Final Project Report: Marine Species Density Data Gap Assessments and Update for the AFTT Study Area, 2020 (Opt. Year 4)

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 2.2 – 2021-11-22

This document should be cited:

Roberts JJ, Schick RS, Halpin PN (2021) Final Project Report: Marine Species Density Data Gap Assessments and Update for the AFTT Study Area, 2020 (Option Year 4). Document version 2.2. Report prepared for Naval Facilities Engineering Command, Atlantic by the Duke University Marine Geospatial Ecology Lab, Durham, NC.

1. Introduction

The U.S. Navy is responsible for compliance with a suite of federal environmental and natural resources laws and regulations intended to conserve marine ecosystems and maintain populations of marine species. Such species include marine mammals such as cetaceans and pinnipeds, which are designated for special protection by the Marine Mammal Protection Act (MMPA). Over the past decade, to assist in environmental planning and in compliance with the MMPA and other relevant laws and regulations, the Navy has funded the Marine Geospatial Ecology Laboratory (MGEL) at Duke University to develop *density surface models* (Hedley & Buckland 2004; Miller et al. 2013) for extant species of marine mammals for several regions.

Density surface models relate the absolute density (individual animals per square kilometer) of species of interest observed during line transect surveys to ecologically relevant oceanographic covariates such as sea surface temperature. After models are fitted, they can be used to predict maps of species density from maps of the covariates. Using surveys conducted by collaborating colleagues at the National Oceanic and Atmospheric Administration's (NOAA) National Marine Fisheries Service (NMFS) and several academic and independent research institutions, we developed and published a foundational suite of density models and maps for the U.S. east coast and Gulf of Mexico (Roberts et al. 2016a) and the Navy's wider Atlantic Fleet Training and Testing (AFTT) area (Mannocci et al. 2017) for over 30 taxa. The Navy then applied the resulting density maps to the development of its AFTT Phase III Environmental Impact Statement (EIS), which documented effects on marine mammal populations predicted to occur during training and testing activities planned for the period November 2018 - November 2025.

Following that founding effort, the Navy initiated the five-year 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 density models and maps using newly available data and methodology, for use in future Navy environmental planning and compliance actions. This culminated in a series of updated models and associated reports for the first four years of the project, known as the "Base Year" (Roberts et al. 2016b) and "Option Years 1-3" (Roberts et al. 2017, 2018, 2020). The focus of Option Years 3 and 4, spanning parts of calendar years 2018-2020, was to implement improvements to our modeling methodology identified during Option Years 1 and 2, and then update a single density model as a means to test the new methods and data. We selected the North Atlantic right whale (NARW) as the species to be updated, as it is one of most critical conservation concerns within the AFTT study area. This effort dovetailed with a strong interest from NOAA Fisheries, which sought specific model improvements to assist with its management

of fishery interactions with right whales. During Option Year 3 and the first half of Option Year 4, we completed development of this new model, known as NARW model version 9 (Roberts et al. 2020). Here, we document the completion of Option Year 4, during which we made two minor updates to the model based on feedback from NOAA Fisheries and the Massachusetts Division of Marine Fisheries (Mass DMF), resulting in versions 10 and 11, respectively, and generated model uncertainty estimates, an important output not able to be included in version 9.

Although the current five-year cooperative agreement ends with this report, the Navy has extended the project via a follow-on cooperative agreement (#N62470-20-2-2011) that ends in 2022. The objective of that agreement is to produce a full suite of final density models for all extant marine mammal species inhabiting the AFTT study area, which the Navy will then utilize in its AFTT Phase IV EIS. We expect our ongoing series of reports to continue through the end of that project, and potentially further if the Navy opts to initiate a new cooperative agreement for the AFTT Phase V EIS.

2. Timeline of the North Atlantic Right Whale Density Model Revisions

The first NARW model we released for management use was version 5, produced for the Phase III Navy Marine Species Density Database (NMSDD) in 2015. That model summarized density over 1995-2014 using surveys that spanned that period. It was eventually published as part of Roberts et al. (2016a), by which time the version number had increased to 5.6, marking incremental changes to the documentation but no change to the model or its density predictions of version 5.

Following that, the Navy initiated the current five-year cooperative agreement to build the Phase IV density models. We developed and released version 7 in Option Year 1 of that project (Roberts et al. 2017). Version 7 incorporated a large quantity of additional data, including the AMAPPS surveys from 2010-2015 and the southeast NARW Early Warning System (EWS) surveys of the 2003/04 - 2015/16 seasons. It summarized density over the period 1998-2016. In 2017, we released version 7 as an update to the NMSDD and to NOAA Fisheries and other interested parties.

In 2019, NOAA approached us about utilizing the NARW density model for fisheries management. A key requirement for NOAA was that density maps include Cape Cod Bay (CCB), Massachusetts, which we had omitted from our models, as the surveys from CCB lacked the perpendicular sighting distance data needed to utilize them under our methodology. In February 2019, Laura Ganley et al. (2019) published abundance estimates for CCB after completing a full reprocessing of all CCB surveys since 1998 to develop the data needed for density and abundance estimation. With Ganley's collaboration, we derived density estimates for CCB from her published abundance estimates and used them to "fill in" the area excluded from our model, comprising the 1600 km^2 of cells of our map grid corresponding to CCB. Simultaneously, we adjusted several of the model's internal geographic boundaries to reduce edge effects in the final predictions. We released this model, known as version 8, in April 2019. Because there were no changes from the version 7 model to densities in the Navy's core areas of interest, we did not update the NMSDD at the time, but released the model directly to NOAA and others interested in the update for other purposes. We did not release new documentation; aside from the CCB and minor boundary changes, version 8 did not add data or make any methodological changes over version 7, and density predictions outside CCB were essentially the same.

After spending much of 2019 assimilating a large quantity of additional survey data from our collaborators, we initiated a full rebuild of the right whale model using this expanded database of surveys and several methodological improvements. This expansion extended our database up through early 2019, and introduced important surveys from new collaborators, including the 2011-2015 Northeast Large Pelagic Survey Cooperative (NLPSC) surveys and the 2017-2018 Marine Mammal Surveys of the Wind Energy Areas (MMS-WEA) from New England Aquarium, as well as the 2017-2018 New York Bight Whale Monitoring Program (NYBWM) surveys from the New York State Department of Environmental Conservation. This work culminated in the May 2020 release of version 9 of the right whale model (Roberts et al. 2020), which was jointly funded under this cooperative agreement and a grant from NOAA Fisheries. Version 9 modeled density from April 2003 through March 2019. At NOAA's request, we summarized this period into two eras: "2003-2009" (i.e., April 2003 - March 2010), representing right whale density before the population entered decline, and "2010-2018" (i.e., April 2010 - March 2019), representing density after the decline started (Pace et al. 2017) and major shifts in the species' distribution began to occur (Davies et al. 2015; Davis et al. 2017; Meyer-Gutbrod et al. 2021).

By this time, the Navy had shifted its deadline for the final Phase IV AFTT model updates to early 2022, and we anticipated that the right whale model would be revised again prior to that deadline, so we did not release an NMSDD update for the version 9 model, but instead released it directly to NOAA and others interested in right whales. We did not anticipate revising the density surfaces until we had incorporated additional survey data. However, feedback from NOAA and Mass DMF prompted two refinements before new data could be introduced, yielding model versions 10 and 11, released in October 2020 and February 2021, respectively. While developing versions 10 and 11, we developed new statistical methods for characterizing model uncertainty, in a methodology-oriented collaboration with the "DenMod" team (https://synergy.st-andrews.ac.uk/denmod/). This facilitated production of version 11.1, which left the density surfaces unchanged from version 11 but added surfaces characterizing model uncertainty.

We are releasing version 11.1 concurrent with this report, which serves as the documentation for versions 10, 11, and 11.1. Each is documented in a subsequent section below. We have provided those models to NOAA and made them available to all interested parties but will hold off on updating the NMSDD until 2022. By that time, we will have accumulated a final batch of survey data for collaborators to be used for the final model update for the AFTT Phase IV EIS. We plan to complete that update in February 2022, and plan to designate it version 12. It is likely that this update will extend the model's temporal range into 2020.

3. Version 10 Model

3.1. Version 9's spring prediction for Massachusetts Bay

In September 2010, Burton Shank of the NOAA Northeast Fisheries Science Center (NEFSC) contacted us about densities predicted by the version 9 model that appeared to be anomalously high in spring months in the vicinity of Massachusetts Bay (Figure 1). Although Massachusetts Bay is a relatively small, inshore area not currently of high interest to the Navy, it encloses the city of Boston, a highly populated area with a large port that is heavily trafficked year-round and therefore of high interest in NOAA management scenarios. Predictions here represented a spatial extrapolation; during the entire 2003-2018 period, none of the surveys available to us had ever fully covered the area.

Although the model predicted high density in Massachusetts Bay, we believed the prediction was generally reasonable based in part on the report of a large aggregation of whales observed by the NEFSC North Atlantic Right Whale Sighting Survey (NARWSS) program in spring 2018. In the data available to us, this aggregation was first reported 21 April 2018 in a directed survey of the nearshore region between Cape Cod Bay and Massachusetts Bay (Figure 2, left). This survey was flown after receiving a report of whales in the area and was conducted in a non-systematic manner, with the objective of finding and photographing all whales in the aggregation. In the data we received, 69 individuals were reported in total, a sizable fraction of the population. Unfortunately, we cannot use directed effort such as this in our model, as its spatiotemporal distribution lacks a sufficient degree of independence from the underlying distribution of the animals. (We did retain a systematic sawtooth survey of Stellwagen Bank conducted later that day. The Stellwagen survey also reported several right whales, which all occurred on transects in the southeast of the surveyed area, closest to Massachusetts Bay.)

On 27 April, NARWSS conducted a long coastal transect from Cape Neddick, Maine to Cape Cod Bay (Figure 2, right). Although the configuration of this transect was less than optimal for density modeling purposes, by virtue of generally paralleling depth gradients and being partially motivated by knowledge of the aggregation observed a week earlier, we judged the inclusion of it to be a reasonable compromise that would allow the model to capture right whale use of the area while reducing the risk of bias that can occur when highly directed, non-systematic data are used. This transect penetrated the farthest into the interior of Massachusetts Bay of all surveys available to us over the 2003-2018 period, so it was particularly valuable for modeling density in this region.

When we extrapolated the resulting model across the full extent of Massachusetts Bay—an area requested by NOAA for their fishery management applications—the model predicted high density in the months of March, April, and May and near-zero density in June (Figure 3 inset maps) and July (not shown). This pattern accorded well with maps of the 15 years of sightings available in NOAA's Right Whale Sighting Advisory System (SAS) database, which showed numerous sightings around Massachusetts Bay in March-May, with a peak in April, and only two sightings in June (Figure 3 main maps).

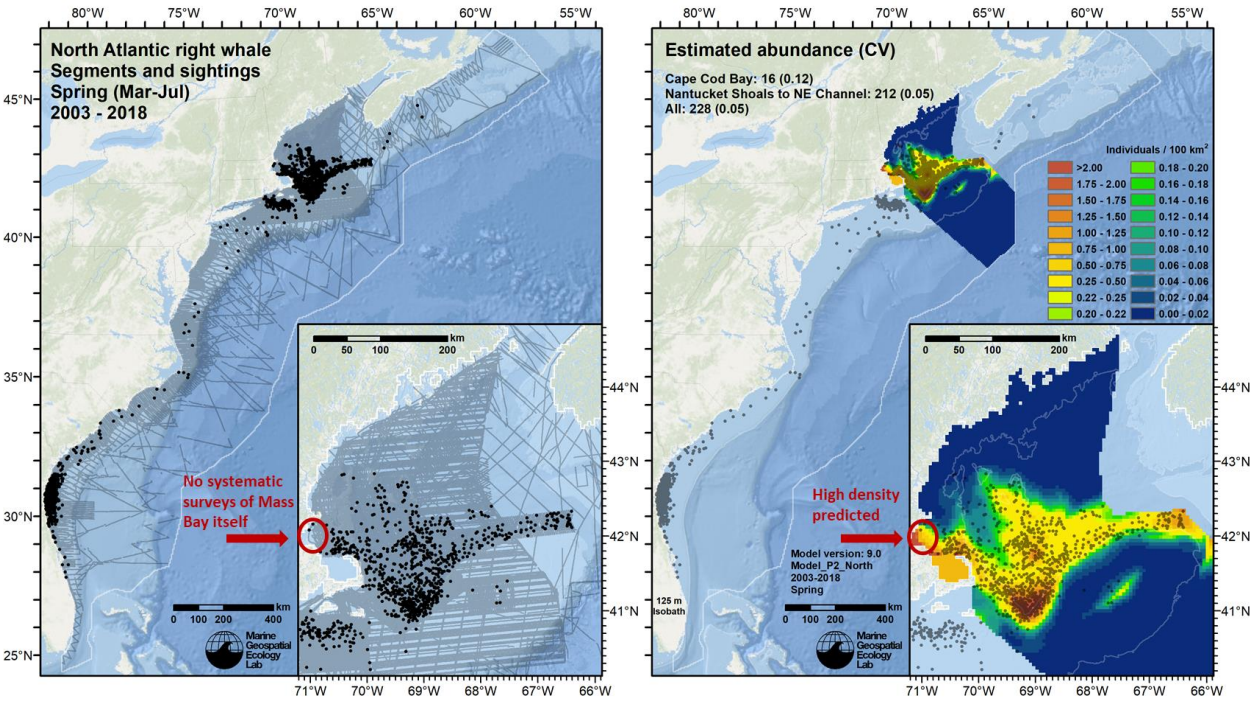


Figure 1. Left: Survey effort and sightings available for the "North of Nantucket Shoals" subregional density model for the "Spring" season, defined for this subregional model as March-July, over the 2003-2018 period. Right: Mean density prediction, averaged over all months and years of the time period. CVs shown here are underestimates (see Section 5.2.1for better estimates).

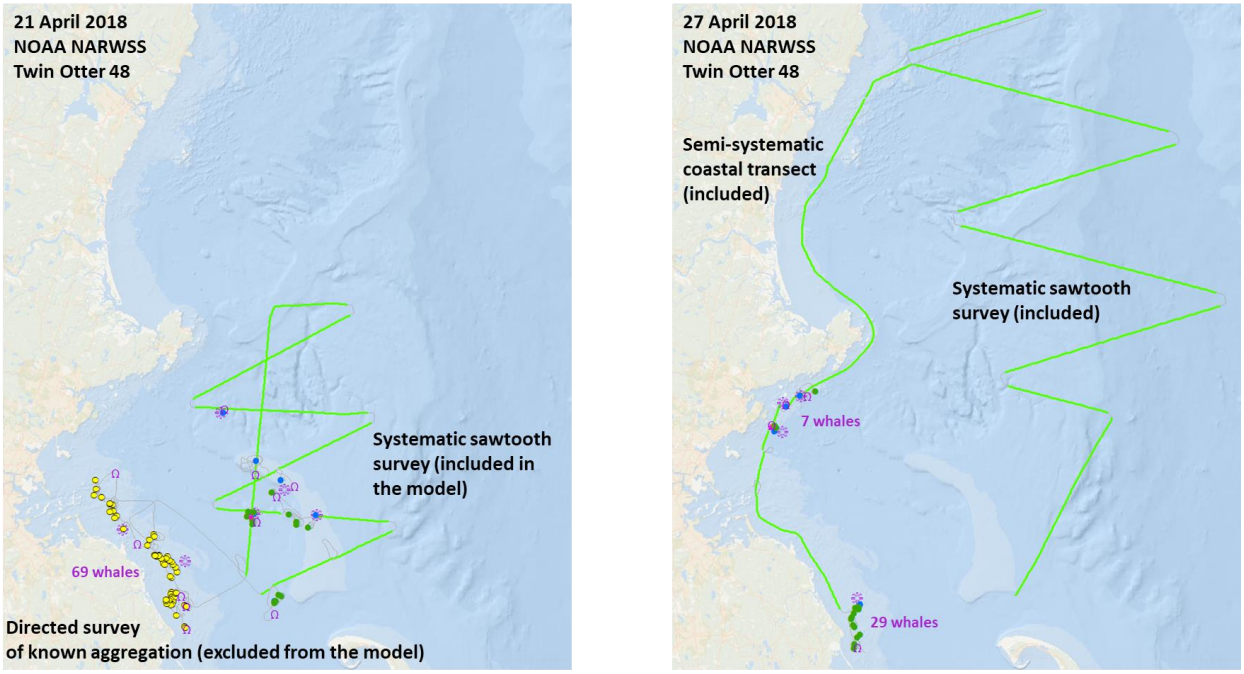


Figure 2. Surveys conducted by the NOAA NEFSC NARWSS aerial survey program that sighted large numbers of right whales in the vicinity of Massachusetts Bay in April 2018.

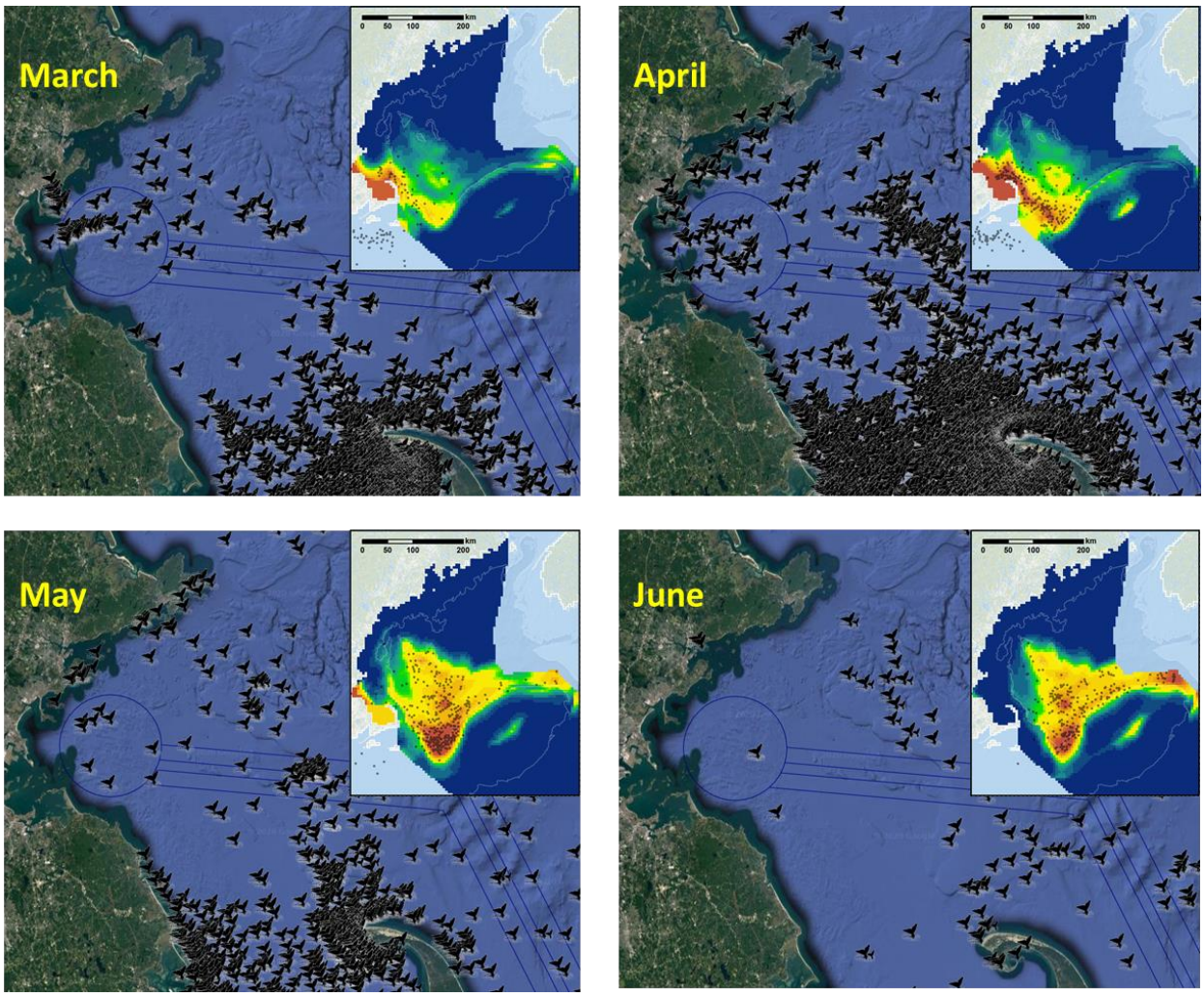


Figure 3. Sightings in NOAA's Right Whale SAS database, 2005-2020, for four spring months. The SAS database includes both systematic survey data, such as those used in our density model (outside CCB) and Ganley et al. (2019) (inside CCB), and opportunistic sightings, such as those reported by surveys we cannot use and by non-scientific sources, such as the Coast Guard and maritime sailors. Because the SAS database includes opportunistic sightings and does not reflect observation effort, it can be safely used to note the presence of whales but not infer their absence nor their density. Insets show corresponding predictions from the version 9 density model, which accounts for observation effort and corrects for differences in detection capability across survey platforms and programs (see Roberts et al. (2020) for a full description).

The SAS database combines opportunistic sightings with sightings from systematic surveys, and does not represent or account for observation effort, so it cannot be reliably used to infer the density of whales, the frequency of their presence, or even that they are absent. (A lack of sightings might mean they were absent, or it might mean that nobody looked for them there at that time.) However, it seemed likely that Massachusetts Bay, with its proximity to Boston, would be subject to a steady amount of opportunistic effort, and thus seemed plausible that the large drop in sightings in SAS from April to June could depict a real pattern. Furthermore, a similar drop was also observed for Cape Cod Bay just to the south, which had been systematically surveyed in January-May by the Center for Coastal Studies (CCS) for over two decades. Thus, after observing that the monthly changes in density extrapolated by the version 9 model were consistent with these other sources of information, we considered the predictions plausible.

Unfortunately, after making the final adjustments to the regional model that predicted density for Massachusetts Bay for these months, we inadvertently skipped some of our diagnostic checks. One of these adjustments, in which we placed a maximum limit on a covariate relationship to try to control a dubious early-season hotspot in the eastern

part of the modeled region, yielded a refitted model that extrapolated extreme densities for the innermost pixels of Massachusetts Bay, which reached over 35 individuals / 100 km^2 at the shallowest pixel (Figure 4). We failed to notice this because we inadvertently skipped our usual check for extreme predictions, and only checked the final prediction maps, which looked ok because in order to to maximize discrimination of realistic density values, the highest color class on the final maps is only 2 individuals / 100 km^2. This value is exceeded only rarely in our experience modeling this species.[1] Thus, using only that map, we could not detect the extreme values. Fortunately, Burton Shank (NOAA NEFSC) detected the problem and asked us to investigate it.

We traced the problem to the "distance to the 125 m isobath" (DistTo125m) covariate (Figure 5). This has almost always been an important covariate in models we have developed for right whales inhabiting the Gulf of Maine in spring. As usual, it estimated higher density on the deeper side of the isobath, indicative of right whales feeding in the Great South Channel and along the northern edge of Georges Bank, and a decrease in density in shallower waters, as well as in distant offshore waters down the slope. However, in shallow waters farthest up from the isobath, the relationship reversed into an upward trend (Figure 5, left). This relationship resulted partly from the most inshore sightings on the 27 April 2018 survey, offshore of Marblehead, Massachusetts. It was reinforced by sightings of several whales at the center of Georges Bank (Figure 1, left), which were also similarly far up-slope from the 125 m isobath. However, we considered the resulting peak in density in Georges Bank (Figure 1, right) to itself be partially extrapolative, as most of the surveying that occurred over Georges Bank happened prior to 2008. (These are the long horizontal survey transects spanning Georges Bank, flown by the NEFSC NARWSS program in 2003-2007, when it was still flying "broadscale" surveys, vs. the more spatially-constrained "sawtooth" surveys flown in recent years.)

In any case, when the fitted model was predicted across the unsurveyed cells of Massachusetts Bay, which were all far inshore of the 125 m isobath (Figure 5, right), the DistTo125m relationship greatly boosted the densities of the most inshore cells. Although such a result was theoretically possible according to the available data, it greatly exceeded densities observed in areas such as Cape Cod Bay or the Great South Channel, which are believed by experts to be some of the highest. It was not plausible that such extreme densities would regularly occur just outside Boston Harbor without being noticed. Therefore, we considered this extreme prediction to be in error and sought to correct it.

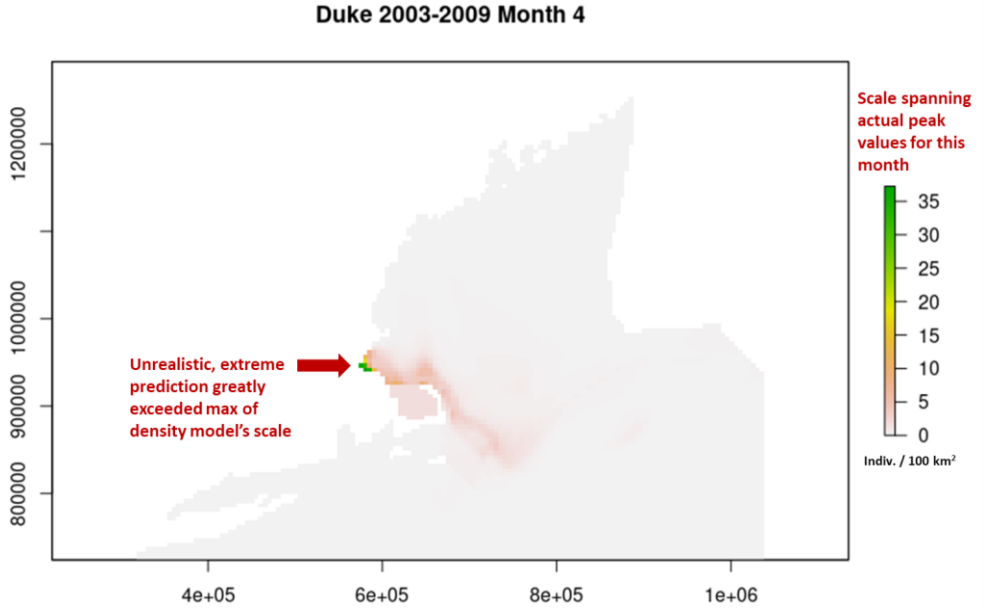


Figure 4. Map showing unrealistic, extreme predictions in the version 9 density model in the vicinity of Massachusetts Bay. Mean monthly density is shown for April of the 2003-2009 era. Note extreme values reaching to 35 individuals / 100 km2. Map courtesy of B. Shank (NOAA NEFSC).

[1] This is not to say that right whales never aggregate in densities higher than 2 whales / 100 km^2. They do. It is to say that the mean monthly density averaged over the model's summarization period of approximately10 years has rarely exceeded this value.

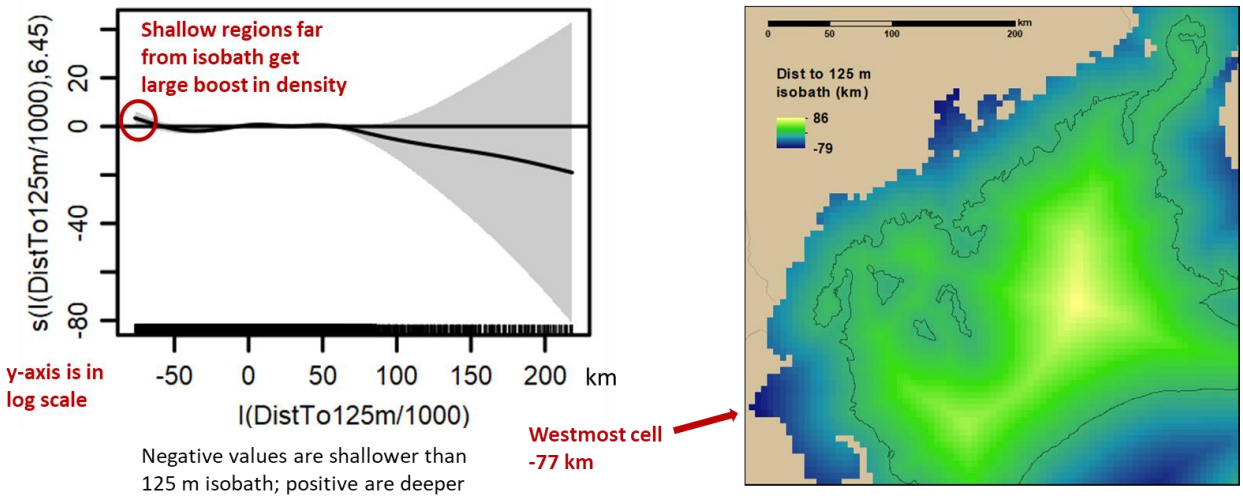


Figure 5. Left: Functional plot from the version 9 density model showing the additive contribution to density from the "distance to the 125 m isobath" (DistTo125m) covariate that caused the extrapolation problem. The y-axis, representing density, is in the natural log scale. The x-axis labels are in km, with negative values representing locations up-slope of the isobath and positive values representing values down-slope. Right: Map of the DistTo125m covariate for the wider Gulf of Maine region. Note the extreme negative values present for Massachusetts Bay, which result in the large density contribution in the circled area of the left plot. The black line shows the 125 m isobath itself.

3.2. Solutions explored

We explored three potential solutions, both independently and in combination. First, we adjusted the model formulation to limit the DistTo125m covariate to a minimum value of -55 km, corresponding to how far available surveys had traversed into Massachusetts Bay. This approach was conceptually similar to "Winsorizing" (Dixon 1960), a technique used to reduce the effect of possibly spurious outliers. Because sampling was so sparse at the extreme values of this covariate, we sought to group the extreme values and treat them similarly, rather than allowing the model to fit a tenuous increasing trend.

Second, we boosted the data available to the model by incorporating aerial surveys flown by the NEFSC NARWSS program in spring 2019. The version 9 model extended through March 2019, but we were able to incorporate two more months of data from NARWSS, allowing a revised model to extend through May 2019. Some of these additional surveys were flown in the vicinity of Massachusetts Bay, which we thought might help to control the extrapolation problem.

Finally, as the third potential solution, we fitted models that excluded the 27 April 2018 coastal transects that sighted right whales just offshore of Marblehead, on the basis that this transect was likely driving the DistTo125m relationship that yielded the extreme extrapolation. By removing it, we thought we might obtain a relationship that would extrapolate more realistically.

3.3. Results

After trying the three potential solutions in isolation and in combination, the best approach proved to be the combination of the first and second, i.e., to limit the DistTo125m covariate to -55 km or larger and to introduce the two more months of newly available aerial survey data from spring 2019. The third solution—to remove the 27 April 2018 transect that sighted right whales near Massachusetts Bay—proved to be too extreme. When we implemented it, either alone or in combination with the other two solutions, the resulting models predicted near-zero density across Massachusetts Bay at all times. This result was unrealistic, given the large aggregation observed in April 2018.

With solutions 1 and 2 implemented, the refitted model successfully eliminated the unrealistic extrapolation in Massachusetts Bay while maintaining a low-to-moderate density in the vicinity (Figure 6) that appropriately reflected the systematic surveying available for density modeling. Mean monthly density predictions for Massachusetts Bay across March-June showed the characteristic rise and fall present in the version 9 model, with a peak in April and a drop to near zero density in June (Figure 7). Mean abundance for the March-July period for the "2003-2018" era dropped by ~20 (Figure 6), reflecting the elimination of the unrealistic extrapolation.

As a side effect, density over central Georges Bank became more dispersed (Figure 6), reflecting the limitation we imposed on the minimum value of the DistTo125m covariate. Given the relatively sparse surveying conducted over Georges Bank since 2007, we do not have a strong reason to prefer either model's result over the other's in this area. However, we slightly prefer the new prediction on the basis that the old one displayed a narrower, more-intense patch directly at the center of Georges Bank, and we know of no such habitat feature that would attract right whales strongly to this area. This question could be resolved in the future by conducting more surveys over the full expanse of Georges Bank; however, the need to address it should be appropriately balanced against other questions competing for survey resources. Given the apparent low importance of central Georges Bank to right whales, other questions may take higher priority. For example, we would consider it much higher priority to conduct additional surveys over Nantucket Shoals in the wedge-shaped area of sparse surveying that extends out from Nantucket, between Great South Channel to the north and waters south of Nantucket to the west (Figure 1). The sparse effort in this "wedge", over the shallowest parts of Nantucket Shoals, is comparable to that of central Georges Bank, yet it seems much more likely that right whales would regularly inhabit it, or at least traverse it during migrations.

