5):1181–1188.

Heide-Jørgensen, M. P., Bloch, D., Stefansson, E., Mikkelsen, B., Ofstad, L. H., and Dietz, R. (2002). Diving behaviour of long-finned pilot whales *Globicephala melas* around the Faroe Islands. *Wildlife Biology*, 8(4):307–313.

Helble, T., D'Spain, G., Campbell, G., and Hildebrand, J. (2013a). Calibrating passive acoustic monitoring: Correcting humpback whale call detections for site-specific and time-dependent environmental characteristics. *Journal of the Acoustical Society of America*, 134(5):400–406.

Helble, T., D'Spain, G., Hildebrand, J., Campbell, G., Campbell, R., and Heaney, K. (2013b). Site specific probability of passive acoustic detection of humpback whale calls from single fixed hydrophones. *Journal of the Acoustical Society of America*, 134(3):2556–2570.

Henderson, E., Hildebrand, J., and Smith, M. (2011). Classification of behavior using vocalizations of Pacific white-sided dolphins (*Lagenorhynchus obliquidens*). *Journal of the Acoustical Society of America*, 130(1):557–567.

Herzing, D. and Johnson, C. (1997). Interspecific interactions between Atlantic spotted dolphins (*Stenella frontalis*) and bottlenose dolphins (*Tursiops truncatus*) in the Bahamas, 1985-1995. *Aquatic Mammals*, 23.2:85–99.

Herzing, D. L. and Elliser, C. R. (2014). Nocturnal feeding of Atlantic spotted dolphins (*Stenella frontalis*) in the Bahamas. *Marine Mammal Science*, 30(1):367–373.

Hildebrand, J., Baumann-Pickering, S., Frasier, K., Tricky, J., Merkens, K., Wiggins, S., M., M., Harris, D., T., M., and Thomas, L. (2015). Passive acoustic monitoring of beaked whale densities in the Gulf of Mexico during and after the Deepwater Horizon oil spill. *Nature Scientific Reports*.

Hildebrand, J. A. (2009). Anthropogenic and natural sources of ambient noise in the ocean. *Marine Ecology Progress Series*, 395(5).

Hildebrand, J. A., Armi, L., and Henkart, P. C. (2012). Seismic imaging of the water-column deep layer associated with the Deepwater Horizon oil spill. *Geophysics*, 77(2):11–16.

Ibsen, S. D., Au, W. W., Nachtigall, P. E., and Breese, M. (2009). Functional bandwidth of an echolocating Atlantic bottlenose dolphin (*Tursiops truncatus*). *The Journal of the Acoustical Society of America*, 125(2):1214–1221.

Ibsen, S. D., Au, W. W., Nachtigall, P. E., DeLong, C. M., and Breese, M. (2007). Changes in signal parameters over time for an echolocating Atlantic bottlenose dolphin performing the same target discrimination task. *The Journal of the Acoustical Society of America*, 122(4):2446–2450.

Jacomy, M., Venturini, T., Heymann, S., and Bastian, M. (2014). ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the *Gephi* Software. *PLoS ONE*, 9(6):e98679.

Jefferson, T. and Schiro, A. (1997). Distribution of cetaceans in the offshore Gulf of Mexico. *Mammal Review*, 27(1):27–50.

Jensen, F. H., Bejder, L., Wahlberg, M., and Madsen, P. T. (2009). Biosonar adjustments to target range of echolocating bottlenose dolphins (*Tursiops* sp.) in the wild. *Journal of Experimental Biology*, 212(8):1078–1086.

Jensen, F. H., Rocco, A., Mansur, R. M., Smith, B. D., Janik, V. M., and Madsen, P. T. (2013). Clicking in shallow rivers: short-range echolocation of Irrawaddy and Ganges River dolphins in a shallow, acoustically complex habitat. *PloS one*, 8(4):e59284.

Kaiser, M. J. and Dodson, R. D. (2008). Trends in plug and abandonment cost in the Gulf of Mexico, 2002 - 2007. *International Journal of Oil, Gas and Coal Technology*, 1(1):24–45.

Kaltenberg, A., Biggs, D., and DiMarco, S. (2007). Deep scattering layers of the northern Gulf of Mexico observed with a ship-board 38-kHz Acoustic Doppler Current Profiler (ADCP). *Gulf of Mexico Science*, 2:97–108.

