Final Project Report: Marine Species Density Data Gap Assessments and Update for the AFTT Study Area, 2018-2020 (Opt. Year 3*)

Cooperative Agreement Number N62470-15-2-8003

Jason J. Roberts, Robert S. Schick, and Patrick N. Halpin

Marine Geospatial Ecology Lab (MGEL), Duke University

Document Version 1.4 – 2020-08-10

This document should be cited:

Roberts JJ, Schick RS, Halpin PN (2020) Final Project Report: Marine Species Density Data Gap Assessments and Update for the AFTT Study Area, 2018-2020 (Option Year 3). Document version 1.4. Report prepared for Naval Facilities Engineering Command, Atlantic by the Duke University Marine Geospatial Ecology Lab, Durham, NC.

1. Introduction

In the United States, national laws protect marine mammals. The Marine Mammal Protection Act (MMPA) prohibits intentional or incidental killing, injuring, or disturbance of marine mammals and specifies the circumstances and rules under which permits may be issued for activities that have the potential to incidentally injure or disturb marine mammals. The Endangered Species Act (ESA) prohibits harm to species threatened with extinction and requires conservation of their habitat. The National Environmental Policy Act (NEPA) specifies a process by which U.S. federal government agencies must evaluate the potential environmental effects of their actions, consider alternatives, and conduct public reviews. For certain actions this may culminate in the development of an Environmental Impact Statement (EIS). Agency actions that involve decisions to issue permits under the MMPA or ESA are often subject to this process.

The U.S. Navy is responsible for compliance with a suite of federal environmental and natural resources laws and regulations that apply to the marine environment, including the MMPA, the ESA, the Magnuson-Stevens Fishery Conservation and Management Act, the Marine Protection, Research and Sanctuaries Act (MPRSA), the Clean Water Act (CWA), Executive Order 13089 on Coral Reef Protection, and NEPA/Executive Order 12114 (EO 12114). Additionally, federal activities that have the potential to affect the state coastal zone are required to be consistent with respective state coastal zone management plans mandated by the Coastal Zone Management Act (CZMA).

To evaluate the potential effects of proposed training, testing, and marine construction activities on marine mammal populations, the Navy requires a detailed understanding of the spatiotemporal distributions of these populations. To facilitate development of the Navy's Atlantic Fleet Training and Testing (AFTT) Phase III EIS, the Navy funded us (the Duke Marine Geospatial Ecology Lab, or MGEL) to develop density surface models (Hedley & Buckland 2004; Miller et al. 2013) for all cetacean and seal species sighted in the AFTT study area (Figure 1) during scientific surveys conducted with protocols compatible with density surface modeling methodology. This culminated in the development and publication of regional models for the U.S. Atlantic coast and northern Gulf of Mexico (Roberts et al. 2016a) and the wider AFTT Study Area (Mannocci et al. 2017b).

* Option Year 3 of this Cooperative Agreement spanned 2018-2019, but the work reported here extended into 2020, part of Option Year 4. We plan to issue a separate report covering the remainder of Option Year 4, when that work is complete.

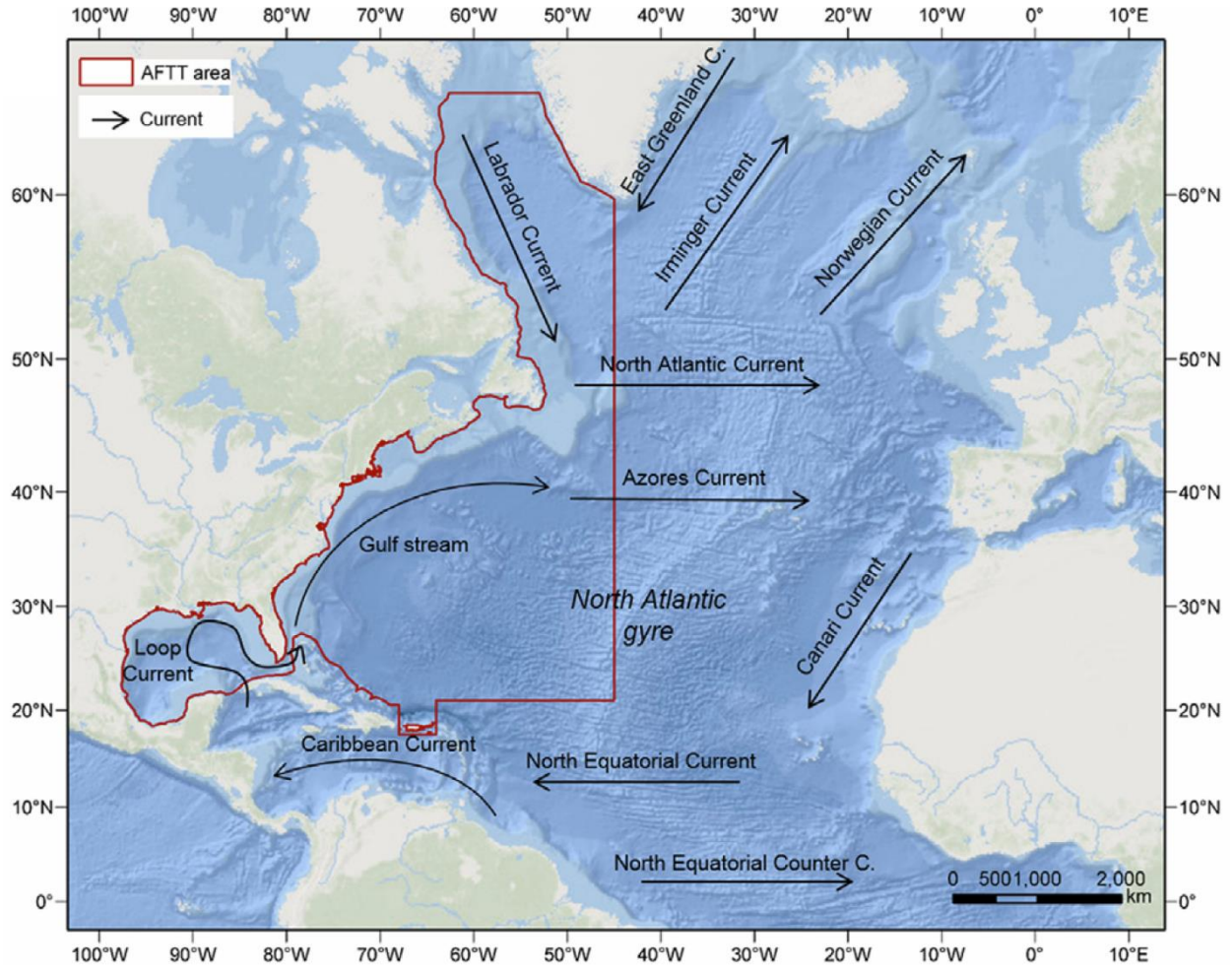


Figure 1. U.S. Navy AFTT Study Area with major current systems. Reproduced from Mannocci et al. (2017b).

The regional Phase III models (Roberts et al. 2016a) spanned areas adjacent to the continental U.S. for which sufficient survey data were available to avoid large geographic extrapolations, while the AFTT-wide models (Mannocci et al. 2017b) addressed the broad unsurveyed areas of the AFTT study area where extrapolations were required. We delivered all of those results to the Navy for use in a new version of the Navy Marine Species Density Database (NMSDD), the authoritative source of marine species density data maintained by the Navy, for the AFTT Phase III EIS (Roberts 2015; Roberts et al. 2015; U.S. Department of the Navy 2017). The Navy then developed the EIS and obtained from the National Marine Fisheries Service (NMFS) an MMPA Letter of Authorization (LOA) to conduct testing and training activities for the period of November 14, 2018 through November 13, 2023 (83 FR 57076).

In 2015, after we completed the Phase III models and the Navy began developing the EIS, the Navy initiated the project "Marine Mammal Density Gap Assessments and Update for the AFTT Study Area", a Cooperative Agreement (#N62470-15-2-8003) with us to prepare revised models using newly available data and methodology. These "second generation" models would be used in the AFTT Phase IV EIS, supplanting those from Phase III. Accordingly, in the Base Year (2015-2016), Option Year 1 (2016-2017), and Option Year 2 (2017-2018) of the Cooperative Agreement, we acquired and integrated a large quantity of additional marine mammal survey data and updated models for 22 taxa for the Atlantic coast, known as the East Coast (EC) region in many of our reports (Roberts et al. 2016b, 2017, 2018). The surveys used for these models extended into 2016 and most updated models were based on data spanning the period 1998-2016, with models for several infrequently-sighted species extending into earlier years. We updated the NMSDD with these new models and released them to NMFS for review and use in other contexts.

At that time, the Navy planned to utilize the 22 updated models to develop the Phase IV EIS, along with additional updates to be made during Option Years 3 and 4. However, in summer of 2018, near the end of Option Year 2, the John S. McCain National Defense Authorization Act for Fiscal Year 2019 amended the MMPA and extended the maximum time period for regulations related to military readiness from five to seven years. The Navy subsequently obtained a revised LOA for the AFTT Phase III analysis with a new end date of November 13, 2025 (84 FR 70712). Given these additional two years, the Navy requested that we prepare a new set of updated models that incorporated additional marine mammal surveys beyond what we had utilized so far. Accordingly, in Option Year 3 of this Cooperative Agreement (spanning August 2018 to August 2019), we suspended production of updated models and canvassed our collaborators for additional data. Then, in Option Year 4 (which started in August 2019), we completed the integration of these data into our modeling infrastructure, and as a test of the new data and associated methodological improvements, we prepared an updated model for the North Atlantic right whale for the EC region.

2. Timeline for Phase IV Density Models

The Navy indicated their Phase IV environmental analysis was likely to begin in early 2022 and that all density models to be used within it must be completed by that time. They requested updates to all models developed for Phase III, including the regional East Coast (EC) and Gulf of Mexico (GOM) models originally published by Roberts et al. (2016a)—some of which were updated in Option Years 1 and 2 (Roberts et al. 2017, 2018)—and the AFTT-wide models published by Mannocci et al. (2017b).

Working backwards from that deadline, accounting for the time required to update all these models, we developed the following timeline in collaboration with Navy Facilities Engineering Command (NAVFAC) Atlantic:

1. In Option Year 3 (2018-2019), we would incorporate new survey data that would expand spatial coverage of key areas and extend our database through the end of 2018. We would also reprocess a number of older surveys to implement several improvements to our data preparation methodology identified in Option Years 1 and 2. (This planned work ended up extending into Option Year 4 (2019-2020), as not all of the new data we desired was available during Option Year 3, and some collaborators proactively contributed data extending into 2019, so we went ahead and processed those data.)

2. In Option Years 3 and 4 (2018-2020), we would also implement various improvements to our modeling methodology identified during Option Years 1 and 2, and then update a single density model as a means to test out the new data and methods, and fix any problems we discovered. We selected the North Atlantic right whale (NARW) as the species to be updated, as it is one of the species of most critical conservation concern within the AFTT study area. This effort dovetailed with a strong interest from NOAA, who sought specific model improvements to assist with their efforts in managing right whales.

3. In the first part (2020-2021) of a new cooperative agreement, we would update the regional GOM models, using new surveys from NOAA's Gulf of Mexico Marine Assessment Program for Protected Species (GoMMAPPS) and the Natural Resources Damage Assessment (NRDA) program that followed the 2010 *Deepwater Horizon* oil spill. Several of these surveys were incorporated into our database during Option Year 3; the remainder would be incorporated in 2020.

4. In the second part (2021-2022) of the new cooperative agreement, we would update the regional EC models for species other than NARW, and the AFTT-wide models. Subject to available time and data, we would incorporate a final tranche of new survey data beforehand, and update the NARW model again.

This report documents work done for item 1 and part of item 2, above.

3. Improvements Requested to the Right Whale Model

As noted above, a primary objective of the modeling exercise documented in this report was to test the data we added and methods we changed after Option Year 2 by completing a full update to one model, North Atlantic right

whale, and fix any problems we discovered. This full run-through of the modeling workflow would facilitate rapid production of the remaining models in the run-up to the Navy's assessment of marine mammal exposures for the AFTT Phase IV EIS. Beyond that goal, we had a number of other goals specific to the right whale model itself. These were mainly based on feedback we received to the prior iterations of that model, known as v7—which was a major update to the Phase III model, completed in 2017 and documented by Roberts et al. (2017)—and v8—which was a minor update undertaken in 2019 to "fill in" Cape Cod Bay with newly-available abundance estimates (Ganley et al. 2019) and correct some minor edge effects present in v7. The v8 model was utilized by NOAA for development of regulations designed to reduce the risk of right whale entanglements in trap and pot fishing gear, and through that process we received a lot of feedback from right whale experts on how the model might be improved. The improvements requested for the new model included:

1. To account for recent changes to right whale distributions, the model should be based on survey data that extend through 2018, or later if possible. In addition to updates from existing collaborators, data should be solicited from two survey programs not used in our prior models:

    a. Aerial surveys of the Massachusetts and Rhode Island Wind Energy Areas led by New England Aquarium (Kraus et al. 2016), spanning 2011-2015 and 2017-2018.

    b. Recent surveys of New York waters, either traditional aerial surveys initiated by the New York State Department of Environmental Conservation (NYS-DEC) in 2017, or digital aerial surveys initiated by the New York State Energy Research and Development Authority (NYSERDA) in 2016, or both.

2. To reflect a view in the right whale community that spatiotemporal patterns in right whale density changed around the time the species entered a decline in approximately 2010, we should consider basing the new model only on recent years. NOAA specifically requested that we build contrasting "before" and "after" models that might illustrate shifts in density, as well as a model spanning both periods, and specifically consider which model would best represent right whale density in the near future.

3. To facilitate better application of the model to near-shore management questions, NOAA and others requested that we extend the spatial extent of the model farther in-shore, particularly north of New York.

4. Many interested parties requested that we increase the resolution of the model beyond 10 km, if possible.

The new model, known as v9 (or version 9), addressed all these requests, as well as various other minor requests not noted here.

4. Survey Data Integrated or Reprocessed

4.1. New survey data integrated

In preparation for completing this analysis, we canvassed continuing collaborators for incremental updates to ongoing survey programs and solicited new collaborations with organizations conducting important surveys compatible with our methodology (see section 5.1) but that we had not previously utilized. Table 1 summarizes the extant survey data included in our prior models and the additional data that was incorporated for this analysis. In total, we added 474,000 km (2416 h) of aerial effort to the extant 2,271,000 km (11,800 h) previously available, and added 12,000 km (691 h) of shipboard effort to the extant 56,000 km (3162 h), boosting the data available for each platform by 21%. Most of the added data were collected in 2016-2019. All continuing collaborators contributed additional data except NJDEP, which did not have an active survey program. Notably, NEFSC, SEFSC, and the NARW EWS teams submitted data through spring of 2019, allowing us to extend the model through that season (see section 5.3.1). Several new collaborators joined, providing new coverage in critical areas. These included: New England Aquarium, which previously contributed data in the southeast and now contributed the NLPSC and MMS-WEA surveys of the northeast, where right whales have been aggregating to feed; NYS-DEC, which contributed two years of monthly baseline monitoring surveys of New York waters; and HDR, which contributed surveys of the Navy's Norfolk Canyon area.

Table 1. Summary of survey effort available for this analysis, 2003-2019. Surveys conducted before April 2003 were not included in this analysis (see section 5.3.1.1). "Extant" effort was included in our prior models. "Added" effort was used by us for the first time in the model presented here. Only "on effort" survey transects are included. Some of this effort (e.g. collected at high Beaufort scales) was subsequently excluded from the analysis (see section 5.2). Table 2 defines acronyms and Table 3 lists publications documenting each survey program. Section 5.3 includes maps of the aggregate survey effort.

			Survey Effort					
			Years		Length (1000 km)		Duration (h)	
Platform	**Organization**	**Program**	**Extant**	**Added**	**Extant**	**Added**	**Extant**	**Added**
Aerial	FWRI	NARW EWS	2003/04-2015/16	2016/17-2018/19	531	114	2789	599
	HDR	Navy Monitoring: Norfolk Canyon		2018		6		
	NEAq	NARW EWS	2003/04-2009/10		236		1259	
		NLPSC		2011-2015		43		227
		MMS-WEA		2017-2018		19		101
	NEFSC	Pre-AMAPPS	2004-2008		35		170	
		AMAPPS	2010-2014	2016-2019	42	48	220	247
		NARWSS	2003-2016	2017-2019	431	41	2248	214
	NJDEP	NJEBS	2008-2009		11		56	
	NYS-DEC	NYBWM		2017-2018		57		291
	SEFSC	MATS	2003-2005		14		71	
		AMAPPS	2010-2015	2016-2019	67	44	347	229
	UNCW	Navy Monitoring: Cape Hatteras	2011-2016	2017	35	3	173	12
		Navy Monitoring: Jacksonville	2009-2016	2017	88	5	422	22
		Navy Monitoring: Norfolk Canyon	2015-2016	2017	11	4	55	19
		Navy Monitoring: Onslow Bay	2007-2011		49		240	
		Right Whale Surveys	2005-2008		115		575	
	VAMSC	MD DNR	2013-2015		16		83	
		VA CZM	2012-2015		21		102	
		Navy Monitoring: VACAPES		2016-2017		20		94
	WLT/SSA/ CMARI	NARW EWS (Skymaster)	2004/05-2012/13		279		1473	
		NARW EWS (Twin Otter)	2003/04-2015/16	2016/17-2018/19	290	70	1517	361
				Aerial Total	**2271**	**474**	**11800**	**2416**
Shipboard	NEFSC	Pre-AMAPPS	2004,2007		4	2	224	112
		AMAPPS	2011,2013,2014	2016	12	4	682	240
	NJDEP	NJEBS	2008-2009		14		728	
	SEFSC	Pre-AMAPPS	2004-2006		15		873	
		AMAPPS	2011,2013	2016	11	6	656	339
				Shipboard Total	**56**	**12**	**3162**	**691**

Table 2. Acronyms of collaborating organizations and survey programs used in this analysis.

Type	Acronym	Meaning
Organization	FWRI	Fish and Wildlife Research Institute (FWRI) of the Florida Fish and Wildlife Conservation Commission (FWC)
	HDR	HDR, Incorporated
	NEAq	New England Aquarium
	NEFSC	NOAA Northeast Fisheries Science Center
	NJDEP	New Jersey Department of Environmental Protection
	NOAA	National Oceanic and Atmospheric Administration
	NYS-DEC	New York State Department of Environmental Conservation
	SEFSC	NOAA Southeast Fisheries Science Center
	UNCW	University of North Carolina Wilmington
	VAMSC	Virginia Aquarium & Marine Science Center
	WLT/SSA/CMARI	Wildlife Trust / Sea to Shore Alliance / Clearwater Marine Aquarium Research Institute
Survey Program	AMAPPS	Atlantic Marine Assessment Program for Protected Species
	MATS	Mid-Atlantic Tursiops Survey
	MD DNR	Maryland Department of Natural Resources Surveys
	MMS-WEA	Marine Mammal Surveys of the Wind Energy Areas
	NARW EWS	North Atlantic Right Whale Early Warning System
	NARWSS	North Atlantic Right Whale Sighting Surveys
	NJEBS	New Jersey Ecological Baseline Study
	NLPSC	Northeast Large Pelagic Survey Collaborative
	NYBWM	New York Bight Whale Monitoring Surveys
	VA CZM	Virginia Coastal Zone Management Surveys
	VACAPES	U.S. Navy Virginia CAPES Operations Area

Table 3. Publications documenting the survey programs used in this analysis. Where possible, a peer-reviewed scientific journal publication is given. One or more relevant technical reports, not subject to formal peer review, may additionally or alternatively be given.

Organization	Program	Publications
FWRI	NARW EWS	(Gowan & Ortega-Ortiz 2014)
HDR	Navy Monitoring: Norfolk Canyon	(Cotter 2019)
NEAq	NARW EWS	(Gowan & Ortega-Ortiz 2014)
	NLPSC	(Kraus et al. 2016; Leiter et al. 2017; Stone et al. 2017)
	MMS-WEA	(Quintana et al. 2018)
NEFSC	Pre-AMAPPS	(Palka 2006)
	AMAPPS	(Palka et al. 2017; Chavez-Rosales et al. 2019)
	NARWSS	(Cole et al. 2007)
NJDEP	NJEBS	(Whitt et al. 2013, 2015)
NYS-DEC	NYBWM	(Tetra Tech & Smultea Sciences 2018; Tetra Tech & LGL 2019)
SEFSC	MATS	None available
	Pre-AMAPPS (shipboard)	(Garrison 2006; Garrison et al. 2010)
	AMAPPS	(Palka et al. 2017; Chavez-Rosales et al. 2019)
UNCW	Navy Monitoring: Cape Hatteras	(McLellan et al. 2018)
	Navy Monitoring: Jacksonville	(Foley et al. 2011)
	Navy Monitoring: Norfolk Canyon	(McAlarney et al. 2018)
	Navy Monitoring: Onslow Bay	(Read et al. 2014)
	Right Whale Surveys	(Torres et al. 2005)
VAMSC	MD DNR	(Barco et al. 2015)
	VA CZM	(Mallette et al. 2014, 2015)
	Navy Monitoring: VACAPES	(Mallette et al. 2017)
WLT/SSA/CMARI	NARW EWS (both aircraft)	(Gowan & Ortega-Ortiz 2014)

4.2. Reprocessing of NARWSS surveys

For the surveys utilized in our analysis, the large majority of right whale sightings reported north of 40 °N were from the North Atlantic Right Whale Sighting Surveys (NARWSS) conducted by the NOAA Northeast Fisheries Science Center (NEFSC). Right whale densities estimated for the northeast depend strongly on these data, so in each modeling cycle, we try to set aside some time to improve our handling of them. During this iteration, we were able to implement four key improvements:

1. Perpendicular sighting distances are now estimated from NARWSS "actual position" (AP) records rather than vertical angles, when the AP records are collected sufficiently quickly after the aircraft breaks from the trackline to circle the sighting. This reduced the amount of heaping in the perpendicular distances, and probably their accuracy as well, yielding a better detection function.

2. Immediate resightings of animals that were just circled are now better recognized and excluded from the analysis. The NARWSS database does not maintain a field specifically flagging such sightings, but they can be identified by searching the sighting comments for specific phrases. After this change was implemented, approximately 180 right whale sightings reported between 2003-2016 were excluded; these represented sightings that were counted twice in our prior models, and thus biased density upward (probably by less than 10%). (We did not run our old code on the 2017-2019 data that were used for the first time in the analysis documented in this report, so we do not know how many sightings from 2017-2019 it would have failed to exclude.)

3. Valid sightings reported at the end of transect lines are now better recognized and included in the analysis. For these sightings, the NARWSS team decided before breaking that they would not return to effort after circling, and logged a sequence of events that made it difficult to automatically determine whether the sightings reported during circling were prompted by the break, and thus should be retained, or were made in response to cues observed while off effort, and thus should be discarded. After this change was implemented, as many as 30 right whale sightings between 2003-2016 were included; these were sightings that were mistakenly excluded from our prior models, and thus biased density downward (probably by less than 5%). (We did not run our old code on the 2017-2019 data that were used for the first time in the analysis documented in this report, so we do not know how many sightings from 2017-2019 it would have failed to include.)

4. To boost survey coverage of inshore areas, as requested by right whale experts and model users, we included on-watch "transit" effort, in which the NARWSS team performed their normal protocol while transiting to or from a designated survey area.

Implementing these improvements, particularly #1, required a complete overhaul of our NARWSS-processing code. After completing this, we reprocessed the entire NARWSS database and manually reviewed every day of surveying (a very time-consuming process). In the following section, we describe the details of improvement #1, for the purpose of documenting our new approach to estimating perpendicular sighting distances. Improvements #2-4 do not require further documentation. Finally, we note that for all analysis presented in this report, we excluded all effort and sightings made in the Gulf of St. Lawrence. Although NARWSS conducted some surveying there in recent years, most flights were directed surveys of known right whale aggregations, and the Gulf of St. Lawrence is currently outside the Navy's AFTT study area. We remain highly interested in including the Gulf of St. Lawrence in future models, but this must be done in collaboration with experts from Canada and a discussion of the prospects for such an effort is beyond the scope of this report.

4.2.1. New algorithm for estimating perpendicular sighting distances for NARWSS

Under the NARWSS survey protocol, after an observer calls out a sighting, the aircraft continues on the trackline until the sighting is perpendicular to the aircraft. The observer then logs an inclination angle measured from hatch marks on the observer's bubble window. The protocol calls for angles to be estimated to the nearest degree, except when the angle is within 1 degree of the horizon, in which case it should be estimated to the nearest 0.1 degree. After this is done, the aircraft breaks from the trackline and circles the sighting to confirm the species identification, group size, and collect photographs.

In prior models, all perpendicular distances were estimated using the inclination angles (and the altitude of the aircraft). This resulted in several problems that complicated the fitting of detection functions. A substantial fraction of angles were rounded to the nearest 5° or 10°, necessitating the binning of distances when fitting the detection function (Buckland et al. 2001). Although the binning was less frequent at inclinations > 75° (i.e. less than 15° down from the horizon), a difference of 1° at these angles can correspond to 100s of meters (m) difference in perpendicular distance. Because NARWSS sighted many distant whales, this resulted in a distribution of distances with spikes at the 1° rounding points (Figure 2). This distribution did not resemble the classic "half a bell curve" shape, in which the count of sightings monotonically decreases as distance increases, that is commonly expected when modeling detection functions. Indeed, the highest number of sightings were recorded at approximately 2600 m, corresponding to an inclination angle of 85° at the target altitude of 750 feet (ft) (Figure 2).

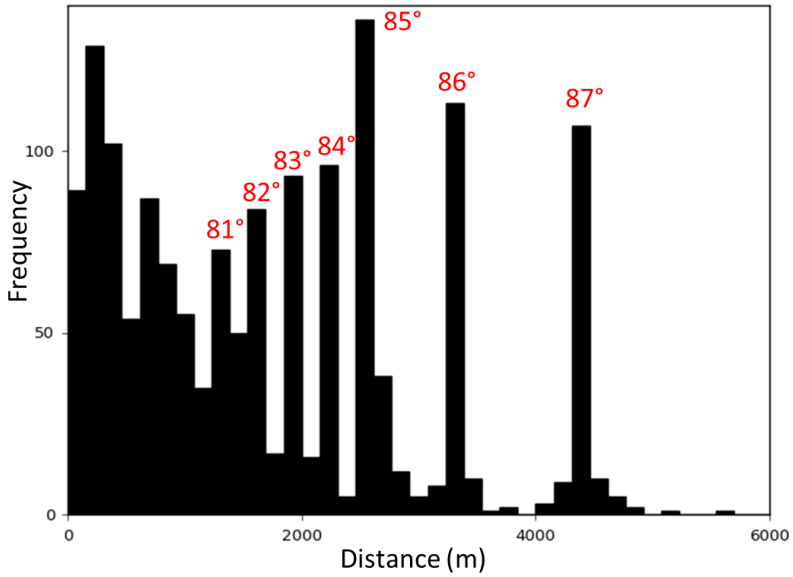


Figure 2. Perpendicular distances estimated from inclination angles for right whales sighted on NARWSS Twin Otter surveys, 2003-2016. Red labels show spikes at 1° rounding intervals. (The small, non-zero numbers of sightings on either side of these spikes results from variation in altitude, not from observers reporting angles in fractional degrees.)

A second problem was an apparent falloff in detections close to the trackline (Figure 2). This was not expected because NOAA's Twin Otters have bubble windows and observers report that large whales close to the trackline are easy to see. However, the NARWSS protocol directs observers to focus attention 1 nautical mile (nmi) out from the trackline to not miss distant right whales (T. Cole, pers. comm.). This could conceivably lead to missing whales close to the trackline, so in prior models we had interpreted the falloff in detections as missed whales and applied left-truncation (Buckland et al. 2001) at an inclination angle of 15°, corresponding to a perpendicular distance of approximately 62 m (Figure 3). This resulted in a loss of about 1% of the sightings.

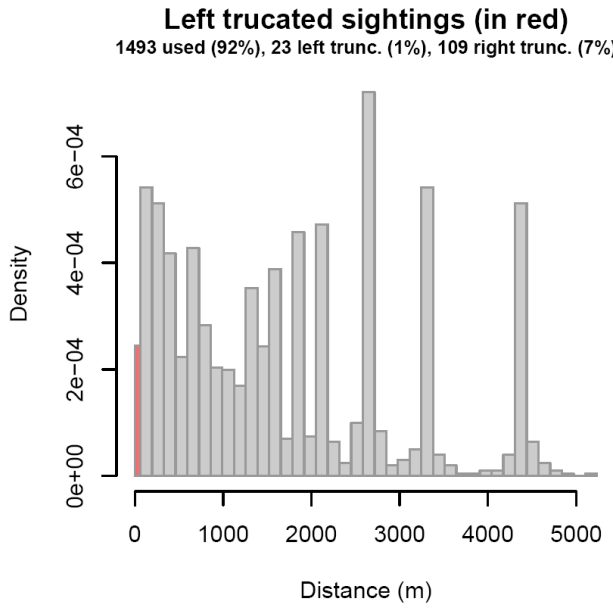


Figure 3. Sightings of right whales that were left truncated (red bar) from the NARWSS Twin Otter surveys, 2003-2016, in prior models.

To address the spikes, had we binned the data using breaks halfway between the heaping angles (Buckland et al. 2001), either at 5° or 10° increments for inclinations < 80°, and at 1° increments for inclinations ≥ 80° and fitted a hazard-rate detection function (Figure 4). This exhibited a low CV but other diagnostics indicated goodness-of-fit problems. Fitting a detection function without the binning scheme had yielded a very similar result (Figure 5).

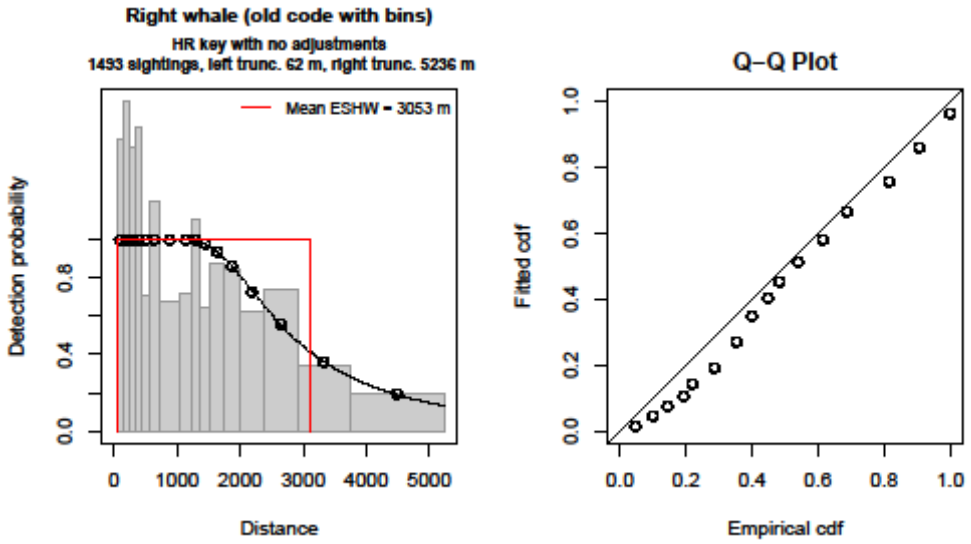


Figure 4. Hazard-rate detection function fitted to the NARWSS right whale sightings shown in Figure 3. A detection function (not shown) similar to this, but with covariates, was used in prior right whale models.

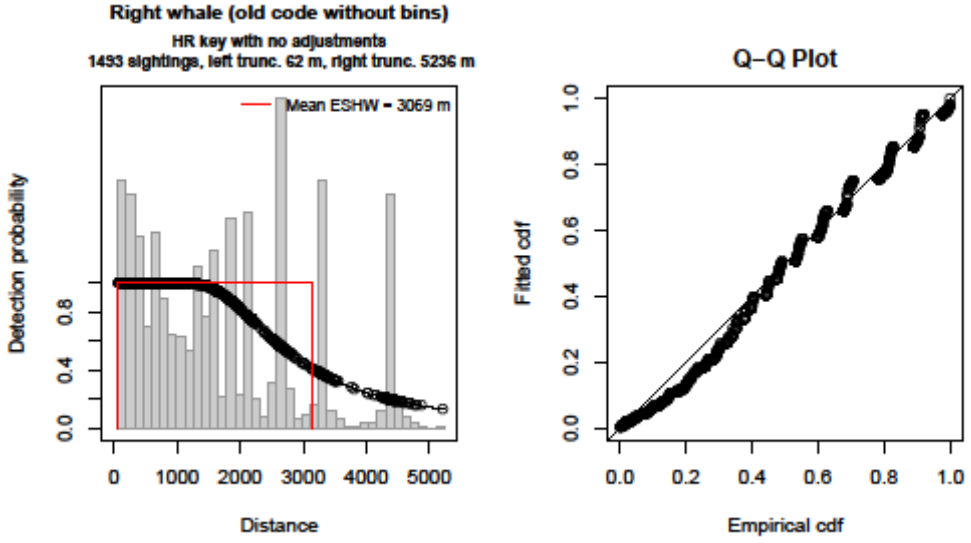


Figure 5. Alternative detection function fitted to the NARWSS right whale sightings shown in Figure 3, but without the binning scheme shown in Figure 4 designed to address the heaped detection angles.

Starting in 2003, NARWSS began collecting the "actual positions" (APs) of groups sighted during circling, by recording GPS coordinates when the aircraft was close to each group (Figure 6). To avoid disturbing the animals, the aircraft would not fly directly over the whales at low altitude, but instead collect the coordinates while passing nearby (C. Khan, pers. comm.). When observers believed the coordinates were especially accurate—typically, when they reached the sighted group quickly—they would additionally log a comment "good for de" [distance estimation]. When

they believed the coordinates were especially likely to be inaccurate—for example, when a large amount of time elapsed between making the sighting and collecting the coordinates—they would log the comment "not good for de".

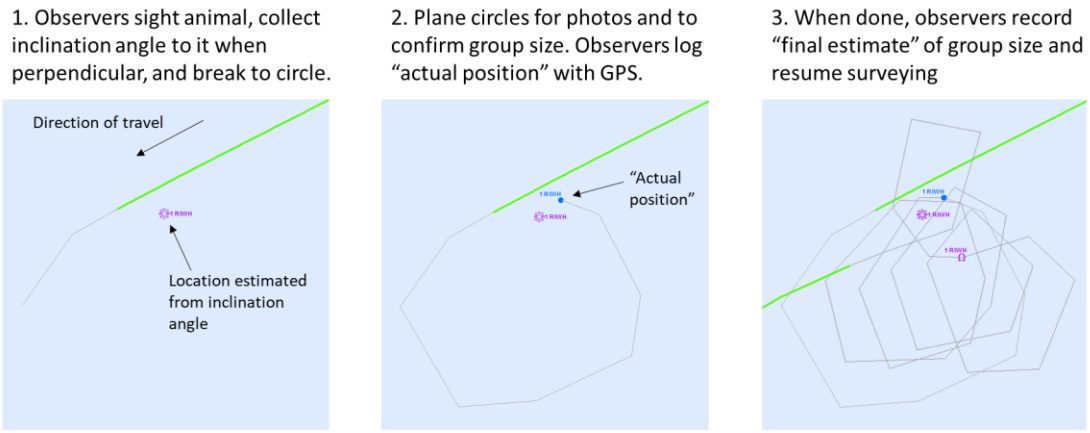


Figure 6. Example sequence of events showing NARWSS observers collecting the inclination angle and actual position of a right whale sighting. This example is from the 20 June 2014 flight of Twin Otter N57RF at 11:41 AM. Green lines are periods when observers were "on watch". Gray lines are periods when observers were circling the sighting, focused on documenting and photographing the animals there.

Although the actual positions were not always collected, and sometimes insufficient data were logged to allow an actual position to be matched with the original sighting cue, there were sufficient data to allow perpendicular distances to right whales to be derived from actual positions for a majority of sightings, if we wanted to do so. However, whales were often not stationary and sometimes a long delay occurred before an AP record was logged, usually because the whale was far from the trackline. This raised the possibility that whales moved substantially from their original position before their locations were obtained, leading to a larger error in position than would be obtained by using the inclination angle instead. After discussing these complications with the NARWSS team, we developed the following algorithm for estimating perpendicular distances and applied it to all NARWSS data collected in 2003 or later:

1. Find the AP record reported during the circle that matches the cue. First match by sighting number. If that is not possible (e.g. the sighting number was not recorded in the comments of the AP record), use the first AP reported during the circle that was of the cued species (or of a more precise taxonomic identification, e.g. an AP of RIWH following a cue of UNLW).

2. If we could not find an AP record, or the one we found was marked "not good for de", use the distance estimate based on the inclination angle.

3. Otherwise, if the AP record was marked "good for de", use the AP-based distance estimate.

4. Otherwise, compare the error that would plausibly result from using the AP-based estimate to the inclination-based estimate, and choose whichever is lower:

    a. First estimate the error in the actual position that would plausibly result from the animal moving before the AP record was logged. Assume that whales move directly toward or away from the trackline (the worst case scenario) at a constant speed of 3.3 km/h, a plausible speed for a travelling whale (C. Khan, pers. comm.), during the entire time between the logging of the cue and the AP. Assume that the whale's AP was displaced from its original location by this amount, which therefore represents the error in perpendicular distance.

    b. Then estimate the plausible error in angle-based distance. Assume that observers might err by 1° in either direction (toward or away from the horizon), a conservative amount based on the NARWSS team's assertion that observers are highly experienced and diligent in recording

accurate angles (C. Khan, pers. comm.). Calculate the mean difference in perpendicular distance that would result from the two possible errors (1° toward the horizon, or 1° away from the horizon) and use that as the error.

c. If the inclination-based error is lower, use the inclination-based perpendicular distance. Otherwise use the AP-based perpendicular distance.

For sightings close to the trackline (e.g. Figure 7), this algorithm often selected the inclination-based distance, as the error in perpendicular distance resulting from a 1° inclination error is very small for inclinations to animals directly below the plane. Conversely, for sightings far from the trackline (e.g. Figure 8), the algorithm often selected the AP-based distances, owing to the large errors in perpendicular distance that result from a 1° error for distant animals. For distant sightings, even when the AP was logged several minutes after the initial cue, the assumed movement of the whale was usually still much smaller than the assumed error in the angle-based perpendicular distance (e.g. Figure 8).

Across the 2003-2016 period, which was used to develop this algorithm, we found that the presumed errors in perpendicular distances estimated from inclination angles were usually smaller than those estimated from AP records for perpendicular distances of about 0-800 m (Figure 9). For distances of about 800-1800 m, errors were lower for each method about half of the time. For distances greater than 1800 m, there was usually so much error in the angle-based estimates that APs were logged in sufficiently short time for AP-based estimates to exhibit smaller errors.

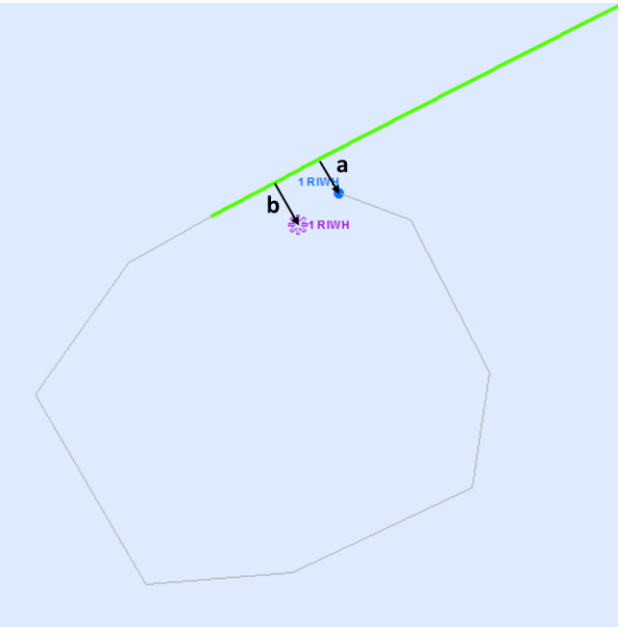


Figure 7. Estimation of perpendicular distances for the example shown in Figure 6. The initial cue (purple star) was sighted at an inclination of 60° while the aircraft was at 750 ft, yielding an inclination-based distance (b) of 132 m, with an error of 5 m, based on a presumed inclination angle error of ±1°. The actual position (blue dot) was logged 74 s after the initial cue at a perpendicular distance (a) of 107 m, with an error of 68 m, based a presumption that the whale travelled at 3.3 km/h perpendicular to the trackline for the 74 s. Because the inclination-based error was lower, we used the inclination-based distance for this sighting.

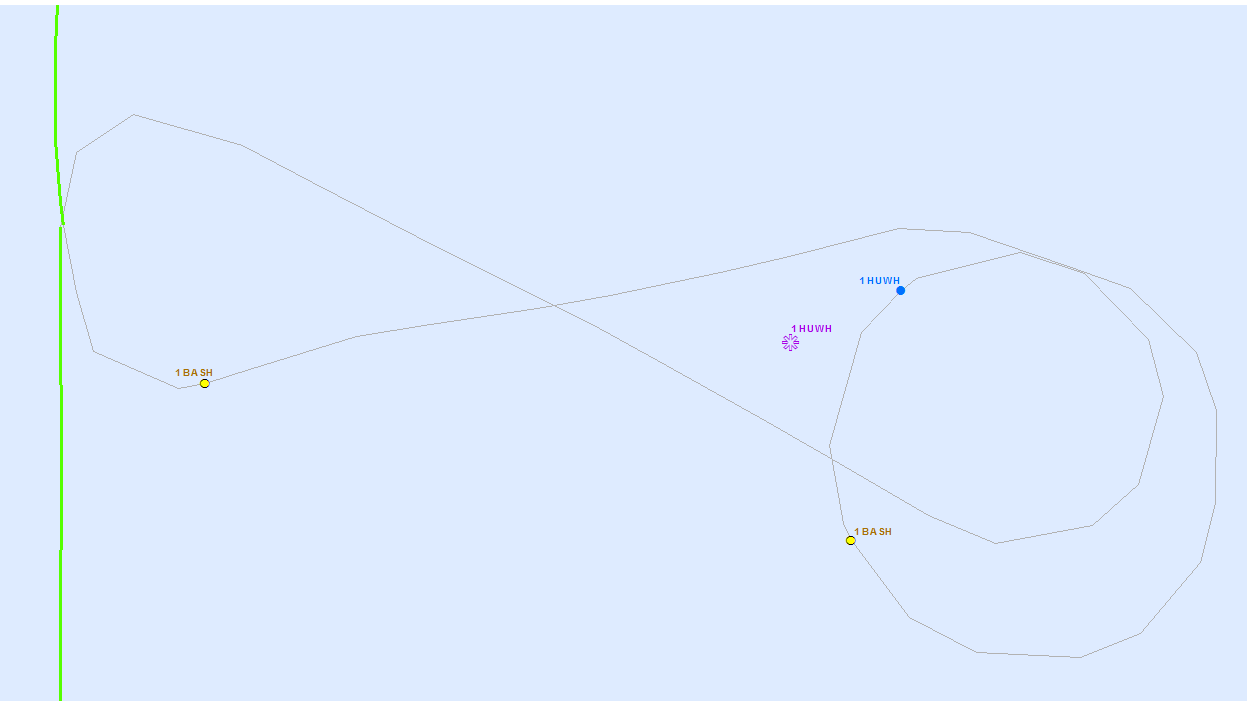


Figure 8. Estimation of perpendicular distances for a humpback whale sighted on the 28 June 2014 flight of Twin Otter N57RF at 10:19 AM. The initial cue (purple star) was sighted at an inclination of 87° while the aircraft was at 750 ft, yielding an inclination-based distance (b) of 4362 m, with an error of 1369 m, based on a presumed inclination angle error of ±1°. The actual position (blue dot) was logged 204 s after the initial cue at a perpendicular distance (a) of 5014 m, with an error of 673 m, based on a presumption that the whale travelled at 3.3 km/h perpendicular to the trackline for the 204 s. Because the AP-based error was lower, we used the AP-based distance for this sighting.

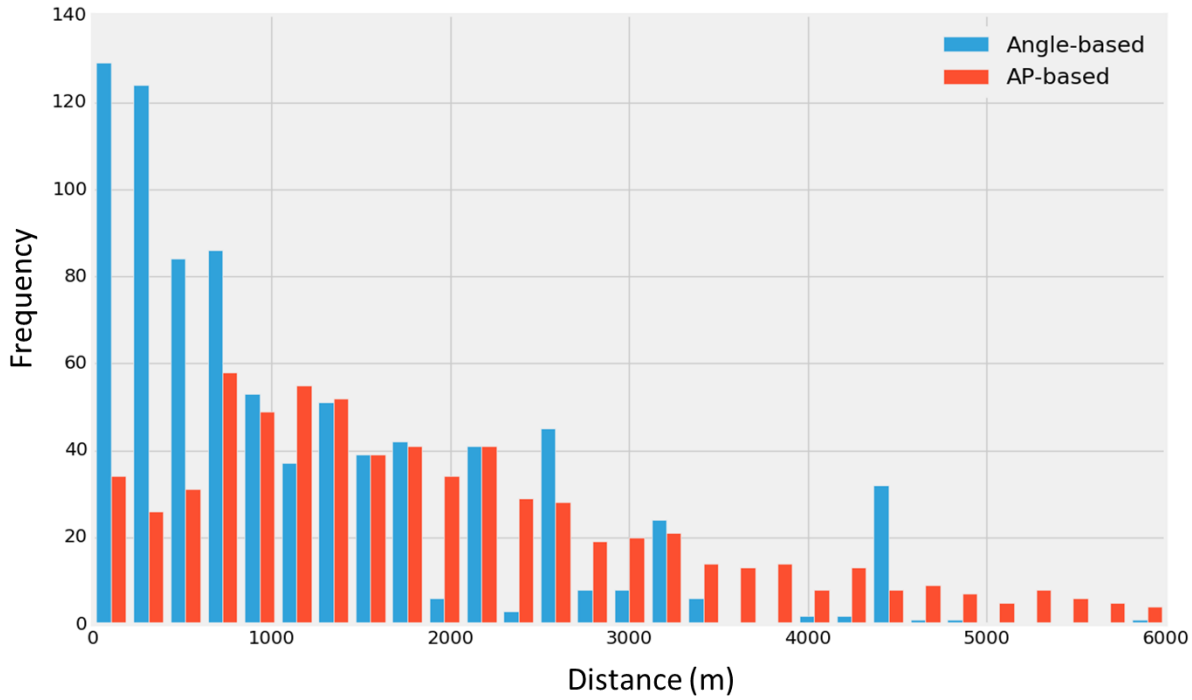


Figure 9. Perpendicular distances estimated by our new algorithm, for right whale sightings from NARWSS Twin Otter surveys, 2003-2016, with a target altitude of 750 ft.

A histogram of perpendicular distances derived by the new algorithm showed shorter spikes at heaping angles and a more even distribution of distances (Figure 10), characteristic of what is expected theoretically. Spikes were not completely eliminated, as inclination angles were still occasionally used for distant sightings, e.g. when observers marked actual positions as "not good for de" or because too much time elapsed between the logging of the initial cue and the AP record. The histogram also showed much less of a drop-off in sightings close to the trackline, supporting the NARWSS team's assertion that observers were not missing sightings as perpendicular distance approached zero, and obviating the need to use left truncation when fitting the detection function.

A preliminary hazard-rate detection function fitted to perpendicular distances estimated with the new algorithm showed better performance in Q-Q analysis (Figure 11) and in other goodness-of-fit diagnostics (not shown) than a similar function fitted to distances estimated only from inclination angles. The preliminary detection function fitted to the new algorithm yielded a significantly smaller effective strip half-width (ESHW) and a narrower "shoulder". This effect persisted even after the covariates were added and the full detection function selection procedure was performed to obtain the final detection function (section 5.2.1.1.5). We suspect this may result from a tendency for NARWSS observers to overestimate the inclination angle for distant sightings. We have not analyzed this systematically, but Figure 12 gives several examples. For large inclination angles our algorithm estimates large errors in angle-based perpendicular distances, as much as hundreds or thousands of meters. In these cases, the algorithm will usually select AP-based distances, making it tolerant to this problem and able to correct a bias toward overestimating inclination angles, if it exists. For this reason, and the other improvements noted above, the new right whale model documented in this report uses the new algorithm for estimating perpendicular distance to right whale sightings reported by NARWSS. Future model updates for other taxa for which NARWSS regularly reported AP records will likely do the same. This is likely to include all "large" whales (all baleen whales except minke whales, plus sperm whales). For other taxa, we will likely have to rely solely on inclination angles. However, these smaller species (e.g. dolphins) tend not to be as detectable far from the trackline. For these closer distances, our algorithm would preferentially select inclination-based distances over AP-based distances. This tendency would be further enhanced when we accounted for the higher speed of the smaller species, which would increase the assumed error in AP-based distances. Thus, we do not anticipate the reliance on inclination-based distances for these smaller taxa will be the problem that it has been for the large whale species.

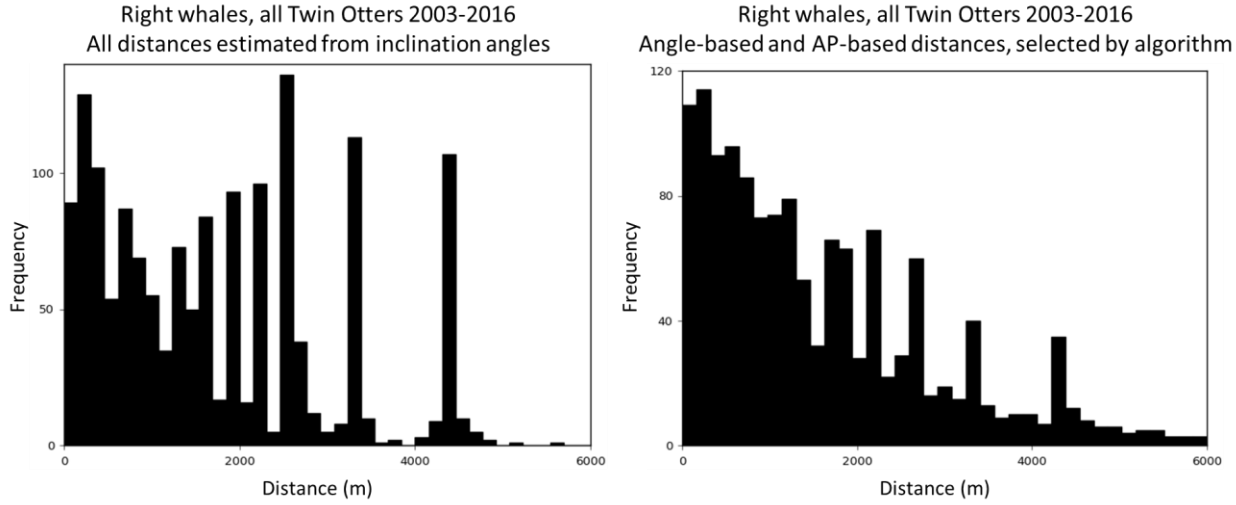


Figure 10. Perpendicular distances for right whale sightings reported by 750-ft NARWSS surveys estimated with inclination angles for all sightings (left), as was done for our prior right whale models, and estimated with the new algorithm (right), as was done for the new right whale model documented in this report. Sightings used for the new algorithm (right) reflect those now excluded or included by our other improvements to our NARWSS processing procedure described at the beginning of this section, resulting in roughly 10% fewer sightings than those previously used (left). In both cases, sightings made in the Gulf of St. Lawrence were excluded.

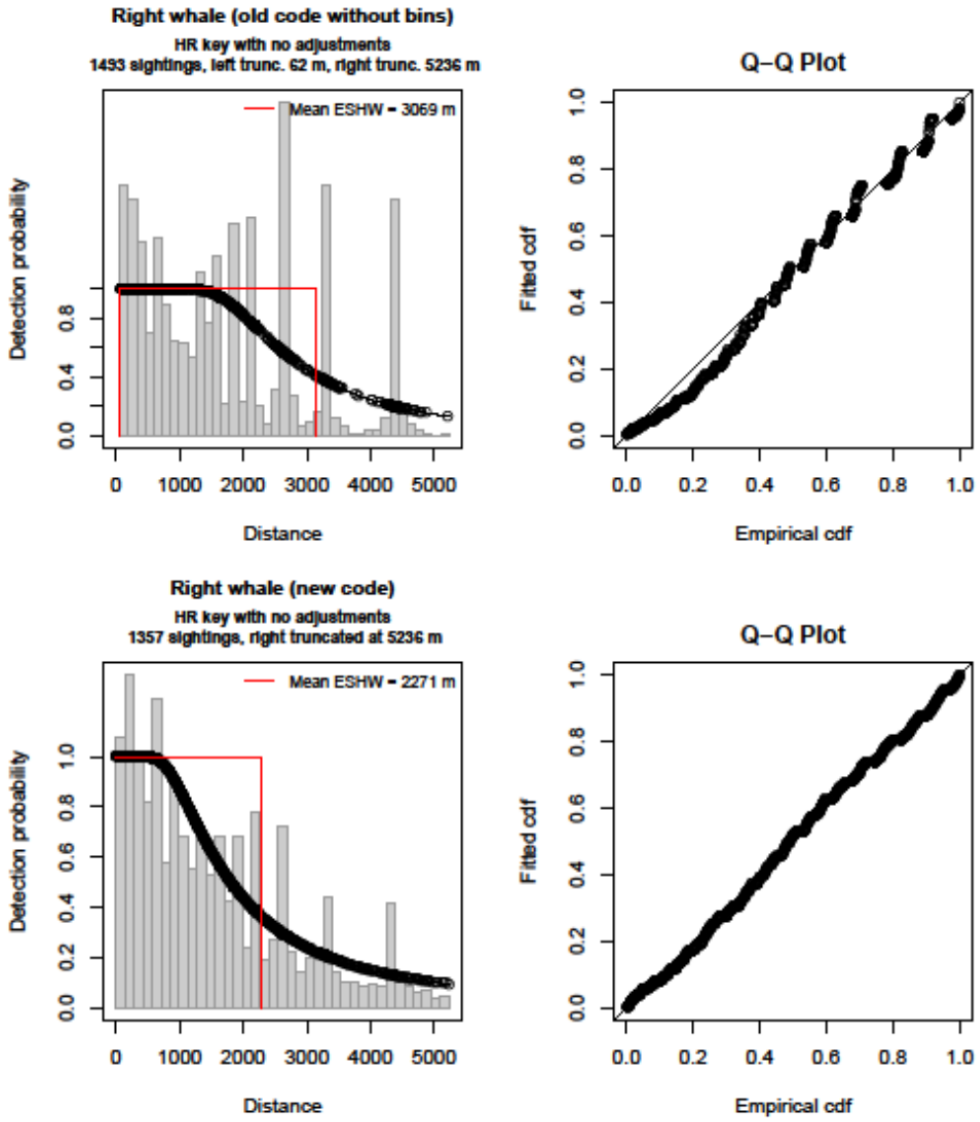
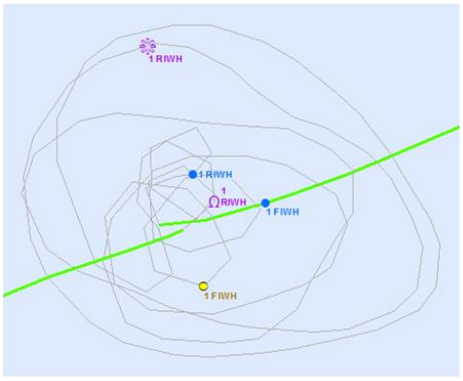
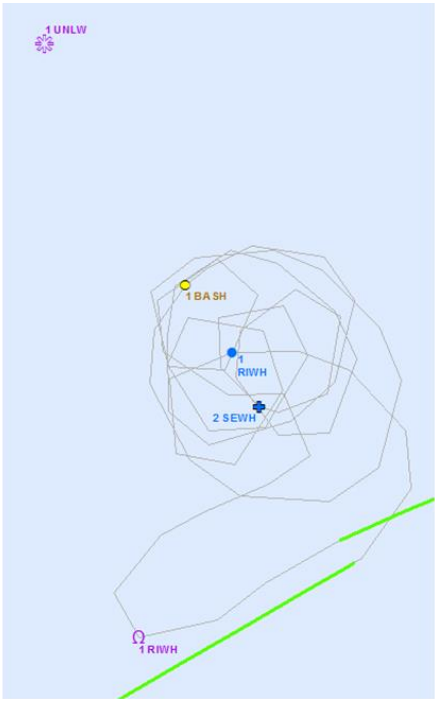


Figure 11. Preliminary hazard-rate detection functions fitted to the right whale sightings shown in Figure 10. The top row shows the function in which all perpendicular distances were estimated from inclination angles (data from Figure 10 left). The bottom row shows the function in which perpendicular distances were estimated with the new algorithm (data from Figure 10 right). To facilitate comparison between the two functions, the top function did not use binning and neither function used covariates. (The final detection function (section 5.2.1.1.5), did use covariates.) Note the decreased heaping (bottom left) and improved Q-Q plot performance (bottom right) that resulted from the new perpendicular distance algorithm, as well as the smaller mean ESHW (bottom left vs. top left).

2014-06-23 11:46 AM NOAA 57

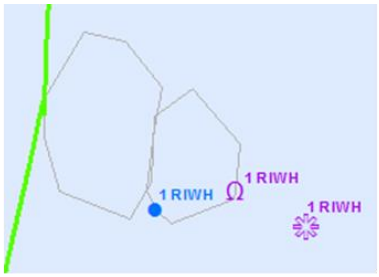


2012-11-16 12:40 PM NOAA 48



2012-04-26 12:05 PM NOAA 57
The 1 RIWH was marked "good for de"

2012-01-04 1:09 PM NOAA 57



2012-01-06 12:38 PM NOAA 57

2012-01-06 11:02 AM NOAA 57

Figure 12. Examples of possible overestimations of inclination angles for sightings of right whales reported during NARWSS surveys. For these, the location of the sighting cue (purple star) estimated from the inclination angle was substantially farther from the trackline than the "actual position" (blue dot) matched to the sighting cue, and usually farther than even the area traversed by the aircraft during circling. Although these examples are all from 2012 and 2014, we do not suggest that this phenomenon is limited to or concentrated within those years.

5. Updated North Atlantic Right Whale Model

The model utilized the two-stage approach known as *density surface modeling* (DSM) (Hedley & Buckland 2004; Miller et al. 2013), implemented as described by Roberts et al. (2016a) and subsequent MGEL reports to the Navy. In general, the same approach was used here unless otherwise noted, and readers should consult those publications for methodological details. We provide an overview in the first section below, and additional details, as they pertain to this specific analysis, in following sections.

5.1. Methods overview

In the first stage of the model, called *detection modeling*, we used traditional *distance sampling* (Buckland et al. 2001) to model the detectability of observed objects (whales, in this case) and correct the survey data for imperfect detection. To account for substantial differences in detectability across survey platforms and programs, we first split the data into subsets according to platform, survey altitude, and other important factors and fitted detection functions to each subset. Then, using findings from the literature, we applied additional corrections to account for the possibility that observers may miss sightings directly on the trackline, either because whales were submerged and unavailable for detection ("availability bias") or because they were simply hard to detect despite being at the surface ("perception bias"). Finally, we split the survey transects into segments and applied the detection functions and corrections to produce the records used to fit the second stage of the model.

In the second stage, we used generalized additive models (GAMs) (Hastie & Tibshirani 1990; Wood 2006) to model the counts of individuals sighted on the survey segments from environmental covariates observed at the segments, as well as, occasionally, space and time. To allow species-environment relationships to differ regionally or seasonally (e.g. for right whales that migrated to southern calving grounds vs. those that overwintered at the northern feeding grounds), we split the data into independent regional and seasonal models, based on what was known of right whale ecology and what patterns we observed in the data. After selecting the best candidate models for each region and season, determined by examining goodness-of-fit statistics, diagnostic plots, total abundance estimates, and spatiotemporal patterns in density predictions, we predicted them across a time series of gridded maps of covariates (e.g., images of sea surface temperature, and so on), to produce a corresponding time series of maps of right whale density and model uncertainty. We then summarized the maps according to the Navy's needs for the NMSDD and NOAA's needs for other management applications.

To facilitate the propagation of uncertainty between modeling stages using the statistical approach described by Bravington et al. (in review) and elaborated by the "DenMod" Working Group (https://synergy.st-andrews.ac.uk/denmod/; Miller et al., in prep), we formulated the model using the "count as response" approach described by Miller et al. (2013). Under this approach, detection functions were fitted to detectability covariates (e.g. Beaufort wind force scale) recorded at survey segments, and then used to estimate an effective strip width (ESW) for each segment from the covariate values observed there. The ESW was multiplied by the segment length to estimate the area effectively surveyed at the segment, which was used as an offset variable in the GAMs. For the response variable, we used the count of animals sighted, inflated by availability and perception bias corrections.

This approach differed from our prior models, in which we fitted detection functions to detectability covariates recorded at the sightings (*sensu* Miller et al. (2013), "DSM with covariates at the observation level") and formulated the model with a Horvitz-Thompson-style estimator. Under that formulation, we used the right truncation distance of the detection functions in place of the ESW when estimating the model offset and inflated the count of animals sighted by detection probability instead. The change in formulation was prompted by our impression that the new uncertainty propagation approach, which the DenMod Working Group was developing at the same time we were developing our model, would be substantially easier to implement when detectability covariates could vary at the segment level rather than the observation level. Whether this is true remains to be seen, but in any case, both formulations are effective approaches to density surface modeling (Miller et al. 2013), and we will have opportunity to investigate this further in the remainder of Option Year 4, when we attempt to apply the DenMod team's uncertainty propagation method to the model (see 6.3.4).

5.2. Detection modeling

Detection modeling consisted of two main steps: 1) developing detection functions that modeled how detectability varied with distance from the trackline and other covariates, and 2) estimating and applying corrections for perception and availability biases, in which animals directly along the trackline were "missed" because they were hard to see or were unavailable while diving.

5.2.1. Detection functions

To fit an effective detection function, a sufficient number of sightings are needed. A common practice is to obtain at least 60-80 sightings, based on a widely-cited recommendation by Buckland et al. (2001). However, this is not a strict requirement—more or fewer may be necessary or sufficient, depending on the purpose (Roberts et al. 2019). When few sightings are available, usually the most effective remedy is to collect more data. When that is not feasible, it is often possible to pool the sightings from similar surveys or species into a single detection function. With sufficient data, a categorical covariate can be used to account for differences between surveys or species. However, this must be done with care; pooling surveys or species that exhibit substantial differences can yield detection functions with high variance or other problems.

To guide pooling decisions, we followed the approach of Roberts et al. (2016a). We organized the surveys into two hierarchies—one for aerial, and one for shipboard—based on the similarity of the surveys' protocols. We then fitted detection functions to branches of the hierarchies for which sufficient sightings were available, ascending to higher nodes when there were too few, and pooling in sightings of similar species when so few sightings were available that we would otherwise be forced to pool strongly incompatible surveys. When pooling species, on advice of collaborators who executed the surveys, we included other "large" whales, defined as all baleen whales except minke whales, plus sperm whales. (We excluded minke whales because they are harder to detect than the larger species (Heide-Jørgensen et al. 2010)). When possible, we used covariates to account for differences between surveys or species. To boost sighting counts, we also included compatible surveys from 1992-2002, which predated the period spanned by the spatial model, which started in 2003.

We fitted detection functions in R with the `mrds` package version 2.2.1. However, the large majority of surveys available for our analysis used a single observer team. Accordingly, we used a single-team modeling approach for all detection functions. For surveys that collected data with two teams, we used data only from the primary team.

For nearly all detection functions, we developed a set of candidate functions that used a variety of formulations and selected one as "best". Usually, this was the candidate that ranked best according to Akaike's information criterion (AIC) (Akaike 1974), discounting those that exhibited problems in goodness-of-fit statistics, Q-Q plots, or other diagnostics. We developed candidates that used both half-normal and hazard-rate forms, using so-called conventional distance sampling (CDS) formulations with adjustment terms recommended by Thomas et al. (2010), and also multi-covariate distance sampling (MCDS) formulations using all combinations of covariates that were available across all surveys that were pooled into the detection function. Occasionally, we also tested uniform functions with cosine adjustments, but these almost always scored poorly compared to half-normal or hazard-rate candidates.

For MCDS candidates, unlike Roberts et al. (2016a), we generally did not limit the number of covariates allowed in a detection function according to the number of sightings available. Instead, based on advice from collaborating distance sampling experts at the University of St. Andrews, we eliminated candidates from contention when the standard error of any covariate coefficient exceeded the estimate for the coefficient, indicating that the influence of that covariate was not supported statistically.

The following sections document the details of the detection hierarchies and detection functions developed for the model.

5.2.1.1. Aerial surveys

Figure 13 shows the detection hierarchy developed for aerial surveys. We split the surveys first by altitude because of the strong, direct effect it has on sighting distances. Then, recognizing the important protocol differences between the NARWSS program and other surveys (see section 4.2)—particularly that NARWSS directed observers to focus attention 1 nmi out from the trackline while others directed observers to guard it more closely—we split the 2003-2016 NARWSS surveys flown at 750 ft into its own node rather than grouping it with the other surveys flown at 750 ft. The other 750 ft surveys, which were mostly carried out in the southeast and mid-Atlantic, reported sightings of only 5 groups of right whales (Figure 14), and 19 groups of large whales of any species (Figure 15). In contrast, the 600 ft surveys spanned the entire study area and sighted more than 700 groups of large whales (Figure 15). Because of this, we pooled the 750 ft surveys (other than NARWSS) with the 600 ft surveys and used the sightings from this 600-750 ft pool for the detection function used for the 750 ft surveys (see section 5.2.1.1.4).

For the surveys at 1000 ft, we split the surveys first according to whether bubble or flat windows were used and then according to the protocols that were used (Figure 13). The right whale "early warning system" (EWS) surveys in the southeast sighted large numbers of right whales, allowing us to fit right-whale-specific detection functions, while the other programs did not, forcing us to fit functions that pooled in other large whale species.

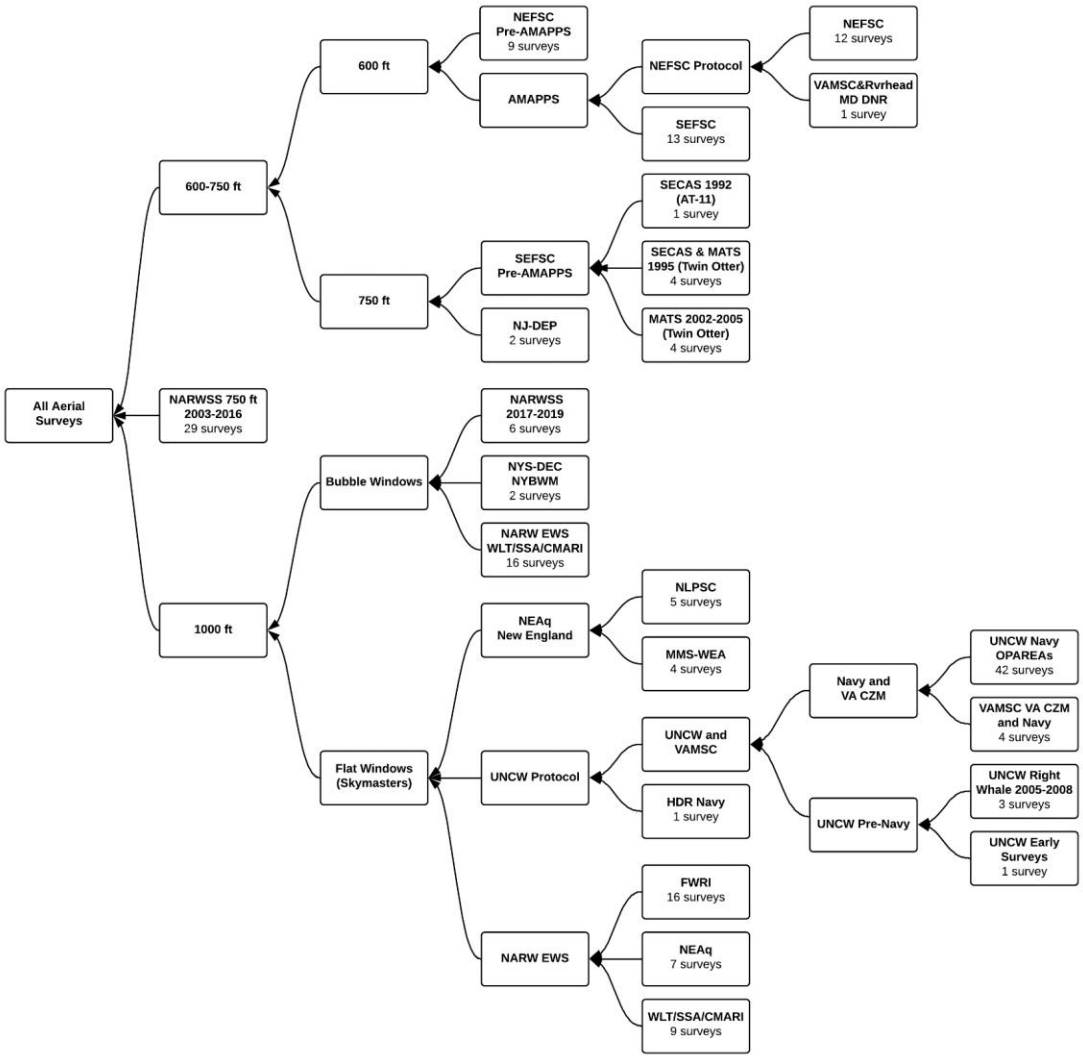


Figure 13. Detection hierarchy developed for aerial surveys.