Finally, the incorporation of additional data from the spring of 2019 added a number of transects and sightings in waters south of New England. Given the importance of this area to many stakeholders, we refitted and re-predicted the model for this region, which resulted in increased density in this region (Figure 8), concentrated in spring months.

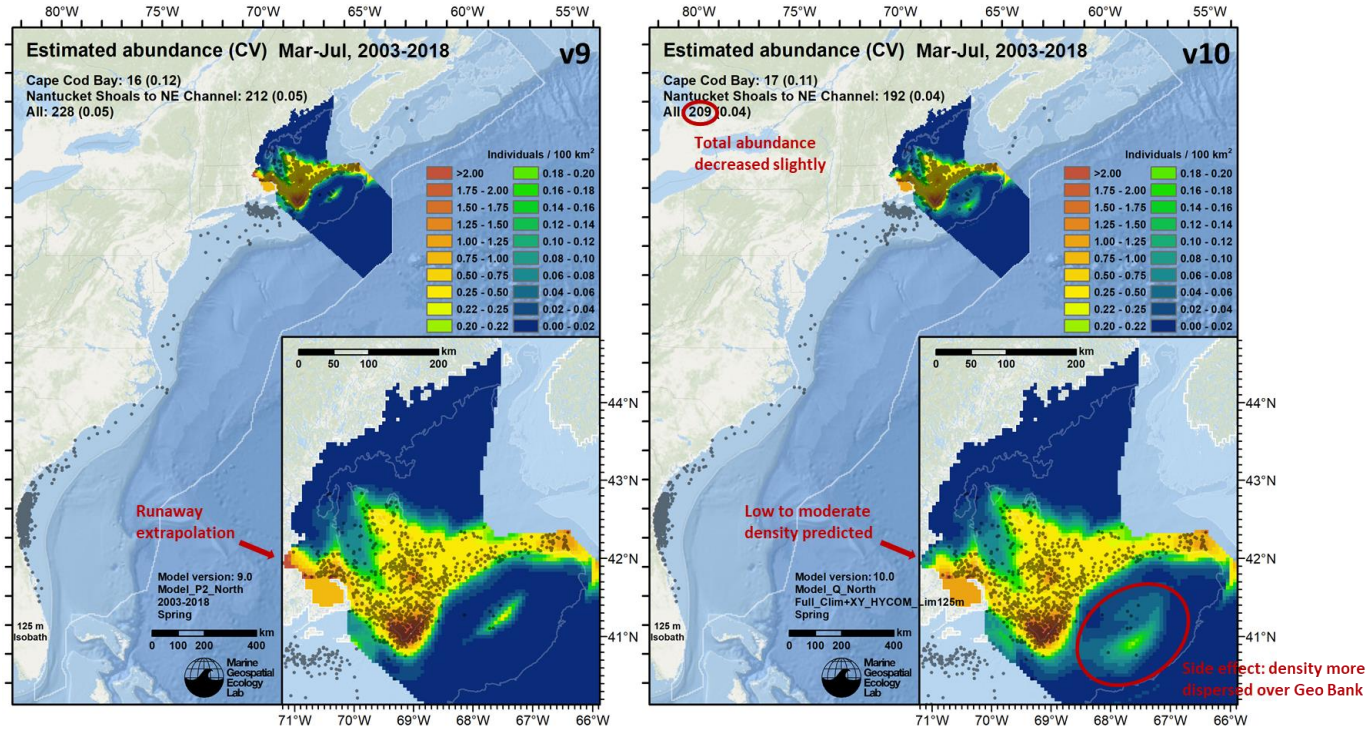


Figure 6. Comparison of version 9 model's March-July mean density prediction (left) showing the runaway extrapolation in Massachusetts Bay to the version 10 model's corresponding prediction (right), showing that low to moderate density is now predicted for Massachusetts Bay, resulting in a slight decrease in total abundance for the period, and a more dispersed density prediction over central Georges Bank. CVs shown here are underestimates (see Section 5.2.1 for better estimates). Predictions are for the 2003-2018 era, which in version 9 spanned April 2003-March 2019, while in version 10 spanned June 2003-May 2019.

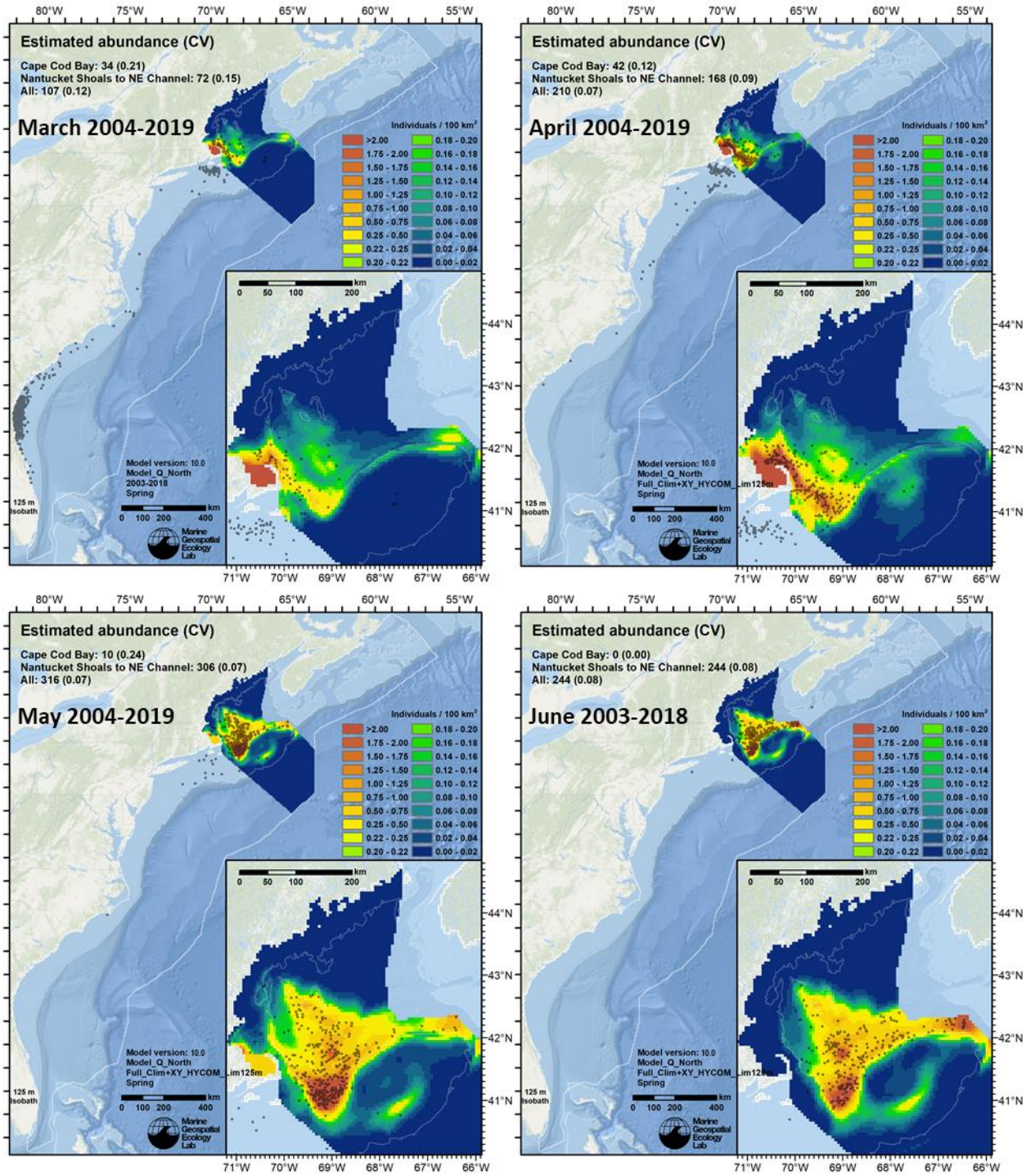


Figure 7. Predictions from model version 10 for March-June of the 2003-2018 era. Note that in Massachusetts Bay, density peaks in April and falls to near zero in June, as in version 9, but no longer reaches extreme values. CVs shown here are underestimates (see Section 5.2.1 for better estimates).

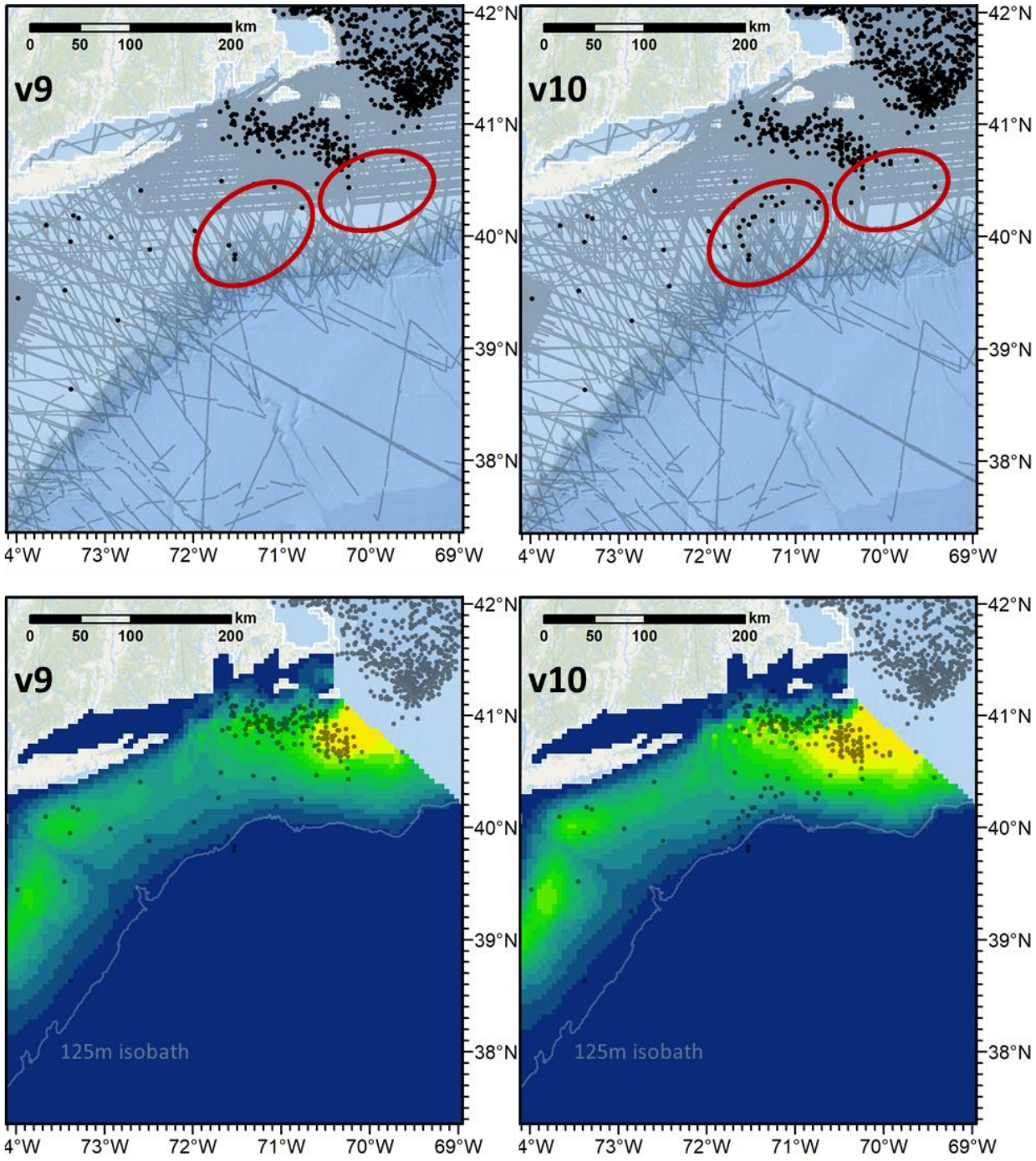


Figure 8. Survey effort and sightings (top) and predicted density (bottom) in the southern New England area for the version 9 (left) and 10 (right) models. Effort and sightings are from all months from January 2003 through May 2019. Predicted densities are year-round means for 2003-2019. Note the presence of additional sightings in circled areas, which occurred in spring 2019 and resulted in increased density in southern New England in version 10 (compare bottom right to bottom left).

We labelled the revised model version 10 and released it for management use to NOAA in October 2020. We did not produce an update to the NMSDD at this time, as the Navy indicated a preference that we incorporate additional data before providing a final NMSDD update in February 2022.

4. Version 11 Model

In early 2021, we were brought into a discussion between NOAA Fisheries and Mass DMF regarding the density estimates used for Cape Cod Bay (CCB). As discussed above, since the version 8 model, we had derived density estimates for CCB from abundance estimates produced by the Center for Coastal Studies (CCS) from the surveys they had flown there. However, CCS had focused its efforts on the months of January through May, encompassing the months that right whales were known to be present in CCB in appreciable numbers. For other months, had to derive estimates using other means.

Over the past two decades, right whales have been observed opportunistically in CCB in December at an increasing frequency, but CCS has not produced abundance estimates for this month because December surveys were only flown occasionally. In the version 8-10 models, we decided with our CCS collaborators that the best approach for estimating densities for December of a given year was to use the estimate for the immediately following January as a proxy, given the lack of data available for December and the possibility that right whales might have started to occupy CCB regularly in December, similar to how they started to occupy it in January in increasing numbers starting in 2011.

During discussions with NOAA regarding proposed fishery regulations, Mass DMF examined counts of sightings from extant December surveys flown by CCS and argued that the data suggested December densities were likely smaller than January's. NOAA raised the issue with us, and we agreed to examine the question again using all available data from CCS since 2003 (corresponding with the starting year of our density model), including still-tentative results from the active survey season up through January 2021. We concluded that the December surveys now available were sufficient for us to produce a better estimate than using January's as a proxy. The attached Appendix documents this investigation and the results. Our final estimates for CCB in December for the 2003-2009 era were 0.000 individuals / 100 km^2 vs. 0.050 in the version 9-10 models, and for the 2010-2018 era were 0.397 vs. 1.282 individuals / 100 km^2.

Although we now consider this a better approach to estimating density for December than our previous approach of using January as a proxy, these results rest on several important assumptions that cannot be validated without collecting additional data for December. They must therefore be interpreted and applied in species management decisions cautiously. All model users interested in right whale densities in CCB in December should carefully consider the caveats and assumptions discussed in the Appendix.

Following this effort, in February 2021 we incorporated the results into the density model, labelling this update version 11, and released it for management use to NOAA. (As with version 10, we did not produce an update to the NMSDD at this time.) The principal change from version 10 to 11 is the adjustment to density for Cape Cod Bay in December.

5. Version 11.1 Model

Version 11.1 does not change the density predictions from version 11; its focus is to provide additional outputs that characterize the uncertainty of the version 11 predictions. The main additions are maps of standard error (SE), the coefficient of variation (CV), and the lower (5%) and upper (95%) intervals of a 90% confidence interval. Also included are tables and plots of estimated total abundance that include CVs or 95% confidence intervals, and CVs that appear on version 11.1 maps have been updated from the tentative values shown on version 11 maps.

The version 11.1 uncertainty outputs account for two sources of variance: the statistical error in spatial model parameter estimates, and the variability in density predictions induced by dynamic covariates, i.e., through ephemeral, seasonal, and interannual variability in the environment. Hereafter, we refer to these two sources as *model uncertainty* and *environmental uncertainty*. In model versions 8 and earlier, our uncertainty outputs only accounted for model uncertainty. In versions 9-11, released while we were developing methods to also account for environmental uncertainty, we did not provide uncertainty outputs. Now, for the first time, in version 11.1 we are also accounting for environmental uncertainty. This is of particular importance for the North Atlantic right whale, which has recently exhibited extreme interannual variability in its distribution (Davis et al. 2017; Ganley et al. 2019; Record et al. 2019; Davies et al. 2019; Quintana-Rizzo et al. 2021; Meyer-Gutbrod et al. 2021). We also plan to apply these methods to

all species modeled for the Navy's AFTT Phase IV project. These models are currently under development and scheduled for completion in February 2022.

In this section we describe the methods used to produce the version 11.1 uncertainty estimates, present the results, and discuss their use and caveats. Finally, we note that these results do not account for all known sources of uncertainty and therefore likely underestimate true variance. We conclude with a discussion of other sources of uncertainty and the prospects for accounting for them in the future.

5.1. Methods

In this subsection, we first provide background material on procedures for fitting and predicting density surface models, which lead up to the uncertainty estimation problem. For full details on how the version 11 right whale model was fitted and predicted, please see Roberts et al. (2020). We then describe the method used to summarize the time series of model predictions into a smaller collection of summary outputs requested by model users. With this, we discuss the estimation and summarization of the two sources of variance that collectively represent uncertainty for the version 11 model. Finally, we briefly discuss other sources of variance that might be incorporated in the future.

5.1.1. Fitting and predicting density surface models

A *density surface model* (Hedley & Buckland 2004; Miller et al. 2013) is a mathematical equation for estimating a species' density—the number of individual animals present per unit area—from associated *covariates* found to correlate with density, such as water temperature. The modeler *fits* (or *trains*) a model by applying statistical algorithms to *training data*. The fitting process determines the structure of the mathematical equation and optimizes the values of the coefficients and other parameters within it. Typically, the training data are a collection of short segments (1-25 km long) obtained by subdividing transects of aerial or shipboard line-transect surveys. Each segment tallies the number of animals sighted and contemporaneous observations describing sighting conditions, such as the weather and sea state. Using these data, the modeler quantifies the surface area of the ocean that was effectively observed during each segment, and then estimates density as the count of animals divided by the area observed. Additional corrections may be applied to account for animals that were likely to be missed by the observation team, either because they were under the water or because the species is simply difficult to detect.

Next, the modeler obtains for each segment the values of the covariates that will be tested during the fitting process to see how well they correlate with density. To obtain them, the modeler usually overlays the segments on maps of covariates produced from satellite sensors or ocean models and extracts the underlying values. Through a systematic, iterative process, the modeler attempts to discover strong correlations between density and the covariates, discarding any that correlate poorly. Once a candidate selection of covariates is identified, the modeler constructs and optimizes the mathematical equation that optimally expresses the correlation. Often, the modeler fits a collection of several such candidate models, each with different combinations of covariates, and ranks them using goodness-of-fit statistics and other criteria before selecting a final model.

Once the final model is ready, it may be used to make predictions. To do so, the modeler plugs a value for each covariate into the mathematical equation and solves it, yielding a predicted value of density for the observed values of the covariates. To predict a map, known as a *density surface*, the modeler obtains a map for each covariate and iteratively predicts across the geographic area of interest, using the covariates' values at each location in turn.

To implement this, the modeler usually defines a geographic grid over which predictions will be made, and then uses a geographic information system or similar software to extract into each grid cell the values of each covariate from its original data source. (Because the original data sources often have different resolutions or other geographic characteristics than the modeler's grid, they often cannot be used directly for prediction, but must be *resampled* instead.) After all the covariates are prepared in this way, the modeler then makes a prediction at each grid cell, yielding a grid of density values that constitutes the density surface.

When models are fitted only to *static* covariates such as the depth or slope of the seafloor, which do not change appreciably over typical model time scales, only one such density surface prediction is needed. The resulting single

map represents a static depiction of density that does not change with time. For species that are not migratory or otherwise highly mobile, a static model may yield results that are reasonable to use in many species management problems. Static models can also be reasonable for mobile species when the modeled period is restricted to one season or timespan during which the species' distribution is relatively stable. Finally, they may also be useful as a fallback when data are too sparse to adequately model the detailed movements of a mobile species. In this latter case, many caveats apply, and substantial caution must be exercised when interpreting or applying the results.

For highly mobile species, including most marine mammals, modelers typically incorporate *dynamic* covariates, of which sea surface temperature (SST) and chlorophyll concentration are two of the most frequently used, owing to their long-term, global availability from satellite remote sensing. Dynamic covariates change with time, and each prediction must be made for a specific *time slice*, a short period of time over which the dynamic covariates are assumed by the model to be stable and unchanging. Thus, there is no "single map," as with a static model; the map obtained depends on the time slice of interest. This substantially complicates the application of the model to species management problems.

To address this complication, we must first consider what users want to do with the density predictions. For the models built under our cooperative agreement with the Navy, the primary applications are to facilitate the estimation of MMPA "takes" of marine mammals that may potentially occur for proposed training and testing activities in the ocean, and to enable model users to assess spatiotemporal patterns in density at locations and time ranges of interest. In both applications, the time ranges of interest usually occur in the future. Ideally, then, models would directly forecast marine mammal densities for these future periods.

At the time of this writing, such forecasting density models were still nascent research yet to be put into operational use. For them to be implemented successfully, several complex challenges must be addressed. Ocean models' outputs must be of sufficient ecological relevance to serve as effective covariates in density models. Ocean models must be capable of forecasting into the required future periods, over the spatial extent required by the density models, and the forecasting accuracy of the ocean models must be demonstrated at the spatial and temporal resolutions of interest for density modeling. The ocean models must also be capable of hindcasting back in time far enough to cover a sufficient range of the density model's training data (the survey segments), or compatible data must be obtained from other sources such as remote sensing, and cross-calibration performed. Density models built on ocean model hindcasts must be shown to be effective at hindcasting density, and modelers must produce convincing evidence that such models can be successfully extrapolated far enough into the future to address user needs.

Until these challenges are overcome and forecasting density models accepted by managers for operational use, we anticipate that model users will continue the traditional practice of utilizing densities of the recent past predicted by non-forecasting density models as a proxy for densities of the near future. Under this approach, modelers identify a period of the past that is likely to be ecologically similar to the future period of interest. Ideally, this past period should be as long as possible so that temporal variability can be comprehensively captured, but the duration is often constrained by practical and ecological factors. The start date of the modeled period is usually limited by how far back in time surveys or covariates are available. For example, covariates related to remotely sensed chlorophyll concentration are usually not available prior to September 1997, when the satellite borne SeaWiFS ocean color sensor began operating. It may also be necessary to avoid extending the model so far back that it encompasses ecosystem regimes that are unlikely to resemble the future. Finally, if the model covers a period long enough to span substantial demographic changes, then density predictions for the past period may substantially overestimate or underestimate the densities of the desired future period, unless this variability is explicitly addressed.

Ideally, the end date of the modeled period is the present day, but in practice it is usually constrained by data availability limitations. Survey data are often not ready for modeling use until the survey team has thoroughly reviewed and cleaned the data, a process that can take several months to more than a year. Similarly, many covariates are often only available after some delay, unless they are all taken from "near real time" remote sensing products or ocean models. (Usually, the model cannot be predicted until they are all available.) Finally, the processes of model fitting, refinement, prediction, and summarization often require days to months to complete. With sufficient resources and software automation, all of these processes, from survey data cleaning to model prediction and summarization, can be shortened to hours or days, if user needs require it. But in practice, for most models, the modeled period often ends months or years before the present date.

Once a suitable modeling period has been solidified and the model fitted, the model may be predicted across the time period of interest, obtaining density surface predictions at every time slice. The resulting time series of density maps may be simply handed over to model users, but in most cases, users require some additional outputs. In particular, they usually request that the time series of predictions be summarized into a smaller number of maps that are easier to utilize in their own modeling or planning processes, and also call for maps estimating the uncertainty or variability in the predictions.

5.1.2. Summarizing density and uncertainty

The topic of how to estimate density surface model uncertainty and summarize density and uncertainty is the focus of a manuscript (Miller et al. in prep) we are drafting with the DenMod working group (https://synergy.st-andrews.ac.uk/denmod/). We anticipate submitting this manuscript to a journal in late calendar year 2021 or early 2022. Below, we provide a brief description of the application of those methods to the right whale density model. For complete details of those methods, please see the manuscript upon publication.

5.1.2.1. User requirements

When we discussed with the Navy how they will use density predictions, they articulated some specific requirements. As discussed above, the predictions should be made over a period of the recent past that is, in our professional judgment, likely to be a reasonable proxy for the relevant near future, which for the Navy was the upcoming 7 years, the period over which their AFTT EIS was likely to apply. They expressed that they are sometimes able to narrow down proposed activities to a specific part of the year, sometimes to a specific month. They requested that predictions for each modeled taxon be summarized into the following outputs:

1. 12 monthly density maps, each giving the mean density for each grid cell for that month averaged over the years predicted.

2. 12 corresponding maps of the coefficient of variation (CV), each grid cell of which expresses the estimated variability or uncertainty around the corresponding mean density (#1 above), accounting for the important sources of variance and uncertainty we were able to address during the time available for the project.

As mentioned above, during the project we were able to account for two such sources of variance and uncertainty: the statistical error in model parameter estimates (*model uncertainty*), and the variance in predictions induced by dynamic covariates (*environmental uncertainty*). A subtle but very important consideration is that the proper estimation of environmental uncertainty, framed here as the CV of the mean, depends on the duration of the proposed human activity. For activities lasting the entire duration of the modeled period, it might be appropriate to estimate CV from the standard error of the means over that period. Conceptually, the CV (output #2) would express how close the mean density (output #1) was likely to be to the true mean over that entire period. If the period was lengthened, the CV would decrease as additional time slices were accumulated, assuming the additional predictions were similar to those before.

We judged this to be the wrong way to estimate environmental uncertainty for the Navy's application. In our professional judgment, human activities (i.e., naval training and testing activities) proposed at a given location were likely to last from a few days to several months, with 1 month being a reasonable approximation of these, and also matching the Navy's requested temporal resolution. Therefore, the estimate of environmental uncertainty for a particular mean monthly density (output #1) is more akin to the standard deviation of the monthly density predictions. Conceptually, it can be expressed as the variability that would result if that month's density was randomly drawn from just one of the years spanned by the modeled period.

Discussions with the broader prospective model user community suggest that the Navy's requirements and desired outputs were likely to meet the needs of many other users as well. In the following section (5.1.2.2), we present the method we used to derive the these outputs using an analytic approach that combined model uncertainty derived using a sandwich estimator (Miller et al. (2013), Appendix B) with environmental uncertainty expressed as the variance of the monthly predictions. This method represents an advance over what was done for the Phase III models and

subsequent updates in 2017-2019, which only accounted for model uncertainty (via a similar sandwich estimator), We anticipate these outputs will be applicable to a wide range of species management and spatial planning problems. However, several users expressed interest in an alternative characterization of uncertainty that used a simulation-based approach, which has certain advantages over the analytic approach. We present this in section 5.1.2.3.

5.1.2.2. Analytic summarization of uncertainty

Full exposition of this approach, including a statistical derivation, is given by Miller et al. (in prep). We thank David L. Miller of the University of St. Andrews for leading the statistical development of this approach. Here, we give a brief description as it pertains to the right whale model.

5.1.2.2.1 For density surfaces

The version 11 model stitched together several regional sub-models. Most of these were generalized additive models (Wood 2017) built with contemporaneous dynamic covariates that had monthly resolution (see Roberts et al. (2020), Section 5.3). For each regional model, one prediction was made for a given month of a given year. Predictions from each model were grouped by month across the years of the three eras 2003-2009, 2010-2018, and 2003-2018, and for each {model, month, era} combination, uncertainty was summarized as follows.

For the focal month, let the estimated abundance (number of individual animals) predicted for grid cell r be $\hat{n}_{rt}$ for years in the era $t = 1 \ldots T$, and the mean abundance over these years be $\bar{n}_r$. The variance of the mean abundance $\mathrm{Var}(\bar{n}_r)$, inclusive of model uncertainty and environmental uncertainty, was estimated:

$$\mathrm{Var}(\bar{n}_r) = \left[\frac{(J-1)}{JT-1}\sum_{t=1}^{T}\mathrm{Var}_t(\hat{n}_{rt})\right] + \left[\frac{J(T-1)}{JT-1}\mathrm{Var}(\hat{n}_t)\right] \quad (1)$$

where $\mathrm{Var}_t(\hat{n}_{rt})$ was the model uncertainty at a particular time slice, which was estimated at each time slice using the sandwich estimator:

$$\mathrm{Var}_t(\hat{n}_{rt}) = \left(A_r \frac{\partial \exp \mathbf{L}_{rt}\beta}{\partial \beta}\Big|_{\beta=\hat{\beta}}\right)\mathbf{V}_{\hat{\beta}}\left(A_r \frac{\partial \exp \mathbf{L}_{rt}\beta}{\partial \beta}\Big|_{\beta=\hat{\beta}}\right)^{\mathsf{T}} \quad (2)$$

where A_r was the area of the cell, $\mathbf{L}_{rt}$ was the prediction matrix (`lpmatrix` in Wood (2017), Section 6.10) for the cell, $\hat{\beta}$ were the model's parameter estimates, and $\mathbf{V}_{\hat{\beta}}$ was the model's covariance matrix. Returning to (1), $\mathrm{Var}(\hat{n}_t)$ was the environmental uncertainty across the prediction period, expressed as the empirical variance of the cell's predicted abundance (driven by interannual variability in the dynamic covariates):

$$\mathrm{Var}(\hat{n}_t) = \frac{1}{T-1}\sum_{t=1}^{T}(\hat{n}_{rt} - \bar{n}_r)^2. \quad (3)$$

Finally, in (1), J was the number of time slices that would ideally be used in the calculation, which may be contrasted to T, which was the actual number of slices. For this analysis, after consulting our collaborating statistician at the University of St. Andrews, we set $J = 30$. Please see Miller et al. (in prep) for a discussion of the J parameter.

To obtain a variance surface for a given month, (1) was calculated across the study area using the years of the era. Thus, for the three eras studied, 2003-2009, 2010-2018, and 2003-2018, T was 7, 9, and 16 years, respectively. It may be seen in (1) that J and T control the relative weight given to the model uncertainty vs the environmental uncertainty in the estimation of total variance. As T increases from the minimum value of 2 for a model that yields non-static predictions, the relative weights rapidly become invariant the choice of J (Figure 9). For the values of T used here, the relative weight is approximately $(T-1)/T$ no matter what value of J is chosen (subject to the necessary constraint that $J \geq T$). For the values of T in this analysis, 7, 9, and 16, the weights of environmental uncertainty were approximately 86, 89, and 94%, respectively. Thus, in the total variance estimates for this model, environmental uncertainty exerted a dominant influence relative to model uncertainty.

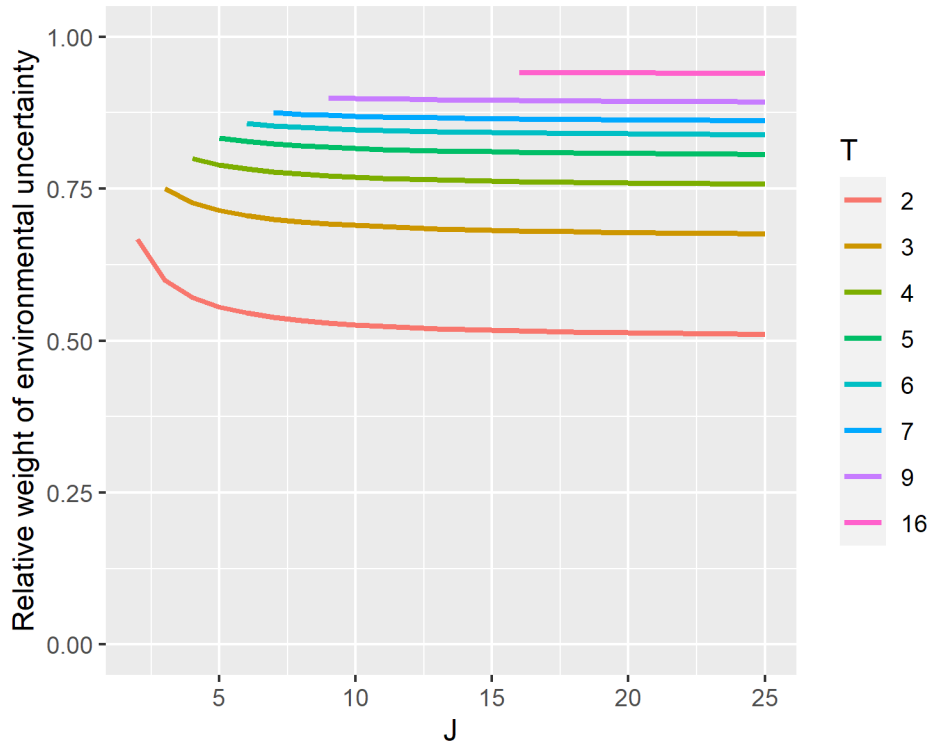


Figure 9. Relative weight given to environmental uncertainty by equation (1) for different values of the J and T parameters. Note how the weight rapidly converges to $(\boldsymbol{T}-\mathbf{1})/\boldsymbol{T}$ as J increases; the corresponding relative weight of model uncertainty converges to $\mathbf{1}/\boldsymbol{T}$. Both weights become rapidly invariant to J as T increases.