Kamminga, C., Cohen, S., and Silber, G. (1996). Investigations on cetacean sonar XI: Intrinsic comparison of the wave shapes of some members of the Phocoenidae family. *Aquatic Mammals*, 22:45–56.

Karnauskas, M., Schirripa, M. J., Kelble, C. R., Cook, G. S., and Craig, J. K. (2013). Ecosystem status report for the Gulf of Mexico. Technical report.

Kessler, J. D., Valentine, D. L., Redmond, M. C., Du, M., Chan, E. W., Mendes, S. D., Quiroz, E. W., Villanueva, C. J., Shusta, S. S., Werra, L. M., et al. (2011). A persistent oxygen anomaly reveals the fate of spilled methane in the deep Gulf of Mexico. *Science*, 331(6015):312–315.

Klatsky, L. J., Wells, R. S., and Sweeney, J. C. (2007). Offshore bottlenose dolphins (*Tursiops truncatus*): Movement and dive behavior near the Bermuda Pedestal. *Journal of Mammalogy*, 88(1):59–66.

Kloepper, L. N., Nachtigall, P. E., Donahue, M. J., and Breese, M. (2012). Active echolocation beam focusing in the false killer whale, *Pseudorca crassidens*. *The Journal of Experimental Biology*, 215(8):1306–1312.

Koblitz, J. C., Amundin, M., Carlström, J., Thomas, L., Carlén, I., Teilmann, J., Tregenza, N., Wennerberg, D., Kyhn, L., Svegaard, S., et al. (2014). Large-scale static acoustic survey of a low-density population: Estimating the abundance of the Baltic Sea harbor porpoise. *The Journal of the Acoustical Society of America*, 136(4):2248–2248.

Kujawinski, E., Soule, M., Valentine, D., Boysen, A., Longnecker, K., and Redmond, M. (2011). Fate of Dispersants Associated with the Deepwater Horizon Oil Spill. *Environmental Science & Technology*, 45(4):1298–1306.

Küsel, E., Mellinger, D., Thomas, L., Marques, T., Moretti, D., and Ward, J. (2011). Cetacean population density estimation from single fixed sensors using passive acoustics. *The Journal of the Acoustical Society of America*, 129:3610.

LeDuc, R., Perrin, W., and Dizon, A. (1999). Phylogenetic relationships among the delphinid cetaceans based on full cytochrome b sequences. *Marine Mammal Science*, 15(3):619–648.

Litchfield, C., Karol, R., Mullen, M., Dilger, J., and Lüthi, B. (1979). Physical factors influencing refraction of the echolocative sound beam in delphinid cetaceans. *Marine Biology*, 52(3):285–290.

Litz, J. A., Baran, M. A., Bowen-Stevens, S. R., Carmichael, R. H., Colegrove, K. M., Garrison, L. P., Fire, S. E., Fougeres, E. M., Hardy, R., Holmes, S., et al. (2014). Review of historical unusual mortality events (UMEs) in the Gulf of Mexico (1990-2009): providing context for the multi-year northern Gulf of Mexico cetacean UME declared in 2010. *Diseases of aquatic organisms*, 112(2):161–175.

Lucke, R. L. (1976). Counting statistics for nonnegligible dead time corrections. *Review of Scientific Instruments*, 47(6):766–767.

Madsen, P., Carder, D., Au, W., Nachtigall, P., Møhl, B., and Ridgway, S. (2003). Sound production in neonate sperm whales (L). *The Journal of the Acoustical Society of America*, 113(6):2988–2991.

Madsen, P., Kerr, I., and Payne, R. (2004a). Echolocation clicks of two free-ranging, oceanic delphinids with different food preferences: False killer whales *Pseudorca crassidens* and Risso's dolphins *Grampus griseus*. *Journal of Experimental Biology*, 207(11):1811–1823.

Madsen, P., Kerr, I., and Payne, R. (2004b). Source parameter estimates of echolocation clicks from wild pygmy killer whales (*Feresa attenuata*) (L). *Journal of the Acoustical Society of America*, 116(4):1909–1912.

Madsen, P. T. and Wahlberg, M. (2007). Recording and quantification of ultrasonic echolocation clicks from free-ranging toothed whales. *Deep Sea Research Part I: Oceanographic Research Papers*, 54(8):1421–1444.

Marques, T., Thomas, L., Martin, S., Mellinger, D., Ward, J., Moretti, D., Harris, D., and Tyack, P. (2013). Estimating animal population density using passive acoustics. *Biological Reviews*, 88(2):287–309.