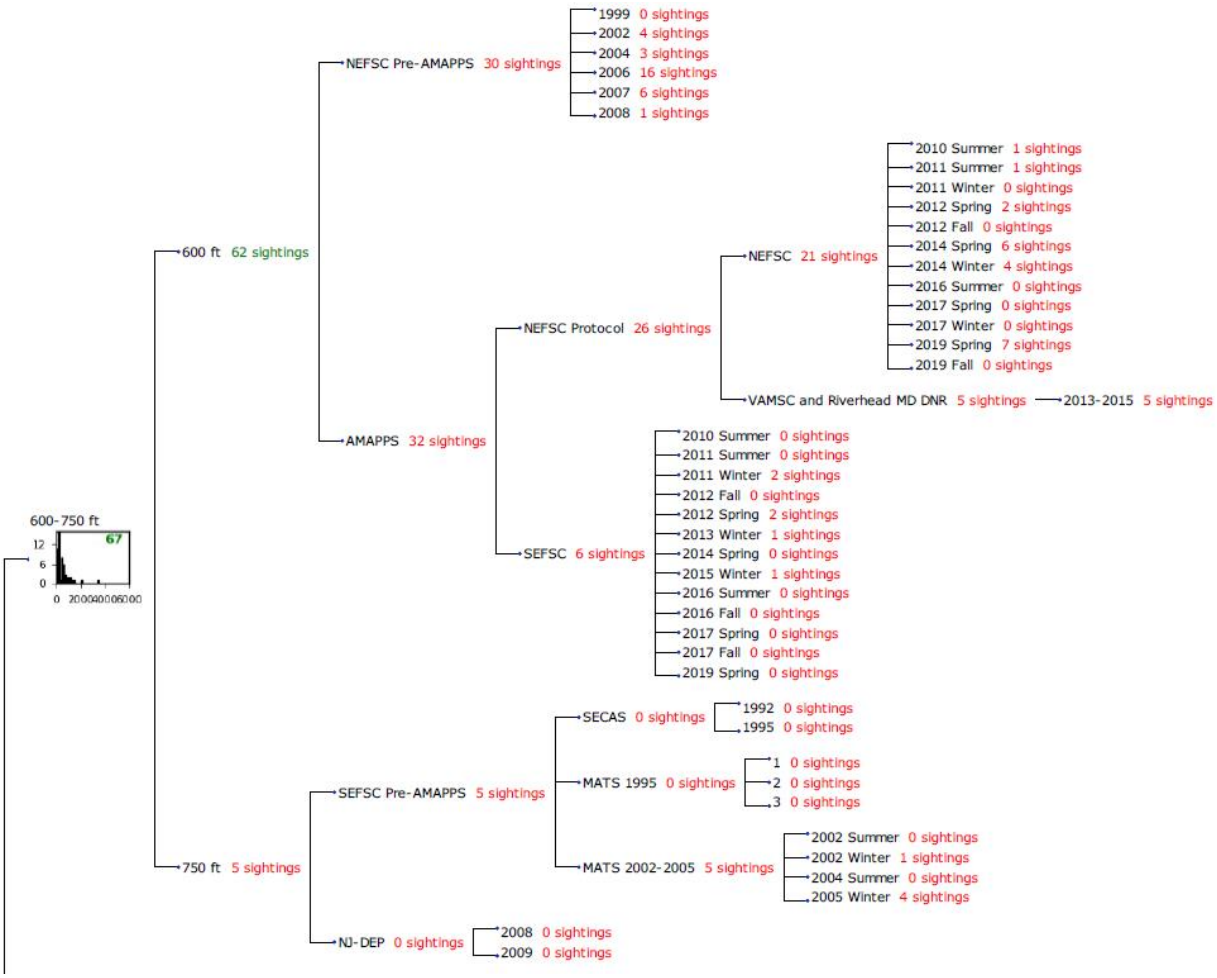


Figure 14. Sightings of right whale groups reported by surveys within the 600-750 ft branch of the detection hierarchy for aerial surveys. Because of these small sighting counts, we fitted no right-whale-specific detection functions to any nodes in this branch of the hierarchy. Instead, we fitted detection functions to pools that included other large whales (Figure 15). In this figure and others like it below, the sighting counts listed for non-leaf nodes are the sum of the counts for leaf nodes below them. Counts of at least 60 sightings are shown in green; less than 60 are shown in red.

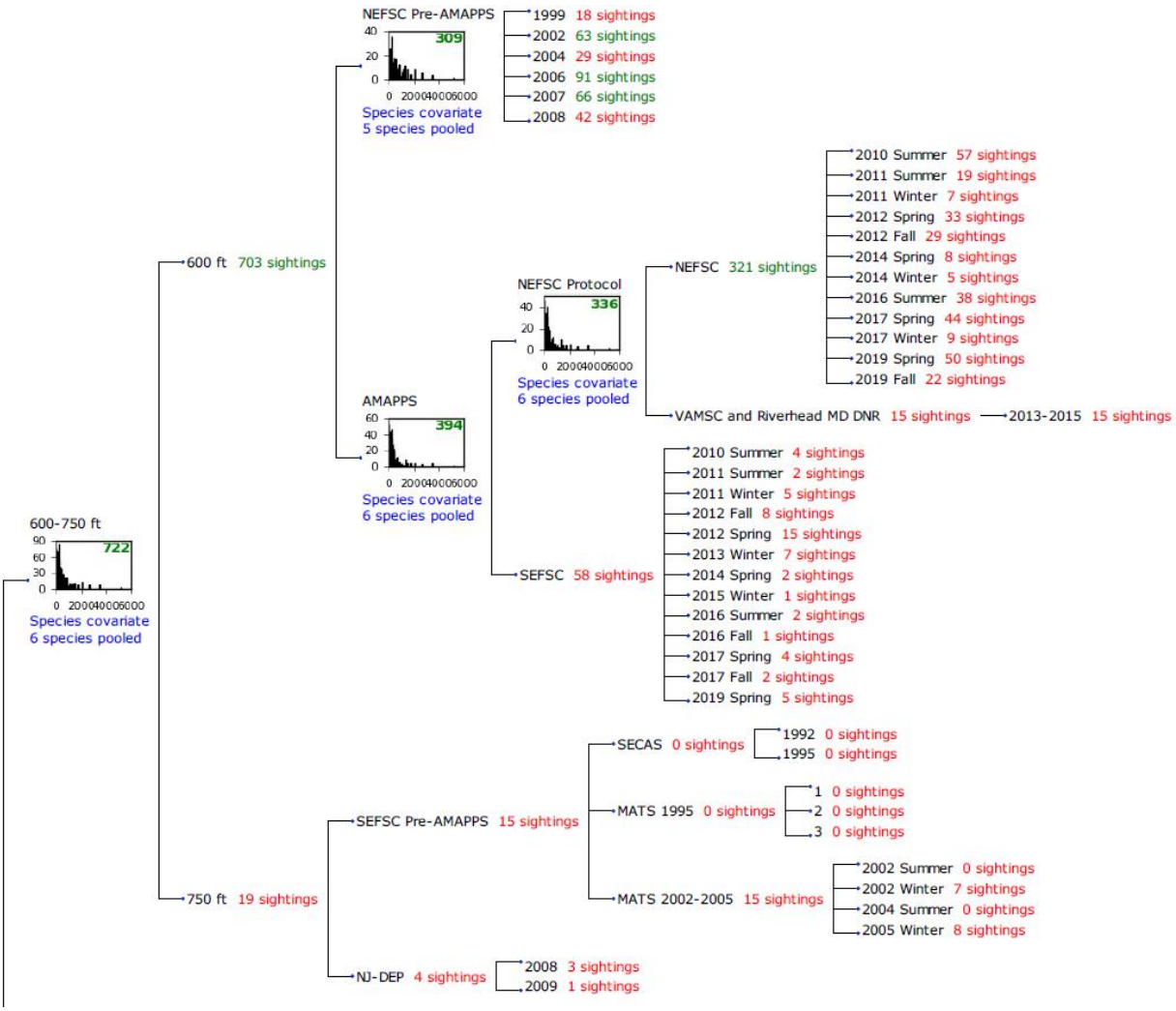


Figure 15. Sightings of large whale groups (all baleen whale species except minke whales, plus sperm whales) reported by surveys within the 600-750 ft branch of the detection hierarchy for aerial surveys. The four histograms show nodes where detection functions were fitted; we present the details of these detection functions below.

5.2.1.1.1 NEFSC Pre-AMAPPS detection function

From NEFSC's pre-AMAPPS marine mammal abundance program, NEFSC provided aerial surveys extending back to 1995. From these, we omitted surveys from 1995-1998 after noticing an anomalous spike in sightings at approximately 2090 m (corresponding to a vertical inclination angle of 85º at 600 ft altitude) that we could not explain. There were insufficient sightings to fit a right-whale-specific detection function, so we pooled all large baleen whales (minke whales excluded) and sperm whales, yielding 309 sightings, of which 30 were right whales. We retained sightings with Beaufort state ≤5 and right-truncated at 1500 m, dropping 26 sightings (8%) and retaining 283 sightings (92%). A scatterplot of perpendicular distance vs. Beaufort state (Figure 16) indicated detection distance was positively correlated with increasing sea state, an illogical result. Because of that, we excluded this covariate from candidate detection functions. NEFSC did not provide detectability covariates other than Beaufort state, but sufficient sightings were available to try species name as a covariate and it was retained in the final detection function (Figure 17). Exploratory analysis revealed similarities in detectability between several species; in the final detection function, we grouped species into two categories: fin and sei whales, which were detected at shorter distances, and humpback, right, and sperm whales, which were detected at longer distances.

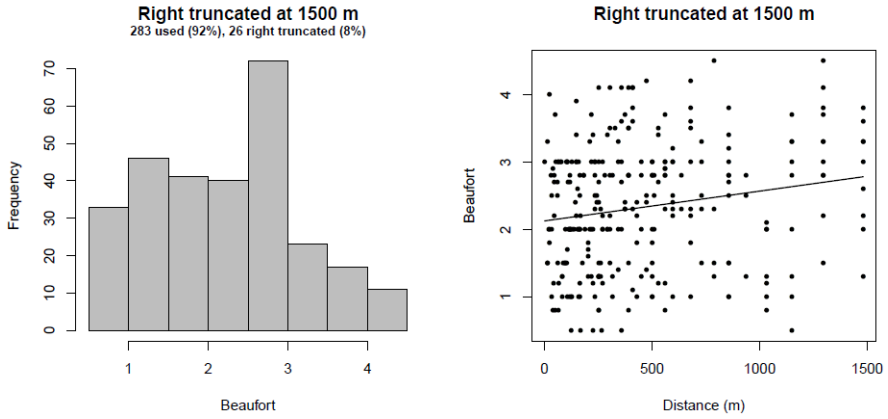


Figure 16. Distribution of Beaufort state (left) and its relationship to perpendicular sighting distance (right) for sightings used in the aerial NEFSC Pre-AMAPPS detection function. (Although the Beaufort scale is a categorical system consisting of whole integers, NEFSC's sightings included measurements as decimal values, such as 2.4.) Note the apparently increasing relationship of distance with Beaufort state (right). Because this relationship is not logical—we would not expect detectability to improve with increasing sea state—we excluded this covariate from candidate detection functions.

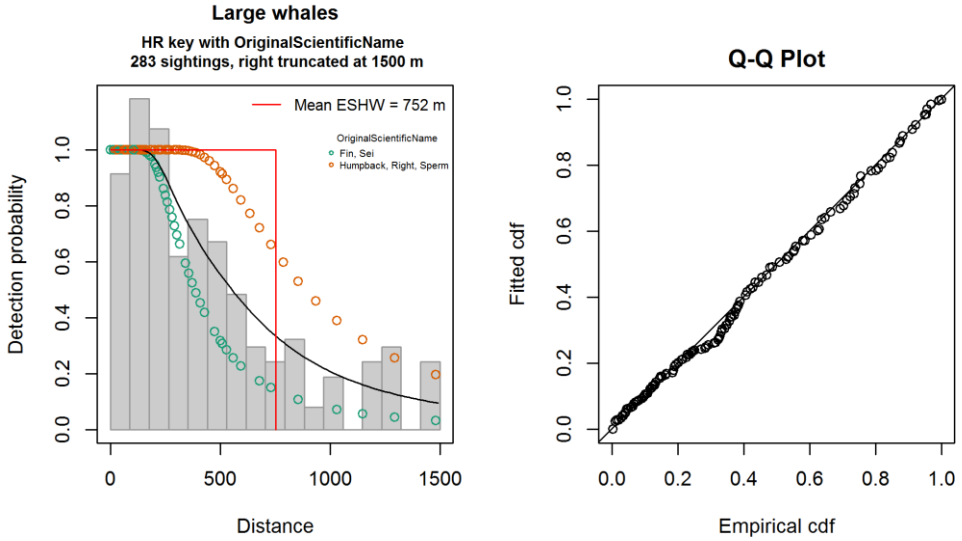


Figure 17. Final detection function applied to the aerial NEFSC Pre-AMAPPS surveys.

5.2.1.1.2 AMAPPS: NEFSC Protocol detection function

All broad-scale marine mammal abundance surveys provided to us by NEFSC since 2010 were from the AMAPPS program, which was jointly designed by NEFSC and SEFSC such that data collected by the two organizations were highly compatible and intended to be analyzed together. However, we did discover an important difference in aerial data collected by the two organizations (see next section) and NEFSC itself generally developed separate detection functions for the two organizations (Palka et al. 2017; Chavez-Rosales et al. 2019), so we did the same. To NEFSC's aerial AMAPPS surveys we added the VAMSC and Riverhead Foundation's survey of Maryland, which was conducted with a protocol identical to the NEFSC AMAPPS survey by some of the same personnel (R. DiGiovanni, pers. comm.).

There were insufficient sightings to fit a right-whale-specific detection function, so we pooled all large baleen whales (minke whales excluded) and sperm whales, yielding 336 sightings, of which 26 were right whales. We retained sightings with Beaufort state ≤5 and right-truncated at 1500 m, dropping 20 sightings (6%) and retaining 316 (94%). As with NEFSC's pre-AMAPPS aerial surveys, a scatterplot of perpendicular distance vs. Beaufort state (Figure 18) indicated detection distance was positively correlated with increasing sea state, an illogical result. Because of that, we excluded this covariate from candidate detection functions. The final detection function used species name as a covariate (Figure 19). Exploratory analysis revealed similarities in detectability between several species; in the final detection function, we grouped species into two categories: blue, fin, and sei whales, which were detected at shorter distances, and humpback, right, and sperm whales, which were detected at longer distances. This effect was not as pronounced as with the NEFSC pre-AMAPPS surveys (Figure 17).

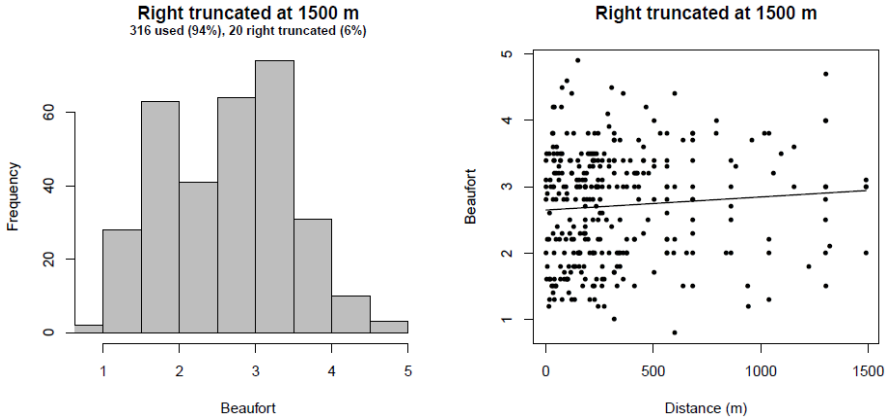


Figure 18. Distribution of Beaufort state (left) and its relationship to perpendicular sighting distance (right) for sightings used in the aerial AMAPPS, NEFSC Protocol detection function. (Although the Beaufort scale is a categorical system consisting of whole integers, NEFSC's sightings included measurements as decimal values, such as 2.4.) Note the apparently increasing relationship of distance with Beaufort state (right). Because this relationship is not logical—we would not expect detectability to improve with increasing sea state—we excluded this covariate from candidate detection functions.

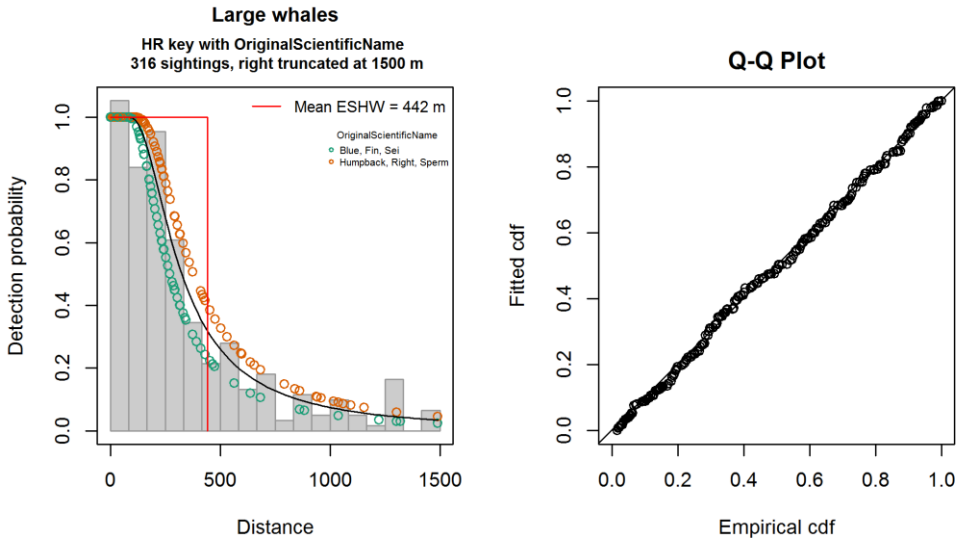


Figure 19. Final detection function applied to the aerial AMAPPS NEFSC Protocol surveys.

5.2.1.1.3 AMAPPS detection function (applied to SEFSC only)

The SEFSC aerial AMAPPS program only reported 58 sightings of large whales from the primary observer teams from summer 2010 through spring 2019, six of which were right whales. In our judgement, this was enough to fit an SEFSC-specific detection function for large whales, but better precision might be obtained by pooling the NEFSC and SEFSC sightings, which were designed by the two organizations to be compatible. However, the AMAPPS annual reports released jointly by the organizations described a subtle protocol difference between the two programs. SEFSC's description included the text:

> During on effort periods (e.g., level flight at survey altitude and speed), observers *searched visually from the track line (0˚) to approximately 50˚ above vertical*. [Emphasis added]

In contrast, NEFSC's description stated:

> The observers viewing from the front two bubble windows and the back side visa window *searched from straight down to the horizon*, with a concentration on waters between straight down (0º) and about 50º up from straight down. [Emphasis added]

In the data we received, NOAA provided perpendicular distances that they had calculated from vertical sighting angles but not the angles themselves. In SEFSC's data, the farthest perpendicular distance was 562 m, corresponding to an inclination angle of 72º up from vertical at the target altitude of 600 ft. This was considerably closer to the trackline than NEFSC's sightings, which we truncated at 1500 m (Figure 19), corresponding to an inclination angle of approximately 83º, dropping 20 sightings beyond this distance (at even larger inclination angles). Although SEFSC's maximum inclination of 72º was considerably larger than their documented target maximum of 50º, the overall pattern in the sighting distances from the two programs was consistent with their documented approach.

Given that SEFSC observers apparently had maintained focus more directly under the airplane, or at least not reported sightings at the higher angles reported by NEFSC observers, it was not appropriate to naïvely pool the SEFSC and NEFSC sightings and fit a detection function at the truncation distance of 1500 m that was appropriate for the NEFSC sightings. The resulting detection function would be biased low, based on the assumption that the probability of SEFSC making a distant sighting was non-zero, when in fact it *was* zero.

To account for this while gaining the benefit of the NEFSC sightings, we pooled NEFSC's sightings with SEFSC's but then truncated at a distance appropriate for SEFSC's behavior. Although SEFSC documented their target maximum angle at 50º, this corresponded to a perpendicular distance of only 218 m, yielding the shortest truncation distance among all programs. We judged this approach too conservative, especially because whale sightings in the southeast portion of the study area were rare and therefore very valuable for use in the spatial model. Instead we truncated at 565 m, the approximate maximum perpendicular distance reported by SEFSC.

To test for differences between the two organizations, we included the organization as a categorical covariate in candidate detection functions, but the difference was statistically insignificant, and detection performance nearly identical, at the 565 m right truncation distance (Figure 20). Although the ratio of data was heavily biased (6:1) toward NEFSC, owing to the rarity of whales in SEFSC's survey area, this result was consistent with the intended design of AMAPPS, which was that NEFSC and SEFSC surveys be as similar and compatible as possible.

For the final detection function, as with the function that only incorporated NEFSC sightings, there were insufficient sightings to fit a right-whale-specific detection function, so we pooled all large baleen whales (minke whales excluded) and sperm whales, yielding 394 sightings of which 32 were right whales. We retained sightings with Beaufort state ≤5 and right-truncated at 565 m, dropping 75 sightings (19%) which were all from NEFSC, and retaining 319 (81%). As with the NEFSC-only detection functions, a scatterplot of perpendicular distance vs. Beaufort state (not shown) indicated detection distance was positively correlated with increasing sea state, so we excluded Beaufort, and the final detection function used species name as a covariate (Figure 21).

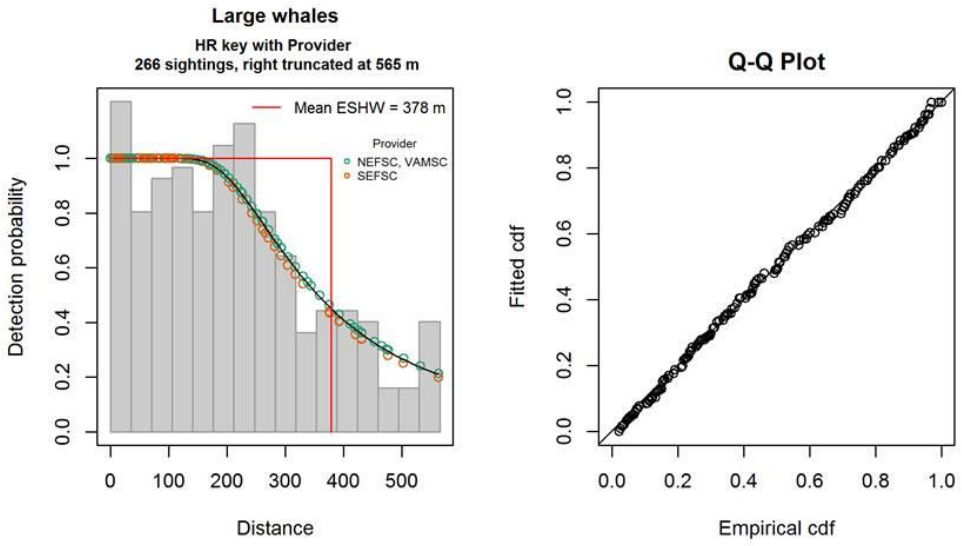


Figure 20. Detection function used to test for differences in detectability of large whales between SEFSC and NEFSC, over the distance range appropriate for SEFSC's observation protocol. No significant difference was detected between two programs, as evidenced by the nearly identical functions traced by the colored circles in the left plot. (Note that this detection function was only used during exploratory analysis, and not selected as "best". Also, this test was performed before NEFSC and SEFSC had provided their final batches of AMAPPS data for our project, so the count of sightings is smaller than was had for our final detection functions.)

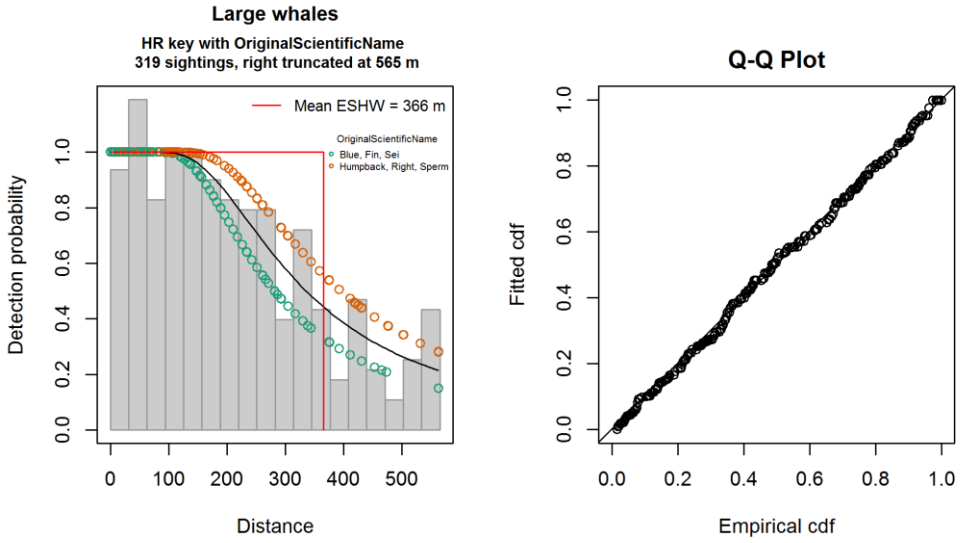


Figure 21. Final detection function applied to SEFSC's AMAPPS surveys. This function pooled AMAPPS data from both SEFSC and NEFSC (see text) but was only applied to SEFSC's AMAPPS surveys. For NEFSC's AMAPPS surveys, we applied the NEFSC-specific detection function (Figure 19) described previously (section 5.2.1.1.2).

5.2.1.1.4 600-750 ft detection function

The detection functions described above covered all the surveys flown at 600 ft, but none flown at 750 ft. Unfortunately, the latter only reported 19 sightings of large whales, 5 of which were right whales. We judged these too few to fit a detection function, and the best course of action was to pool in the 600 ft surveys (rather than pool in

non-large-whale species), and a fit a detection function for all data from the 600-750 ft pool. The resulting pool had 722 sightings of large whales (all baleen whales except minke whales, plus sperm whales), of which 67 were right whales. Because Beaufort state was not available for all the 750 ft surveys, we excluded it as a covariate. We did test altitude and species name; the former was found not to be significant (there were only 19 sightings at 750 ft), but the latter was significant. In the final detection function (Figure 22), we grouped species into three two categories, based on exploratory analysis of the data: sei whales, which were detected only at relatively short distances; blue and fin whales, which were detected at farther distances; and humpback, right, and sperm whales, which were detected at the farthest distances. This pattern is consistent with the sizes and behavior patterns of these species. This detection function should be viewed cautiously, as most of the sightings were obtained from NEFSC surveys flown at 600 ft, and then applied to NJ-DEP and SEFSC (pre-AMAPPS) surveys flown at 750 ft.

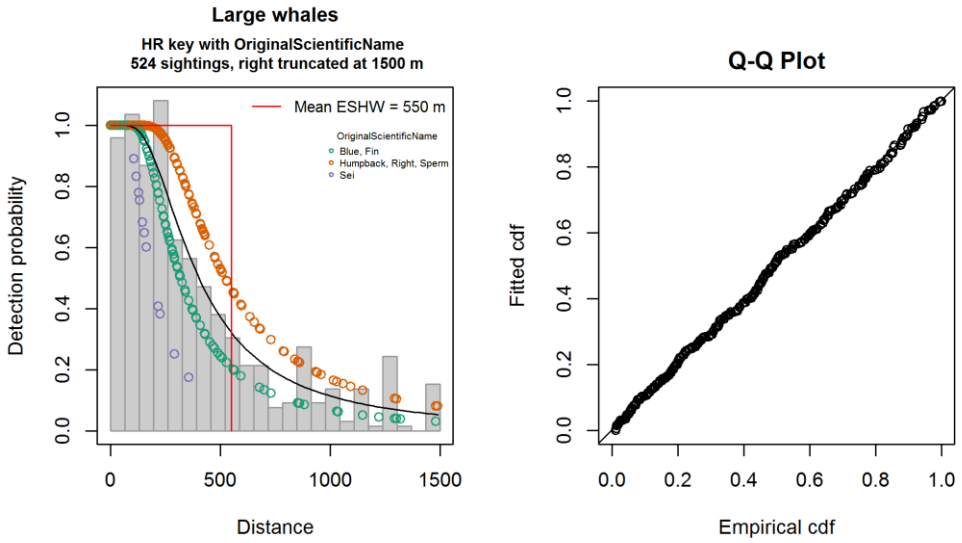


Figure 22. Final detection function applied to aerial surveys flown at 750 ft (other than NARWSS).

5.2.1.1.5 NARWSS 750 ft (2003-2016) detection function

In 2003, NEFSC's NARWSS program began flying exclusively on NOAA's de Havilland Twin Otter aircraft, abandoning the Grumman Goose and Widgeon that were used during the initial years of the program (1998-2002). From 2003-2016, NARWSS flew a target altitude of 750 ft (Cole et al. 2007). Because the NARWSS protocol directed observers to focus attention 1 nmi out from the trackline to not miss distant right whales (T. Cole, pers. comm.), we did not pool NARWSS surveys with others flown at 750 ft.

After implementing the aforementioned changes to our processing of NARWSS data (see section 4.2) we excluded survey effort made in Beaufort state 6 or higher, NARWSS quality code x (defined as "unacceptable"), or visibility < 3 nmi. This represented a wider range of observing conditions than we accepted for other surveys, which we were willing to do for NARWSS because of the large number of right whale sightings available from the program, which facilitated testing the effects of sub-optimal conditions via detection function covariates. We also excluded effort made in the Gulf of St. Lawrence (this may be included in a future model, if we model right whale distributions in Canada more comprehensively). After exclusions, we were left with 1399 sightings of right whale groups, which were sufficient to fit a right-whale-specific detection function.

NARWSS, with a protocol designed specifically for large whales, reported many sightings at long distances, allowing us to adopt a much larger truncation distance than for the other NOAA surveys, which had protocols designed to minimize the chance of missing small marine mammals such as harbor porpoises (e.g. by flying lower and focusing attention at smaller vertical inclination angles). For our NARWSS right-whale-specific detection function, we used a truncation distance of 5236 m, corresponding to a vertical inclination angle of 87.5° at 750 ft altitude. As discussed above, NARWSS rounded vertical angles less than 89° to the nearest degree. The angle 87.5° fully encompassed the

range of perpendicular distances (3738-5236 m) likely to be represented by a reported angle of 87°. At this angle/distance, a resulting 52 sightings (4%) were right-truncated and 1347 (96%) were retained. As discussed in section 4.2, after we implemented the processing improvements for NARWSS surveys, we found it was no longer necessary to left-truncate NARWSS data, as we had done in prior analyses (Roberts et al. 2016a).

The final detection function (Figure 25) included two covariates. The first, Beaufort state, was best modeled as a two-level categorical covariate, with Beaufort 3 showing detectability at greater distances than other Beaufort states (Figure 23). NARWSS observers indicated this "hump-shaped" pattern in detectability at moderate sea states was plausible, as a moderate (but not excessive) "texture" to the water surface helped observers sight the dorsal sides of large whales. The second covariate, visibility, was modeled as continuous, with distant sightings being reported only when visibility was high (Figure 24).

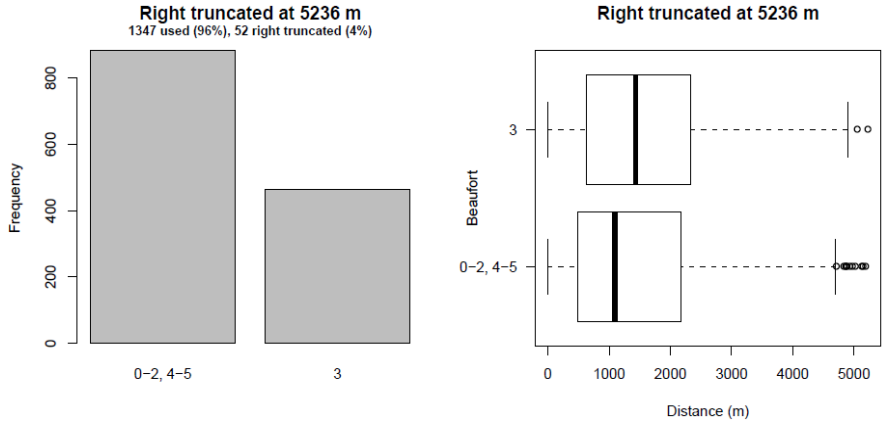


Figure 23. Frequency of categorized Beaufort state (left) and the distribution of perpendicular sighting distances by Beaufort state (right) for sightings used in the 750 ft NARWSS (2003-2016) detection function.

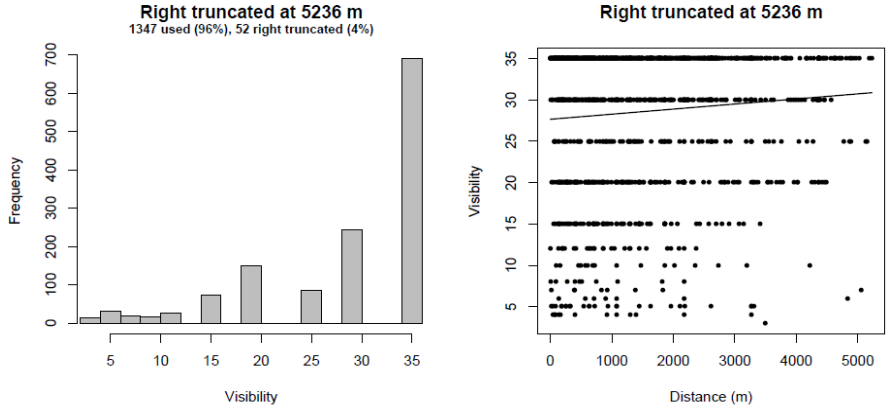


Figure 24. Distribution of visibility (left) and its relationship to perpendicular sighting distance (right) for sightings used in the 750 ft NARWSS (2003-2016) detection function.

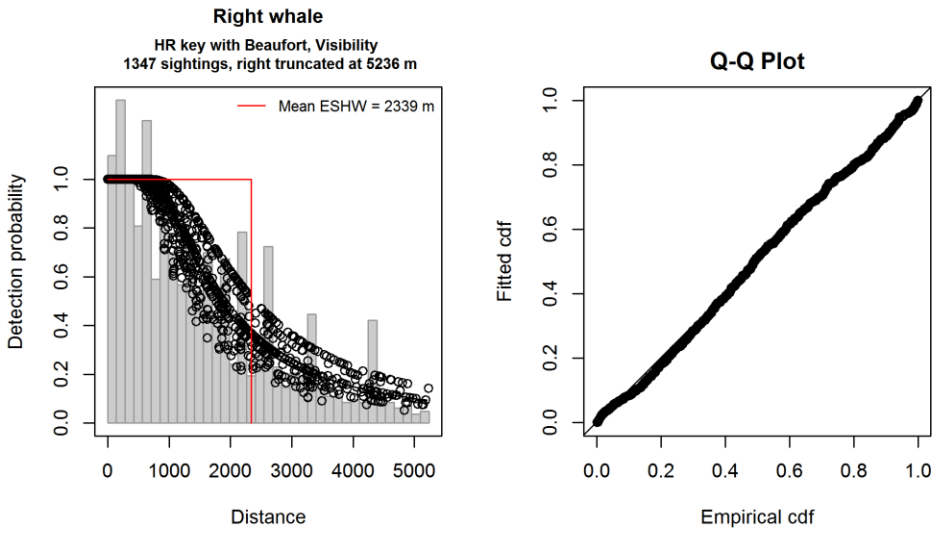


Figure 25. Final detection function applied to 750 ft NARWSS surveys (2003-2016).

5.2.1.1.6 NARWSS 1000 ft (2017-2019) detection function

Starting in 2017, NARWSS changed their protocol to fly at 1000 ft and usually only record sightings of large whales. On certain surveys, minke whales were also recorded. For our analysis, the NARWSS team contributed data collected from 2017 through May 2019. Given the strong, direct effect of altitude on perpendicular sighting distance, we did not pool the 2017-2019 surveys with those made earlier at 750 ft. As we did with the earlier surveys, we excluded survey effort made in Beaufort state 6 or higher, NARWSS quality code x (defined as "unacceptable"), visibility < 3 nmi, or in the Gulf of St. Lawrence. After exclusions, we were left with 179 sightings of right whale groups, after right-truncating the sightings at 4984 m, corresponding to a vertical sighting angle of 86.5°, halfway between the 86° and 87° bins to which sighting angles were rounded. This was a sufficient count of sightings to fit a right-whale-specific detection function, but when we did so, the mean ESHWs from candidate detection functions (not shown here) were about 1700 m, which was substantially less than the mean ESHW (2339 m) of the detection function we selected as best for the earlier surveys conducted at 750 ft. This result did not make sense—we expected that mean ESHW would increase, rather than decrease, after the survey raised its target altitude from 750 to 1000 ft. (Note that some of the decrease can be explained by the use of a shorter truncation distance, but not all of it.)

To reduce the chance that this effect resulted from the much lower quantity of data currently available for the 1000 ft. protocol, we pooled in other species of large baleen whales (minke whales excluded) and sperm whales, and refitted the detection candidate functions. After exclusions, this left us with 897 sightings of large whales, including the 179 right whales. The final detection function utilized three covariates: visibility, quality code, and species name. As with the earlier 750 ft surveys, detection distances increased as visibility increased (Figure 26). Detection distances were also higher when the quality of sighting conditions were coded "excellent" rather than "good", "moderate", or "poor" (Figure 27). Finally, humpback whales exhibited the farthest sighting distances, consistent with their behavior and size, followed by bowhead, fin and right whales, and lastly by sei whales (Figure 28). No sperm whales were sighted.

The mean ESHW (2068 m) of the final detection function (Figure 29) was closer to ESHW (2339 m) of the 750 ft surveys' detection function than were the right-whale-specific detection functions that we initially tried, but it was still smaller. At this time, we cannot explain why this is the case. It is possible that as more sightings are collected at the 1000 ft altitude, detection distances will increase. We plan to revisit this question in future modeling efforts.

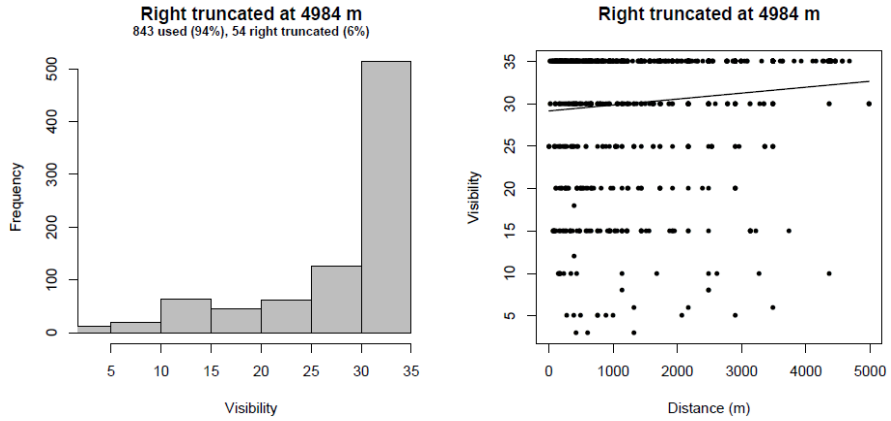


Figure 26. Distribution of visibility (left) and its relationship to perpendicular sighting distance (right) for sightings used in the 1000 ft NARWSS (2017-2019) detection function.

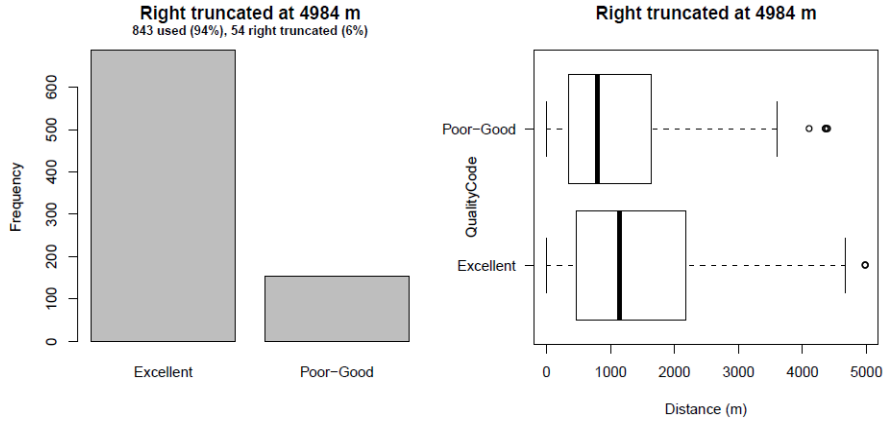


Figure 27. Frequency of sighting quality codes (left) and the distribution of perpendicular sighting distances by sighting quality (right) for sightings used in the 1000 ft NARWSS (2017-2019) detection function.

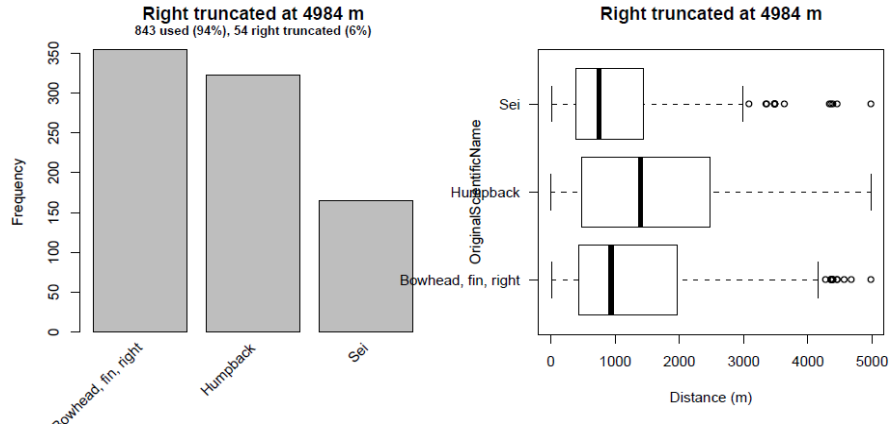


Figure 28. Frequency of species identifications (left) and the distribution of perpendicular sighting distances by species identification (right) for sightings used in the 1000 ft NARWSS (2017-2019) detection function.

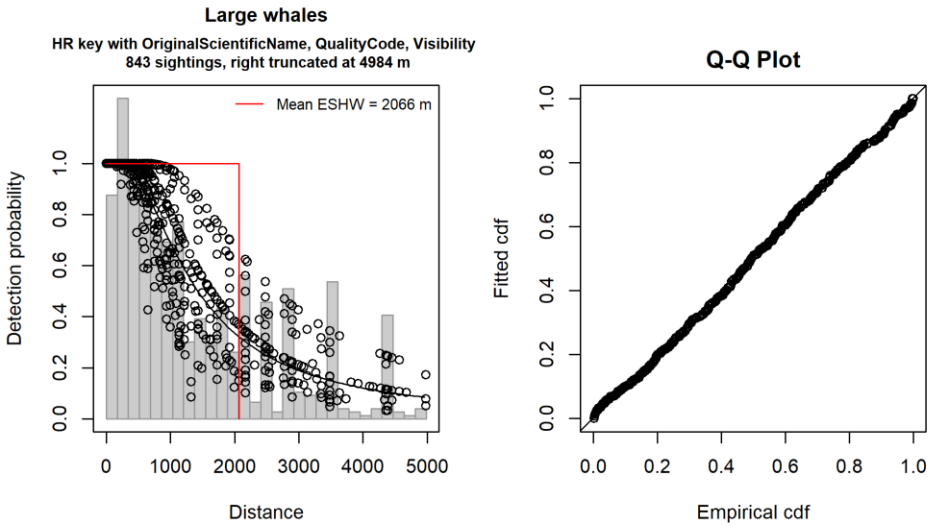


Figure 29. Final detection function applied to 1000 ft NARWSS surveys (2017-2019).

5.2.1.1.7 NYS-DEC NYBWM detection function

The New York State Department of Environmental Conservation (NYS-DEC) New York Bight Whale Monitoring (NYBWM) program flew from March 2017 through February 2020, but only the first two years of the program (March 2017 through February 2019) were available for our analysis (Tetra Tech & Smultea Sciences 2018; Tetra Tech & LGL 2019). Surveys were performed on a Partenavia P.68 with bubble windows. We restricted our analysis to effort collected in Beaufort 5 or less and quality levels of "excellent", "good", "moderate", and "fair", while excluding "poor". Even under this wide range of conditions, only 9 right whales were sighted, so we pooled in large baleen whales (excluding minke whales) and sperm whales, raising the sighting count to 126. We right-truncated the data at 2500 m, dropping 5 sightings (4%).

Despite the bubble windows, histograms of sighting distances showed a drop-off in sightings close to zero distance, suggesting some sightings close to the trackline were missed. The observer team confirmed this effect, indicating that the Partenavia's bubble windows were not particularly deep, making it difficult to maintain close attention below the plane (D. Ireland, pers. comm.). To account for this, we left-truncated the sightings at 100 m, corresponding to a vertical inclination angle of roughly 18° at the target altitude of 1000 ft, and dropping 6 sightings (5%), leaving 115 (91%) available for the detection function.

Exploratory analysis suggested possible influences (not shown here) from two covariates. Beaufort state displayed a similar relationship to that seen with the 750 ft NARWSS surveys, with farthest detection distances at Beaufort 3. The sightings with quality levels of "excellent" or "good" had farther distances than those with quality levels of "moderate" or "fair". However, detection functions that included these covariates did not score as well on AIC, and we ended up selecting a hazard-rate detection function with no covariates or adjustments as best (Figure 30). It is possible that if the third year of sightings were included (in a future analysis), there would be enough that the additional precision afforded by the covariates would yield a better-scoring AIC. We plan to revisit this question in future modeling efforts.

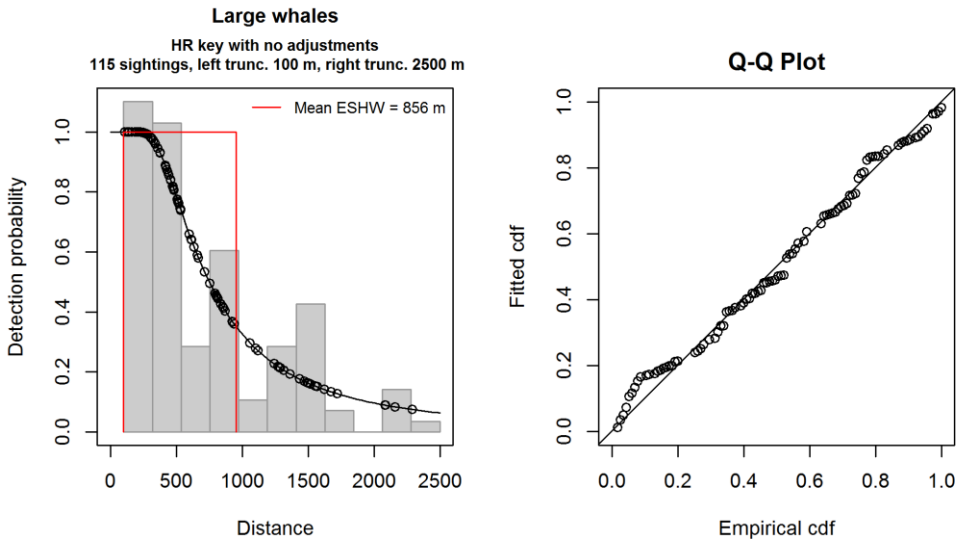


Figure 30. Final detection function applied to the NYS-DEC NYBWM surveys.

5.2.1.1.8 NEAq New England detection function

The New England Aquarium (NEAq) conducted two multi-year series of aerial surveys of waters off southern New England, encompassing a region known as the Massachusetts and Rhode Island Wind Energy Areas (Mass/RI WEAs). The first program, known as the Northeast Large Pelagic Survey Collaborative (NLPSC) (Kraus et al. 2016), spanned 2011-2015. The second program, known as the Marine Mammal Surveys of the Wind Energy Areas (MMS-WEA), was initiated in 2017 (Quintana et al. 2018) and continued into at least 2020. For our analysis, we had the entire NLPSC corpus of surveys and the MMS-WEA surveys for 2017 and 2018. Given the similarities between the programs—MMS-WEA was designed to be a direct follow-on to NLPSC—we pooled them together. For MMS-WEA, we retained the "general" and "condensed" surveys, but excluded the "directed" surveys, which were not sufficiently systematic and were directed to known aggregations of right whales.

Between them, the two programs reported 97 sightings of right whales (Figure 31), which we judged enough to fit a right-whale-specific detection function. However, under the survey protocol that was used, observers did not collect vertical angles with a clinometer or similarly precise method. Instead, angles were rounded into several coarse ranges, necessitating a "binned" or "heaped" analysis (Buckland et al. 2001). When fitting candidate detection functions, we used the bins defined specifically for the NLPSC surveys by Kenney (2019) (under his section 8.98: STRIP).

This problem was compounded by the use of a flat-windowed aircraft (Cessna Skymaster) that resulted in a drop-off in sightings close to the trackline. To try to improve the precision of candidate detection functions, we opted to pool in sightings of other large baleen whales (except minke whales) plus sperm whales (Figure 32), with the idea that it might enable additional covariates to be selected, even if it would not directly mitigate the heaping problem by increasing the variety of perpendicular distance values.

We set the left truncation distance to 71 m, based on prior research that measured the "blind spot" beneath the aircraft for the NLPSC surveys as a 142 m wide strip (Taylor et al. 2014). No sightings were reported in bins closer than this. We set the right truncation distance to 3704 m, corresponding to the outer limit of Kenney's (2019) 1-2 nmi bin, yielding five bins for the detection function. We excluded effort made when the Beaufort state was greater than 4 or the weather code (WX) was something other than "clear", "gray", or "haze". This retained 308 large whale sightings, of which 97 were right whales. The final detection function (Figure 35) utilized two covariates: Beaufort state and glare. Perhaps surprisingly, the largest detection distances were reported at Beaufort 3-4 (Figure 33); however, this peak was consistent with what we observed with the NARWSS program. For glare, the largest distances were reported in "severe" conditions (Figure 34), consistent with observer comments that strong glare backlights whale blows, making them visible at long distances.

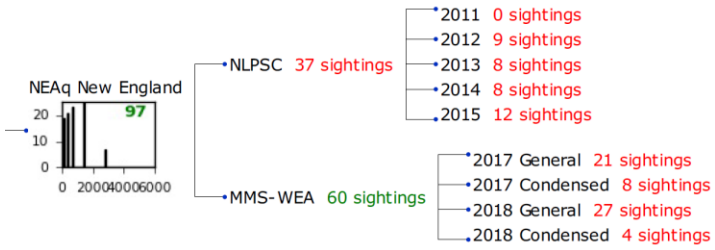


Figure 31. Sightings of right whale groups reported by the NLPSC and MMS-WEA surveys.

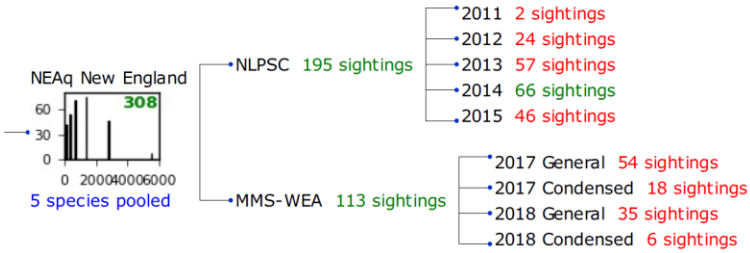


Figure 32. Sightings of large whale groups (all baleen whale species except minke whales, plus sperm whales) reported by the NLPSC and MMS-WEA surveys. These sightings were used to fit the detection function.

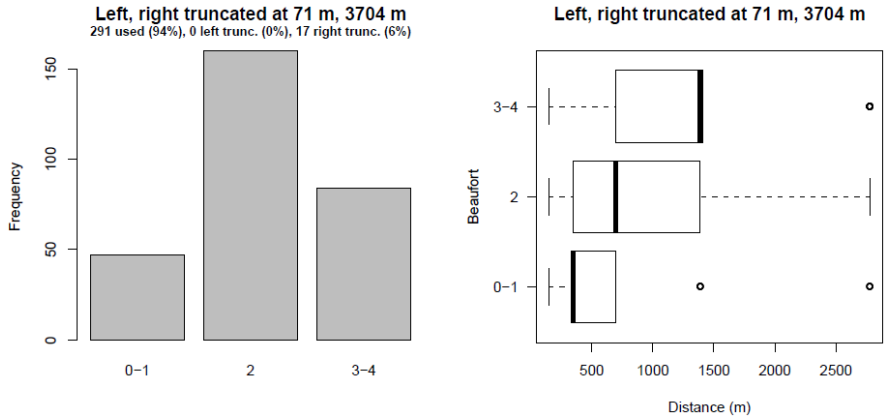


Figure 33. Frequency of categorized Beaufort state (left) and the distribution of perpendicular sighting distances by Beaufort state (right) for sightings used in the NEAq New England surveys' detection function.

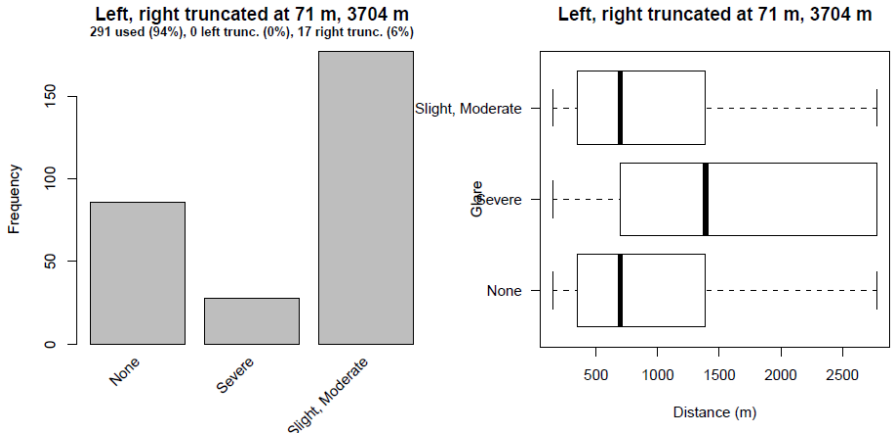


Figure 34. Frequency of categorized glare (left) and the distribution of perpendicular sighting distances by glare category (right) for sightings used in the NEAq New England surveys' detection function.

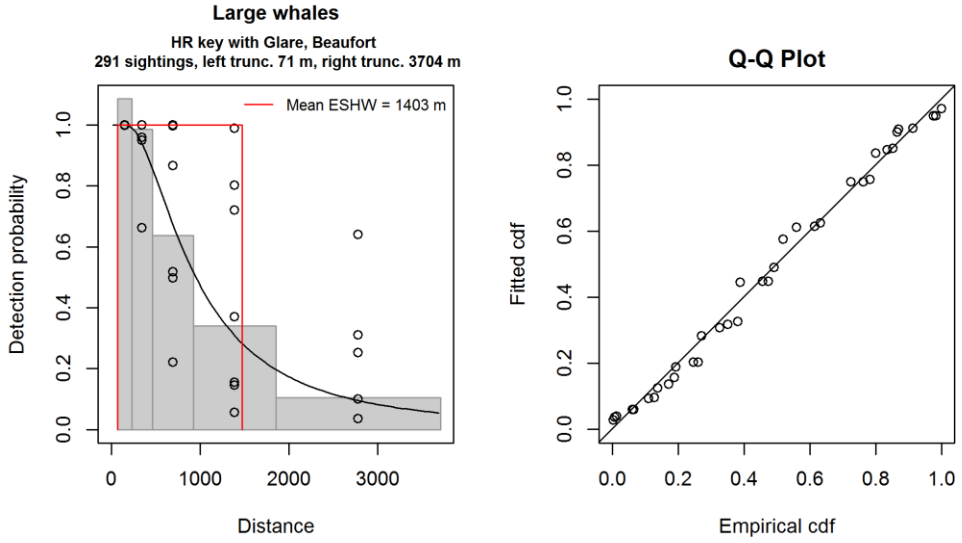


Figure 35. Final detection function applied to the NEAq New England surveys.

5.2.1.1.9 UNCW Protocol detection function

For more than 15 years, the Navy has funded a group of organizations that included the University of North Carolina, Wilmington (UNCW), the Virginia Aquarium & Marine Science Center (VAMSC), and HDR, Inc. to survey various Navy operating areas in the mid-Atlantic and southeast U.S. These organizations adopted a common closing-mode survey protocol flown at a target altitude of 1000 ft, usually with flat-windowed Cessna Skymaster aircraft, with perpendicular distances estimated for every marine mammal group usually by overflying it and collecting a GPS position. Certain surveys measured declination angles instead, using marks on windows or wing struts.

Collectively, between 2002-2018, these surveys sighted only 36 right whales (Figure 36). We judged this insufficient for fitting a detection function, especially because some protocol differences did exist between the earliest surveys and those that came later. To boost sighting counts, we pooled in large baleen whales (except minke whales) and sperm whales, yielding 198 sightings in total (Figure 37), including the 36 right whales.

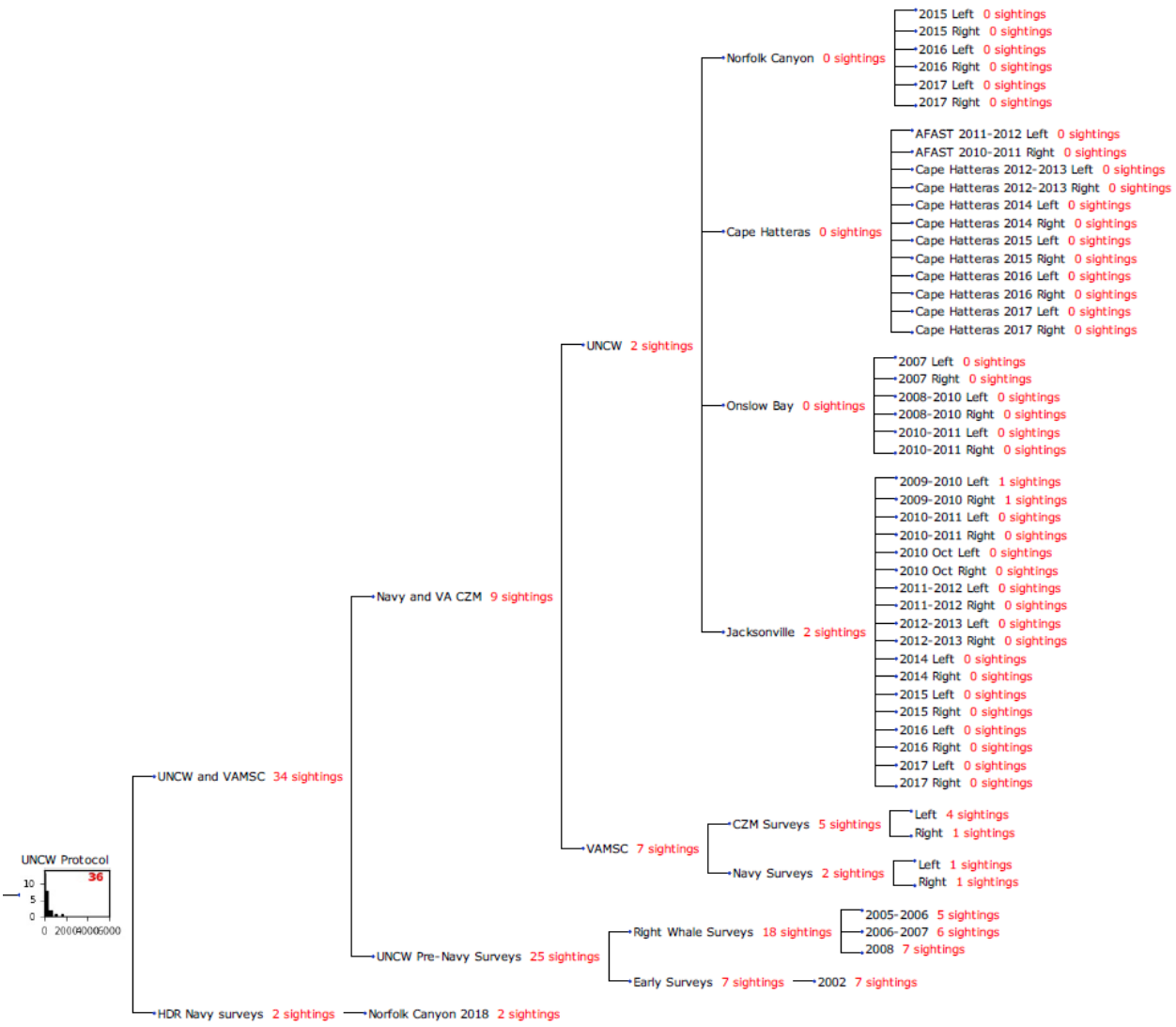


Figure 36. Sightings of right whale groups reported by surveys within the UNCW Protocol branch of the detection hierarchy for aerial surveys. Because of these small sighting counts, we fitted no right-whale-specific detection functions to any nodes in this branch of the hierarchy. Instead, we fitted detection functions to pools that included other large whales (Figure 37).

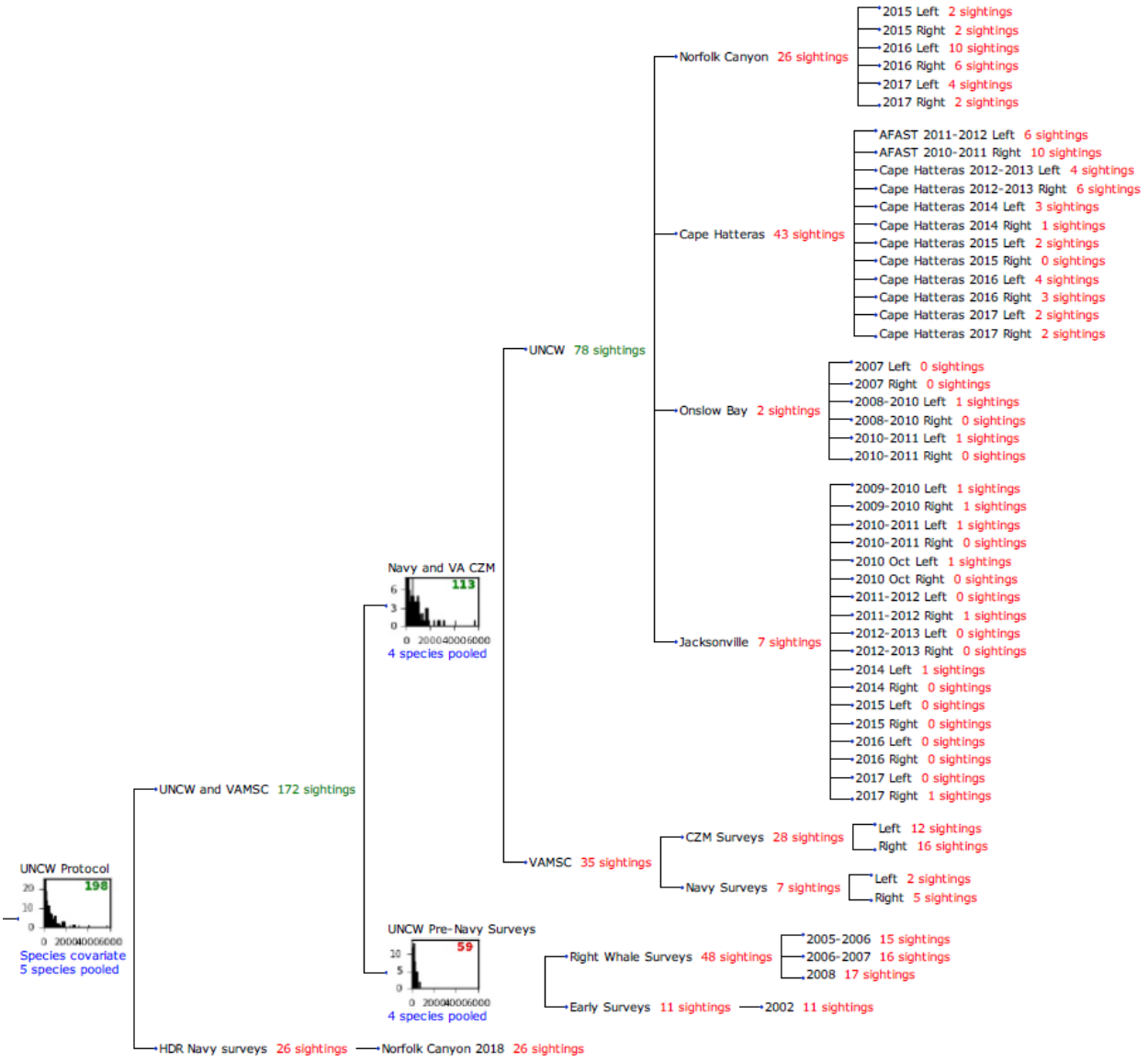


Figure 37. Sightings of large whale groups (all baleen whale species except minke whales, plus sperm whales) reported by surveys within the UNCW Protocol branch of the detection hierarchy for aerial surveys. The three histograms show nodes where detection functions were fitted; we present the details of these below.

Within this branch of the hierarchy, we fitted three detection functions. For the surveys conducted by UNCW and VAMSC, we split what we called the UNCW Pre-Navy Surveys from the more recent surveys conducted by UNCW and VAMSC. The Pre-Navy surveys were focused more on monitoring right whales migrating through in-shore areas than were the subsequent surveys, which focused on Navy operating areas that spanned the shelf break. The older surveys also had some protocol differences that appeared to yield shorter detection distances (Figure 37). We fit separate detection functions for these two pools. This left only the HDR Navy surveys, for which we only had one year of data, the 2018 survey of the Norfolk Canyon study area. From this we only had 26 sightings, which was too few to fit a detection function. Being uncertain about which UNCW-protocol surveys were the best proxy to pool with the HDR surveys, we pooled all of them (but applied the resulting detection function just to the HDR surveys).

In the current section, we document the "UNCW Protocol" detection function, developed for and applied only to the HDR surveys. In the two following sections (5.2.1.1.10 and 5.2.1.1.11), we document the two other detection functions, developed for and applied to UNCW and VAMSC surveys, as shown in Figure 37.

For the UNCW Protocol detection function, we right-truncated at 2000 m, dropping 7 sightings (4%) and retaining 191 (96%) for fitting the detection function. In the data provided to us, HDR did not include any detectability covariates (not even Beaufort state), leaving species identification as the only covariate to try. Exploratory analysis suggested a strong difference in detection distances for right and sei whales as compared to humpback, fin, and sperm whales (Figure 38). This differed from the northern surveys, where right whales were more similar to humpback and sperm whales. One possible interpretation was that many right whales sighted on these surveys of the mid-Atlantic and southeast were migrating, and perhaps migratory behaviors made them harder to detect than the behaviors exhibited at the northern feeding grounds. Alternatively, we noted that most of the right whale sightings (25 of 36) were reported by the UNCW Pre-Navy Surveys, which exhibited generally shorter detection distances than the more recent UNCW and VAMSC surveys. Perhaps protocol differences between the surveys were the cause for this phenomenon. Still, the Pre-Navy surveys only sighted 59 whales in total, making it hard to draw definitive conclusions.

Ultimately, we elected to include species name as a covariate under the categorization shown (Figure 38). The "best" detection function (Figure 39) retained this covariate, confirming its strong correlation with detection probability, and scored 9.8 AIC units better than the second-ranked candidate detection function, which included no covariates. In general, if adding a covariate improves a model by 2 or more AIC units, then the addition of the covariate is considered a good tradeoff in goodness-of-fit vs. parsimony, so we selected the function that included the covariate. We plan to revisit this in the future once HDR delivers more data, which might facilitate additional alternatives.

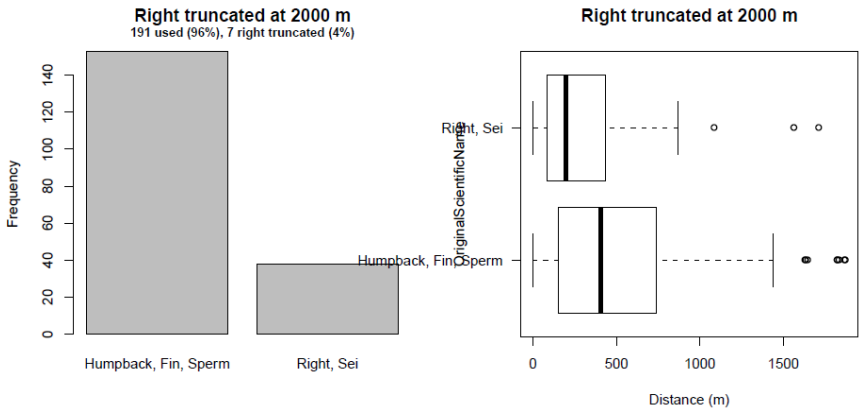


Figure 38. Frequency of species identifications (left) and the distribution of perpendicular sighting distances by species identification (right) for sightings used in the UNCW Protocol detection function.

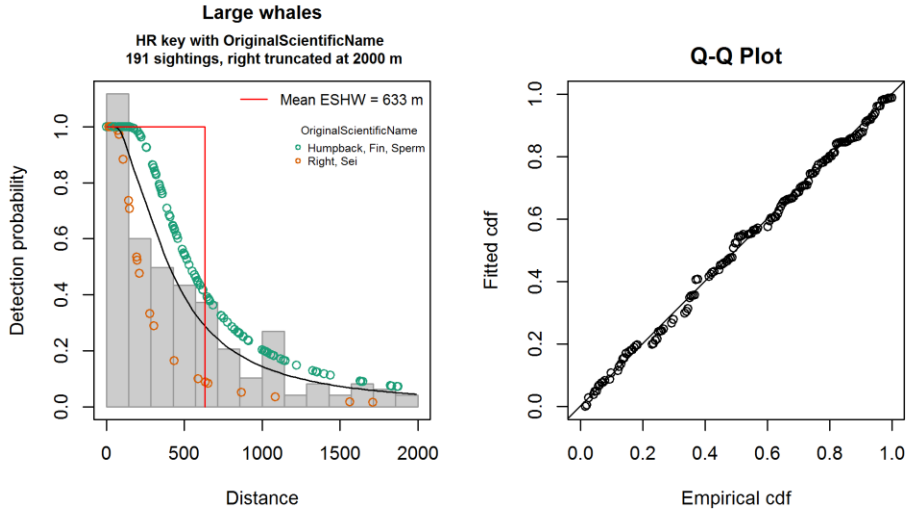


Figure 39. Final UNCW Protocol detection function, applied only to the HDR Navy surveys.