From the monthly variance surfaces, we produced, as outputs, monthly surfaces of:

- The standard error (SE) of density, computed as the square root of variance, then rescaled from the cell's abundance to density in our preferred unit of individuals / 100 km^2. Although we termed this statistic the "standard error", it is, perhaps, conceptually closer to a standard deviation, given the dominance of environmental uncertainty in its formulation, as we discussed above.
- The coefficient of variation (CV), computed as SE divided by mean density.
- Approximate asymptotic 90% confidence intervals, estimated from the CV as $CI_{lower} = density/\exp(1.645\sqrt{\log(1+CV^2)})$ and $CI_{upper} = density * \exp(1.645\sqrt{\log(1+CV^2)})$.

Several regional models used only static or climatological dynamic covariates (see Roberts et al. (2020), Section 5.3). These were models for which we lacked sufficient data to confidently model interannual variability. For these models, density predictions did not change according to year (for models with climatological covariates, predictions did change from month to month). We produced the same outputs listed above, but the variance only accounted for model uncertainty, estimated with (2). As a consequence, the variance and the derived uncertainty outputs were underestimated, and were generally lower than regions for which environmental uncertainty was able to be estimated. We discuss this further in section 5.2.2.2 below.

Finally, for Cape Cod Bay, for which density estimates were derived from independent abundance estimates produced by Ganley et al. (2019) and follow-on work (see section 4 above), we estimated SE as the standard deviation of the monthly density estimates for the years of the eras analyzed.

5.1.2.2.2 For abundance estimates

To estimate the standard errors for regional abundance estimates (and then the CVs that appear in parentheses next to abundances in many of our density maps), we used the same approach as for surfaces but adjusted the sandwich operator (2) to summarize variance across the cells representing the region of interest, replacing the scalar A_r with the vector $\mathbf{A}$, representing the area of all cells that were summarized (this area was the same for all cells), and the 1-row vector $\mathbf{L}_{rt}$ with the multi-row vector $\mathbf{L}$, representing the prediction matrix (`lpmatrix`) for all cells being summarized. This calculation accounted for covariance between cells, an important consideration given that covariate values across adjacent or nearby cells are often similar.

To estimate the standard error of the summed abundance of two or more regions, e.g., for the total abundance of the full study area, we assumed the regional abundances were independent, added their variances, and took the square root. This assumption is not likely to hold in reality—for example, density north and south of Cape Hatteras might be negatively correlated, as some years many whales travel south to the calving grounds, and in other years they remain in the mid-Atlantic or New England—but we have not tried to assess the effect this might have on the variance of a total abundance estimate. In any case, in the future, we may attempt to end our practice of using multiple independent regional models and instead use hierarchical GAMs (Pedersen et al. 2019) that encompass the entire study area but use the region as a categorical interaction term, to allow relationships to vary between regions. If this were successful, it would obviate the need to combine variances from multiple models.

5.1.2.3. Simulation-based summarization of uncertainty

A drawback to the idea of summarizing uncertainty into a single surface (e.g., SE or CV) is that it treats the cells as fully independent units, when in reality, nearby cells are usually similar to each other in their covariate values. This presents a problem when a model user needs to utilize a set of plausible alternative density surfaces that "could have occurred" if uncertainty was accounted for. It is usually not satisfactory to generate alternative surfaces by iterating across the study area and drawing random density values for each cell according to its estimated mean and standard error. This will simply lead to noisy surfaces that show unrealistically high local variability (like "television static") but revert to the mean surface when smoothed spatially (e.g., with a moving window average). Nor is it satisfactory to generate a series of quantile maps of density (e.g., each percentile from 5% to 95%). Under this approach, the spatial distribution again resembles the mean, but is scaled up or down roughly according to the quantile, often yielding unrealistically extreme total abundances, both high and low.

Instead of these approaches, a method is needed to simulate plausible alternative density surfaces in a way that allows nearby cells to covary realistically. In the simulated surfaces, local variability would be more realistic, the population sizes would likely be less extreme, and the density would "shift around" from surface to surface according to the uncertainty in model parameters and variability in dynamic covariates through time. Once produced, the set of simulated surfaces could be summarized into the summary surfaces and statistics described in Section 5.1.2.2 in a straightforward manner, with more accurate confidence intervals than the approximate asymptotic method used there. Finally, for users needing to assess the uncertainty of a higher-level model they have built atop the density model, the complete set of simulated surfaces could be provided instead of the summaries, allowing the users to assess their own model's uncertainty using Monte Carlo simulation.

After such a set of plausible alternative density surfaces were simulated, the summary surfaces and statistics described in Section 5.1.2.2 above can be computed in a straightforward manner, as we discuss below. A further benefit is that more accurate confidence intervals can be obtained by directly summarizing the set of alternative density predictions rather than using the approximate asymptotic method of Section 5.1.2.2. Finally, users needing to assess the uncertainty of a higher-level model they have built atop the density model can obtain realistic results by executing their model multiple times over the simulated density surfaces in a Monte Carlo simulation.

The development of a method for simulating plausible alternative density surfaces has been a focus of the DenMod working group and their forthcoming manuscript (Miller et al. in prep). We attempted to apply this method to the version 11 density model but were only partially successful. We determined at the time of our application that the method required additional work to overcome the problems we encountered, which were also experienced by colleagues at NOAA Southwest Fisheries Science Center (SWFSC) when they attempted to apply it their cetacean

density models of the California Current Ecosystem (Becker et al. 2020), and by colleagues at the University of St. Andrews and NOAA SEFSC when they attempted to apply it to the MGEL-led AFTT Phase III density models for the Gulf of Mexico (Roberts et al. 2016a). Although we hoped to base our summary outputs on simulated density surfaces, rather than using the analytical method of Section 5.1.2.2, and to release the simulated surfaces to model users, we were not able to sufficiently mitigate these problems and had to defer the use of the simulation based method until the future, when the DenMod collaborators can develop a general solution.

In this section, we briefly describe the simulation method (please see Miller et al. (in prep) upon publication for complete details) and then outline our experimental application of it to part of the right whale model. Section 5.2.2 below gives the results.

5.1.2.3.1 Simulation and summarization of density surfaces

For each sub-region of the study area modeled with a generalized additive model (GAM), we performed the following procedure. Let T be the number of time slices to be simulated, computed here as the number of months of the year included in the model (maximum of 12) multiplied by the years in the summarized era. Let B be the number of alternative density simulations per time slice. (The DenMod team suggests between 100-1000 simulations; our specific choices appear in following sections.) After the model is fitted:

1. Extract posterior estimates of the GAM's parameters, $\hat{\beta}$, and covariance matrix, $\mathbf{V}_{\hat{\beta}}$.
2. From those, sample $b = 1, \ldots, B$ alternative sets of parameters, β_b, representing alternative models.
3. For each time slice, $t = 1, \ldots, T$:
    a. Form the prediction matrix (`lpmatrix`), $\widetilde{\mathbf{X}}_t$, for this time slice.
    b. For each alternative model, $b = 1, \ldots, B$ generated above:
        i. Calculate the predicted density surface, $\widehat{\mathbf{N}}_{b,t} = \exp(\widetilde{\mathbf{X}}_t \beta_b)$, rescaled to the appropriate density unit (typically individuals / 100 km^2 for our models).
        ii. Store $\widehat{\mathbf{N}}_{b,t}$ for this slice-model combination.
4. Summarize the per slice-model predicted density surfaces ($\widehat{\mathbf{N}}_{b,t}$) by computing the appropriate summary statistic (typically mean, standard deviation, or, for 90% confidence limits, the 5% and 95% quantiles) for all alternative models over suitable subsets of t (e.g., for each of the 12 months).

Under this approach, model uncertainty is captured by the alternative parameter estimates, which will yield a range of predictions for a given set of covariate values, and environmental uncertainty is captured by iterating through all of the time slices of the focal era.

For models that used only static covariates, which do not change in time, as in Test Case 1 in the following section, or used only climatological dynamic covariates, which changed by month but not year, the same procedure was executed but only for one time slice. In effect, T was 1, but the rest of the procedure was the same. In these cases, as with the analytic approach of Section 5.1.2.2, the results captured model uncertainty but not environmental uncertainty.

5.1.2.3.2 Test Case 1: a simple univariate model to demonstrate the approach

To demonstrate that results for the simulation approach could match those of the analytic approach for a very simple scenario, and to work out the mechanics of programming, we first designed a test case that used a small sample of the right whale data in a model having a single static covariate. For the survey data, we used the segments from the version 9 model (the most recent model at the time of this test) for the "Nantucket Shoals to Northeast Channel" region for March-July (Figure 10).

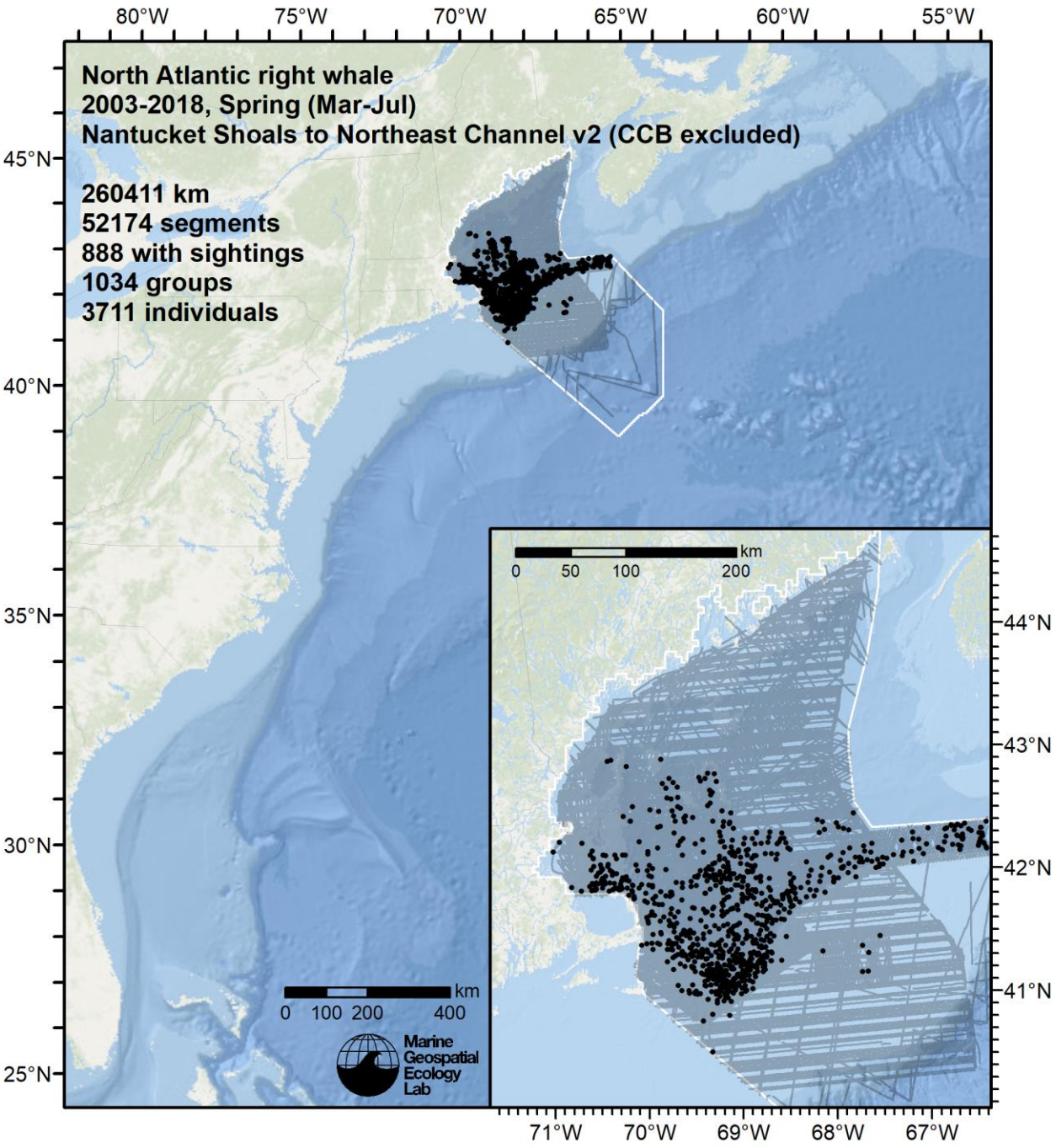


Figure 10. Survey segments and sightings used in simulation-based uncertainty summarization Test Case 1.

We fitted the model with Depth as the only covariate, using a thin-plate regression spline with our usual model parameterization (see Roberts et al. (2020) for details). We then summarized uncertainty surfaces using both the analytic (Section 5.1.2.2.1) and simulation-based (Section 5.1.2.3.1) approaches. For the simulations, we used 1000 alternative models ($B = 1000$). Because there was only one static covariate, there was only one time slice ($T = 1$).

5.1.2.3.3 Test Case 2: the “South of Hatteras Island” region for “Winter” (October-May)

The results of Test Case 1, shown in Section 5.2.2.1 below, suggested that the simulation-based approach could be successful, so long as we did not assume that model parameters would have multivariate-normal distributions, but instead more accurately elucidated the parameters’ distributions with Metropolis-Hastings sampling, or something similar. For Test Case 2, we applied the approach we refined in Test Case 1 to the “South of Hatteras Island” regional model for the “Winter” period, defined for this model as October-May, inclusive. This regional model was developed as part of model version 9 and not revised in the versions 10 or 11. We selected this region as our second test case, as it had the largest number of survey segments and sightings available (Figure 11) across all of the regions, with the best interannual replication of surveying (at least in the core calving habitat where density was highest). Also, this region’s model was one of the most complex, with a single static environmental covariate (Distance to Shore), 6 contemporaneous dynamic environmental covariates, a bivariate smooth of x and y coordinates (to account for static spatial variability not captured by environmental covariates), and the calving year as a factor term. This latter term was designed to account for interannual variations in right whale density at the calving grounds that we believed were not related to contemporaneous ocean conditions but to antecedent effects that are currently not easy to represent directly in the model, such as foraging success in prior summers.

To avoid lengthy delays in completing the sampling, prediction, and summarization of this complex model, while still obtaining sufficient alternative models to allow reasonable estimates of uncertainty, we limited the number of simulations to 200 ($B = 200$). For the other parameters of the `gam.mh` function in R package `mgcv`, we used the same values as in Test Case 1.

5.2. Results

5.2.1. Analytic summarization of uncertainty

Here we present the principal outputs of this uncertainty analysis: the uncertainty surfaces for the three analyzed eras, plotted as maps in a manner similar to our traditional density maps. For brevity, we only show the coefficient of variation, with the corresponding density map alongside. The other outputs (standard error and upper and lower confidence intervals) are also available but not included here.

For all eras, the predictions started in June of the first year listed and ended in May of the year after the last year listed. For example, for the 2003-2009 era, predictions started in June 2003 and ended in May 2010. (Please see Roberts et al. (2020) for more details.) Therefore, below, the maps will start with June and end with May.

Total abundance estimates for all regions with non-zero abundance that were modeled with GAMs have CVs estimated via this method. These CVs replace those that appeared in the version 11 maps, although total abundances remain the same.

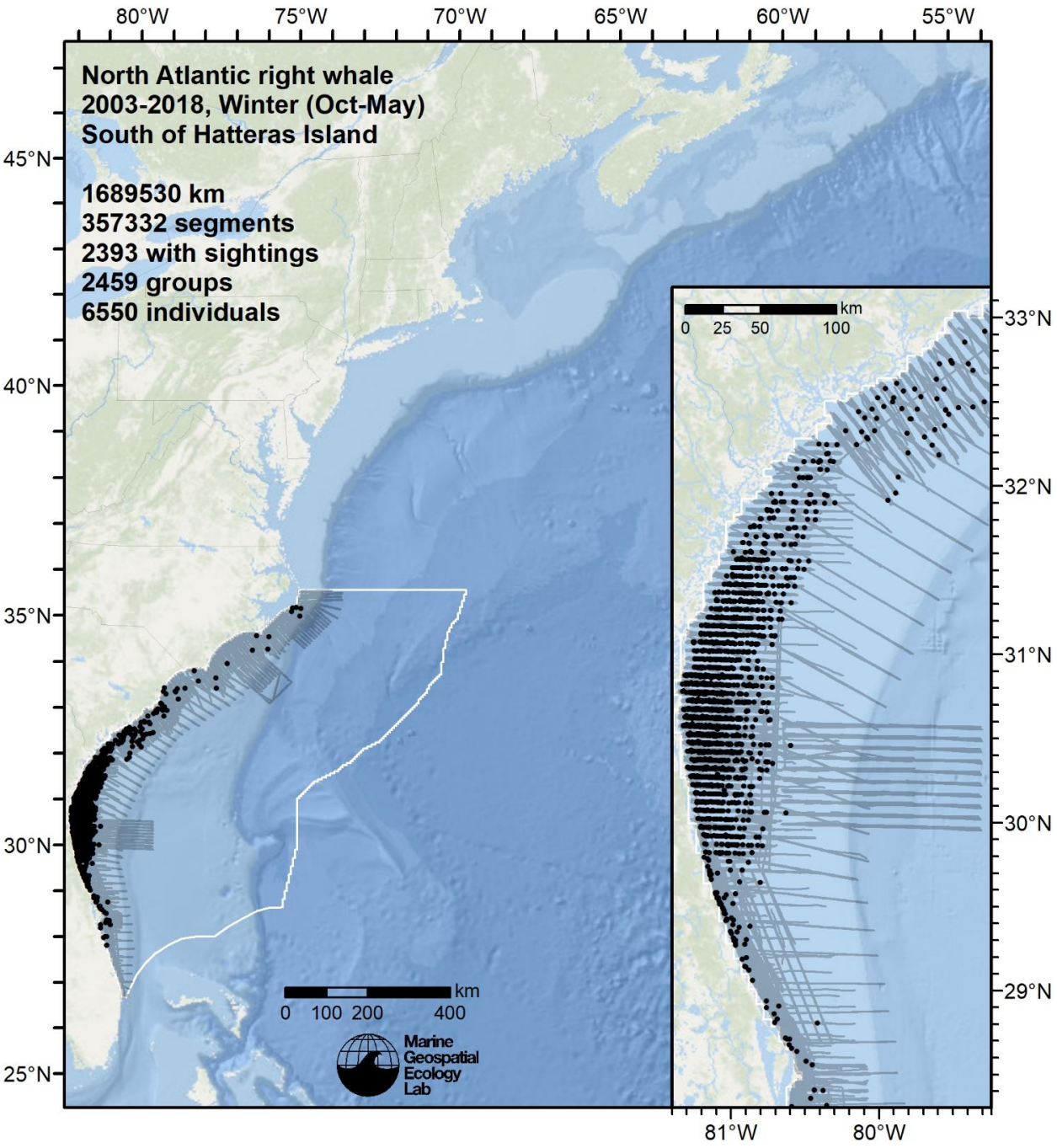


Figure 11. Survey segments and sightings used in the "South of Hatteras Island" regional model for the "Winter" (October-May) season. See Roberts et al. (2020) for model details. This model was used as simulation-based uncertainty summarization Test Case 2.

5.2.1.1. 2003-2009 era

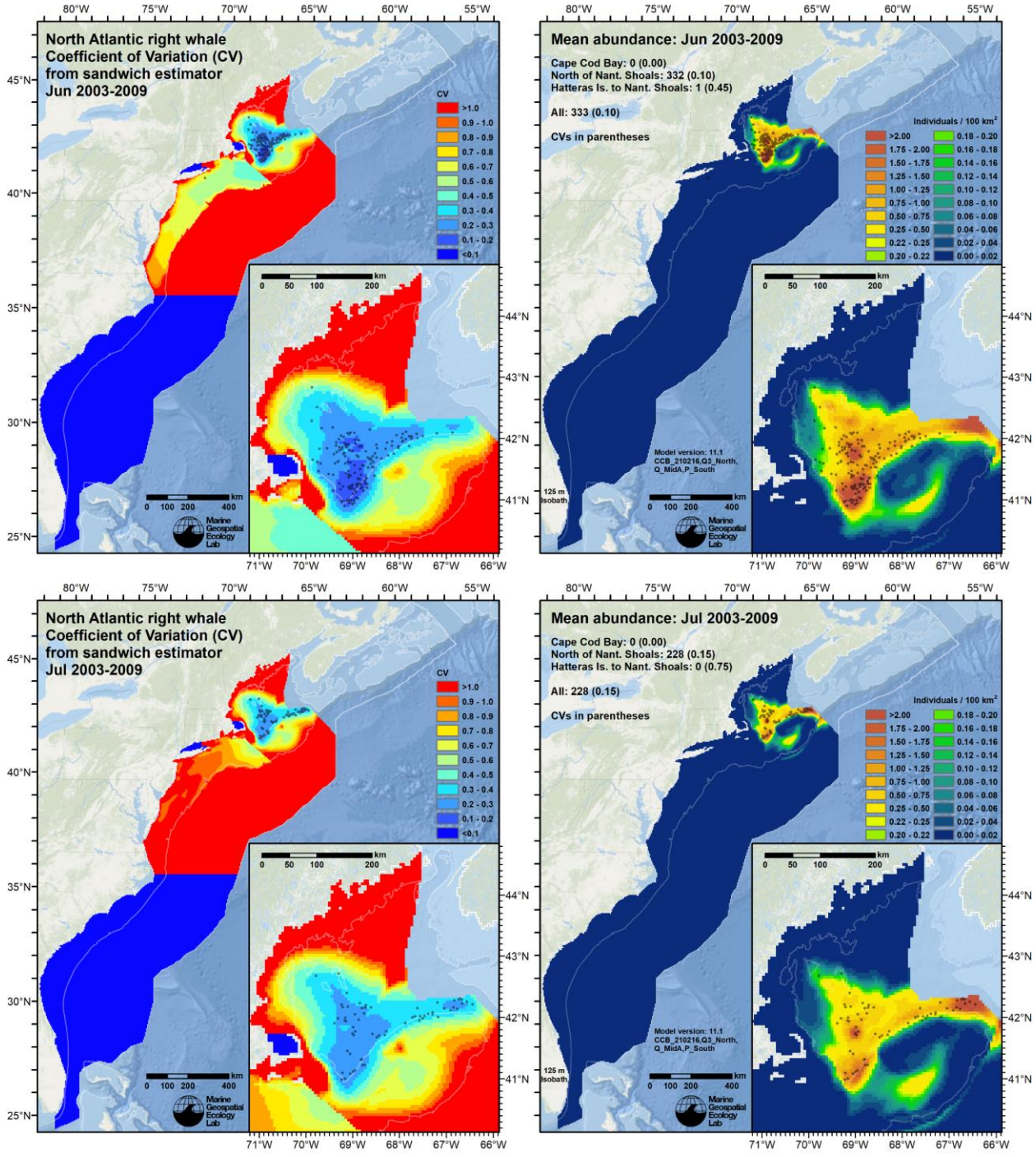


Figure 12. Monthly maps (rows) of estimated CV (left column) and mean density (right column) for the 2003-2009 era, spanning June 2003 through May 2010, with sightings overlaid. Note that survey effort, not shown here, was heterogeneous; a lack of sightings does not necessarily indicate that whales were absent. (Effort maps are included with the files that accompany this report.)

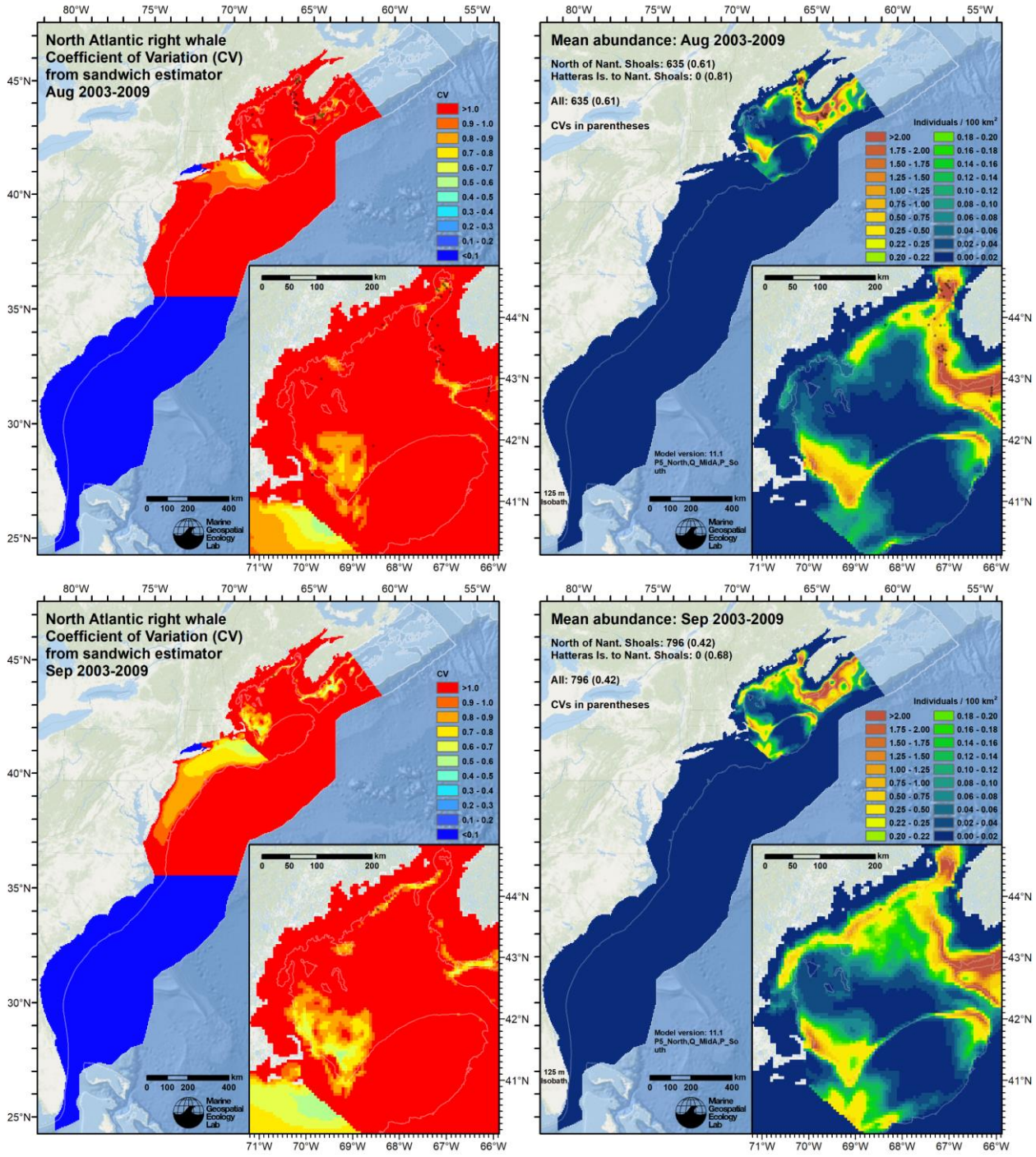


Figure 12. Continued from previous page.

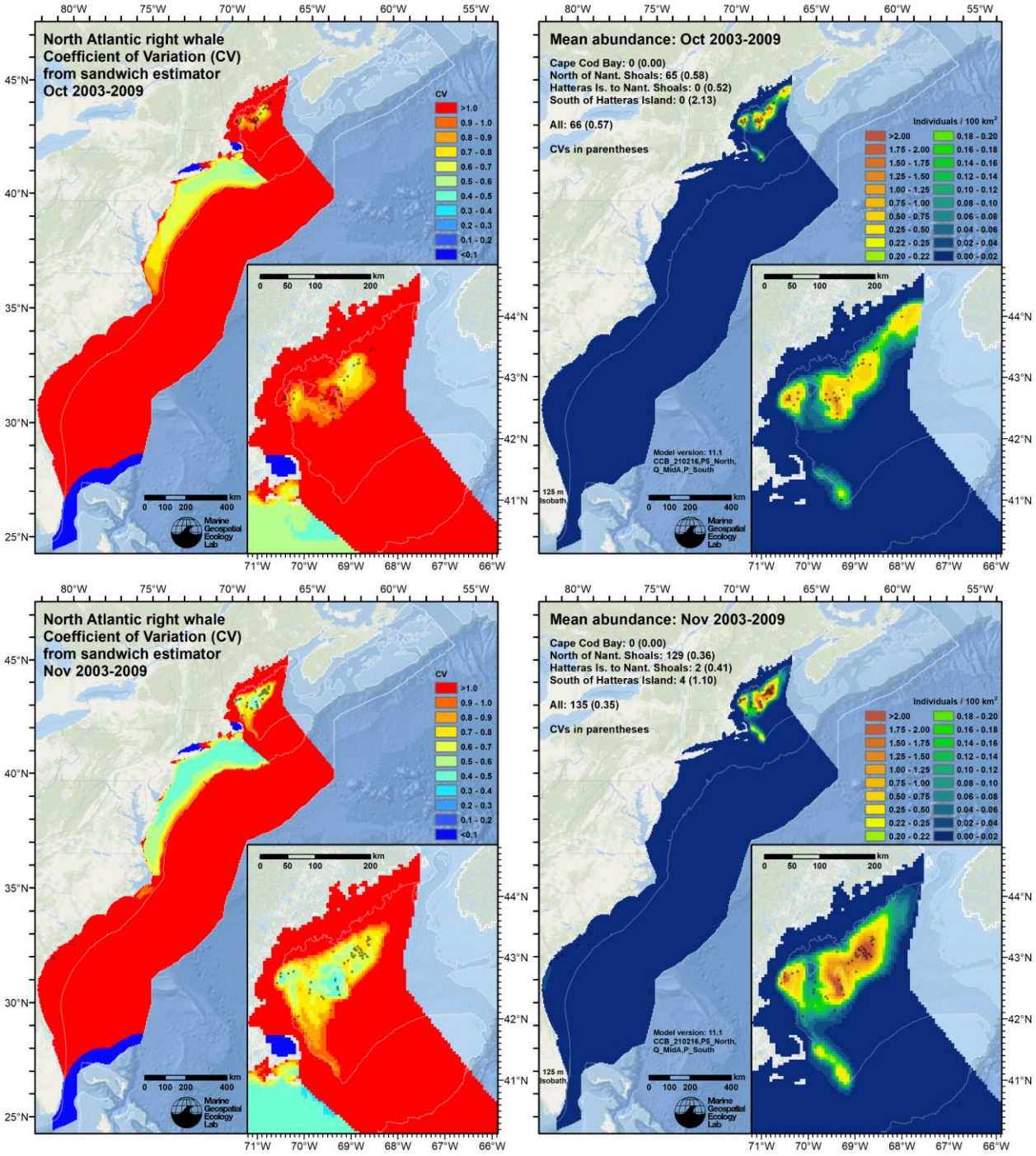


Figure 12. Continued from previous page.