Marques, T. A., Munger, L., Thomas, L., Wiggins, S., and Hildebrand, J. A. (2011). Estimating North Pacific right whale *Eubalaena japonica* density using passive acoustic cue counting. *Endangered Species Research*, 13(3):163–172.

Marques, T. A., Thomas, L., Ward, J., DiMarzio, N., and Tyack, P. L. (2009). Estimating cetacean population density using fixed passive acoustic sensors: An example with Blainville's beaked whales. *Journal of the Acoustical Society of America*, 125(4):1982–1994.

Massey Jr, F. J. (1951). The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American statistical Association*, 46(253):68–78.

Mate, B. R., Rossbach, K. A., Nieukirk, S. L., Wells, R. S., Blair Irvine, A., Scott, M. D., and Read, A. J. (1995). Satellite-monitored movements and dive behavior of a bottlenose dolphin (*Tursiops truncatus*) in Tampa Bay, Florida. *Marine Mammal Science*, 11(4):452–463.

McDonald, M. A. and Fox, C. G. (1999). Passive acoustic methods applied to fin whale population density estimation. *The Journal of the Acoustical Society of America*, 105(5):2643–2651.

McDonald, M. A., Hildebrand, J. A., and Webb, S. C. (1995). Blue and fin whales observed on a seafloor array in the Northeast Pacific. *Journal of the Acoustical Society of America*, 98(2 part 1):712–721.

Meade, R. (1996). Setting: Geology, hydrology, sediments, and engineering of the Mississippi River. *US Geological Survey Circular*, pages 13–30.

Mellinger, D. and Heimlich, S. (2013). Introduction to the special issue on methods for marine mammal passive acoustics. *Journal of the Acoustical Society of America*, 134(3):2381–2382.

Mellinger, D., Stafford, K., Moore, S., Dziak, R., and Matsumoto, H. (2007). An overview of fixed passive acoustic observation methods for cetaceans. *Oceanography*, 20(4):36–45.

Metropolis, N. and Ulam, S. (1949). The Monte Carlo Method. *Journal of the American Statistical Association*, 44(247):335–341.

Michel, J., Owens, E. H., Zengel, S., Graham, A., Nixon, Z., Allard, T., Holton, W., Reimer, P. D., Lamarche, A., White, M., et al. (2013). Extent and degree of shoreline oiling: Deepwater Horizon oil spill, Gulf of Mexico, USA. *PloS one*, 8(6):e65087.

Minamikawa, S., Watanabe, H., and Iwasaki, T. (2013). Diving behavior of a false killer whale, *Pseudorca crassidens*, in the Kuroshio-Oyashio transition region and the Kuroshio front region of the western North Pacific. *Marine Mammal Science*, 29(1):177–185.

Miyazaki, N., Kusaka, T., and Nishiwaki, M. (1973). Food of *Stenella coeruleoalba*. *Scientific Reports of the Whales Research Institute*, 25:265–75.

Moore, P., Dankiewicz, L., and Houser, D. (2008). Beamwidth control and angular target detection in an echolocating bottlenose dolphin (*Tursiops truncatus*). *Journal of the Acoustical Society of America*, 124(5):3324–3332.

Mountain, D. C., Anderson, D., Voysey, G., and Brughera, A. (2013). The effects of sound in the marine environment workbench: A simulation tool to predict the impact of anthropogenic sound on marine mammals. *The Journal of the Acoustical Society of America*, 133(5):3494–3494.

Müller-Karger, F. E., Smith, J. P., Werner, S., Chen, R., Roffer, M., Liu, Y., Muhling, B., Lindo-Atichati, D., Lamkin, J., Cerdeira-Estrada, S., et al. (2014). Natural Variability of Surface Oceanographic Conditions in the Offshore Gulf of Mexico. *Progress in Oceanography*.

Müller-Karger, F. E., Walsh, J. J., Evans, R. H., and Meyers, M. B. (1991). On the seasonal phytoplankton concentration and sea surface temperature cycles of the Gulf of Mexico as determined by satellites. *Journal of Geophysical Research: Oceans (1978–2012)*, 96(C7):12645–12665.

Mullin, K. (2007). Abundance of cetaceans in the oceanic Gulf of Mexico based on 2003-2004 ship surveys. *Available from: NMFS, Southeast Fisheries Science Center, PO Drawer*, 1207.