It is worth noting that, similar to Roberts et al. (2016a), we did not apply left truncation in the UNCW Protocol detection function or the two that came below it in the hierarchy, despite these surveys occurring with Cessna Skymaster aircraft that had flat windows. This differed from our treatment of some of the other surveys that used Skymasters with flat windows, e.g. the NEAq New England surveys (section 5.2.1.1.8). It also differed from McLellan et al. (2018), who applied a left truncation of 149 m, although that study was of beaked whales. As in our prior analysis (Roberts et al. 2016a), we detected no drop-off in sighting counts close to the trackline. Given this lack of empirical evidence that left truncation was needed, coupled with the rarity of right whale sightings reported during these programs and the mid-Atlantic generally (making them especially important to retain for use in the spatial model), we again decided against applying left-truncation in these detection functions.

5.2.1.1.10 Navy and VA CZM detection function

For surveys conducted by UNCW and VAMSC in Navy operating areas, and by VAMSC in areas of interest to the State of Virginia, we retained survey effort where Beaufort state was 4 or less, and weather conditions were classified as "clear", "partly cloudy", "continuous layer of clouds", or "haze" (we excluded "rain", "fog", and "other"). Under these conditions, the pooled surveys reported 113 large whales, 9 of which were right whales. Exploratory analysis (not shown) suggested that the Beaufort state and visibility covariates might influence detectability, but AICs of candidate detection functions ranked them worse than a detection function with no covariates, indicating that effects were statistically marginal, probably owing to the relatively small number of sightings available. We also investigated using a species name covariate, but the distributions of distances for non-right-whale species were similar, and only 9 right whales were sighted, which we judged insufficient for defining a category. (We noted with some interest that the distribution of distances for these 9 was lower than the other species.) For the final detection function (Figure 40), we selected the candidate which scored the best on AIC, which was a half-normal function with no covariates.

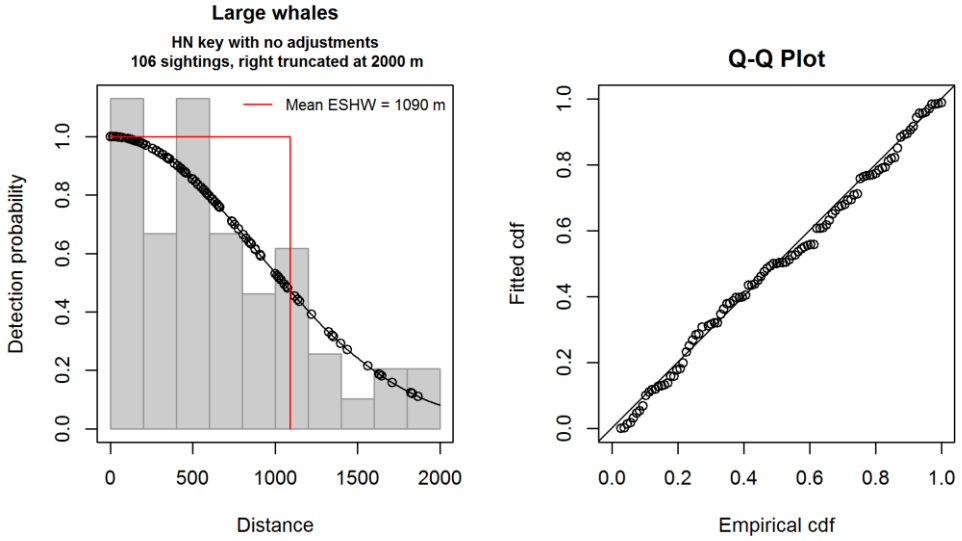


Figure 40. Final detection function applied to the Navy and VA CZM surveys.

5.2.1.1.11 UNCW Pre-Navy Surveys detection function

On some of UNCW's earlier surveys, perpendicular distances were estimated from vertical angles obtained by the closest 10° declination bin marked on wing struts when the sighting was perpendicular to the aircraft. This necessitated a "binned" or "heaped" analysis (Buckland et al. 2001). We placed cutpoints at the 10° intervals that defined the edges of the bins and used 70° as the right-truncation angle, corresponding to a perpendicular distance of ~838 m, leaving 7 bins for fitting the detection function. At this truncation distance, only 59 sightings were available, but 25 were right whales. Exploratory analysis suggested detection probability decreased as Beaufort state increased (Figure 41), which was inconsistent with what we observed on surveys in the northeast but consistent with some of the right whale Early Warning System (NARW EWS) surveys discussed in following sections. We judged that there

were too few sightings to test for a species effect on detection probability, especially given that bins were used rather than clinometers. The final detection function was a hazard-rate function with Beaufort state as a continuous covariate as opposed to a categorical covariate (Figure 42).

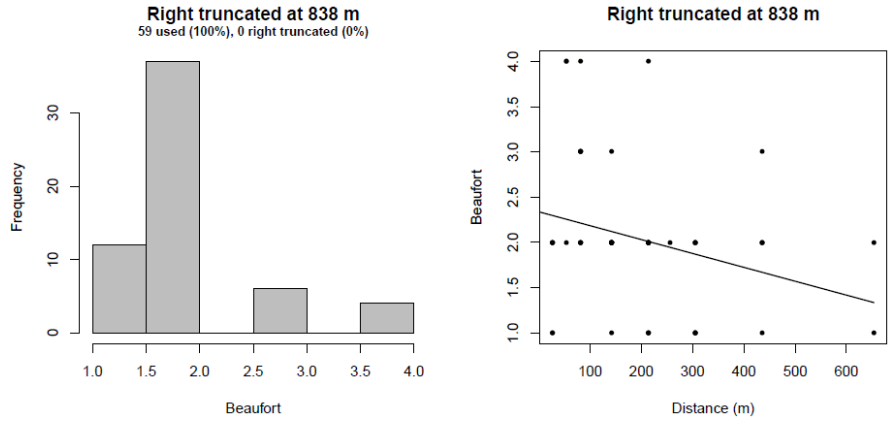


Figure 41. Distribution of Beaufort state (left) and its relationship to perpendicular sighting distance (right) for sightings used in the UNCW Pre-Navy Surveys' detection function.

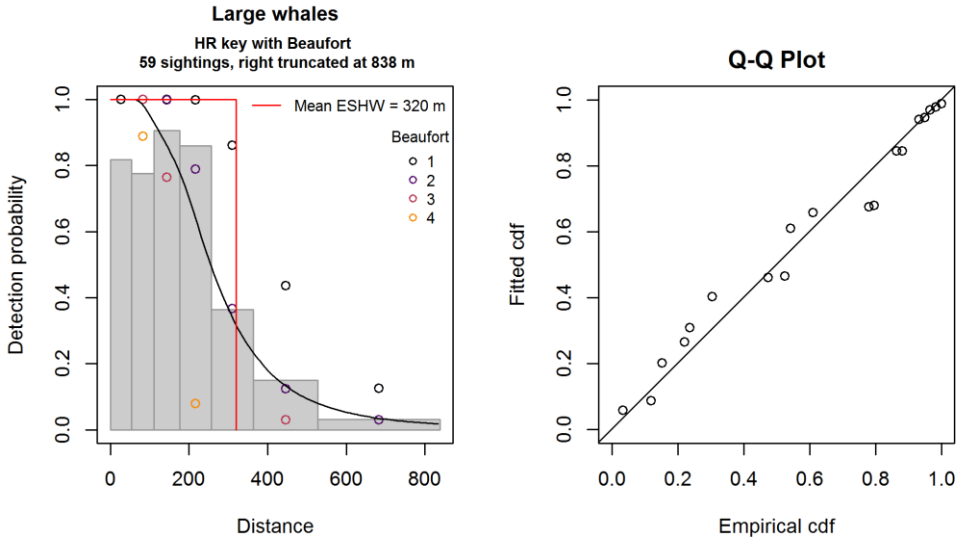


Figure 42. Final detection function applied to the UNCW Pre-Navy Surveys.

5.2.1.1.12 FWRI detection function

Since the 1990s, collaborators participating in the right whale Early Warning System (NARW EWS) have surveyed the right whale calving grounds every winter, focusing on the core months of December-March, with some effort coming outside this period, and the core regions in Florida and Georgia, with substantial but less intense effort in the Carolinas. Most surveys occurred with flat-windowed Cessna Skymaster aircraft, with flights in the vicinity of Georgia made with bubble-windowed de Havilland Twin Otters. In the 2003/04 calving season, the teams adopted a common flight altitude (1000 ft) and protocol (two observers, closing mode, sighting distances estimated from GPS locations of sightings). Based on recommendations from our collaborators, who emphasized the importance of the standardization that occurred starting in 2003/04, we began our integration of the NARW EWS surveys at that season. For the model presented in this report, we included data extending through the end of the 2018/19 season.

The NARW EWS teams collectively reported over 2500 sightings of right whale groups for the modeled period (2003/04-2018/19), which enabled us to fit right-whale-specific detection functions. (Indeed, the surveys sighted few other large whales, so pooling them in offered a very limited benefit for the purpose of modeling right whales anyway.) To account for differences in team protocols, personnel, and aircraft, and regional differences in right whale behavior (e.g. whales transiting in South Carolina vs. calving in Florida) we fitted separate detection functions for each team. Across all teams, we excluded effort made in Beaufort state 4 or higher, in keeping with our prior models that used these data (Roberts et al. 2018) and with the analysis approach historically taken by the Fish and Wildlife Research Institute (FWRI), e.g. by Gowan et al. (2014). This was a conservative criterion relative to what we did with other surveys, but so much effort was available that we judged it a reasonable tradeoff to exclude some made in rougher seas in order reduce uncertainty in the resulting detection functions. We also excluded effort when visibility was less than 3 nmi or the weather code (WX) was something other than "clear", "gray", or "haze".

We also excluded any "verification" flights or effort segments that were directed toward locations where right whales were recently reported, and retained effort only executed on systematic transects. Occasionally, "verification" sightings were made during the course of systematic transects. Starting with the 2010/11 season, FWRI systematically tracked all such sightings, for all teams. We excluded all of these sightings but retained the systematic transects they occurred on (which often sighted other whales, not previously known). For seasons prior to 2010/11, we excluded sightings we judged likely to be verifications, through a combination of manual review and classification modeling, using observer comments logged with the sightings and checking parameters such as the distance and time taken to reach the sighting. Finally, some surveys executed a systematic shore-paralleling transect close to shore at the beginning or end of the day. Because these ran parallel to important ecological gradients (e.g. the seafloor depth), we excluded them to minimize the chance of biasing in the spatial model, except in the south where effort was very sparse.

In this section, we document the detection function for surveys conducted by FWRI, and document those for the other NARW EWS teams in following sections. After exclusions, surveys conducted by FWRI reported 818 sightings of right whale groups (Figure 43). FWRI used a flat-windowed Cessna Skymaster, and histograms of perpendicular sighting distances (Figure 44) showed a characteristic drop-off in sighting frequency close to the trackline indicating that poor visibility of the area below the plane had likely caused sightings to be missed. Accordingly, we applied left-truncation when fitting candidate detection functions. After exploratory analysis, we selected a left-truncation distance of 142 m, corresponding to a vertical inclination angle of 25° at 1000 ft altitude, coincidentally very similar to that (149 m) used by McLellan et al. (2018) for the same aircraft type and altitude, albeit for beaked whales.

After left-truncation, which dropped 35 sightings (4%), and right-truncation, which was done at 4800 m and dropped 20 sightings (2%), we were left with 763 sightings for fitting the detection function. Exploratory analysis suggested that covariates for Beaufort state, glare, weather, and cloud cover might correlate with detection probability. The final detection function (Figure 47) included Beaufort state and the weather code (WX). Lower detection distances were observed in higher sea states (Beaufort 2-3) than in lower sea states (Beaufort 0-1) (Figure 45). Highest detection distances were observed when the weather was classified as "clear", followed by "haze" and then "gray" (Figure 46).

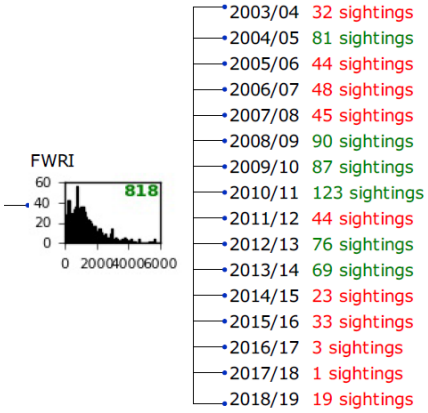


Figure 43. Sightings of right whale groups reported by NARW EWS surveys conducted by FWRI.

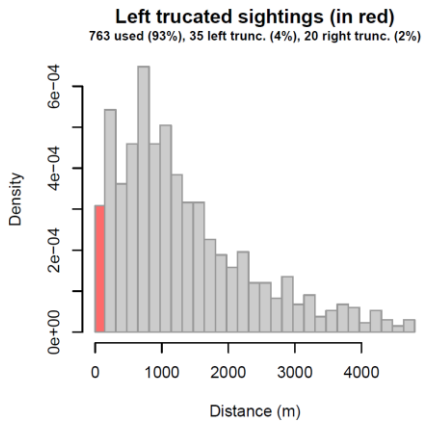


Figure 44. Distribution of perpendicular sighting distances for right whale sightings reported by FWRI surveys. Note the fall-off in detections close to 0 m. We left-truncated the sightings at 142 m, dropping the sightings shown in red.

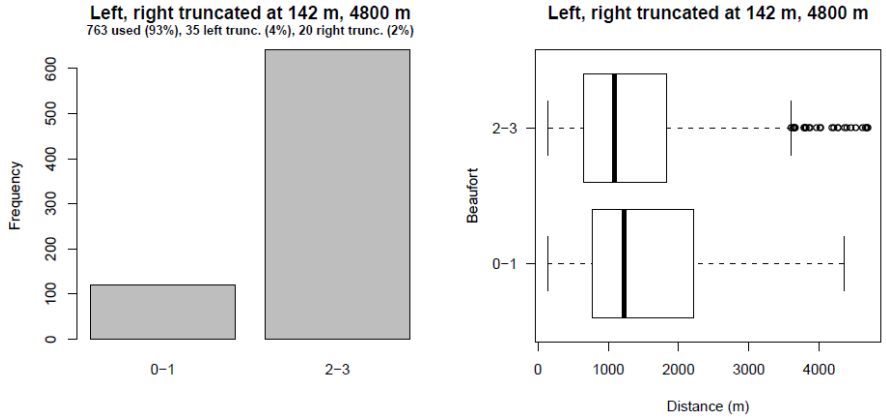


Figure 45. Frequency of categorized Beaufort state (left) and the distribution of perpendicular sighting distances by Beaufort state (right) for sightings used in the FWRI detection function.

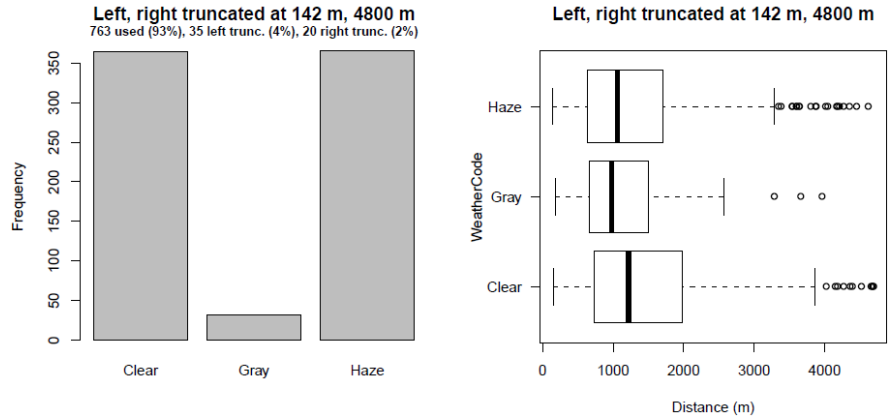


Figure 46. Frequency of weather codes (left) and the distribution of perpendicular sighting distances by weather codes (right) for sightings used in the FWRI detection function.

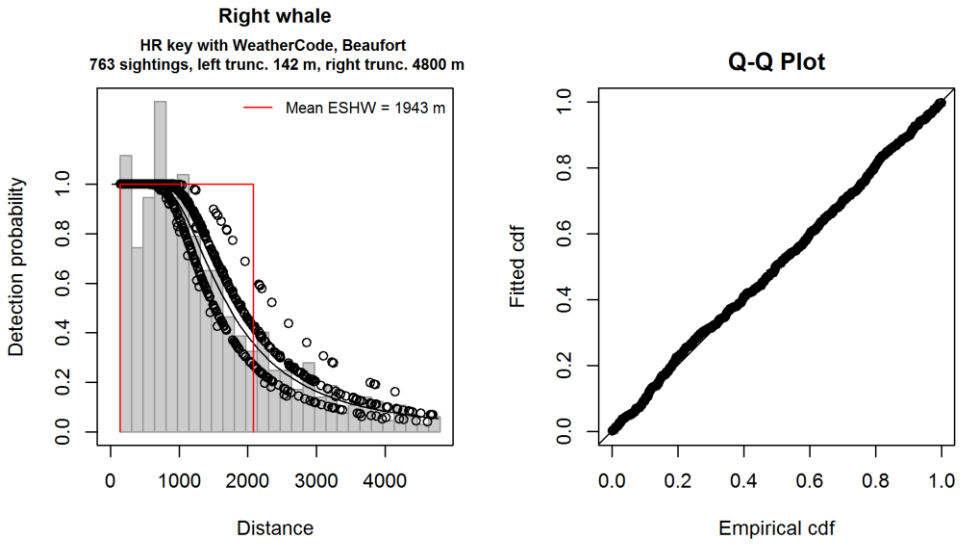


Figure 47. Final detection function applied to the FWRI surveys.

5.2.1.1.13 NEAq NARW EWS detection function

The New England Aquarium conducted NARW EWS surveys of core calving habitat during the seasons 2003/04 through 2009/10 (Figure 48). Owing to the large number of right whales that visited the calving grounds during that period, NEAq reported the largest number of right whale sightings among the EWS teams we analyzed (Figure 48).

Like FWRI, NEAq flew a Cessna Skymaster with flat windows, but in contrast, histograms of sighting counts by detection distances did not show a reduced count close to the trackline. Instead, sighting counts were slightly "spiked" at very short distances (Figure 48, Figure 51), which would suggest that some sightings were rounded to a distance of zero. We observed this phenomenon with the WLT/SSA/CMARI Twin Otter surveys and were able to track down the root cause (see section 5.2.1.1.15), but the effect was much less pronounced with NEAq's surveys and we could not identify a root cause. Ultimately, we could not explain why NEAq's surveys showed no drop-off in detection rates below the aircraft despite the flat windows, and speculate that perhaps the pilots were cueing observers to sightings on the trackline.

We right-truncated at 4800 m, dropping 38 sightings (4%), leaving 907 sightings for fitting the detection function. Exploratory analysis suggested that covariates for Beaufort state, weather, and cloud cover might correlate with detection probability. The final detection function (Figure 51) included Beaufort state and cloud cover. Lower detection distances were observed in Beaufort state 3 than in Beaufort states 0-2 (Figure 49). Higher detection distances were observed when the cloud cover was classified as "clear" or "scattered" than "broken" or "overcast" (Figure 50).

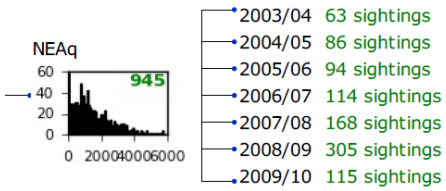


Figure 48. Sightings of right whale groups reported by NARW EWS surveys conducted by New England Aquarium.

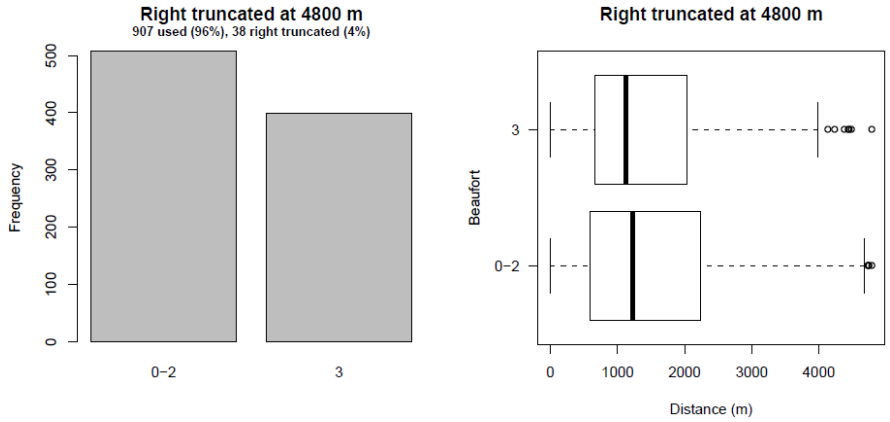


Figure 49. Frequency of categorized Beaufort state (left) and the distribution of perpendicular sighting distances by Beaufort state (right) for sightings used in the NEAq NARW EWS detection function.

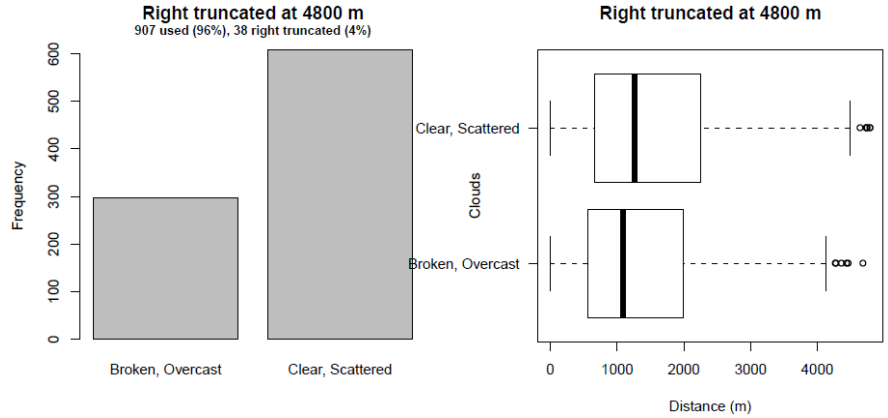


Figure 50. Frequency of categorized cloud cover code (left) and the distribution of perpendicular sighting distances by cloud cover code (right) for sightings used in the NEAq NARW EWS detection function.

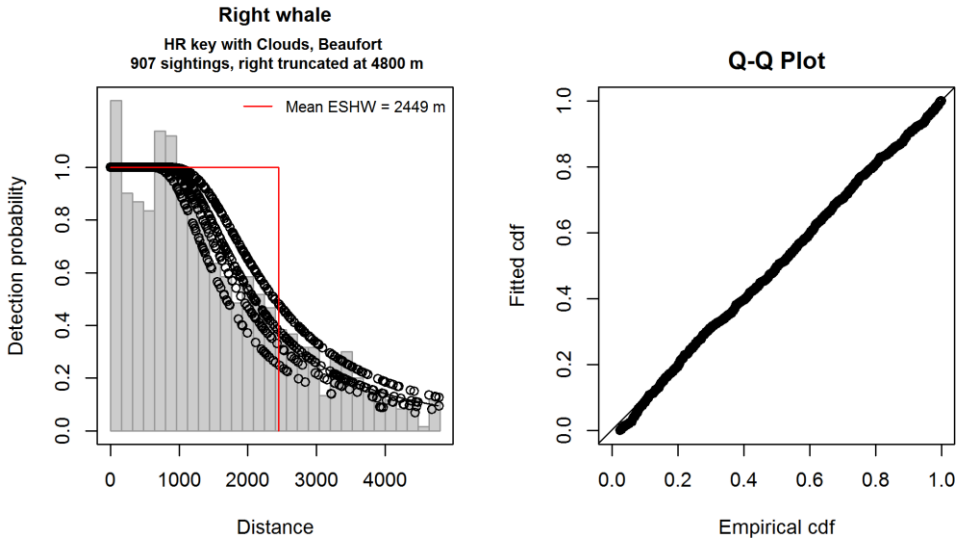


Figure 51. Final detection function applied to the NEAq NARW EWS surveys.

5.2.1.1.14 WLT/SSA/CMARI Skymaster detection function

In South Carolina and, less frequently, North Carolina, the collaborator first known as Wildlife Trust (WLT), then the Sea to Shore Alliance (SSA), and then the Clearwater Marine Aquarium Research Institute (CMARI) conducted aerial surveys using a Cessna Skymaster. These surveys began in the 2004/05 season and were discontinued after the 2012/13 season. They sighted substantially fewer right whales than the surveys to the south, but nonetheless played a very important role in helping to determine how density varied approaching the core of the calving grounds.

After applying the same exclusions we applied to the FWRI surveys (section 5.2.1.1.12), we obtained 175 sightings of right whale groups (Figure 52). Similar to FWRI, histograms of perpendicular sighting distances (Figure 53) showed a characteristic drop-off in sighting frequency close to the trackline indicating that poor visibility of the area below the plane had likely caused sightings to be missed. Accordingly, we applied left-truncation when fitting candidate detection functions, using the same 142 m truncation distance as we did with FWRI.

After left-truncation, which dropped 9 sightings (5%), and right-truncation, which was done at 3500 m and dropped 11 sightings (6%), we were left with 155 sightings for fitting the detection function. The final detection function (Figure 56) included Beaufort state and glare. A "hump-shaped" relationship between detection distances and sea state was observed, with highest distances found at Beaufort state 2 (Figure 54). Similar to the NEAq surveys in New England, highest distances were observed with "severe" glare, followed by "none" or "slight", and finally "moderate" (Figure 55), again supporting the suggestion that strong glare backlights whale blows, making them visible at long distances.

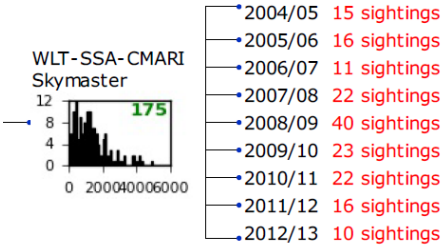


Figure 52. Sightings of right whale groups reported by NARW EWS surveys conducted in Cessna Skymasters by Wildlife Trust (WLT), Sea to Shore Alliance (SSA), and Clearwater Marine Aquarium Research Institute (CMARI).

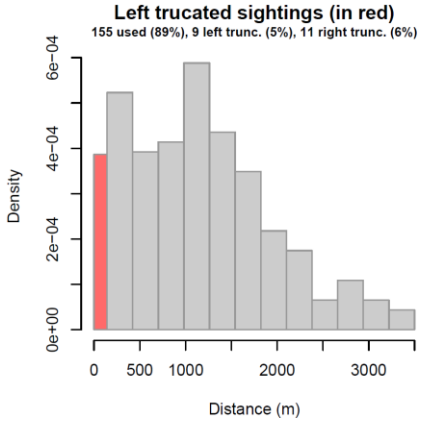


Figure 53. Distribution of perpendicular sighting distances for right whale sightings reported by WLT/SSA/CMARI Skymaster surveys. Note the fall-off in detections close to 0 m. We left-truncated the sightings at 142 m, dropping the sightings shown in red.

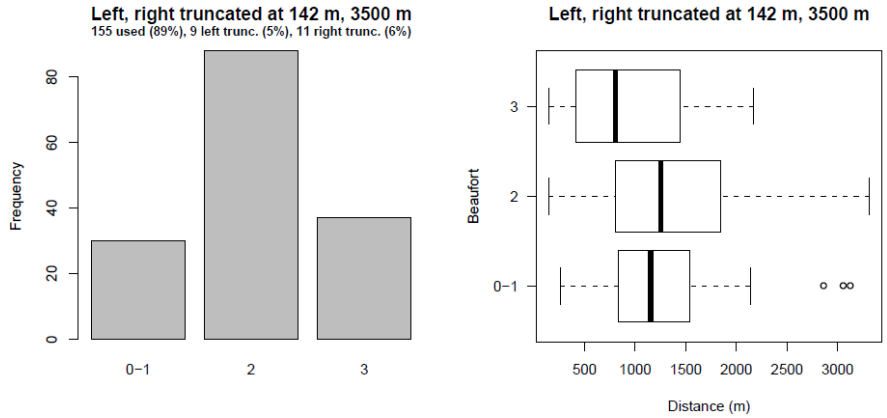


Figure 54. Frequency of categorized Beaufort state (left) and the distribution of perpendicular sighting distances by Beaufort state (right) for sightings used in the WLT/SSA/CMARI Skymaster surveys' detection function.

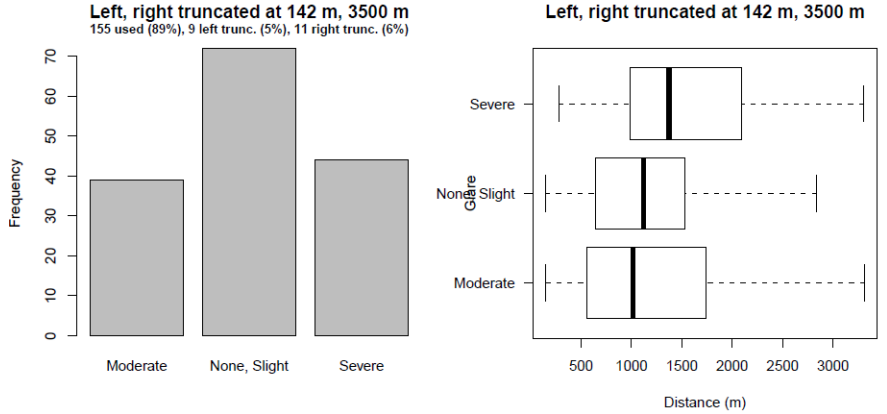


Figure 55. Frequency of glare codes (left) and the distribution of perpendicular sighting distances by glare codes (right) for sightings used in the WLT/SSA/CMARI Skymaster surveys' detection function.

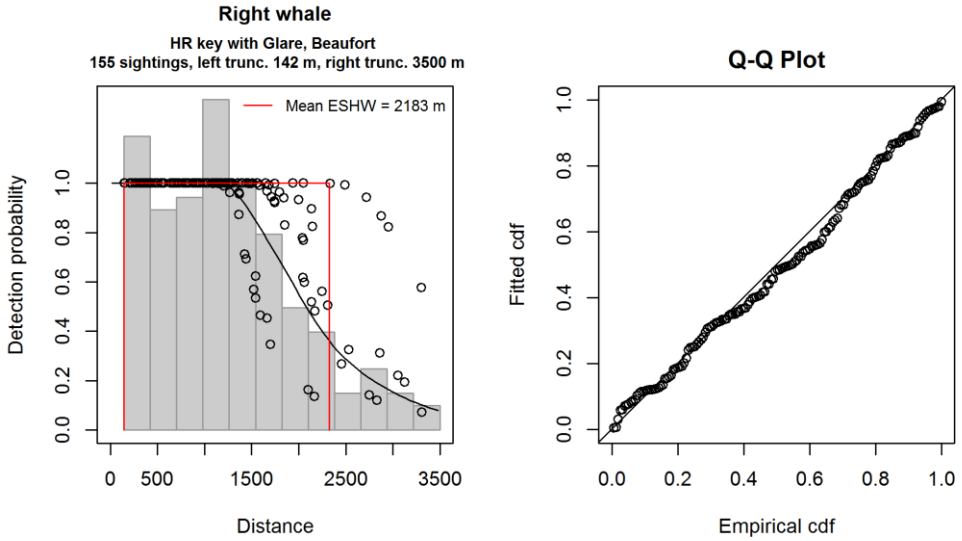


Figure 56. Final detection function applied to the WLT/SSA/CMARI Skymaster surveys.

5.2.1.1.15 WLT/SSA/CMARI Twin Otter detection function

WLT/SSA/CMARI also conducted surveys with a de Havilland Twin Otter, in the vicinity of Georgia, between the FWRI and NEAq surveys conducted to the south and their own surveys conducted with a Skymaster to the north. Unlike the Skymaster, the Twin Otter had bubble windows, which appeared to obviate the need for left truncation, consistent with the Twin Otter surveys conducted by other collaborators farther north. However, the WLT/SSA/CMARI surveys exhibited a different problem: a spike in detections near distance 0 (Figure 57, Figure 59).

In shipboard surveys, a spike in detections at very close distances can indicate an animal movement problem in which animals approached the vessel, e.g. dolphins coming to bow-ride, before observers detected them. But this explanation made no sense for an aerial survey. Upon further investigation, we traced the spike to an unusual number of sightings with a distance of *exactly* 0. In practice, a distance of exactly 0 should be impossible on these surveys, for which perpendicular distance was calculated trigonometrically from the trackline to a GPS point logged while overflying the sighting. The probability of a GPS point being obtained precisely along the trackline is vanishingly small. It turned out that for these distance 0 sightings, the GPS point had been logged while the plane was still on the trackline, before it broke to circle, and many of the observer comments for these sightings had exclamations similar to "whale surfaced directly under the plane!"

Technically, this practice of rounding to zero violated one of the assumptions of distance sampling, that distances be accurately estimated (Buckland et al. 2001). Although it is likely that these "directly below the plane" sightings had distances close to zero, the rounding introducing a small bias. We consulted with distance sampling experts at the University of St. Andrews to determine the appropriate remedy. They reviewed our candidate detection functions and noted with approval that the hazard-rate model "chopped through" the spike (Figure 59) and provided a realistic shoulder, so we kept this result. Most likely, the sightings at distance 0 would have been spread over a few hundred meters on either side of the trackline, reducing the count in the left-most distance bin and boosting the ones next to it.

After applying the same exclusions we applied to the FWRI surveys (section 5.2.1.1.12), we obtained 635 sightings of right whale groups (Figure 57). After right-truncation, which was done at 3500 m and dropped 13 sightings (2%), we were left with 622 sightings for fitting the detection function. The final detection function (Figure 59) included Beaufort state as the only covariate. Detection distances increased as sea state increased, which was evident in exploratory analysis (Figure 58) as well as the final detection function. The peak at Beaufort state 3 was consistent with what we observed with the NARWSS surveys.

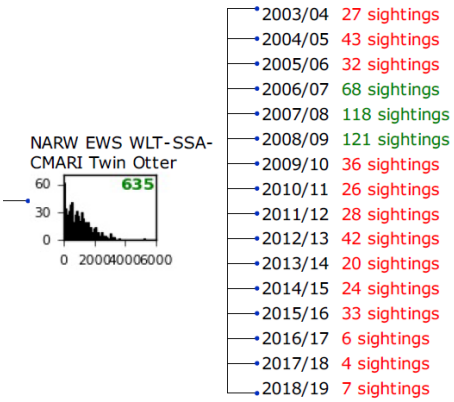


Figure 57. Sightings of right whale groups reported by NARW EWS surveys conducted in de Havilland Twin Otters by WLT, SSA, and CMARI.

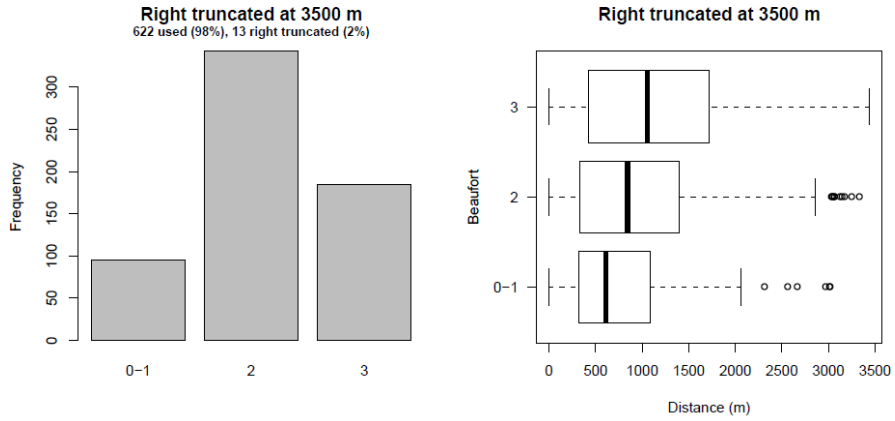


Figure 58. Frequency of categorized Beaufort state (left) and the distribution of perpendicular sighting distances by Beaufort state (right) for sightings used in the WLT/SSA/CMARI Twin Otter surveys' detection function.

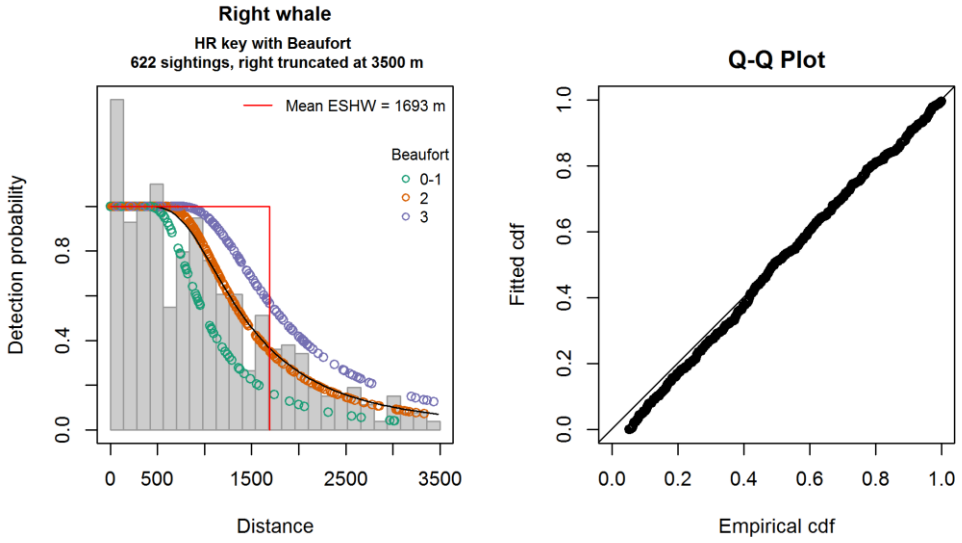


Figure 59. Final detection function applied to the WLT/SSA/CMARI Twin Otter surveys.

5.2.1.2. Shipboard surveys

Figure 60 shows the detection hierarchy developed for shipboard surveys. We split the surveys first by the visual observation method (pedestal-mounted binoculars or naked eyes), then by organization, and then by survey program. Relatively few right whales were sighted (Figure 60), owing to most shipboard surveys having occurred far offshore over the continental slope and abyss, which is not known to be prime right whale habitat. There were three main exceptions, all from NEFSC. First, the HB 07-09 cruise, provided by NEFSC to us for the first time in this analysis (it was not available for our prior right whale models), was a summer survey of the northern Gulf of Maine that was focused on harbor porpoises but also sighted many right whales near the lower Bay of Fundy and along Nova Scotia. HB 07-09 was a two-team survey in which one team used naked eyes and the other used binoculars; we used the binocular team's sightings for the analysis here. The GU 14-02 cruise was conducted in spring and included some transects on the continental shelf, sighting a fair number of right whales. Finally, AJ 99-02 was a naked eye, harbor porpoise survey, also in the northern Gulf of Maine, but was performed in 1999, which was before 2003, the starting year of our spatial model, so we ended up excluding it from our analysis.

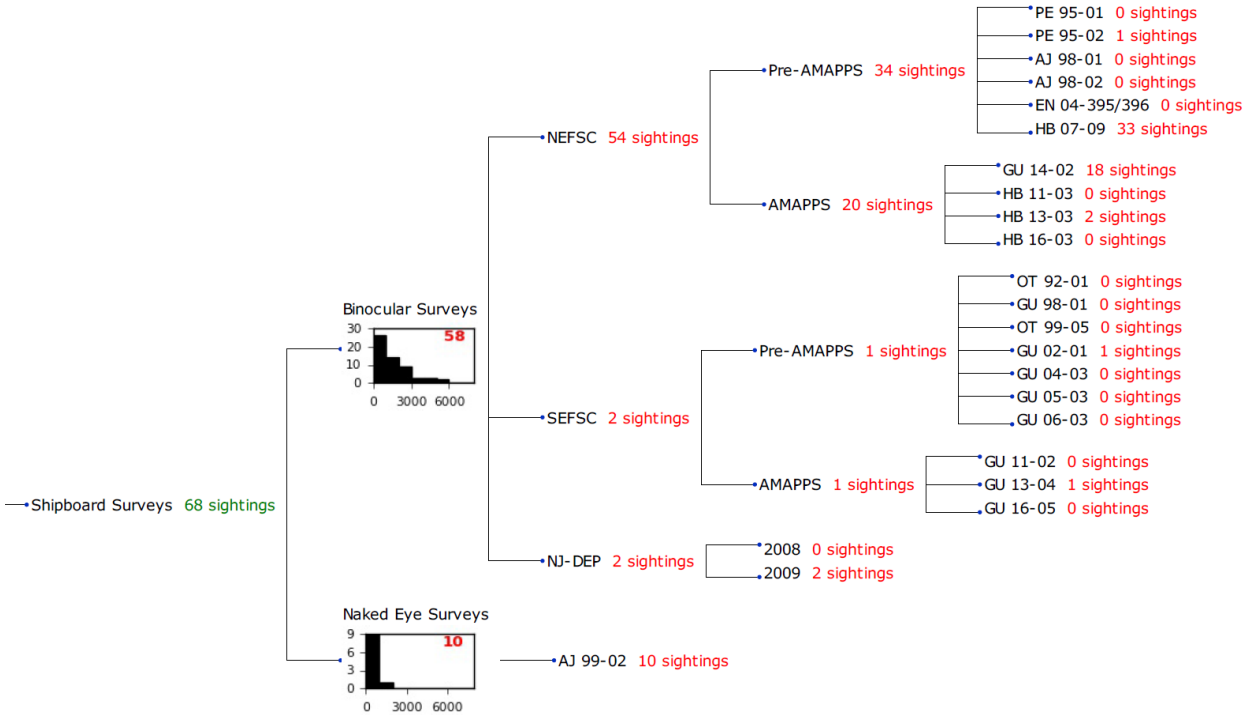


Figure 60. Detection hierarchy for shipboard surveys, shown here with counts of sightings of right whale groups. Because of these small sighting counts, we fitted no right-whale-specific detection functions in this hierarchy. Instead, we fitted detection functions to pools that included other large whales (Figure 61). Note that the detection hierarchy included one naked eye survey, but this was not used as it occurred before 2003, the starting year of our spatial model.

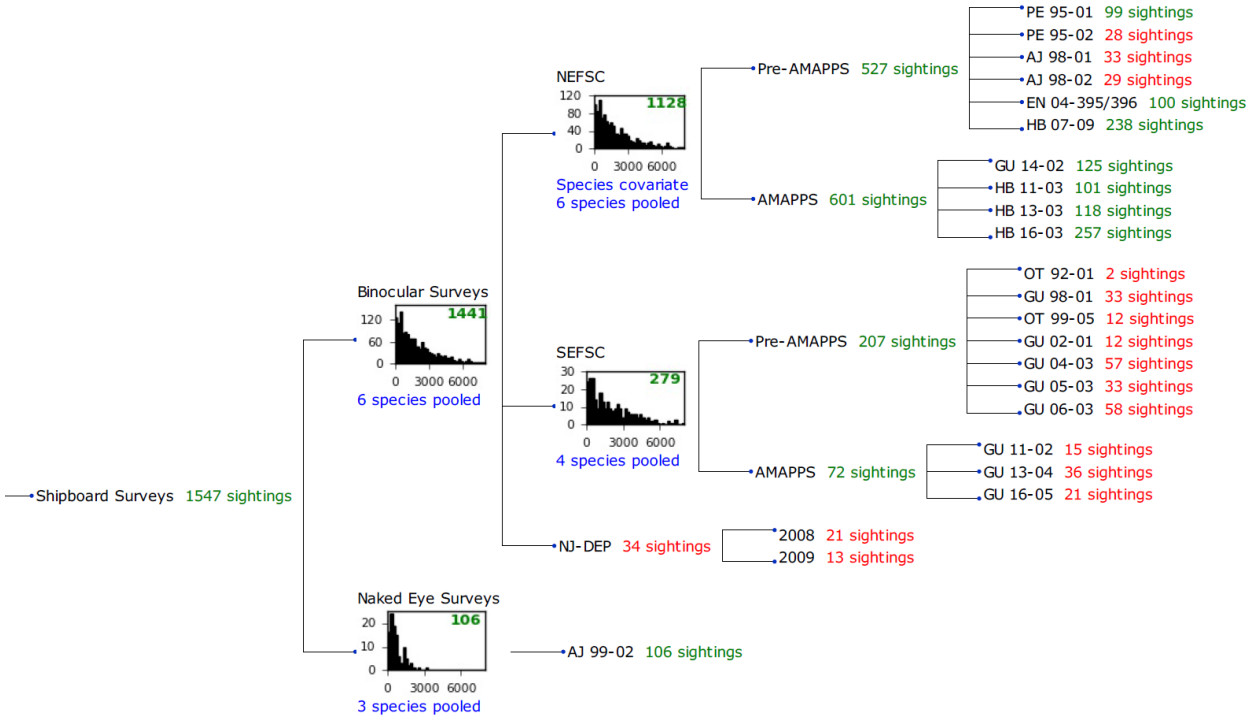


Figure 61. Sightings of large whale groups (all baleen whale species except minke whales, plus sperm whales) reported by shipboard surveys. The histograms show nodes where detection functions were fitted; we present the details of these detection functions below. (As noted above, Naked Eye Surveys were not actually used in this model.)

Although 58 right whales were sighted on surveys that utilized pedestal-mounted binoculars (Figure 60), we judged this a marginal number for fitting detection functions and pooled in other large whale species (all baleen whales except minke whales, plus sperm whales). This yielded over 1000 sightings across the surveys (Figure 61), with a large majority coming from NEFSC, owing to more baleen whales occurring in the northern part of the U.S. Atlantic EEZ, where they feed. We considered fitting detection functions at the survey-program level, i.e. pre-AMAPPS vs. AMAPPS, as we had done with aerial surveys, but we found fewer differences in shipboard protocols in the pre-AMAPPS and AMAPPS eras than we did with aerial surveys, so we fitted detection functions at the organizational level instead (NEFSC and SEFSC) and used a covariate to differentiate between eras. The era covariate was selected, indicating statistically significant differences; for both organizations, detection distances were higher for AMAPPS. We also tried survey vessel as a covariate, but found it did not perform as well as organization, and ultimately it was not selected. Finally, because NJ-DEP did not report enough sightings for its own organization-level detection function, we fitted a detection function to all binocular surveys and applied it just to NJ-DEP.

For all shipboard detection functions, we excluded effort made in Beaufort state 6 or higher. We describe the details of each of the three final shipboard detection functions in sub-sections below.

5.2.1.2.1 NEFSC detection function

After exclusions, the NEFSC binocular surveys reported 1128 sightings of large whale groups (Figure 61), which included all baleen whales except minke whales, plus sperm whales. After right-truncation, which was done at 6500 m and dropped 39 sightings (3%), we were left with 1089 sightings for fitting the detection function. The final detection function (Figure 64) included species identification and survey era as covariates. Detection distances were higher for humpback and sperm whales than for blue, fin, right, or sei whales (Figure 62), consistent with what observers reported anecdotally about these species and with most results for aerial surveys. Detection distances were higher for the AMAPPS surveys than those that came before (Figure 63), perhaps reflecting improved methods, observer training, or equipment during AMAPPS. Although Beaufort state was tested in candidate detection functions, it tended to be dropped from the selection procedure, and exploratory plots (not shown) suggested a nearly flat relationship between detection distance and Beaufort state.

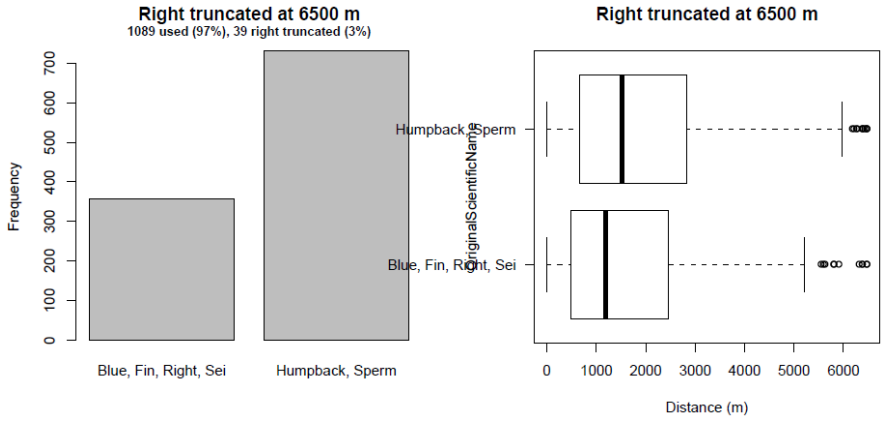


Figure 62. Frequency of species identifications (left) and the distribution of perpendicular sighting distances by species identification (right) for sightings used in the NEFSC binocular surveys detection function.

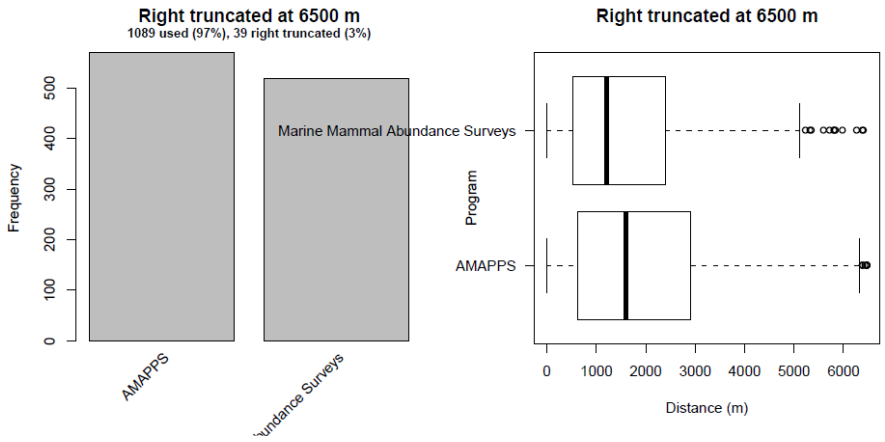


Figure 63. Frequency of sightings by survey era (left) and the distribution of perpendicular sighting distances by survey era (right) for sightings used in the NEFSC binocular surveys detection function. In this detection function, "Marine Mammal Abundance Surveys" is equivalent to "Pre-AMAPPS" in the detection hierarchy diagram.

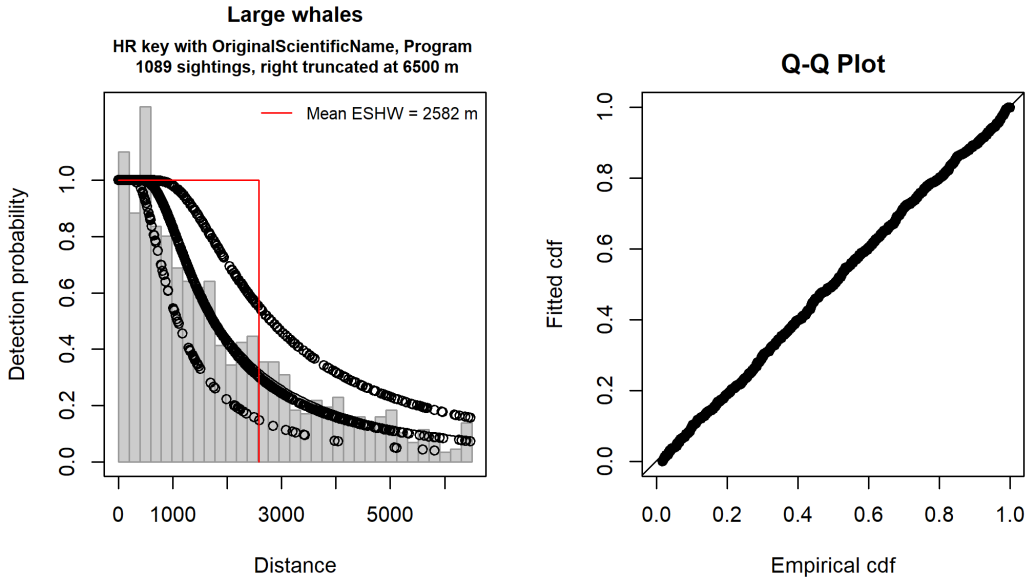


Figure 64. Final detection function applied to the NEFSC binocular surveys.

5.2.1.2.2 SEFSC detection function

After exclusions, the SEFSC binocular surveys reported 279 sightings of large whale groups (Figure 61), which included all baleen whales except minke whales, plus sperm whales. This was substantially lower than NEFSC, even though SEFSC conducted as much or more shipboard effort as NEFSC (Table 1). The reason is simply that the southeast U.S. is not as suitable habitat for baleen whales as the northeast, at least during the summer when most shipboard surveys occurred. After right-truncation, which was done at 6000 m and dropped 13 sightings (5%), we were left with 266 sightings for fitting the detection function. The final detection function (Figure 67) included Beaufort state and survey era as covariates. Detection distances were higher for lower sea states (Figure 65), and higher for the AMAPPS surveys than those that came before (Figure 66), as occurred with NEFSC. We did not test species identification, as over 90% of the sightings were of sperm whales. The detection function may therefore be biased if sperm whales are not a reliable proxy for right whales. This can be tested or addressed in the future if additional non-sperm-whale sightings accumulate on future SEFSC surveys.

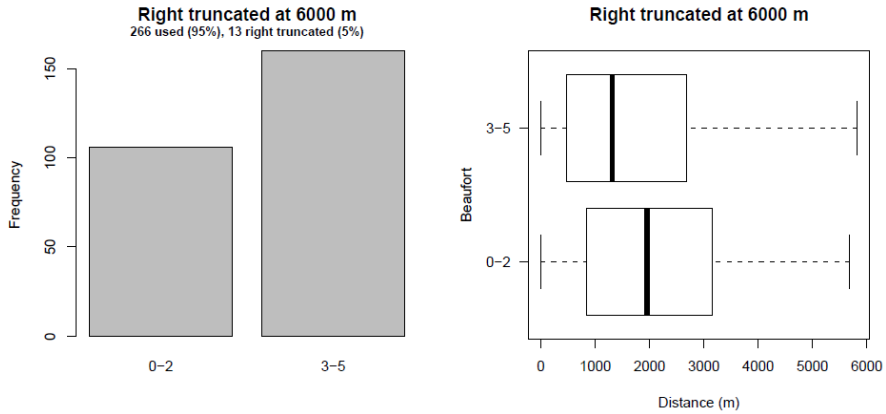


Figure 65. Frequency of categorized Beaufort state (left) and the distribution of perpendicular sighting distances by Beaufort state (right) for sightings used in the SEFSC binocular surveys detection function.

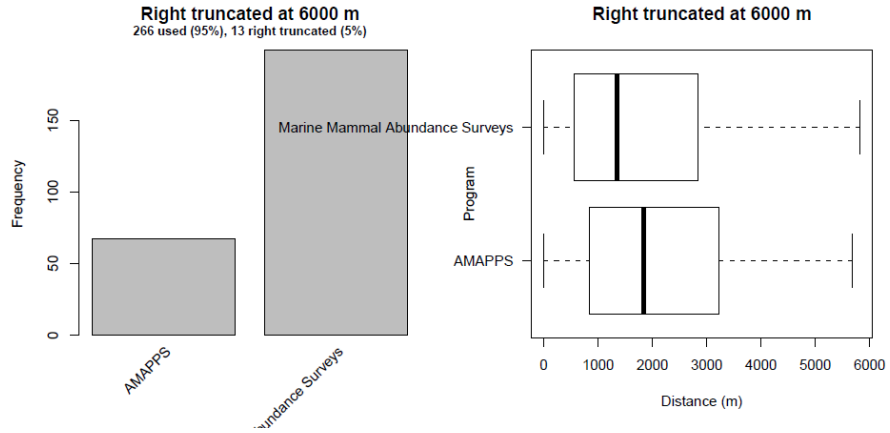


Figure 66. Frequency of sightings by survey era (left) and the distribution of perpendicular sighting distances by survey era (right) for sightings used in the SEFSC binocular surveys detection function. In this detection function, "Marine Mammal Abundance Surveys" is equivalent to "Pre-AMAPPS" in the detection hierarchy diagram.

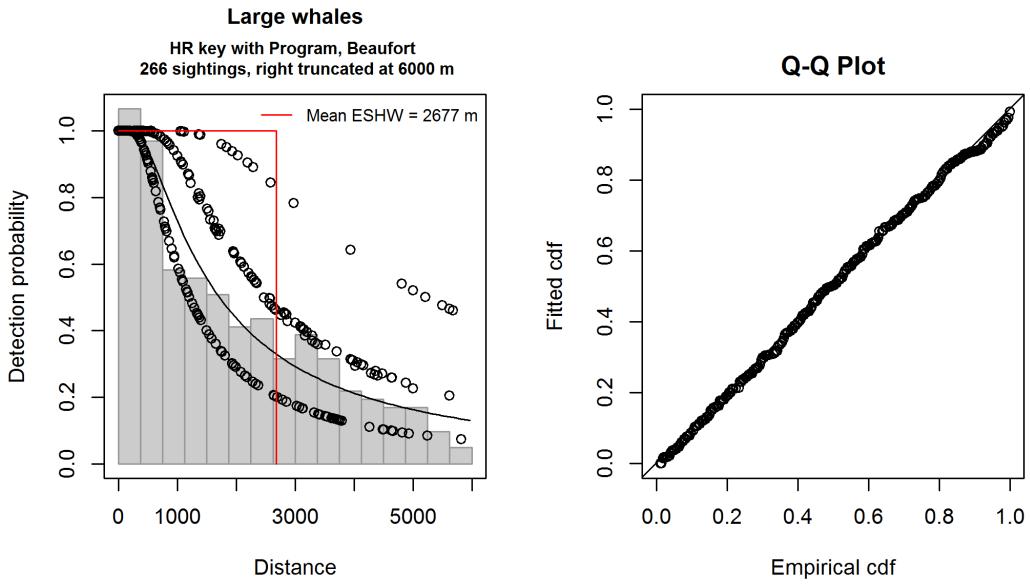


Figure 67. Final detection function applied to the SEFSC binocular surveys.

5.2.1.2.3 Binocular Surveys detection function

To model detectability for the NJ-DEP surveys, we used all binocular surveys as a proxy, being unsure whether NEFSC or SEFSC would be more suitable. After exclusions, the combined binocular surveys reported 1443 sightings of large whale groups (Figure 61), most of which were from NEFSC. As above, this included all baleen whales except minke whales, plus sperm whales. After right-truncation, which was done at 6500 m and dropped 54 sightings (4%), we were left with 1389 sightings for fitting the detection function. The final detection function was a hazard-rate model with a 2nd order polynomial adjustment and no covariates (Figure 68). NJ-DEP did not provide any covariates, so the only one we could try was species identification, which did not score as high in AIC or other statistics.

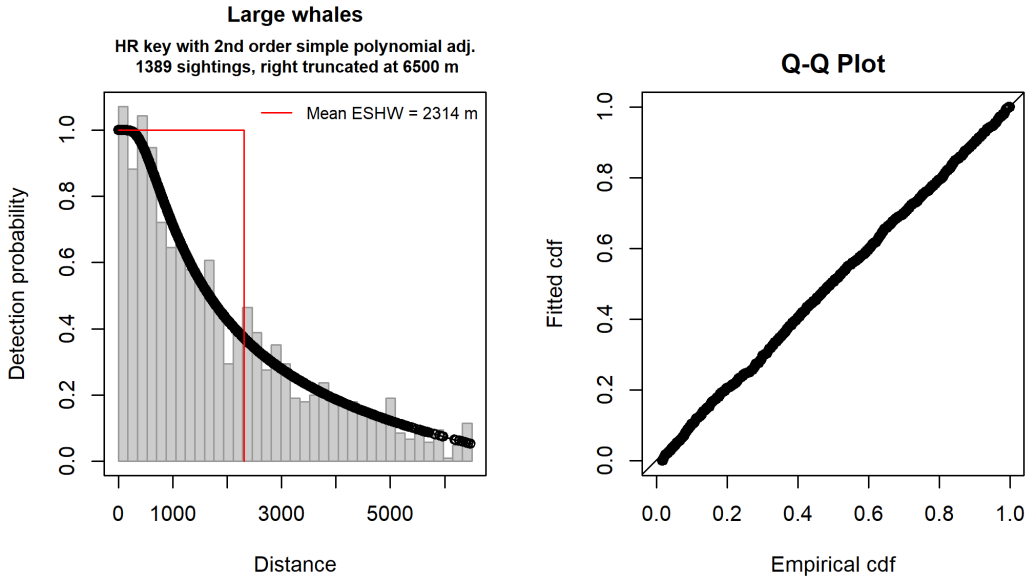


Figure 68. Final detection function for the Binocular Surveys node, applied only to the NJ-DEP surveys.

5.2.2. Perception and availability bias corrections

An important assumption of distance sampling is that objects that occur at distance 0 (directly along the survey trackline) are always detected (Buckland et al. 2001). This is known as the $g(0)=1$ assumption, signifying that the probability of detection, g, at distance 0 is 1. When this assumption is violated and $g(0)<1$, density will be biased low (that is, the number of animals present will be underestimated).

On surveys for marine animals, this bias generally occurs in two ways. *Perception bias* occurs when animals are not detected by the observer team because they are hard to see—e.g. because they are small or otherwise inconspicuous. *Availability bias* occurs because the animals are not available to be seen, even though they are present—e.g. because they are submerged. The two biases are often assumed to be independent, and modelers typically estimate a g(0) for each—i.e. $g(0)_{perception}$ as the probability that an animal at the surface will be sighted if it is at distance 0, and $g(0)_{availability}$ as the probability that an animal along the trackline will be at the surface while it is within visual range—and then multiply the detection function by each one, reducing the integrated probability of detection and effective strip width accordingly. Alternatively, the counts of detected animals can be divided by each one, inflating the counts to account for animals that were "missed", which is what we did for this analysis, to facilitate the use of sighting-specific g(0) estimates.

To ensure the biases are independent of the surveying process, appropriate field methods must be used to ensure animals do not respond to the survey platform before they are sighted. On aerial surveys, this is achieved by flying at sufficiently high speed or altitude, so that the animals do not have a chance to respond before observers are close or are not disturbed by the presence of the aircraft high above. On shipboard surveys, it is accomplished by enabling observers to scan far ahead of the vessel, e.g. by positioning them as high as possible and equipping them with binoculars.

A recommended approach for assessing perception bias is to utilize an independent observer or multiple observer teams and match up the sightings reported by each to characterize how frequently sightings were missed. The most commonly-used variations of this approach are known generally as *mark recapture distance sampling* (MRDS; Burt et al. 2014). Many of the shipboard surveys available for our analysis used one of these variations. Most of the aerial surveys did not, owing to the logistical difficulty of carrying sufficient personnel in a small aircraft or the financial challenge of affording a suitable aircraft and so many observers. The main exceptions were the NOAA AMAPPS aerial surveys, which carried two independent observer teams on de Havilland Twin Otters. Prior to AMAPPS, the NEFSC broad-scale marine mammal abundance surveys used a "circle-back" method to characterize detection probability using a single team by making multiple observations of sighted groups (Palka 2006).

Because of this heterogeneity we used a single-team approach for our analysis. For all surveys we only used effort and sightings from the team that our collaborators designated as the primary team. Then, to correct for perception bias, we consulted our collaborators and searched the literature for studies that had estimated it, selected estimates that best matched according to the platform, species, and surveyor organization or location, and applied the various estimates to the appropriate surveys after fitting detection functions using traditional single-team methodology (not MRDS).

To estimate availability bias for aerial surveys, we surveyed the literature for right whale dive data and used Laake et al.'s (1997) estimator to calculate regional, survey-program-specific corrections. For shipboard surveys, we assumed that availability bias would be negligible (i.e. $g(0)_{availability} = 1$) due to the ships operating at slow speeds during the surveys. Right whales are known to undertake moderately long dives while feeding in deep waters—e.g. Baumgartner et al. (2017) reported a mean dive duration of 13.4 min (SD=2.8) at Jeffreys Ledge, the longest duration of six regions studied—but even at the longest durations they would still surface within visual range of passing survey ships, which all travelled at 10 knots or less. For example, at that speed, the ship would travel about 4.14 km in the 13.4 min of the longest diving whales at Jeffreys Ledge, which was less than the right truncation distance on all shipboard detection functions.

Below, section 5.2.2.1 details the perception bias corrections we used for shipboard surveys, and section 5.2.2.2 the perception and availability bias corrections we used for aerial surveys.

5.2.2.1. Bias corrections for shipboard surveys

As discussed above, we assumed that availability bias was negligible for shipboard surveys for right whales. For perception bias, we selected survey-specific estimates when they were available, and otherwise used the NMFS AMAPPS estimates as the best available proxy (Table 4). No right-whale-specific estimates were available, owing to few right whales being sighted on shipboard surveys, so we used estimates for other large whales. This, and the use of AMAPPS estimates on non-AMAPPS surveys, likely introduced an unknown amount of bias. We suspect the bias resulting from using AMAPPS estimates on non-AMAPPS surveys is likely to be small, as those non-AMAPPS surveys were conducted using similar protocols by many of the same personnel, often on the same vessels. We anticipate that this potential source of bias may eventually be eliminated in future models, when it is practicable to drop the pre-AMAPPS shipboard surveys from the analysis. (We believe it likely that NEFSC and SEFSC will continue to estimate perception bias on all future surveys using the AMAPPS method, or a method evolved from it.)

Table 4. Perception bias estimates used for shipboard surveys.

Surveys	**$g(0)_{perception}$ (CV)**	**Source**
NEFSC EN 04-395/396	0.940 (0.610)	Palka (2006), Table 5: 2004, upper team, estimate for fin or sei whales (no right whale estimate available)
All other NEFSC	0.513 (0.136)	Palka et al. (2017), Appendix I, Table 3.2: NE-shipboard group 1 (NEFSC AMAPPS)
All SEFSC	0.472 (0.228)	Palka et al. (2017), Appendix I, Table 3.2: SE-shipboard group 1 (SEFSC AMAPPS)
NJ-DEP	0.513 (0.136)	NEFSC AMAPPS estimate as proxy

5.2.2.2. Bias corrections for aerial surveys

For aerial surveys, we estimated both perception and availability bias. This was considerably more complicated than for shipboard surveys, owing to the diversity of aerial survey programs.

5.2.2.2.1 Perception bias

For perception bias, the only relevant estimates available in the literature were Palka et al.'s (2017) estimates for the NEFSC and SEFSC AMAPPS surveys conducted from 2010-2013 on de Havilland Twin Otter aircraft using two-team MRDS methodology, and Palka's (2006) estimates for the NEFSC pre-AMAPPS surveys conducted from 1999-2004 on the same aircraft but using a single team with circle-backs to estimate bias. We used Palka et al.'s (2017) estimates for all surveys. We did this even for the NEFSC pre-AMAPPS surveys, rather than using Palka's (2006) estimates for them, for two reasons. First, the two-team MRDS approach is a more modern method and is what NMFS prefers to use now (presumably, they believe it to be better than the circle-back approach, otherwise they would have continued with that). Second, the circle-back approach actually estimates both perception bias and availability bias as a single g(0). Because Palka's (2006) circle-back estimate was for all species of large whales, it was based on aggregate diving behavior of all those species. We preferred to base the availability component of g(0) on right-whale-specific dive behavior, to account for known differences in diving in different regions (see next section 5.2.2.2.2). The following pseudocode describes how we assigned a perception bias correction to each sighting.

```
If group size ≤ 2:
   If it was an NEFSC survey other than NARWSS:
      Use NEFSC AMAPPS estimate for large whales: g(0)perception = 0.503 (CV=0.168)
   Otherwise:
      Use SEFSC AMAPPS estimate for large whales: g(0)perception = 0.898 (CV=0.091)
Otherwise:
   Assume g(0)perception = 1
```

The AMAPPS perception bias estimates were taken from Palka et al. (2017), Appendix I, Table 3.2. The large majority of sightings made during those surveys were of 1 or 2 whales, so we only applied the corrections to sightings with those group sizes. Our assumption of $g(0)_{perception} = 1$ for groups of 3 or more was based on Hain et al's (1999) finding that larger groups of right whales were substantially easier to detect than smaller groups, and that a number of other analyses of other species of large whales surveyed by air estimated or assumed perception bias was negligible for groups of any size (Forney et al. 1995; Carretta et al. 2000; Heide-Jørgensen et al. 2012; Hansen et al. 2019).

Our logic applied the NEFSC estimate to NEFSC's broad-scale marine mammal abundance surveys, and SEFSC's to all other surveys, including NEFSC's NARWSS surveys. We believe that NEFSC's AMAPPS estimate is unusually low, not only because it is inconsistent with the studies referenced above that found aerial perception bias to be negligible for large whales, but also because it is lower even than Palka's (2006) estimate (g(0)=0.53, CV=0.54) that incorporated availability bias in addition to perception bias (using the "circle-back" method). Because of this unusual difference, we believed the SEFSC AMAPPS estimate to be a better proxy for the remaining surveys. Also, the SEFSC estimate (0.898) was relatively high, which was similar to analyses noted above that estimated or assumed g(0)=1.

Finally, we caution that density generally varies directly with the inverse of g(0); for example, if g(0) is halved, density is doubled. This highlights the importance of estimating perception bias as accurately as possible, especially when abundance estimation is the primary goal of an analysis. We strongly encourage all collaborators to adopt methods for estimating perception bias, which would allow us to use program-specific estimates in the future, rather than applying NMFS's estimates to all programs. That said, we believe our treatment here was better than the traditional alternative, which is to assume that g(0)=1 and thereby underestimate density.

5.2.2.2.2 Availability bias

Availability bias for aerial surveys depends on a number of factors. Important considerations include the diving patterns of species of interest, the speed and altitude of the aircraft, and the field of view available to the observer (as determined by the aircraft windows, obstructions, etc.). Because availability bias can vary considerably based on these factors, we attempted to account for as many as feasible in the time available for this project.

To do this, we relied on the approach of Laake et al. (1997), as formulated by Robertson et al. (2015):

$$a(x) = \frac{\bar{s}}{\bar{s}+\bar{d}} + \frac{\bar{d}\left[1 - e^{\{-t(x)/\bar{d}\}}\right]}{\bar{s}+\bar{d}}$$

Here, $a(x)$ is the availability bias correction factor at perpendicular distance x; the value $a(0)$ is the $g(0)_{availability}$ that we sought to estimate. $\bar{s}$ and $\bar{d}$ are average surface and dive intervals for the species, and $t(x)$ is the time that a parcel of water at distance x remains in view of the observer. Under this approach, the first fraction $\left(\frac{\bar{s}}{\bar{s}+\bar{d}}\right)$ represents the probability that the animal is at the surface when it enters the observer's field of view, and the second fraction represents the probability that it is submerged but will surface while within the field of view as the observer passes over. Please see Laake et al. (1997) for further details.