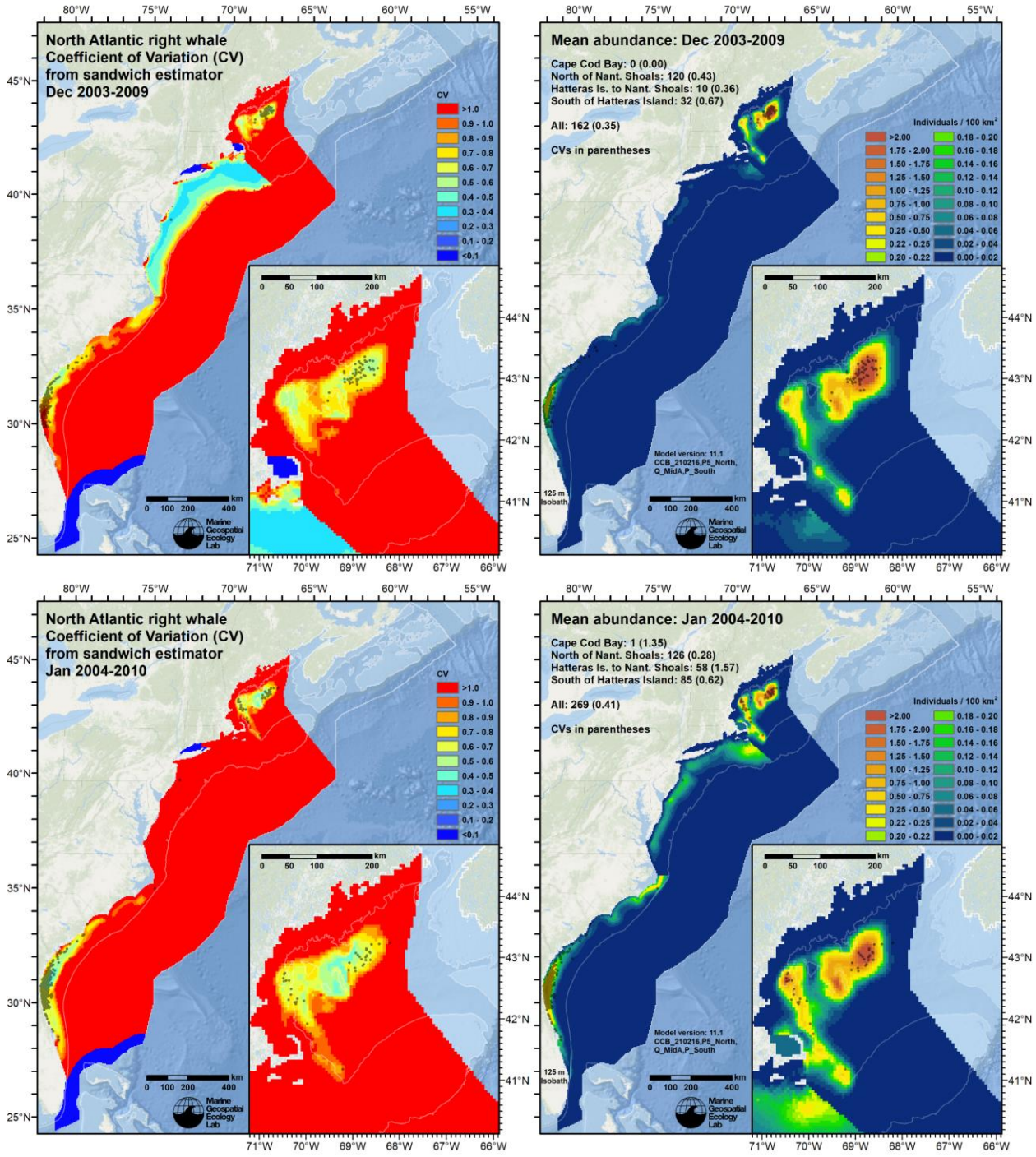


Figure 12. Continued from previous page.

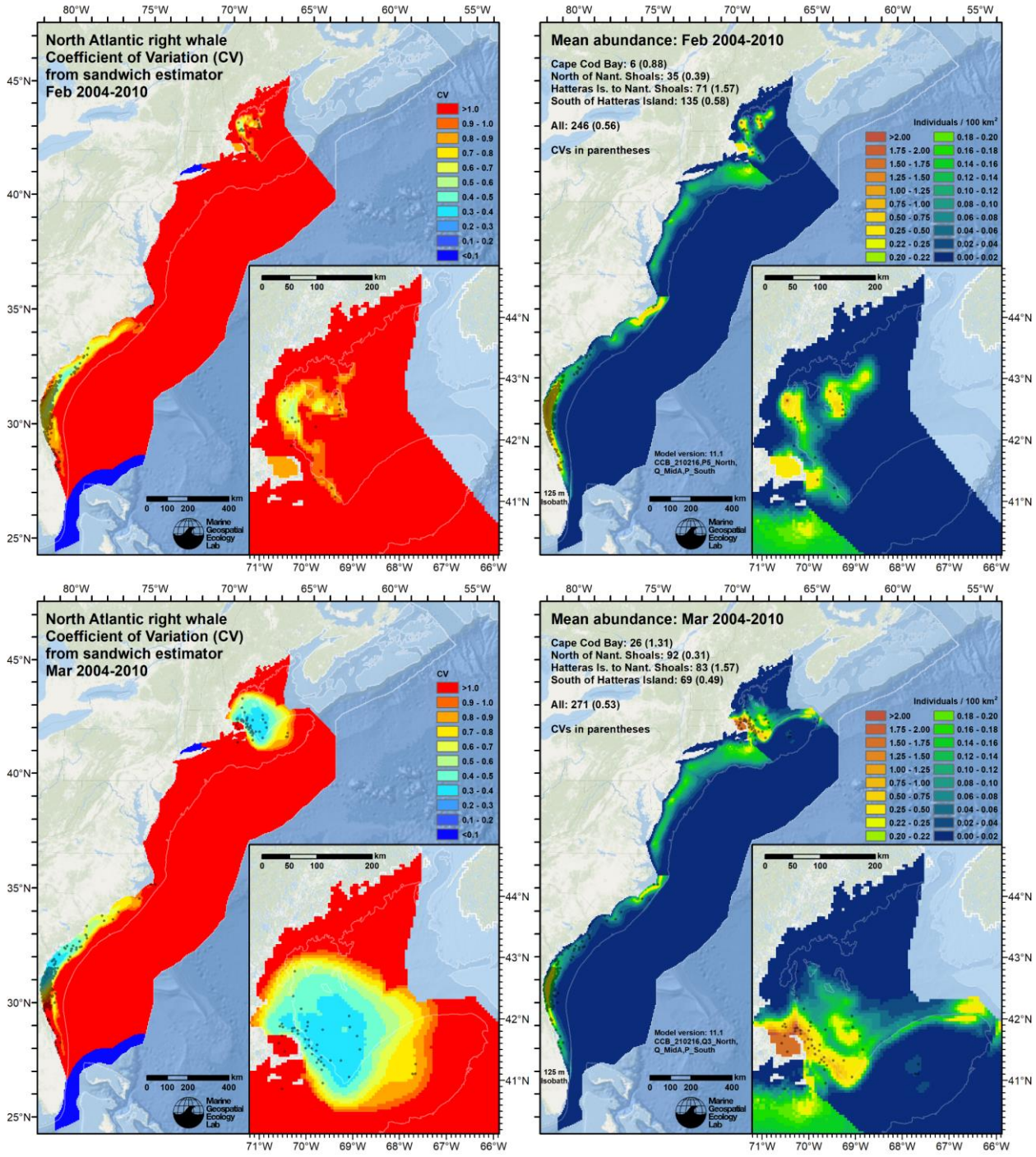


Figure 12. Continued from previous page.

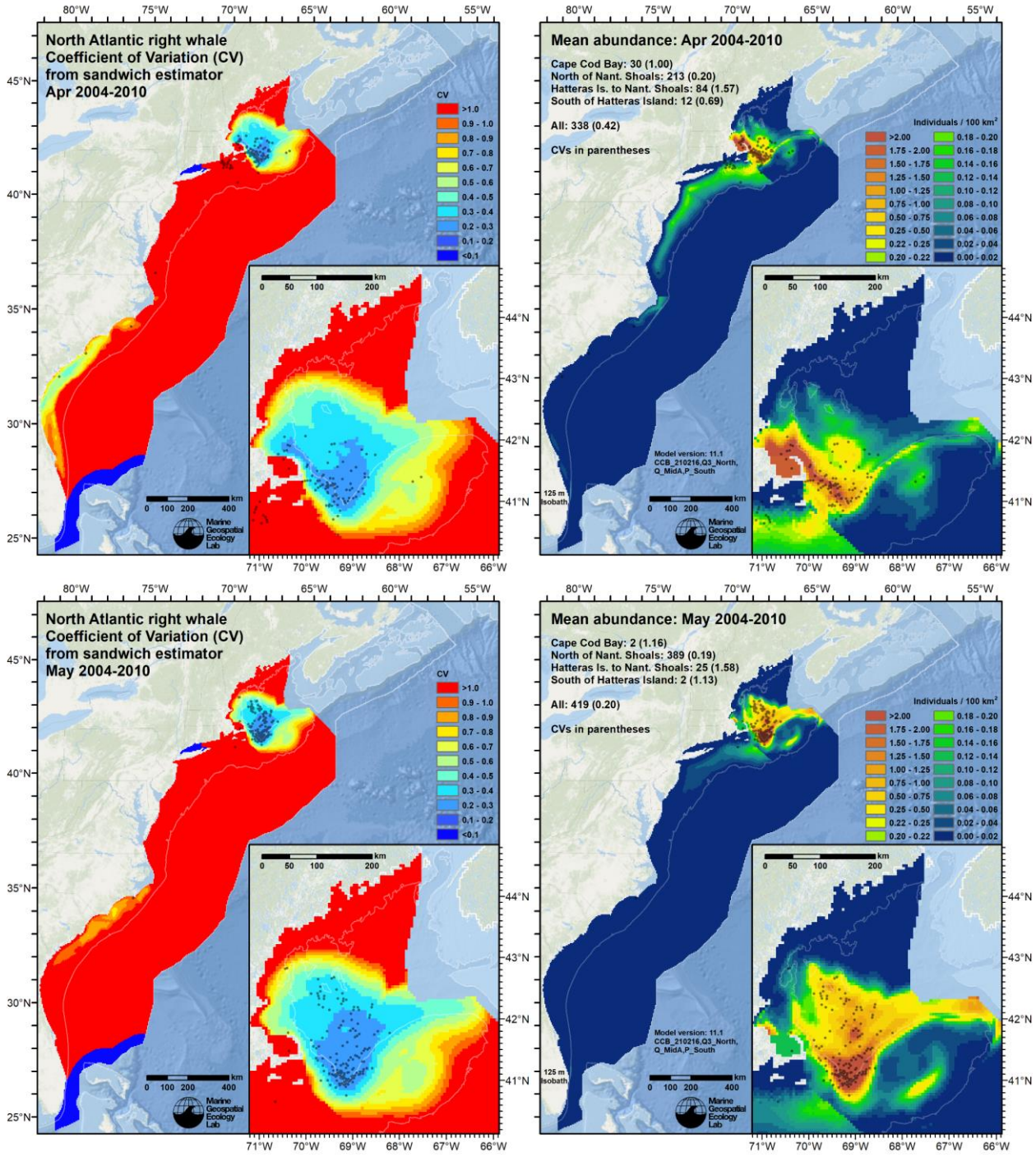


Figure 12. Continued from previous page.

5.2.1.2. 2010-2018 era

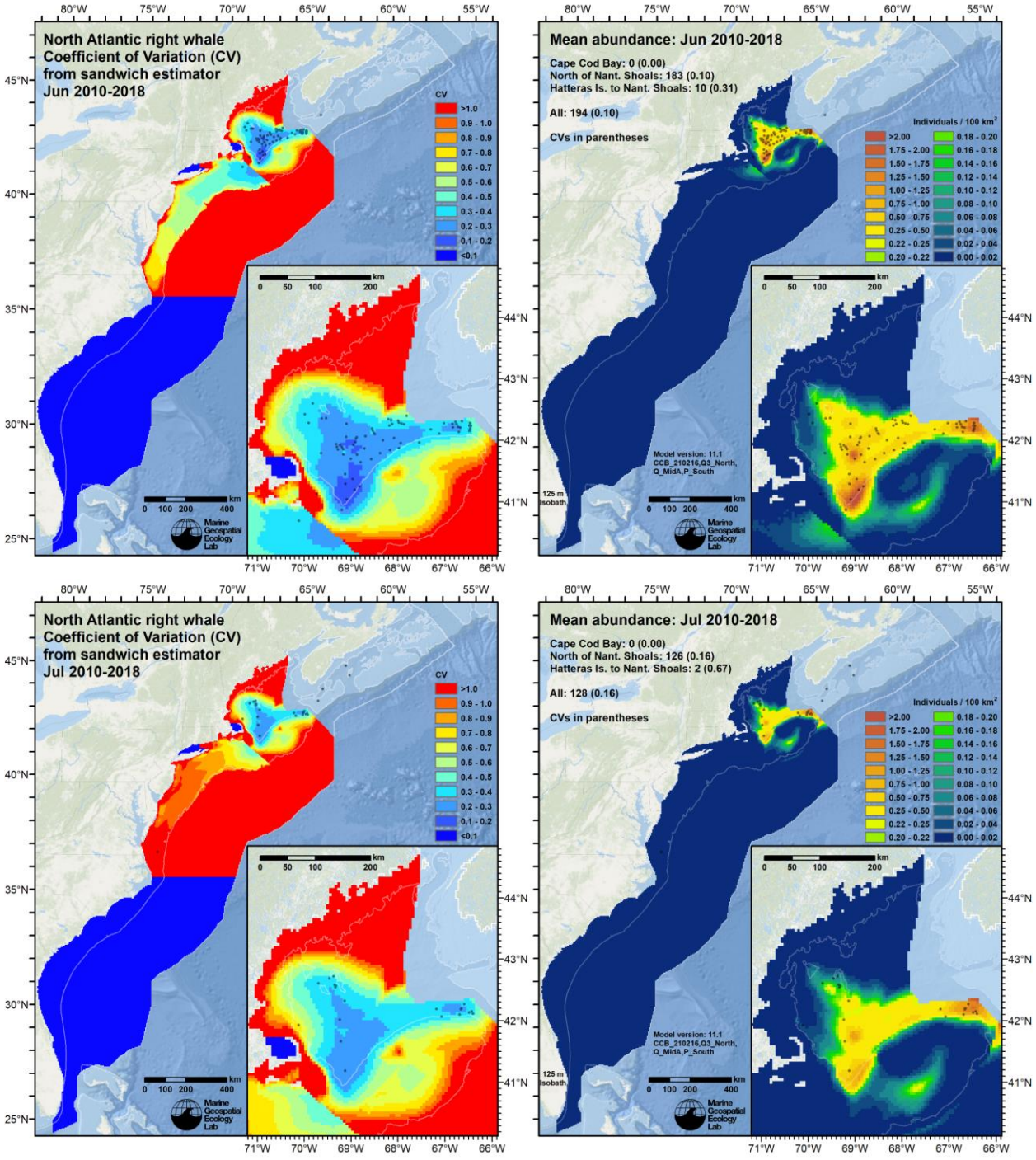


Figure 13. Monthly maps (rows) of estimated CV (left column) and mean density (right column) for the 2010-2018 era, spanning June 2010 through May 2019, with sightings overlaid. Note that survey effort, not shown here, was heterogeneous; a lack of sightings does not necessarily indicate that whales were absent. (Effort maps are included with the files that accompany this report.)

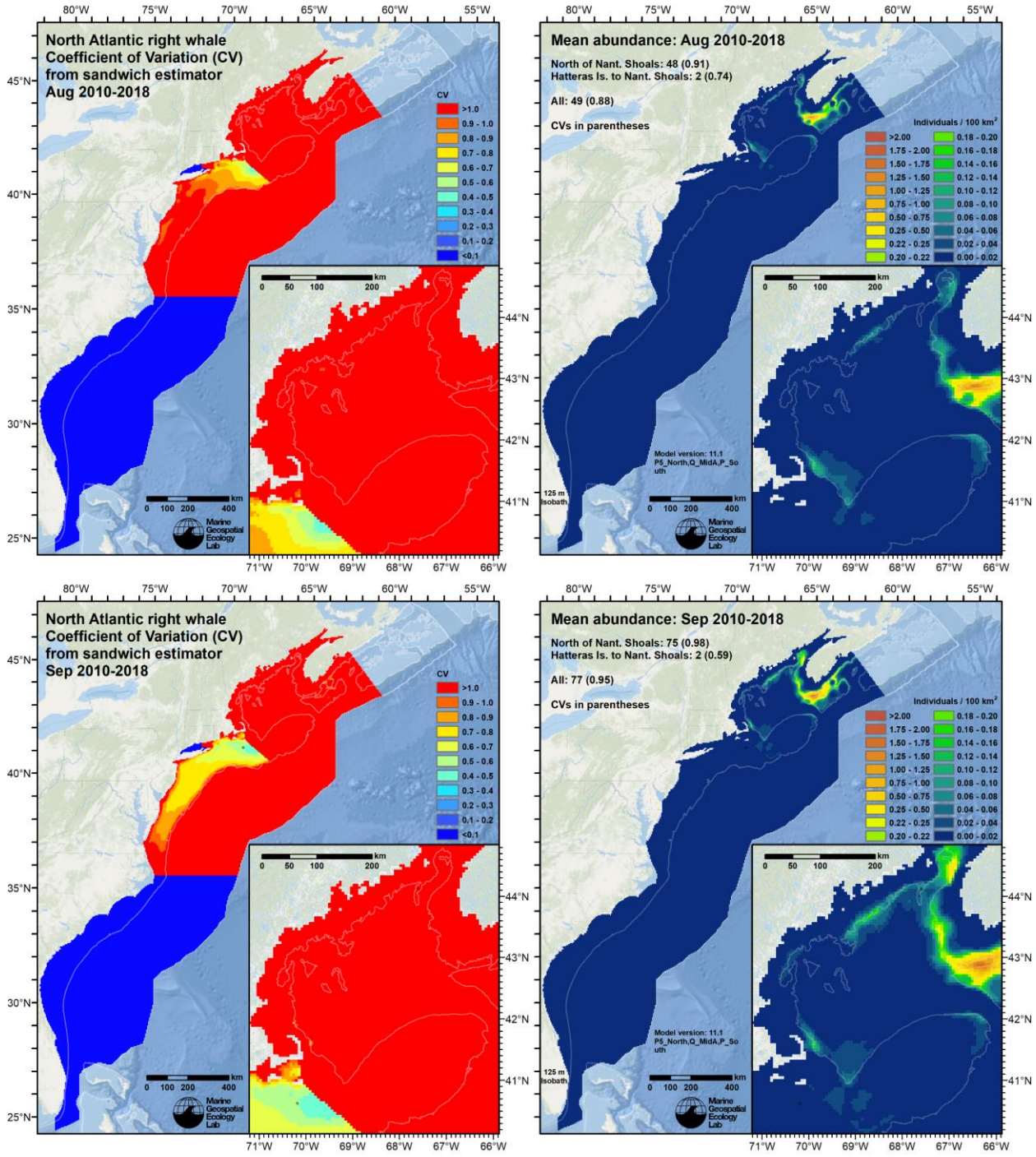


Figure 13. Continued from previous page.

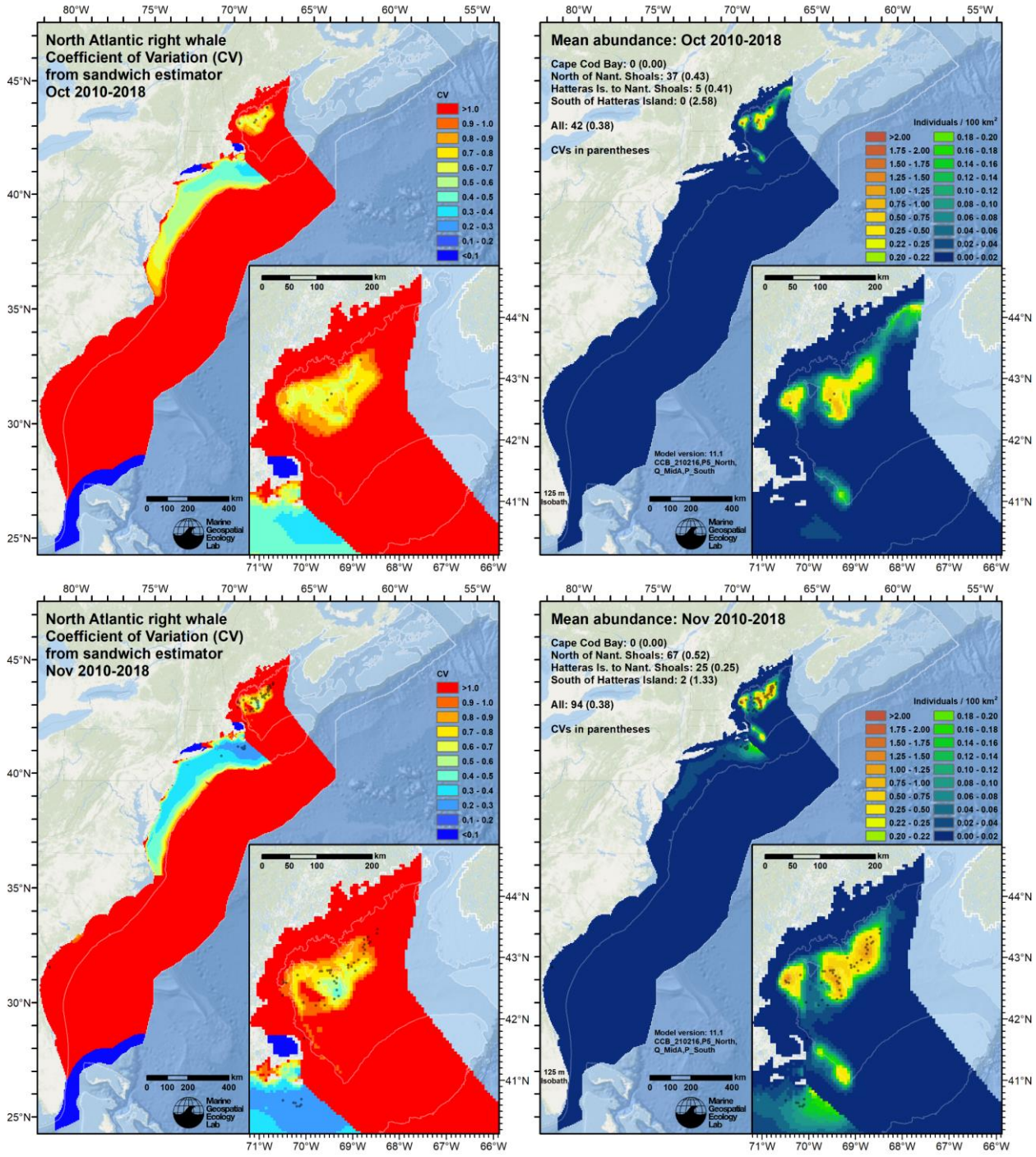


Figure 13. Continued from previous page.

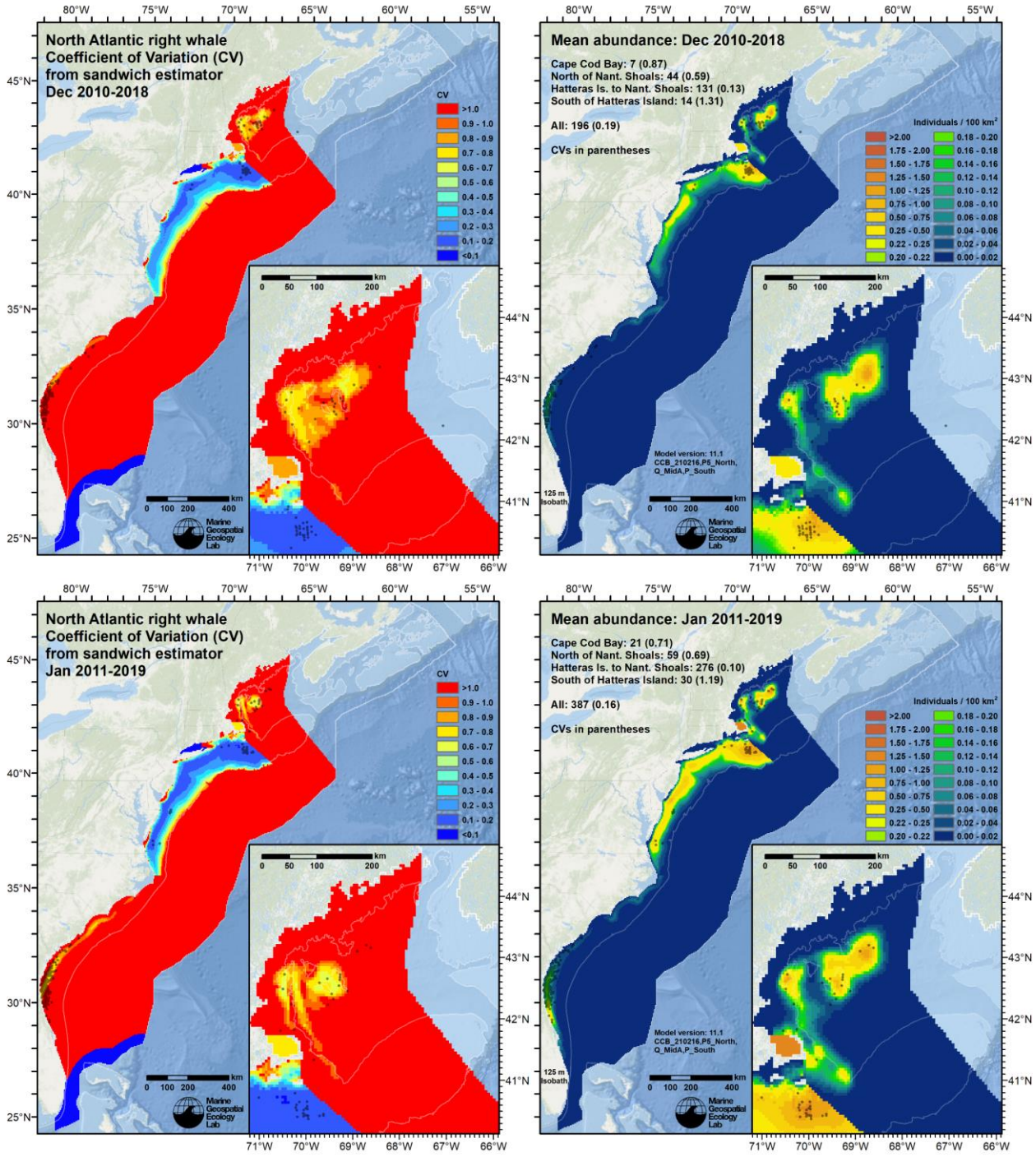


Figure 13. Continued from previous page.

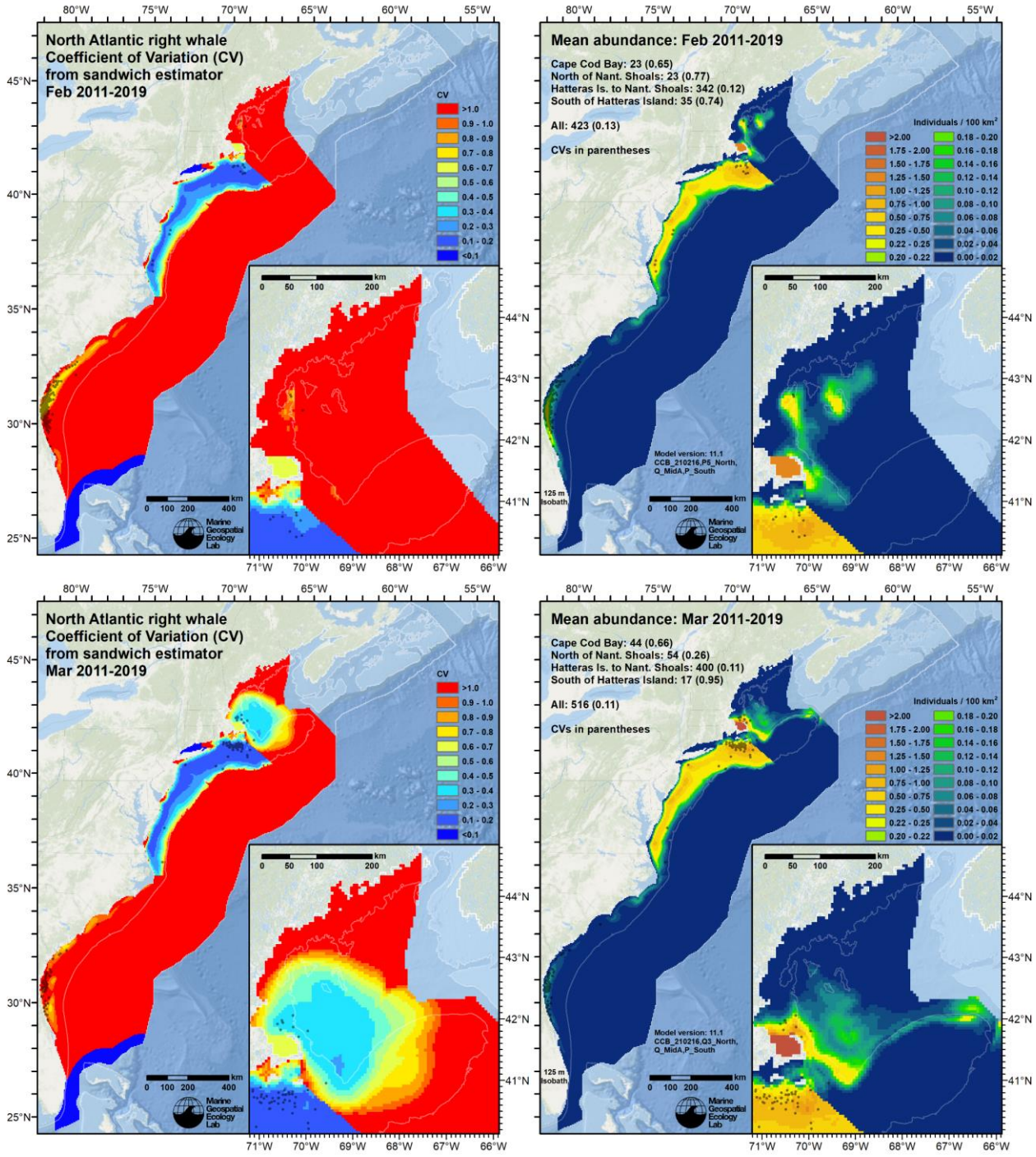


Figure 13. Continued from previous page.

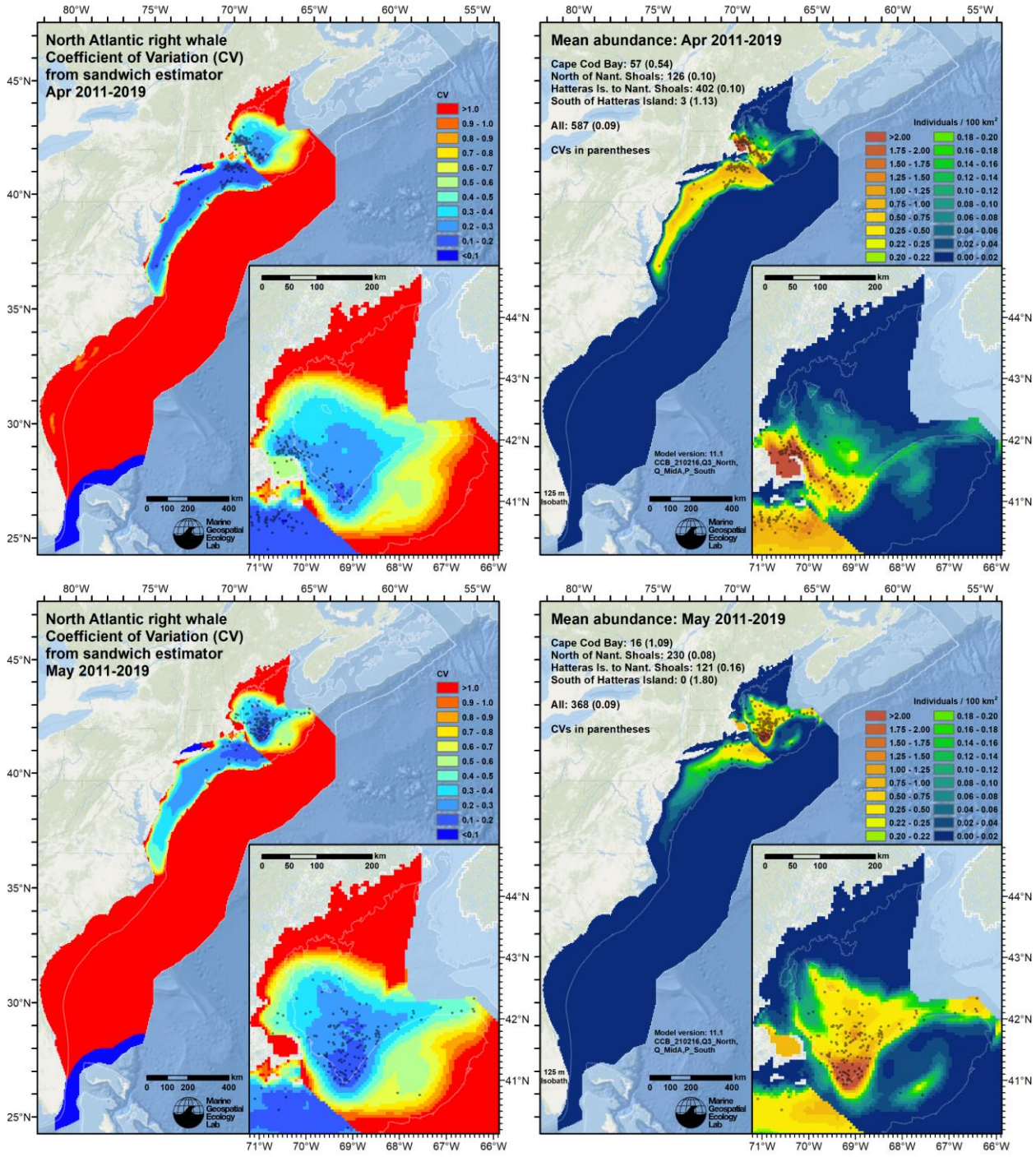


Figure 13. Continued from previous page.