Mullin, K. and Fulling, G. (2004). Abundance of cetaceans in the oceanic northern Gulf of Mexico, 1996-2001. *Marine Mammal Science*, 20(4):787–807.

Mullin, K. D., Jefferson, T. A., Hansen, L. J., and Hoggard, W. (1994). First sightings of melon-headed whales (*Peponocephala electra*) in the Gulf of Mexico. *Marine Mammal Science*, 10(3):342–348.

Murphy, S. J., Hurlburt, H. E., and O'Brien, J. J. (1999). The connectivity of eddy variability in the Caribbean Sea, the Gulf of Mexico, and the Atlantic Ocean. *Journal of Geophysical Research: Oceans (1978-2012)*, 104(C1):1431–1453.

Neff, J., Sauer, T., and Hart, A. (2011). Bioaccumulation of hydrocarbons from produced water discharged to offshore waters of the US Gulf of Mexico. In *Produced Water*, pages 441–477. Springer.

Newman, M. E. (2004). Detecting community structure in networks. *The European Physical Journal B-Condensed Matter and Complex Systems*, 38(2):321–330.

Newman, M. E. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103(23):8577–8582.

Nicot, J.-P. (2009). A survey of oil and gas wells in the Texas Gulf Coast, USA, and implications for geological sequestration of CO_2. *Environmental Geology*, 57(7):1625–1638.

Norris, K. S. (1968). The evolution of acoustic mechanisms in odontocete cetaceans. In Drake, E. T., editor, *Evolution and Environment*, pages 297–324. Yale University Press New Haven.

Nosal, E. and Frazer, L. (2006). Track of a sperm whale from delays between direct and surface-reflected clicks. *Applied Acoustics*, 67(11-12):1187–1201.

Nosal, E. and Frazer, L. (2007). Modified pairwise spectrogram processing for localization of unknown broadband sources. *IEEE Journal of Oceanic Engineering*, 32(3):721–728.

Olson, P. and Reilly, S. (2002). Pilot whales *Globicephala melas* and *G. macrorhynchus*. In WF Perrin, B. W. and Thewissen, J., editors, *Encyclopedia of marine mammals*, pages 898–903. Academic Press, San Diego, California.

O'Sullivan, S. and Mullin, K. D. (1997). Killer whales (*Orcinus orca*) in the northern Gulf of Mexico. *Marine Mammal Science*, 13(1):141–147.

Oswald, J., Barlow, J., and Norris, T. (2003). Acoustic identification of nine delphinid species in the eastern tropical Pacific Ocean. *Marine Mammal Science*, 19(1):20–37.

Passow, U., Ziervogel, K., Asper, V., and Diercks, A. (2012). Marine snow formation in the aftermath of the Deepwater Horizon oil spill in the Gulf of Mexico. *Environmental Research Letters*, 7(3):035301.

Pearson, K. (1920). Notes on the history of correlation. *Biometrika*, pages 25–45.

Perrin, W. F. (2001). *Stenella attenuata*. *Mammalian Species*, pages 1–8.

Perrin, W. F. and Hohn, A. A. (1994). Pantropical spotted dolphin *Stenella attenuata*. In *Handbook of marine mammals*, volume 5. Academic Press.

Peterson, C. H., Kennicutt II, M. C., Green, R. H., Montagna, P., Harper, Jr, D. E., Powell, E. N., and Roscigno, P. F. (1996). Ecological consequences of environmental perturbations associated with offshore hydrocarbon production: a perspective on long-term exposures in the Gulf of Mexico. *Canadian Journal of Fisheries and Aquatic Sciences*, 53(11):2637–2654.

Peterson, M. S. and Lowe, M. R. (2009). Implications of cumulative impacts to estuarine and marine habitat quality for fish and invertebrate resources. *Reviews in Fisheries Science*, 17(4):505–523.

Pitman, R. L. and Stinchcomb, C. (2002). Rough-toothed dolphins (*Steno bredanensis*) as predators of mahimahi (*Coryphaena hippurus*). *Pacific Science*, 56(4):447–450.

Porter, M. and Bucker, H. (1987). Gaussian-beam tracing for computing ocean acoustic fields. *Journal of the Acoustical Society of America*, 82(4):1349–1359.

Porter, M. B. (1992). The KRAKEN normal mode program. Technical report, DTIC Document.

Porter, M. B. and Liu, Y.-C. (1994). Finite-element ray tracing. *Theoretical and computational acoustics*, 2.