We obtained surface and dive intervals from diving studies in the literature. When feasible, we accounted for difference in diving behavior based on overall geographic location (e.g. calving grounds vs. feeding grounds) and group composition (e.g. singletons vs. mother-calf pairs vs. surface active groups (SAGs)). Because these factors could vary on a per-sighting basis, we developed an availability correction for each sighting individually. We were interested in availability at x=0, so we needed to estimate $t(0)$, the total time-in-view on the trackline at perpendicular distance zero. We derived this parameter from the speed and altitude of the survey. Below, we first discuss the estimation of $t(0)$ and then the application of surface and dive intervals and resulting availability estimates.

Estimation of t(0)

We based our estimates of $t(0)$ on results reported by Robertson et al. (2015), who estimated availability corrections for bowhead whales sighted from de Havilland Twin Otters flown at a target altitude of 305 m (1000 ft) altitude and speed of 220 km/h (119 knots). To do this, they fitted linear models to the times an object was in view forward and aft of perpendicular as a function of perpendicular distance, using field data collected by trained marine

mammal observers. To obtain this training data, they flew trial passes by a static structure placed at randomly determined perpendicular distances. They asked observers to execute their "normal" search pattern and recorded the times the object first came into view, passed perpendicular, and passed out of view behind the plane. Under these conditions, their procedure estimated that at the mean speed (62.31 m/s) and altitude (305 m) of the analysis, the total time-in-view on the trackline *t(0)* was 37.78 s, based on a search sector spanning 37.4° forward to 121.2° aft, also estimated by the model.

To adapt this result to the surveys used in our analysis, which occurred at different speeds and altitudes, we used the expression:

$$t(0) = t_R(0)\frac{v_R}{v}\frac{h}{h_R}$$

where $t(0)$ was the survey-specific value we sought to estimate, $t_R(0)$ was Robertson's time-in-view estimate at distance 0 (37.78 s), v_R and h_R were the speed (62.31 m/s, or 121.1 kts) and altitude (305 m, or 1001 ft), respectively, of Robertson's experiment, and v and h were the speed and altitude, respectively, of our survey. This approach assumed that $t(0)$ would scale linearly with altitude (e.g. doubling altitude would double time-in-view) and inverse-linearly with speed (e.g. doubling speed would halve time-in-view). Table 5 shows our results.

Table 5: $t(0)$ estimates for surveys used in our analysis.

Altitude (m)	Speed (kts)	$t(0)$ (s)	Surveys
183	95	28.90	VAMSC and Riverhead Maryland DNR
183	110	24.96	NEFSC pre-AMAPPS; NEFSC and SEFSC AMAPPS
229	100	34.36	NARWSS 2003-2016
229	110	31.23	SEFSC pre-AMAPPS
229	120	28.63	NJ-DEP
305	87	52.60	VAMSC VA CZM and Navy VACAPES
305	100	45.76	NARWSS 2017-2019, NEAq, NYS-DEC, all SEUS NARW, UNCW
305	110	41.60	HDR

We caution that our approach, although based on empirical results, assumed that Robertson's results, which involved a specific organization, aircraft type, and survey protocol, were a reasonable proxy to use for our surveys when scaled using the method we described. Although no empirical analyses similar to Robertson's were available in the literature for our surveys, we found two relevant applications of the Laake et al. (1997) approach, to which our results may be compared. Palka (pers. comm.) developed similar availability corrections for the AMAPPS surveys, which used similar aircraft to Robertson, and possibly some of the same personnel, but flew at a lower altitude and slower speed. Palka's analysis assumed a hemispherical, forward-facing field of view and based $t(0)$ estimates on the maximum forward distance animals were sighted for the 2010-2013 AMAPPS surveys (sometimes the maximum perpendicular distance was used instead). For fin, sei, and right whales, which were analyzed as a guild for the purpose of deriving $t(0)$, Palka reported a maximum distance of 1490 m. Under the assumed field of view, this is the along-track distance that a parcel of water directly below the plane would be sufficiently visible to an observer to allow them to sight a large whale. To obtain this result, Palka assumed an altitude of 183 m and 105 kts[1]. For comparison, if we rescale Robertson's results using the method above, we obtain $t(0) = 26.15$ s and a distance of 1412 m, which differs by less than 7% of Palka's result.

[1] Although Palka (pers. comm.) used a speed of 105 kts for the AMAPPS surveys, in our analysis we used 110 kts, the speed consistently reported in the AMAPPS annual reports. This discrepancy is currently unresolved but can be clarified for future models, with Palka's assistance.

The second relevant application of Laake et al. (1997) we identified was from McLellan et al. (2018), who estimated availability corrections for beaked whales sighted on UNCW surveys that occurred near Cape Hatteras, NC from May 2011-November 2015. These surveys were conducted at 305 m and 100 kts from Cessna Skymasters. In a similar analysis to Palka's, McLellan et al. estimated an along-track visible distance of 973.4 m. For comparison, rescaling Robertson's results using our method, we obtain $t(0) = 45.76$ s and a distance of 2354 m, approximately 2.4 times McLellan's result. This substantial disparity may relate to the marked difference in detectability between beaked whales and baleen whales. For example, in our most recent analysis of beaked whales (Roberts et al. 2017), we estimated an ESHW of 478 m for beaked whales sighted on surveys that used UNCW's protocol, 98% of which came from UNCW's Navy surveys. But in this report (section 5.2.1.1.10 above) we estimated an ESHW of 1090 m for baleen whales sighted on those surveys, approximately 2.3 times that of beaked whales.

Surface and dive intervals, and estimation of availability corrections

To determine the surface and dive intervals—the $\bar{s}$ and $\bar{d}$ parameters of the availability bias correction factor equation shown above—we utilized estimates from the literature that considered geographic location, group size, and group composition. The following pseudocode describes our approach.

```
If south of Cape Hatteras, use intervals from Hain et al. (1999):
   If group size = 1:
      Use mean intervals for single juvenile (s=2.14, d=3.17 min)
   Otherwise, if group size = 2 and a calf was present (or presence was unknown):
      Use mean intervals for mother/calf pair (s=5.44, d=2.25 min)
   Otherwise:
      Use mean intervals for Surface Active Group (SAG) (s=7.76, d=1.45 min)

Otherwise, if between Cape Hatteras and Canada, use intervals from CETAP (1982):
   If group size = 1:
      Use mean intervals for Lone (s=3.20, d=7.42 min)
   Otherwise, if group size = 2 and a calf was present:
      Use mean intervals for Calf (s=1.94, d=2.50 min)
   Otherwise:
      Use mean intervals for Group (s=1.81, d=3.69 min)

Otherwise, use intervals from Baumgartner (2017) and Baumgartner and Mate (2003):
   Use mean dive interval from Bay of Fundy and Roseway Basin (d=10.7 min)
   If a calf was present:
      Use surface interval for calves and females with calves (s=5.39 min)
   Otherwise:
      Use surface interval for all others excluding pregnant female (s=3.13 min)
```

In the literature we found sufficiently detailed studies of right whale diving behavior in the calving grounds near Florida (Hain et al. 1999) and in important feeding grounds in the Gulf of Maine (CETAP 1982; Nieukirk 1992; Winn et al. 1995; Baumgartner & Mate 2003; Baumgartner et al. 2017) and Cape Cod Bay (Ganley et al. 2019). Cape Hatteras is the location where the Gulf Stream deviates away from the continental shelf out into the wider Atlantic, and thus forms an ecoregional border between the cold, nutrient-rich waters to the north that flow from Canada, and the warm, less-rich waters to the south that are strongly influenced by the Gulf of Mexico. We surmised that right whales south of Cape Hatteras would exhibit diving behavior more similar to that observed on the calving grounds than on the feeding grounds. Accordingly, we obtained estimates for this area from Hain et al. (1999). We note that these authors provided data for a single juvenile but not a single adult. Therefore, we are using the juvenile estimate for all singletons. (There were two sightings of singleton calves; these were included in this category.) Also, their SAG estimate is for groups of size 2 or more, and they did not provide estimates groups not classified as SAGs. Thus, we used the SAG estimate for all pairs where a calf was known not to be present and for all groups size 3 or greater.

North of Cape Hatteras, Baumgartner and Mate (2003) reported relatively long dive times for deep basins in Canada compared to those reported for shallower basins in the southern Gulf of Maine, e.g. by CETAP (1982). Considering this, we split the sightings approximately at the U.S.-Canada border and applied separate dive data to the

two regions. We defined the split between them as the line that passes through 42.5 N, 66.0 W in the Northeast Channel and 44.0 N, 67.5 W in the northern Gulf of Maine, so that all of Georges Bank would be aggregated into the southern region.

For the southern region, labelled in the pseudocode above as "between Cape Hatteras and Canada", we used data from CETAP (1982), Special Topic D, Table 5. These authors reported mean durations of surface activity bouts and inter-bout dives for right whales observed at Stellwagen Bank, east of the elbow of Cape Cod, and the Great South Channel in May of 1980 and 1981. These are all areas where right whales regularly feed. We used CETAP's estimates for Lone whales, Calves, and Groups to derive availability corrections.

CETAP reported that cows "closely attended" their calves and exhibited similar but slightly longer dive intervals than the calves. From this, we believed it likely that the cows synchronized their diving with the calves, and therefore used the calf intervals to represent the pair. CETAP's group estimate was for groups of 2 or 3 whales; groups larger than 3 were not tracked, owing to the difficulty in reliably identifying more than three individuals in a continuous tracking scheme. We assumed that whales in larger groups would behave similarly to groups of 2 or 3. Finally, CETAP's surface and dive intervals were for individuals, not groups. Following McLellan et al. (2018) and Palka (in prep.), we assumed that whales in groups dived asynchronously, which increased the probability that at least one would be on the surface at any given moment. So for groups of two or more, we estimated availability of the group itself using the following expression (McLellan et al. 2018):

$$a(0)_{group} = 1 - (1 - a(0)_{individual})^{GroupSize}$$

For this region we also considered data from Winn et al. (1995), who reported shorter mean inter-bout dive durations than CETAP of 4.76 min and 1.54 min for whales monitored with radio tags in the Great South Channel in 1989 and 1990, respectively. However, they did not provide surface activity bout durations, so we could not estimate availability corrections from these data. Similarly, Baumgartner et al. (2017) reported a mean dive duration of 2.7 minutes for 22 whales tagged in Great South Channel but no surface intervals. Given the variability in diving behavior reported across these studies, we urge additional study of right whale diving to comprehensively characterize how best to estimate availability bias. As an extreme example of high variability, Ganley et al. (2019) conducted 87 focal follows on 22 days in 2016 and 2017 in Cape Cod Bay during the months of January-April and reported that mean monthly percent surface time ranged from 16% in January to 55% in April.

In the northernmost region, Baumgartner and Mate (2003) reported mean surface and dive intervals of 3.13 min and 12.17 min for feeding right whales monitored with time-depth recorders in the lower Bay of Fundy and Roseway Basin in 2000-2001. The authors also collected *Calanus finmarchicus* stage 5 copepodites, believed to be the preferred prey of right whales in this region, and reported right whales undertaking deep dives to the bottom mixed layer where copepods were aggregated. Nieukirk (1992) reported a similar estimate of 21.65% time spent at the surface for 8 right whales tagged with satellite tags in the Bay of Fundy in 1989-1990. Baumgartner et al. (2017) updated the 2003 paper to provide mean dive intervals of all dives, not just feeding dives, but did not update the surface intervals. Accordingly, we used dive intervals from the 2017 paper and surface intervals from the 2003 paper. As with the CETAP data, Baumgartner's data were for individuals, so for groups of two or more, we estimated availability of the group itself using the method described above.

Across all regions, the resulting $g(0)_{availability}$ correction factors ranged from a high of greater than 0.999 for large feeding aggregations observed mainly in the Gulf of Maine and southern New England, to a low of 0.256 for single whales observed in the northern region where deep diving has regularly been observed. Across all 4162 sighted groups that were retained for the spatial model (rather than being left- or right-truncated), the mean $g(0)_{availability}$ correction factor was 0.737.

For this analysis, we did not attempt to estimate uncertainty for availability bias correction factors. This is something that should eventually be done, however it might not be possible without access to the original diving data and a complex collaborative analysis, and for such results to be useful, we also require a method to propagate these uncertainty estimates to the spatial modeling stage, when they are estimated on an individual sighting basis. This is an ongoing research topic for the "DenMod" initiative funded by the Navy Living Marine Resources program. When suitable methodology becomes available from that initiative, we can consider estimating and propagating uncertainty in availability bias correction factors.

5.3. Spatial modeling

In this section, we present the methodological details of fitting and predicting the spatial model, the detailed results, the final predicted density surfaces, and a comparison of those surfaces to passive acoustic monitoring data collected independently.

5.3.1. Methods

In this section, we cover the overall approach we took in creating this new right whale model, covering most of the methodological details and decisions (some are left to region-specific sections that follow this one).

5.3.1.1. Addressing the recent substantial changes to right whale abundance and distribution

It is now widely recognized that the right whale population entered a sustained decline around the start of the past decade. A comprehensive state-space photo-ID capture-recapture analysis estimated a 99.99% probability that the population's trajectory post-2010 was a decline (Pace et al. 2017). Several other studies reported large changes in distribution patterns around the same time (Davis et al. 2017; Ganley et al. 2019; Record et al. 2019; Davies et al. 2019). In April 2019, at NOAA's request we presented the version 8 (v8) right whale model (Roberts et al. 2017) to the Atlantic Large Whale Take Reduction Team (ALWTRT) to solicit feedback on its use for mitigating risk that right whales would become entangled in vertical fishing lines in U.S. trap and pot fisheries. ALWTRT members expressed strong concerns that the v8 model, which spanned 1998-2016, did not sufficiently reflect present-day abundance levels and distribution patterns. They cited, as examples, evidence that right whales were now occurring in large numbers south and west of Nantucket Shoals, from winter to early summer, and in lower numbers in the Gulf of Maine. These patterns had not been seen in the 1998-2009 period. The ALWTRT members called for a model update that included data through at least the end of 2018, and that was tuned to give predictions more resembling recent patterns than historical patterns. Around this time, NOAA requested that we develop our next model to have separate "before" and "after" predictions that summarized density patterns before and after the point in time that major changes occurred.

In response to this request, we first considered how these changes might best be modeled. Right whales present a particular challenge, in that they have undergone large changes in abundance and distribution, but we currently lack essential components needed to comprehensively model this. We suspect these changes were initially driven by changes to prey distribution or abundance, and then compounded by cumulative health effects (e.g. fewer females obtaining sufficient food resources to reproduce frequently) and new anthropogenic sources of injury and mortality as whales relocated to new feeding areas (e.g. the Gulf of St. Lawrence, where many recent entanglements have occurred). Fully accounting for these effects would require comprehensive prey distribution data (e.g. a time series of zooplankton density surfaces spanning the spatiotemporal extent of our model) and mechanisms for incorporating cumulative health effects and anthropogenic impacts into our model. These are major research tasks that are beyond the scope of our Navy modeling work. While we may eventually be able to progress on those problems via other initiatives, for the current update to the right whale model we required a simpler approach that was tractable in the time available. Accordingly, we utilized a methodology similar to that of the v8 model but adapted it represent changes in a "before" and "after" approach in line with what NOAA requested.

First, we surveyed the literature and relevant unpublished materials to determine when to define the "before" and "after" periods. In Pace et al.'s (2017) analysis, the peak of right whale abundance occurred in 2010, with 2011 slightly lower. At the 2019 North Atlantic Right Whale Consortium (NARWC) annual meeting, R. Pace and S. Kraus presented an update to this model that extended it from 2015 to 2018; in the update, 2011 was slightly higher than 2010. In Davis et al.'s (2017) summarization of right whale acoustic detections spanning 2004-2014, to demonstrate changes to spatiotemporal patterns in acoustic activity, the authors split their data into the periods 2004-2010 and 2011-2014. In Davies et al.'s (2019) analysis of right whale sightings in the Bay of Fundy, the authors noted in 2010 a marked increase in the percentage of survey days each year in which no right whales were sighted, and a substantial decrease in annual integrated encounter rate. Record et al. (2019) compared right whale sightings per unit effort to prey abundance in three areas around the Gulf of Maine and summarized results into two contrasted periods, 2004-2008 and 2012-2016. In Ganley et al.'s (2019) analysis of monthly right whale abundance in Cape Cod Bay, peak

seasonal abundance started regularly exceeding 100 whales in April 2008, and large numbers of whales (>50) started regularly appearing at the beginning of the season in January 2011. Large springtime feeding aggregations of right whales were first reported off southern New England in April 2010, near Rhode Island Sound (Figure 69; https://nefsc.noaa.gov/press_release/2011/SciSpot/SS1102/), and have been reported south of Martha's Vineyard, Nantucket Island, or Nantucket Shoals every year since. Conversely, 2011 was the last year that whales were sighted in substantial numbers in fall in the central Gulf of Maine by systematic surveys (Figure 70). Finally, for the calving grounds, by applying photo ID capture-recapture methods to photographed sightings from 2014-2015, Gowan et al. (2019) documented higher calving rates in 2001-2011, younger age at first reproduction in 2005-2011, and a large decrease from 2010 to 2011 in probability of photographic capture of non-breeding females at the winter calving grounds. Subsequently, a comprehensive analysis of photo ID data from the 1997/98-2018/19 calving seasons indicated that the total number of unique individuals sighted south of the Virginia / North Carolina border between December-March peaked in the 2009/10 calving season, (T. Gowan, pers. comm.; Figure 71). The rate of unique individuals per survey flown also peaked in 2009/10 (Figure 71).

Collectively, these reports show that the various changes in distribution did not happen precisely in synchrony with each other or the start of the decline, but instead over several years. For our analysis, we selected spring 2010 as the temporal split-point between the "before" and "after" periods, on the basis of 2010 being the peak or near peak in Pace's models, of 2009/10 being the calving season with the largest number of individuals visiting the southeast, and of 2010 being a year that major changes in feeding behavior occurred in both the Bay of Fundy and southern New England. Spring 2010 was also roughly in the middle of the other changes we discussed above.

A second major change to the model, made in response to calls to focus it away from historical years and more on the present, was to shift it to start later in time. The v8 model started in 1998; we started the v9 model in 2003. This was the first year that that there was fairly comprehensive surveying of both the southern calving grounds in winter and the northern feeding grounds in spring and summer from programs that were sufficiently systematic to use under our methodology. (On advice from T. Gowan, we elected not to use southeast right whale Early Warning System (EWS) surveys prior to the 2003/04 calving season, as survey protocols and data collected prior to that season were not standardized across the EWS teams, making those prior seasons difficult to use.) Hereafter in this report, we refer to the "before" period as the *2003-2009 era*, comprising the period of spring 2003 through the end of the winter 2009/10 calving season, and the "after" period as the *2010-2018 era*, comprising the spring of 2010 through the end of the winter 2018/19 calving season. (Thus the 2003-2009 era extended into 2010, and the 2010-2018 era into 2019.)

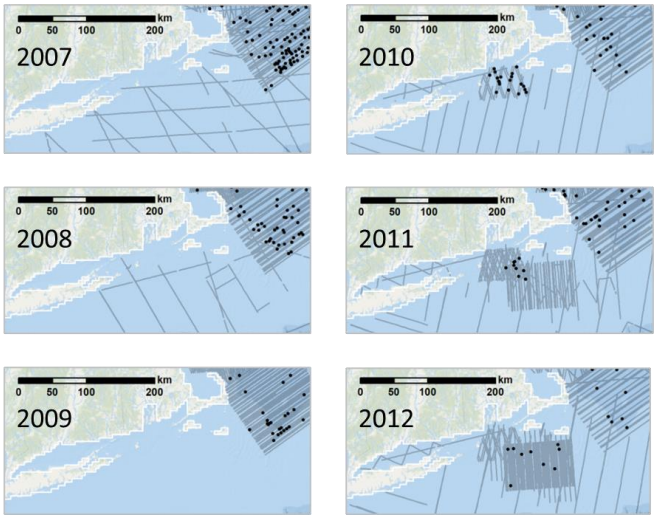


Figure 69. Survey effort and sightings of right whales in the southern New England area for 2007-2012, showing the first appearance of large numbers of sightings in spring 2010. Although systematic survey effort was relatively sparse in this area prior to 2010, extensive shipping occurs in this area and we believe it is likely that if large aggregations occurred in prior years they probably would have been noticed and surveyed.

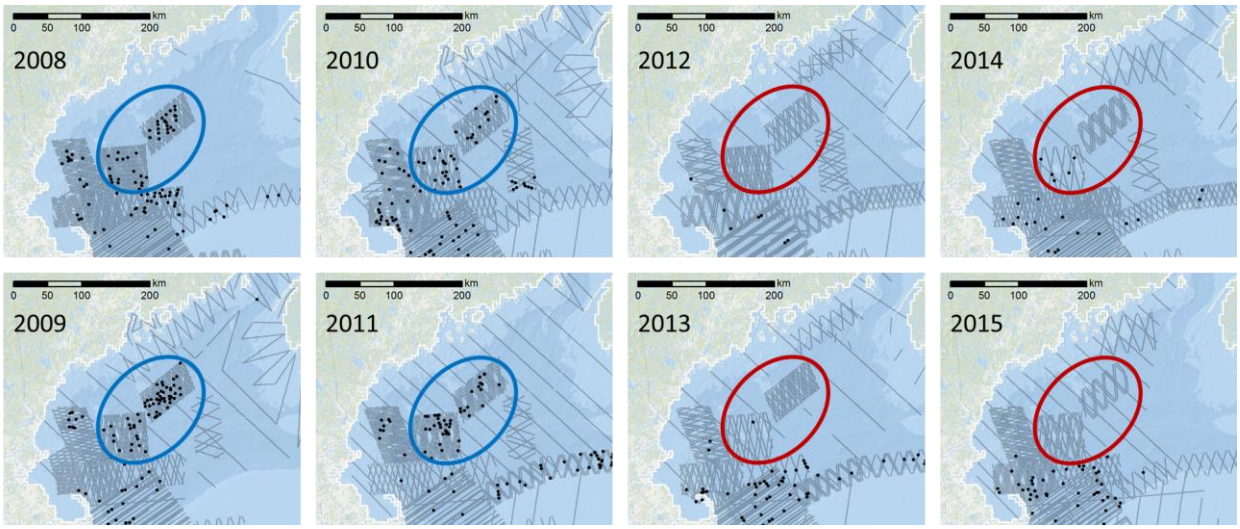


Figure 70. Survey effort and sightings of right whales in central Gulf of Maine for 2008-2015, showing the regular occurrence right whales in fall through 2011 (blue circles) and absence in years thereafter (red circles). Each map shows effort and sightings from October of the year indicated through September of the following year.

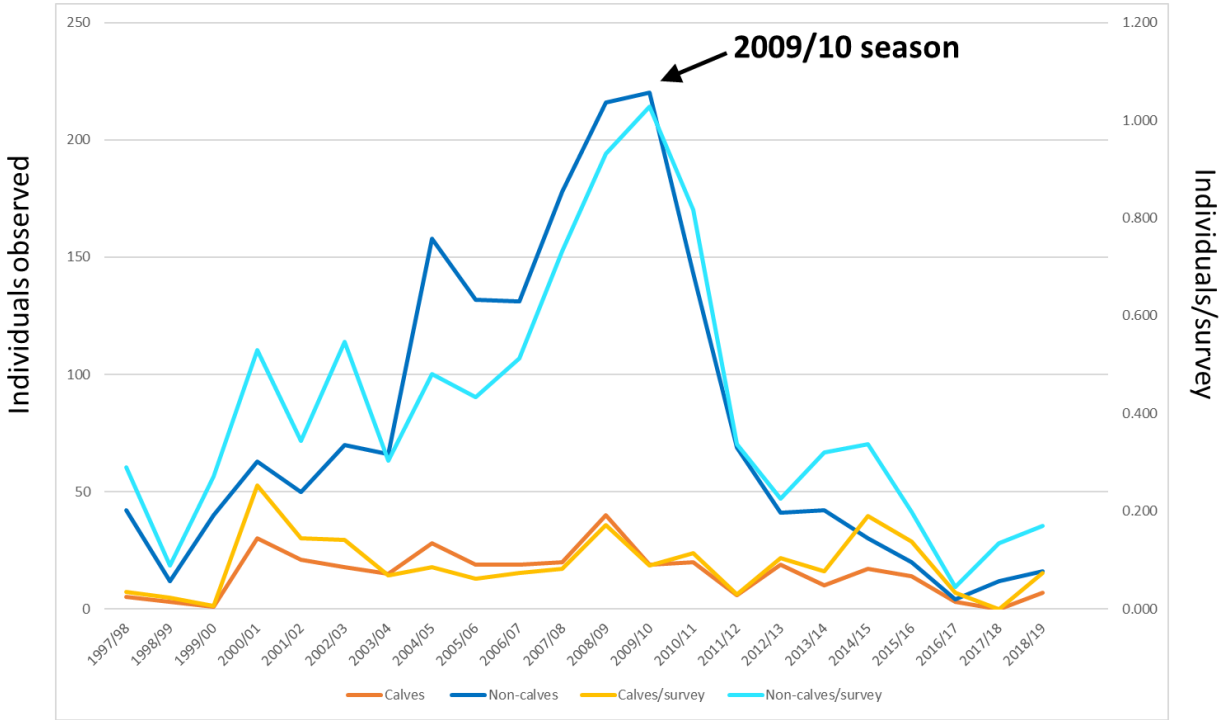


Figure 71. Count of unique individual right whales (red and dark blue) and rate of unique individuals per survey (yellow and light blue), sighted on surveys south of the Virginia / North Carolina border during December-March. Data courtesy of T. Gowan, compiled from the NARWC photo ID database (circa November 2016) and right whale Southeast Implementation Team presentations.

5.3.1.2. Study area, regions, and seasons

To meet the needs of the Navy and other model users, we needed to offer predictions at a monthly timestep, where sufficient data made this possible, and fall back to coarser temporal resolutions where data were sparse. An important

problem is that right whales in different seasons or locations are likely to exhibit different relationships to environmental conditions. For example, a pregnant female that has migrated to the calving grounds to give birth is likely to associate with a different range of water temperatures than a juvenile that is overwintering in the Gulf of Maine or southern New England. Similarly, whales might associate with waters having different levels of productivity during winter than during summer. To account for this, we split the study area spatially into regions (Figure 72), and each region temporally into seasons, and modeled each region's season independently, allowing different species-environment relationships to be modeled for each one. This was a similar procedure to our prior models (Roberts et al. 2016a, 2017, 2018), except that here we split first by region, then season, to allow seasons to vary between regions.

We split the study area first at Hatteras Island. In our previous model, we placed this split at Cape Fear, based on it being designated by NOAA as the northernmost limit of the calving Critical Habitat area. However, as mentioned above (in section 5.2.2.2.2), Cape Hatteras is at the border between two major ecoregions. We proceeded under the ecological assumption that whales that travel south of Cape Hatteras (e.g. to reach the traditional calving grounds) are likely to correlate differently with the environment than those north of Cape Hatteras, even if calving does not generally occur north of Cape Fear. We first tested models that placed the split directly at Cape Hatteras but this resulted in extreme edge effects (not shown) between the models on either side of the split, so we moved the split north to Rodanthe, North Carolina, the easternmost point of Hatteras Island, to mitigate this. We excluded the extreme southern part of the study area from our analysis. Although we will eventually provide predictions for this area for the Phase IV models, the v9 model omits it. Right whales occasionally move into this area, even into the Gulf of Mexico (Moore & Clark 1963), but this is relatively rare, although it happened again very recently, in March, 2020. We would like to further consult with our collaborators before deciding how to address this area in Phase IV.

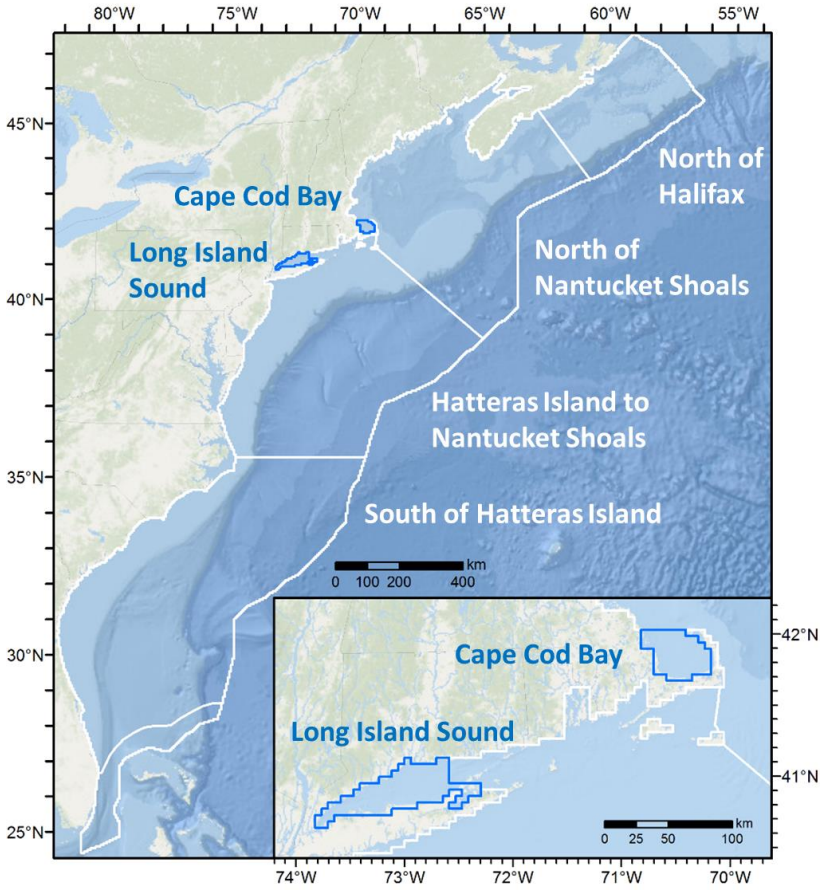


Figure 72. Extents of regional models in the overall v9 right whale model. Models were fitted for three regions: South of Hatteras Island, Hatteras Island to Nantucket Shoals, North of Nantucket Shoals. North of Halifax and the extreme southern part of the study area (unlabeled) were omitted from the analysis (see text). Densities for Cape Cod Bay were obtained based on Ganley et al.'s (2019) results. Right whales were assumed to be absent from Long Island Sound.

North of Cape Hatteras, we split the study area again at Nantucket Shoals, a very shallow area that separates the Great South Channel, an area where right whales reliably aggregated to feed in spring in the 2003-2009 era, from the so-called "South of the Islands" area south of Nantucket Island and Martha's Vineyard. Right whales are now regularly aggregating in this area in winter and spring. We initially attempted to keep these two areas together in a single model, but this resulted in extreme unrealistic predictions in one area or the other, depending on which covariates were retained by the model. We found that more realistic models were obtained by grouping the South of the Islands area with the mid-Atlantic, and thus placed the split at Nantucket Shoals. However, we ultimately concluded that the South of the Islands area is ecologically distinct from the mid-Atlantic area to the south, and investigating the best means for modeling these areas is a top priority for us for future updates to this model (see section 6.3.1).

Within the region between Cape Hatteras and Nantucket Shoals, we split off Long Island Sound, delineating the inner shallower region using bathymetric maps, and excluded this region from our model. The only systematic surveys of Long Island Sound that we are aware of are from AMAPPS; the only cetacean reported in the area we excluded was harbor porpoise, during winter. Opportunistic sightings of humpback whales and belugas have occasionally been reported since 2015, and a fin whale was reported in 1993. Based on the heavy boat traffic and lack of opportunistic sightings, we assumed that right whale density was zero within the area we excluded.

North of Nantucket Shoals, we split the study area a final time on the Scotian Shelf near Halifax. Survey effort was sparse northeast of Halifax in spring and summer (Figure 73) and did not reach beyond this point in fall and winter (Figure 74). We are omitting the area northeast of Halifax from the v9 model but ultimately will include it in the final model used for Phase IV, after we have had more time to study how better to model it. We are hopeful that collaborations can be established with Canadian survey organizations that would allow us to introduce additional data for Canadian waters, which would boost our ability to model this area.

Finally, as we have done previously, we split Cape Cod Bay (CCB) off from the region north of Nantucket Shoals and used external abundance estimates from our collaborators to derive densities for it. For the years 2003-2017, we derived densities from the abundance estimates published by Ganley et al. (2019). As we did with the v8 model, we divided their monthly abundance estimates by the size of their study area (as specified to us by L. Ganley) to obtain monthly density estimates, then assigned this value uniformly to the cells of our grid that we designated as CCB. (In total, these cells spanned 1600 km^2.) For the years 2018-2019, abundance estimates were not yet available from these authors. Instead, on their advice, we obtained monthly right whale sighting counts from the Center for Coastal Studies (B. McKenna, pers. comm.) and rescaled them to abundance using the correlative relationship published in Ganley et al. (2019) that was intended for this purpose. Finally, because Ganley et al. only provided estimates for January-May, we set, for all years, the December density to that of January, the June density to that of May, and predicted the other months from the wider density models for the region north of Nantucket Shoals.

5.3.1.3. Accounting for inter-era and interannual effects

To simplify model implementation and to facilitate comparison between the two eras (2003-2009 and 2010-2018), NOAA suggested that we fit fully independent models to the eras and a third model to the entire 2003-2018 period. We examined this possibility and found the idea of fitting two independent era models to be problematic, owing to the spatiotemporal differences in effort between the eras (Figure 73, Figure 74).

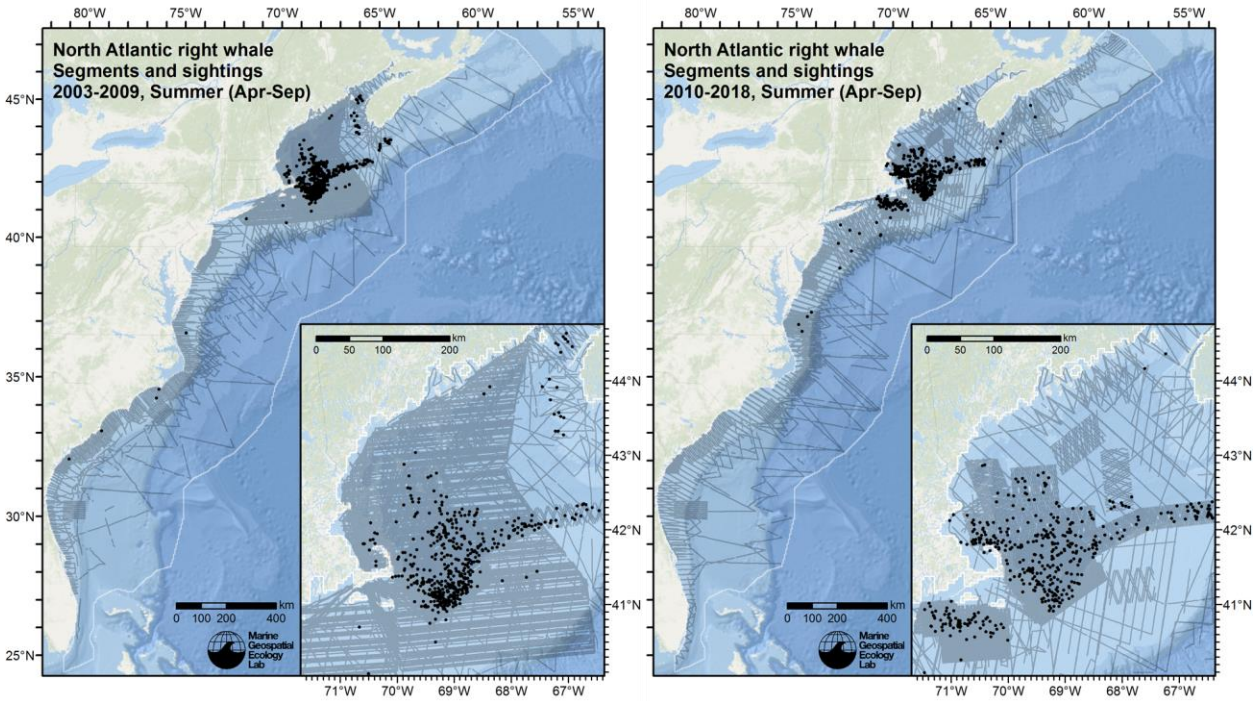


Figure 73. Comparison of survey effort and sightings available for this analysis for the two analyzed eras (2003-2009, 2010-2018) for the months of April-September. Note that the recent era has less broad-scale coverage in the Gulf of Maine (right inset) than the older era (left inset), owing to NARWSS ending broad-scale surveys in 2007.

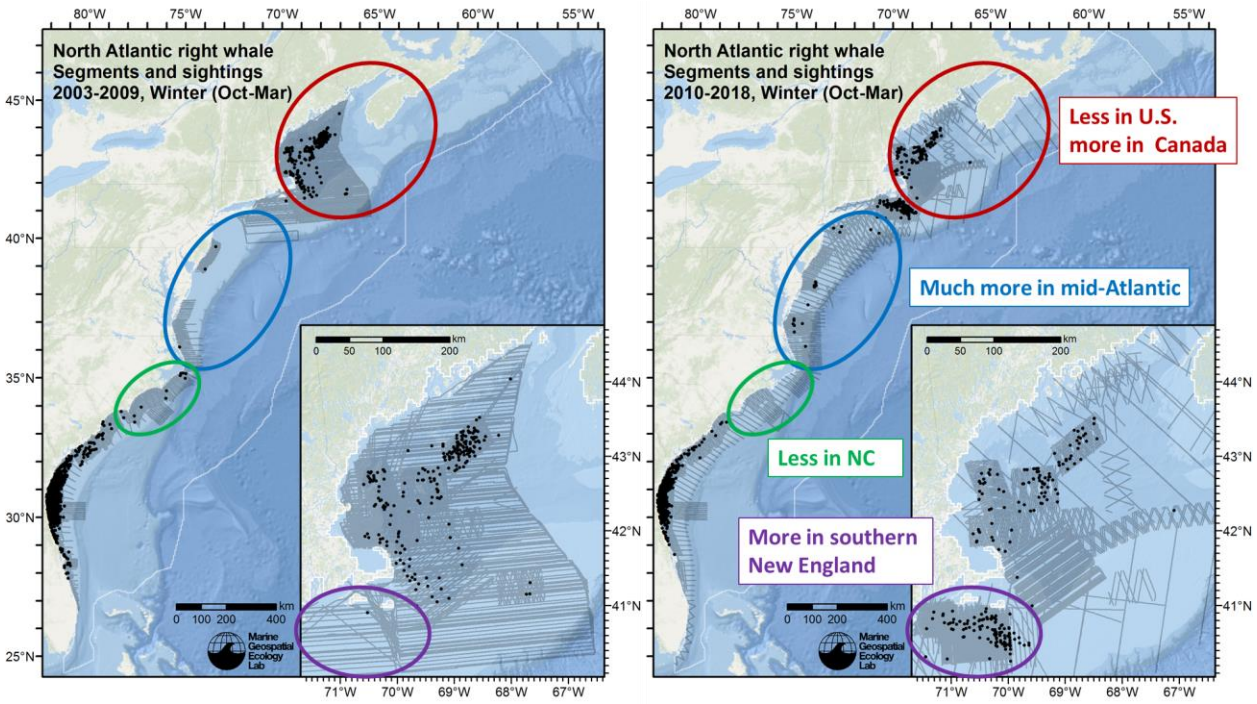


Figure 74. Comparison of survey effort and sightings available for this analysis for the two analyzed eras (2003-2009, 2010-2018) for the months of October-March. As for April-September (Figure 73), the recent era has less broad-scale coverage in the Gulf of Maine (right inset) than the older era (left inset). Colored circles and text highlight other important differences in effort or sightings.

For example, in April-September, the Gulf of Maine region would be easier to model in the older era (Figure 73), especially in the Bay of Fundy and elsewhere in Canadian waters, where many more right whale sightings were available than in the recent era. By contrast, in October-March, the recent era would be much easier to model in the mid-Atlantic (Figure 74) than the older era, which received very little survey coverage in this region.

In our judgment, differences such as these precluded the possibility of fitting fully independent models for the two eras; too many regions would be left with insufficient effort to be modeled with anything other than stratified design-based line transect density estimates that would yield uniform densities over certain patches, and empty areas elsewhere. Instead we proceeded from the assumption that right whales may not have changed habitat preferences between the two eras, but rather, the habitat itself changed (e.g. via climate change) and the whales moved to follow the best habitat. Under this view, a model containing data from both eras should be able to reproduce changes in distributions if it is given effective covariates. However, we lacked what we believed to be one of the most important covariates—prey density—so we utilized several alternative approaches to account for inter-era changes while keeping all of the data in the same model.

First, in regions that had consistent replication of effort every year of the complete study period (2003-2018), we included the year as a categorical covariate, following the approach of Gowan and Ortega-Ortiz (2014). The only region that qualified for this approach was the region south of Hatteras Island, where the right whale EWS teams consistently covered the core calving habitat in Florida and Georgia every year from December-March with multiple flights per month. Coverage in the northern part of this area, in North Carolina, was not as consistent across the time series, with less effort occurring in the 2010-2018 era (Figure 74), but there were still several years of heavy coverage. There was virtually no coverage beyond the 200 m isobath (e.g. over the Blake Plateau) during the core winter months, but this was not a major concern: this very warm area is not believed to be right whale habitat, and the on-shelf coverage extended far enough into this area to establish a strong decreasing gradient in right whale density going into it. By including year as a categorical covariate, we allowed the model to account for interannual variations that were not driven by contemporaneous oceanographic conditions, but instead by more complex, temporally-lagged effects that we cannot currently represent, such as the cumulative influence of foraging success on reproductive success. We present the results for this model in section 5.3.2.1 below.

Second, for regions that did not have consistent replication of effort every year, but still had some replication when summarized by era, we included the era itself as a two-level categorical covariate. The idea here was that if area-wide abundance was generally higher or lower in the second era as compared to the first, then the era covariate could account for it, leaving the oceanographic covariates to account for spatial distribution and higher-frequency temporal changes (e.g. seasonal movements). All regions qualified for this approach, so we tested it in all of them (except south of Hatteras Island, where we used year as a categorical covariate). We present the results for these models in sections 5.3.2.2 and 5.3.2.3 below.

As a final means to help account for inter-era and interannual variability, we based nearly all models on contemporaneous dynamic covariates. We discuss this in the next section.

5.3.1.4. Covariates

There are two types of dynamic (i.e. time-varying) covariates frequently used in species distribution models. The first, contemporaneous covariates, have covariate values resolved to a particular year. For example, if a model was based on survey data that spanned two years and used remotely-sensed sea surface temperature (SST) as a contemporaneous covariate, survey data from year 1 would be matched to SST images from year 1, and survey data from year 2 would be matched to SST images from year 2. Species distribution models fitted with contemporaneous covariates can thus potentially account for both intra-annual (e.g. seasonal) and interannual changes.

The second type, climatological covariates, are long term averages of contemporaneous covariates. Climatological covariates divide the calendar year into shorter time slices such as days, weeks, months or seasons, and for each slice, apply a summary statistic such as the mean to many years of observations made during that slice to estimate the long-term state. Climatologies smooth out interannual variability while maintaining intra-annual variability and preserving spatial gradients. In the example model above, if a climatological representation of SST, e.g. the monthly climatological average, was used as a covariate, all years of the survey would be matched to the same

climatological SST images, e.g. by month. For more about contemporaneous and climatological covariates, please see Mannocci et al. (2017a).

In our previous right whale models, we compared models fitted with daily-resolution contemporaneous covariates to models fitted with 8-day climatological covariates (i.e., a series of 46 images, each representing an 8-day period of the year, averaged over many years). For most regions and seasons, climatological models performed slightly better according to goodness-of-fit statistics, and we ended up basing our final models mostly on climatological covariates. This allowed the models to reproduce intra-annual variations but not interannual changes. While this result was statistically supported under the methodology we defined for the Phase III modeling process, it was somewhat unsatisfactory for right whales, which had been shown to exhibit interannual variability in spatial distribution that was believed to be driven by interannually-varying oceanographic conditions (Gowan & Ortega-Ortiz 2014), and were known at that time to have undergone two decades of population growth (Pettis & Hamilton 2014), even though the subsequent decline had not been fully recognized yet.

After completing the v8 model in 2019, we contemplated this outcome and concluded that it might result from using contemporaneous covariates with too fine a temporal resolution. For the Phase III models, we had selected daily resolution data to try to capture as many ephemeral effects as possible, but it is possible they were simply too noisy. Although right whales can relocate quickly, they are also flexible animals that are adapted to a variety of conditions, so they may just ignore day-to-day fluctuations and instead respond to slower, more sustained signals. If so, the rapid fluctuations contained in the daily covariates would manifest as statistical error in density predictions, which might have put the contemporaneous covariates at a disadvantage under our model selection process.

To address this in the v9 model, we tested contemporaneous covariates at both daily and monthly resolution, in addition to climatological covariates at monthly resolution. (We switched from 8-day to monthly resolution for practical reasons: some covariates were available at monthly but not 8-day. We did not have time to attempt a comparison of 8-day to monthly, to determine which was more optimal. The question of what level of smoothing would be best remains an open question for future investigation.) After this change, in version 9 the monthly contemporaneous covariates performed best in the goodness-of-fit statistics we use, except in the region north of Nantucket Shoals in the spring and summer models. Following sections present the details of each model.

Table 6 lists the covariates we prepared for the v9 model. To be usable for our model, covariates needed to be: 1) plausibly correlated with cetacean habitat (ideally, reported as such in the literature); 2) available as gridded surfaces spanning the study area; 3) available for the entire modeled period, spring 2003 to spring 2019; 4) available at 25 km spatial resolution or higher; and 5) available at least at monthly contemporaneous resolution, and ideally at daily resolution to facilitate the daily vs. monthly comparison. We favored products in which the covariate producer solved any gap-filling problems, as is the case with so-called Level 4 remote sensing products as well as ocean model outputs. Although we have previously employed several techniques to fill gaps caused by clouds and other phenomena, our preference is for oceanographers and similar experts to handle this job, using expertise they have specifically developed for the products in question.

After downloading each product, we resampled it to the model's grid, filled any remaining gaps with Laplacian interpolation, and prepared monthly climatologies (when they were not already available from the covariate producer). Using the resampled grids, we extracted covariate values as follows. First, we matched the segments to the resampled grids using the appropriate ranges of dates. (For example, to extract the monthly contemporaneous chlorophyll covariate, we matched all segments from January 2010 to the monthly chlorophyll grid for January 2010, all February 2010 segments to the February 2010 grid, and so on). Next, for each grid to which some segments had been matched, we overlaid the segments' centroids on the grid using GIS and extracted the value of the cell under each centroid using linear interpolation. We proceeded through all of the segments and covariates in this way until each segment had a value for each covariate at both the two contemporaneous resolutions and the climatological resolution.

Table 6. Candidate covariates for the spatial models. All covariates were resampled or averaged to the spatial and temporal resolutions of the analysis, and its map projection (see section 5.3.1.5). Each model only considered the covariates appropriate for the modeled region. In later sections of this report, physical oceanographic and biological covariates prefaced by "Clim" refer to their climatological formulations. All distances were calculated using a fast-marching algorithm that accounted for shortest paths around islands and peninsulas.

Type	Covariates	Resolution	Time range	Description
Spatiotemporal	x, y	5 km		Easting (m) and northing (m) representing geographic location in the projected coordinate system of the analysis.
	Year			Year (integer), as a categorical covariate.
	Era			Era, as a categorical covariate: 2003-2009 or 2010-2018.
Static	Depth, Slope	30 arc sec		Seafloor depth (m) and slope, derived from SRTM30-PLUS global bathymetry (Becker et al. 2009), which was also used to derive the remaining static covariates below.
	DistToShore, DistTo125m, DistTo300m	30 arc sec		Distance to the closest shoreline (m), excluding Bermuda and Sable Island, and various ecologically relevant isobaths.
	DistToCCB	30 arc sec		Distance to Cape Cod Bay (m). Used only in models north of Nantucket Shoals (see 5.3.2.3).
	Fetch_50km	30 arc sec		Wind fetch (km), estimated as mean straight-line distance to land from the center of each grid cell in each of 12 directions (spaced evenly at 30º increments), with a maximum allowed distance of 50 km. Used to distinguish inshore areas (e.g. Nantucket Sound) from more oceanic areas, in certain models.
Physical oceanographic	SST_MUR	1 km, daily	2003-2019	Sea surface temperature (SST) (°C) from the ver. 4 Multiscale Ultrahigh Resolution (MUR) L4 analysis (Chin et al. 2017).
	SST_HYCOM, BotT_HYCOM	0.08°, 3-hour	2003-2019	Sea surface and bottom temperature (°C) from the HYCOM GOFS 3.1 GLBv0.08 3-hourly ocean model (Bleck 2002). SST_HYCOM and SST_MUR were both tested but never used in the same model.
	SSS_HYCOM, BotS_HYCOM	0.08°, 3-hour	2003-2019	Sea surface and bottom salinity (PSU) from HYCOM.
	UV_HYCOM	0.08°, 3-hour	2003-2019	Current speed ($m\ s^{-1}$) from HYCOM.
	WindSpeed	0.25°, 6-hour	2003-2019	Wind speed ($m\ s^{-1}$) from the L3.0 Cross-Calibrated Multi-Platform (CCMP) Version 2.0 surface vector winds (Atlas et al. 2011).
Biological	Chl	4 km, daily	2003-2019	Chlorophyll *a* mass concentration ($mg\ m^{-3}$) from the CMEMS L4 Copernicus-GlobColour processor (CMEMS product code OCEANCOLOUR_GLO_CHL_L4_REP_OBSERVATIONS_009_082).

5.3.1.5. Spatial resolution and map projection

At NOAA's request, we extended the spatial extent of the model farther inshore than we did for the v7 and v8 models. This necessitated two important changes. First, to better resolve the complex topography of the New England coastline, we increased the spatial resolution from 10 km to 5 km. Second, to increase the data available close to shore and within inshore areas of New England, we reprocessed the entire database of NARWSS surveys, 2003-2019, to include all on-watch transiting effort conducted in these areas (Figure 75). We also reprocessed NEFSC broad-scale surveys to recover on-effort sections conducted in inshore areas (e.g. Penobscot Bay) that we had excluded in previous modeling cycles.

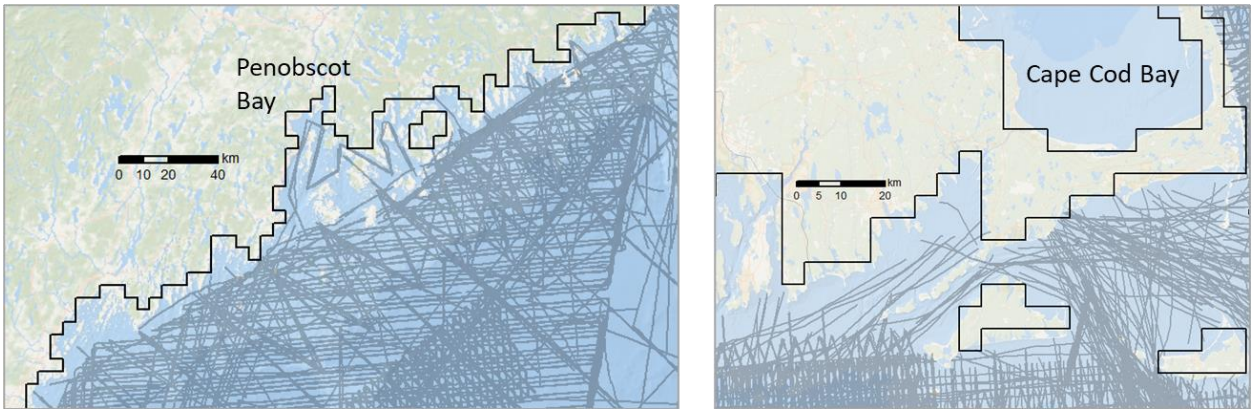


Figure 75. Close-up views of inshore areas of Maine (left) and southern Massachusetts (right) showing penetration of survey segments into inshore areas.

We used the same Albers equal area map projection and grid cell registration used by Roberts et al. (2016a) and all subsequent modeling efforts. 4 cells of the new 5 km grid fit precisely within 1 cell of the old 10 km grid.

5.3.1.6. Model fitting, selection, and summarization

We omitted data collected when the Beaufort wind force scale was greater than 5, except in the southeast, where we omitted data when Beaufort was greater than 3, in keeping with Gowan and Ortega-Ortiz's (2014) analysis of this region. To avoid model optimization issues that can sometimes occur when segments are very short, we omitted survey segments less than 1 km in length (2473 such segments and 11 sightings were omitted). After these omissions, 591,206 segments and 4162 sightings (comprising 12,581 individuals) were retained for the analysis, with a mean segment length of 4.993 km (SD=0.351 km). We present maps of these data in section 5.3.2 below.

In general, we used the model fitting, selection, prediction, and summarization approach of Roberts et al. (2016a). Regional, seasonal models were fitted as generalized additive models (GAMs) with the `mgcv` package 1.89-28 in R 3.6.1. The response variable was the abundance observed on the survey segment, corrected for perception and availability biases. We used thin plate regression splines with shrinkage smoothers [`s(...,bs="ts")`] for all univariate smoothed terms and for bivariate smooths of spatial location. We fitted other bivariate smooths, e.g. interactions between oceanographic covariates, used in select models, with tensor product smoothers [`te(...,bs="ts")`]. We used restricted maximum likelihood (REML) as the smoothing parameter estimation method. We used the log of effective area surveyed (as estimated from detection functions) as an offset. We used the Tweedie distribution for all models [`family=tw()`], to account for overdispersion.

For each regional, seasonal model we usually fitted multiple candidate models that compared contemporaneous and climatological covariates. If sufficient data were available, we usually included a full suite of covariates (Table 6) and relied on shrinkage smoothers to remove insignificant terms from each model. When a candidate model had insignificant covariates ($p > 0.05$) or covariates with estimated degrees of freedom less than 0.85, we removed all such

covariates and refitted the model, repeating this process until no such covariates were left. We avoided including conceptually similar or highly correlated covariates in the same model and fitted multiple candidates for the appropriate combinations instead. In general, the main case of this involved sea surface temperature, in which the MUR SST and HYCOM SST covariate were almost perfectly correlated ($r > 0.98$). When insufficient data were available such that fitting a model using a full suite of covariates proved problematic (e.g. it exhibited obvious overfitting problems), we resorted to a more manual approach to model selection. The exact procedure depended on the region modeled; see section 5.3.2 for details.

We checked models using summary statistics, Q-Q plots, plots of randomized quantile residuals and observed vs. predicted abundances, and examined predictions for obvious unrealistic effects (e.g. excessive predictions and unrealistic artifacts or patterns). When multiple candidates were available that did not show such problems, we usually selected the candidate that exhibited the lowest REML score.

We predicted models across the entire temporal extent of the modeled period, April 2003 to March 2019. We summarized the predictions into 12 monthly maps giving the mean densities at each pixel averaged over multiple years. We performed this summarization for both eras, 2003-2009 (which was April 2003 through March 2010) and 2010-2018 (which was April 2010 through March 2019), as well as the full modeled period, 2003-2018 (which was April 2003 through March 2019). Final density surface maps are given in section 5.3.3.

5.3.1.7. Model uncertainty

To estimate uncertainty, we used the method of Roberts et al. (2016a), quantifying uncertainty as coefficients of variation (CV), or standard errors (SE), that represented the uncertainty in the GAM parameter estimates. These estimates account for the quantity of survey effort available, the number of segments with sightings, and the strength of the correlations between the model covariates and abundances observed on the survey segments. But, owing to limitations with the statistical methodology of density surface modeling, we were not able to include several important known sources of uncertainty in model predictions, including: 1) uncertainty in detection function parameters, 2) uncertainty in perception and availability bias correction factors, and 3) natural variability in dynamic covariates. Thus, the uncertainty metrics that we present here underestimate true uncertainty.

Until such time that we can incorporate some of these additional sources into our final uncertainty estimates, we urge caution interpreting our results (although we note they are compatible with all previous models we have produced). Recently, the "DenMod" Working Group and its collaborators have developed methods for propagating uncertainty from sources 1 and 3 into the final predictions, and we plan on applying these methods immediately in Option Year 4 of this Cooperative Agreement. Accordingly, we do not present uncertainty maps in this report, but instead will include them in the Option Year 4 report, after that work is complete.

5.3.2. Results

In this section, we present the results for the three modeled regions (Figure 72), proceeding from south to north.

5.3.2.1. South of Hatteras Island

South of Cape Hatteras, North Carolina, survey effort was heaviest in the months of December-March in the vicinity of Florida, Georgia, and South Carolina within approximately 100 km of shore, the focal area and months surveyed by the southeast right whale Early Warning System programs. Navy OPAREAs of Jacksonville (FL), Onslow Bay (NC), and Cape Hatteras itself were also surveyed monthly for many years by UNCW as part of the Navy's marine species monitoring program. Finally, starting in 2010, NOAA Fisheries conducted aerial surveys of waters out to 200 m depth once or twice a year as part of the AMAPPS program, covering different seasons in different years. Offshore shipboard surveys occurred less frequently and always in summer (at least during 2003-2019). Prior to 2010, similar NOAA surveys occurred on a less frequent basis.

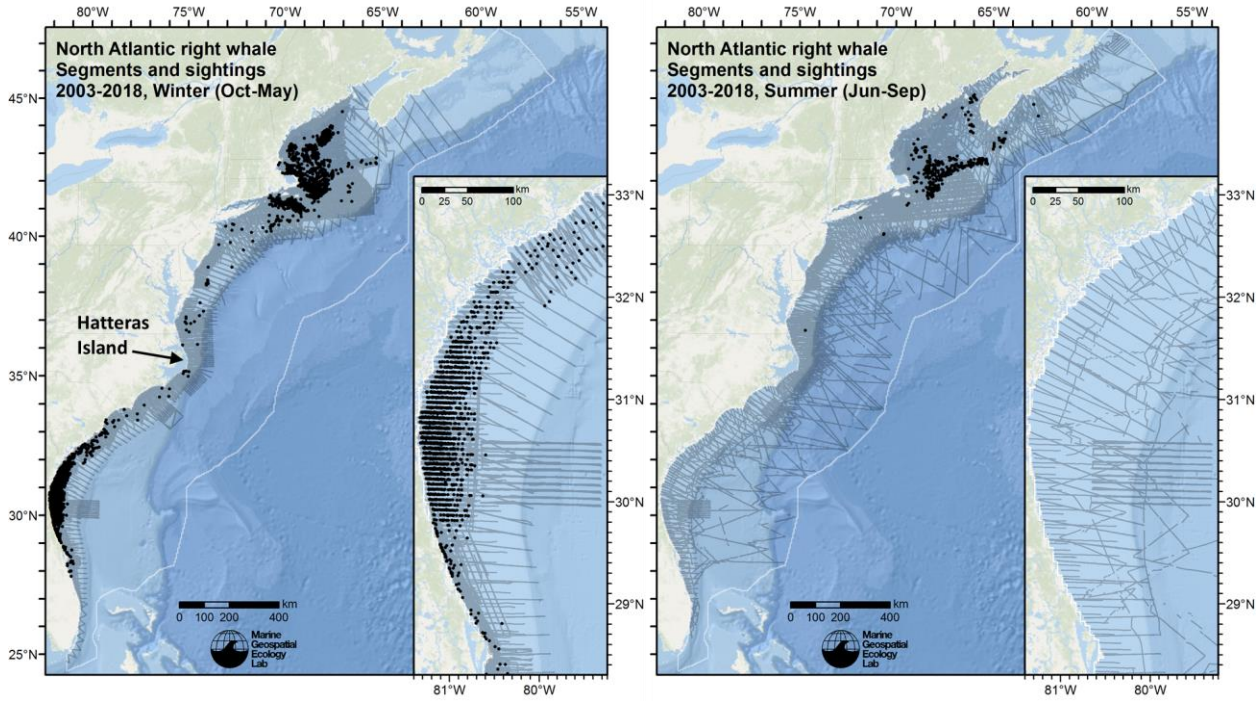


Figure 76. Survey segments and right whale sightings during the "Winter" (October-May) and "Summer" (June-September) seasons defined for modeling south of Hatteras Island. Figure 77 shows the segments and sightings used to fit the model, which spanned both eras.

Right whales were sighted the months of November-May, with most in December-March, when most calving activity was known to occur. No sightings were reported during June-October, and a comprehensive synthesis of passive acoustic monitoring prepared for 2004-2018 (G. Davis, pers. comm.) detected no right whales during these months save a brief detection in Onslow Bay, NC. (However, Hodge et al. (2015) reported detections during each of these months in 2012 in either North Carolina or Georgia or both.) Based on the lack of summertime sightings, we split the year into two seasons for this region (Figure 76): "Winter" (October-May) and "Summer" (June-September). For Winter, we fitted a spatial model, discussed below; for Summer, we assumed the species was absent and manually set density to zero. We included October in the Winter period because surveying was relatively light during this month, so as a precaution we grouped it with the other months in which right whale presence might be predicted.

We caution that there is a very small but non-zero chance that a right whale will be present in this region in June-September (as evidenced, e.g., by the acoustic detections reported by Hodge et al. (2015)). Our modeling process is not optimized to estimate a representative density for these very rare events. Doing so, e.g. by including the entire year in a single model, and thereby allowing some very-low, non-zero density to be predicted for months of high summer, can reduce the precision of estimates made for other times of the year, when right whale density is greater. For management problems that require very-low, non-zero estimates for these very rare situations, e.g. to evaluate the chance that a right whale might be harmed by some activity occurring in Florida in August, we recommend that an alternative approach to our density model be used to estimate the probability of that very rare event.

For the Winter model (Figure 77), we allowed the GAM to select from a full suite of covariates. The core calving habitat in Florida and Georgia was intensely surveyed every year during the core months of December-March, so we included the calving year as a categorical covariate (2003/04, 2004/05, …, 2018/19) to account for interannual variability that did not correlate well with contemporaneous dynamic covariates (e.g. variations driven by foraging success in the northeast earlier in the year). We also included a bivariate smooth of spatial location (x,y) to capture spatial variability not represented by the other covariates. This practice is generally not recommended when a model will be predicted across unsurveyed areas, such as for October-May for large regions of the Blake Plateau, lower continental slope, and abyssal plain. However, these areas are not believed to be good habitat for right whales, and the

resulting model predicted very low density when extrapolated over these unsurveyed areas, meeting this expectation, so we retained the spatial location covariate to improve the model's performance in areas that were intensely sampled.

The final model (Figure 78) retained nine covariates (Figure 79), including the bivariate smooth of spatial location and the categorical calving year covariate. The models fitted to contemporaneous covariates at monthly resolution outperformed the models that used contemporaneous daily covariates or climatological covariates. The model fitted with the MUR SST slightly outperformed the model that used HYCOM SST. We urge caution interpreting the functional relationships (Figure 79) of complex additive models such as this, however we noted that many of the relationships were consistent with what is believed to be known about right whale calving habitat. For example, the relationship with SST was unimodal with a peak between 14-15 ºC, consistent with that reported by Gowan and Ortega-Ortiz (2014). Relationships indicated higher density at low wind and current speeds, consistent with Good's (2008) view that right whales seek calm waters for calving. Relationships with distance to shore and chlorophyll concentration indicated higher density close to shore, with a decrease within about 10 km of shore. The bivariate smooth of spatial location appeared to act contrary to this, with density increasing away from shore, particularly in the north. We suspect the model was using this predictor to capture some of the detail around the Cape Hatteras area. Finally, the variations in the categorical Year covariate somewhat resembled annual counts of unique non-calves present in the study area (compare Figure 79, lower-left, with Figure 71, dark blue line). This suggests the covariate was used as anticipated, to capture interannual variability not represented by contemporaneous oceanographic covariates.

A Q-Q plot of deviance residuals (Figure 78) for the model showed large residuals beyond about the 6th quantile. We found deviance residuals could be reduced to some degree (but not eliminated) by switching from a Tweedie to a negative binomial distribution (results not shown), which might be better adapted to handling the overdispersion of this case, at least when implemented by the R `mgcv` package used to fit the models. However, when we predicted the negative binomial models, the predictions appeared to be under-smoothed (i.e. too patchy or fragmented), suggesting that the models were overfitted. Colleagues at the NOAA Southwest Fisheries Science Center described a similar experience, in which negative binomial models of large whales yielded more extreme predictions of abundance than Tweedie models. Given our limited time, and that the Tweedie distribution has been widely used in cetacean density surface models (Roberts et al. 2016a; Mannocci et al. 2017a; Virgili et al. 2019; Becker et al. 2019; Chavez-Rosales et al. 2019), we deferred future investigation of the negative binomial distribution to future updates.

Mean density predictions (Figure 80) showed a zone of peak density roughly 10-30 km offshore centered approximately on the Florida-Georgia border at the core of the calving grounds, roughly consistent with Gowan and Ortega-Ortiz's (2014) model. A second, less-dense zone was predicted along North Carolina between Cape Lookout and Hatteras Island, consistent with the view that this area forms a migratory "bottleneck" where the continental shelf narrows substantially and the Gulf Stream flows very close to shore, constricting the area through which migrating right whales are presumed to pass. (However, we urge caution with this interpretation until it can be further investigated with additional data). Mean abundance during the 2010-2018 era (corresponding to the 2010/11-2018/19 calving seasons) was estimated at only 31% of the abundance of the 2003-2009 era (corresponding to 2003/04-2009/10), reflecting the decreases in migration to the calving grounds and in calving rates that occurred in the latter era (Gowan et al. 2019).

Across all years, peak abundance south of Hatteras Island was usually predicted to occur in February (Figure 81), but occasionally in January or March. Consistent with passive acoustic monitoring (Davis et al. 2017), moderate to high density was predicted for December-March, low density for November and April, and virtually zero density predicted for October and May. Highest abundance was predicted during the years of 2007/08-2010/11, consistent with the overall view that the population peaked around 2010 and then simultaneously entered decline and shifted distribution patterns (see 5.3.1.1).