5.2.1.3. 2003-2018 era

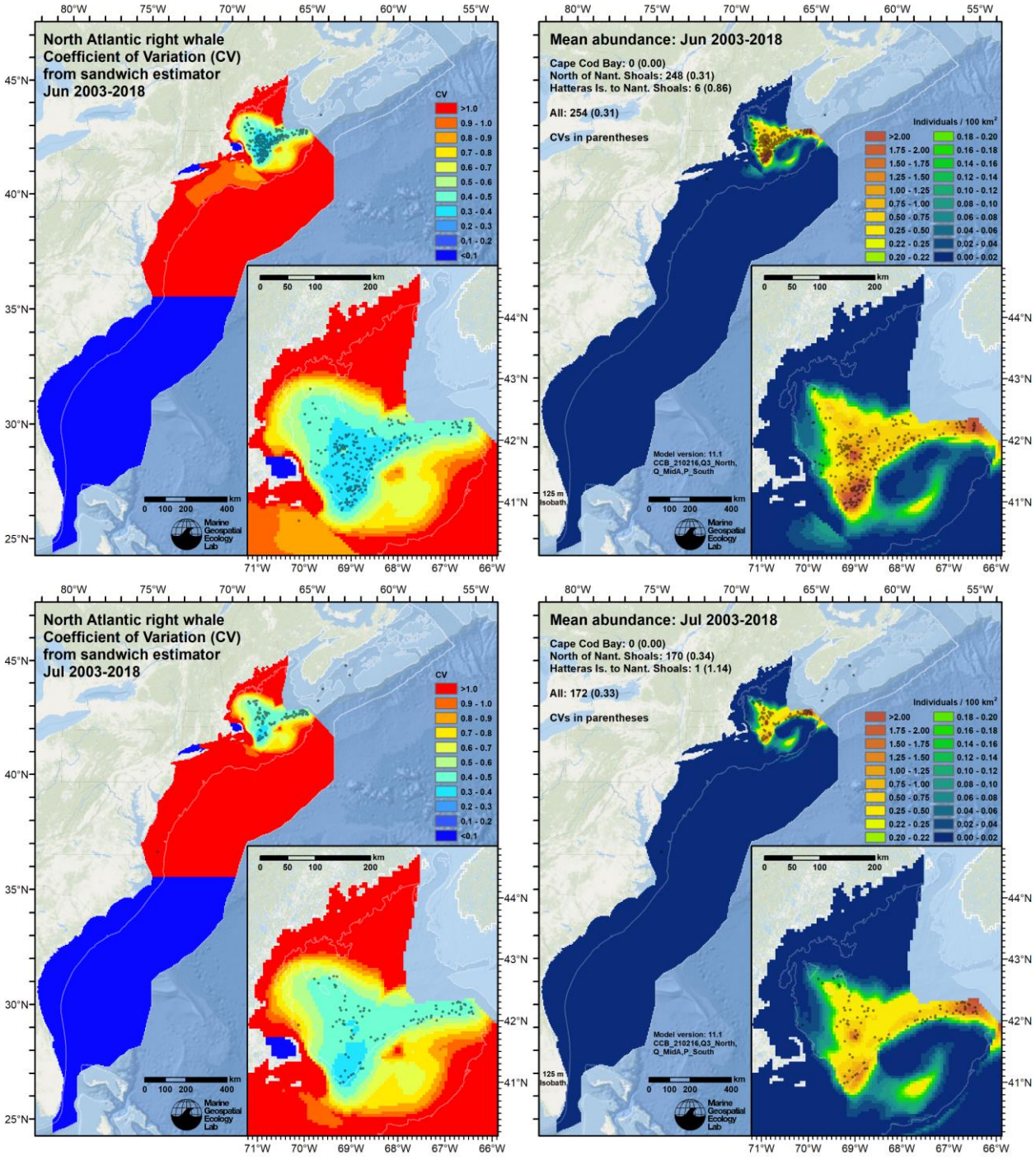


Figure 14. Monthly maps (rows) of estimated CV (left column) and mean density (right column) for the 2003-2018 era, spanning June 2003 through May 2019, with sightings overlaid. Note that survey effort, not shown here, was heterogeneous; a lack of sightings does not necessarily indicate that whales were absent. (Effort maps are included with the files that accompany this report.)

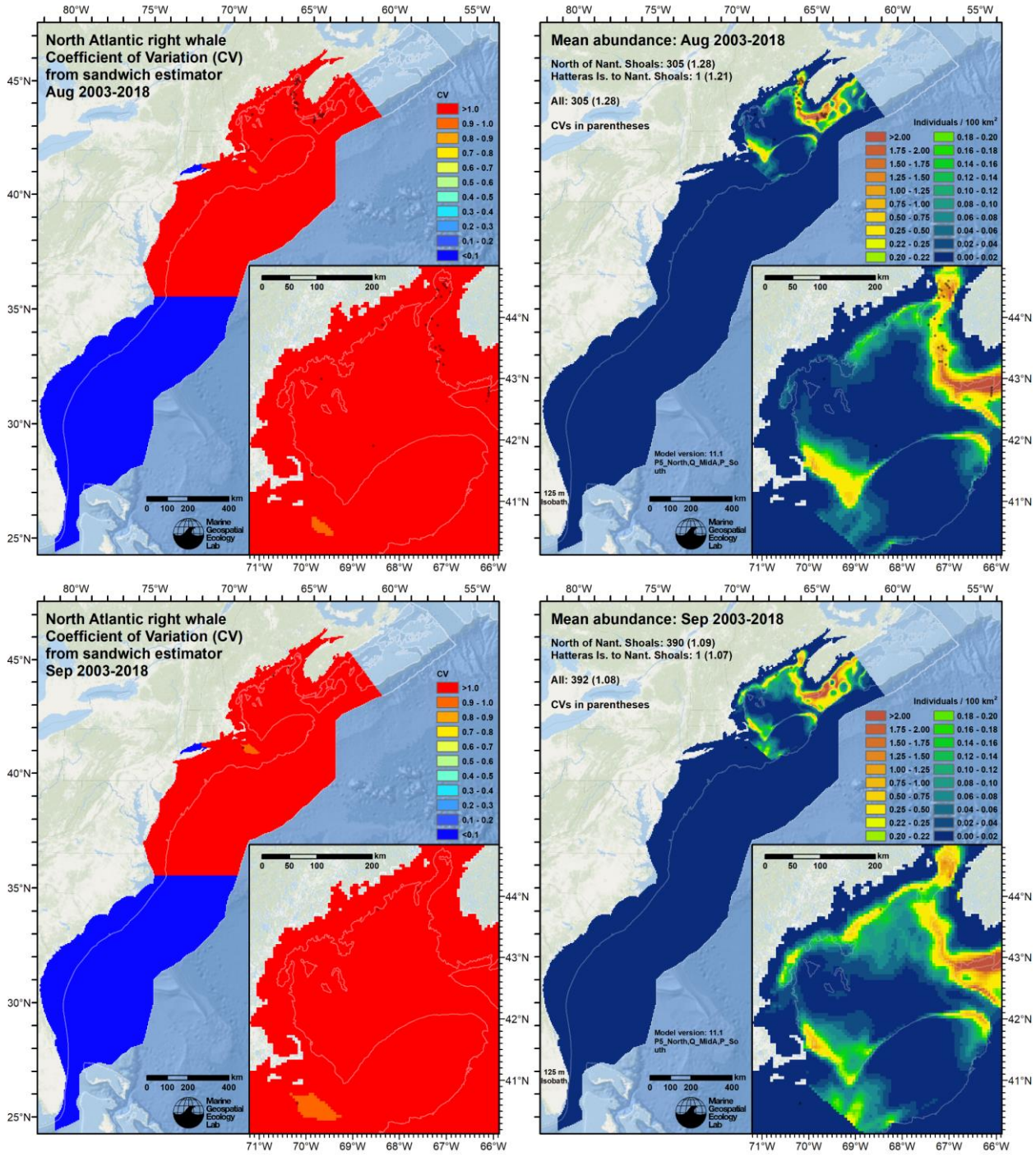


Figure 14. Continued from previous page.

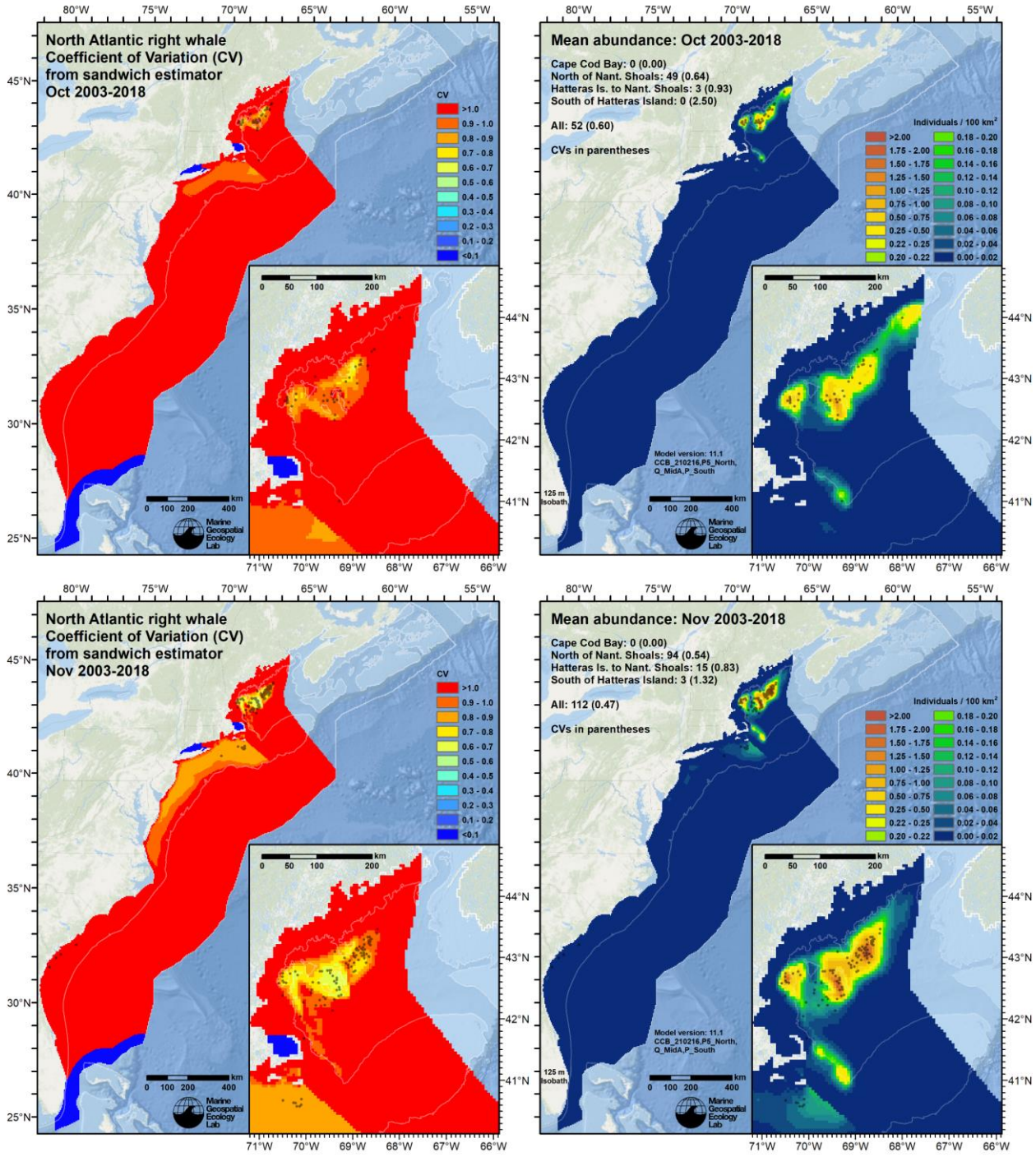


Figure 14. Continued from previous page.

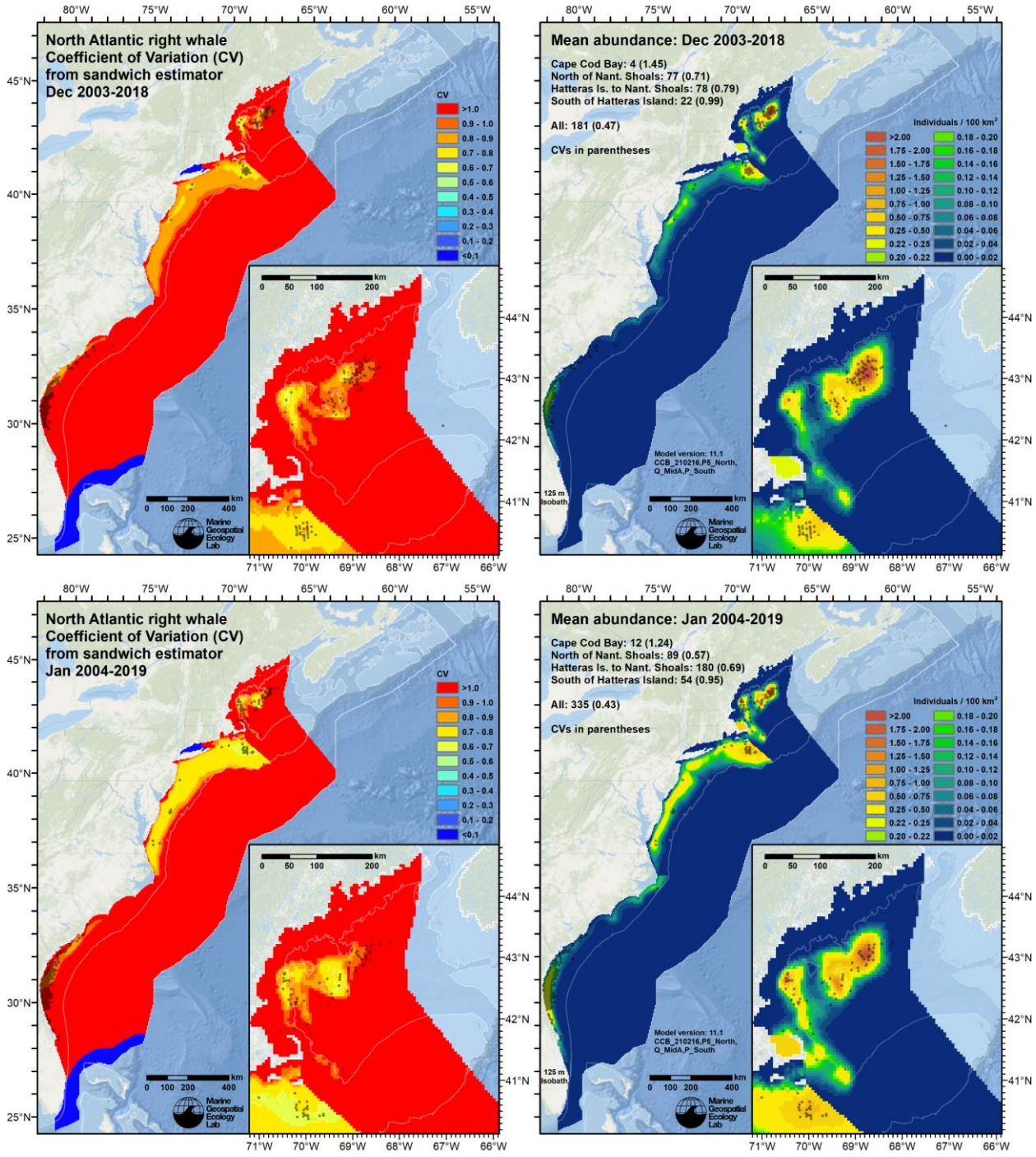


Figure 14. Continued from previous page.

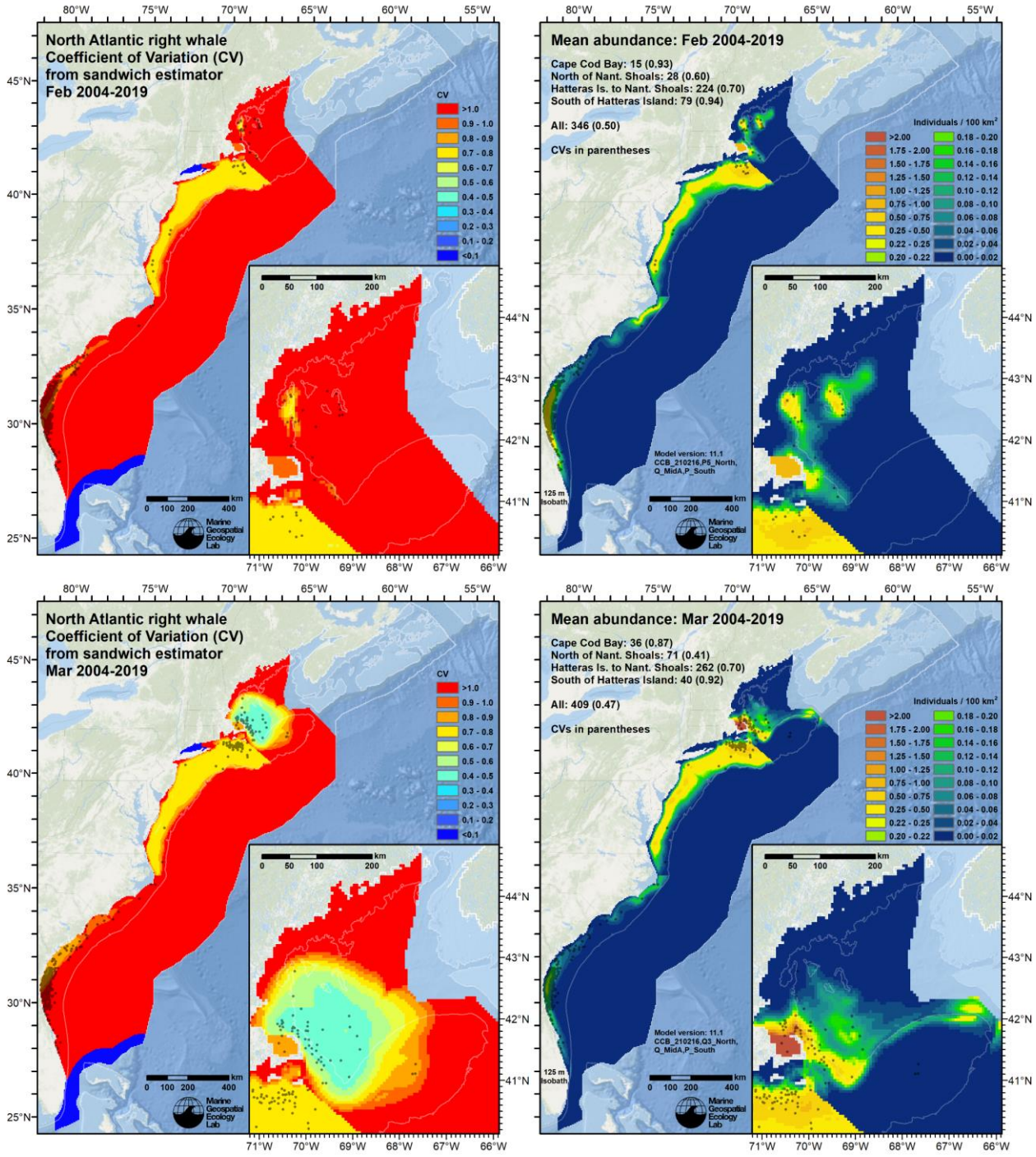


Figure 14. Continued from previous page.

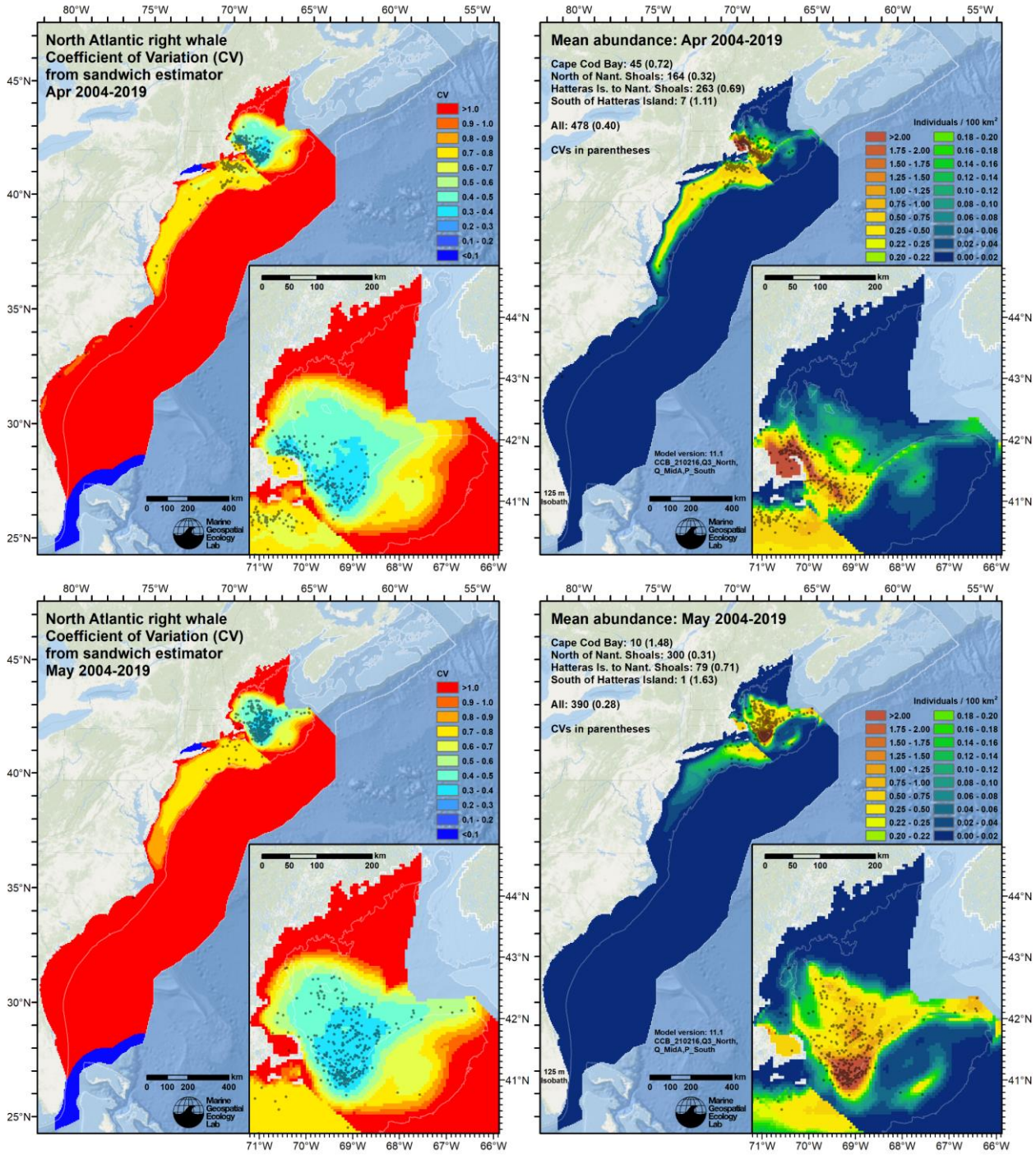


Figure 14. Continued from previous page.

5.2.2. Simulation-based summarization of uncertainty

5.2.2.1. Test Case 1: a simple univariate model to demonstrate the approach

As we expected, Depth was retained as statistically significant covariate (Figure 15, top) but explained little deviance (2.8%), reflecting the tendency of right whales to avoid both very shallow and very deep depths, but the need for additional covariates to define their habitat more precisely. The fitted relationship (Figure 15, bottom left) showed an extreme fall-off in density as depths increased beyond about 500 m (this is difficult to see in the plot), with widely expanding confidence intervals. The density prediction (Figure 16) reflected this relationship, with a density of zero predicted over the continental slope and the abyssal plain, low density predicted at very shallow depths near shore and the top of Georges Bank, and moderate density predicted across the intermediate depths everywhere else. The total estimated abundance of 515 (CV=0.04) was reasonable for such a crude model, considering that much of the population is believed to have inhabited this area during these months. The implausibly low CV reflects the large amount of data used but assumes that the model is "correct" insofar as it adequately accounts for spatiotemporal patterns in density. This is obviously not the case, as neither seasonal nor interannual variability is captured.

```
Family: Tweedie(p=1.311)
Link function: log

Formula:
IndividualsCorrected ~ offset(log(SegmentArea)) + s(Depth, bs = "ts")

Parametric coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -24.954      5.745  -4.344  1.4e-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(Depth) 3.596      9 8.887  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.00225   Deviance explained =  2.8%
-REML = 6707.5  Scale est. = 15.86     n = 52173
```

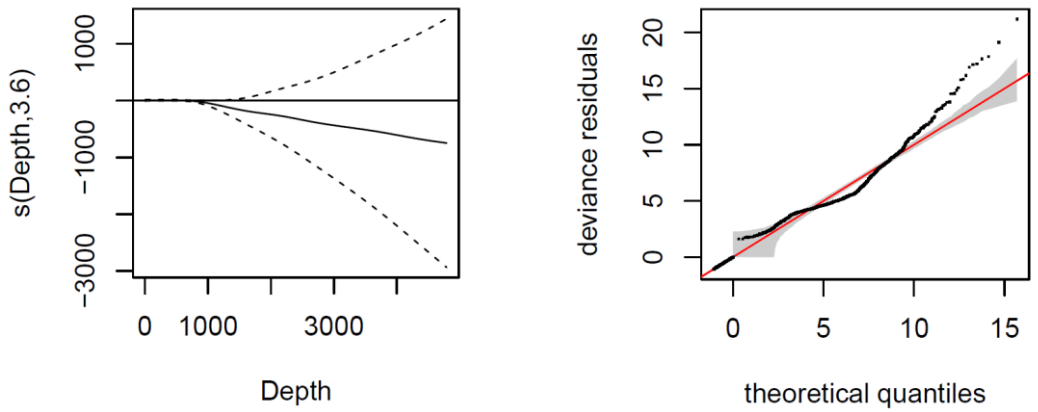


Figure 15. Diagnostic output for Test Case 1's model, including the model summary (top), functional plot for the single covariate (bottom left), and a Q-Q plot of deviance residuals (bottom right). Note that the y-axis of the functional plot is in natural log scale, the default for models fitted with a log link function.

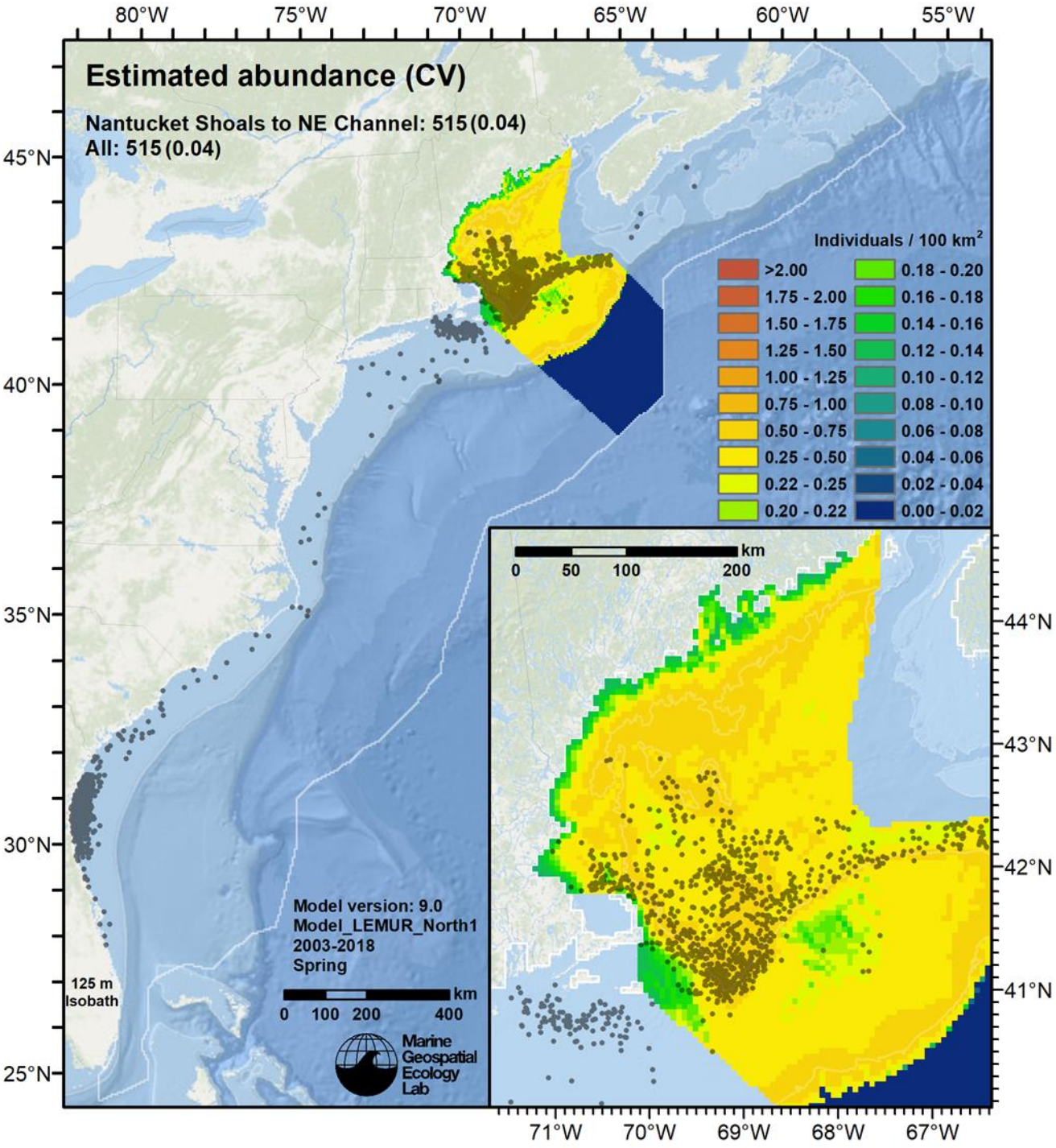


Figure 16. Density predicted by the Test Case 1 model, plotted using the color scale used consistently for the right whale model versions 8-11.

In areas where appreciable density was predicted, the analytic summarization yielded standard errors that were very low relative to their estimates except along a 1-2 cell strip along the continental slope (Figure 17). CVs were correspondingly low, except over the slope and abyss where near-zero density was predicted. (This pattern of high CVs where density is near zero is common in GAM-based density surface models and does not indicate cause for concern.) These uncertainty statistics again presume model correctness and do not reflect environmental variability.

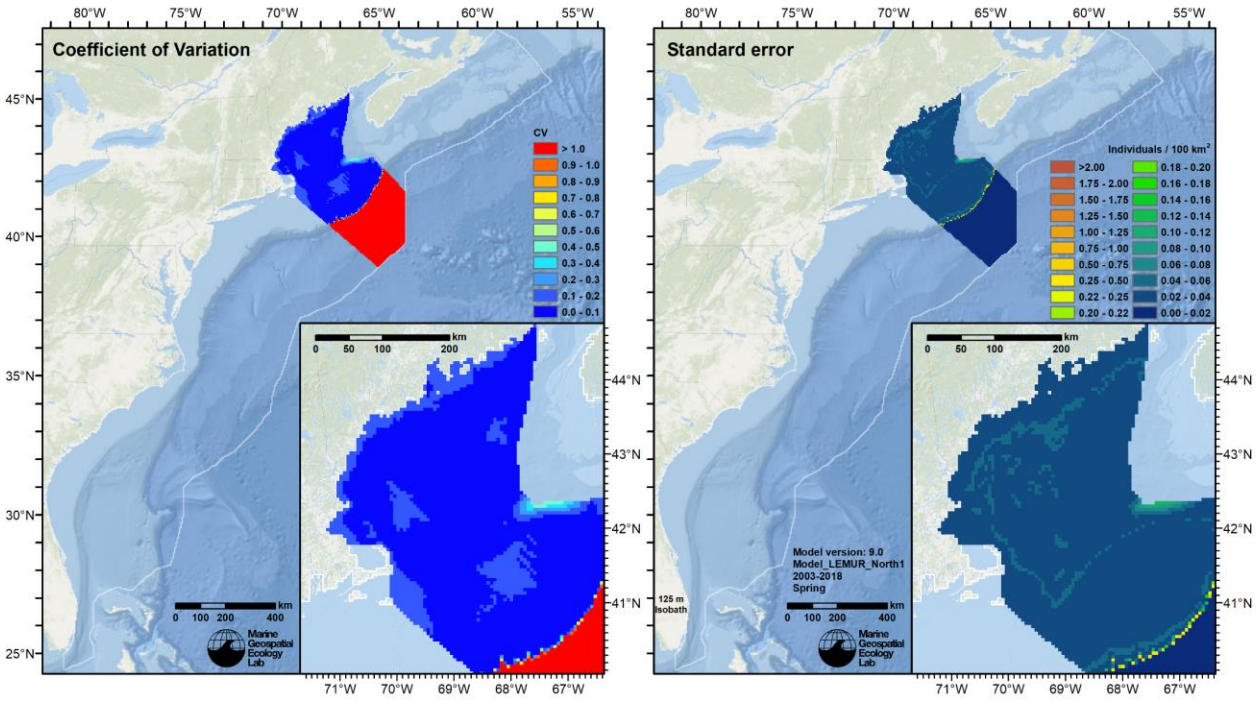


Figure 17. Standard error (right) and the corresponding CV (left), predicted for the Test Case 1 model and obtained by the analytic summarization method.