Powell, L. A. (2007). Approximating variance of demographic parameters using the delta method: a reference for avian biologists. *The Condor*, 109(4):949–954.

R Core Team (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rabalais, N., Turner, R., and Wiseman, W. (2002). Gulf of Mexico hypoxia, aka "The dead zone". *Annual Review of Ecology and Systematics*, 33:235–263.

Rankin, S. and Barlow, J. (2007). Sounds recorded in the presence of Blainville's beaked whales, *Mesoplodon densirostris*, near Hawai'i. *The Journal of the Acoustical Society of America*, 122(1):42–45.

Rasmussen, M. H., Miller, L. A., and Au, W. W. (2002). Source levels of clicks from free-ranging white-beaked dolphins (*Lagenorhynchus albirostris* Gray 1846) recorded in Icelandic waters. *The Journal of the Acoustical Society of America*, 111(2):1122–1125.

Rasmussen, M. H., Wahlberg, M., and Miller, L. A. (2004). Estimated transmission beam pattern of clicks recorded from free-ranging white-beaked dolphins (*Lagenorhynchus albirostris*). *The Journal of the Acoustical Society of America*, 116(3):1826–1831.

Rayleigh, L. (1887). *The Theory of Sound, Vol. II.* Dover Publications, New York,.

Reddy, C. M., Arey, J. S., Seewald, J. S., Sylva, S. P., Lemkau, K. L., Nelson, R. K., Carmichael, C. A., McIntyre, C. P., Fenwick, J., Ventura, G. T., et al. (2012). Composition and fate of gas and oil released to the water column during the Deepwater Horizon oil spill. *Proceedings of the National Academy of Sciences*, 109(50):20229–20234.

Reilly, S. B. (1990). Seasonal changes in distribution and habitat differences among dolphins in the eastern tropical Pacific. *Marine Ecology Progress Series*, 66(1-2):1–11.

Roberts, B. L. and Read, A. J. (2015). Field assessment of C-POD performance in detecting echolocation click trains of bottlenose dolphins (*Tursiops truncatus*). *Marine Mammal Science*, 31(1):169–190.

Roch, M., Baumann-Pickering, S., Batchelor, H., Hwang, D., Sirovic, A., Hildebrand, J., Berchok, C., Cholewiak, D., Munger, L., Oleson, E., Van Parijs, S., Risch, D., and Soldevilla, M. (2013). Tethys: A workbench and database for passive acoustic metadata. In *Oceans - San Diego, 2013*, pages 1–5.

Roch, M., Klinck, H., Baumann-Pickering, S., Mellinger, D., Qui, S., Soldevilla, M., and Hildebrand, J. (2011). Classification of echolocation clicks from odontocetes in the Southern California Bight. *Journal of the Acoustical Society of America*, 129(1):467–475.

Roch, M., Soldevilla, M., Burtenshaw, J., Henderson, E., and Hildebrand, J. (2007). Gaussian mixture model classification of odontocetes in the Southern California Bight and the Gulf of California. *Journal of the Acoustical Society of America*, 121(3):1737–1748.

Roch, M. A., Stinner-Sloan, J., Baumann-Pickering, S., and Wiggins, S. M. (2015). Compensating for the effects of site and equipment variation on delphinid species identification from their echolocation clicks. *The Journal of the Acoustical Society of America*, 137(1):22–29.

Ryerson, T., Aikin, K., Angevine, W., Atlas, E., Blake, D., Brock, C., Fehsenfeld, F., Gao, R.-S., de Gouw, J., Fahey, D., et al. (2011). Atmospheric emissions from the Deepwater Horizon spill constrain air-water partitioning, hydrocarbon fate, and leak rate. *Geophysical Research Letters*, 38(7).

Scott, M. D. and Chivers, S. J. (2009). Movements and diving behavior of pelagic spotted dolphins. *Marine Mammal Science*, 25(1):137–160.

Seber, G. (1982). *The estimation of animal abundance*. Griffin London.

Sen, P. K. (1968). Estimates of the regression coefficient based on Kendall's tau. *Journal of the American Statistical Association*, 63(324):1379–1389.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*, volume 26. CRC press.

Simard, P., Hibbard, A. L., McCallister, K. A., Frankel, A. S., Zeddies, D. G., Sisson, G. M., Gowans, S., Forys, E. A., and Mann, D. A. (2010). Depth dependent variation of the echolocation pulse rate of bottlenose dolphins (*Tursiops truncatus*). *The Journal of the Acoustical Society of America*, 127(1):568–578.