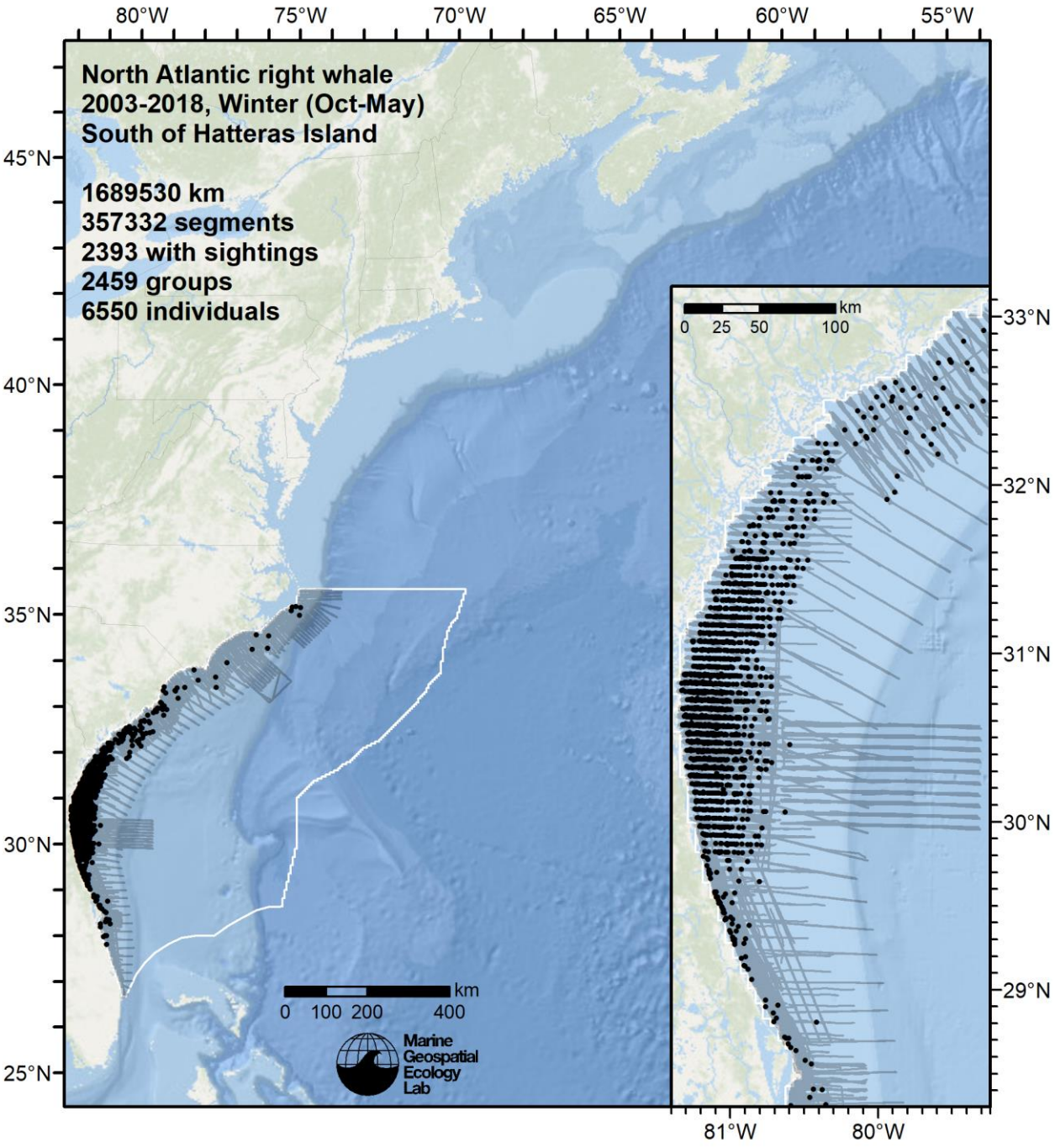


Figure 77. Segments and sightings used to fit the Winter model for the South of Hatteras Island region.

```
Family: Tweedie(p=1.085)
Link function: log
Formula:
IndividualsCorrected ~ offset(log(SegmentArea)) + s(x, y, bs = "ts", k = 50) + s(pmax(I(DistToShore/1000), 2.5), bs = "ts") +
    s(BotT_HYCOM_Monthly, bs = "ts") + s(BotS_HYCOM_Monthly, bs = "ts") + s(UV_HYCOM_Monthly, bs = "ts") +
    s(log10(Chl_Monthly), bs = "ts") + s(WindSpeed_Monthly, bs = "ts") + Year + s(SST_MUR_Monthly, bs = "ts")

Parametric coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -22.4247     0.2870 -78.145  < 2e-16 ***
Year2004/05   0.8243     0.1284   6.419 1.38e-10 ***
Year2005/06   0.2952     0.1395   2.116  0.03434 *
Year2006/07   1.0661     0.1344   7.932 2.16e-15 ***
Year2007/08   1.7029     0.1241  13.722  < 2e-16 ***
Year2008/09   1.0045     0.1289   7.792 6.63e-15 ***
Year2009/10   0.1603     0.1335   1.201  0.22989
Year2010/11  -0.1180     0.1495  -0.789  0.43008
Year2011/12   0.5119     0.1616   3.169  0.00153 **
Year2012/13   0.3060     0.1482   2.065  0.03893 *
Year2013/14  -0.3309     0.1564  -2.115  0.03440 *
Year2014/15  -1.1593     0.1925  -6.022 1.72e-09 ***
Year2015/16  -0.1440     0.1751  -0.823  0.41066
Year2016/17  -1.5998     0.3655  -4.376 1.21e-05 ***
Year2017/18  -3.0619     0.4131  -7.412 1.25e-13 ***
Year2018/19  -2.2478     0.2590  -8.680  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                                   edf Ref.df      F  p-value
s(x,y)                         23.4814     49 10.349  < 2e-16 ***
s(pmax(I(DistToShore/1000), 2.5))  5.3010      9 15.332  < 2e-16 ***
s(BotT_HYCOM_Monthly)              4.3877      9 11.978  < 2e-16 ***
s(BotS_HYCOM_Monthly)              1.0724      9  4.919 3.73e-12 ***
s(UV_HYCOM_Monthly)                0.7782      9  0.340 0.040905 *
s(log10(Chl_Monthly))              1.0433      9  1.178 0.000648 ***
s(WindSpeed_Monthly)               3.4020      9  6.070 3.76e-14 ***
s(SST_MUR_Monthly)                 4.7007      9 15.865  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.0173   Deviance explained = 19.5%
-REML =  16544  Scale est. = 4.436     n = 357307

Method: REML   Optimizer: outer newton
full convergence after 16 iterations.
Gradient range [-0.0003886187,4.789546e-05]
(score 16543.81 & scale 4.435952).
Hessian positive definite, eigenvalue range [0.2513791,27308.47].
Model rank =  128 / 128
Basis dimension (k) checking results. Low p-value (k-index<1) may
indicate that k is too low, especially if edf is close to k'.
                                   k'    edf k-index p-value
s(x,y)                         49.000 23.481    0.86   0.040 *
s(pmax(I(DistToShore/1000), 2.5))  9.000  5.301    0.90   0.010 **
s(BotT_HYCOM_Monthly)              9.000  4.388    0.93   0.275
s(BotS_HYCOM_Monthly)              9.000  1.072    0.93   0.560
s(UV_HYCOM_Monthly)                9.000  0.778    0.93   0.545
s(log10(Chl_Monthly))              9.000  1.043    0.92   0.270
s(WindSpeed_Monthly)               9.000  3.402    0.92   0.185
s(SST_MUR_Monthly)                 9.000  4.701    0.91   0.085 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

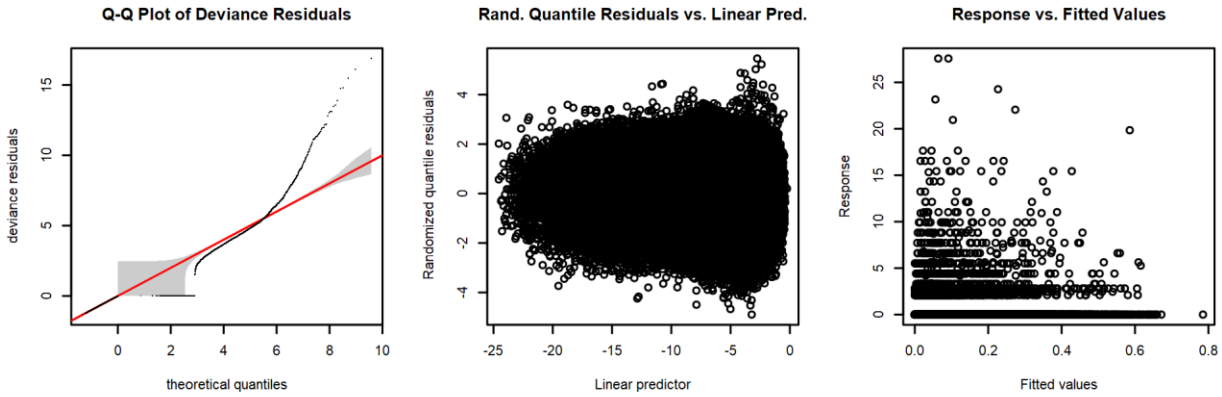


Figure 78. Statistical output for the Winter model for the South of Hatteras Island region.

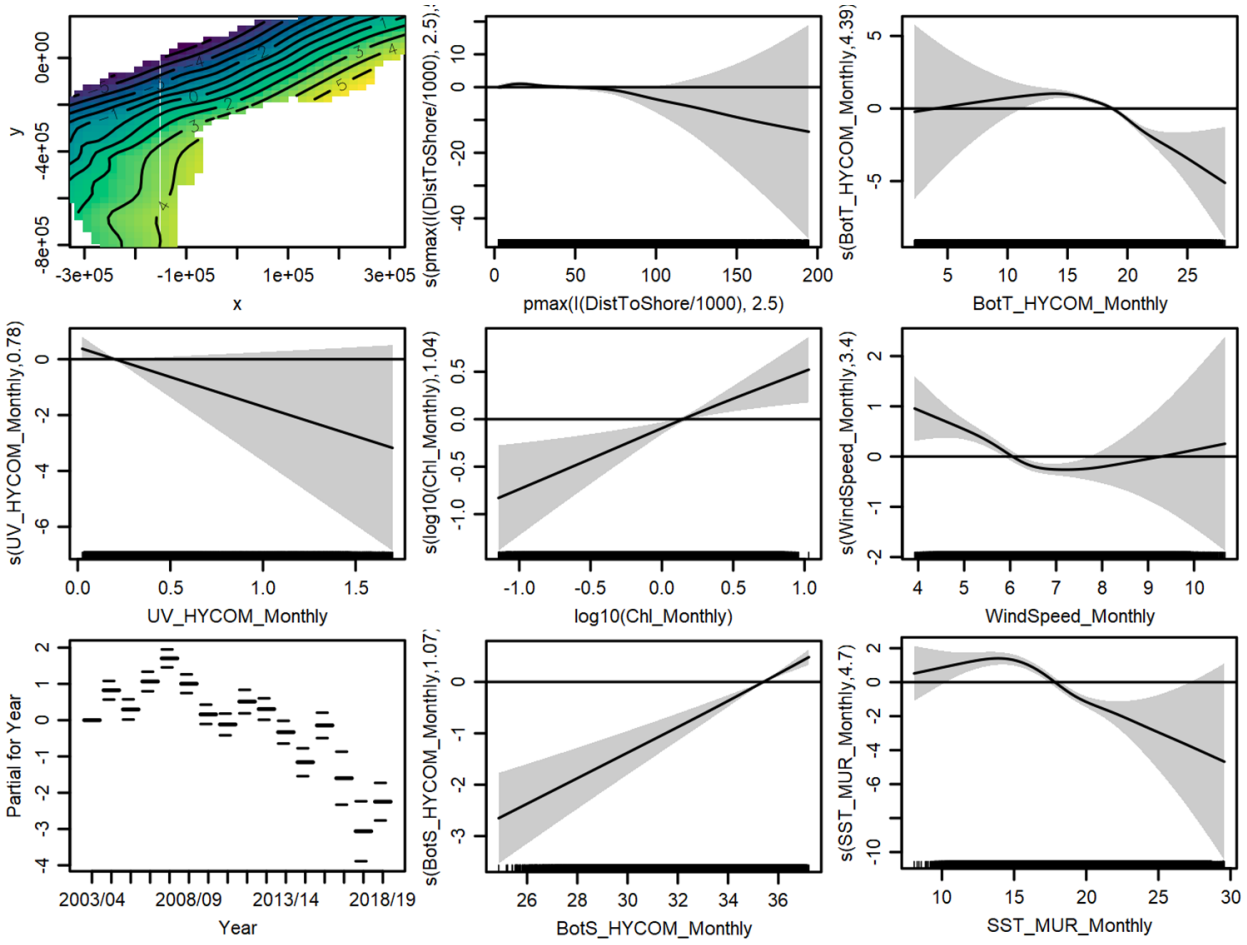


Figure 79. Functional plots for the Winter model for the South of Hatteras Island region.

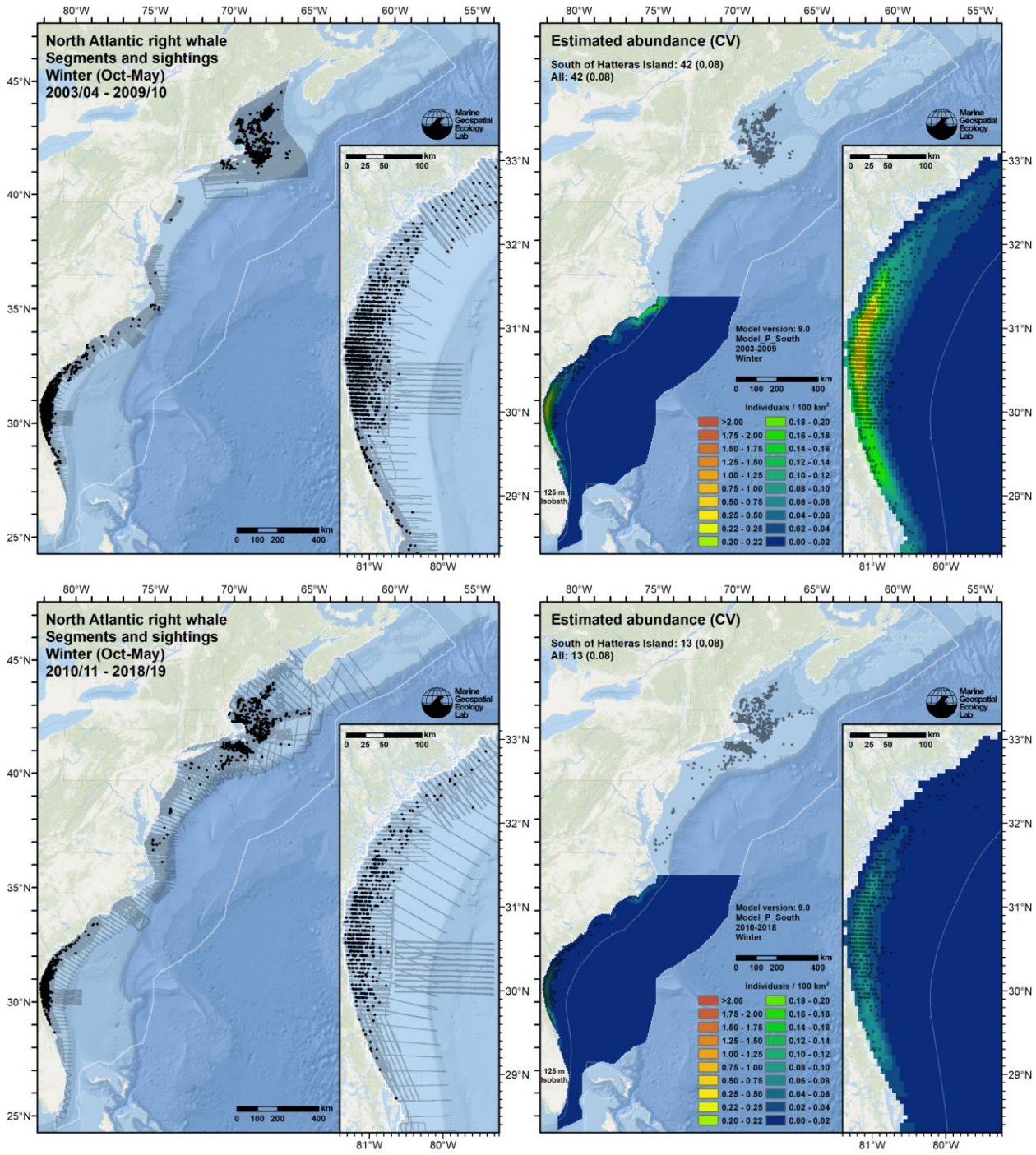


Figure 80. Survey effort and right whale sightings (left maps) and mean density predictions with sightings overlaid (right maps), for the 2003-2009 (top row) and 2010-2018 (bottom row) eras, for the Winter season (October-May). Effort and sightings are all of those that occurred during those eras during October-May. Density predictions represent the mean across all those years and months.

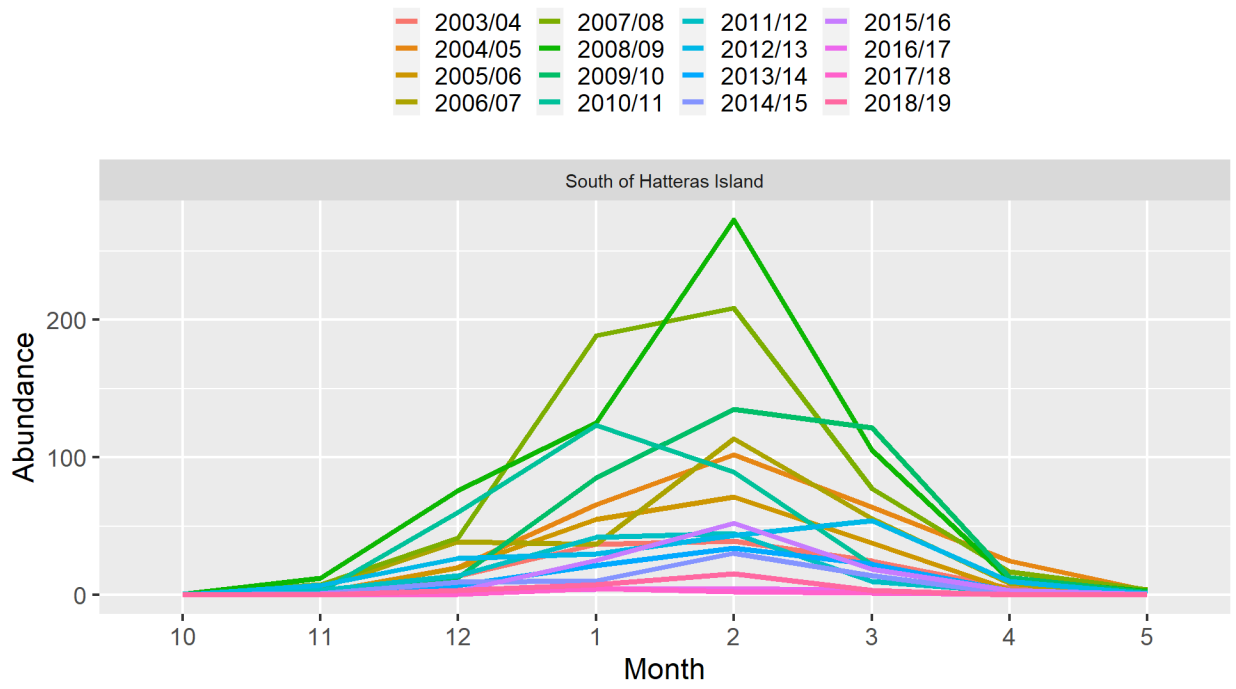


Figure 81. Predicted right whale abundance south of Hatteras Island for all modeled years and months, October-May.

5.3.2.2. Hatteras Island to Nantucket Shoals

Between Cape Hatteras and Nantucket Shoals, during the 2003-2009 era, a substantial portion of the continental shelf was only surveyed between June and August (Figure 82). The surveying that did occur outside of these months was restricted to regions north of 40 ºN (surveyed by the NARWSS program in 2003-2007), coastal New Jersey (surveyed by NJ-DEP in 2008-2009), and coastal Virginia and North Carolina (surveyed by UNCW and SEFSC in 2005-2008). In the 2010-2018 era, comprehensive coverage was provided throughout the year by both the broad-scale AMAPPS program and regional programs by NARWSS, New England Aquarium, NYS-DEC, VAMSC, UNCW, and HDR, although these programs were active during different ranges of years.

As we discussed in section 5.3.1.3 above, the lack of effort during the 2003-2009 era during non-summer months, when right whales were likely to be present in the mid-Atlantic, was a major impediment to modeling this era. The gaps in coverage precluded addressing inter-era effects with the year as a categorical covariate, and less than 10 sightings were reported in the 2003-2009 era, which precluded any possibility of modeling the era independently. However, more than 200 sightings were reported during the 2010-2018 era. This increase is likely to have resulted both from the increase in survey coverage but also an increase in the use of the mid-Atlantic area by right whales, which has also been observed with passive acoustic monitoring (Davis et al. 2017). To facilitate modeling the 2003-2009 era and also allow for an inter-era difference in overall abundance in the region, we fitted a model to the entire 2003-2018 period (Figure 83) but included the era as a categorical covariate. We did not split the data into seasons, on the basis that passive acoustic monitoring had detected right whales throughout the region in all months (Salisbury et al. 2015; Kraus et al. 2016; Davis et al. 2017). Instead, we fitted a single model to all data.

As an additional complication, sightings were concentrated in the northern part of the region, beginning in 2010 off Rhode Island Sound (Figure 69) and then progressing in later years to waters south of Martha's Vineyard, Nantucket Island, and Nantucket Shoals. However, this "South of the Islands" area also received some of the highest survey effort in the region. So, while right whale experts had noted that this area regularly saw high densities of right whales in recent years, it was not immediately clear how these densities differed from those farther south (e.g. off New York). To help correct for differences in survey effort, we first attempted to use as a covariate a bivariate smooth of spatial location. Although this helped distinguish the region of apparently higher density to the north, it resulted in

extreme overfitting in the south, causing patches of elevated density to bloom around individual sightings. Similar results were obtained when we removed the spatial smooth and instead allowed the model to consider a full suite of covariates. Ultimately, we tested a large array of models with just a few covariates, without the spatial smooth.

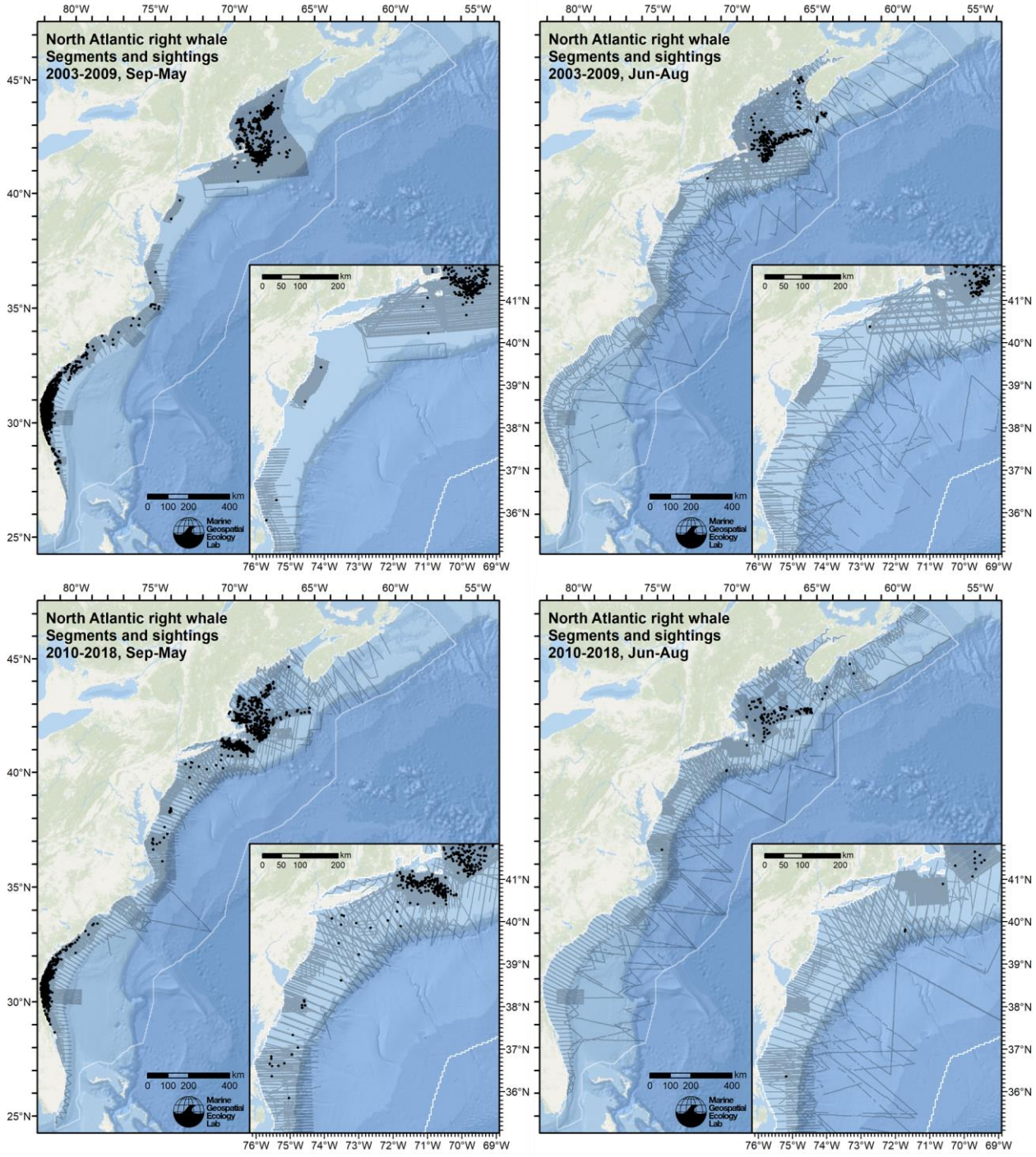


Figure 82. Survey segments and right whale sightings for the 2003-2009 (top row) and 2010-2018 (bottom row) eras, for the non-summer (left column) and summer (right-column) months. Insets showing the marked inter-era differences in effort and sightings in the region between Hatteras Island and Nantucket Shoals. Figure 83 shows the segments and sightings used to fit the model, which spanned both eras.

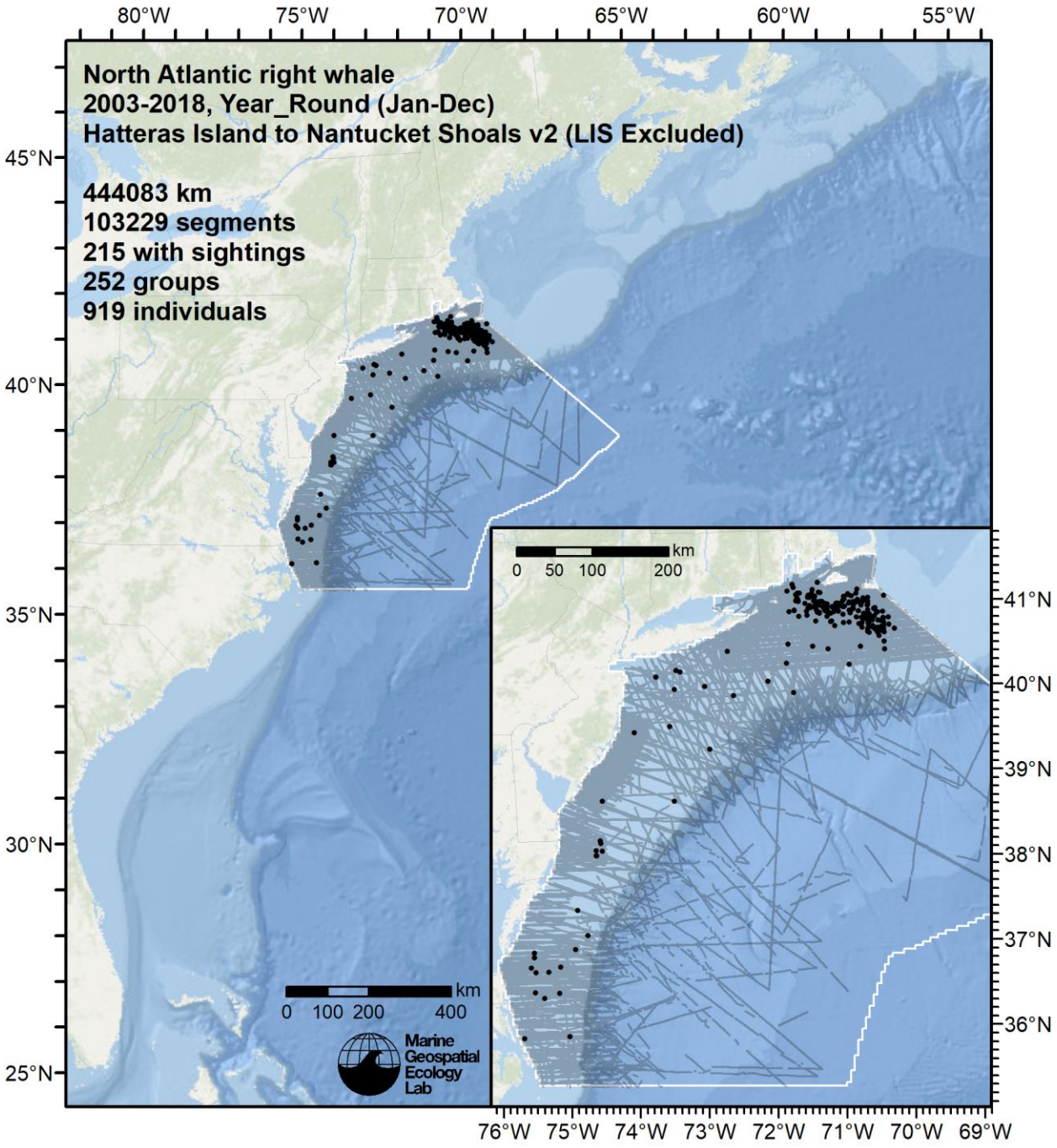


Figure 83. Segments and sightings used to fit the model for the Hatteras Island to Nantucket Shoals region.

For candidate models, we initially tested models having Depth, Era, and all combinations of two dynamic covariates. To boost the possibility of distinguishing the South of the Islands area from the shelf to the south, we tested each pair of dynamic covariates in two ways: first as two univariate additive terms, and alternately as a single bivariate interaction term (using a tensor product smoother). All of these models predicted high density in either Nantucket Sound, Buzzards Bay, or Narragansett Sound, where no sightings occurred on surveys we used, nor did acoustic monitoring (in Nantucket Sound, as reported by Davis et al. (2017)), nor opportunistic sightings (for all three areas, as reported by the NOAA Right Whale Sighting Advisory System) indicate high right whale presence. To help prevent

overprediction of density in these inshore areas, we added the Fetch covariate, to help distinguish them from the open ocean areas South of the Islands, where right whales were frequently sighted in recent years.

The models fitted to contemporaneous covariates at a monthly resolution again outperformed the models that used contemporaneous daily covariates or climatological covariates, as occurred in the South of Hatteras Island region. Most of the best-scoring models included either surface or bottom salinity in an interaction with another dynamic covariate. We selected the model that achieved the best REML score, which retained Depth, Era, Fetch, and an interaction between sea surface salinity (SSS) and MUR SST (Figure 84, Figure 85). (The model with HYCOM SST performed essentially the same but explained 0.2% less deviance.)

The selected model indicated decreasing density with increasing depth but also with decreasing fetch, yielding highest density in shallow waters that were not too close to shore or enclosed by land. The SSS:SST interaction indicated highest density in an envelope defined by low temperature (<15 ºC) and moderate salinity (31-34 PSU), which likely characterizes the winter and springtime waters of the mid-Atlantic shelf. The categorical Era covariate exhibited a strong effect, with the coefficient for the 2010-2018 era estimated at 2.729 on the scale of the link function, indicating that density in the 2010-2018 era was roughly 15x higher than in the 2003-2009 era, when all other covariates were held constant.

Mean density predictions (Figure 86, Figure 87) and monthly abundance summaries (Figure 88, Figure 89) for the two eras reflected this large difference. Monthly predictions suggested right whales first occupy the region in appreciable density in December, at the onset of winter, and then rise to peak abundance in March before falling to near zero in June. This prediction accords well with monthly patterns in sightings in the 2010-2018 era (see Figure 106 for the complete sequence of monthly predictions). However, passive acoustic monitoring during this era reported detections South of the Islands every month of the year (see section 5.3.4).

As might be expected, the model predicted highest densities in the South of the Islands region, where right whales have been observed aggregating in large groups in recent years. However, during the peak February-April period, the model also predicted relatively high density down the continental shelf, nearly to Hatteras Island, such that during the 2010-2018 era, mean abundance for the modeled area for these three months was 354, 410, and 324, comprising 75% or more of the population. We suspect the model overpredicted density in the mid-Atlantic area, especially in 2010 and 2011, for which peak abundance was predicted at nearly 600 whales (Figure 89). During these years, springtime right whale aggregations were sighted in Rhode Island Sound, while in subsequent years, when lower total abundance was predicted, aggregations occurred progressively farther to the east. Determining how best to model the area between Hatteras Island and Nantucket Shoals remains a priority for future updates to our right whale model (see section 6.3.1).

```
Family: Tweedie(p=1.228)
Link function: log

Formula:
IndividualsCorrected ~ offset(log(SegmentArea)) + Era + s(Depth, bs = "ts") + s(Fetch_50km, bs = "ts") +
    te(SST_MUR_Monthly, SSS_HYCOM_Monthly, bs = "ts")

Parametric coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   -28.144      1.199 -23.469  < 2e-16 ***
Era2010-2018    2.729      0.363   7.518  5.6e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                                    edf Ref.df     F  p-value
s(Depth)                         0.8767      9 0.733   0.0058 **
s(Fetch_50km)                    0.9863      9 2.577 5.82e-07 ***
te(SST_MUR_Monthly,SSS_HYCOM_Monthly) 7.3870     24 7.524  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.00991   Deviance explained = 33.7%
-REML = 1775.6  Scale est. = 14.004    n = 103229

Method: REML   Optimizer: outer newton
full convergence after 10 iterations.
Gradient range [-1.733265e-07,6.765326e-08]
(score 1775.597 & scale 14.00368).
Hessian positive definite, eigenvalue range [0.3710646,1512.061].
Model rank =  44 / 44

Basis dimension (k) checking results. Low p-value (k-index<1) may
indicate that k is too low, especially if edf is close to k'.

                                    k'    edf k-index p-value
s(Depth)                         9.000  0.877    0.78   0.005 **
s(Fetch_50km)                    9.000  0.986    0.81   0.020 *
te(SST_MUR_Monthly,SSS_HYCOM_Monthly) 24.000  7.387    0.68  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

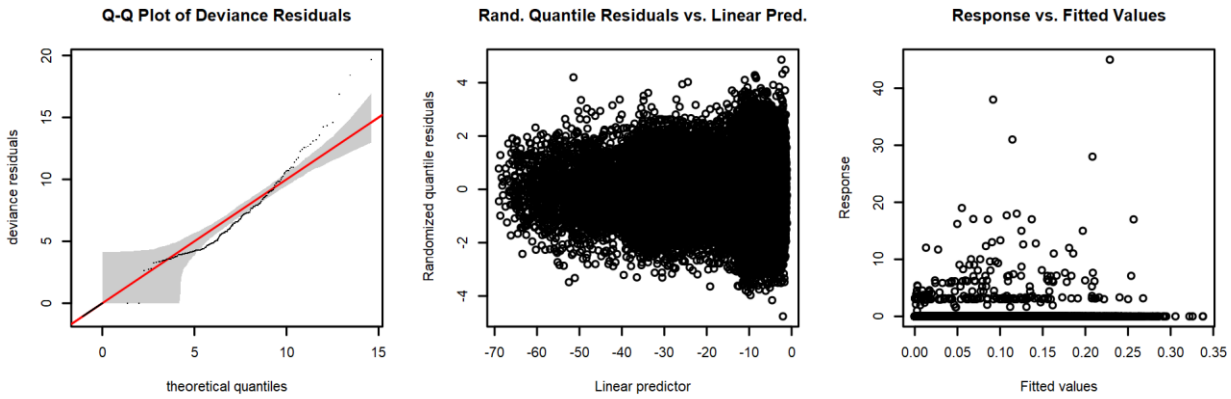


Figure 84. Statistical output for the model for the Hatteras Island to Nantucket Shoals region.

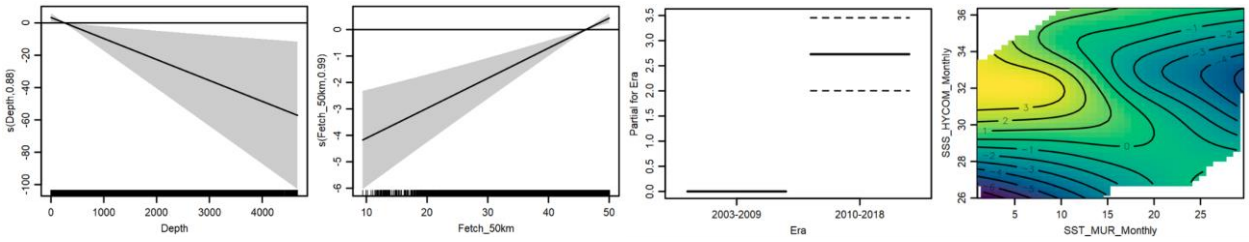


Figure 85. Functional plots for the Hatteras Island to Nantucket Shoals region.

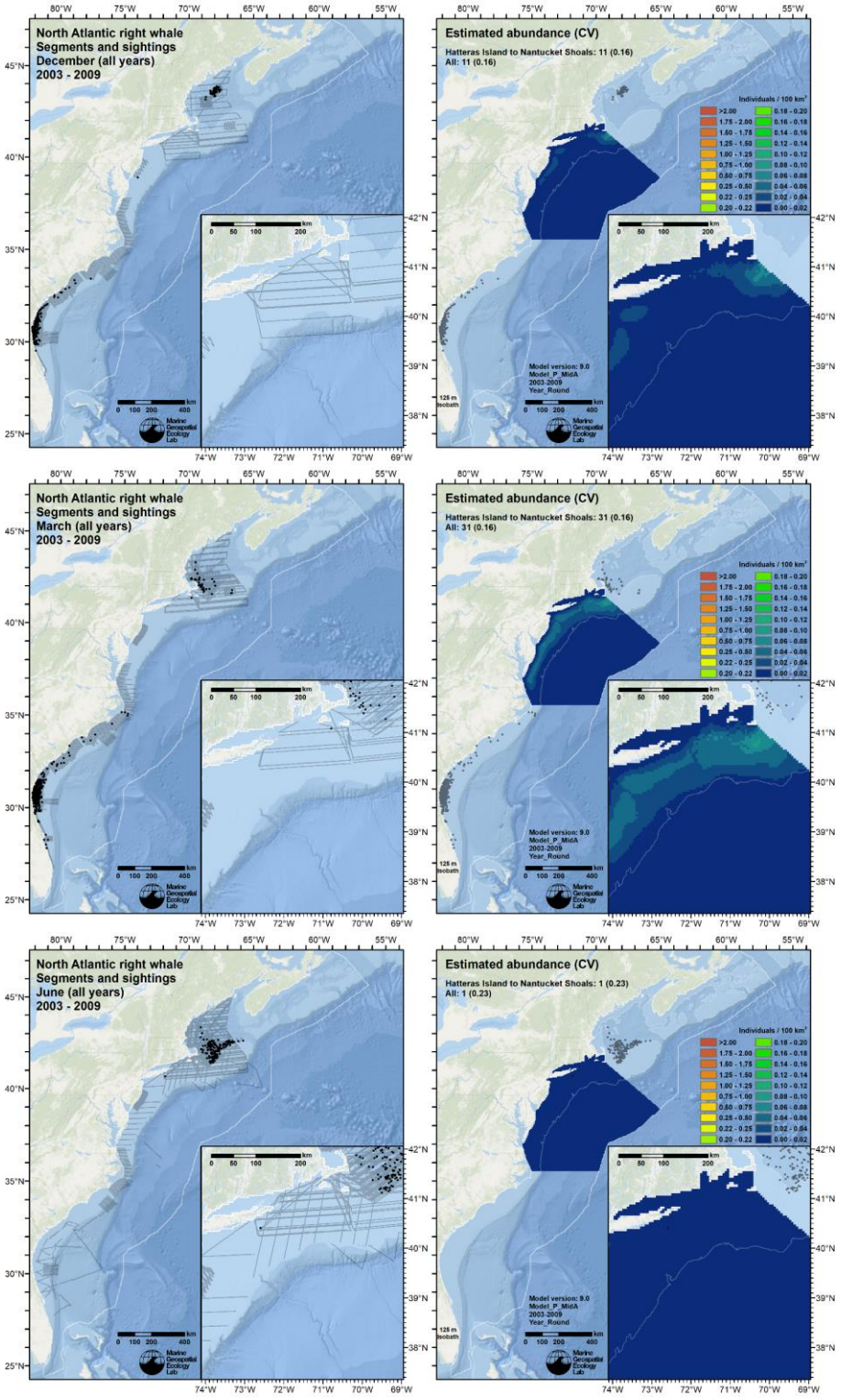


Figure 86. Survey effort and right whale sightings (left maps) and mean density predictions with sightings overlaid (right maps) for the 2003-2009 era, for December (top row), March (middle row), and June (bottom row), representing the beginning, middle, and end of the main seasonal period of right whale presence between Cape Hatteras and Nantucket Shoals. Effort and sightings are cumulative, and density predictions represent means, across all years of the era for the indicated month. Note how low density predictions are compared to the 2010-2018 era (Figure 87).

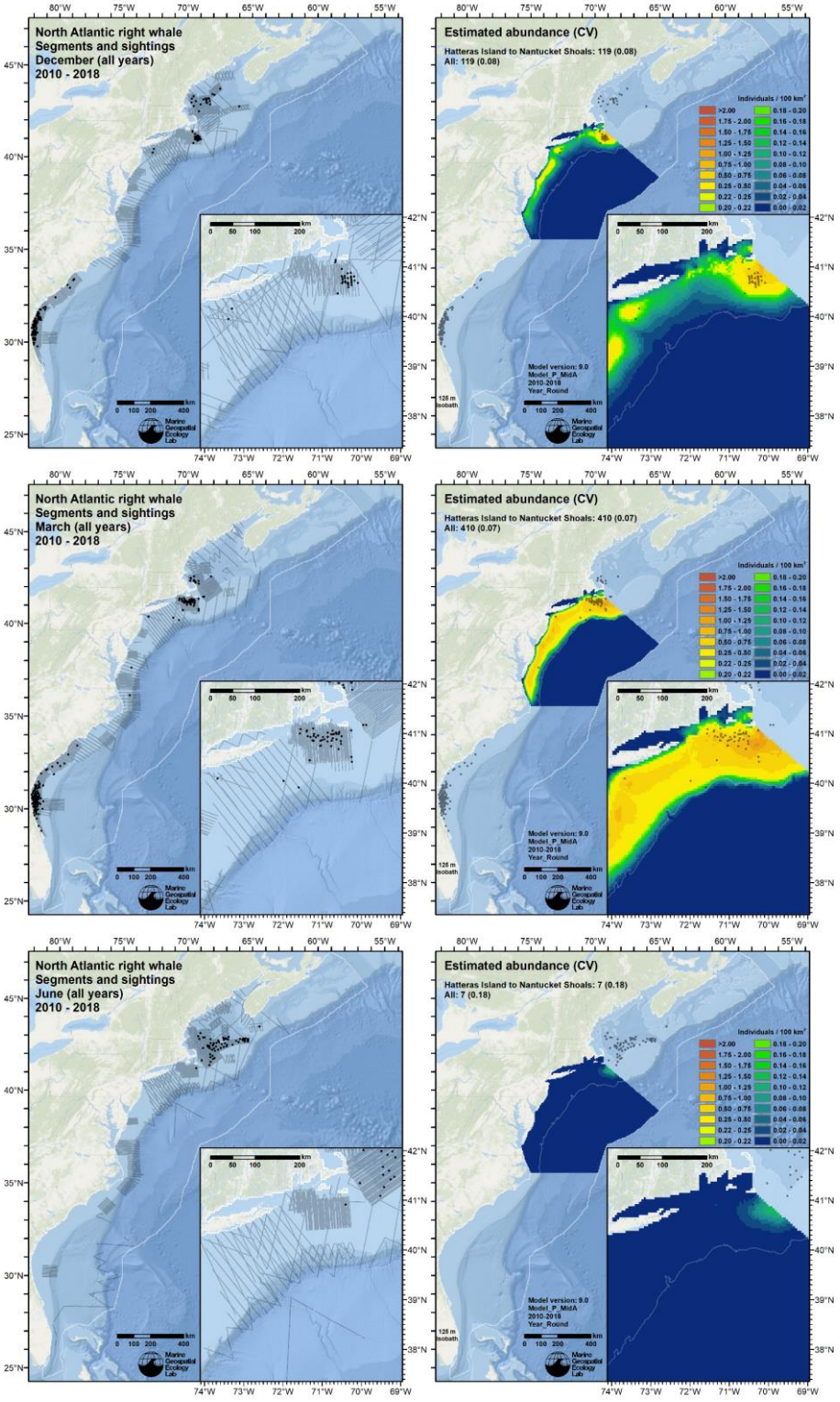


Figure 87. Survey effort and right whale sightings (left maps) and mean density predictions with sightings overlaid (right maps) for the 2010-2018 era, for December (top row), March (middle row), and June (bottom row), representing the beginning, middle, and end of the main seasonal period of right whale presence between Cape Hatteras and Nantucket Shoals. Effort and sightings are cumulative, and density predictions represent means, across all years of the era for the indicated month. Note how high density predictions are compared to the 2003-2009 era (Figure 86).

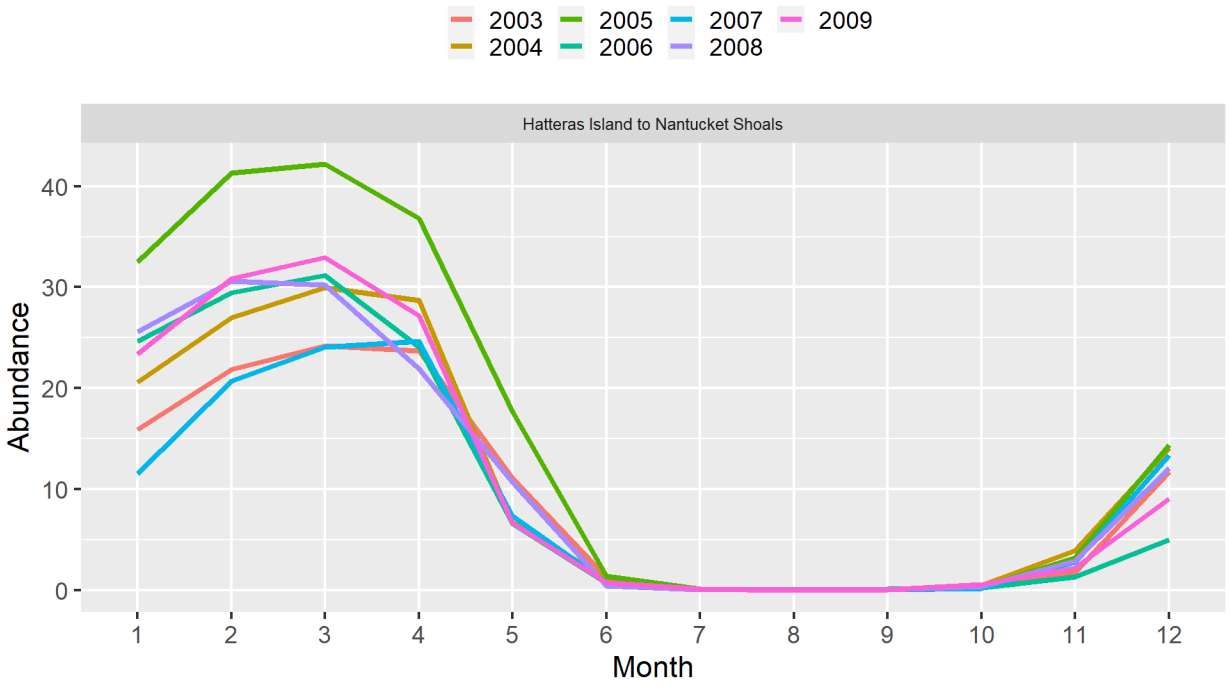


Figure 88. Predicted right whale abundance for the 2003-2009 era between Hatteras Island and Nantucket Shoals.

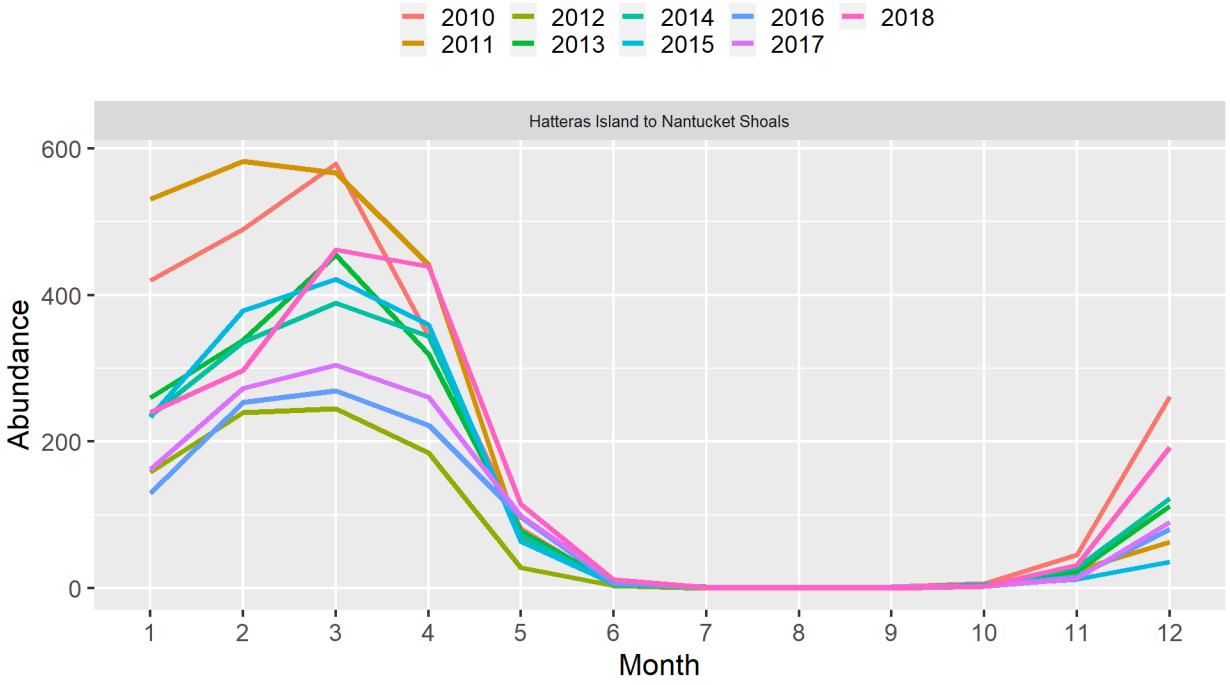


Figure 89. Predicted right whale abundance for the 2010-2018 era between Hatteras Island and Nantucket Shoals. Note that the y-axis scale is different than that of Figure 88.

5.3.2.3. North of Nantucket Shoals

Prior to the changes to right whale distribution patterns (see section 5.3.1.1), a common view had emerged of an annual, counterclockwise cycle of right whales around the Gulf of Maine region, that seemed to hold for the 2003-2009 era. Starting in winter, right whales occupy Cape Cod Bay to feed, emerging in spring into the Great South Channel, where they feed into early summer (Kenney et al. 2001). Then, as summer progresses, they move along the northern edge of Georges Bank and shift north to Canadian basins in the Bay of Fundy, Roseway Basin, and further north on the Scotian Shelf (Kenney et al. 2001; Brillant et al. 2015). In late summer, fall, and early winter, they continue around to the central and western Gulf of Maine (Weinrich et al. 2000; Cole et al. 2013), before returning to Cape Cod Bay or migrating south to the calving grounds (Kenney et al. 2001). This cycle is evident in the sightings reported by surveys conducted during the 2003-2009 era, principally in the data from the NARWSS program, which collected the bulk of the survey data for the region. Some evidence of the cycle can be seen in the surveys of 2010-2018 as well, but changes in the distribution of survey effort (Figure 73, Figure 74), especially in the NARWSS program itself, which stopped flying broad-scale surveys in 2007, confound a simple comparison of the eras. However, there is no question that substantial changes have occurred in the Gulf of Maine, as revealed in passive acoustic monitoring and by visual surveys not able to be used for density surface models such as ours (Davis et al. 2017; Record et al. 2019; Davies et al. 2019).

The heterogeneous sampling in the systematic surveys available for our model precluded the use of year as a categorical covariate to capture interannual variability not captured by dynamic covariates. As a fallback, like we did with the Hatteras Island to Nantucket Shoals region, we used Era as a two-level categorical covariate instead. Right whale movements around the Gulf of Maine are believed to track zooplankton prey distributions, at least in winter, spring, and summer, but we lacked an effective covariate representing prey distribution. Modeling these movements strictly from the readily-available covariates we had obtained (section 5.3.1.4) proved difficult, particularly given the lack of homogenous survey effort of the entire Gulf of Maine region, so we split the year into three seasons that we modeled independently, allowing different covariate relationships to be fitted for different parts of the cycle.

5.3.2.3.1 Spring (March-July)

We defined the "Spring" season for this model as March-July, representing the period in which right whales move out of Cape Cod Bay (March-April), across the Great South Channel (April-June), and along the northern edge of Georges Bank and north to Canada (June-July). The monthly sequences of sightings generally reflected this pattern during both eras (Figure 105, Figure 106), although it was more pronounced for the 2003-2009 era. A large number of sightings were available (Figure 90), but survey effort was generally restricted to U.S. waters, except along the northern edge of Georges Bank. During the 2010-2018 era, the AMAPPS and NARWSS programs extended further into Canada but only in July (Figure 98). Ultimately, this lack of temporally replicated coverage in Canada proved problematic for modeling north of Georges Bank, and to avoid aberrant predictions in parts of Canada that had low temporal replication, we clipped the spatial extent of the model closely to the survey transects conducted in the U.S. and the northern edge of Georges Bank (Figure 90).

The large number of sightings allowed us to fit a model that used a full suite of covariates. However, we found that none of the available dynamic covariates correlated well with the easterly movement of whales across the Great South Channel. To capture this movement, we introduced a bivariate interaction between day of year and the distance to the top of Cape Cod Bay (using a tensor product smoother). Finally, given that the spatial extent of the model was closely limited to the extent of the survey effort, we also included a bivariate smooth of spatial location as a candidate covariate.

The final model (Figure 91) retained seven covariates (Figure 92), including the bivariate smooth of spatial location, the interaction between day of year and distance to Cape Cod Bay, and the categorical Era covariate. Models built with climatological covariates outperformed those with contemporaneous covariates, in contrast with models built for the southern regions. As a result, the only covariate that varied between the eras was the Era covariate itself, for which the estimate for the 2010-2018 era was -0.57586 on the scale of the link function. Thus, the density pattern predicted for the 2010-2018 era was the same as the 2003-2009 era (Figure 93), just uniformly lower by approximately 40%, with a mean abundance of 158 in 2010-2018 vs 282 in 2003-2009. This result is consistent with the view that fewer right whales are lingering in the Great South Channel to feed, but instead moving straight to Canada.

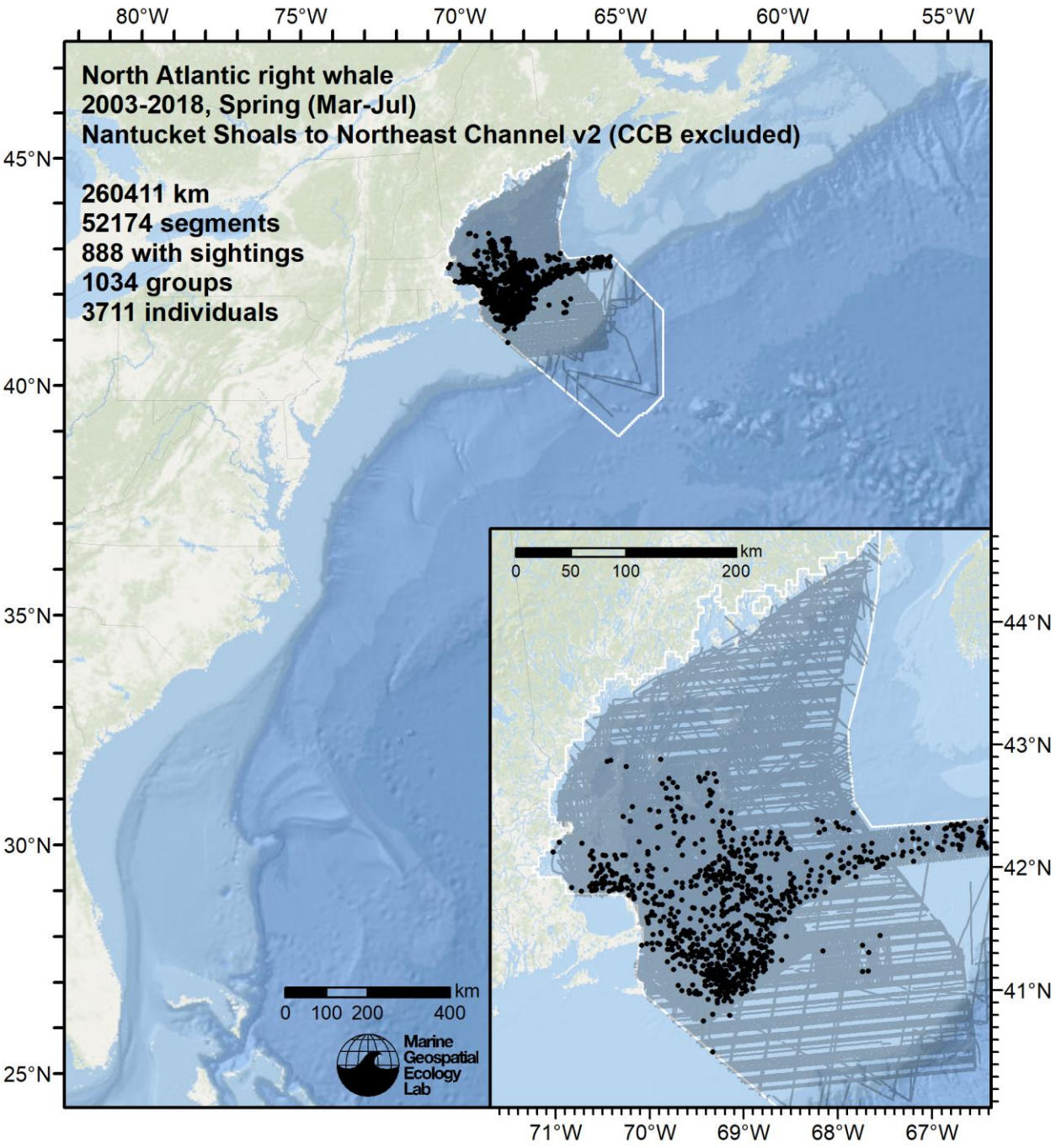


Figure 90. Segments and sightings used to fit the Spring model for the North of Nantucket Shoals region.

The dynamic oceanographic covariates retained by the model suggested peak density at about 9 ºC surface temperature and 5 ºC bottom temperature, and high densities at both very low chlorophyll concentrations, representing offshore waters, and relatively high concentrations, representing very inshore waters (e.g. near Cape Cod). The DayOfSeason:DistToCCB interaction showed a peak close to Cape Cod Bay one month into the season (in April), when right whales appeared in large numbers west of the Great South Channel, and a progressive movement away from the Bay as the season progressed. The bivariate smooth of spatial location appears to have been used to boost density along the northern edge of Georges Bank, likely because no other covariate was able to distinguish this area from others of lower density.

```
Family: Tweedie(p=1.259)
Link function: log
Formula:
IndividualsCorrected ~ offset(log(SegmentArea)) + Era + s(x, y, bs = "ts", k = 50) + s(I(DistTo125m/1000), bs = "ts") +
    te(I(DistToCCB/1000), DayOfSeason, bs = "ts") + s(ClimBotT_HYCOM_Month bs = "ts") + s(ClimChl_Monthly, bs = "ts") +
    s(ClimSST_HYCOM_Monthly, bs = "ts")

Parametric coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  -21.22138    0.21487 -98.763  < 2e-16 ***
Era2010-2018  -0.57586    0.08071  -7.135 9.81e-13 ***
---
Signif. Codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                                    edf Ref.df     F  p-value
s(x,y)                           27.803     49 7.555  < 2e-16 ***
s(I(DistTo125m/1000))             6.453      9 7.846  < 2e-16 ***
te(I(DistToCCB/1000),DayOfSeason) 11.464     24 4.395  < 2e-16 ***
s(ClimBotT_HYCOM_Monthly)         5.266      9 7.910  < 2e-16 ***
s(ClimChl_Monthly)                6.474      9 7.725 1.22e-15 ***
s(ClimSST_HYCOM_Monthly)          3.493      9 4.050 3.53e-09 ***
---
Signif. Codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.0397   Deviance explained =   30%
-REML = 6135.8  Scale est. = 10.486    n = 52173

Method: REML   Optimizer: outer newton
full convergence after 14 iterations.
Gradient range [-0.002778961,0.001135086]
(score 6135.792 & scale 10.48638).
Hessian positive definite, eigenvalue range [0.185948,4539.382].
Model rank =  111 / 111
```

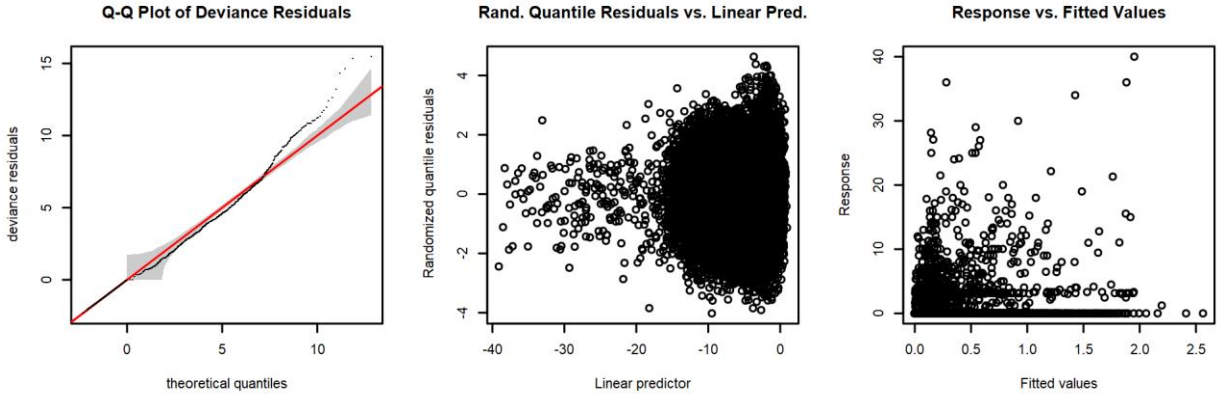


Figure 91. Statistical output for the Spring model for the North of Nantucket Shoals region.

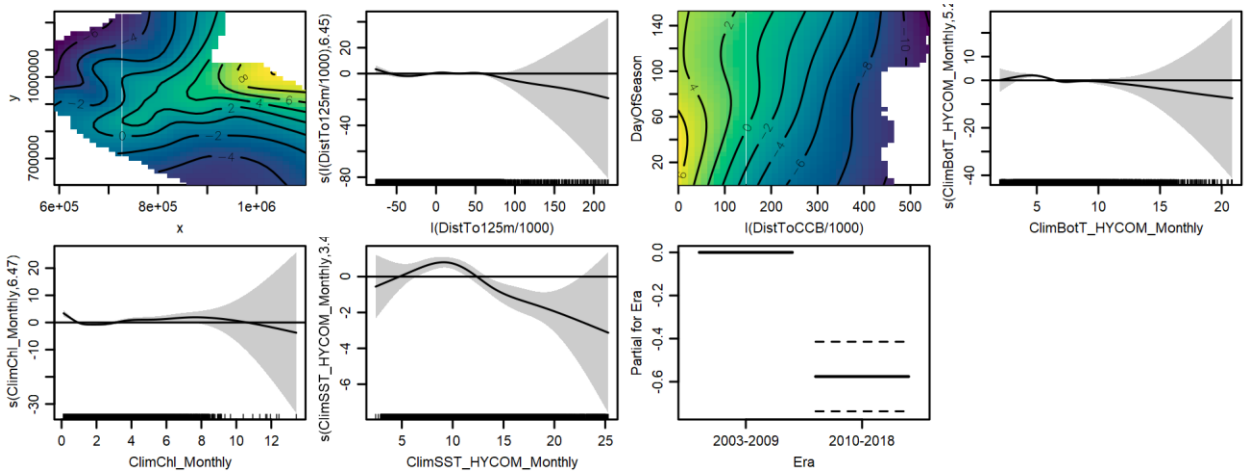


Figure 92. Functional plots for the Spring model for the North of Nantucket Shoals region.

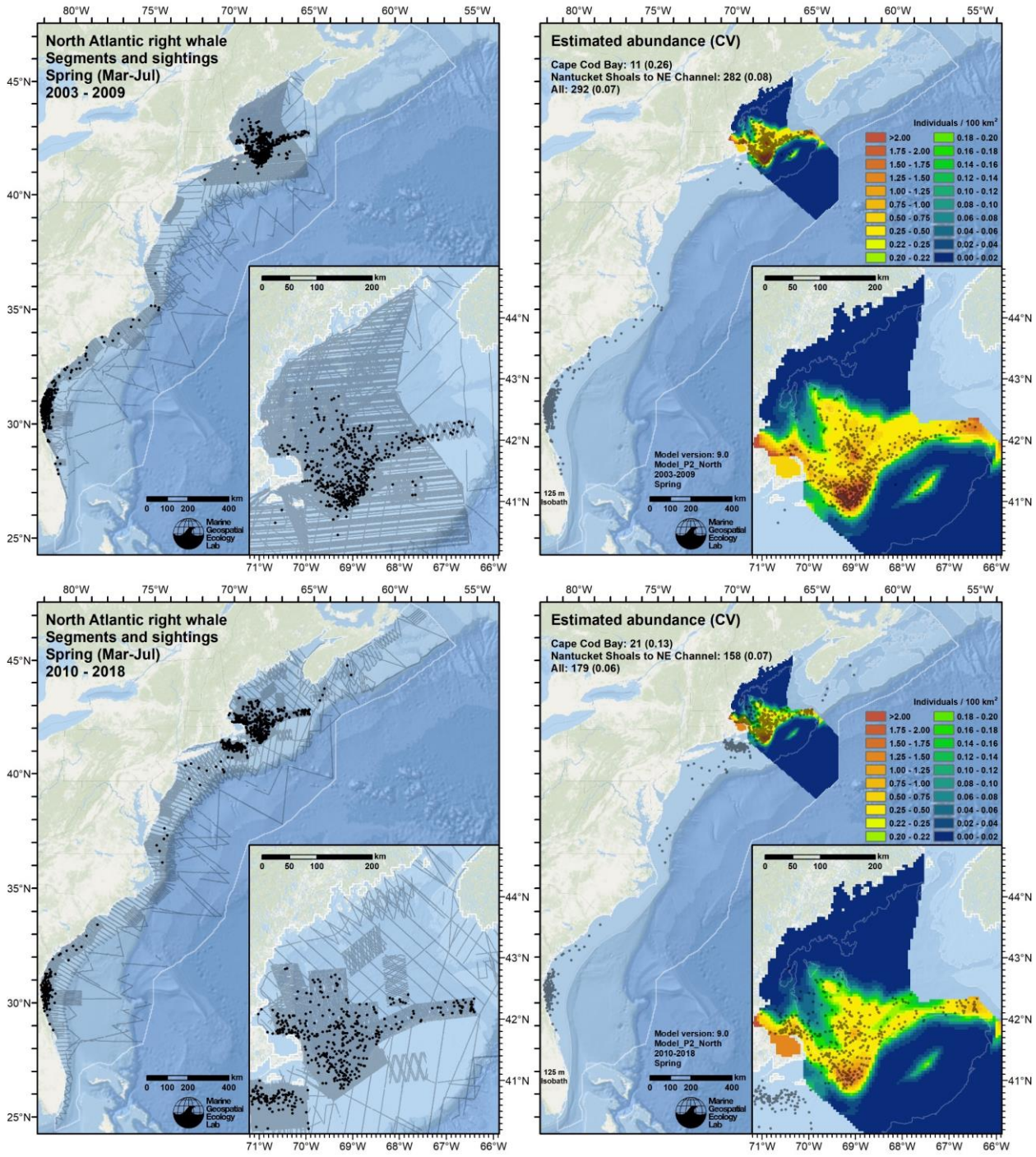


Figure 93. Survey effort and right whale sightings (left maps) and mean density predictions with sightings overlaid (right maps), for the 2003-2009 (top row) and 2010-2018 (bottom row) eras, for the Spring season (March-July). Effort and sightings are all of those that occurred during those eras during March-July. Density predictions represent the mean across all those years and months.

5.3.2.3.2 Summer (August-September)

The "Summer" period of August and September was the only time of year that survey effort covered the Gulf of Maine, Bay of Fundy, and the southern Scotian Shelf during both eras (Figure 94). As noted above, we explored including July in this season rather than the "Spring" season, which would have contributed additional effort in Canadian waters in the 2010-2018 era. However, this made the Spring model more difficult to fit, and that three-month Summer model resulted in substantial predictions in July near the lower Bay of Fundy, in the Grand Manan Basin (GMB), a location where right whales were reliably known to aggregate in large numbers in late summer to feed. This seemed contrary to the findings of Davies et al.'s (2019) study of this area, which reported "In the 2000s, right whales immigrated into the GMB later in the summer (after 1 August, as indicated by increased PFE [power function exponent]), and did not accumulate into high densities until September." It went on to report "Since 2014, the animals appear to have been arriving earlier again, yet not staying resident in the GMB in most of these recent years." Our interpretation was that density in the GMB in July 2010-2018 was probably low, and the best approach was to place July with the Spring model, which we did not extrapolate into Canada beyond the northern edge of Georges Bank.

Formulating this model was very difficult. Although most areas of the continental shelf and slope to 2000 m depth received some effort in both eras (Figure 95), temporal replication was relatively sparse across this 19-year period, consisting of a few surveys by NARWSS, NEFSC (both AMAPPS and pre-AMAPPS), and New England Aquarium (two days of NLPSC aerial surveys), yielding only 41 sightings over about 51,000 km total effort. Sightings were concentrated in two well-known areas—the GMB and the vicinity of Roseway Basin—but a third concentration occurred between the two areas, along the eastern slope of the Scotian Shelf, west of Yarmouth, sighted by an NEFSC shipboard survey in 2007 (HB 07-09). Five more sightings were scattered around the Gulf of Maine, along the Northern Coastal Shelf, Cape Cod, and west of Georges Basin. It was therefore desirable that model predictions reflect the apparent high densities where sightings were concentrated, but also exhibit non-zero densities around the Gulf of Maine where sightings were scattered. As a further complication, only two sightings were reported in the 2010-2018 era. This, no doubt, reflected decreased use of once-important areas, such as the GMB (Davies et al. 2019). But effort was very sparse, and opportunistic sightings reported to NOAA's Right Whale Sighting Advisory System indicated that right whales had still been sighted regularly in GMB in the 2010-2018 era.

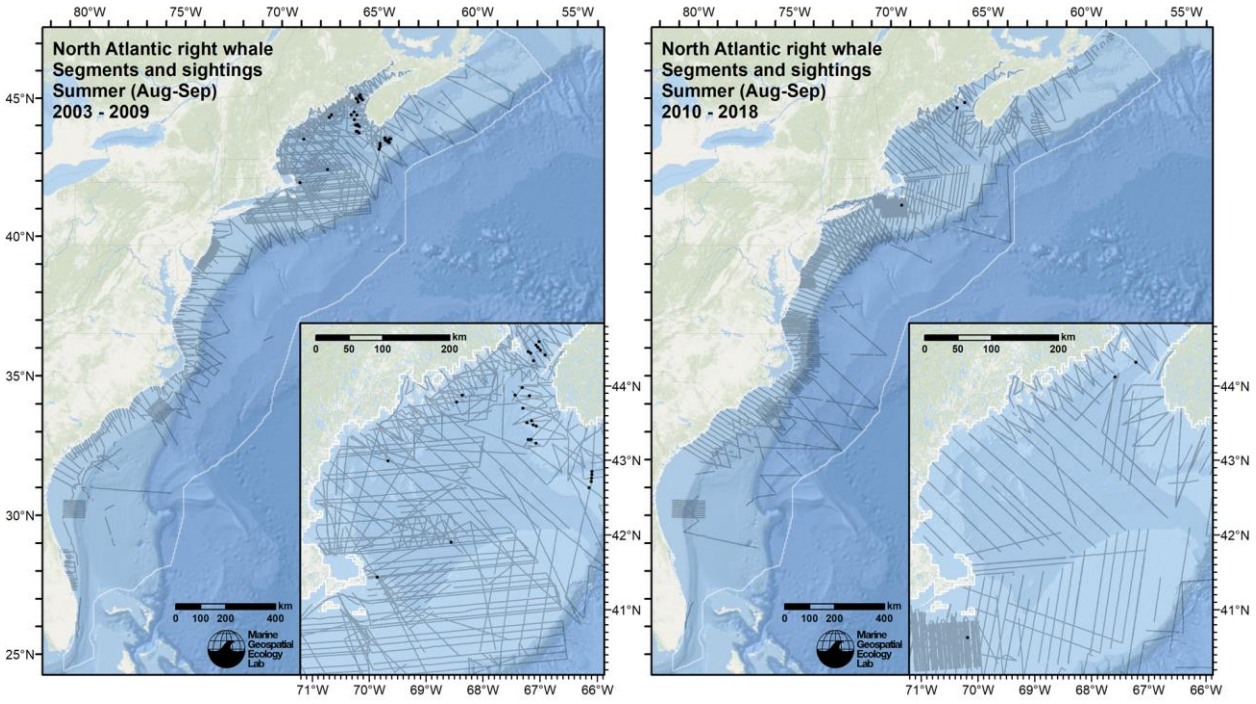


Figure 94. Survey segments and right whale sightings during the "Summer" (July-August) season defined for modeling north of Nantucket Shoals. Figure 95 shows the segments and sightings used to fit the model, which spanned both eras. Note that only two sightings were reported north of Nantucket Shoals in the 2010-2018 era.