In our first attempt to apply the simulation-based method, we simulated 1000 alternative model parameterizations β_b under the assumption that parameter simulations should be distributed multivariate-normally about their means; that is, $\beta_b \sim N(\hat{\beta}, \mathbf{V}_{\hat{\beta}})$. A plot of the individual simulated models roughly reproduced the default confidence intervals output by R package `mgcv` (Figure 15) for the Depth covariate's functional relationship, as well as Nychka-type intervals (Figure 18). However, this assumption of multivariate-normality proved to be problematic, as it led to many implausibly extreme density estimates. For example, of the total abundance estimates for the first 20 simulated models, 15 might be considered reasonable, ranging from 495-1074, while the remaining 5 were implausibly extreme, ranging from 4.994×10^{23} to infinity:

```
Nsim[1:20]
```

```
##  [1]  5.255342e+02  5.110453e+02 4.089944e+173  5.228244e+02  5.004754e+46
##  [6]  5.275519e+02  4.994185e+23  5.078991e+02  5.033711e+02  5.056545e+02
## [11]  4.949608e+02  5.649067e+02           Inf 3.697967e+190  5.178185e+02
## [16]  5.151887e+02  1.073806e+03  5.243233e+02  5.107127e+02  5.027090e+02
```

Of the 1000 models simulated, 112 yielded total abundances of infinity. The 888 that remained ranged from 460-1.496×10^{307} with a median of 529 but a mean of 1.736×10^{304}. A histogram of the 888 estimates showed an extreme positive skew (Figure 19). Further examination of the distribution showed that estimates > 600 were very rare and widely distributed (Figure 20, left), while estimates < 600 exhibited only a moderate skew (Figure 20, right), resembling a log-normal distribution, as might be expected for count models. Of the 1000 simulations, 667 had abundance estimates < 600, with a mean of 520, vs. 515 for the fitted model that utilized the point estimates for the model parameters. The CV of the 667 estimates was 0.0461, vs. 0.0417 from the analytic approach.

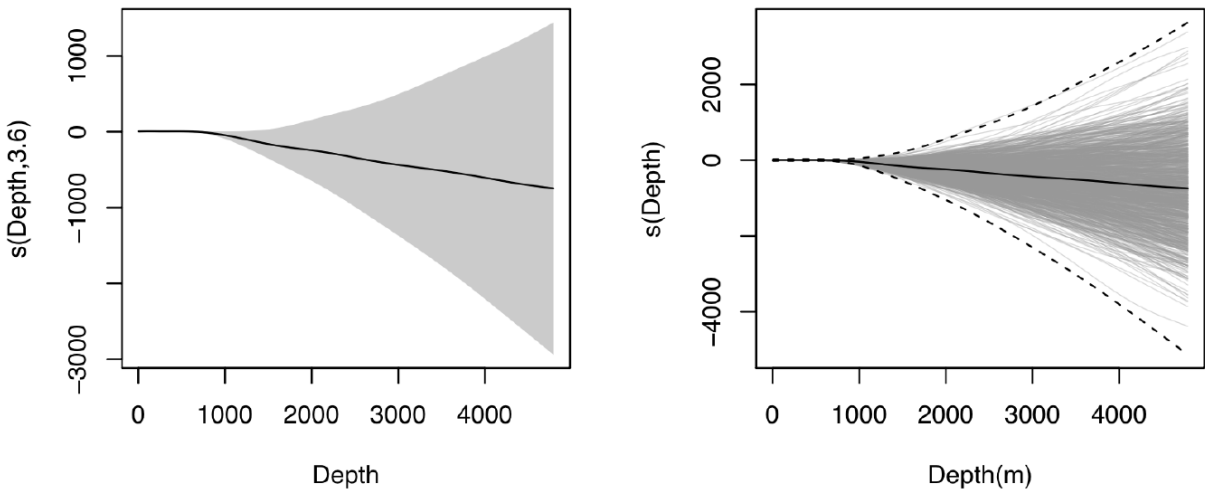


Figure 18. Left: Functional plot of Test Case 1 model's Depth covariate using Nychka-type confidence intervals. Right: Plot of 200 random model simulations (faint grey lines) with the default confidence intervals (dashed lines) output by R package `mgcv` (the same intervals are shown in Figure 15). Figure courtesy of D. L. Miller.

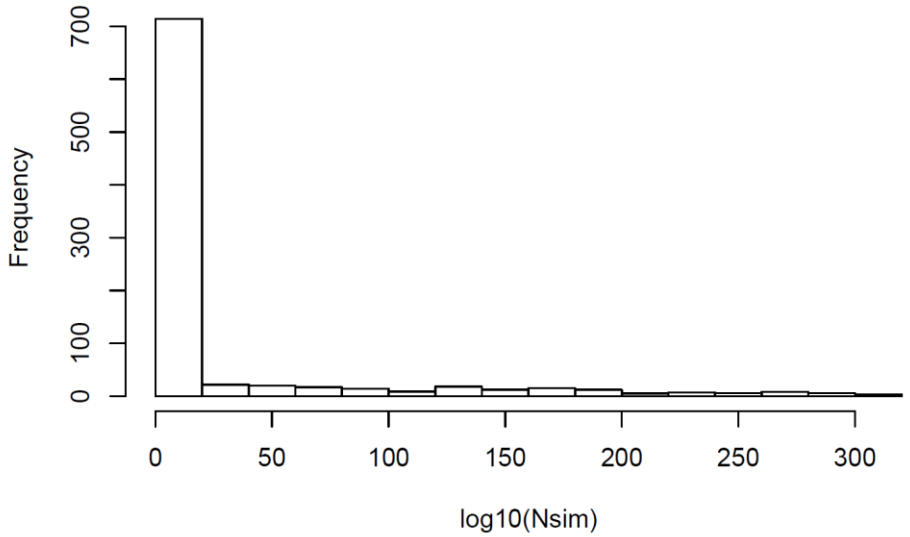


Figure 19. Histogram of the 888 non-infinite abundances from the simulated models. Note the x-axis is in $\log_{10}$ scale.

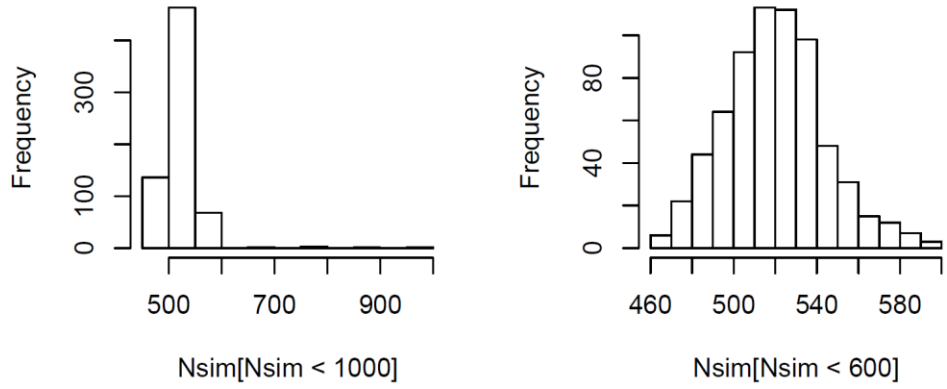


Figure 20. Histograms of abundances less than 1000 (left) and 600 (right). Here, the x-axis is in linear scale.

The obvious conclusion from this was model parameter simulations should not be distributed multivariate-normally and that we needed to devise a better approach to sampling their actual distributions. We made a second attempt using Metropolis-Hastings sampling, as implemented by the newly-implemented `gam.mh` function in R package `mgcv` (as initially implemented by S. Wood with a correction in May 2020 by L. J. Thomas), with parameters `burn=2000, t.df=40, rw.scale=0.05`. (We do not necessarily advocate these settings for general use; please consult the `mgcv` documentation and statistical experts for advice.) As before, we performed 1000 simulations.

This time, the simulated models displayed an asymmetric distribution about the fitted relationship and default confidence intervals (Figure 21). Importantly, functional plots of simulated models appeared not to exceed zero at deep depths. Because the model uses a log link function, when the model's response (abundance for the segment or grid cell) is calculated, predictions of the Depth term that are even moderately negative on the link scale will essentially add nothing to the model's intercept. The intercept was approximately -25 for this model (Figure 15), thus we anticipated that predictions for all simulated models would be essentially zero.

This expectation was almost correct, but of the 1000 simulations, 3 still yielded unrealistic, extreme total abundances (Figure 22). When these were trimmed, the remaining 997 simulations displayed a distribution (Figure 23) similar to that of the 667 simulations with abundance < 600 that we obtained with the multivariate normal assumption (Figure 20, right). The mean (518.1) and CV (0.042) of the 997 simulations were closer to those obtained analytically (515 and 0.0417) than were those of the 667 simulations that assumed multivariate normality (520 and 0.0461).

Maps of per-cell summarizations of SE and CV of the 997 simulated models (Figure 24) that were retained showed strong similarity to maps produced with the analytic method (Figure 17). The main area of deviation was over the slope and abyss, where densities and standard errors estimated by both methods were extremely close to zero but the simulation-based approach yielded bands of lower CV than the analytic method. Given that CVs are unstable when both the mean and standard error are close to zero, this difference is not important.

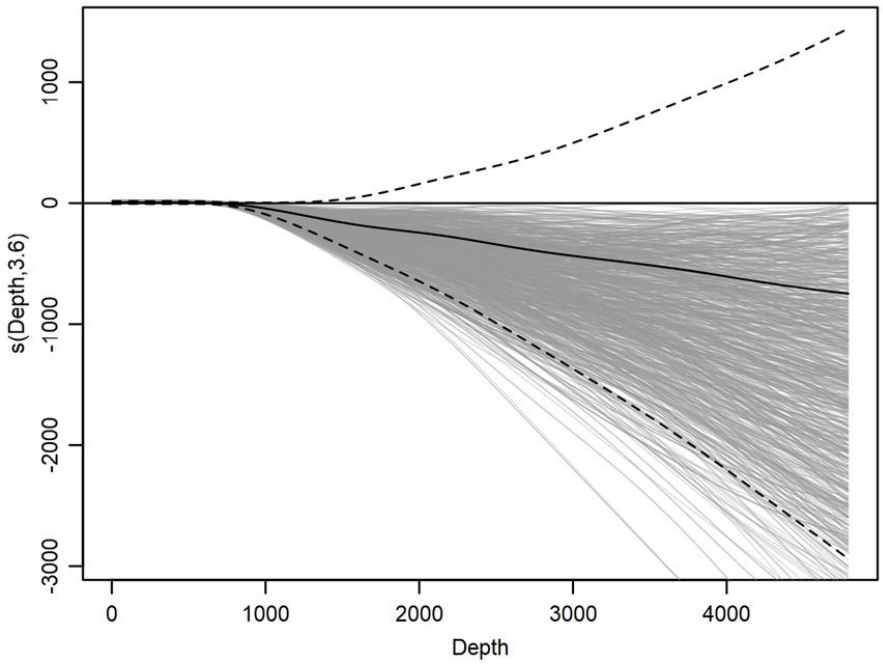


Figure 21. Functional plot of the Test Case 1 model's Depth term with a random selection of 200 model simulations (faint grey lines) generated with Metropolis-Hastings sampling, with the default confidence intervals (dashed lines) output by R package `mgcv` (the same intervals are shown in Figure 15). Note how for deeper depths, all simulations appear not to exceed zero, suggesting that they will not yield extreme predictions of density and abundance.

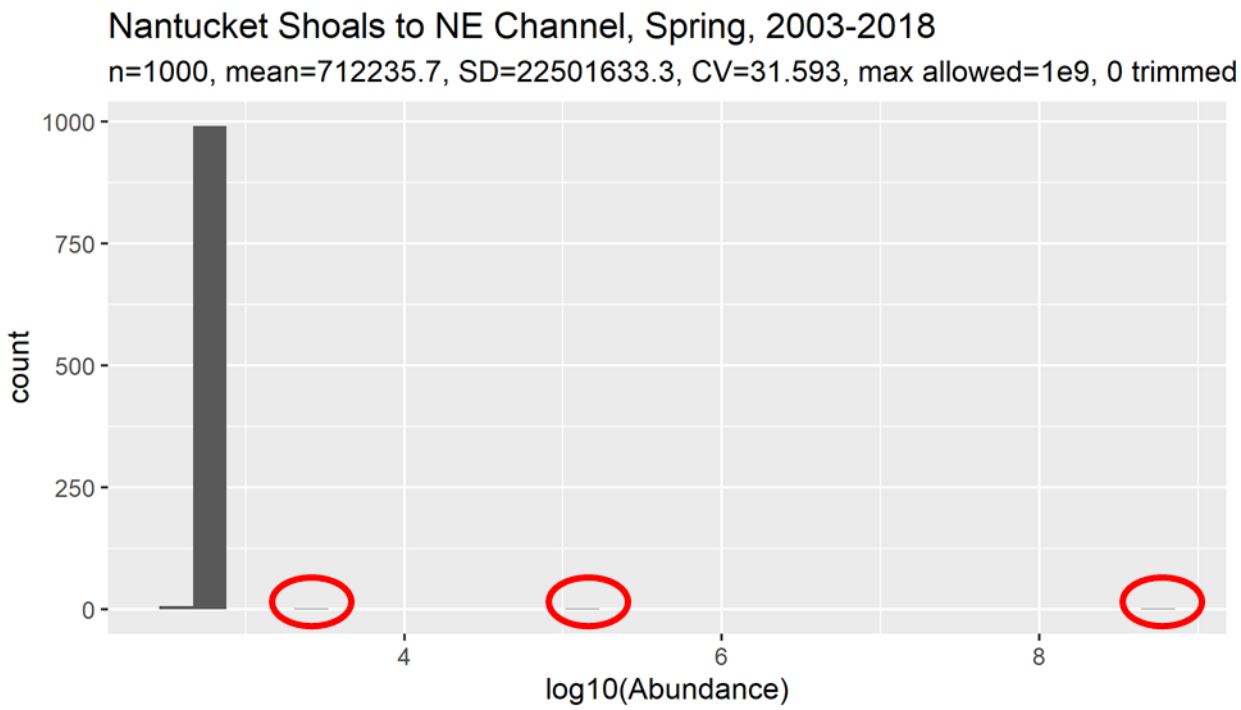


Figure 22. Histogram of abundance estimates from the 1000 model simulations generated with Metropolis-Hastings sampling. Note the three extreme estimates that were obtained (red circles).

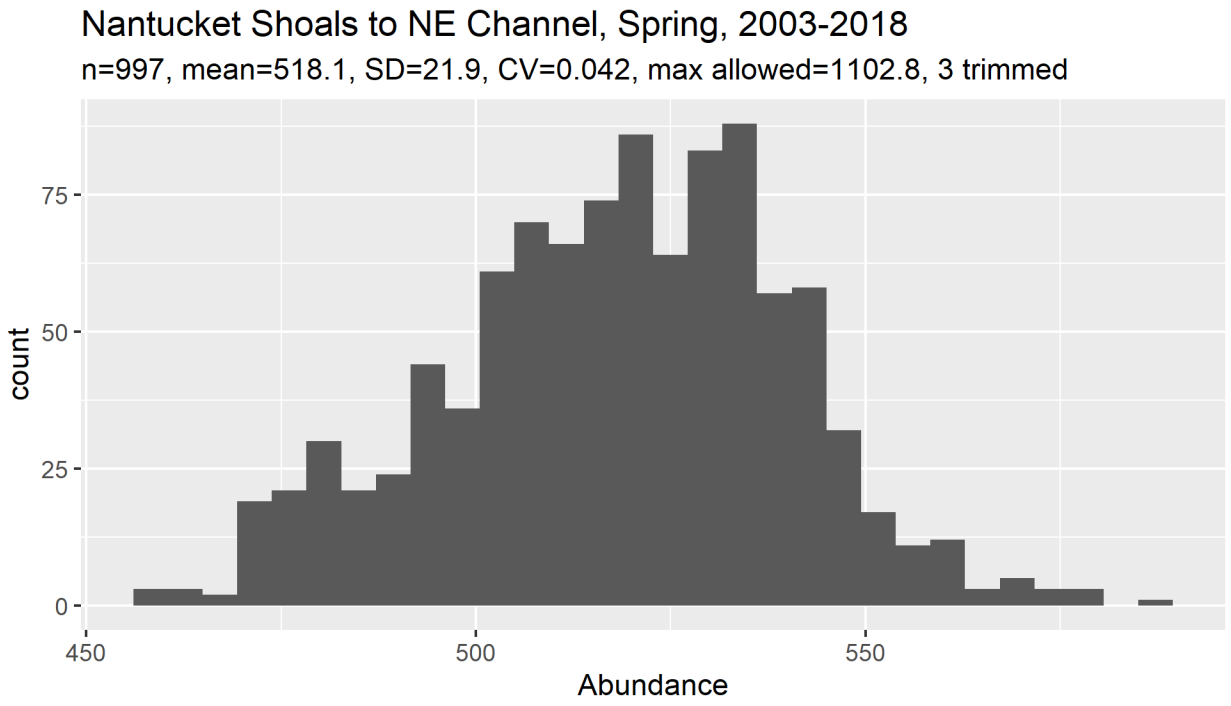


Figure 23. Histogram of abundance estimates of the 997 model simulations that remained after the 3 extreme results shown in Figure 22 were trimmed from the set. Here, the x-axis is in linear scale.

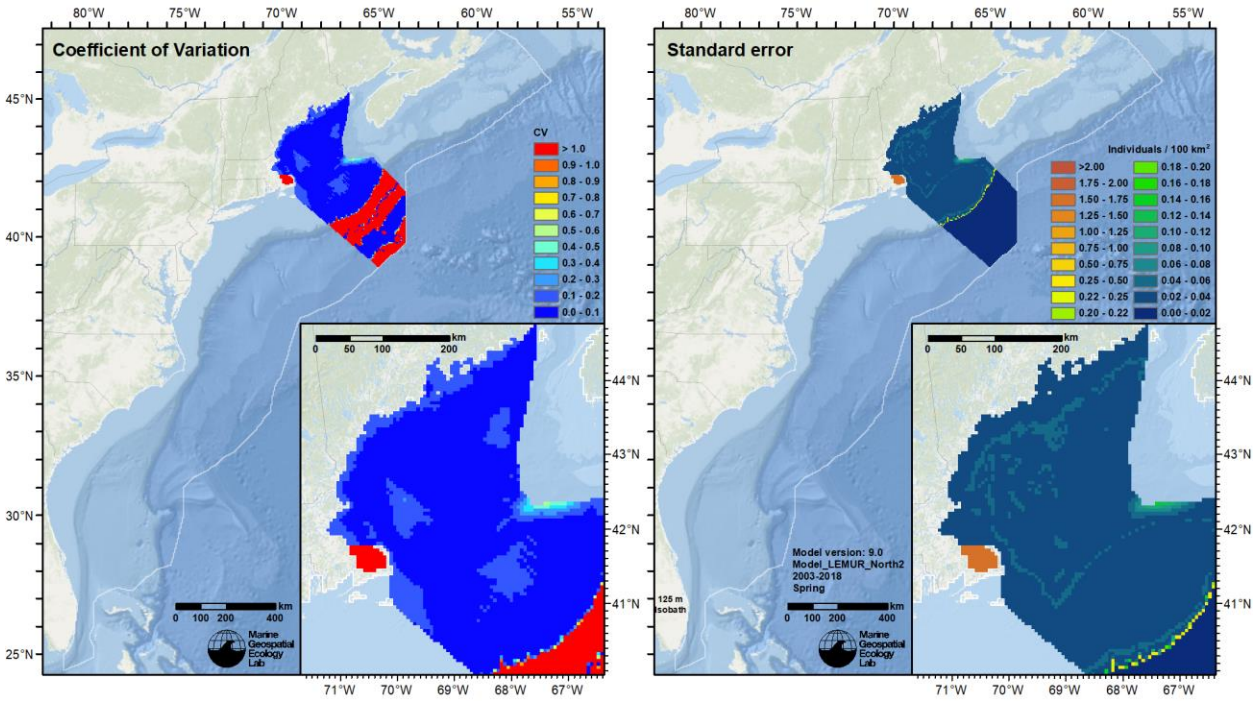


Figure 24. Standard error (right) and the corresponding CV (left), for the Test Case 1 model, summarized on a per-cell basis from the 997 random model simulations generated with Metropolis-Hastings sampling and retained for analysis. Note the strong similarity of these results to those obtained by the analytic summarization method (Figure 17). (Results for Cape Cod Bay, produced by a different model, were inadvertently retained in these plots and should be ignored.)

5.2.2.2. Test Case 2: the "South of Hatteras Island" region for "Winter" (October-May)

For the 200 models simulated here, the functional relationships of the 7 univariate model terms displayed more symmetric distributions about their point estimates (Figure 25) than occurred with the single univariate term in Test Case 1 (Figure 21). Several of them extended far toward the upper interval or exceeded it. This raised the prospects for obtaining a greater number of unrealistic density predictions than we obtained in Test Case 1, particularly for terms with larger y-axis limits or terms that would require a greater degree of extrapolation beyond sampled ranges. Our concern was particularly acute for the DistToShore term, which met both of these conditions and showed several simulations with upward trends that were close to or above zero at the maximum sampled value (roughly 190 km).

Distance covariates such as DistToShore can be problematic when extensive spatial extrapolation is necessary (Mannocci et al. 2017) because values in unsampled regions (e.g., far from shore) often greatly exceed values in sampled regions (e.g., close to shore). However, when this model was developed (in version 9), we found DistToShore to be important for preventing excessive density predictions directly adjacent to shore that were not supported by patterns in sightings. Given the utility of this, and that fitted relationship trended rapidly to zero at high values (Figure 25, top left, heavy line), we judged it suitable for extrapolation across the full "South of Hatteras Island" area. However, as we show later in this section, this term appeared to contribute problematic simulations that stymied our attempt to apply the simulation-based approach to uncertainty estimation.

In any case, our objective was not to refit any of the GAMs that together comprised the overall right whale model, only to characterize their uncertainties. Therefore, we proceeded with the approach prototyped in Test Case 1: Metropolis-Hastings sampling followed by the trimming of extreme results. An important difference was that we now had contemporaneous dynamic covariates, necessitating prediction and summarization of a series of time slices. For the 200 models predicted across the 8 modeled months, this comprised 11,200 density surface predictions for the 7-year 2003-2009 era, and 14,400 predictions for the 9-year 2010-2018 era.

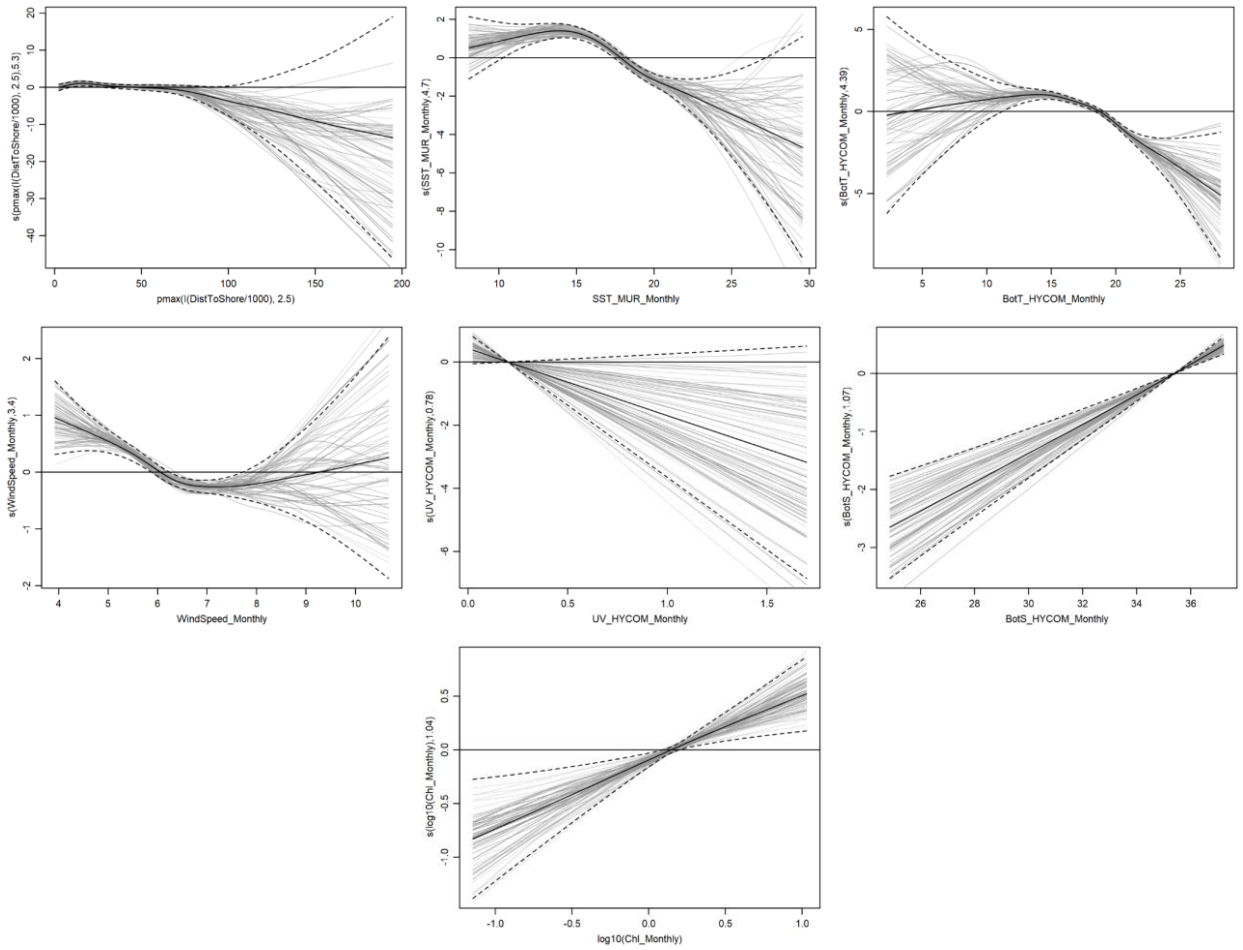


Figure 25. Functional plots of the Test Case 2 model's 7 univariate terms for the 200 simulated models (faint grey lines) generated with Metropolis-Hastings sampling, with the default confidence intervals (dashed lines) output by R package `mgcv` (the same intervals are shown in Roberts et al. (2020), Figure 79).

These predictions yielded a substantially greater proportion of extreme results than occurred in Test Case 1 (Figure 26). For both eras, more than 6% of the predicted density surfaces exhibited total abundances greater than 1 billion whales. After these were excluded, the mean monthly abundance still exceeded 2 million whales for both eras, due to remaining predictions that were less than 1 billion but still extreme (Figure 26). Nevertheless, the bulk of the predictions were within a reasonable range, so we endeavored to develop a systematic criterion for trimming them further.

We first thought to exclude some upper percentile of the predicted surfaces, but with over 6% already excluded, it seemed unsatisfactory to simply expand the percentage to an arbitrary level that seemed appropriate. We had done that previously for density models for the Arctic (Cañadas et al. 2020) and found that the percentile required to achieve a reasonable result varied between 15-30% from model to model, (although the Arctic analysis assumed that parameter estimates were distributed multivariate-normally, rather than using Metropolis-Hastings sampling). We settled on a two-step approach that first trimmed density surfaces having total abundances more than 100 times the median, and then trimmed any having greater than twice the 95th percentile of those that remained. Our goal was to first eliminate surface predictions that were patently implausible, then trim the very end of any long tail that remained. The thresholds used in this approach were still arbitrary but had the advantage that they could be applied to any model without developing a model-specific percentile through trial and error, or other model-specific tuning.

The new trimming approach eliminated 10.5% of the predicted surfaces from the 2003-2009 era, and 10.7% from the 2010-2018 era (Figure 27). Mean monthly abundances for the eras across the October-May "Winter" season were

48.6 and 13.8, respectively. The large difference between the eras is consistent with findings that calving dropped to pre-2000 levels in 2010 (Meyer-Gutbrod et al. 2021) and migration to the calving grounds decreased substantially after 2011 (Gowan et al. 2019). The season-wide means for both eras are low because they included months in which density is believed to be very low. By contrast, for February, often the month in which abundance was predicted to be highest in the region, the means were 150.7 and 39.0 for the two eras, respectively (Figure 28).

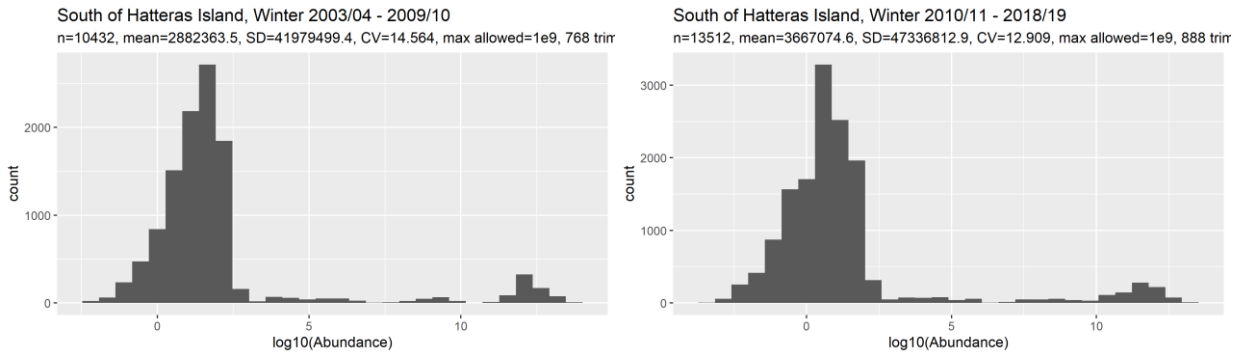


Figure 26. Histograms of abundance estimates for Test Case 2, for the 2003-2009 era (left) and 2010-2018 era (right). Note that the x-axis is on $\log_{10}$ scale. The subtitles show the statistics that resulted when abundance estimates greater than 1 billion whales were trimmed. Note the large number of extreme predictions that still remained (e.g., between 1000 and 1 billion). The large number of very low predictions (e.g. < 0 on $\log_{10}$ scale) were not considered problematic, as the model included months in which right whales were expected to be largely absent from the area.

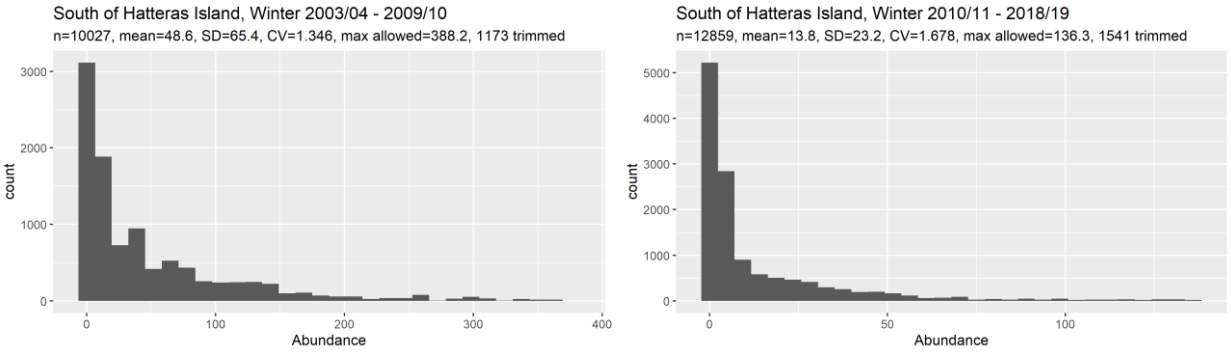


Figure 27. Histograms of abundance estimates for Test Case 2 after additional extreme predictions were trimmed (see text). Here, the x-axis is in linear scale.