Širović, A., Hildebrand, J., Wiggins, S., McDonald, M., Moore, S., and Thiele, D. (2004). Seasonality of blue and fin whale calls and the influence of sea lee in the Western Antarctic Peninsula. *Deep-Sea Research Part II-Topical Studies in Oceanography*, 51(17-19):2327–2344.

Soldevilla, M., Henderson, E., Campbell, G., Wiggins, S., Hildebrand, J., and Roch, M. (2008). Classification of Risso's and Pacific white-sided dolphins using spectral properties of echolocation clicks. *Journal of the Acoustical Society of America*, 124(1):609–624.

Soldevilla, M., Wiggins, S., and Hildebrand, J. (2010). Spatio-temporal comparison of Pacific white-sided dolphin echolocation click types. *Aquatic Biology*, 9(1):49–62.

Spier, C., Stringfellow, W. T., Hazen, T. C., and Conrad, M. (2013). Distribution of hydrocarbons released during the 2010 MC252 oil spill in deep offshore waters. *Environmental Pollution*, 173:224–230.

Stafford, K., Mellinger, D., Moore, S., and Fox, C. (2007). Seasonal variability and detection range modeling of baleen whale calls in the Gulf of Alaska, 1999-2002. *Journal of the Acoustical Society of America*, 122(6):3378–3390.

Stanley, D. R. and Wilson, C. A. (2000). Seasonal and spatial variation in the biomass and size frequency distribution of the fish associated with oil and gas platforms in the northern Gulf of Mexico. *U.S. Department of the Interior, Minerals Management Service, Gulf of Mexico OCS Region*.

Thiel, H. (1950). A rank-invariant method of linear and polynomial regression analysis, Part 3. In *Proceedings of Koninalijke Nederlandse Akademie van Weinenschatpen A*, volume 53, pages 1397–1412.

Thomas, L., Buckland, S., Rexstad, E., Laake, J., Strindberg, S., Hedley, S., Bishop, J., Marques, T., and Burnham, K. (2010). Distance software: design and analysis of distance sampling surveys for estimating population size. *Journal of Applied Ecology*, 47(1):5–14.

Tiemann, C. O., Porter, M. B., and Frazer, L. N. (2004). Localization of marine mammals near Hawai'i using an acoustic propagation model. *The Journal of the Acoustical Society of America*, 115(6):2834–2843.

Turner, R. E. (1990). Landscape development and coastal wetland losses in the northern Gulf of Mexico. *American Zoologist*, 30(1):89–105.

Turner, R. E. (1997). Wetland loss in the northern Gulf of Mexico: multiple working hypotheses. *Estuaries*, 20(1):1–13.

Turner, R. E. and Rabalais, N. N. (1994). Coastal eutrophication near the Mississippi river delta. *Nature*, 368(6472):619–621.

Urick, R. J. (1967). *Principles of underwater sound for engineers*. Tata McGraw-Hill Education.

US Navy, L. M. (2009). GDEM-V QAV Analysis and Delivery (N62306-08-F-7504). Technical report.

Valentine, D. L., Fisher, G. B., Bagby, S. C., Nelson, R. K., Reddy, C. M., Sylva, S. P., and Woo, M. A. (2014). Fallout plume of submerged oil from Deepwater Horizon. *Proceedings of the National Academy of Sciences*, 111(45):15906–15911.

Van Parijs, S. M., Smith, J., and Corkeron, P. J. (2002). Using calls to estimate the abundance of inshore dolphins: a case study with Pacific humpback dolphins *Sousa chinensis*. *Journal of Applied Ecology*, 39(5):853–864.

Venn-Watson, S., Garrison, L., Litz, J., Fougeres, E., Mase, B., Rappucci, G., Stratton, E., Carmichael, R., Odell, D., Shannon, D., et al. (2015). Demographic clusters identified within the Northern Gulf of Mexico common bottlenose dolphin (*Tursiops truncatus*) Unusual Mortality Event: January 2010-June 2013. *PloS one*, 10(2):e0117248.

Ward, J., Jarvis, S., Moretti, D., Morrissey, R., DiMarzio, N., Johnson, M., Tyack, P., Thomas, L., and Marques, T. (2011). Beaked whale (*Mesoplodon densirostris*) passive acoustic detection in increasing ambient noise. *Journal of the Acoustical Society of America*, 129(2):662–669.