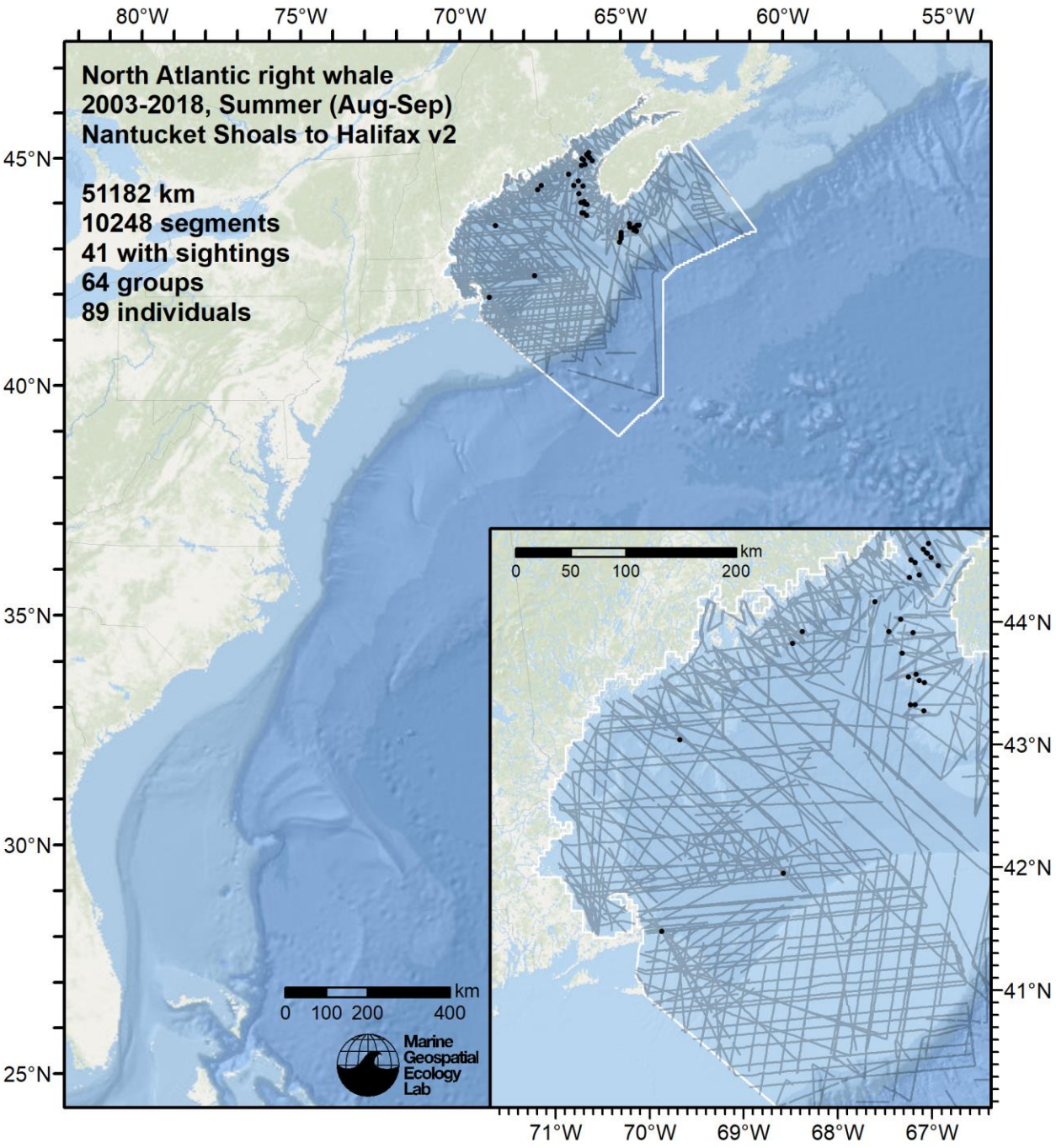


Figure 95. Segments and sightings used to fit the Summer model for the North of Nantucket Shoals region.

To address the apparent decrease in density between eras we included Era as a categorical covariate, as in the Spring model. Because the modeled area was nearly entirely covered at least once on both eras, we initially included a bivariate smooth of spatial location, but models built with this covariate were all severely overfitted. For example, the best scoring climatological-covariate model explained 70.4% of deviance with only spatial location, era, and bottom salinity. This model, and the others that used spatial location, predicted circular hotspots of density around each scattered sighting, which we judged unrealistic. We then removed the spatial location covariate and allowed the model to consider a full suite of static and dynamic covariates instead and relied on `mgcv`'s shrinkage smoothing to sort out which covariates provided the best correlations.

```
Family: Tweedie(p=1.276)
Link function: log
Formula:
IndividualsCorrected ~ offset(log(SegmentArea)) + Era + s(I(DistToShore/1000), bs = "ts") + s(I(DistTo300m/1000), bs = "ts") +
    s(ClimBotS_HYCOM_Monthly, bs = "ts") + s(ClimSST_MUR_Monthly, bs = "ts")

Parametric coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  -22.5346     0.6419 -35.104  < 2e-16 ***
Era2010-2018  -2.8955     0.6746  -4.292 1.78e-05 ***
---
Signif. Codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Approximate significance of smooth terms:
                           edf Ref.df     F  p-value
s(I(DistToShore/1000))    0.9183      9 0.887  0.00253 **
s(I(DistTo300m/1000))     1.0541      9 4.731 1.52e-11 ***
s(ClimBotS_HYCOM_Monthly) 3.2716      9 3.811 2.23e-08 ***
s(ClimSST_MUR_Monthly)    0.9918      9 2.003 1.03e-05 ***
---
Signif. Codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
R-sq.(adj) =  0.0188   Deviance explained = 50.6%
-REML = 328.05  Scale est. = 16.299    n = 10244

Method: REML   Optimizer: outer newton
full convergence after 11 iterations.
Gradient range [-2.45806e-07,1.152948e-07]
(score 328.0521 & scale 16.29889).
Hessian positive definite, eigenvalue range [0.3461328,226.1339].
Model rank =  38 / 38
Basis dimension (k) checking results. Low p-value (k-index<1) may
indicate that k is too low, especially if edf is close to k'.
                           k'   edf k-index p-value
s(I(DistToShore/1000))    9.000 0.918    0.86   0.040 *
s(I(DistTo300m/1000))     9.000 1.054    0.84   0.025 *
s(ClimBotS_HYCOM_Monthly) 9.000 3.272    0.80   0.005 **
s(ClimSST_MUR_Monthly)    9.000 0.992    0.89   0.900
---
Signif. Codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

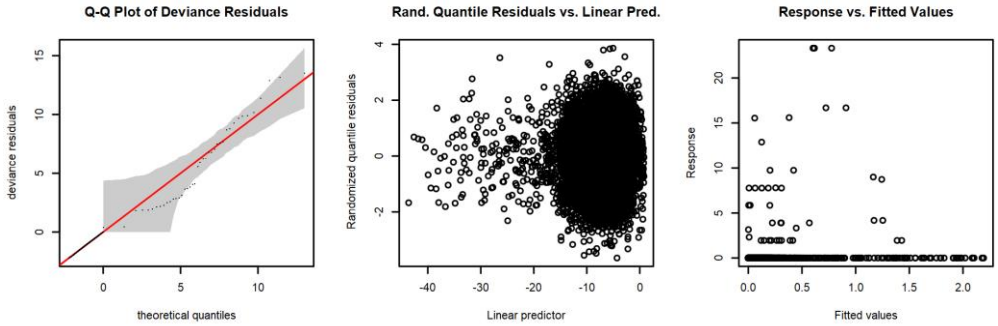


Figure 96. Statistical output for the Summer model for the North of Nantucket Shoals region.

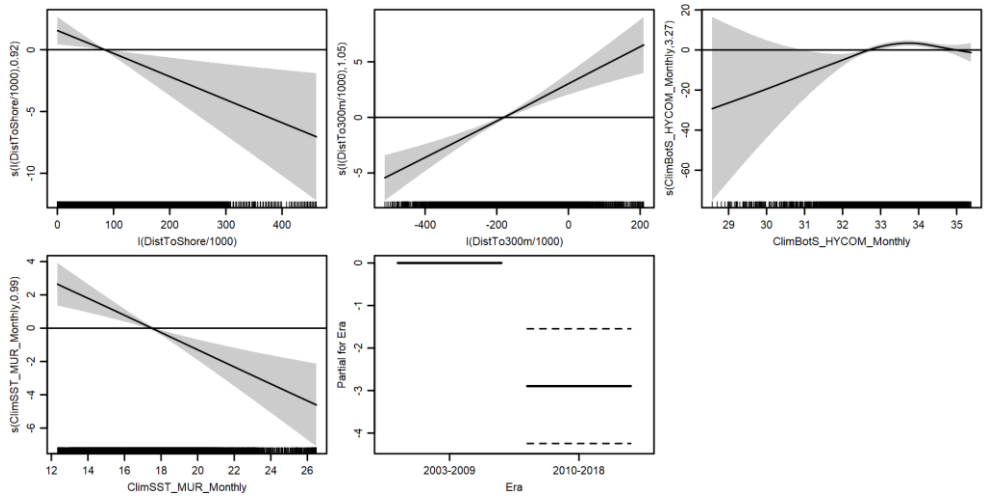


Figure 97. Functional plots for the Summer model for the North of Nantucket Shoals region.

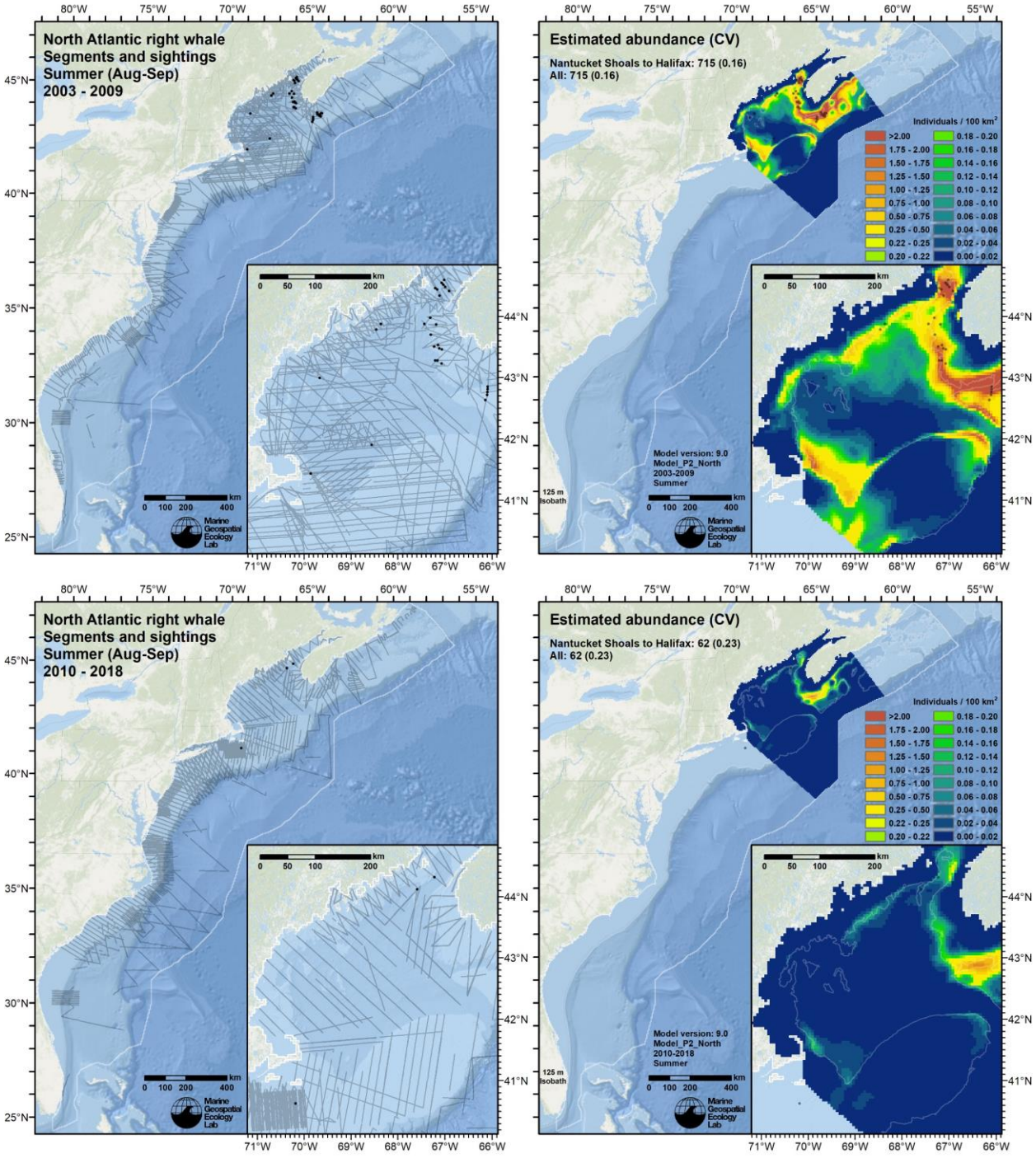


Figure 98. Survey effort and right whale sightings (left maps) and mean density predictions with sightings overlaid (right maps), for the 2003-2009 (top row) and 2010-2018 (bottom row) eras, for the Summer season (August-September). Effort and sightings are all of those that occurred during those eras during August-September. Density predictions represent the mean across all those years and months.

Models fitted with contemporaneous covariates generally scored higher in goodness of fit statistics than models fitted with climatological covariates but predicted total abundance between 1000-2000 whales. Investigation revealed that contemporaneous models were too tightly fitted to conditions that occurred in the years that surveying happened; when these models were extrapolated to unsurveyed years, the resulting predictions were too high. Because of this,

we ruled out the use of contemporaneous covariates for this model and selected as our final choice the climatological model that scored best in goodness of fit tests.

The final model (Figure 96, Figure 97) retained two "distance to isobath" covariates, fitting linear relationships that established the basic pattern of density around the Gulf of Maine. It fitted an inverse linear relationship to the dynamic SST_MUR covariate, indicating decreasing density with increasing surface temperature, and a "hump-shaped" relationship to bottom salinity, which was correlated with conditions in the northern Gulf of Maine around the edge of the Scotian Shelf. Finally, the model retained the Era covariate as a very strong effect; the estimate for the 2010-2018 era was -2.8955 on the scale of the link function, indicating that density for the 2010-2018 era would be <10% of the density of the 2003-2009 era, when all other covariates were held equal.

The model predicted highest densities in the basins around the Scotian Shelf, where right whales were historically known to feed (Kenney et al. 2001), and moderate density around the Gulf of Maine roughly along the 125 m isobath. As with the Spring model, because climatological covariates were used, the Era covariate itself was the only covariate that varied between the eras, so the density pattern predicted for the 2010-2018 era was the same as the 2003-2009 era, just uniformly lower by approximately 91% (Figure 98). This result is consistent with the view that in the 2010-2018 era right whales continued to briefly visit feeding hotspots of the 2003-2009 era but did not linger; instead whales moved on to other areas having better foraging opportunities (e.g. the Gulf of St. Lawrence).

The final summarized abundance for the 2003-2009 period was 715 right whales, which is clearly an overestimate. Although this could be a matter of statistical uncertainty, which was underestimated (see section 5.3.1.7.), we identified three other possible sources for the problem. First, temporal replication in surveying was relatively poor and it is possible that one of the surveys came across an unusual aggregation of right whales by chance. In particular, the NEFSC shipboard survey HB 07-09, mentioned above, sighted a large number of right whales in a relatively unusual location. These sightings appeared to influence the model to predict high density around the entire edge of the Scotian Shelf, rather than in just the historically high areas (e.g. GMB and Roseway Basin). Second, many of the sightings at Roseway Basin came from an NEFSC aerial survey which had an unusually strong (i.e. $g_{perception}(0) << 1$) perception bias correction factor (see section 5.2.2.2.1). Because this was a Canadian basin where long dives had been observed, the availability bias correction factor was also strong (i.e. $g_{availability}(0) << 1$). Together, these corrections very strongly inflated the abundance of right whales in the vicinity of Roseway Basin. Finally, the model lacked an effective prey covariate, which would likely help distinguish regions of excellent feeding habitat from regions of lower prey availability.

Given this overestimate and the various data limitations and modeling compromises made in this model, we urge caution in applying the results. In future updates, we plan to revisit how to best model this region (see section 6.3). However, given the critically endangered status of the North Atlantic right whale, our overestimate of density may be viewed as precautionary when used to assess possible human impacts to the species, such as when permitting activities under the U.S. MMPA. In any case, the overestimate applies to predictions for the 2003-2009 era, which are not likely to be used for management purposes (see section 6.2).

5.3.2.3.3 Winter (October-February)

We allocated the remaining months to a "Winter" season, spanning October-February, in which large numbers of right whales were sighted in the central and western Gulf of Maine, both in the systematic surveys used here (Figure 99) and in opportunistic datasets (Wikgren et al. 2014). Systematic effort covered the U.S. EEZ fairly well in the 2003-2009 era, but was much more patchy in the 2010-2018 era, when the NARWSS program no longer conducted broad-scale surveys (Figure 99). The 2010-2018 era did include some surveying in Canadian waters, mainly by the AMAPPS program. However, given the relatively sparse coverage, with no sightings near the Bay of Fundy, where right whales were known to regularly occur until at least October (Davies et al. 2019) and the NOAA Right Whale Sighting Advisory System showed them occurring through November, we excluded Canadian waters from this season and clipped the model extent to the U.S. EEZ (Figure 100).

Sufficient sightings were reported within this region to try the normal shrinkage-based model selection procedure. The temporal heterogeneity in surveying precluded the use of year as a categorical covariate; we used Era instead, as in the models above. Fewer sightings were reported as the season progressed, so, as with the Spring model, we included a bivariate interaction between day of year and the distance to Cape Cod Bay. To address spatial patterns not captured by this or the other covariates, we included a bivariate smooth of spatial location.

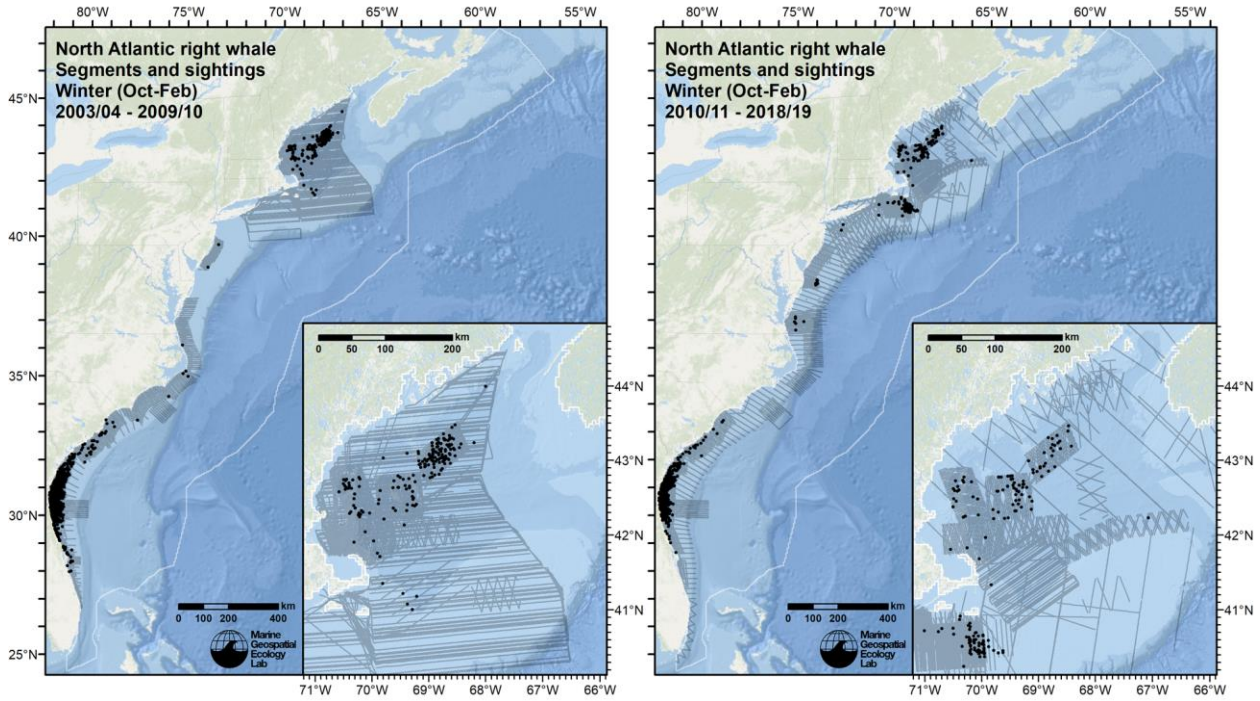


Figure 99. Survey segments and right whale sightings during the "Winter" (October-February) season defined for modeling north of Nantucket Shoals. Figure 100 shows the segments and sightings used to fit the model, which spanned both eras.

The final model (Figure 101) retained seven covariates (Figure 102), including the bivariate smooth of spatial location and the interaction between day of year and distance to Cape Cod Bay, but discarded the Era covariate. However, models fitted to monthly contemporaneous covariates substantially outperformed those fitted to climatological covariates, explaining 5% more deviance. (They also outperformed models fitted to daily contemporaneous covariates, by a lesser degree.) The model appeared to use the spatial location covariate to establish the overall pattern of density in the central and western Gulf of Maine, and the DayOfSeason:DistToCCB covariate to boost density far from Cape Cod Bay at the beginning of the season and reduce it at the end, by which time whales presumably have moved out of this modeled region, e.g. into Cape Cod Bay or south of Nantucket Shoals. The model fitted "hump-shaped" relationships to surface and bottom salinity, indicating highest density at moderate salinity values, and to current speed, indicating highest density at relatively low speeds. The model fitted an inverse linear relationship with SST, indicating decreasing density with increasing temperature.

Mean seasonal abundance predicted for the 2003-2009 era was about twice that of the 2010-2018 era (Figure 103), even though the final model included neither a year covariate nor an era covariate. This striking difference resulted completely from relationships fitted to contemporaneous oceanographic covariates. We have not attempted to determine which relationships specifically drove this result, but Gulf of Maine surface temperatures have been warming since at least 1980, with a particularly strong trend between 2003 and at least 2013 (Pershing et al. 2015), and our model predicted a strong inverse correlation between density and SST (Figure 102). Although the differences between eras was driven by oceanography—in contrast to the Spring and Summer models above, in which the Era covariate provided the only source of change—the mean predicted density pattern was similar between the two eras (Figure 103), just lower in the latter era. Finally, despite not retaining the Era covariate, the model predicted a marked decline when the second era started in 2010/11 (Figure 104), around the time that abundance in Cape Cod Bay increased substantially in these months. This result provides further evidence that right whale distributions changed markedly around the time the population entered decline (as discussed in section 5.3.1.1).

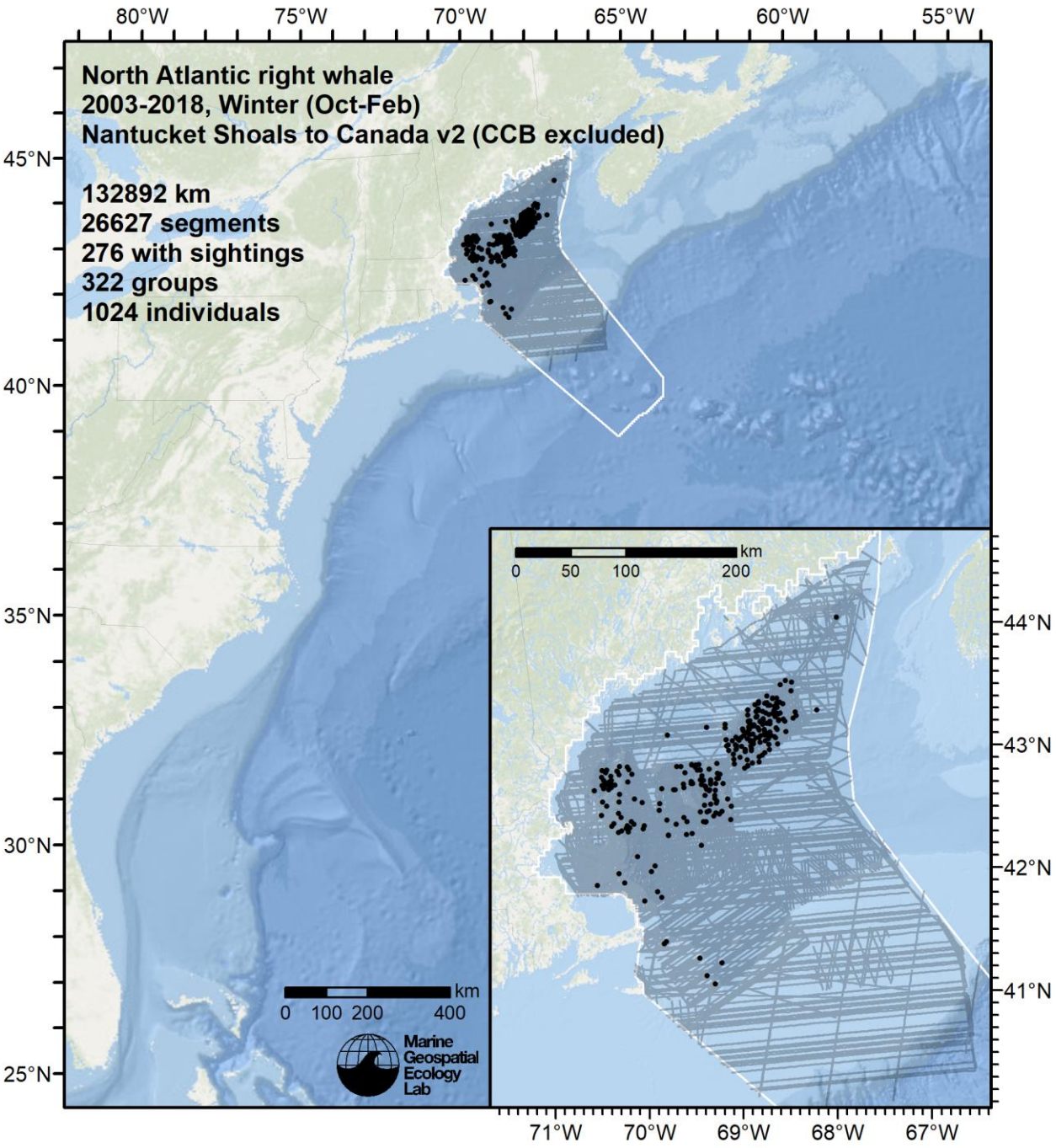


Figure 100. Segments and sightings used to fit the Winter model for the North of Nantucket Shoals region.

```
Family: Tweedie(p=1.201)
Link function: log
Formula:
IndividualsCorrected ~ offset(log(SegmentArea)) + s(x, y, bs = "ts", k = 50) + s(I(DistTo125m/1000), bs = "ts") +
    te(I(DistToCCB/1000), DayOfSeason, bs = "ts") + s(SSS_HYCOM_Monthly, bs = "ts") + s(BotS_HYCOM_Monthly, bs = "ts") +
    s(UV_HYCOM_Monthly, bs = "ts") + s(SST_HYCOM_Monthly, bs = "ts")

Parametric coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  -23.958      1.092  -21.94   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                                   edf Ref.df     F  p-value
s(x,y)                         27.9449     49 2.997  < 2e-16 ***
s(I(DistTo125m/1000))           0.9627      9 0.930 0.000743 ***
te(I(DistToCCB/1000),DayOfSeason) 11.8895     24 3.358 2.20e-15 ***
s(SSS_HYCOM_Monthly)            2.7348      9 2.986 4.05e-07 ***
s(BotS_HYCOM_Monthly)           4.1762      9 3.856 2.30e-08 ***
s(UV_HYCOM_Monthly)             2.6364      9 1.090 0.007048 **
s(SST_HYCOM_Monthly)            1.1506      9 2.771 2.11e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.0407   Deviance explained = 37.3%
-REML = 1968.7  Scale est. = 8.6324    n = 26627