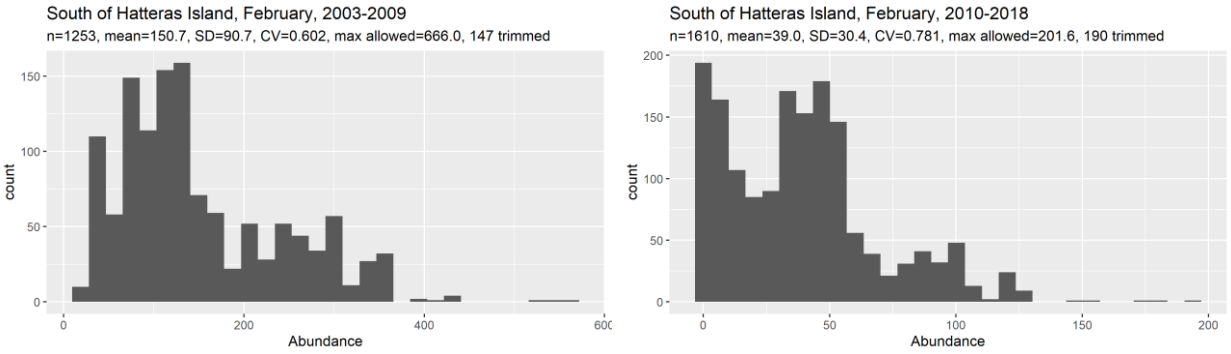


Figure 28. Histograms of abundance estimates for the month of February—often the month of highest predicted abundance at the calving grounds—for Test Case 2 after extreme predictions were trimmed as in the previous figure.

Maps of monthly standard error for the two eras (Figure 29, Figure 30), summarized from the density surface predictions that remained after trimming, showed very strong resemblance to SE obtained through the analytical method over the continental shelf, defined here to be waters shallower than 125 m (compare left and right insets in the referenced figures). This suggested the simulation-based method yielded reliable results in regions with sufficient surveying. However, beyond the shelf (waters deeper than 125 m), results were less consistent. In most regions, both methods estimated standard errors very close to zero. But in several small patches far offshore of Florida, and one location along the continental slope off North Carolina, the simulation-based method estimated high standard errors, comparable to those occurring on the shelf. These patches reoccurred across the months for both eras, at the same locations but with varying intensity. This suggested that static covariates such as DistToShore or the bivariate smooth of spatial location were driving the relationship, but that the intensity was being regulated by a dynamic covariate such as SST or wind speed.

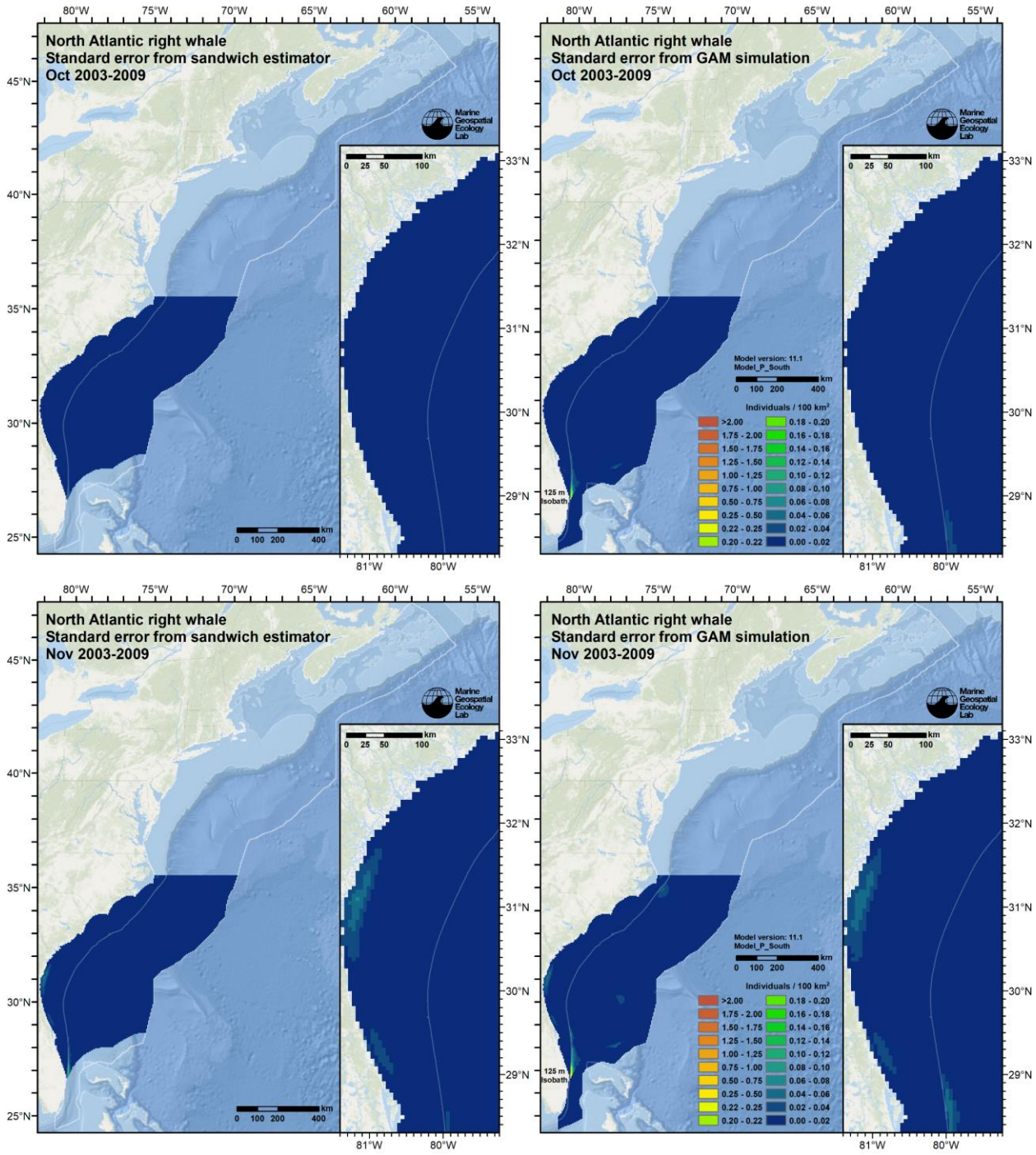


Figure 29. Monthly maps (rows) of standard error from the analytical method (left column) and the simulation-based method (right column) for Test Case 2 for the 2003-2009 era, spanning October 2003 through May 2010. Note the patches of higher standard error estimated by the simulation-based method that occur on the deeper (right) side of the 125 m isobath (gray line), particularly in the December-April period.

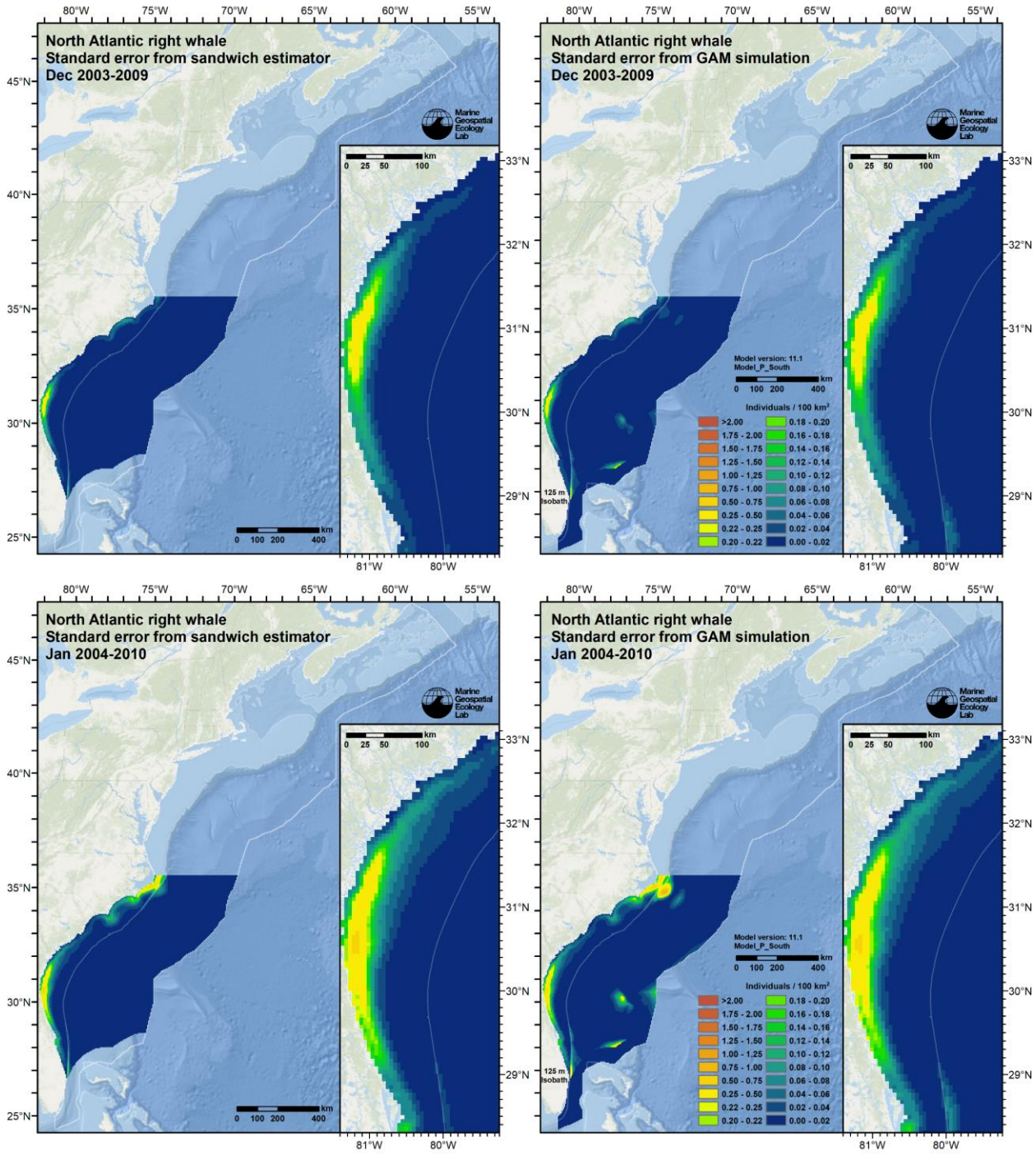


Figure 29. Continued from previous page.

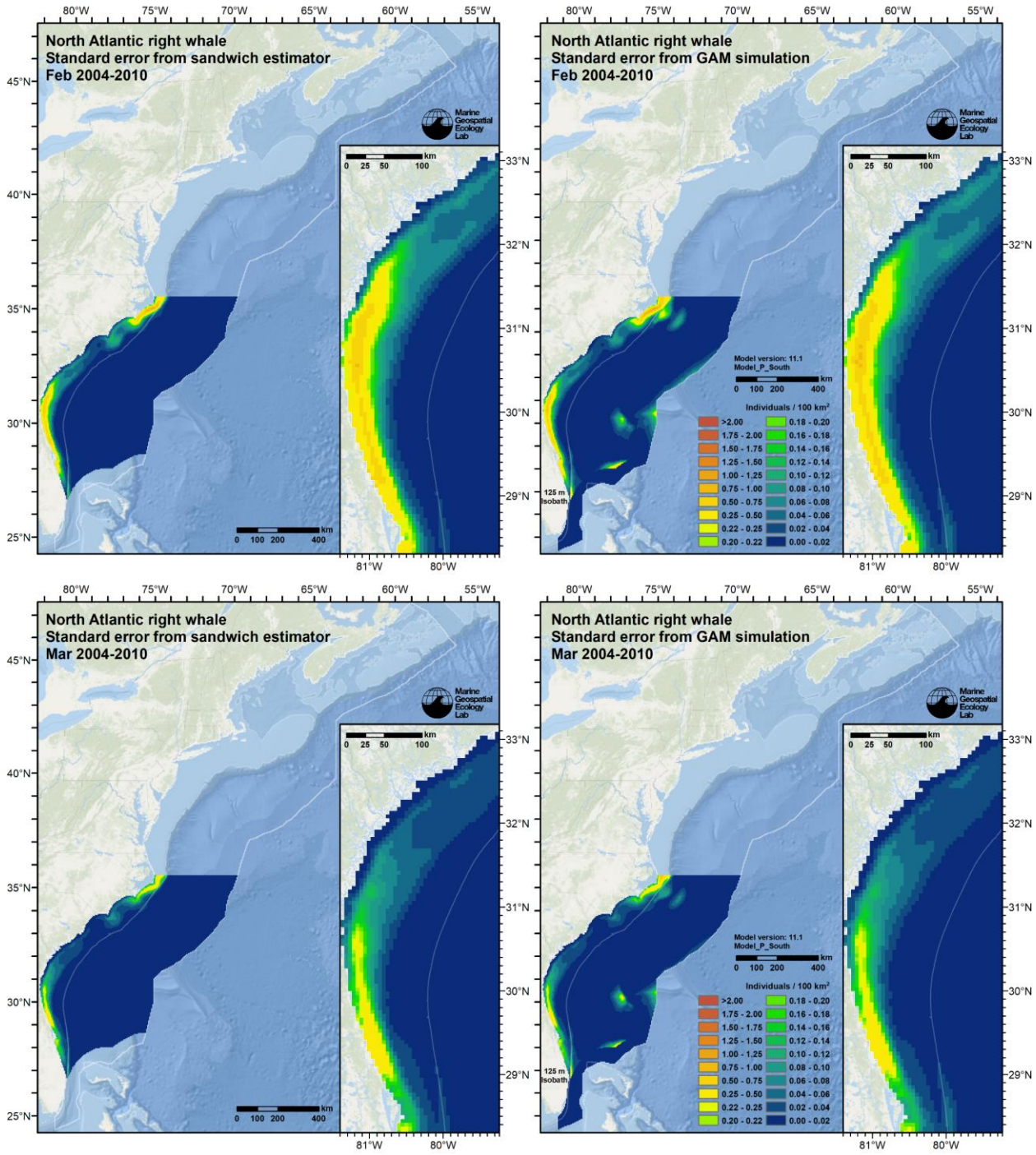


Figure 29. Continued from previous page.

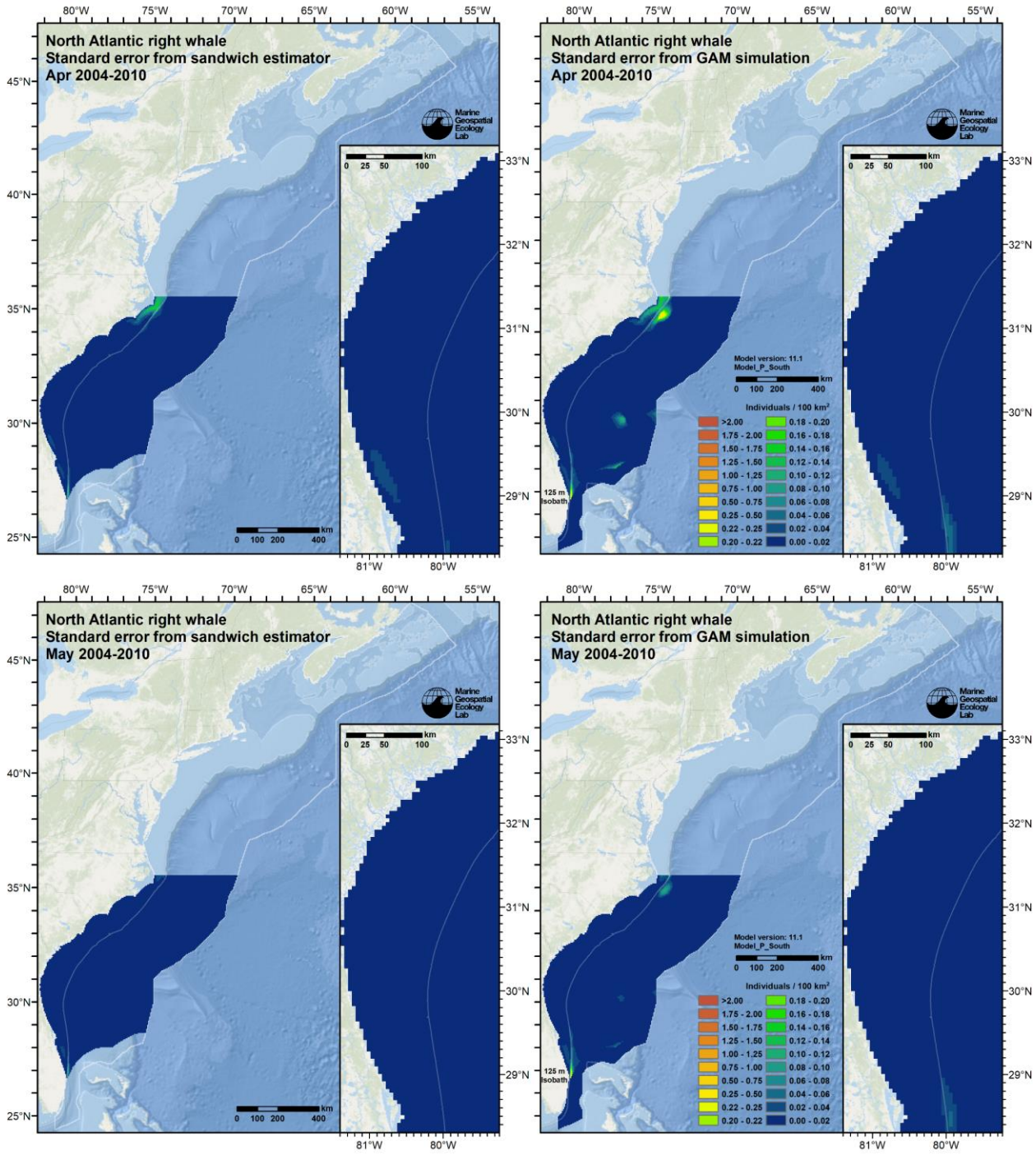


Figure 29. Continued from previous page.

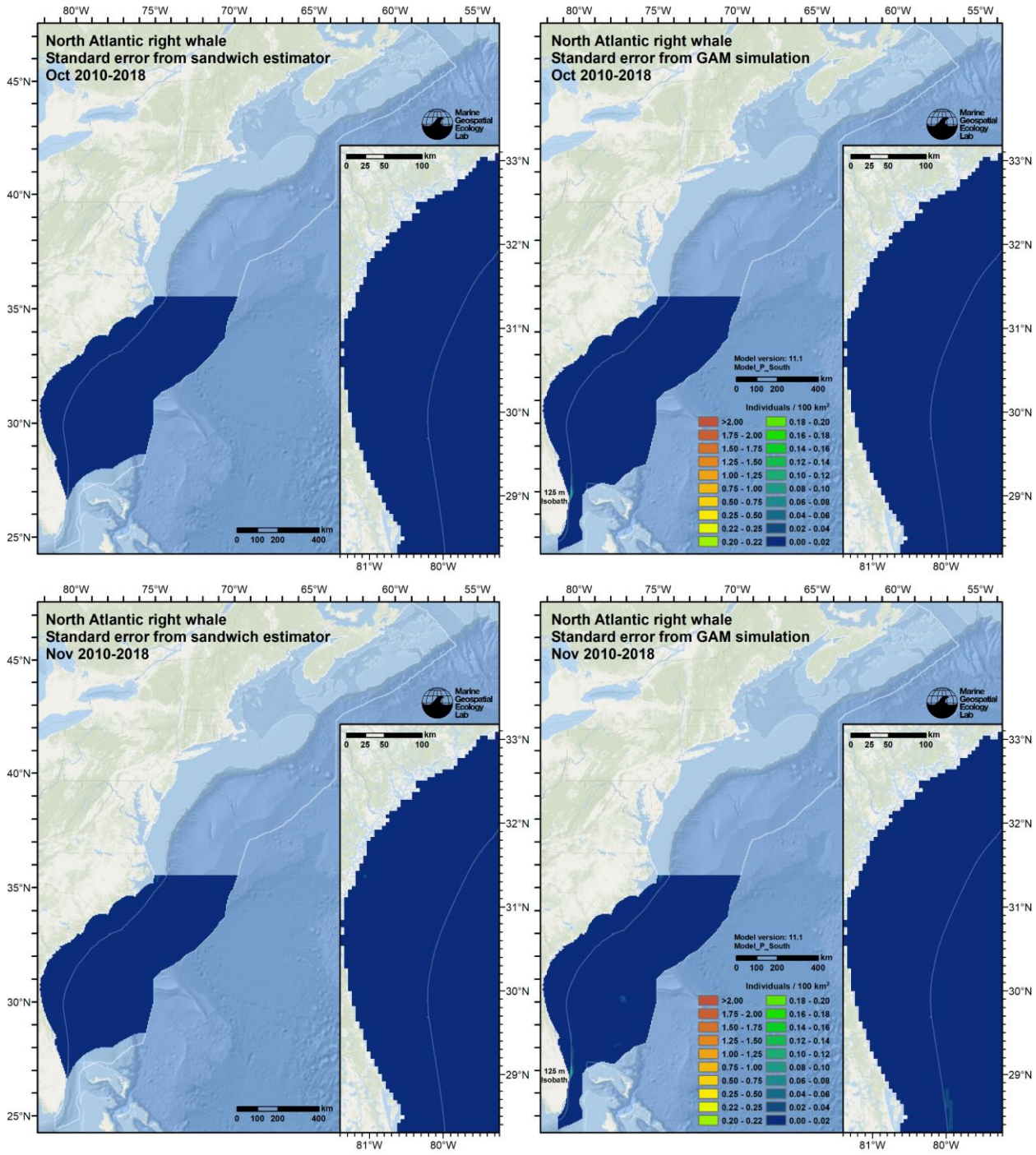


Figure 30. Monthly maps (rows) of standard error from the analytical method (left column) and the simulation-based method (right column) for Test Case 2 for the 2010-2018 era, spanning October 2003 through May 2010. Note that standard errors are lower for this era than for the 2003-2009 era (Figure 29), owing to the overall higher density in the region during that era, but as in that era, the simulation-based method shows patches of higher SE than the analytical method, mainly on the deeper (right) side of the 125 m isobath (gray line), but also on the shallower side between Cape Lookout and Hatteras Island, NC, in January-March.

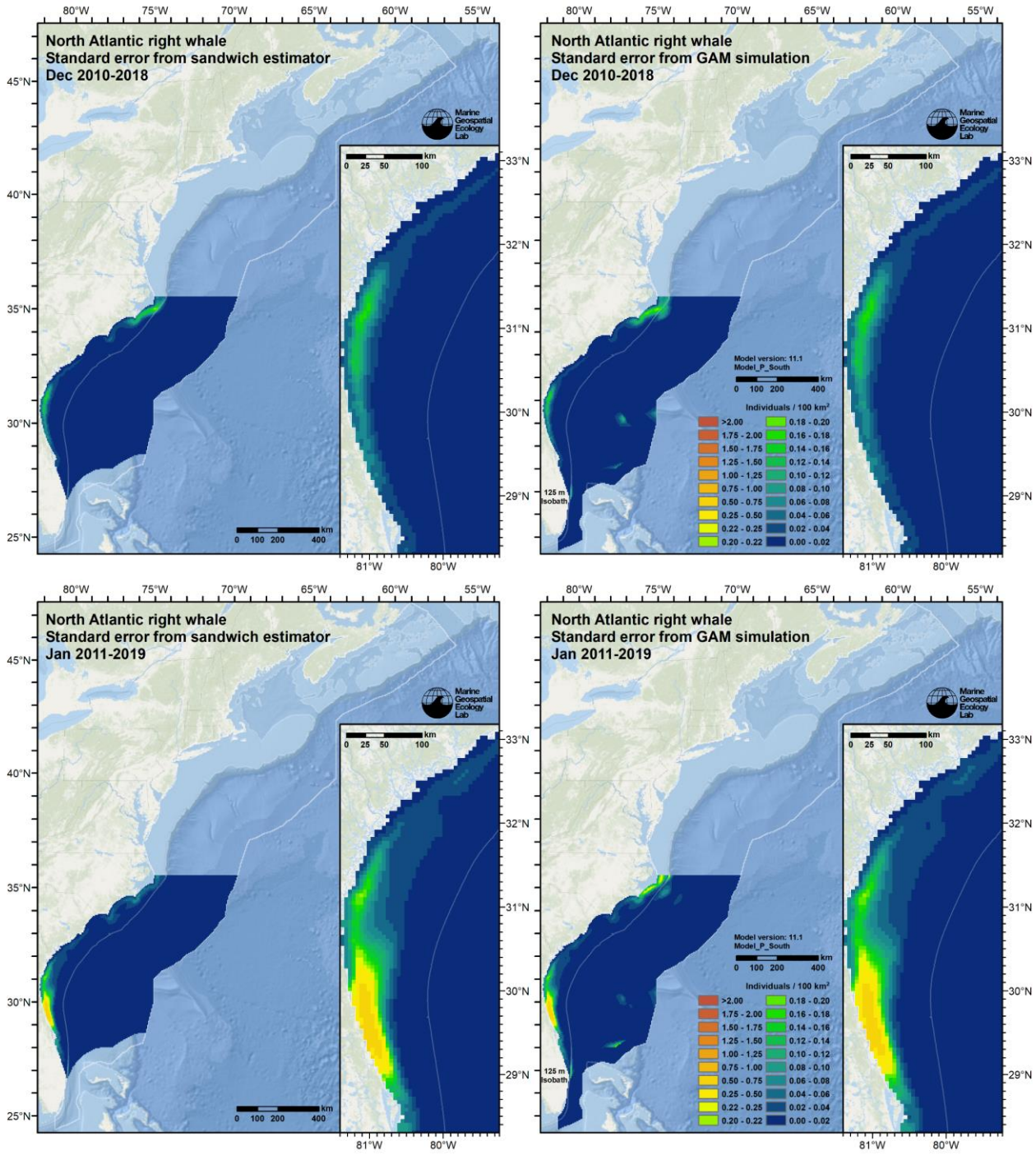


Figure 30. Continued from previous page.

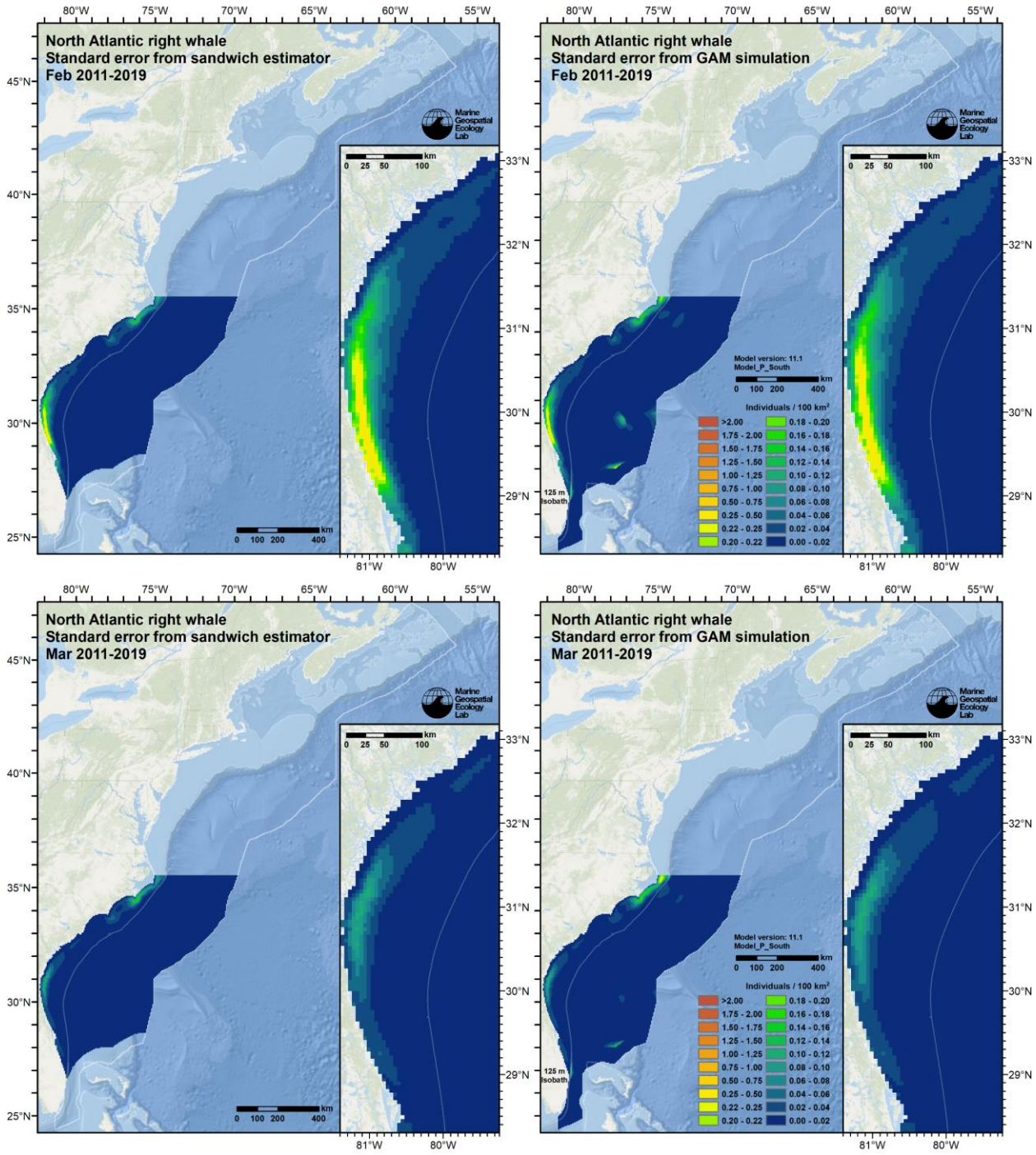


Figure 30. Continued from previous page.

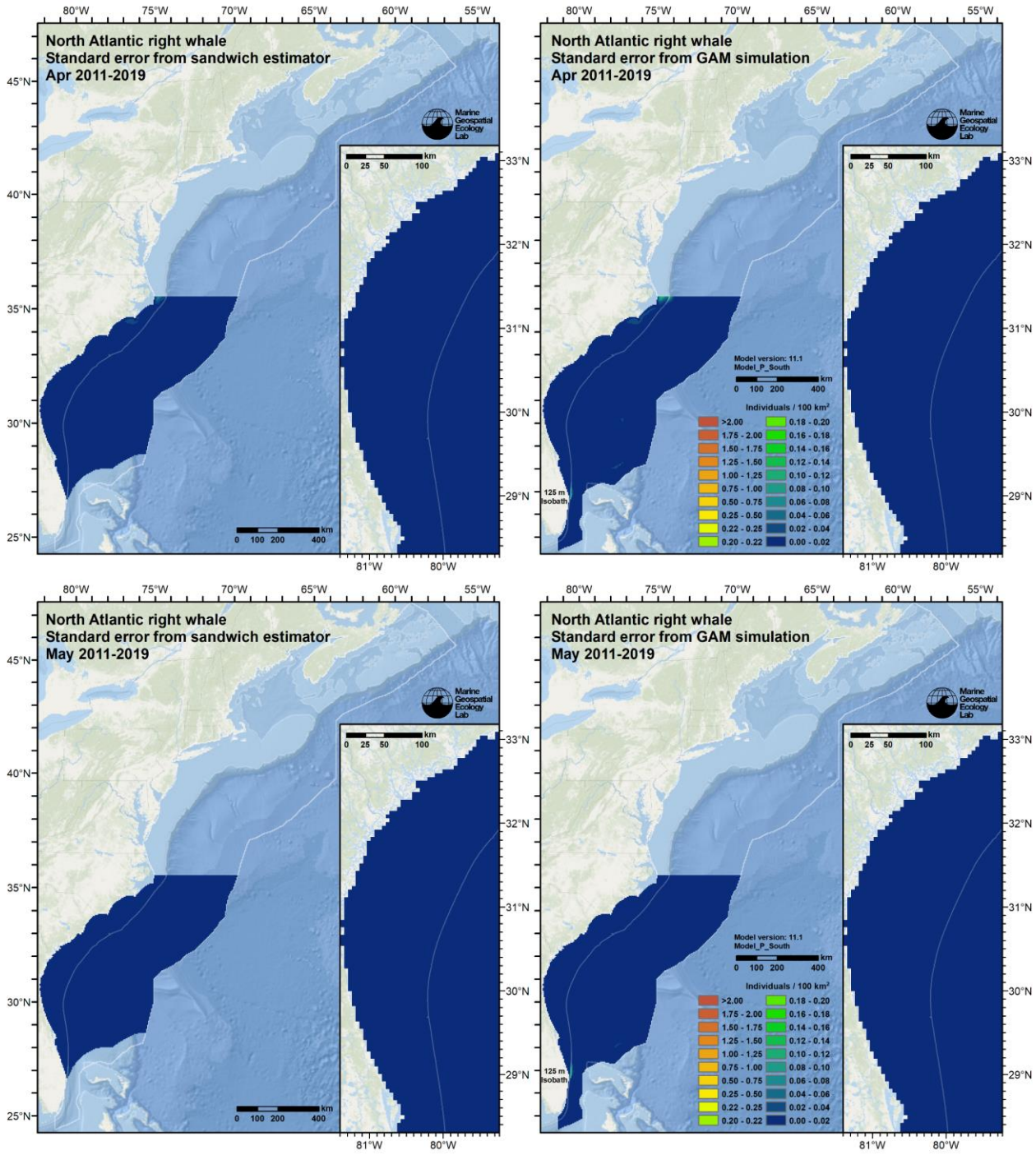


Figure 30. Continued from previous page.