Waring, G., Josephson, E., Maze-Foley, K., and Rosel, P. (2013). U.S. Atlantic and Gulf of Mexico Marine Mammal Stock Assessments – 2012. Technical report, NOAA Tech Memo NMFS NE 223.

Watkins, W. A. and Schevill, W. E. (1971). Four hydrophone array for acoustic three-dimensional location. Report Reference No. 71-60, Woods Hole Oceanographic Institution.

Watkins, W. A. and Schevill, W. E. (1974). Listening to Hawaiian Spinner Porpoises, *Stenella* cf. *longirostris*, with a Three-Dimensional Hydrophone Array. *Journal of Mammalogy*, 55(2):319–328.

Wells, R. S., Early, G. A., Gannon, J. G., Lingenfelser, R. G., and Sweeney, P. (2008). Tagging and tracking of rough-toothed dolphins (*Steno bredanensis*) from the March 2005 mass stranding in the Florida Keys. Technical Memorandum NMFS-SEFSC 574, NOAA.

Wells, R. S., Manire, C. A., Byrd, L., Smith, D. R., Gannon, J. G., Fauquier, D., and Mullin, K. D. (2009). Movements and dive patterns of a rehabilitated Risso's dolphin, *Grampus griseus*, in the Gulf of Mexico and Atlantic Ocean. *Marine Mammal Science*, 25(2):420–429.

Wiggins, S., Frasier, K., Henderson, E., and Hildebrand, J. (2013). Tracking dolphin whistles using an autonomous acoustic recorder array. *Journal of the Acoustical Society of America*, 133(6):3813–3818.

Wiggins, S. and Hildebrand, J. (2007). High-frequency Acoustic Recording Package (HARP) for broad-band, long-term marine mammal monitoring. In *Symposium on Underwater Technology and Workshop on Scientific Use of Submarine Cables and Related Technologies, 2007*, pages 551 – 557. IEEE, Institute of Electrical and Electronics Engineers, Tokyo, Japan.

Wiggins, S., McDonald, M., and Hildebrand, J. (2012). Beaked whale and dolphin tracking using a multichannel autonomous acoustic recorder. *Journal of the Acoustical Society of America*, 131(1):156–163.

Williams, R., Gero, S., Bejder, L., Calambokidis, J., Kraus, S. D., Lusseau, D., Read, A. J., and Robbins, J. (2011). Underestimating the damage: interpreting cetacean carcass recoveries in the context of the Deepwater Horizon/BP incident. *Conservation Letters*, 4(3):228–233.

Wilson, R., Nelson, J., Balmer, B., Nowacek, D., and Chanton, J. (2013). Stable isotope variation in the northern Gulf of Mexico constrains bottlenose dolphin (*Tursiops truncatus*) foraging ranges. *Marine biology*, 160(11):2967–2980.

Wiseman, W. J., Sturges, W., and Kumpf, H. (1999). Physical oceanography of the Gulf of Mexico: Processes that regulate its biology. In Kumpf, H., Steidinger, K., and Sherman, K., editors, *The Gulf of Mexico Large Marine Ecosystem: Assessment, sustainability and management*, pages 77–92. Blackwell Science, New York.

Wormuth, J., Ressler, P., Cady, R., and Harris, E. (2000). Zooplankton and micronekton in cyclones and anticyclones in the northeast Gulf of Mexico. *Gulf of Mexico Science*, 18(1):23–34.

Würtz, M., Poggi, R., and Clarke, M. R. (1992). Cephalopods from the stomachs of a Risso's dolphin (*Grampus griseus*) from the Mediterranean. *Journal of the Marine Biological Association of the United Kingdom*, 72(04):861–867.

Zimmer, W. M. (2011). *Passive acoustic monitoring of cetaceans*. Cambridge University Press New York.

Zimmer, W. M., Harwood, J., Tyack, P. L., Johnson, M. P., and Madsen, P. T. (2008). Passive acoustic detection of deep-diving beaked whales. *The Journal of the Acoustical Society of America*, 124(5):2823–2832.

Zimmer, W. M., Tyack, P. L., Johnson, M. P., and Madsen, P. T. (2005). Three-dimensional beam pattern of regular sperm whale clicks confirms bent-horn hypothesis. *The Journal of the Acoustical Society of America*, 117(3):1473–1485.