Method: REML   Optimizer: outer newton
full convergence after 16 iterations.
Gradient range [-3.305313e-06,5.218913e-06]
(score 1968.67 & scale 8.632449).
Hessian positive definite, eigenvalue range [0.2947686,1647.347].
Model rank =  119 / 119
```

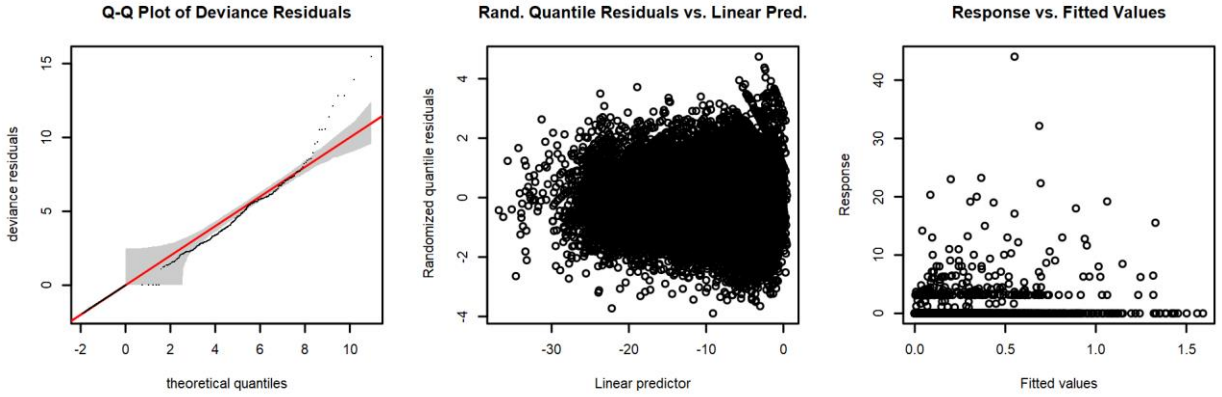


Figure 101. Statistical output for the Winter model for the North of Nantucket Shoals region.

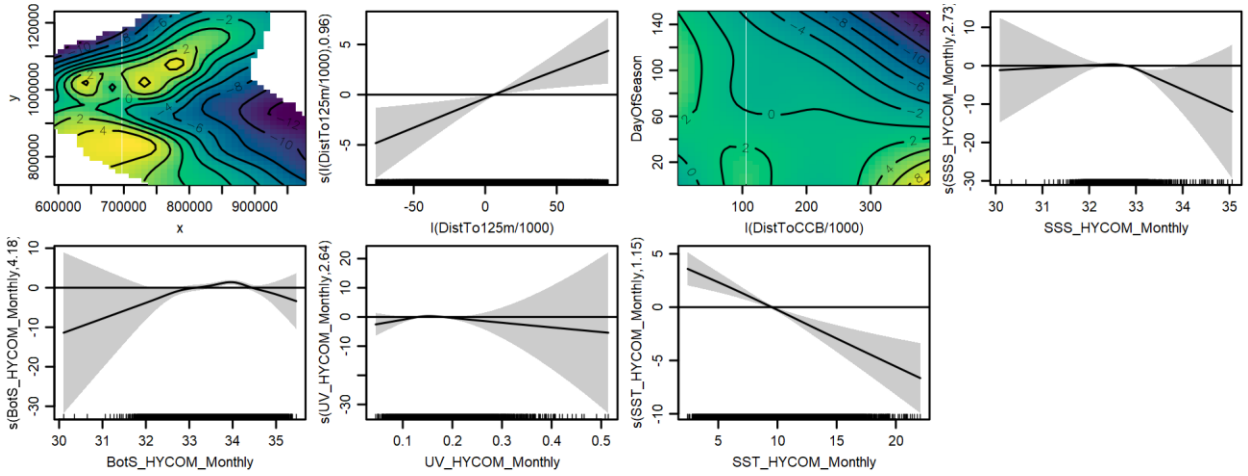


Figure 102. Functional plots for the Winter model for the North of Nantucket Shoals region.

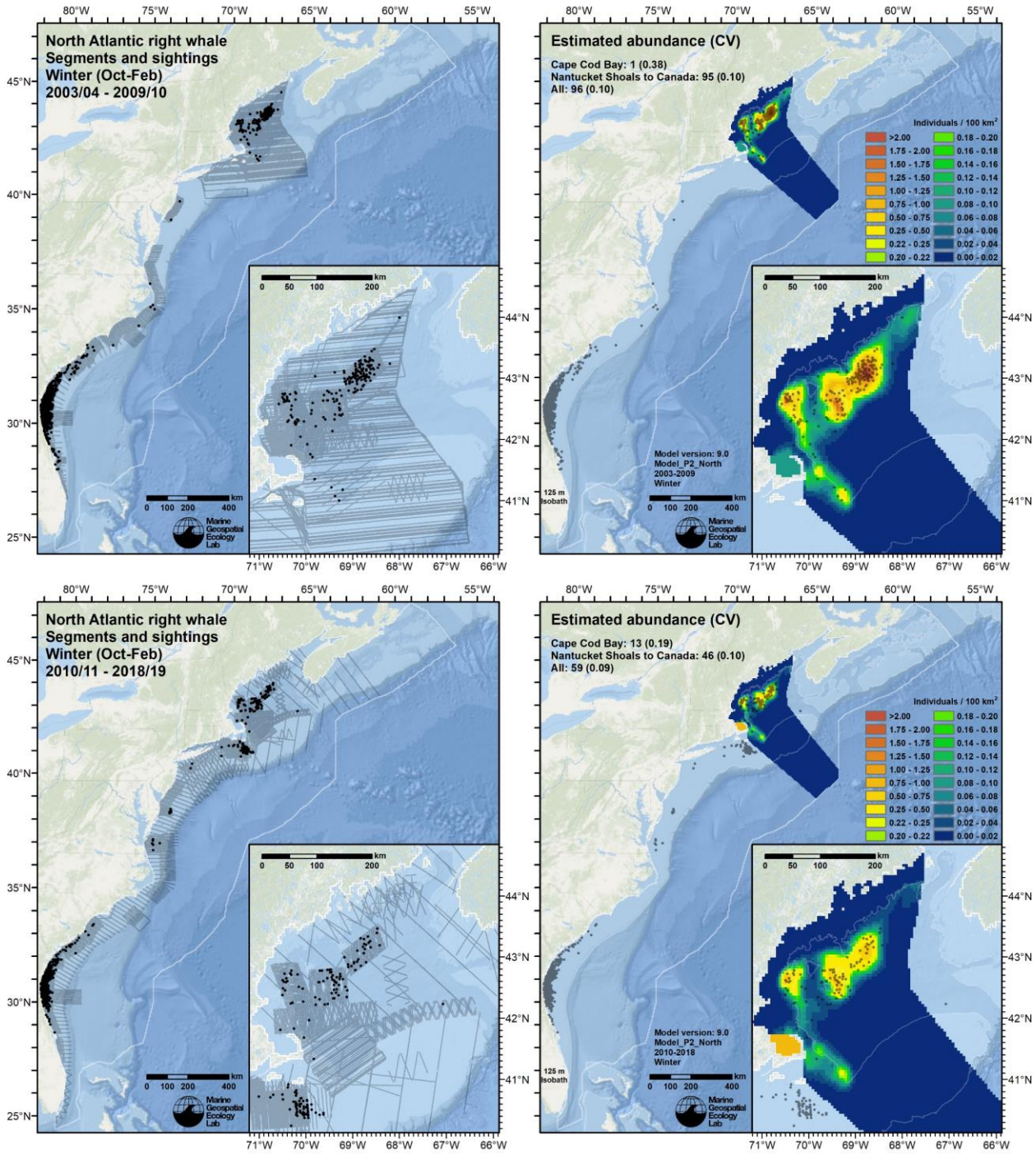


Figure 103. Survey effort and right whale sightings (left maps) and mean density predictions with sightings overlaid (right maps), for the 2003-2009 (top row) and 2010-2018 (bottom row) eras, for the Winter season (October-February). Effort and sightings are all of those that occurred during those eras during October-February. Density predictions represent the mean across all those years and months.

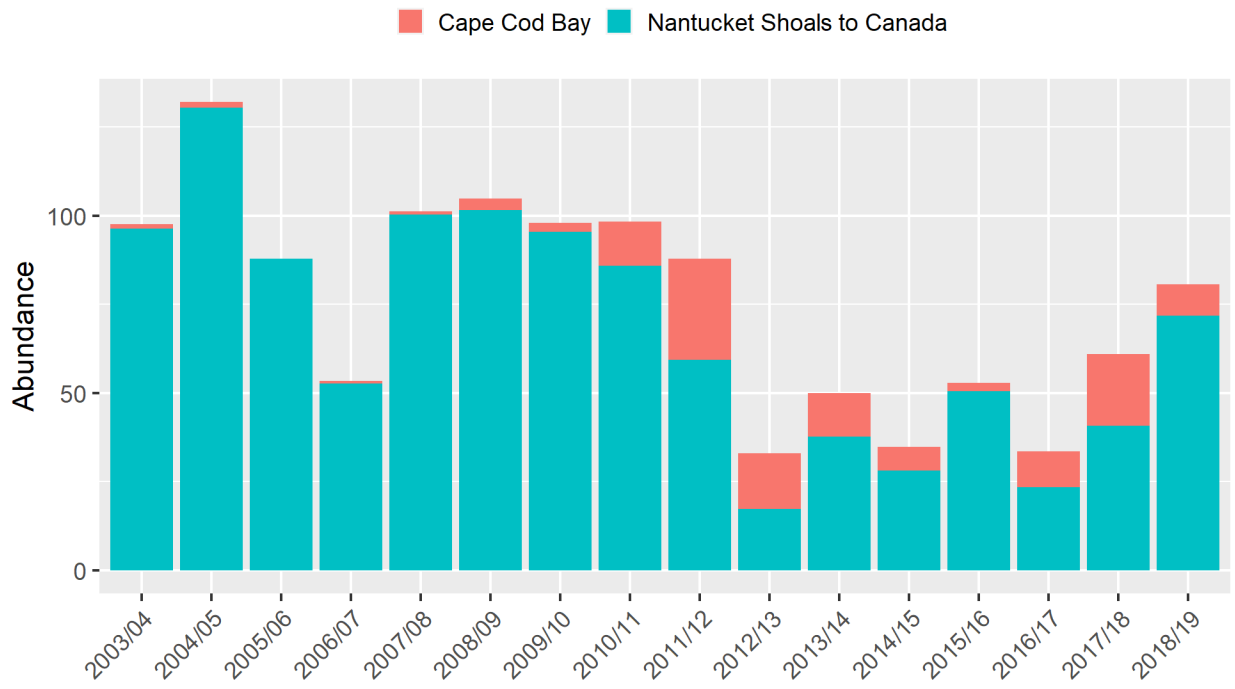


Figure 104. Mean abundance for the North of Nantucket Shoals region (blue) predicted by the seasonal model for "Winter" (October-February) and for Cape Cod Bay (red) from Ganley et al. (2019) (see section 5.3.1.2).

5.3.3. Final density surfaces

In this section, we present maps of the final density surfaces that combine predictions for all the regions. For each of the two eras, 2003-2009 (Figure 105) and 2010-2018 (Figure 106), there are 12 density maps, one for each month, that represent mean density predictions for that month during the era. We also include maps for the combined 2003-2018 period (Figure 107), for comparison. For right whale management purposes, we recommend either the 2010-2018 or 2003-2018 densities; we discuss this further in section 6.2 below. For the Navy Phase IV EIS, final density surfaces are not needed until 2022. We anticipate making at least one more update to this model by that time, therefore we do not anticipate the predictions presented here will ultimately be used for the Phase IV EIS.

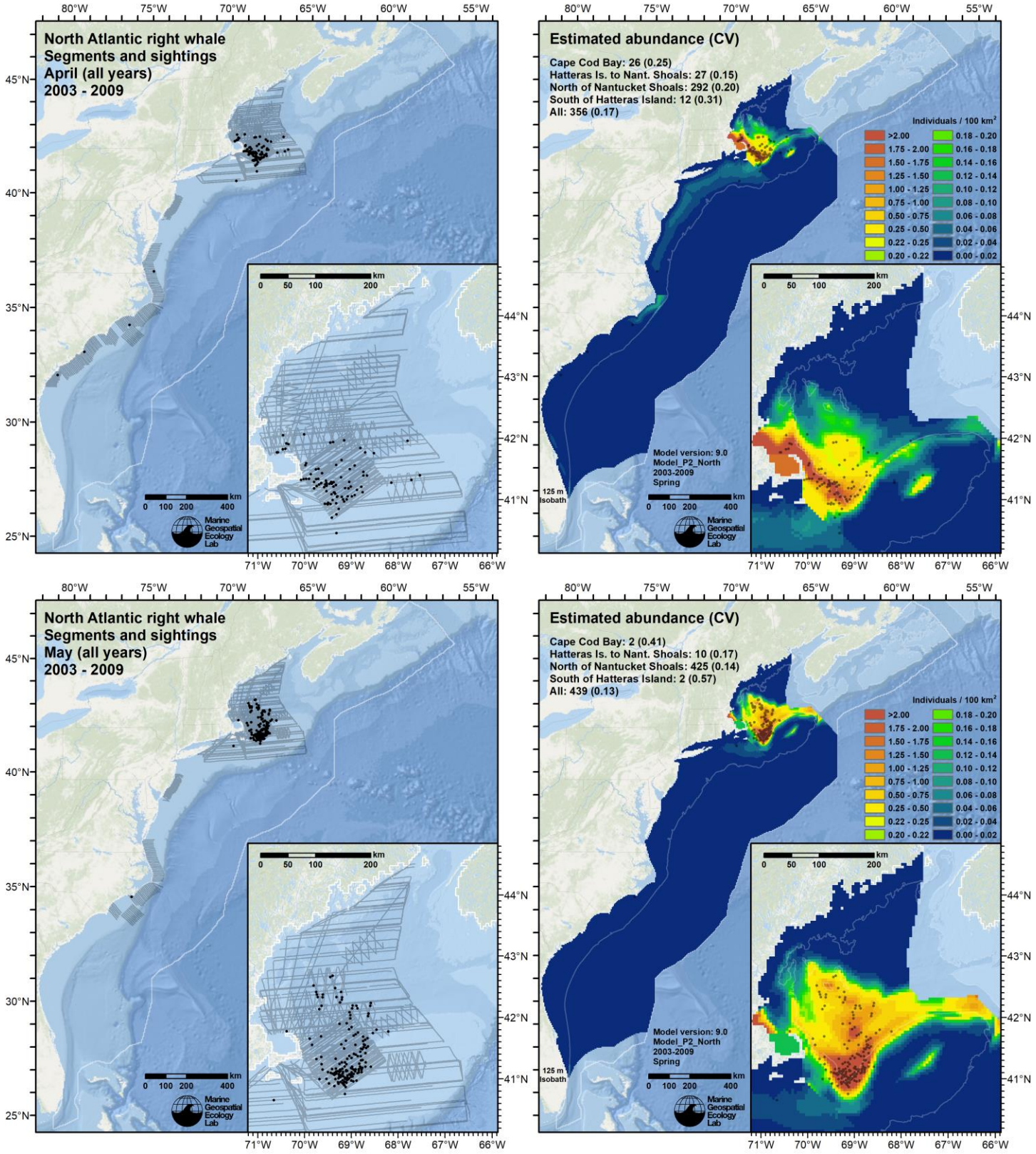


Figure 105. Monthly maps (rows) of survey effort and right whale sightings (left column) and mean density predictions with sightings overlaid (right column), for the "2003-2009" era, which spanned April 2003 through March 2010. Effort and sightings are all of those that occurred during the specified month during the era. Density predictions represent the mean for the specified month over the era.

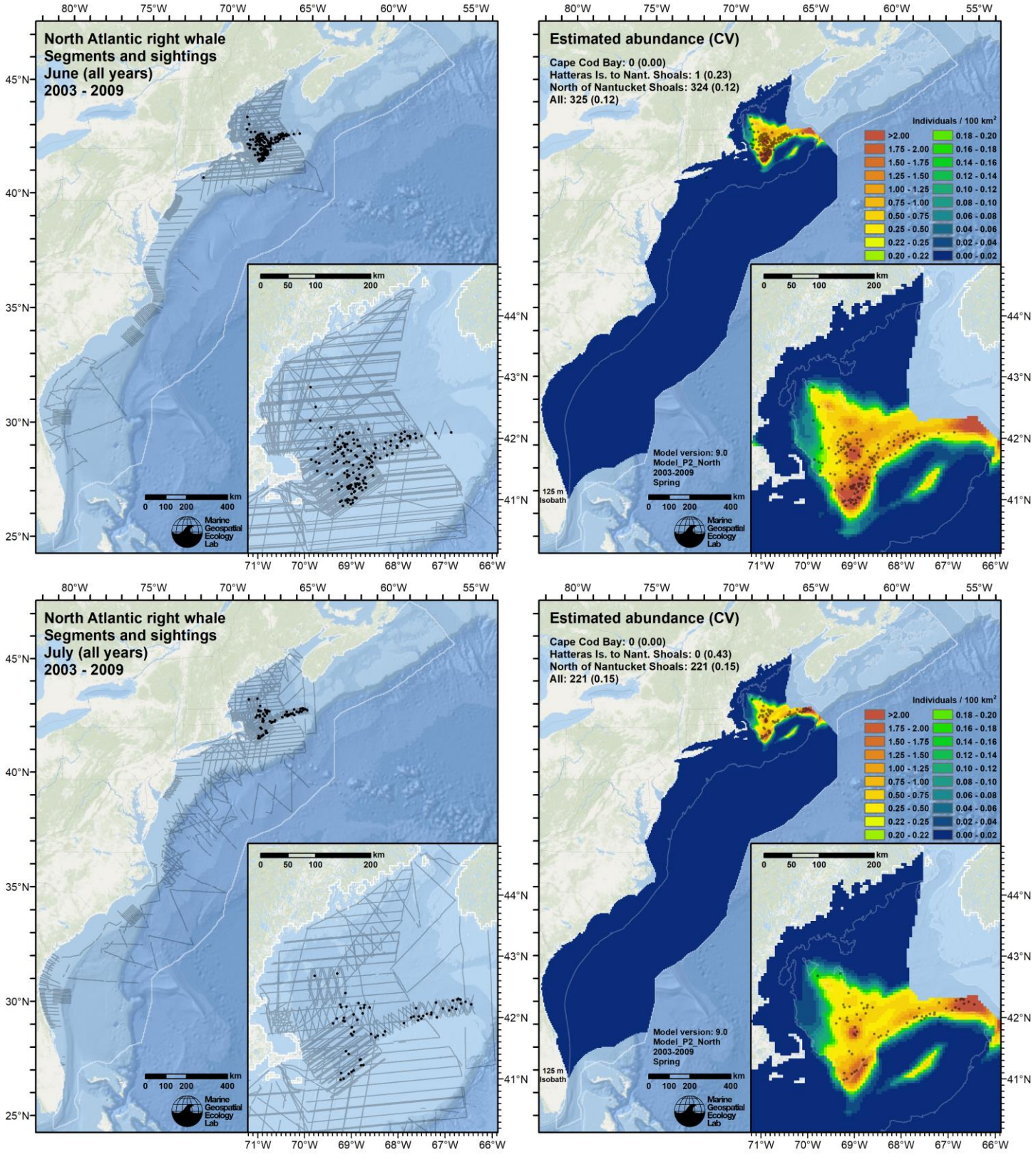


Figure 105. Continued from previous page.

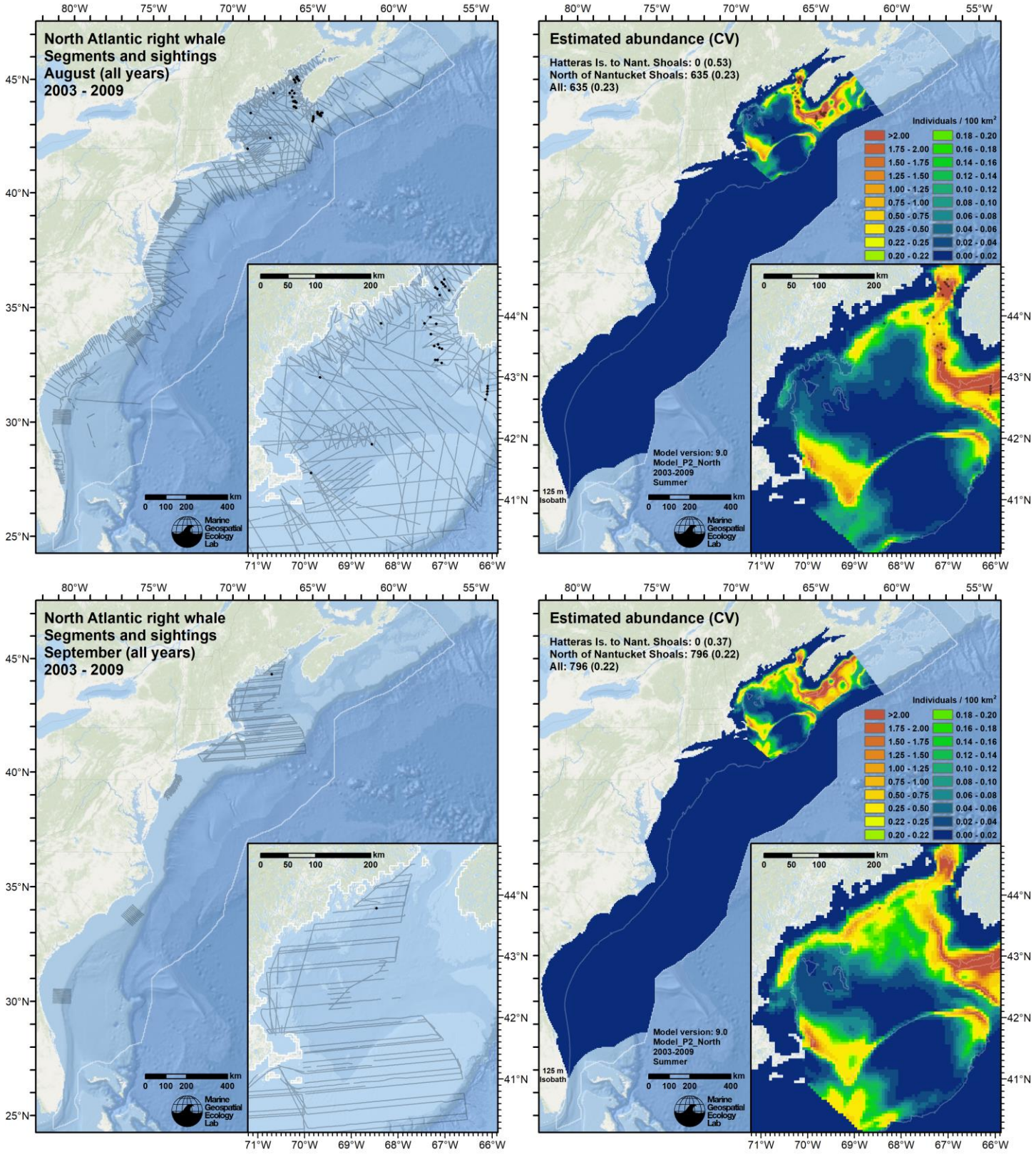


Figure 105. Continued from previous page.

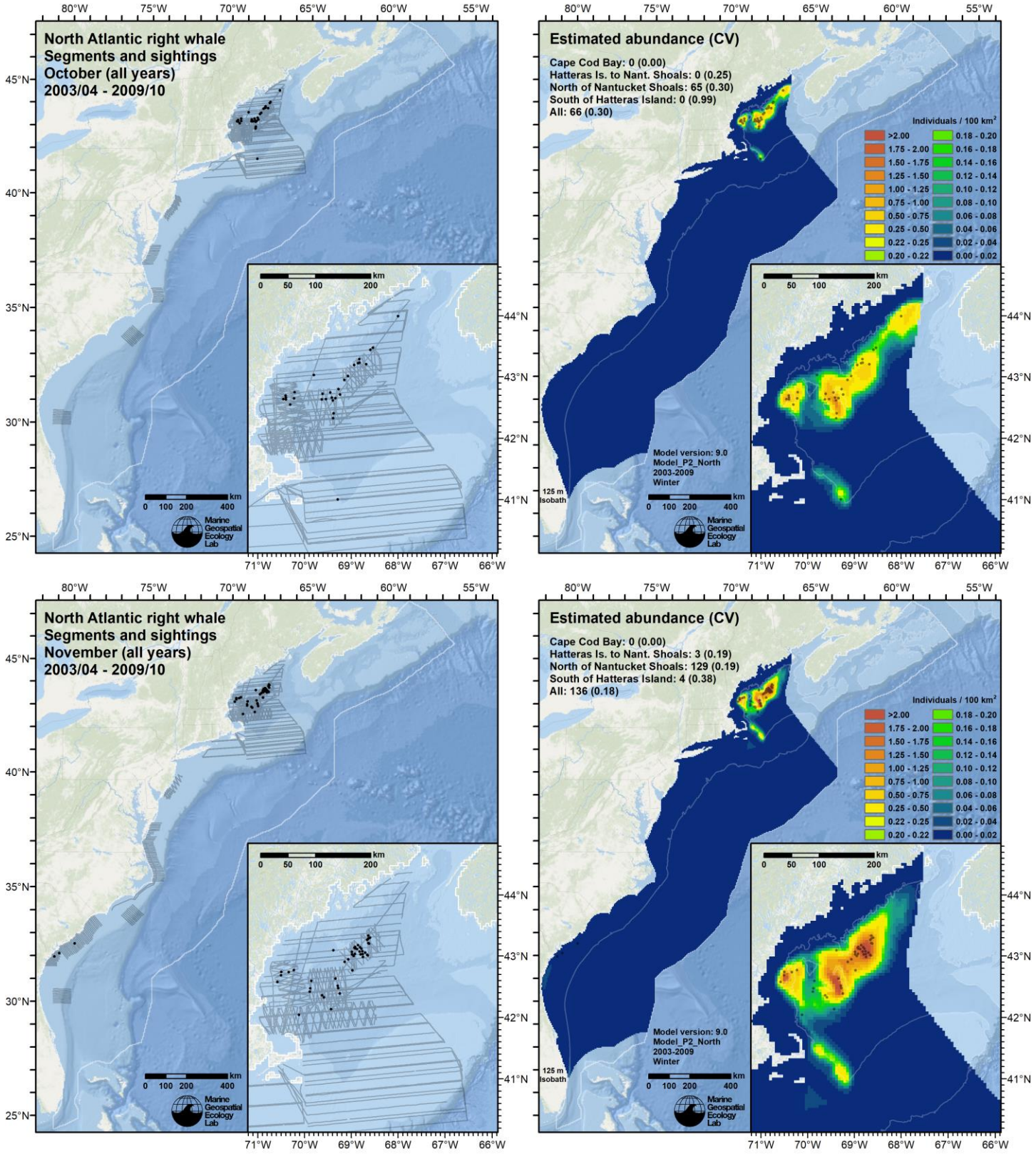


Figure 105. Continued from previous page.

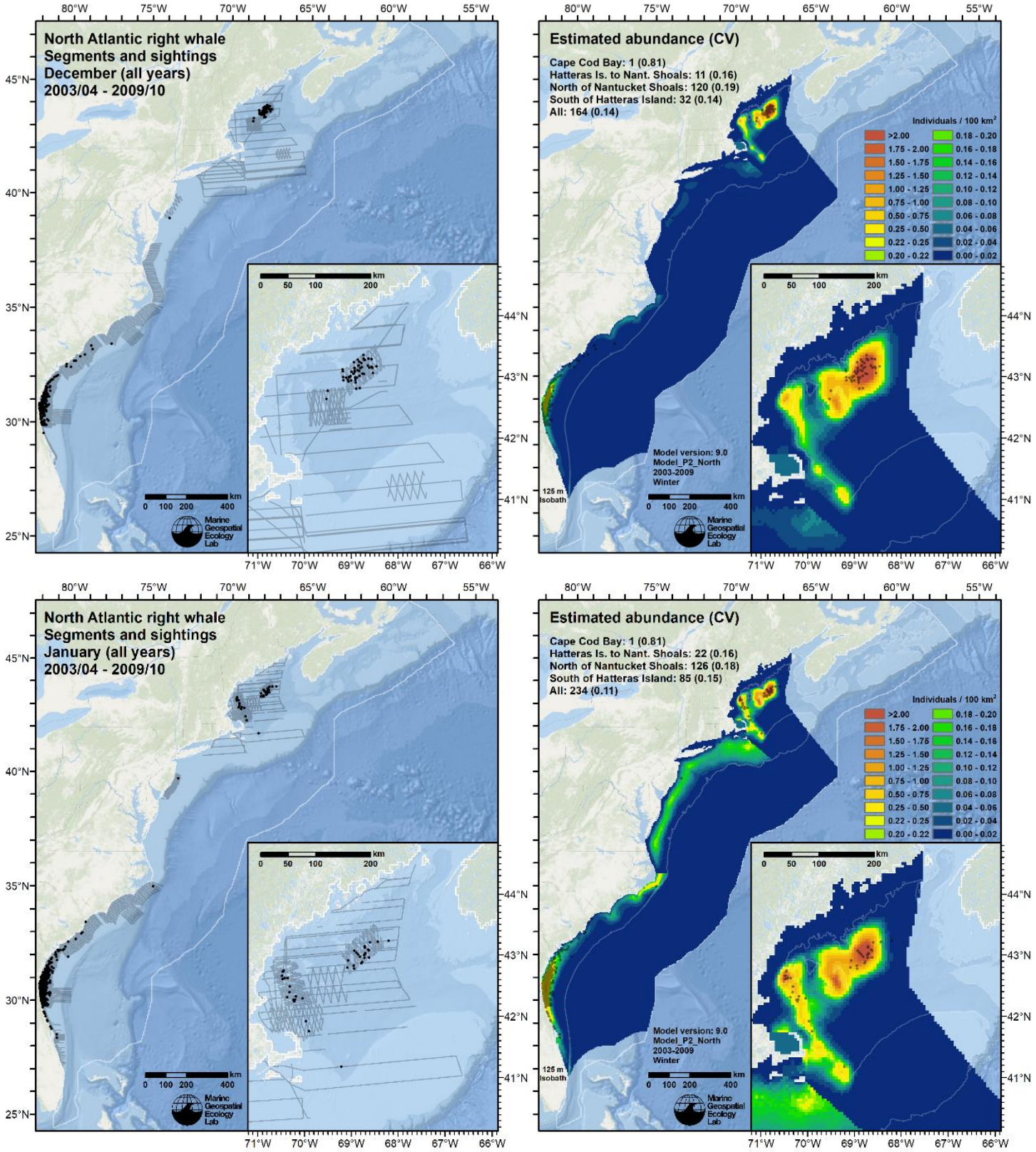


Figure 105. Continued from previous page.

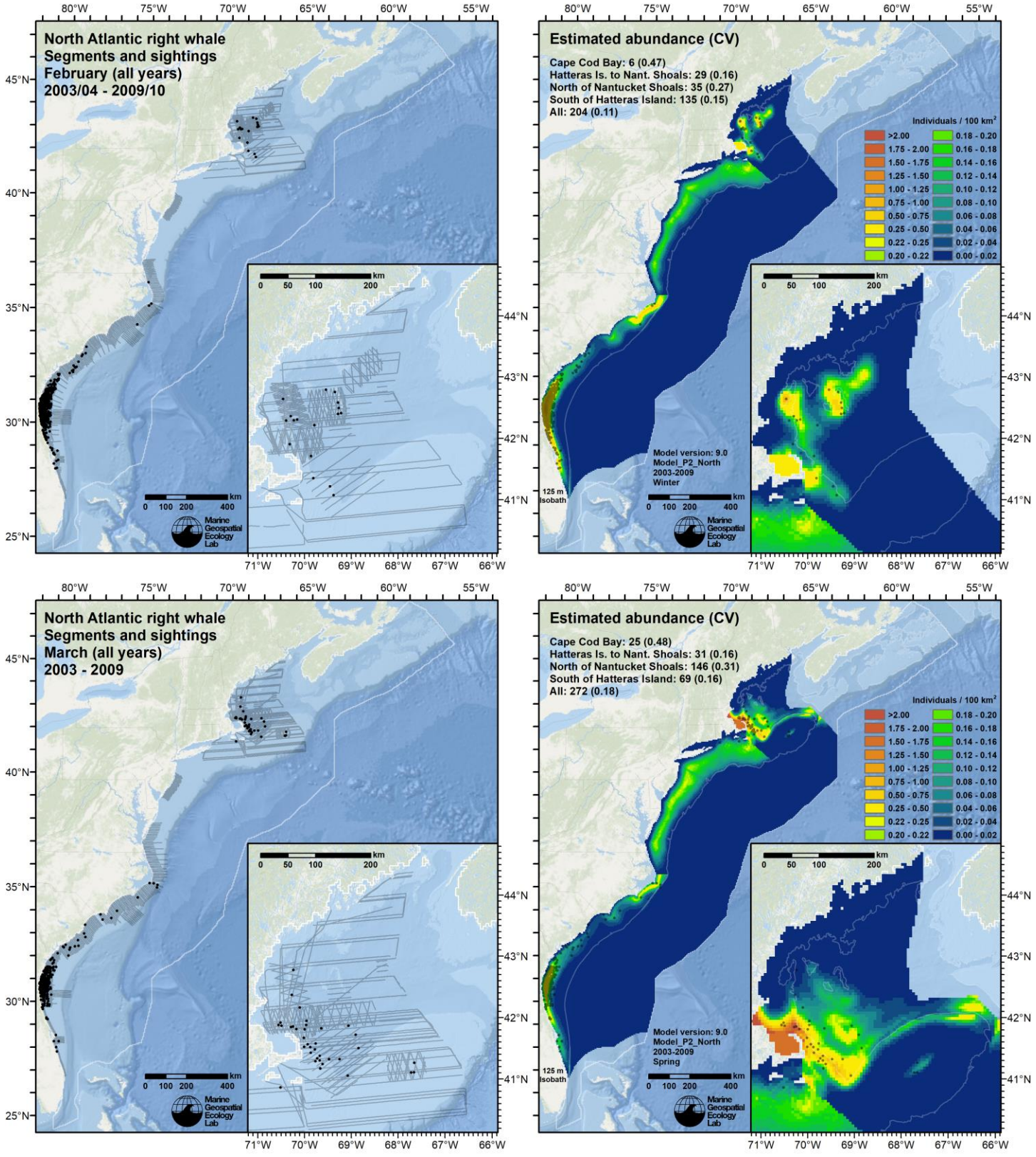


Figure 105. Continued from previous page.

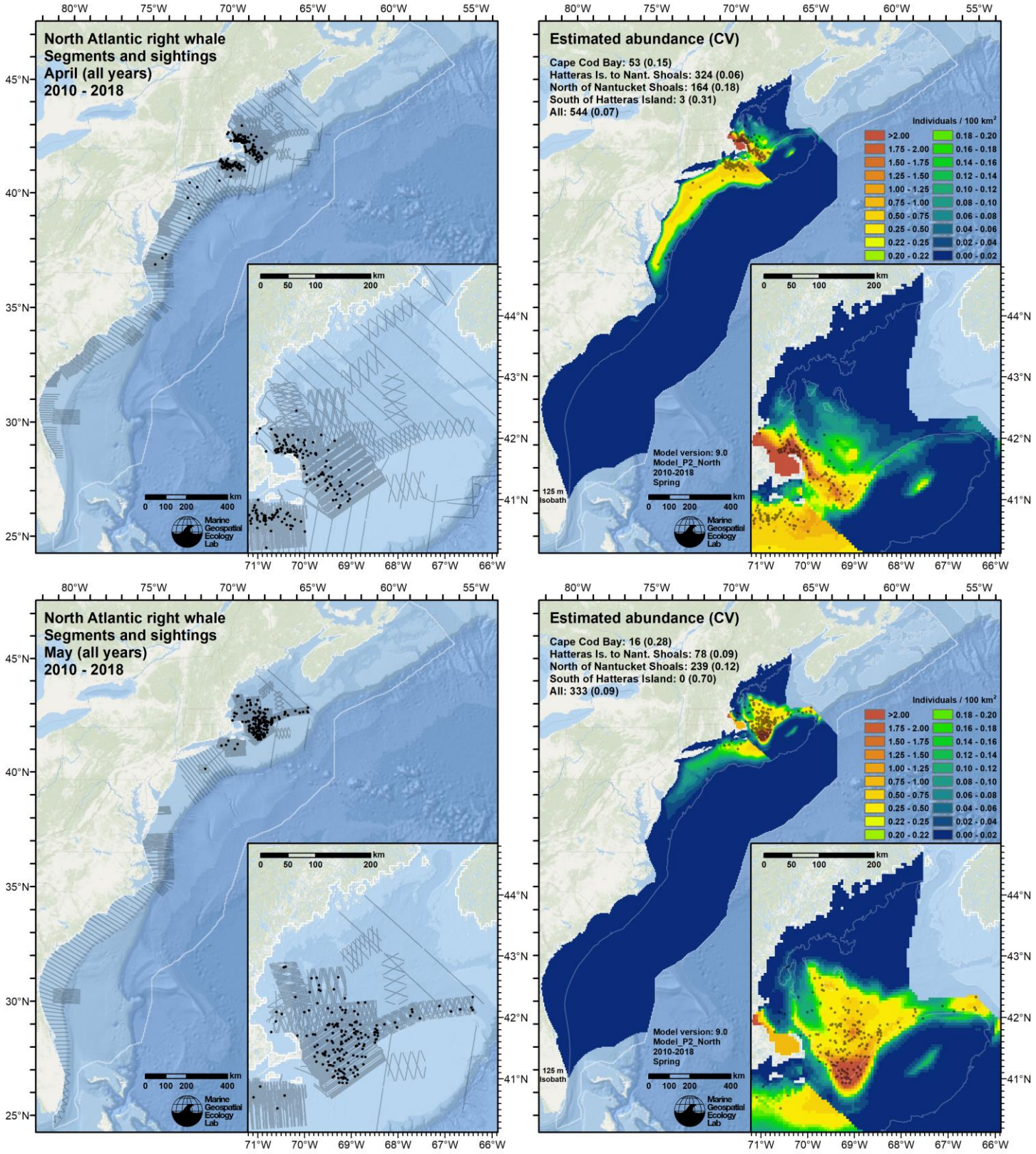


Figure 106. Monthly maps (rows) of survey effort and right whale sightings (left column) and mean density predictions with sightings overlaid (right column), for the 2010-2018 era, which spanned April 2010 through March 2019. Effort and sightings are all of those that occurred during the specified month during the era. Density predictions represent the mean for the specified month over the era.

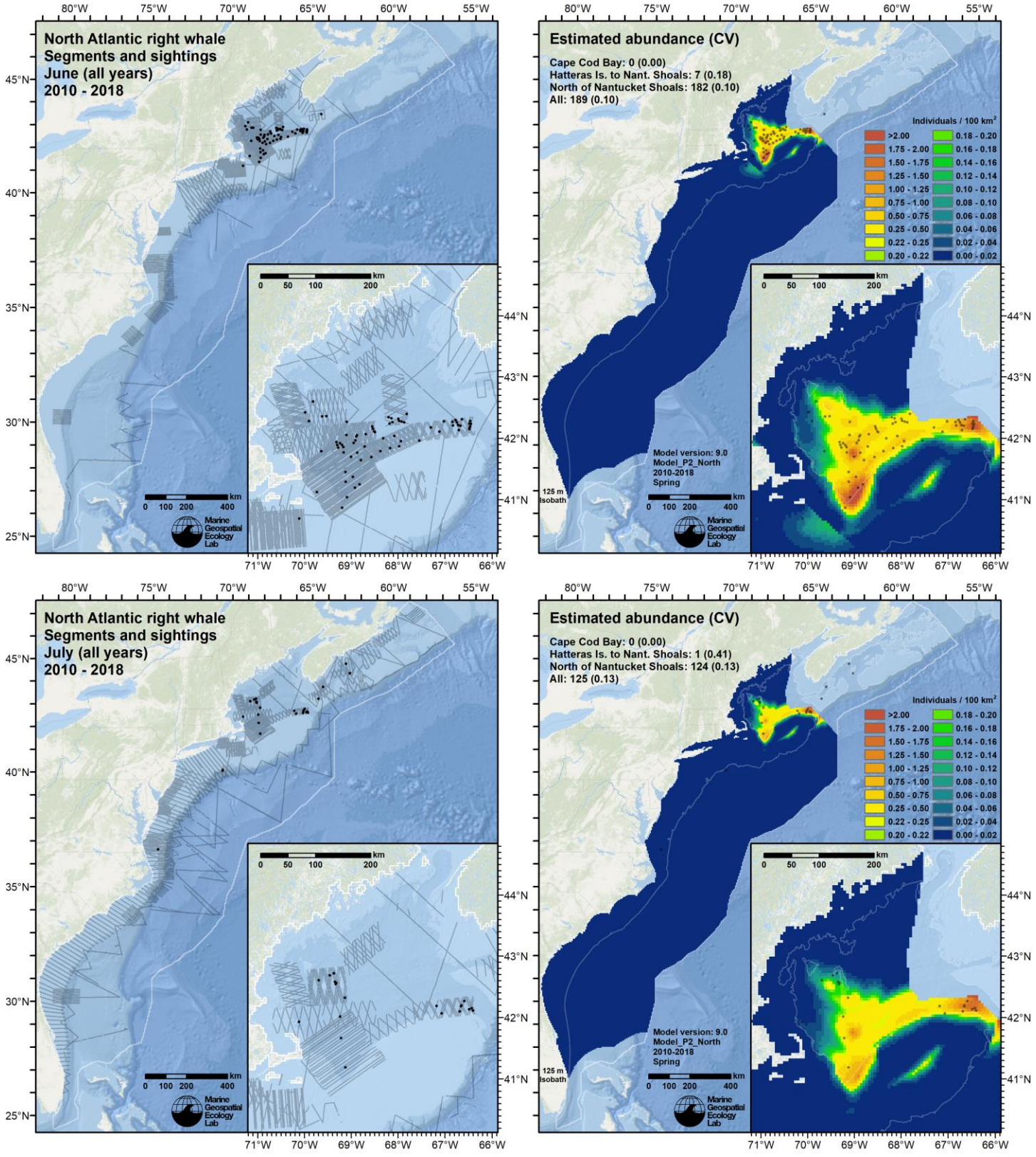


Figure 106. Continued from previous page.

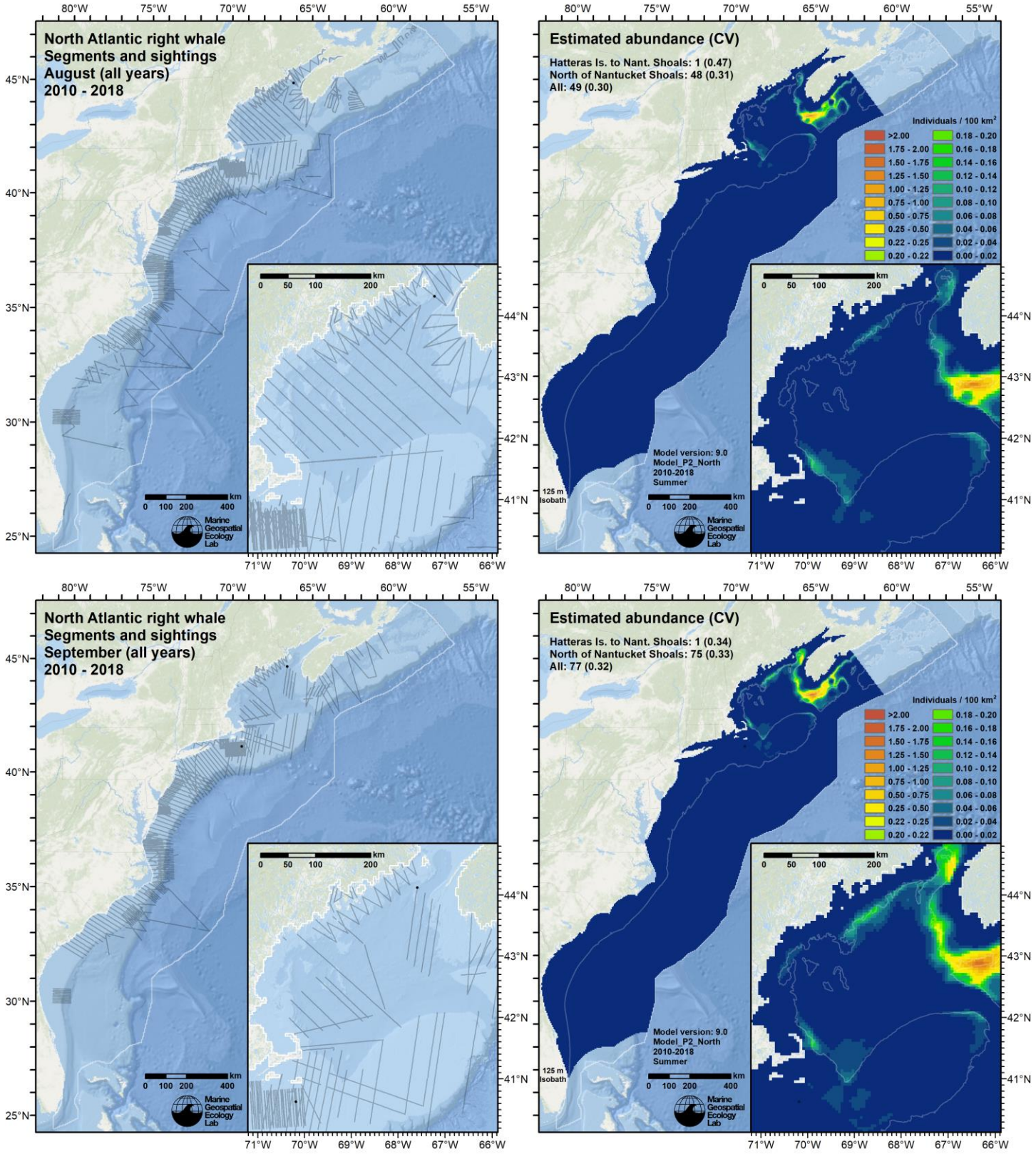


Figure 106. Continued from previous page.

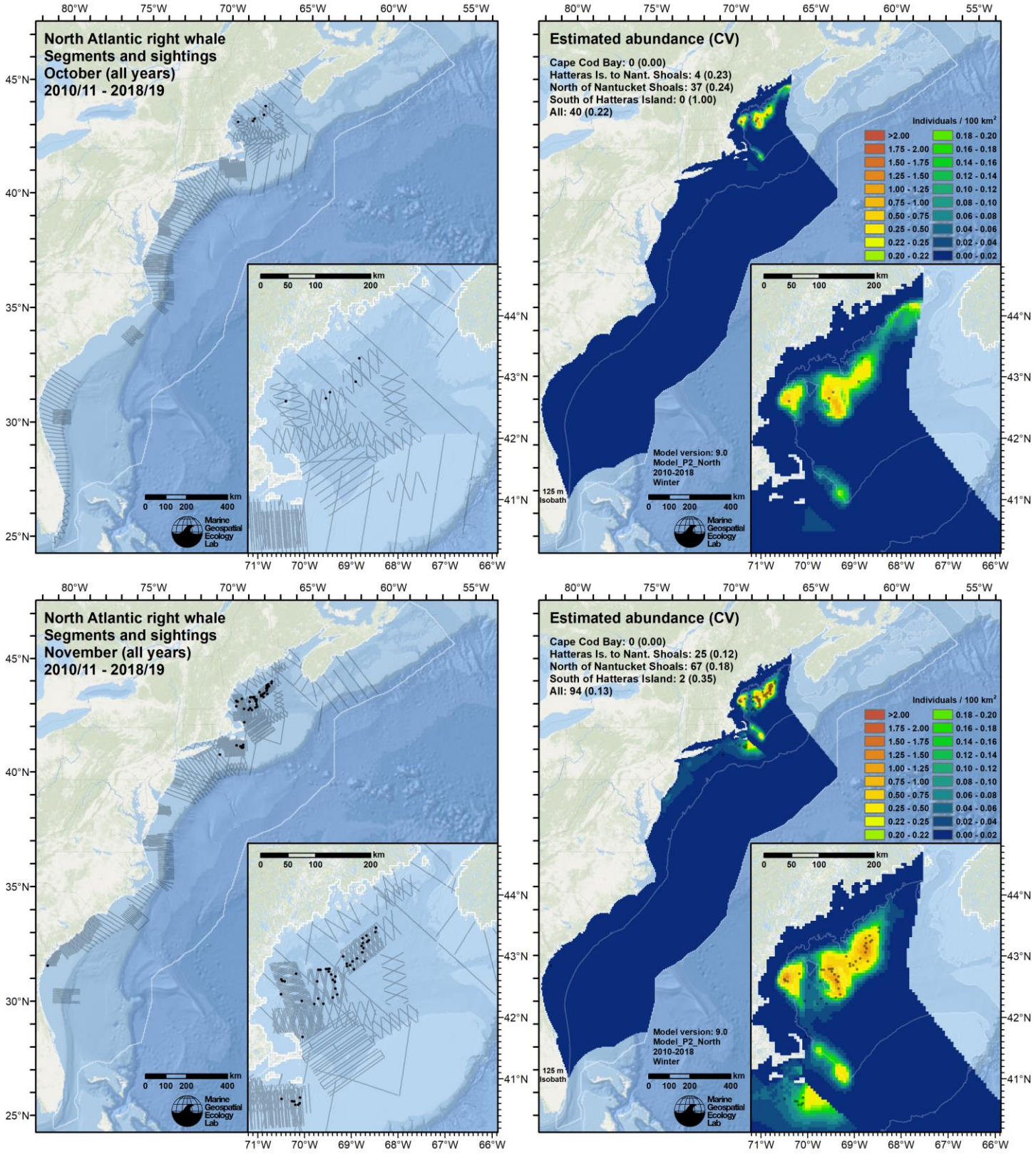


Figure 106. Continued from previous page.

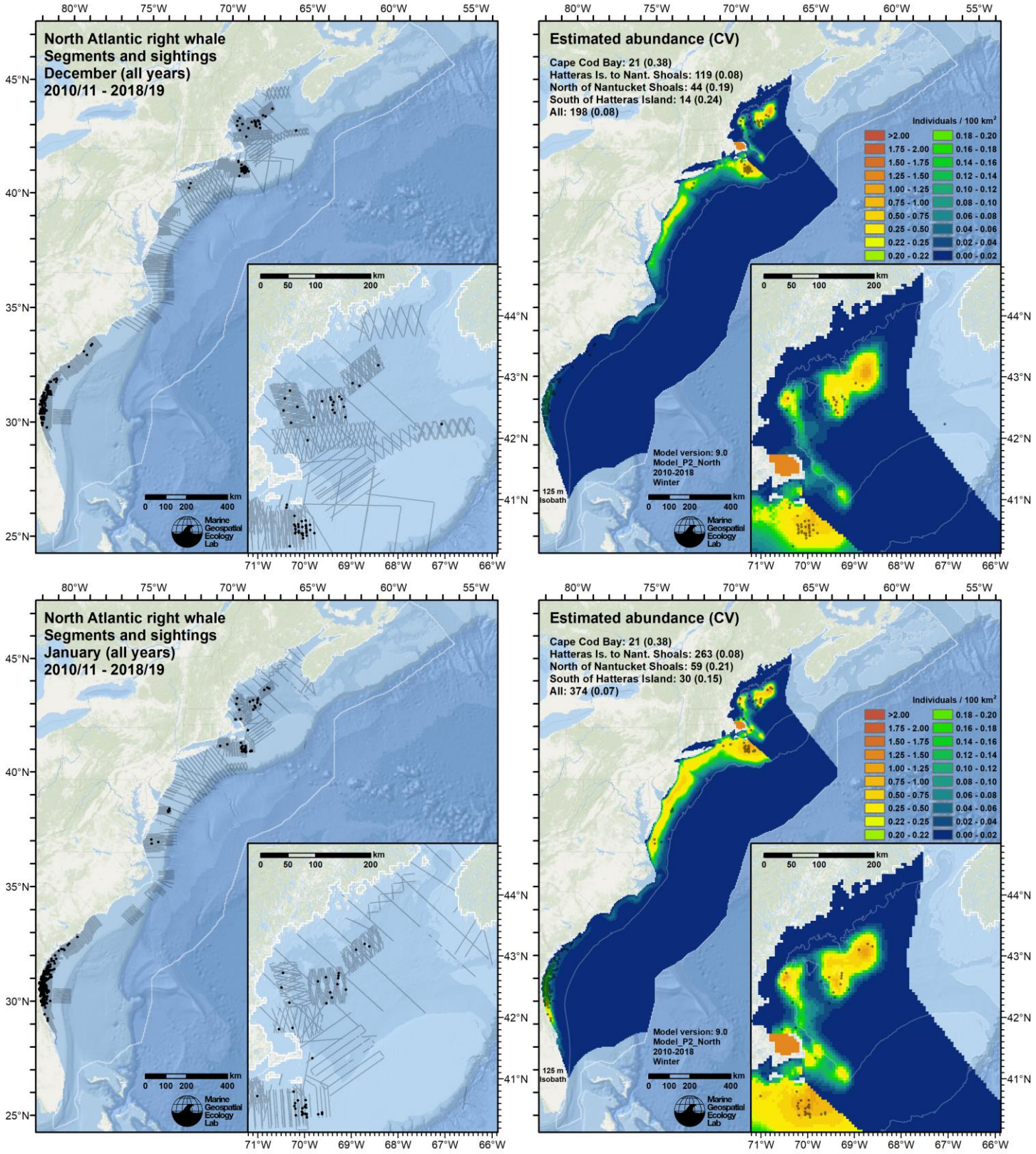


Figure 106. Continued from previous page.

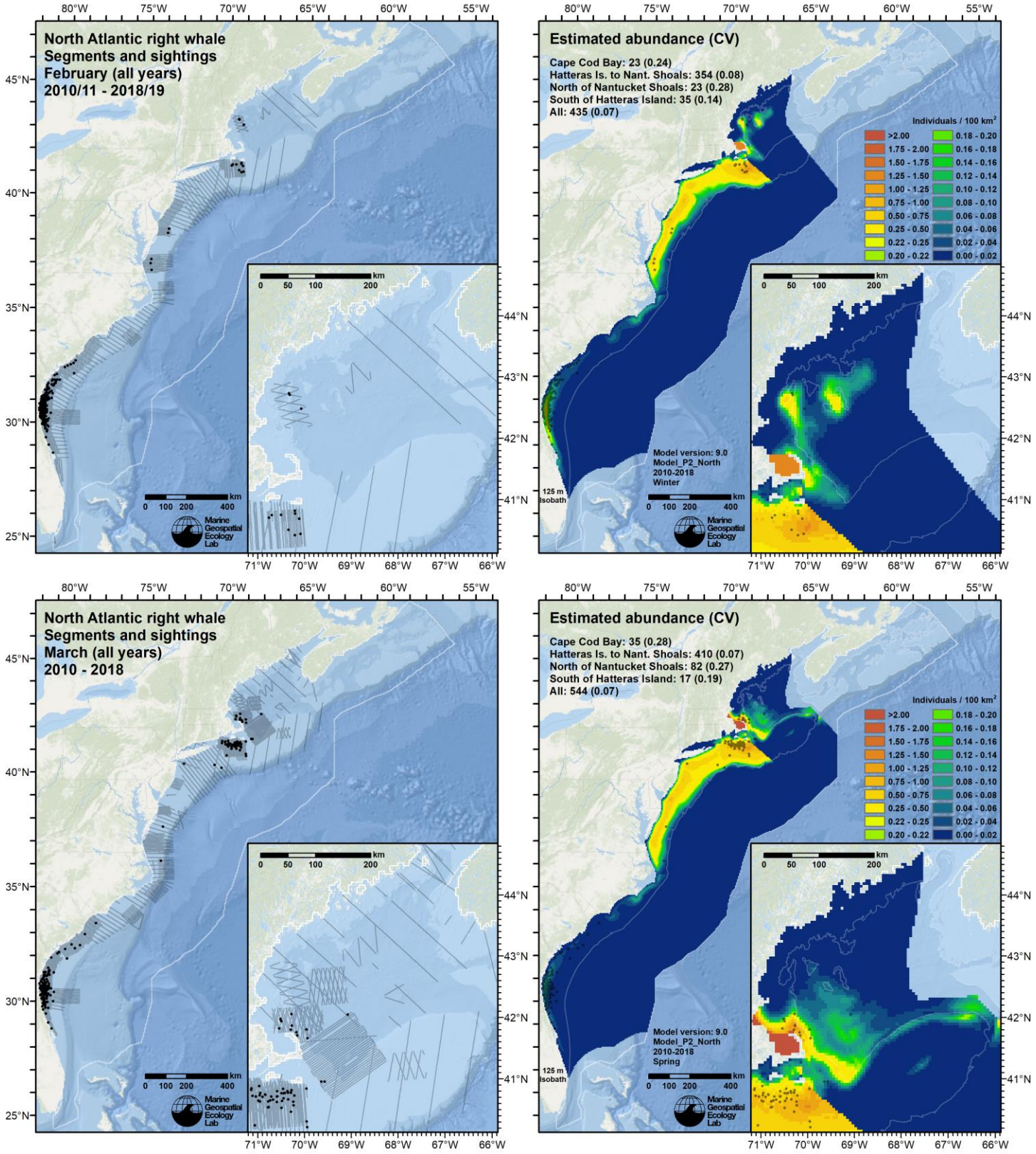


Figure 106. Continued from previous page.

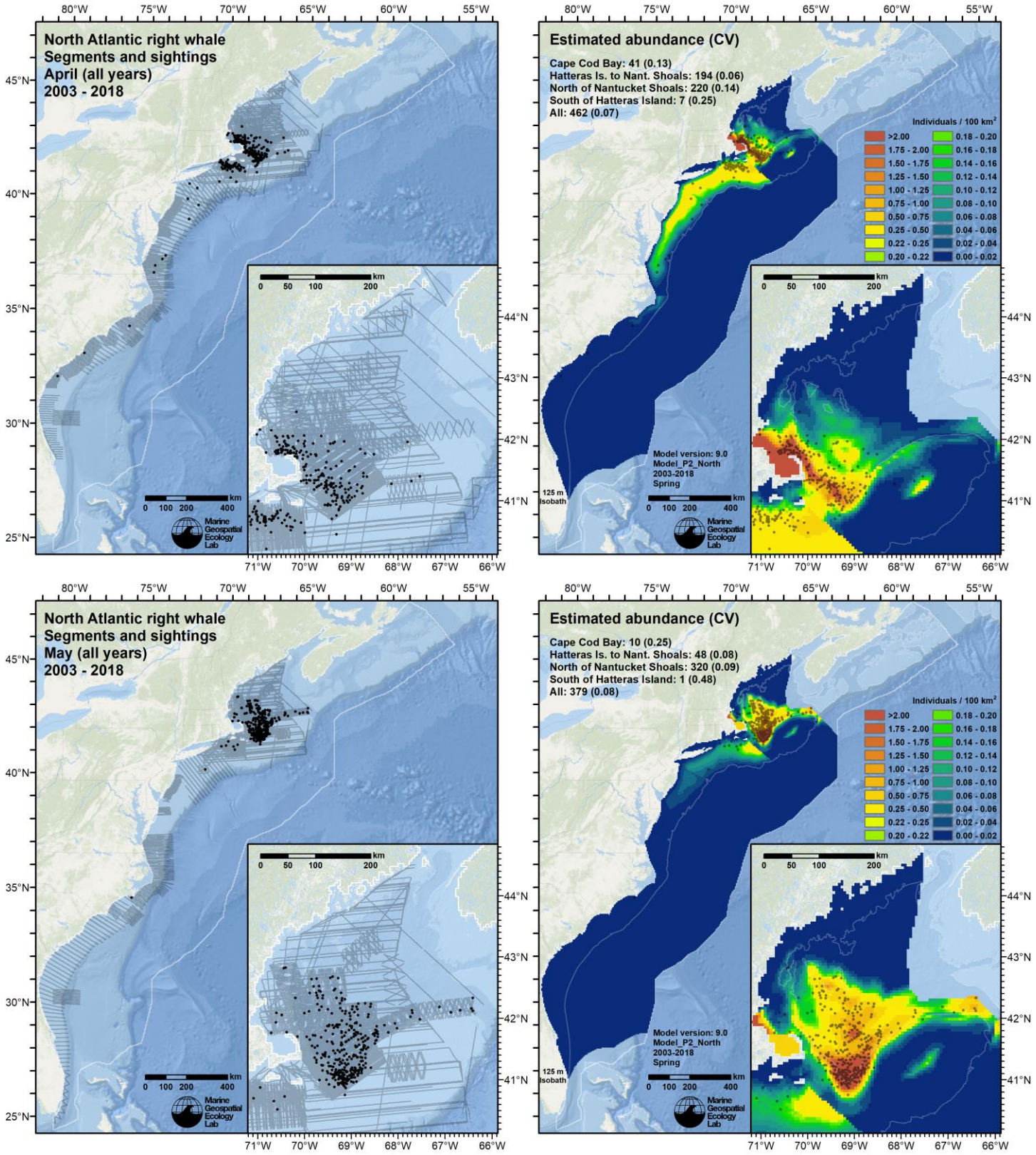


Figure 107. Monthly maps (rows) of survey effort and right whale sightings (left column) and mean density predictions with sightings overlaid (right column), for the 2003-2018 era, which spanned April 2003 through March 2019. Effort and sightings are all of those that occurred during the specified month during the era. Density predictions represent the mean for the specified month over the era.

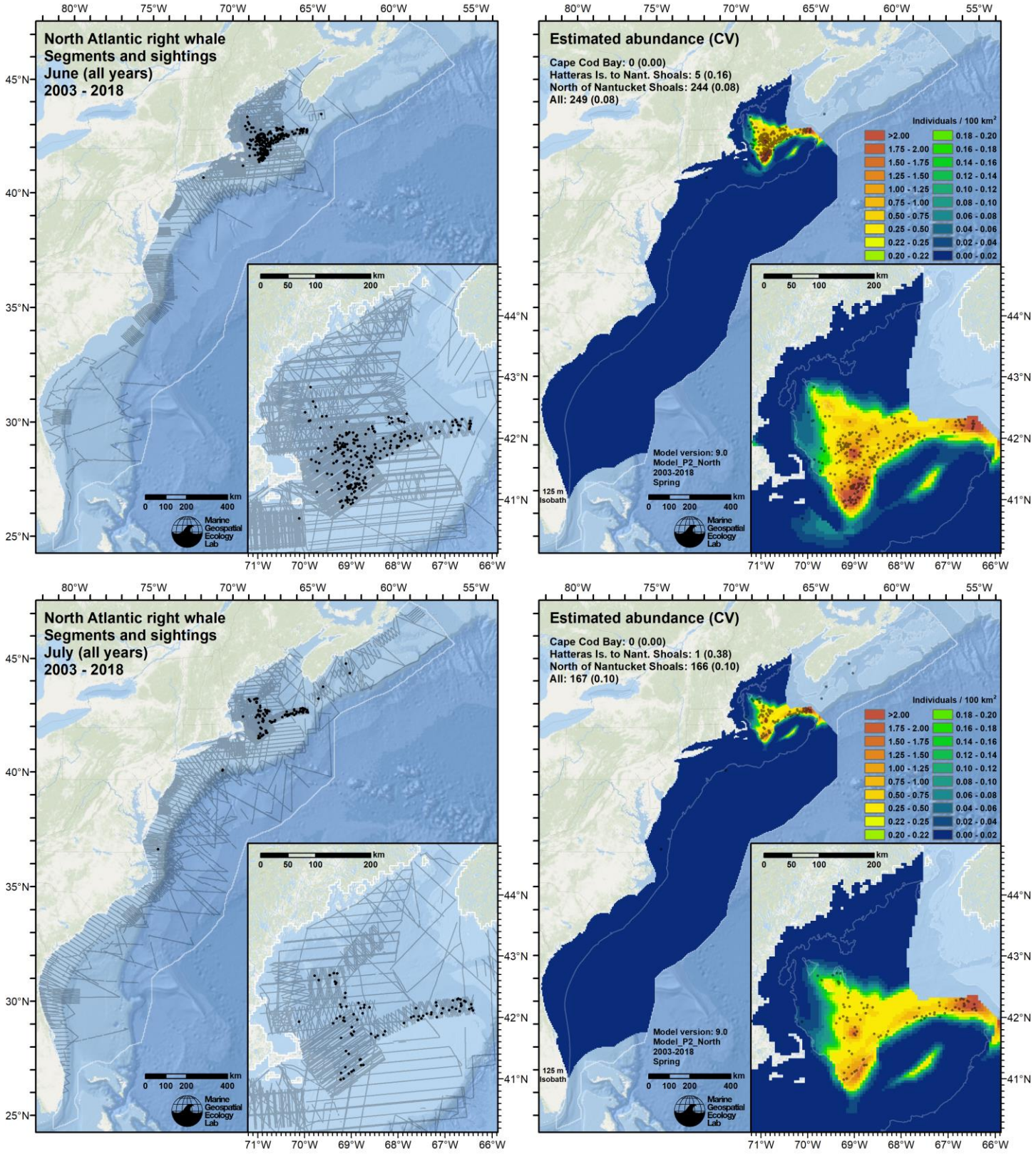


Figure 107. Continued from previous page.

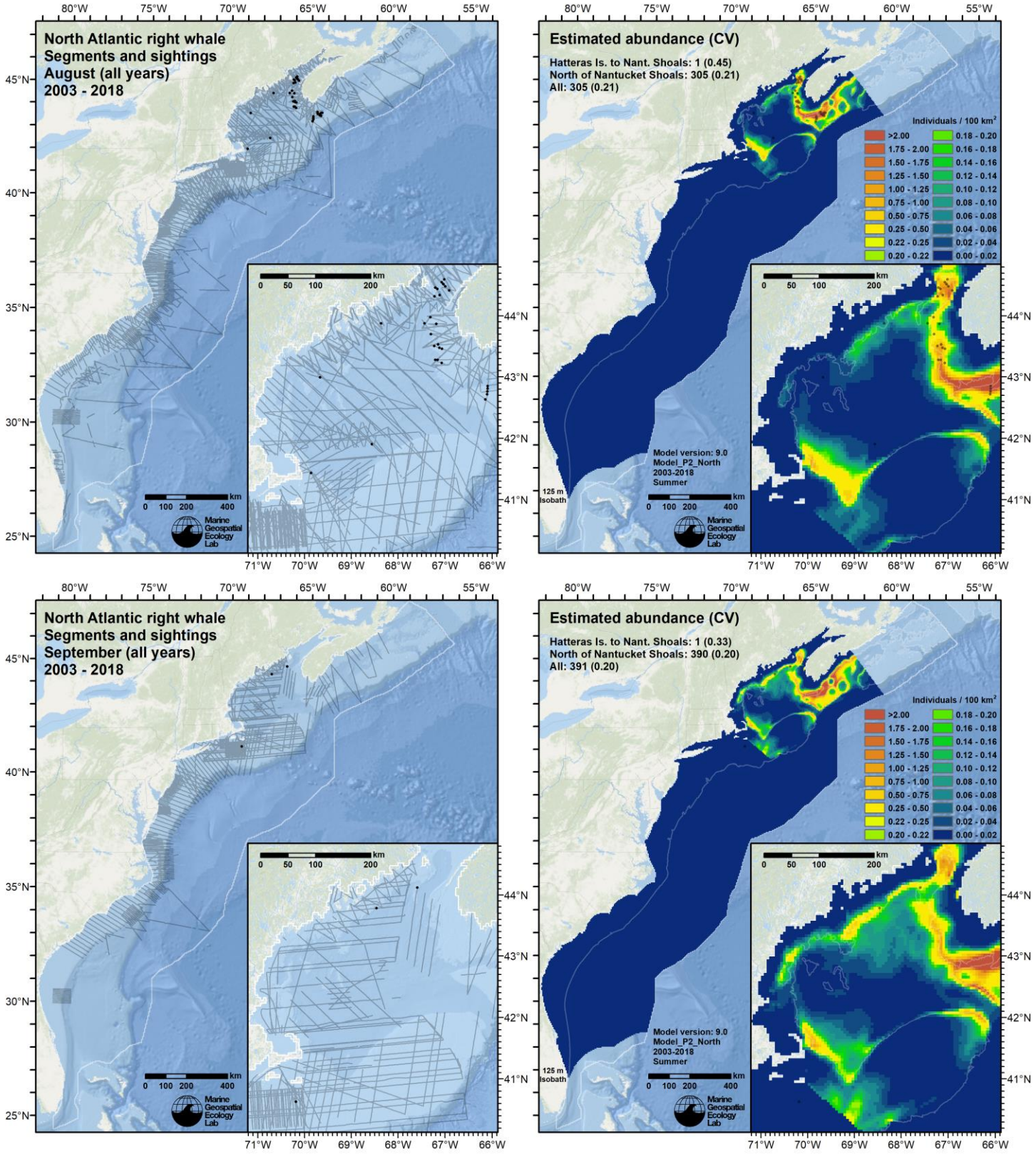


Figure 107. Continued from previous page.

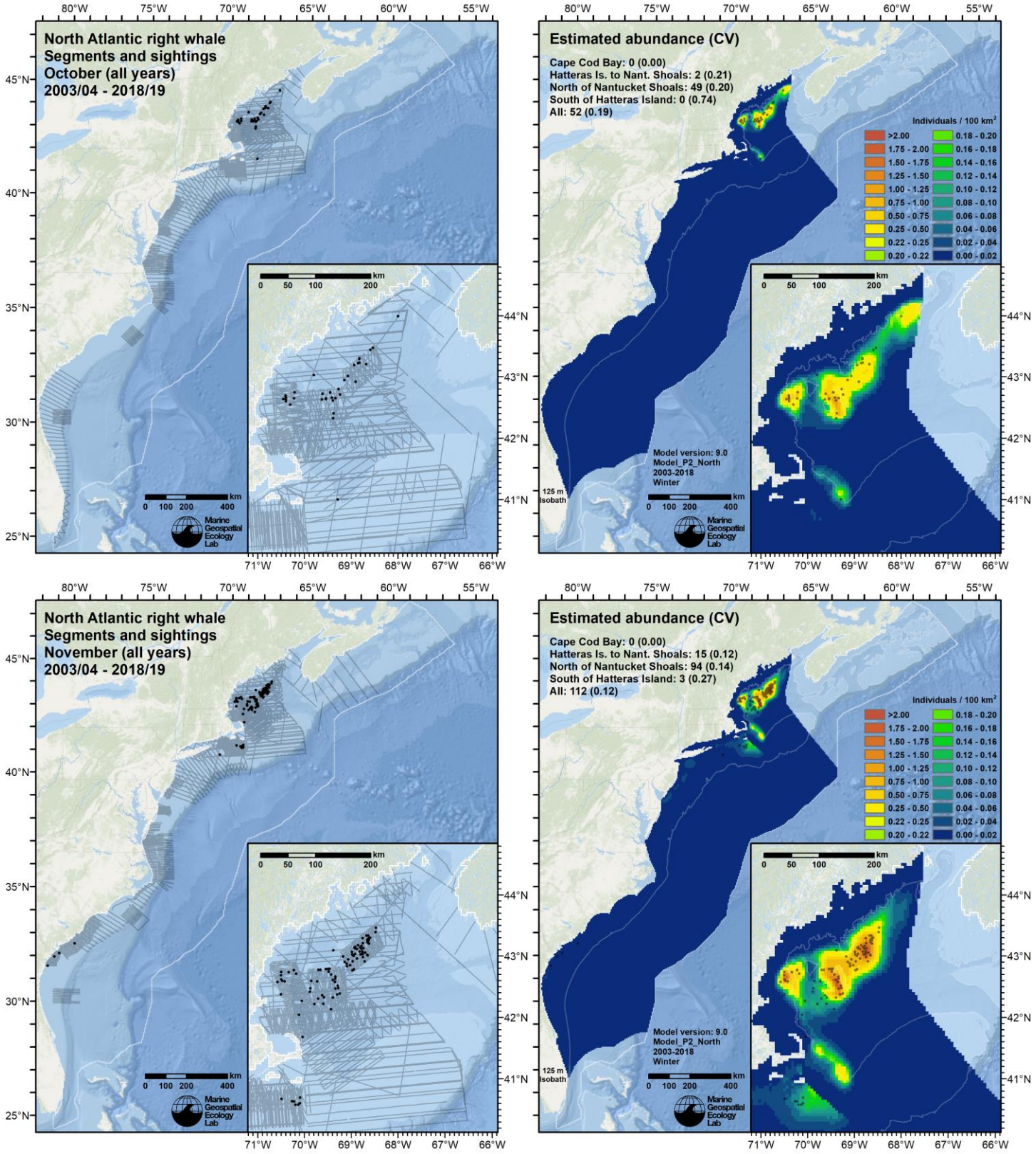


Figure 107. Continued from previous page.

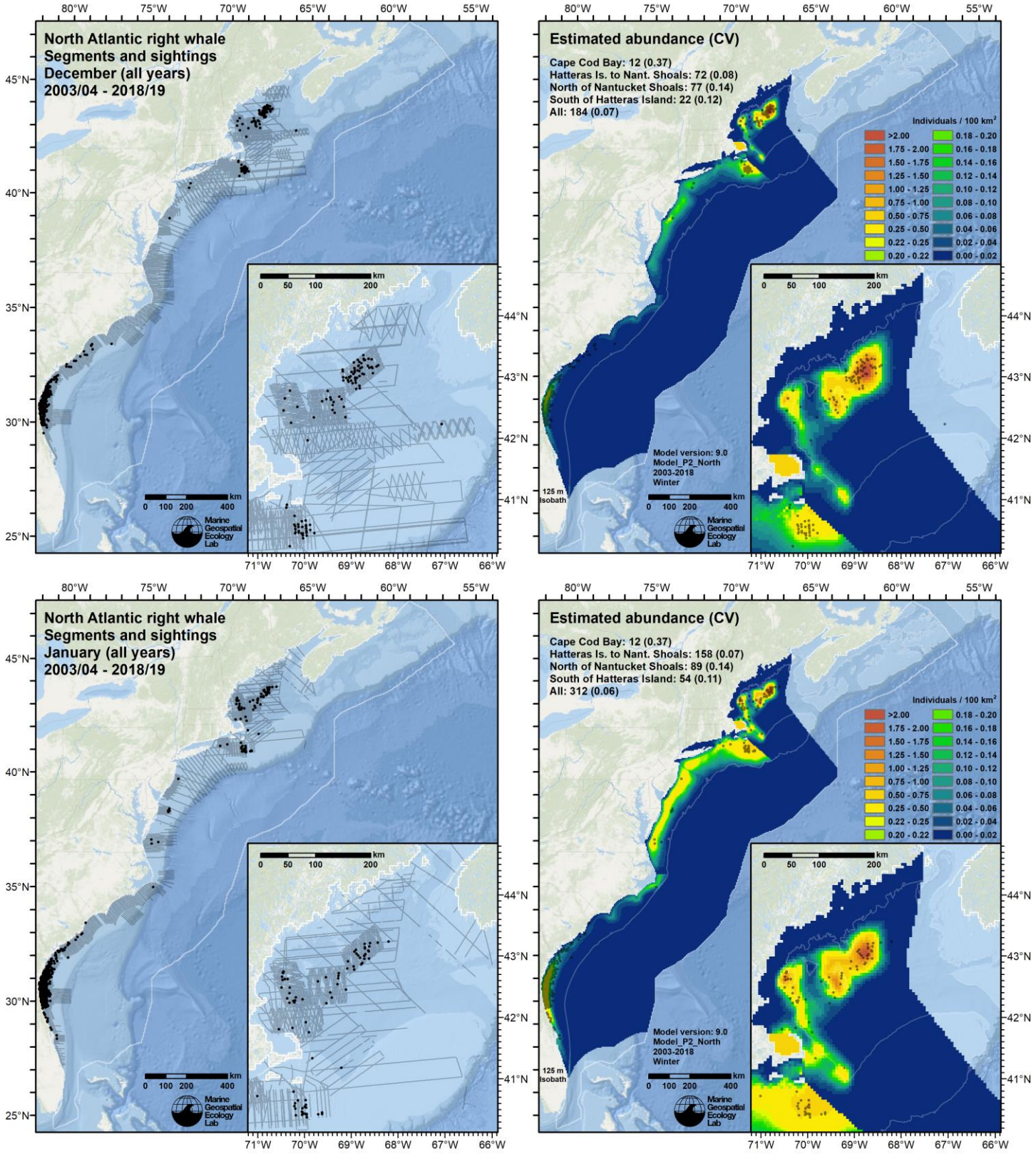


Figure 107. Continued from previous page.

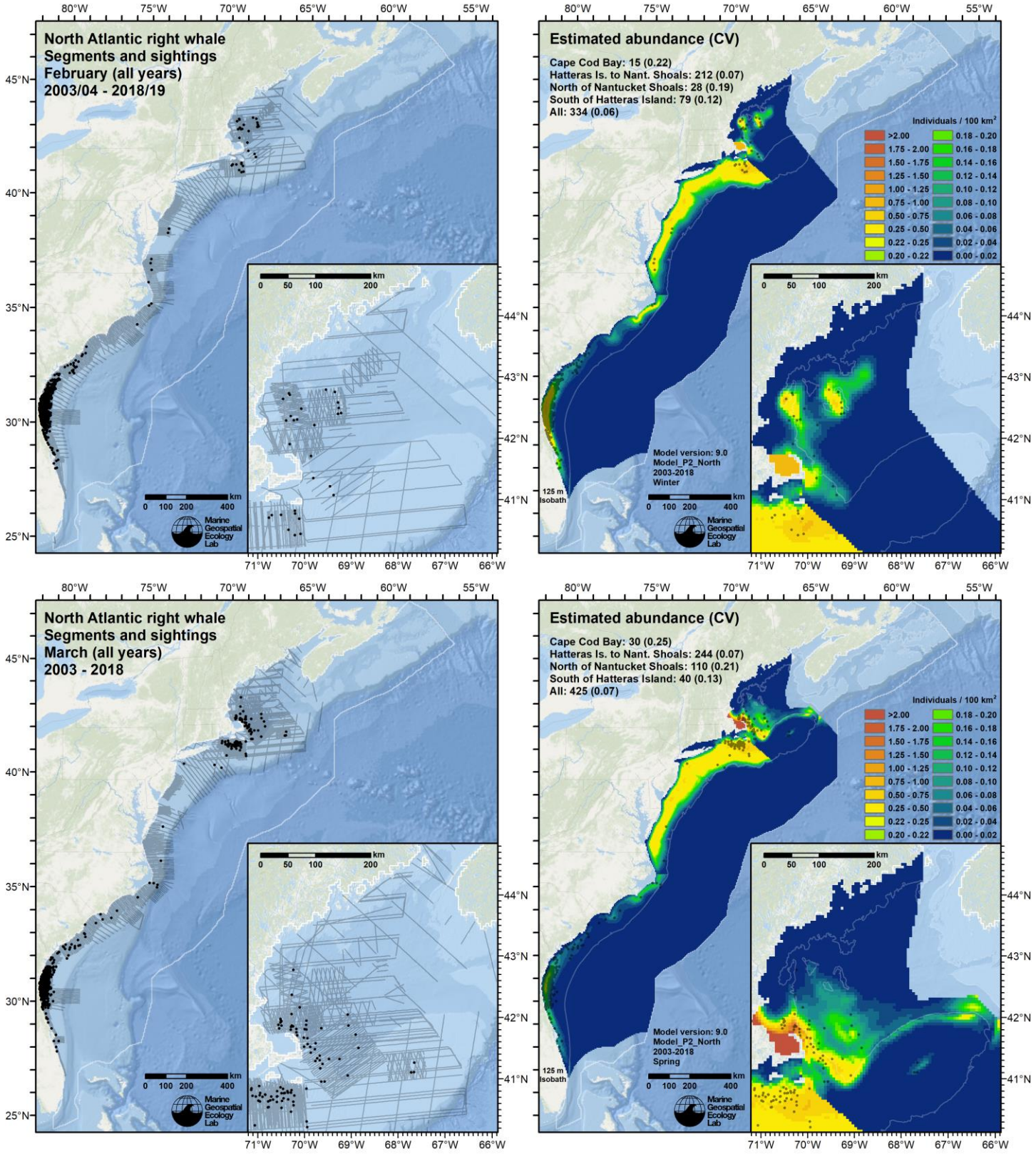


Figure 107. Continued from previous page.

5.3.4. Comparison to passive acoustic monitoring

For the past several years, Genevieve Davis and Sofie Van Parijs of NOAA NEFSC have coordinated the aggregation of a large database of hydrophone recordings contributed by many collaborators, and led several studies that characterized the spatiotemporal distributions of baleen whales, including right whales, revealed through passive acoustic detections of their vocalizations (Davis et al. 2017, 2020). At the present time, these data are not directly usable under "density surface modeling" methodology (Hedley & Buckland 2004; Miller et al. 2013); doing so would require, at minimum, sufficient data to produce detection functions for the hydrophones, and a detailed understanding of right whale calling rates, which may differ between locations, age classes, behaviors, or other factors (Cusano et al. 2018). (For a detailed discussion of density estimation methodology for passive acoustical data, see Marques et al. (2013)). Those problems may eventually be solved, or new methods developed that allow acoustic detections to be incorporated anyway, but in any case, the acoustic data represent an independent, systematic source of detections of right whale presence that may be used to assess density model predictions in a semi-quantitative way. We first completed such an assessment for the v7 right whale model (Roberts et al. 2017). Here, with the assistance of and permission from G. Davis, we present an update to that assessment, comparing the v9 model's predictions presented above to acoustic data provided by Davis for the period August 2004 through June 2019. The acoustic data include those used in the NEFSC-led studies (Davis et al. 2017, 2020) as well as new data that may not yet be published.

For each day each hydrophone was deployed, Davis and her collaborators determined whether or not right whales were acoustically "present", defined for the purpose of this analysis as having at least three good up-calls detected on that day. Using these per deployment, per-day acoustic presence/absence results, we computed for each deployment, for each month it was deployed, the fraction of days of the month right whales were acoustically present. We then plotted the results overtop monthly maps of mean density predicted for the 2003-2009 (Figure 108), 2010-2018 (Figure 109), and 2003-2018 (Figure 110) eras.

In total, the comparison included 430 hydrophone deployments, each represented in the referenced figures as a black dot, indicating no detections that month, or a red circle, indicating that detections did occur (with the size of the circle indicating the fraction of days with detections). Deployment durations ranged from 10-436 days in the water, with a mean of 135 days (SD=75.2 days). The figures do not indicate which year a deployment occurred, but instead present all deployments that were active for at least 10 days that month during any year of the era. Fewer deployments occurred during the 2003-2009 era (n=145) and they were shorter (mean duration=71.4 days, SD=44.8 days) than during the 2010-2018 era (n=285, mean duration=167 days, SD=66.4 days). We included deployments through June 2019 in the 2010-2018 and 2003-2018 series of maps.

We completed this assessment after all aspects of the v9 density model were complete, therefore modeling decisions made while we were completing the v9 model were not influenced by having viewed the results of this assessment. However, our v9 modeling decisions did incorporate lessons learned from the circa-2017 comparison of v7 model's predictions (Roberts et al. 2017) to acoustic data from 2004-2016.

During the 2003-2009 era (Figure 108), in general, the density model rarely failed to predict at least a small amount of density wherever right whales were detected acoustically within the spatial extent of predictions. The main exceptions to this came in the months of October and November, in which acoustic detections were reported along New York and New Jersey, but density predictions were essentially zero. During the 2010-2018 era (Figure 109), when much more acoustic data were available, matchup between the density predictions and acoustic detections was also generally good, except in three areas: 1) Along the shelf break, particularly in the months of February-July, acoustic detections were occasionally made but predicted density was essentially zero. 2) In the "South of the Islands" area, acoustic detections were made year-round (least frequently in August) but density predictions were essentially zero from July-October. 3) In coastal Maine, close to shore, a string of hydrophones deployed in fall of 2010 detected right whales occasionally from October-December, but density was predicted to be zero.

Given these discrepancies, we urge caution applying the density model predictions in these areas during the months noted. We also encourage collaborators to increase visual surveying and acoustic monitoring of these locations during those months so that a better understanding and better models can be obtained in the future.

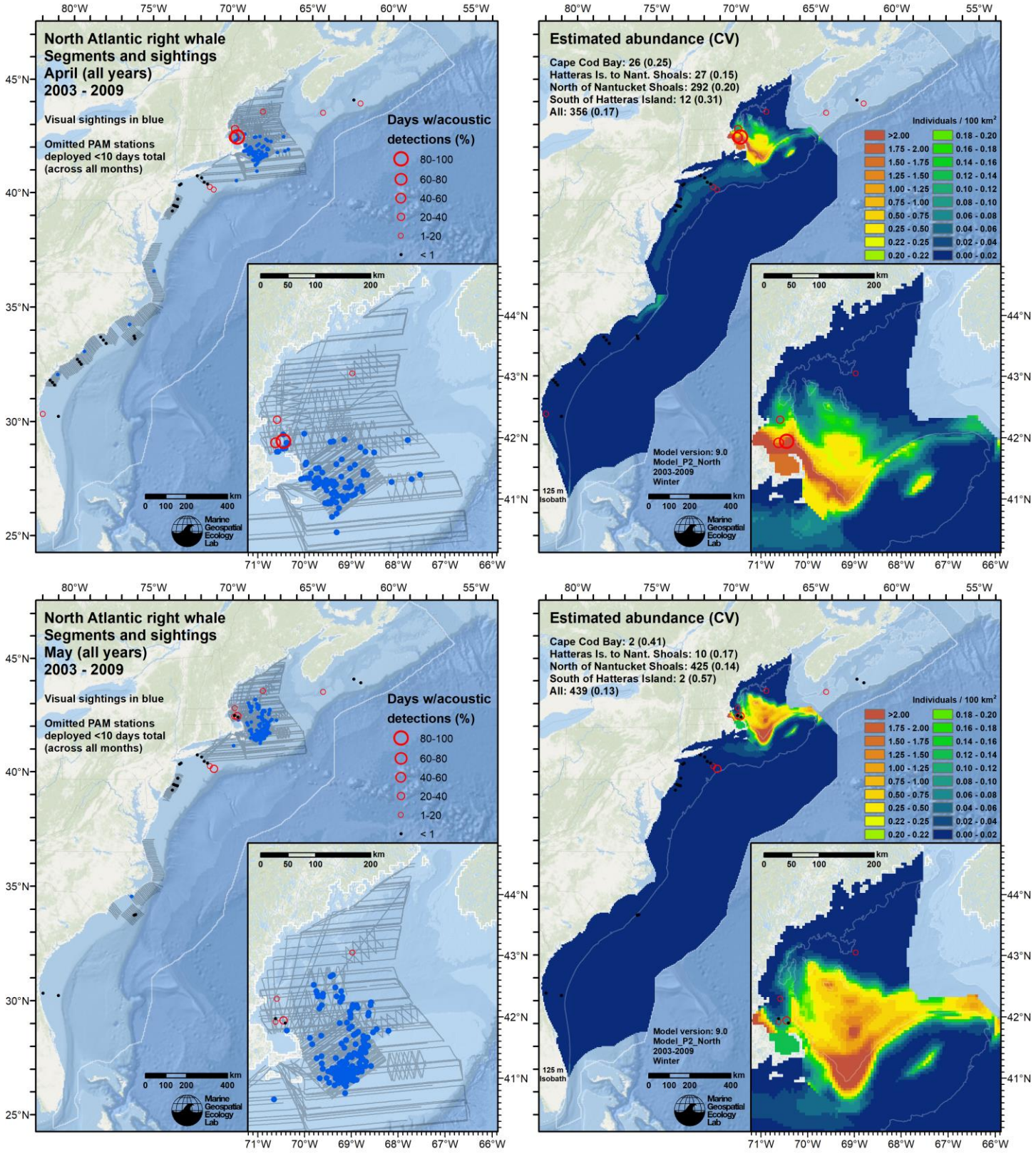


Figure 108. Monthly maps (rows) of survey effort, right whale sightings, and acoustic detections (left column) and of mean density predictions with acoustic detections overlaid (right column), for the "2003-2009" era, which spanned April 2003 through March 2010. Effort and sightings are all of those that occurred during the specified month during the era. Density predictions represent the mean for the specified month over the era. Acoustic data from August 2004 through March 2010, courtesy of G. Davis and her collaborators (Davis et al. 2017). Most hydrophones (black dots and red circles) were only active during one year of the era.

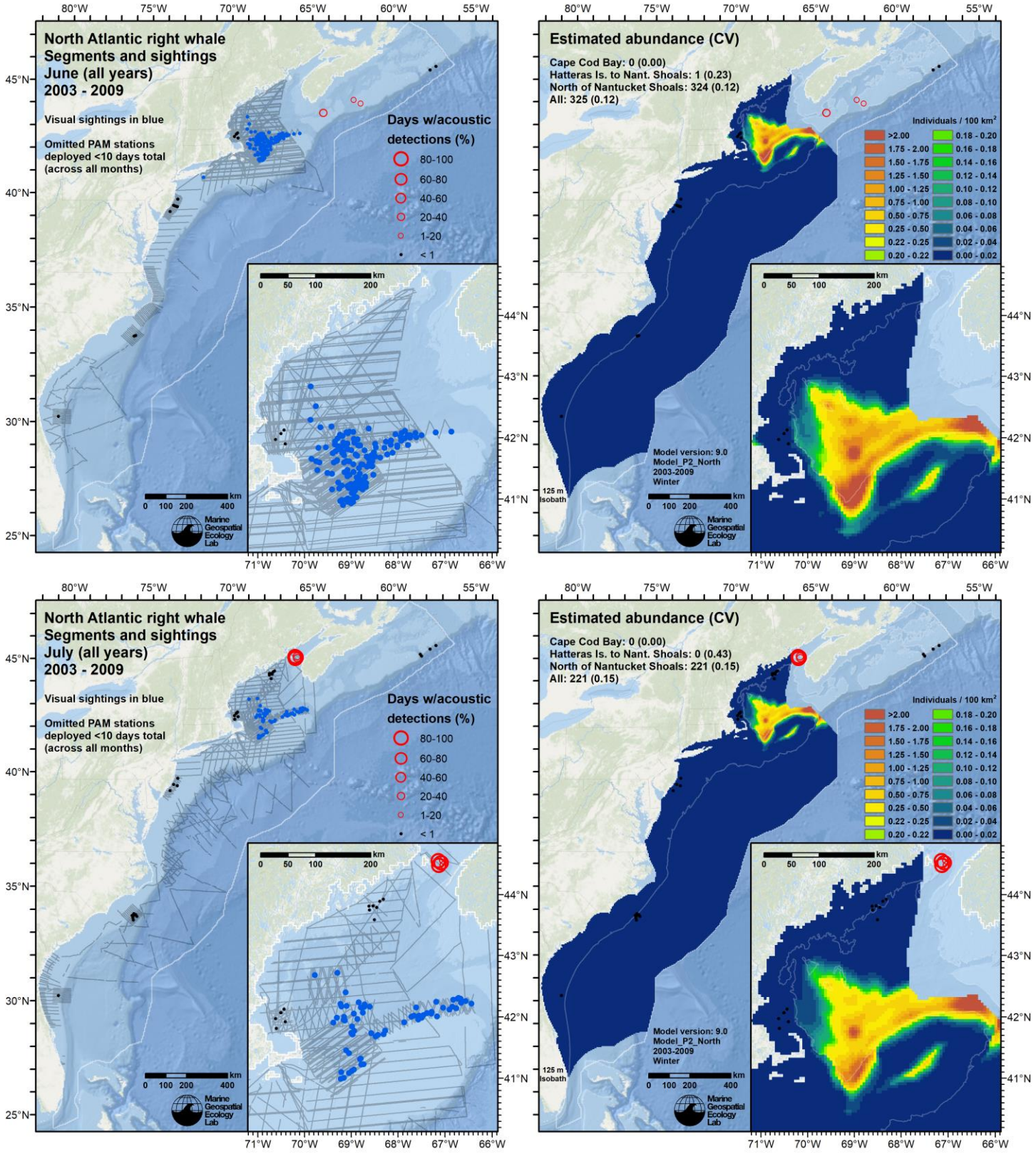


Figure 108. Continued from previous page.

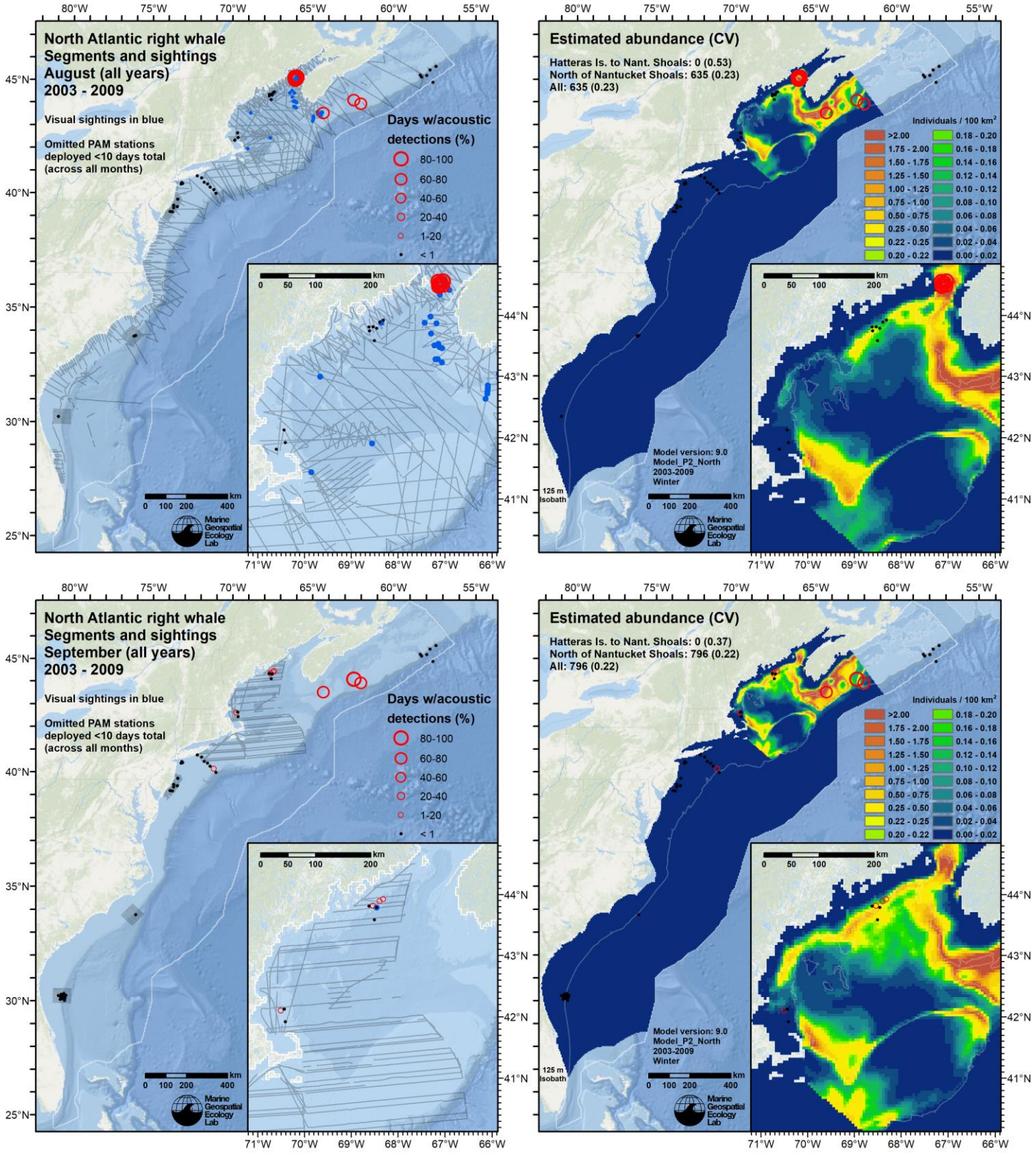


Figure 108. Continued from previous page.

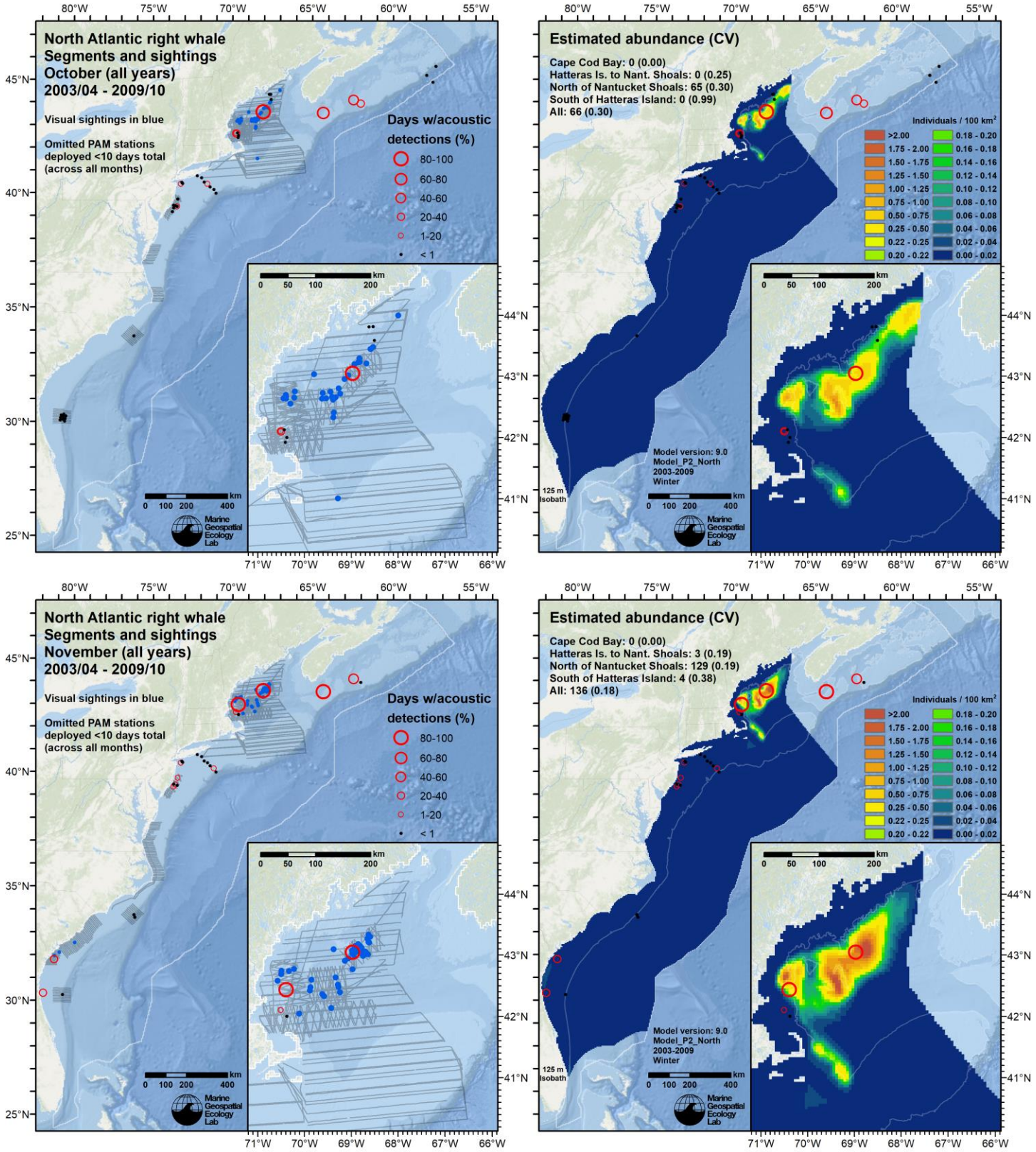


Figure 108. Continued from previous page.

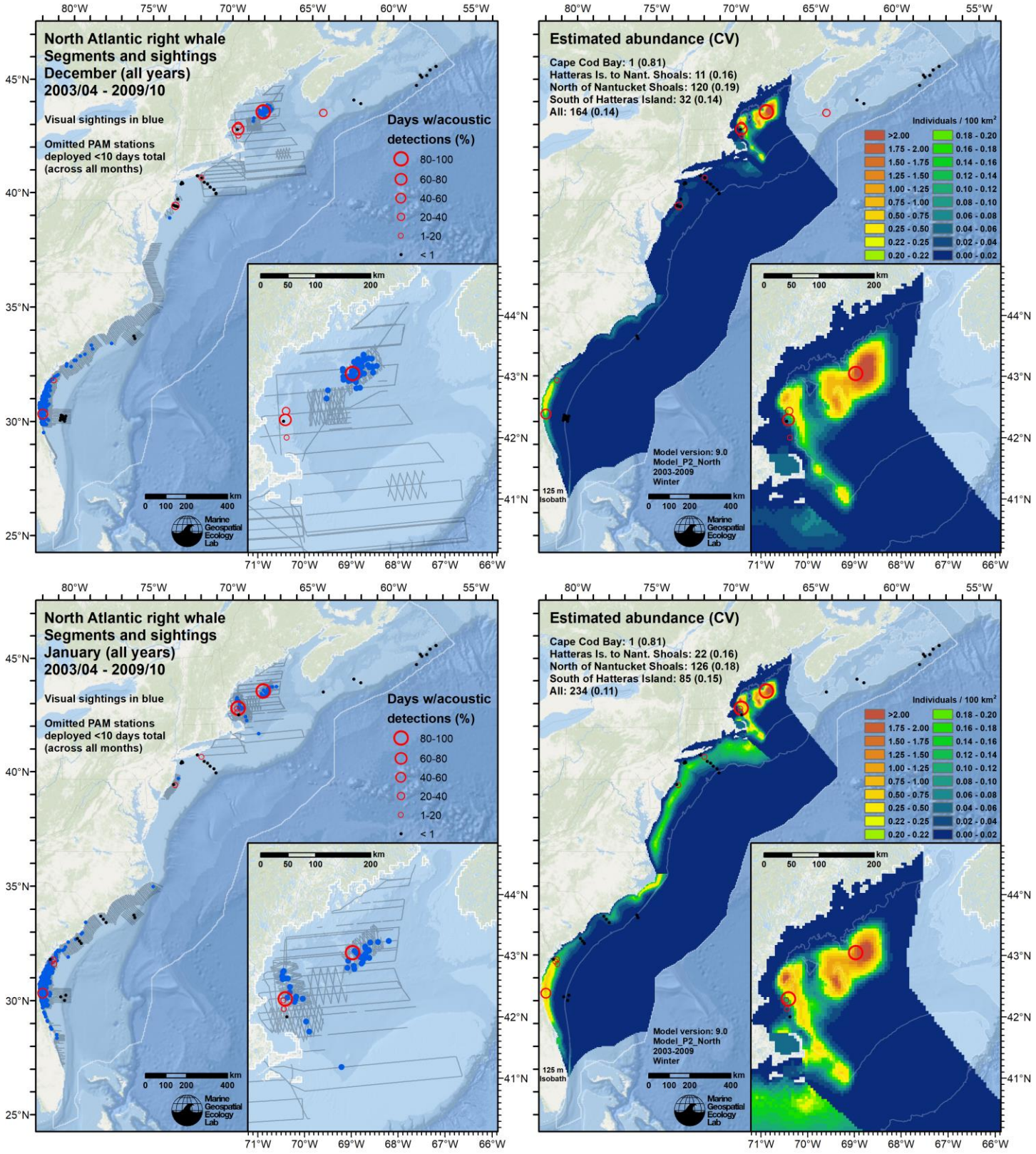


Figure 108. Continued from previous page.

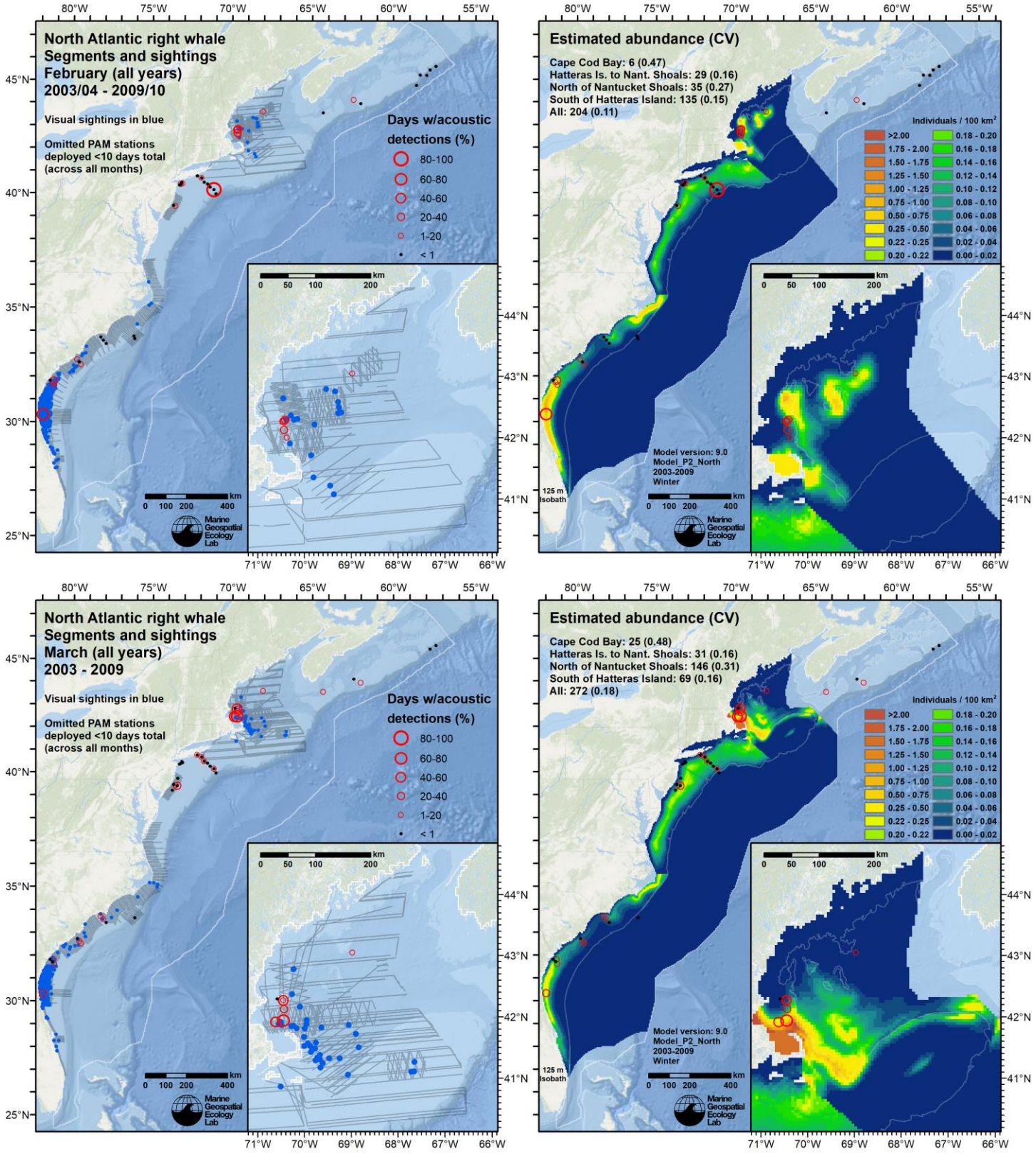


Figure 108. Continued from previous page.

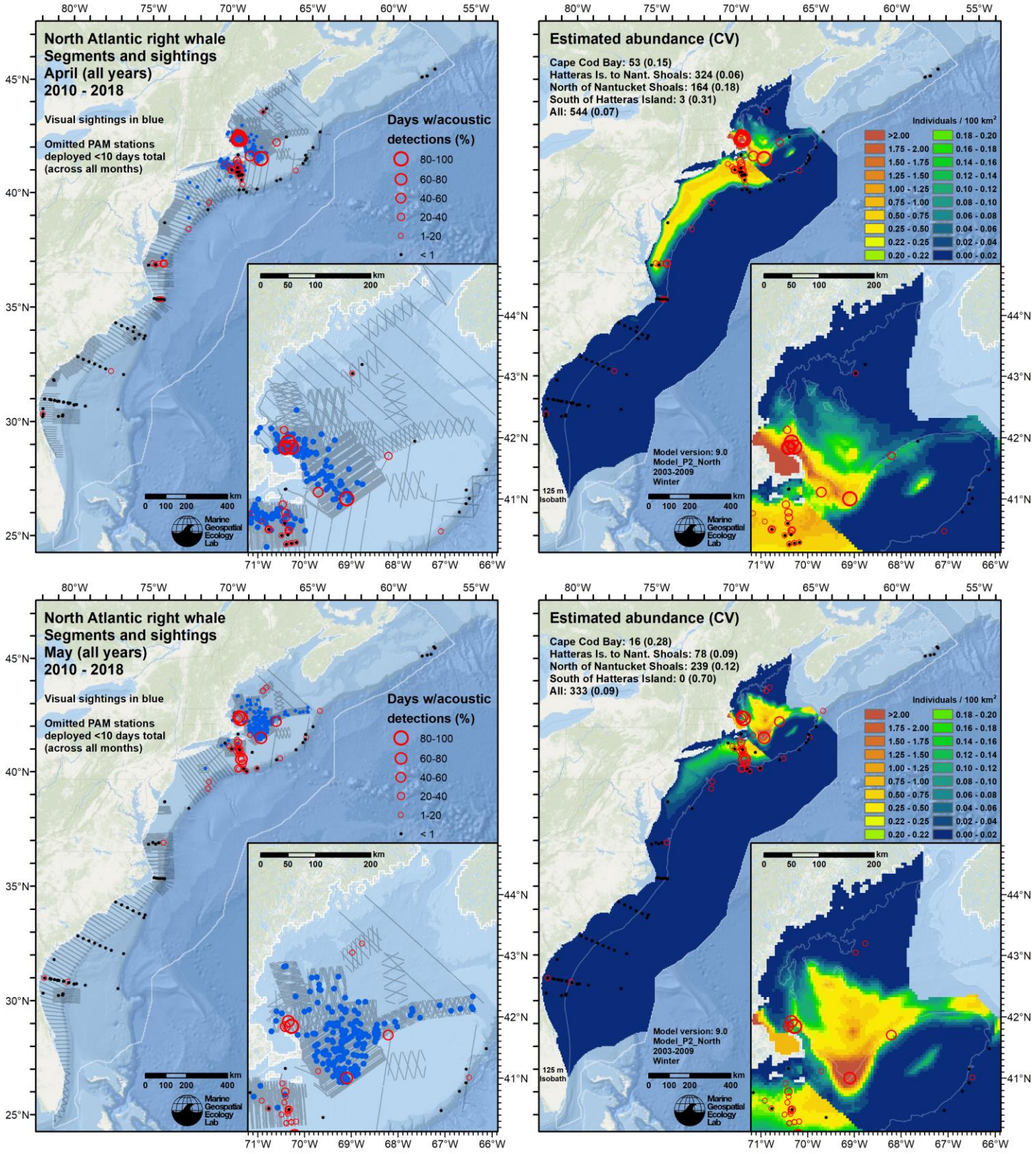


Figure 109. Monthly maps (rows) of survey effort, right whale sightings, and acoustic detections (left column) and of mean density predictions with acoustic detections overlaid (right column), for the "2010-2018" era, which spanned April 2010 through March 2019. Effort and sightings are all of those that occurred during the specified month during the era. Density predictions represent the mean for the specified month over the era. Acoustic data from April 2010 through June 2019, courtesy of G. Davis and her collaborators (Davis et al. 2017). Most hydrophones (black dots and red circles) were only active during one year of the era.

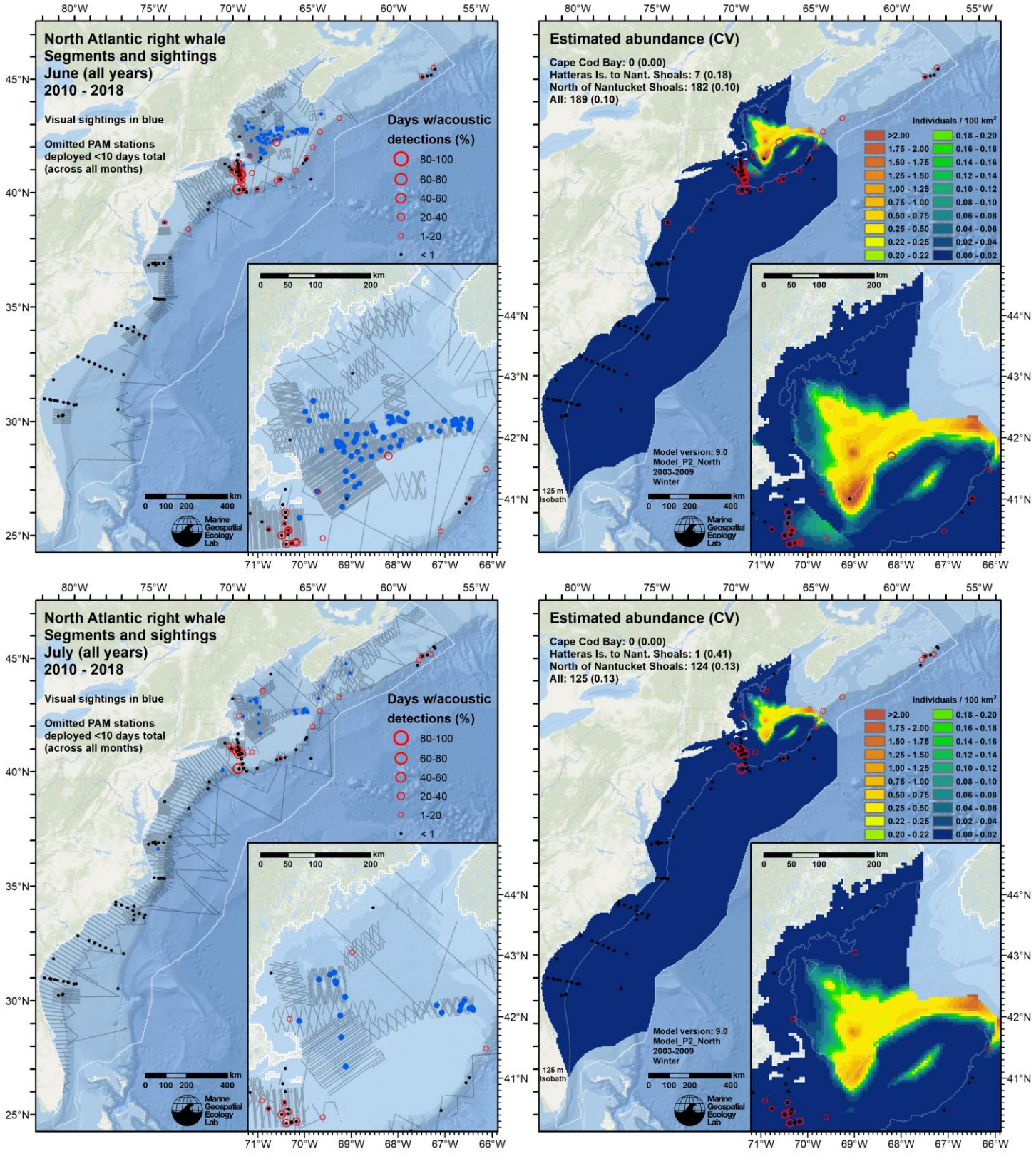


Figure 109. Continued from previous page.

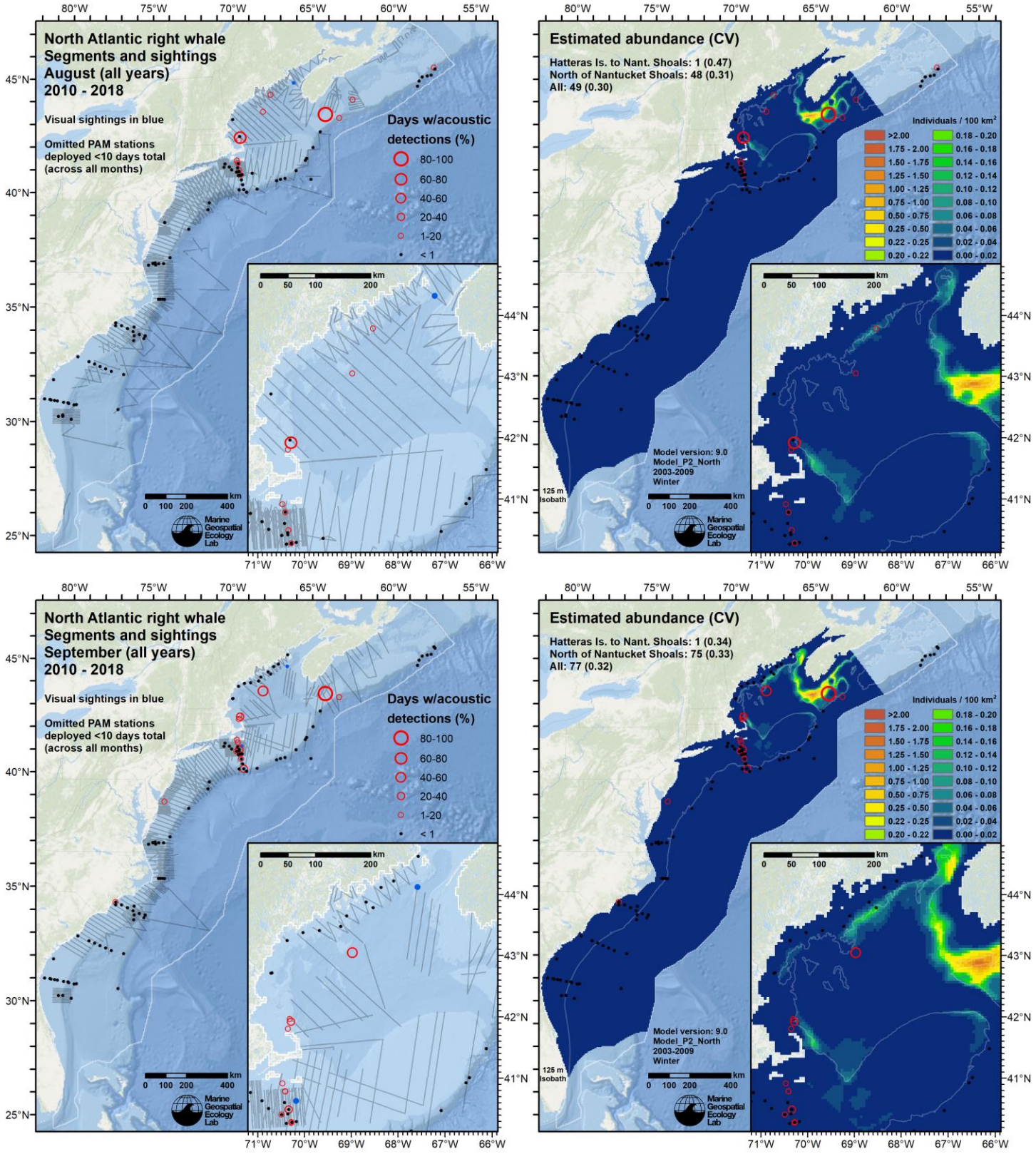


Figure 109. Continued from previous page.

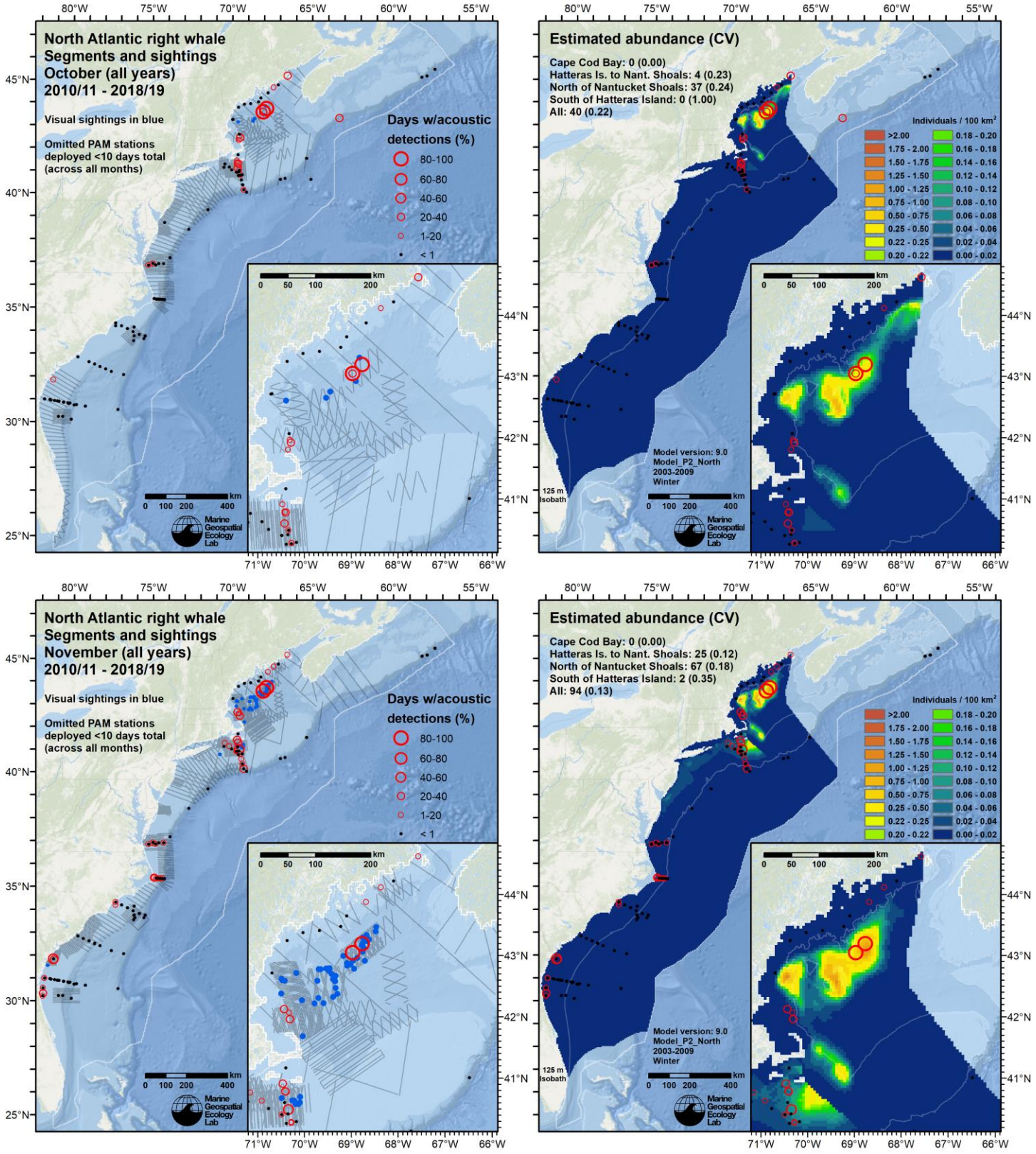


Figure 109. Continued from previous page.

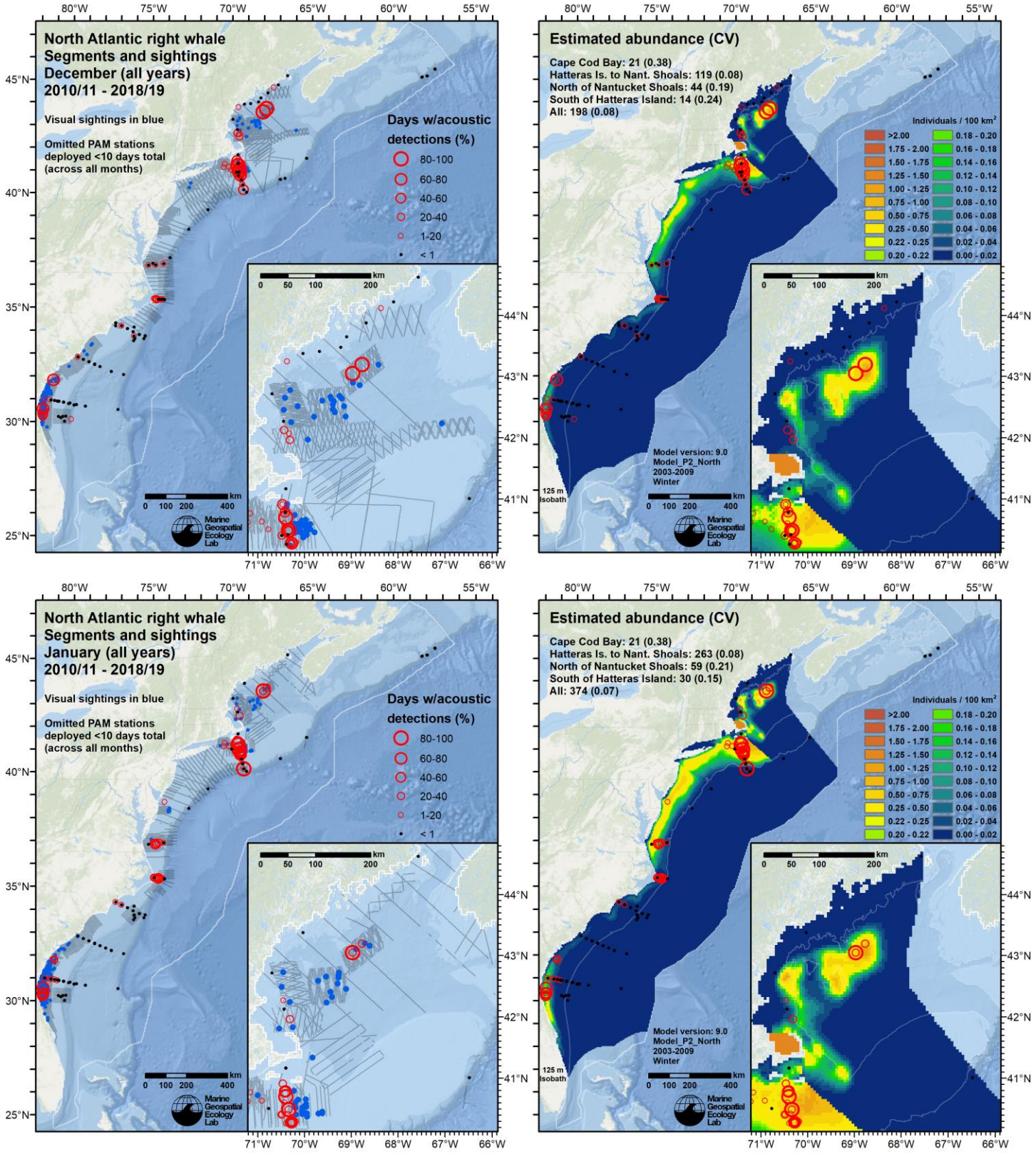


Figure 109. Continued from previous page.

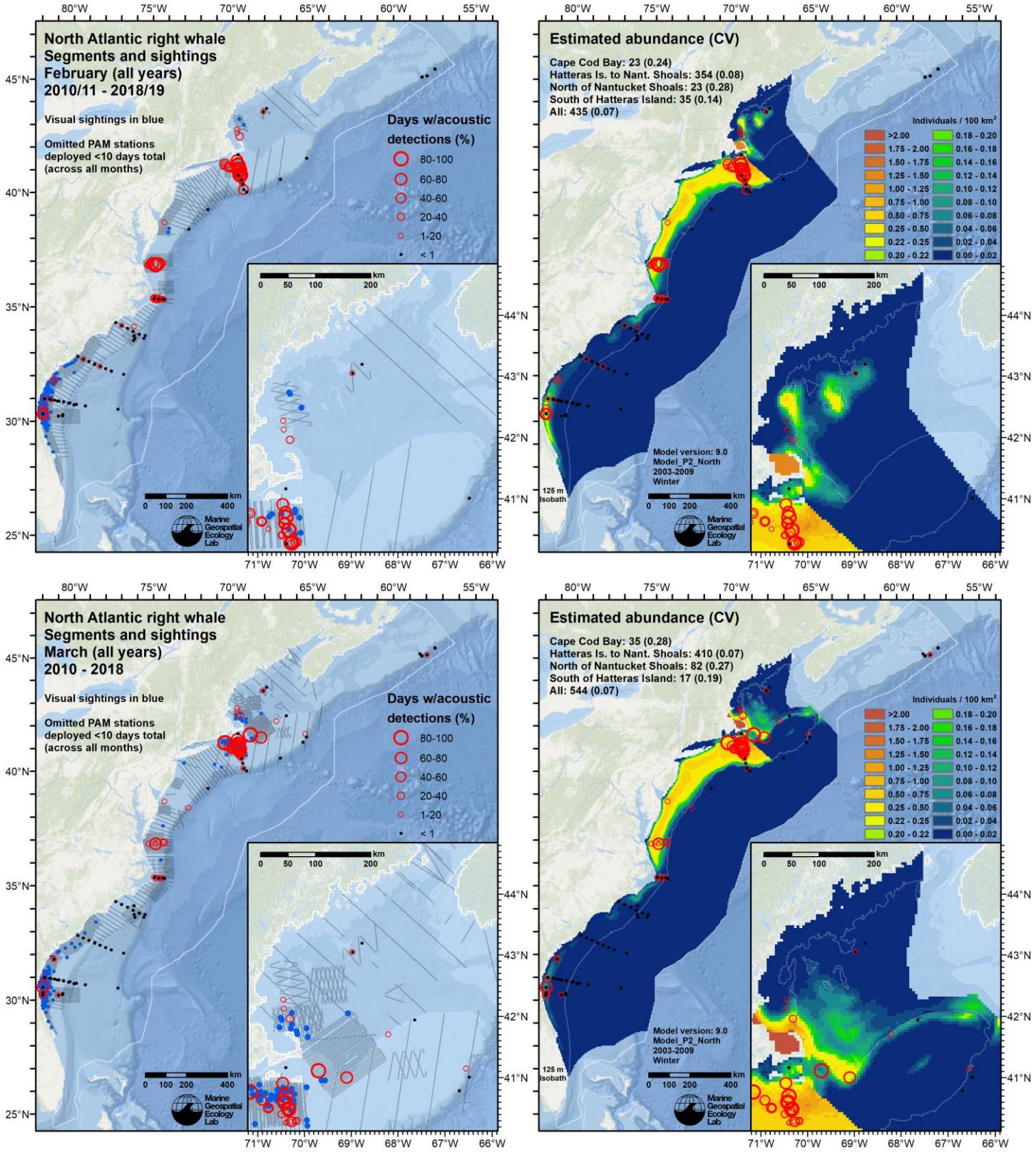


Figure 109. Continued from previous page.

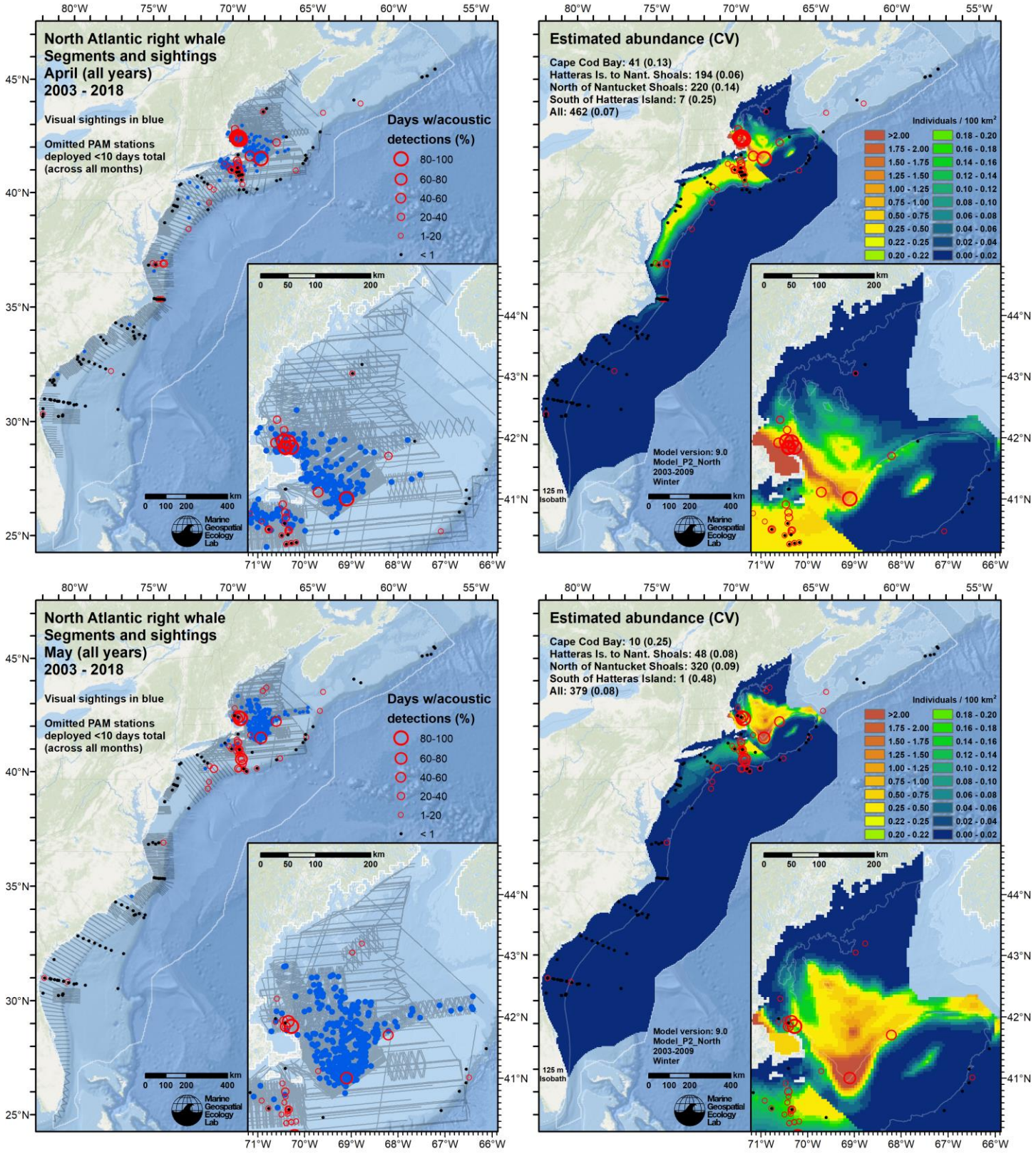


Figure 110. Monthly maps (rows) of survey effort, right whale sightings, and acoustic detections (left column) and of mean density predictions with acoustic detections overlaid (right column), for the "2003-2018" era, which spanned April 2003 through March 2019. Effort and sightings are all of those that occurred during the specified month during the era. Density predictions represent the mean for the specified month over the era. Acoustic data from August 2004 through June 2019, courtesy of G. Davis and her collaborators (Davis et al. 2017). Most hydrophones (black dots and red circles) were only active during one year of the era.

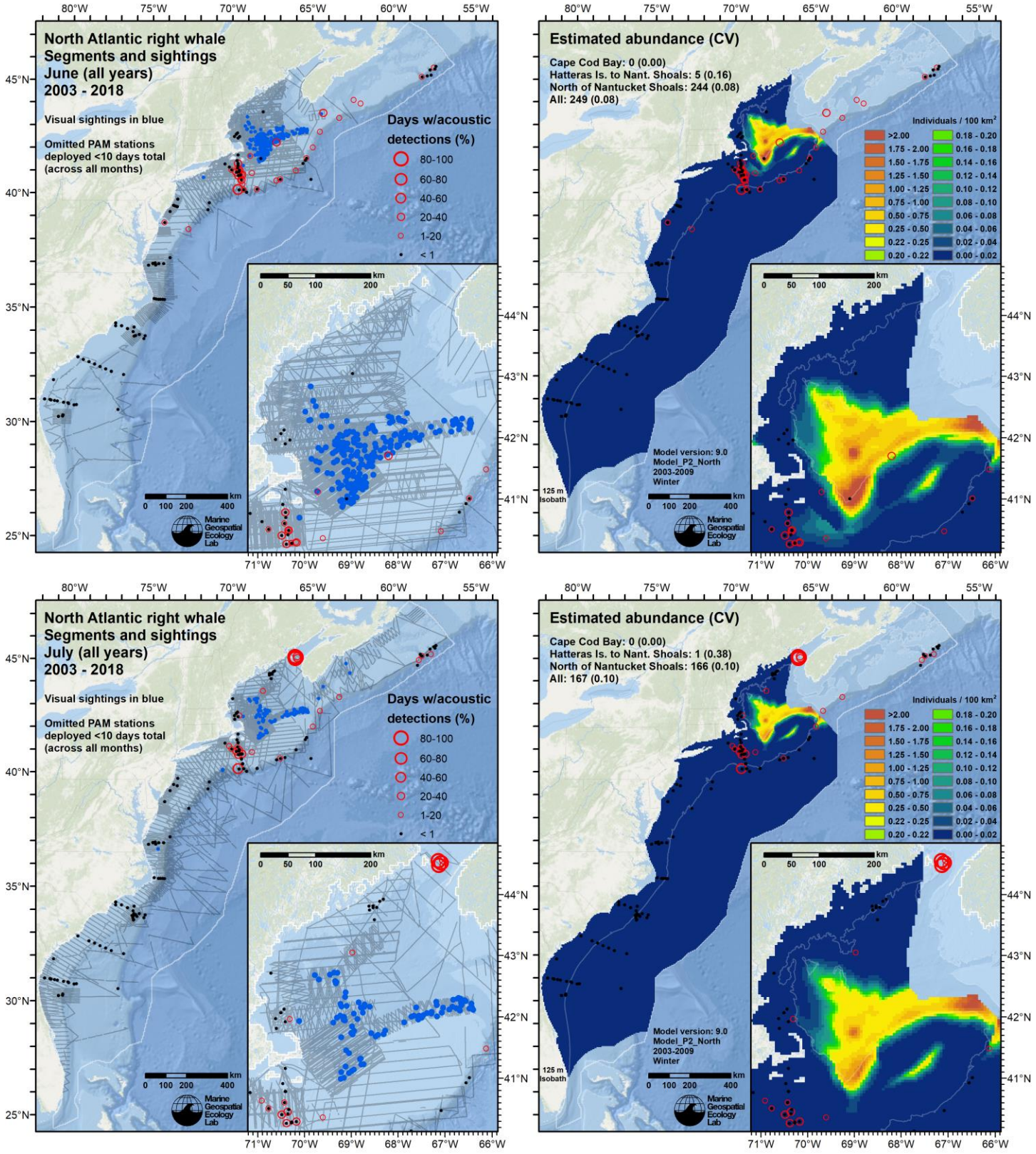


Figure 110. Continued from previous page.

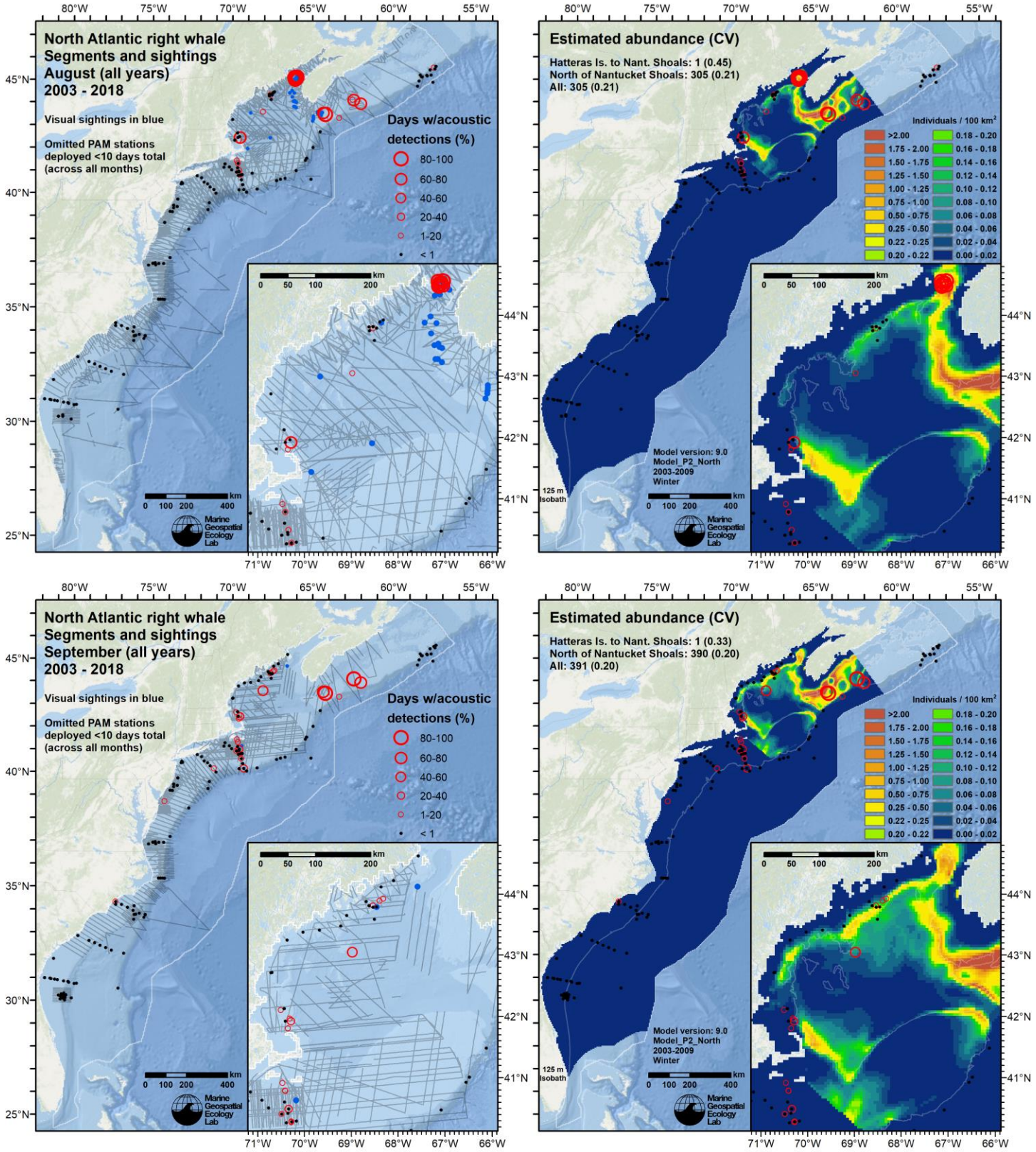


Figure 110. Continued from previous page.

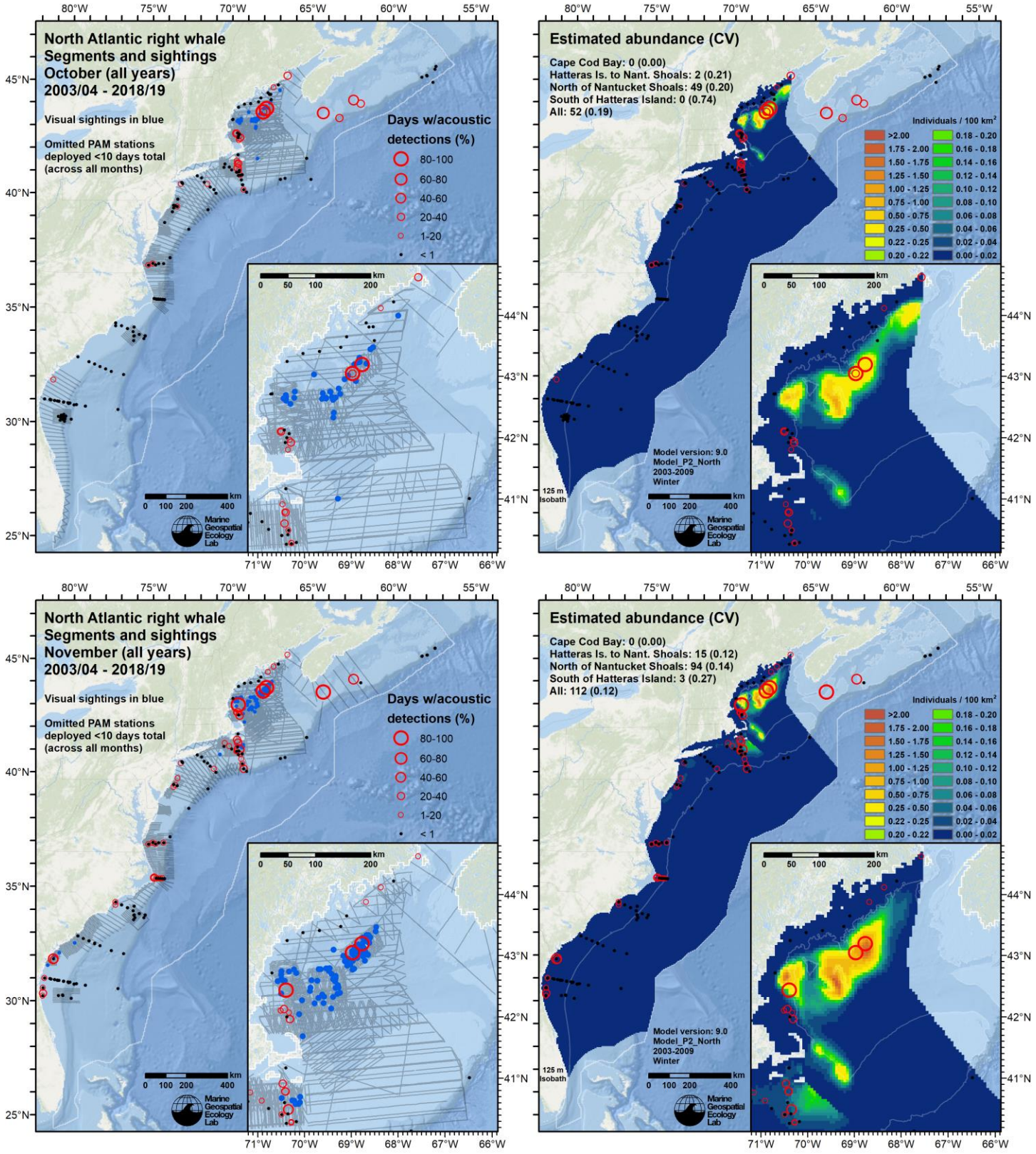


Figure 110. Continued from previous page.

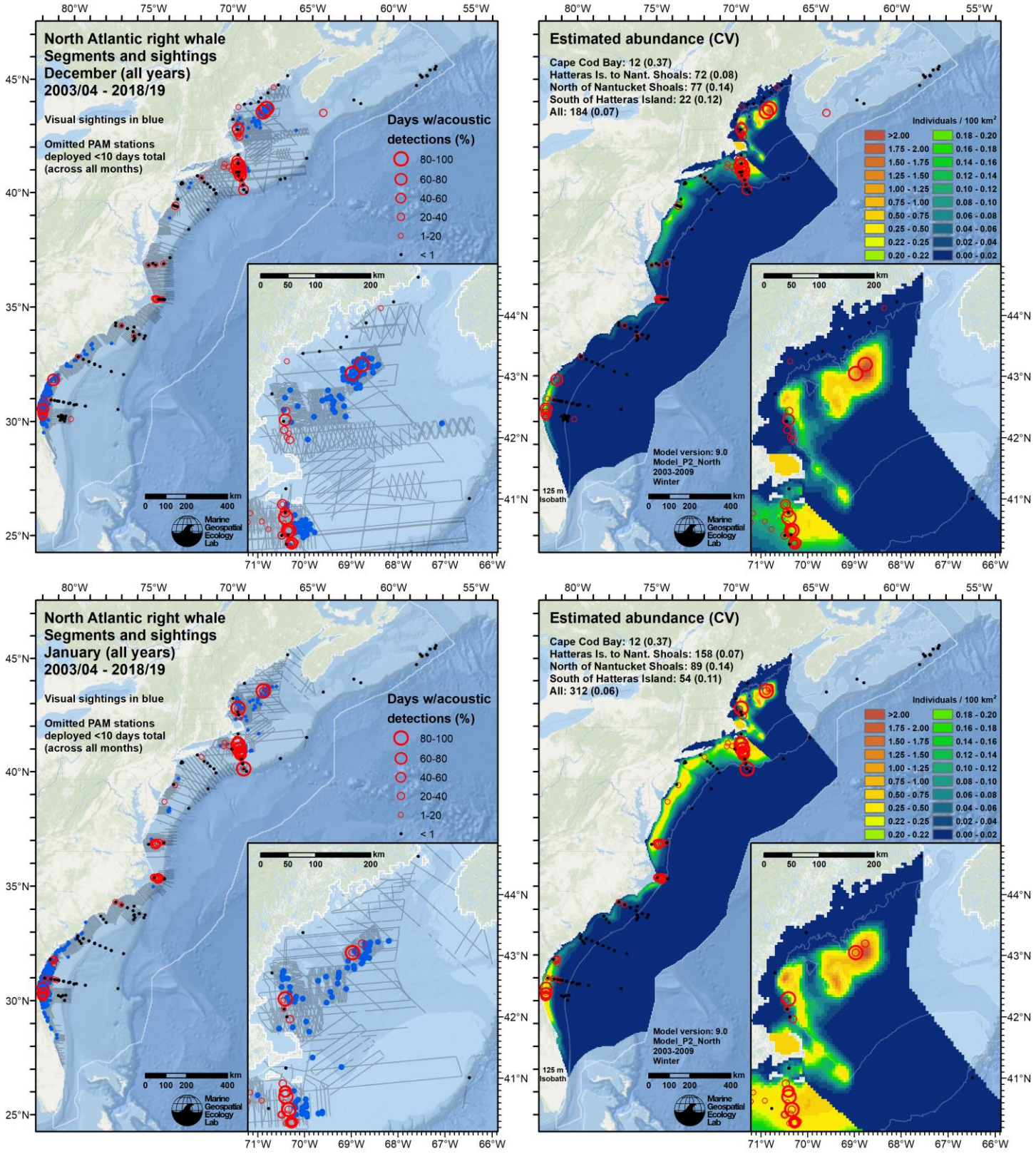


Figure 110. Continued from previous page.

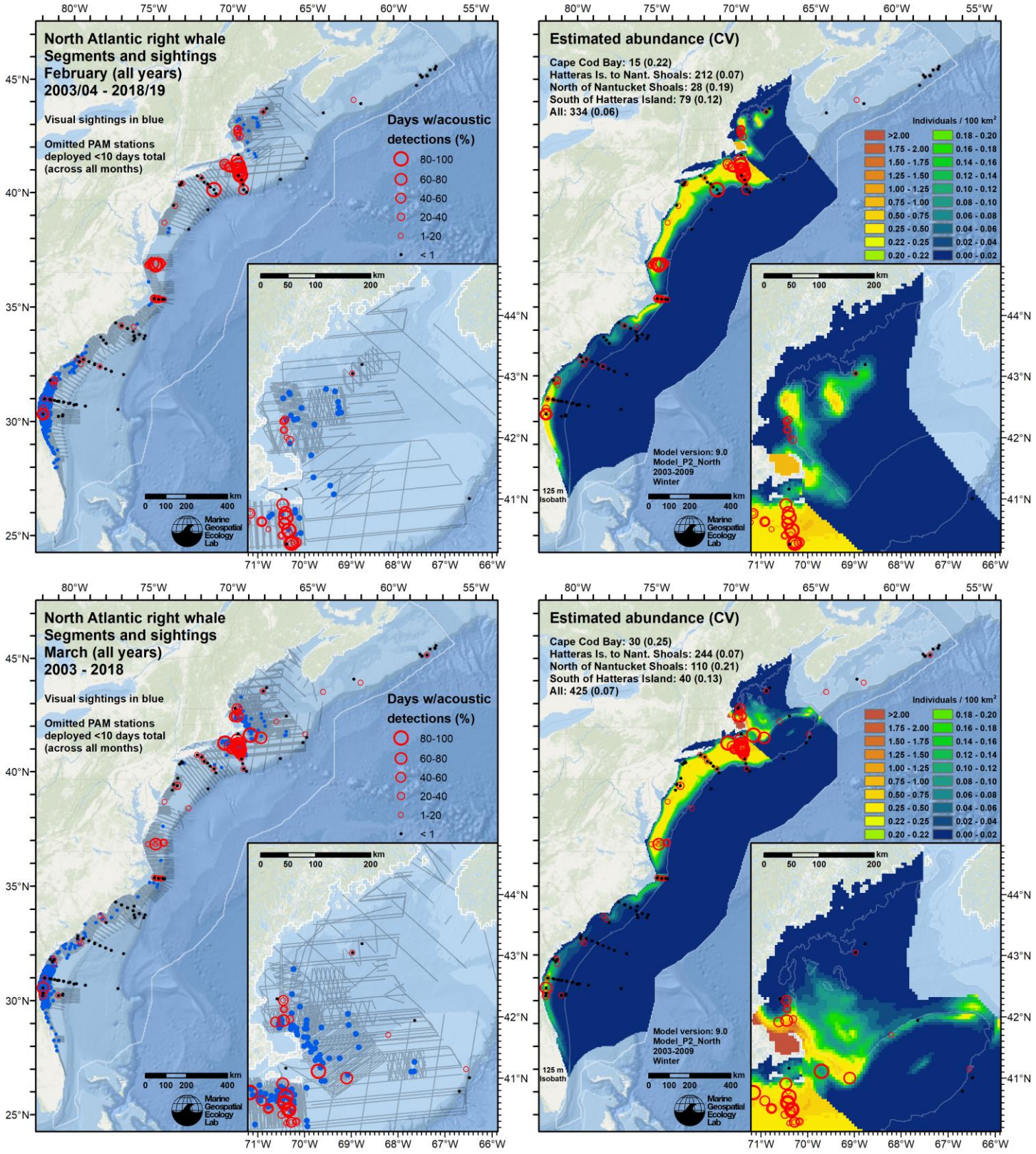


Figure 110. Continued from previous page.

6. Discussion

6.1. Intended uses of this model

During the past year, interested parties outside the U.S. Navy occasionally inquired about the appropriate uses for our marine mammal density models. These models are mainly intended for science and management problems in which: 1) an absolute density estimate (individuals / km^2) or rigorous effort correction is needed (as occurs when combining data from multiple survey programs), and 2) monthly climatological maps of average density are satisfactory representation of species distributions. Examples where these cases are often both true include:

- Development of MMPA take estimates, either for Letters of Authorization for the Navy or Incidental Harassment Authorizations for other users.
- Development of plans, regulations, or mitigations that have medium or large spatial or temporal scales, e.g. tens of thousands of square kilometers and at least a month long.
- Communication of basic patterns in cetacean species distributions in U.S. waters.
- Exploratory analysis in which the analyst needs species distribution maps but does not yet know the level of spatiotemporal precision that will ultimately be required. ("You have to start somewhere.")
- Development of inter-species comparative maps (e.g. species diversity maps).

These models are generally not intended for the following uses:

- Abundance estimation. When suitable photo ID data are available, capture-recapture methods, e.g. Pace et al. (2017), are likely to be more accurate; otherwise standard line-transect abundance estimates that carefully account for perception and availability bias, as are often (but not always) used for NOAA Stock Assessment Reports, may be a better choice.
- Fine-scale management, e.g. for study areas <500 km^2 in area. In these cases, a more focused study specifically tailored to the area and questions of interest may be appropriate. (However, our models may be suitable as a fallback, depending on the area and questions.)
- Nowcasting and dynamic management. Currently, our models produce climatological density predictions averaged over a number of years. These may be poor predictions for where animals are currently or will be in the near future. Our models could be adapted to perform well at providing nowcasts or short-term forecasts by incorporating right whale observations and detections in near real time, should that application become of interest to the Navy or other prospective users.
- Extrapolation to distant unsurveyed areas or the far future, except when explicitly designed for that purpose. The right whale model presented here, and the other regional density models described in Roberts et al. (2016a), were not designed for such extrapolation. The AFTT-wide models of Mannocci et al. (2017b) were explicitly designed for extrapolation across unsurveyed portions of the AFTT. It is possible to use a variation of the methodology presented here to build models for extrapolation into the far future, but that is beyond the scope of this project and the Navy's current interests.
- Rigorous explanation of right whale habitat preferences. The goal of our models is prediction, not explanation. Using multi-variable additive models for explanation can be unsatisfying, and we urge caution in overinterpreting our models this way.

6.2. Which predictions should be used for right whale management?

In this report, we present the predictions of a model fitted to data from 2003-2018, a period over which right whale abundance and distribution patterns are believed to have changed (see section 5.3.1.1). We took explicit steps (see section 5.3.1.3) to account for these changes in regional models (see section 5.3.2) that made up the overall model and summarized predictions into two eras, 2003-2009 and 2010-2018, as well as the entire 2003-2018 period (see section 5.3.3). Substantial changes were predicted between the eras, consistent with much of the evidence recently published about right whale abundance and distribution. Our general recommendation is to use the predictions for the 2010-2018 era for applications involving the density of right whales in the present day or for the next few years.

For users who need to assess the potential impact of proposed human activities on right whales and intend our models to serve as densities from which impacts will be quantified (e.g. by multiplying density by something representing the impact), and who want to be especially careful to minimize those impacts, we suggest using the 2003-2018 predictions instead for regions south of Hatteras Island or north of Nantucket Shoals. Doing so will incorporate the higher densities predicted for the 2003-2009 period, which plausibly represent a state that right whales could return to in the future, if conditions reverted to that era.

Similarly, for users who are confident that our model overpredicts density in the 2010-2018 period for the region between Hatteras Island and Nantucket Shoals, we suggest using the 2003-2018 predictions, which will dampen the predictions from 2010-2018 with those from 2003-2009. However, we urge extreme caution in taking this approach, and suggest instead you inquire with us about the latest progress in modeling the mid-Atlantic region.

6.3. Anticipated updates to this model

In the near future, or for the completion of the Navy AFTT Phase IV modeling process, we anticipate making the following improvements and releasing an updated model.

6.3.1. Improve the model for the Hatteras Island to Nantucket Shoals region

We believe this model may have overpredicted density in the 2010-2018 period, mainly in the mid-Atlantic region, south of the large aggregation of sightings in the "South of the Islands" area (see section 5.3.2.2). Possible solutions to investigate include:

- Split the "South of the Islands" area off from the mid-Atlantic area to the south. We experimented with this for the current model but were not able to obtain a good result for the mid-Atlantic, but one might emerge after further experimentation.
- Develop new covariates that can distinguish the "South of the Islands" area from the region to the south. Our collaborators gave some good suggestions during our May 2020 review workshop, including trying anomalies or some of the ocean color processing done by Thomas et al. (2003).
- Examine the surveys in the region and remove any that were even slightly directed to known aggregations of right whales. We had already removed non-systematic verification flights flown by NARWSS and New England Aquarium, but it is possible that systematic flights by NARWSS to check the status of dynamic management areas were too biased toward times that right whales might be present. If so, this would bias density high.

6.3.2. Improve the "Summer" model for the region north of Nantucket Shoals

This model clearly overpredicted density in the 2003-2009 period (see section 5.3.2.3.2). One possible reason is the very strong corrections for both perception and availability bias that were used. We propose to examine whether these were realistic, and whether reasonable predictions could be obtained by adjusting them.

6.3.3. Reduce edge effects between regions

The final density surfaces (see section 5.3.3) clearly show "edge effects" between modeled regions, especially where models meet at Nantucket Shoals. Suggestions for reducing this effect include:

- During model fitting, include some data from outside the area that will be predicted, so that the training data extends beyond the region of prediction. This may help avoid predicting extreme values at the edges of predicted areas. (Survey segments near the edges of regional models may end up being used by models on both sides of the edge, but this will not result in extra density being predicted.)

- During model prediction, define a band of cells where models meet and average predictions between the two within the band, weighted by how close the cells are to each side.

6.3.4. Apply new uncertainty propagation methods to improve uncertainty estimates

Currently, model uncertainties are quantified as coefficients of variation (or standard errors) that represent the errors in GAM parameter estimates. These errors factor in the quantity of survey effort available, the number of segments with sightings, and the strength of the correlations between the model covariates and abundances observed on the survey segments. But, owing to limitations with the statistical methodology of density surface modeling, we were not able to account for several important known sources of uncertainty into model predictions, including: 1) uncertainty in detection function parameters, 2) uncertainty in perception and availability bias correction factors, and 3) variability in dynamic covariates. Recently, the "DenMod" Working Group and its collaborators have developed methods for propagating uncertainty from sources 1 and 3 into the final predictions. We are applying these methods as part of the work done for Option Year 4 of this cooperative agreement. We will continue to work with the DenMod Working Group to eventually obtain a solution for source 2.

7. Acknowledgements

Above all, we thank the observers, scientists, engineers, pilots, captains, and crews who collected and shared line transect surveys and environmental covariates with us; thank you for the opportunity to analyze the data you produced. We also thank the funding agencies and program managers for making these surveys possible. We are very grateful to the scientists who collaborated with us on this project. While not coauthors on this report to the Navy, they will be coauthors on the journal manuscript that will eventually result from this project. They are, alphabetically, Susan Barco, Mark Baumgartner, Tim Cole, Peter Corkeron, Mark Cotter, Genevieve Davis, Rob DiGiovanni Jr., Laura Ganley, Lance Garrison, Caroline Good, Tim Gowan, Bob Kenney, Amy Knowlton, Christian Khan, Kate MacNair, Bill McLellan, Ann Pabst, Debi Palka, Dan Pendleton, Ester Quintana, Jessica Redfern, Meghan Rickard, Melanie White, and Ann Zoidis. We are very grateful to Beth Josephson and Laura Dias for meticulously preparing NOAA survey data for our use and responding to our questions. Thanks to Genevieve Davis, Sofie Van Parijs, and their collaborators for allowing us to compare our density model predictions to their acoustic detections. Thanks to David L. Miller, Elizabeth Becker, and Karin Forney for analysis advice. Danielle Jones, Laura Sparks, and Andrew DiMatteo provided very helpful comments on drafts of this report. Finally, we are especially grateful to U.S. Navy Fleet Forces Command and NOAA Fisheries for funding the development of this model, and Navy Facilities Engineering Command Atlantic for managing this Cooperative Agreement.

8. References

Akaike H. 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control **19**:716–723.

Atlas R, Hoffman RN, Ardizzone J, Leidner SM, Jusem JC, Smith DK, Gombos D. 2011. A Cross-calibrated, Multiplatform Ocean Surface Wind Velocity Product for Meteorological and Oceanographic Applications. Bulletin of the American Meteorological Society **92**:157–174.

Barco SG, Burt L, DePerte A, Digiovanni, Jr. R. 2015. Marine Mammal and Sea Turtle Sightings in the Vicinity of the Maryland Wind Energy Area July 2013-June 2015, VAQF Scientific Report #2015-06. Page 93. Virginia Aquarium & Marine Science Center Foundation, Virginia Beach, VA.

Barlow J, Sexton S. 1996. The effect of diving and searching behavior on the probability of detecting track-line groups, go, of long-diving whales during linetransect surveys. Page 21. Administrative Report LJ-96-14. NOAA National Marine Fisheries Service, Southwest Fisheries Center.

Baumgartner M, Wenzel F, Lysiak N, Patrician M. 2017. North Atlantic right whale foraging ecology and its role in human-caused mortality. Marine Ecology Progress Series **581**:165–181.

Baumgartner MF, Mate BR. 2003. Summertime foraging ecology of North Atlantic right whales. Marine Ecology Progress Series **264**:123–135.

Becker EA, Forney KA, Redfern JV, Barlow J, Jacox MG, Roberts JJ, Palacios DM. 2019. Predicting cetacean abundance and distribution in a changing climate. Diversity and Distributions **25**:626–643.

Becker JJ et al. 2009. Global Bathymetry and Elevation Data at 30 Arc Seconds Resolution: SRTM30_PLUS. Marine Geodesy **32**:355–371.

Bleck R. 2002. An oceanic general circulation model framed in hybrid isopycnic-Cartesian coordinates. Ocean Modelling **4**:55–88.

Brillant SW, Vanderlaan AS, Rangeley RW, Taggart CT. 2015. Quantitative estimates of the movement and distribution of North Atlantic right whales along the northeast coast of North America. Endangered Species Research **27**:141–154.

Buckland ST, Anderson DR, Burnham KP, Laake JL, Borchers DL, Thomas L. 2001. Introduction to Distance Sampling: Estimating Abundance of Biological Populations. Oxford University Press, Oxford, UK.

Burt ML, Borchers DL, Jenkins KJ, Marques TA. 2014. Using mark-recapture distance sampling methods on line transect surveys. Methods in Ecology and Evolution **5**:1180–1191.

Carretta JV, Lowry MS, Stinchcomb CE, Lynn MS, Cosgrove R. E. 2000. Distribution and abundance of marine mammals at San Clemente Island and surrounding offshore waters: results from aerial and ground surveys in 1998 and1999. NOAA ADMINISTRATIVE REPORT LJ-00-02.

CETAP. 1982. A characterization of marine mammals and turtles in the mid- and North Atlantic areas of the U.S. outer continental shelf, final report, Cetacean and Turtle Assessment Program, University of Rhode Island. Page 576. #AA551-CT8-48. Bureau of Land Management, Washington, DC.

Chavez-Rosales S, Palka DL, Garrison LP, Josephson EA. 2019. Environmental predictors of habitat suitability and occurrence of cetaceans in the western North Atlantic Ocean. Scientific Reports **9**:5833.

Chin TM, Vazquez-Cuervo J, Armstrong EM. 2017. A multi-scale high-resolution analysis of global sea surface temperature. Remote Sensing of Environment **200**:154–169.

Cole T, Gerrior P, Merrick RL. 2007. Methodologies of the NOAA National Marine Fisheries Service Aerial Survey Program for Right Whales (Eubalaena glacialis) in the Northeast U.S., 1998-2006. CRD 07-02. U.S. Department of Commerce, Woods Hole, MA. Available from http://www.nefsc.noaa.gov/publications/crd/crd0702/ (accessed May 30, 2014).

Cole TV, Hamilton P, Henry AG, Duley P, Pace III RM, White BN, Frasier T. 2013. Evidence of a North Atlantic right whale Eubalaena glacialis mating ground. Endang Species Res **21**:55–64.

Cotter MP. 2019. Aerial Surveys for Protected Marine Species in the Norfolk Canyon Region: 2018–2019 Final Report. Page 75. HDR, Inc., Virginia Beach, VA.

Cusano DA, Conger LA, Van Parijs SM, Parks SE. 2018. Implementing conservation measures for the North Atlantic right whale: considering the behavioral ontogeny of mother-calf pairs. Animal Conservation. Available from http://doi.wiley.com/10.1111/acv.12457 (accessed June 3, 2019).

Davies K, Brown M, Hamilton P, Knowlton A, Taggart C, Vanderlaan A. 2019. Variation in North Atlantic right whale Eubalaena glacialis occurrence in the Bay of Fundy, Canada, over three decades. Endangered Species Research **39**:159–171.

Davis GE et al. 2017. Long-term passive acoustic recordings track the changing distribution of North Atlantic right whales (Eubalaena glacialis) from 2004 to 2014. Scientific Reports **7**. Available from http://www.nature.com/articles/s41598-017-13359-3 (accessed June 5, 2019).

Davis GE et al. 2020. Exploring movement patterns and changing distributions of baleen whales in the western North Atlantic using a decade of passive acoustic data. Global Change Biology:gcb.15191.

Foley HJ, Holt RC, Hardee RE, Nilsson PB, Jackson KA, Read AJ, Pabst DA, McLellan WA. 2011. Observations of a western North Atlantic right whale (Eubalaena glacialis) birth offshore of the protected southeast U.S. critical habitat. Marine Mammal Science **27**:E234–E240.

Forney KA, Barlow J, Carretta JV. 1995. The abundance of cetaceans in California waters. Part 2: Aerial surveys in winter and spring of 1991 and 1992. Fishery Bulletin **93**:15–26.

Ganley L, Brault S, Mayo C. 2019. What we see is not what there is: estimating North Atlantic right whale Eubalaena glacialis local abundance. Endangered Species Research **38**:101–113.

Garrison LP. 2006. Cruise Results: NOAA Ship Gordon Gunter Cruise GU-06-03 (038). Page 27. Mississippi Laboratories, NMFS, NOAA, U.S. Dept. of Commerce, Pascagoula, Mississippi. Available from http://www.concurinc.com/PLTRT/Nov2006/8.%20GU-06-03%20CruiseReport.pdf (accessed November 3, 2016).

Garrison LP, Martinez A, Maze-Foley K. 2010. Habitat and abundance of cetaceans in Atlantic Ocean continental slope waters off the eastern USA. Journal of Cetacean Research and Management **11**:267–277.

Good CP. 2008. Spatial ecology of the North Atlantic right whale (Eubalaena glacialis).

Gowan TA, Ortega-Ortiz JG. 2014. Wintering Habitat Model for the North Atlantic Right Whale (Eubalaena glacialis) in the Southeastern United States. PLoS ONE **9**:e95126.

Gowan TA, Ortega-Ortiz JG, Hostetler JA, Hamilton PK, Knowlton AR, Jackson KA, George RC, Taylor CR, Naessig PJ. 2019. Temporal and demographic variation in partial migration of the North Atlantic right whale. Scientific Reports **9**:353.

Hain JHW, Ellis SL, Kenney RD, Slay CK. 1999. Sightability of right whales in coastal waters of the southeastern United States with implications for the aerial monitoring program. Pages 191–207 Marine Mammal Survey and Assessment Methods. Balkerma, Rotterdam.

Hansen RG, Boye TK, Larsen RS, Nielsen NH, Tervo O, Nielsen RD, Rasmussen MH, Sinding MHS, Heide-Jørgensen MP. 2019. Abundance of whales in West and East Greenland in summer 2015. NAMMCO Scientific Publications **11**. Available from https://septentrio.uit.no/index.php/NAMMCOSP/article/view/4689 (accessed April 17, 2019).

Hastie T, Tibshirani R. 1990. Generalized Additive Models. Chapman and Hall, London, UK.

Hedley SL, Buckland ST. 2004. Spatial models for line transect sampling. Journal of Agricultural, Biological, and Environmental Statistics **9**:181–199.

Heide-Jørgensen MP et al. 2012. Rate of increase and current abundance of humpback whales in West Greenland. The journal of cetacean research and management **12**.

Heide-Jørgensen MP, Witting L, Laidre KL, Hansen RG, Rasmussen M. 2010. Fully corrected estimates of common minke whale abundance in West Greenland in 2007. Journal of Cetacean Research and Management **11**:75–82.

Hodge KB, Muirhead CA, Morano JL, Clark CW, Rice AN. 2015. North Atlantic right whale occurrence near wind energy areas along the mid-Atlantic US coast: implications for management. Endanger Species Res **28**:225–234.

Kenney RD. 2019. The North Atlantic Right Whale Consortium Database: A Guide for Users and Contributors Version 5. Page 152. University of Rhode Island, Narragansett, RI.

Kenney RD, Mayo CA, Winn HE. 2001. Migration and foraging strategies at varying spatial scales in western North Atlantic right whales: a review of hypotheses. Journal of Cetacean Research and Management **2**:251–260.

Kraus SD et al. 2016. Northeast Large Pelagic Survey Collaborative Aerial and Acoustic Surveys for Large Whales and Sea Turtles, OCS Study BOEM 2016-054. US Department of the Interior, Bureau of Ocean Energy Management, Sterling, Virginia. Available from https://www.boem.gov/RI-MA-Whales-Turtles/ (accessed November 4, 2016).

Laake JL, Calambokidis J, Osmek SD, Rugh DJ. 1997. Probability of Detecting Harbor Porpoise From Aerial Surveys: Estimating g(0). Journal of Wildlife Management **61**:63–75.

Leiter S, Stone K, Thompson J, Accardo C, Wikgren B, Zani M, Cole T, Kenney R, Mayo C, Kraus S. 2017. North Atlantic right whale Eubalaena glacialis occurrence in offshore wind energy areas near Massachusetts and Rhode Island, USA. Endangered Species Research **34**:45–59.

Mallette SD, Lockhart GG, McAlarney RJ, Cummings EW, McLellan WA, Pabst DA, Barco SG. 2014. Documenting Whale Migration off Virginia's Coast for Use in Marine Spatial Planning: Aerial and Vessel Surveys in the Proximity of the Virginia Wind Energy Area (VA WEA), VAQF Scientific Report 2014-08. Page 89. Virginia Aquarium & Marine Science Center Foundation, Virginia Beach, VA.

Mallette SD, Lockhart GG, McAlarney RJ, Cummings EW, McLellan WA, Pabst DA, Barco SG. 2015. Documenting Whale Migration off Virginia's Coast for Use in Marine Spatial Planning: Aerial Surveys in the Proximity of the Virginia Wind Energy Area (VA WEA) Survey/Reporting Period: May 2014 - December 2014, VAQF Scientific Report 2015-02. Page 49. Available from http://www.deq.virginia.gov/Portals/0/DEQ/CoastalZoneManagement/FundsInitiativesProjects/task95-02-13.pdf (accessed November 3, 2016).

Mallette SD, McAlarney RJ, Lockhart GG, Cummings EW, Pabst DA, McLellan WA, Barco SG, Aquarium V. 2017. Aerial Survey Baseline Monitoring in the Continental Shelf Region of the VACAPES OPAREA: 2016 Annual Progress Report. Page 100. Virginia Aquarium & Marine Science Center Foundation, Virginia Beach, VA.

Mannocci L et al. 2017a. Temporal resolutions in species distribution models of highly mobile marine animals: Recommendations for ecologists and managers. Diversity and Distributions **23**:1098–1109.

Mannocci L, Roberts JJ, Miller DL, Halpin PN. 2017b. Extrapolating cetacean densities to quantitatively assess human impacts on populations in the high seas. Conservation Biology **31**:601–614.

Marques TA, Thomas L, Martin SW, Mellinger DK, Ward JA, Moretti DJ, Harris D, Tyack PL. 2013. Estimating animal population density using passive acoustics: Passive acoustic density estimation. Biological Reviews **88**:287–309.

McAlarney R, Cummings E, McLellan W, Pabst A. 2018. Aerial Surveys for Protected Marine Species in the Norfolk Canyon Region: 2017 Annual Progress Report. Page 72. University of North Carolina Wilmington, Wilmington, NC.

McLellan WA, McAlarney RJ, Cummings EW, Read AJ, Paxton CGM, Bell JT, Pabst DA. 2018. Distribution and abundance of beaked whales (Family Ziphiidae) Off Cape Hatteras, North Carolina, U.S.A.: DISTRIBUTION AND ABUNDANCE OF BEAKED WHALES. Marine Mammal Science. Available from http://doi.wiley.com/10.1111/mms.12500 (accessed May 8, 2018).

Miller DL, Burt ML, Rexstad EA, Thomas L. 2013. Spatial models for distance sampling data: recent developments and future directions. Methods in Ecology and Evolution **4**:1001–1010.

Moore JC, Clark E. 1963. Discovery of Right Whales in the Gulf of Mexico. Science **141**:259.

Nieukirk S. 1992, December 7. Satellite Monitored Dive Characteristics of the Northern Right Whale. Eubalaena glacialis. Master's Thesis. Oregon State University. Available from http://ir.library.oregonstate.edu/xmlui/handle/1957/36164.

Pace RM, Corkeron PJ, Kraus SD. 2017. State-space mark-recapture estimates reveal a recent decline in abundance of North Atlantic right whales. Ecology and Evolution. Available from http://doi.wiley.com/10.1002/ece3.3406.

Palka DL. 2006. Summer abundance estimates of cetaceans in US North Atlantic navy operating areas. 06–03, Northeast Fish Sci Cent Ref Doc. U.S. Department of Commerce, Woods Hole, MA. Available from http://www.nefsc.noaa.gov/publications/crd/crd0603/crd0603.pdf (accessed March 5, 2014).

Palka DL et al. 2017. Atlantic Marine Assessment Program for Protected Species: 2010-2014 (OCS Study BOEM 2017-071). US Dept. of the Interior, Bureau of Ocean Energy Management, Washington, DC. Available from https://www.boem.gov/espis/5/5638.pdf.

Pershing AJ et al. 2015. Slow adaptation in the face of rapid warming leads to collapse of the Gulf of Maine cod fishery. Science **350**:809–812.

Pettis HM, Hamilton PK. 2014. North Atlantic Right Whale Consortium 2014 annual report card. Page 9. North Atlantic Right Whale Consortium, Boston, MA. Available from https://www.narwc.org/uploads/1/1/6/6/116623219/2014_report_card.pdf.

Quintana E, Kraus S, Baumgartner M. 2018. Megafauna Aerial Surveys in the Wind Energy Areas of Massachusetts and Rhode Island with Empashis on Large Whales: Preliminary Draft Submitted October 2018. Page 61. New England Aquarium, Boston, MA.

Read AJ et al. 2014. Occurrence, distribution and abundance of cetaceans in Onslow Bay, North Carolina, USA. Journal of Cetacean Research and Management **14**:23–35.

Record N et al. 2019. Rapid Climate-Driven Circulation Changes Threaten Conservation of Endangered North Atlantic Right Whales. Oceanography **32**. Available from https://tos.org/oceanography/article/rapid-climate-driven-circulation-changes-threaten-conservation-of-endangere (accessed July 16, 2020).

Roberts JJ. 2015. Estimates of bottlenose dolphin density for estuaries in the AFTT area for the Phase III NMSDD, Document version 1.2. Duke University Marine Geospatial Ecology Lab, Durham, NC.

Roberts JJ et al. 2016a. Habitat-based cetacean density models for the U.S. Atlantic and Gulf of Mexico. Scientific Reports **6**:22615.

Roberts JJ, Mannocci L, Halpin PN. 2015. Marine mammal density models for the U.S. Navy Atlantic Fleet Training and Testing (AFTT) study area for the Phase III Navy Marine Species Density Database (NMSDD), Document Version 1.1. Duke University Marine Geospatial Ecology Lab, Durham, NC.

Roberts JJ, Mannocci L, Halpin PN. 2016b. Final Project Report: Marine Species Density Data Gap Assessments and Update for the AFTT Study Area, 2015-2016 (Base Year), Document Version 1.0. Page 21. Duke University Marine Geospatial Ecology Lab, Durham, NC.

Roberts JJ, Mannocci L, Halpin PN. 2017. Final Project Report: Marine Species Density Data Gap Assessments and Update for the AFTT Study Area, 2016-2017 (Opt. Year 1), Document Version 1.4. Page 87. Duke University Marine Geospatial Ecology Lab, Durham, NC.

Roberts JJ, Mannocci L, Schick RS, Halpin PN. 2018. Final Project Report: Marine Species Density Data Gap Assessments and Update for the AFTT Study Area, 2017-2018 (Opt. Year 2), Document Version 1.2. Page 114. Duke University Marine Geospatial Ecology Lab, Durham, NC.

Roberts JJ, Miller DL, Becker EA, Bouchet PJ, Halpin PN, Thomas L. 2019. Questions Frequently Asked by Marine Mammal Density Surface Modelers. Available from https://doi.org/10.17605/OSF.IO/5EZA8 (accessed June 1, 2020).

Robertson FC, Koski WR, Brandon JR, Thomas TA, Trites AW. 2015. Correction factors account for the availability of bowhead whales exposed to seismic operations in the Beaufort Sea:10.

Salisbury DP, Clark CW, Rice AN. 2015. Right whale occurrence in the coastal waters of Virginia, U.S.A.: Endangered species presence in a rapidly developing energy market. Marine Mammal Science:n/a-n/a.

Stone KM, Leiter SM, Kenney RD, Wikgren BC, Thompson JL, Taylor JKD, Kraus SD. 2017. Distribution and abundance of cetaceans in a wind energy development area offshore of Massachusetts and Rhode Island. Journal of Coastal Conservation **21**:527–543.

Taylor JKD, Kenney RD, LeRoi DJ, Kraus SD. 2014. Automated Vertical Photography for Detecting Pelagic Species in Multitaxon Aerial Surveys. Marine Technology Society Journal **48**:36–48.

Tetra Tech, LGL. 2019. Year 2 Annual Survey Report for New York Bight Whale Monitoring Aerial Surveys, March 2018 - February 2019. Page 192. New York State Department of Environmental Conservation, East Setauket, NY. Available from http://www.dec.ny.gov/docs/fish_marine_pdf/mmaeran2.pdf.

Tetra Tech, Smultea Sciences. 2018. Year 1 Annual Survey Report for New York Bight Whale Monitoring Aerial Surveys, March 2017 - February 2018. Page 115. New York State Department of Environmental Conservation, East Setauket, NY. Available from http://www.dec.ny.gov/docs/fish_marine_pdf/mmaeran1.pdf.

Thomas AC, Townsend DW, Weatherbee R. 2003. Satellite-measured phytoplankton variability in the Gulf of Maine. Continental Shelf Research **23**:971–989.

Thomas L, Buckland ST, Rexstad EA, Laake JL, Strindberg S, Hedley SL, Bishop JRB, Marques TA, Burnham KP. 2010. Distance software: design and analysis of distance sampling surveys for estimating population size. Journal of Applied Ecology **47**:5–14.

Torres LG, Mclellan WA, Meagher E, Pabst DA. 2005. Seasonal distribution and relative abundance of bottlenose dolphins, Tursiops truncatus, along the US mid-Atlantic coast. Journal of Cetacean Research and Management **7**:153.

U.S. Department of the Navy. 2017. U.S. Navy Marine Species Density Database Phase III for the Atlantic Fleet Training and Testing Area: Final Technical Report. Page 293. Naval Facilities Engineering Command Atlantic, Norfolk, VA.

Virgili A et al. 2019. Combining multiple visual surveys to model the habitat of deep-diving cetaceans at the basin scale: Large-scale modelling of deep-diving cetacean habitats. Global Ecology and Biogeography **28**:300–314.

Watwood SL, Miller PJ, Johnson M, Madsen PT, Tyack PL. 2006. Deep-diving foraging behaviour of sperm whales (Physeter macrocephalus). Journal of Animal Ecology **75**:814–825.

Weinrich MT, Kenney RD, Hamilton PK. 2000. Right whales (Eubalaena glacialis) on Jeffreys Ledge: A habitat of unrecognized importance? Marine Mammal Science **16**:326–337.

Whitt A, Dudzinski K, Laliberté J. 2013. North Atlantic right whale distribution and seasonal occurrence in nearshore waters off New Jersey, USA, and implications for management. Endangered Species Research **20**:59–69.

Whitt AD, Powell JA, Richardson AG, Bosyk JR. 2015. Abundance and distribution of marine mammals in nearshore waters off New Jersey, USA:16.

Wikgren B, Kite-Powell H, Kraus S. 2014. Modeling the distribution of the North Atlantic right whale Eubalaena glacialis off coastal Maine by areal co-kriging. Endangered Species Research **24**:21–31.

Winn HE, Goodyear JD, Kenney RD, Petricig RO. 1995. Dive patterns of tagged right whales in the Great South Channel. Continental Shelf Research **15**:593–611.

Wood S. 2006. Generalized additive models: an introduction with R. CRC press.