5.3. Discussion

In this analysis, we prepared uncertainty estimates for version 11 of the right whale density model using two methods. Both methods accounted for *model uncertainty*, defined as the statistical error in model parameter estimates, and, when dynamic covariates were used, for *environmental uncertainty*, defined as the variance in predictions induced by the dynamic covariates, e.g., interannually. The first method used an analytic approach that yielded the summary maps and statistics requested by the Navy (Section 5.2.1). The second method used a simulation-based approach that yielded an additional useful output: a collection of plausible alternative density surfaces that "could have happened" if uncertainty was accounted for (Section 5.2.2.2). The intent was that model users could incorporate these alternative surfaces into their own simulations (e.g., with a Monte Carlo approach) to characterize the uncertainty of their own models in a way that better accounts for spatial and temporal covariance.

Unfortunately, we were only partially successful in applying the second approach to the right whale model. In our focal test case, the regional "South of Hatteras Island" model for the October-May "Winter" season (Section 5.2.2.2), over 10% of the simulated alternative density surfaces exhibited anomalous patches of extreme density. Outside of these anomalies, summaries of the simulated surfaces matched summaries obtained with the analytic method. However, for the simulated surfaces to be fit for general use, we need to debug and correct these anomalies, which will require several additional months of analysis and experimentation. Given that, and that the analytic results meet the Navy's needs, we are releasing the analytic uncertainty summaries now as version 11.1 of the model and deferring the simulation-based outputs to a future release.

We start this section by interpreting the analytic summarizations of uncertainty and the implications for model users. We then discuss the simulation-based method and the prospects for utilizing it successfully in the future. We conclude with a brief discussion of other sources of uncertainty not accounted for here.

5.3.1. Interpretation of analytic uncertainty estimates

Here, we discuss the summary CV maps (Figure 12, Figure 13, Figure 14) for each regional sub-model, in the same order the models were originally presented in Roberts et al. (2020), Section 5.3.2, which documented the models and the decisions that went into developing them. For brevity, we omit explicit references to figure numbers for the CV maps except in special circumstances.

5.3.1.1. South of Hatteras Island

In this region, right whales are mainly present in winter and early spring and absent at other times. Accordingly, the year was modeled in two seasons, "Summer" and "Winter". In Summer months, defined for this region as June through September, no right whales were sighted, and density was assumed to be zero. (See Roberts et al. (2020), Section 5.3.2.1, for a discussion of this and other modeling decisions and details.) CV was therefore also set to zero.

In the Winter months, defined as October through May, a complex model with static, dynamic, spatial, and temporal covariates was fitted. Predictions for all eras showed similar intraannual patterns in density, but high interannual variability. Over the continental shelf, where virtually all sightings occurred, predicted density was near zero during the beginning months (October and November) and ending month (May) of the season. During these months, CV exceeded 1 in most locations, as usually occurs in GAM-based density surface models when density predictions approach zero. This also happened during all months over the continental slope and abyssal plain, where right whales are believed to occur very rarely. The high CVs at these locations and during the beginning and ending months are not a cause for concern: both the mean and standard error (SE) were both very low, meaning that it was very unlikely that there was any appreciable density even when uncertainty was accounted for, and the high CV was a side effect of the very low SE exceeding the even-lower mean. Whenever CVs are high, the SE should then be examined to determine if the magnitude of the error is high enough to be of concern. If not, the high CVs can be ignored.

For the continental shelf during December-April, the location and months typically inhabited by right whales, CVs ranged from 0.3 to more than 1. CVs were lower in the northern part of the core calving habitat, from 30-33 °N,

and higher in the southern part, south of 30 °N, reflecting greater interannual variability in the south, consistent with evidence that whales typically distribute further south in relatively cold years (Keller et al. 2006, 2012; Gowan & Ortega-Ortiz 2014).

CVs were considerably lower in the 2003-2009 era (maps in Figure 12; total abundances in Figure 31) than in the 2010-2018 era (maps in Figure 13, total abundances in Figure 32), reflecting the decline in density at the calving grounds following the 2009/10 season and the higher annual CVs in the latter era (Figure 33). The general pattern of rising density in the first era followed by a fall in the second was supported by a similar pattern in annual counts of unique, photographically-identified individuals spotted south of Virginia (Figure 34).

It is possible that as additional years of data are accumulated into future model updates, the model uncertainty component of SE and CV will decrease. However, if interannual variability remains high, as we anticipate will happen until right whales again achieve reliable success in summertime foraging and calving activity returns to higher levels, the environmental uncertainty component of SE and CV will remain high.

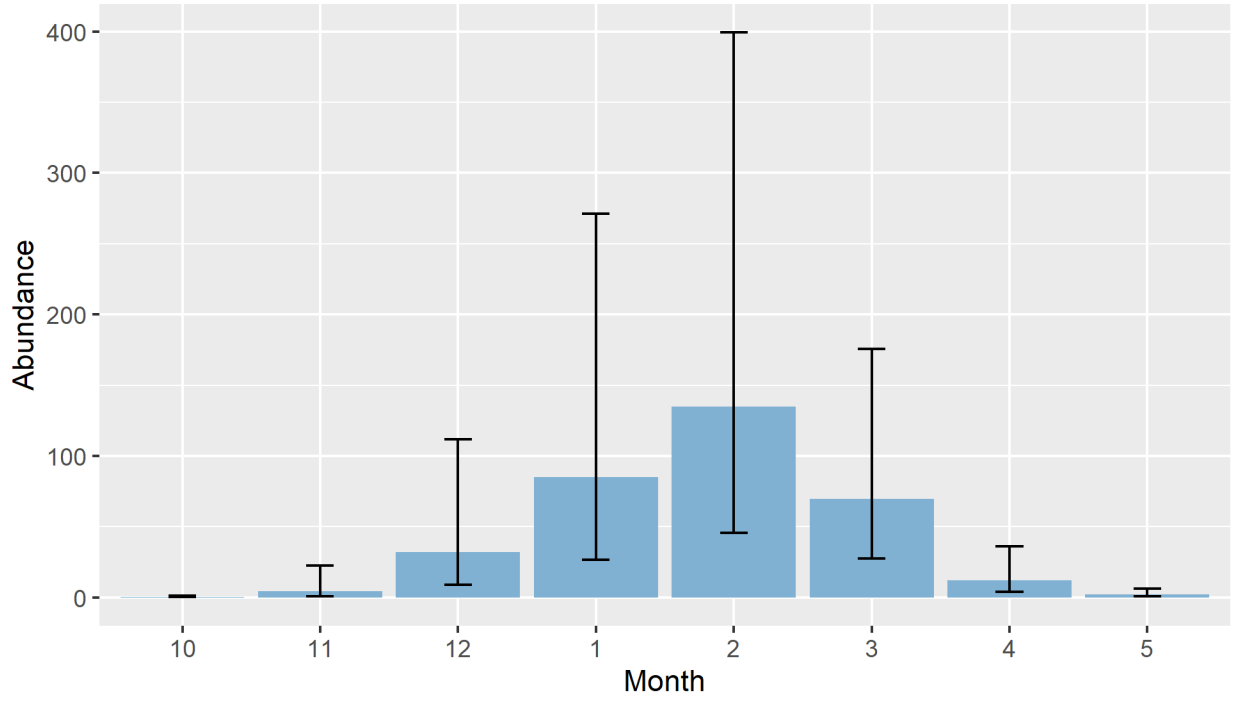


Figure 31. Monthly mean total abundance for the "South of Hatteras Island" region, with approximate asymptotic 95% confidence intervals, for the 2003-2009 era. Wide confidence intervals indicate high CVs (text in Figure 12 gives abundance and CV values). These were driven mainly by interannual variability in right whale density (Figure 33).

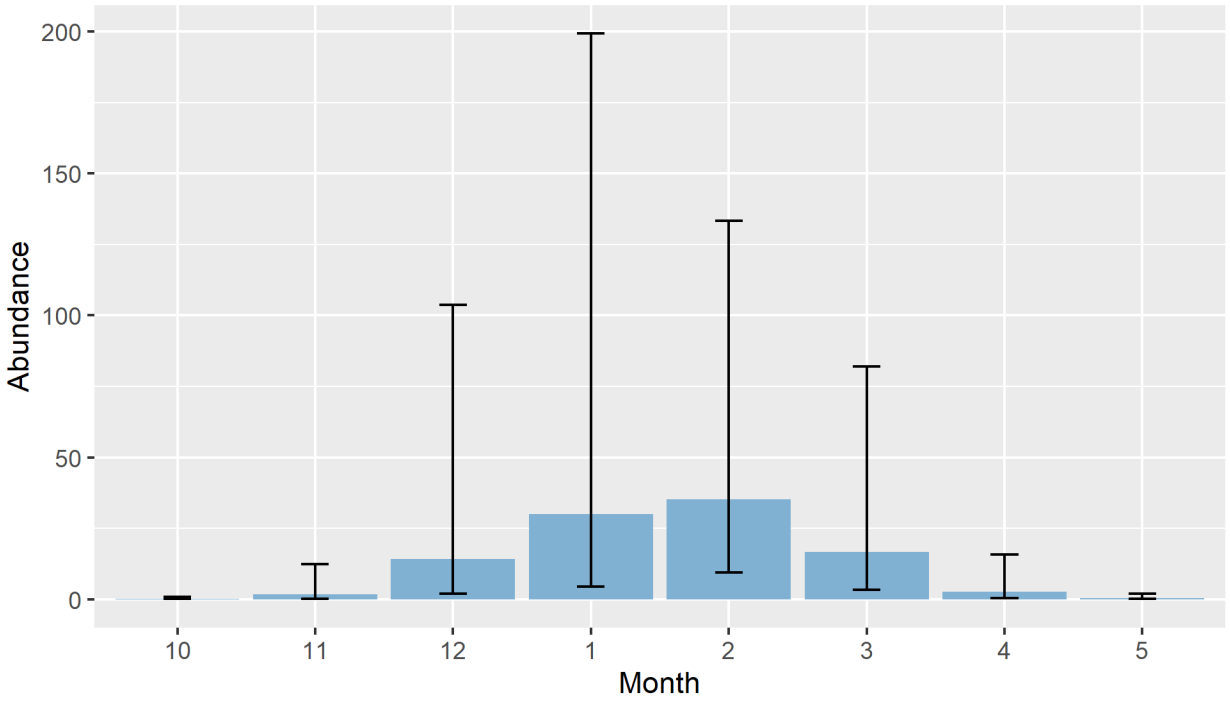


Figure 32. Monthly mean total abundance for the "South of Hatteras Island" region, with approximate asymptotic 95% confidence intervals, for the 2010-2018 era. Note that the y-axis has a lower limit than the previous figure. The wider confidence intervals relative to the means than in the previous figure indicate even higher CVs.

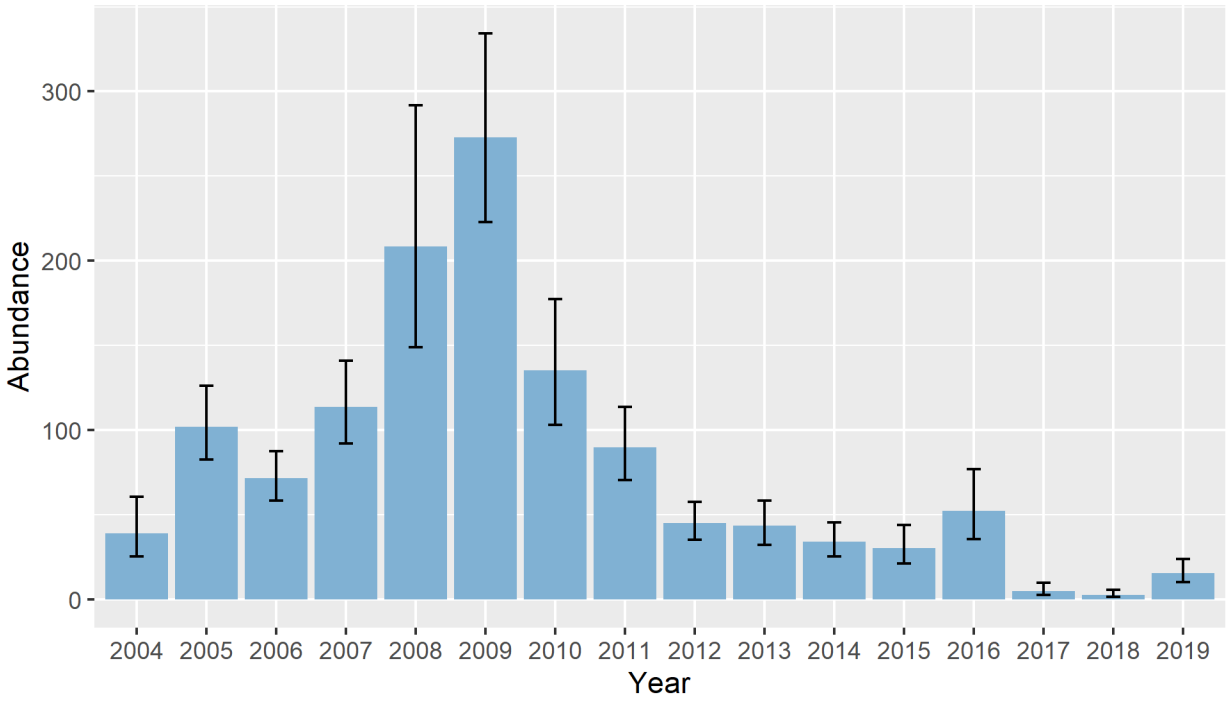


Figure 33. Estimated total abundance for the "South of Hatteras Island" region in February from 2004-2019, typically the month of peak abundance in the region. Whiskers show approximate asymptotic 95% confidence intervals.

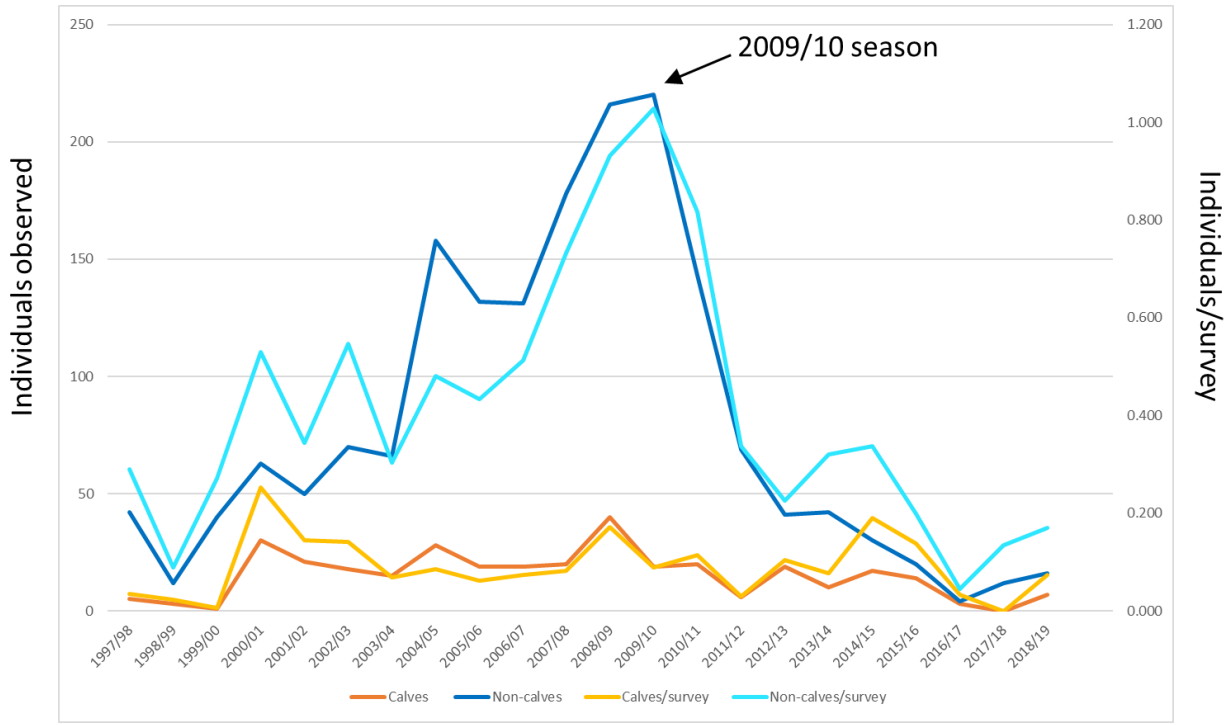


Figure 34. Total counts and sighting rates of unique, photographically-identified individuals south of Virginia, by survey season (defined as December of the first year listed through March of the following year). Data courtesy of T. Gowan (FWRI), compiled from North Atlantic Right Whale Consortium photo ID database (as of November 2016) and Southeast Implementation Team presentations.

5.3.1.2. Hatteras Island to Nantucket Shoals

In this region, the substantial differences in survey effort between the focal eras, particularly the lack of any effort in the 2003-2009 era over large portions of the continental shelf during the months of September-May, presented a major impediment to modeling this region. This problem was compounded by a substantial rise in the species' use of this area in the 2010-2018 era (Quintana-Rizzo et al. 2021). This change was much harder to model than the decline in the calving area ("South of Hatteras Island" discussed in the previous section), because the calving area benefitted from consistent and dense replication of survey effort across the 2003-2018 period, facilitating the use of contemporaneous dynamic covariates as well as the use of Year as a temporal factor to account for unexplained interannual differences. This proved infeasible for the mid-Atlantic "Hatteras Island to Nantucket Shoals" region, at least with the data and covariates available at the time. The models we tried that used contemporaneous covariates all overestimated right whale abundance (>800 whales). As a fallback, we switched to climatological covariates and used the Era as a binary factor to allow for an overall inter-era difference in density. Roberts et al. (2020), Section 5.3.2.2, discusses these issues in more detail.

An important consequence of falling back to climatological dynamic covariates is that predictions varied only from month to month and not interannually. Because of this, the SEs and CVs for the mid-Atlantic could only account for model uncertainty, because interannual variability was unmodeled and unknown. Evidence suggests that interannual variability was high in the mid-Atlantic, and thus these SEs and CVs based only on model uncertainty represent underestimates of the "true" variability that occurred. For example, sighting rates for the offshore wind energy development areas off Rhode Island and Massachusetts increased substantially from 2012-2019 (Quintana-Rizzo et al. 2021), strongly suggesting high interannual variability in density.

Let us consider the 2010-2018 era first, as the 2003-2009 era had a complication that is better addressed second. Predicted densities for the region were highest during the months of December-May, after whales returned from summer feeding grounds. During these months, in the 2010-2018 era, total abundance was estimated to exceed 25

whales in every month (Figure 36) and CVs of total abundance ranged from 0.10-0.16 (shown in text in maps in Figure 13). These relatively low values accounted only for error in the model parameter estimates and would undoubtably be higher if interannual variability was able to be accounted for. Similarly, maps of per-cell CV were less than 0.5 for most of the study area, with a substantial area of each month less than 0.25. This resulted in a striking difference in the CV maps between regions, in which the mid-Atlantic appears mostly blue to yellow, and the regions around it show much more yellow and red.

In the 2003-2009 era, total abundances were much lower (Figure 35) and CVs were much higher (shown in text in maps in Figure 12). The CV for total abundance for December was 0.36, while for January-May it was 1.57-1.58. This markedly higher uncertainty in January-May resulted from our decision to consider all of 2010 to be part of the 2010-2018 era during model fitting, but then to include the January 2010 - May 2010 predictions in the "2003-2009" era, which ran from June 2003 - May 2010. Our justification for considering 2010 to be part of the 2010-2018 era during model fitting was driven by the first appearance of large aggregations of right whales feeding off Rhode Island Sound in spring of 2010, which we took to be an indicator that the pattern of distribution had changed. In any case, as a result, the "2003-2009" era included the first five months of 2010, which featured markedly higher densities in the mid-Atlantic than in prior years (which were all the same as each other because climatological covariates were used). This resulted in the 2003-2009 summarization including some "inter-era" variability, which drove up CVs.

There is widespread acceptance in the right whale community that right whale density between Hatteras Island and Nantucket Shoals was higher in 2010-2018 than in 2003-2009. This belief comes not just from visual surveys but also passive acoustic monitoring (Davis et al. 2017). Now that we have obtained a concurring result in this version of the density model, in future versions we may opt to exclude modeling 2003-2009 for this region as an objective, and instead focus only on 2010 and later, when temporal replication in effort was higher, which might allow us to try contemporaneous covariates again. If that was successful, we could account for interannual variability, similar to what was done for the calving grounds. Meanwhile, we advise model users to exercise particular caution with this region, especially in waters south and east of Rhode Island, Martha's Vineyard, Nantucket, and Nantucket Shoals. These areas have seen rapidly rising use by right whales over the 2010-2018 era (Quintana-Rizzo et al. 2021). Although we estimate density to be highest in December-May, sightings have been reported in June-November in increasing numbers in recent years. We anticipate that our future models will predict higher density in June-November after we incorporate additional data collected in the region.

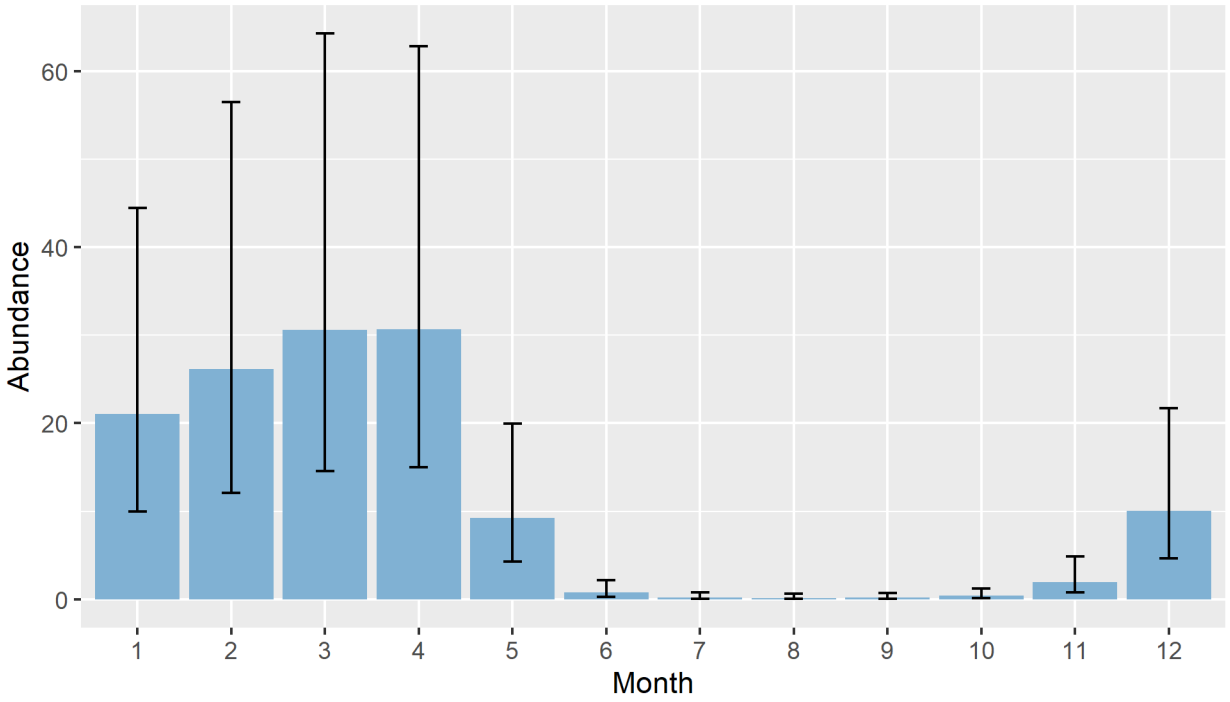


Figure 35. Monthly mean total abundance for the "Hatteras Island to Nantucket Shoals" region, with approximate asymptotic 95% confidence intervals, for the 2003-2009 era. The confidence intervals here do not account for interannual variability (see text).

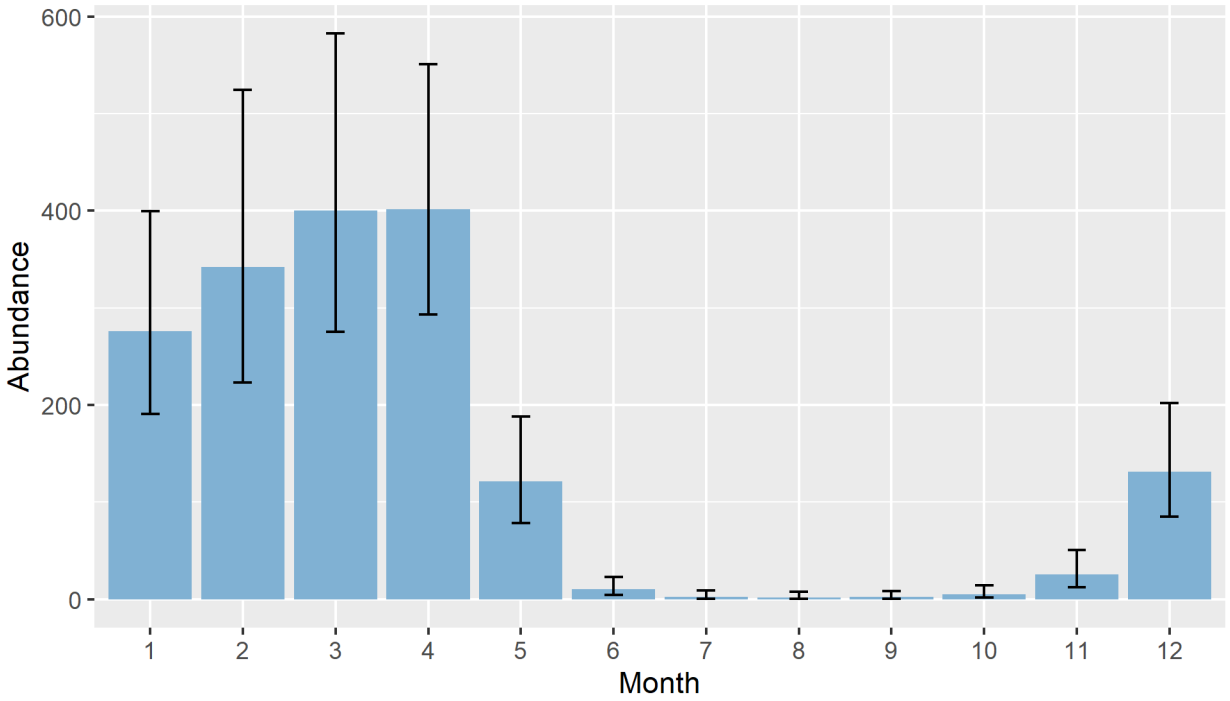


Figure 36. Monthly mean total abundance for the "Hatteras Island to Nantucket Shoals" region, with approximate asymptotic 95% confidence intervals, for the 2010-2018 era. The confidence intervals here do not account for interannual variability (see text). Note the y-axis has a higher limit than the previous figure.

5.3.1.3. North of Nantucket Shoals

Historically, in this region, right whales were believed to migrate around the Gulf of Maine and nearby basins in an annual counterclockwise cycle. This cycle seemed to hold for the 2003-2009 era but was disrupted somewhat in the 2010-2018 era as increasing numbers of whales travelled in summer to the Gulf of St. Lawrence to feed instead of to traditional closer locations such as Roseway Basin and the lower Bay of Fundy. It proved difficult to model these movements strictly from the covariates readily available for the version 9 model, particularly given a lack of homogenous survey effort across the 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. Please see Roberts et al. (2020), Section 5.3.2.3, for details. Here, we discuss the uncertainty estimates for each season in turn.

5.3.1.3.1 Spring (March-July)

In this region and season, the degree of heterogeneity and temporal replication in survey effort fell between that of the "South of Hatteras Island" region, where heterogeneity was low and replication high, and the "Hatteras Island to Nantucket Shoals" region, where the opposite applied. North of Nantucket Shoals, the NARWSS program provided replicated coverage of areas known to be important to right whales while the AMAPPS program contributed occasional broad-scale surveys providing full coverage of the area. This allowed fairly complex models to be fitted.

As with the "South of Hatteras Island" region, the model selection procedure evaluated models both with contemporaneous dynamic covariates and with climatological dynamic covariates. Models with climatological covariates scored better, explaining more deviance and achieving better goodness of fit statistics, and we ultimately selected the top-scoring climatological model. Similar to the "Hatteras Island to Nantucket Shoals" model, this model utilized Era as a binary factor to account for substantial differences in density between the eras, with densities of the 2003-2009 era (corresponding to the March-July period of the years 2004-2010) predicted to be roughly three times higher than densities of the 2010-2018 era (corresponding to March-July of 2011-2019).

As with the "Hatteras Island to Nantucket Shoals" model, the use of climatological covariates meant that uncertainty estimates were based only on model uncertainty, and interannual variability was not modeled or accounted for (even though inter-era variability was addressed by the binary factor). While it is possible that density patterns were consistent enough to posit a climatological-scale species-environment relationship (e.g., based on migration behaviors that were learned or evolved from long-term environmental patterns) we believe it more likely that densities actually varied interannually but that we lacked the survey data or covariates to capture this variability. At the southeast calving grounds, where survey data was dense and highly replicated, density shifts were well-correlated with water temperature, while bulk interannual variability could be captured with Year as a categorical covariate. In the northeast, density shifts were likely related to prey availability—a covariate we didn't have—while spatiotemporal replication in effort was insufficient to allow the use of Year as a categorical covariate.

In any case, the use of climatological covariates resulted in CVs that are likely to be underestimates. For the 2003-2009 era, CVs of total abundance ranged from 0.10-0.31 (Figure 12), while for 2010-2018 they ranged from 0.08-0.26 (Figure 13). For the combined 2003-2018 era, they were higher, 0.31-0.41 (Figure 14), reflecting the addition of inter-era variability captured by the binary factor covariate. (A similar effect occurred in the "Hatteras Island to Nantucket Shoals" region for the 2003-2018 era, although we did not discuss it in the previous section.)

Maps of CVs showed similarly low values (0.10-0.50) in areas where appreciable density was predicted, with a notable exception being in March along the northeastern tip of Georges Bank, in the Northeast Channel, where moderate density was predicted and CVs were greater than 1. This gives reason to doubt the model's prediction here in this month; future models would benefit from additional survey data here and other parts of Canadian waters in late winter and early spring. In areas where density was predicted to be very low, CVs also reached 1 or higher, but these areas showed low standard error and thus the high CVs are not cause for concern.

5.3.1.3.2 Summer (August-September)

In this region and season, it is generally believed that the bulk of the right whale population feeds in Canadian waters. In the 2003-2009 era, the important feeding grounds included Grand Manan Basin[2], Roseway Basin, and other zooplankton-aggregating basins of the Scotian Shelf (Davies et al. 2015). In the 2010-2018 era, use of these traditional feeding grounds decreased substantially, and a large portion of the population shifted to feeding in the Gulf of St. Lawrence instead (Davies et al. 2019; Meyer-Gutbrod et al. 2021). This two-month season was very difficult to model, with only 41 sightings available, and only two of them occurring in the 2010-2018 era. Please see Roberts et al. (2020), Section 5.3.2.3.2, for discussion of the difficulties and our approach to addressing them.

Despite the sparse effort and few sightings available, we tested models with both contemporaneous and climatological dynamic covariates, as we described in detail Roberts et al. (2020). Models with contemporaneous covariates scored slightly better on goodness of fit tests but yielded excessive predictions, averaging 1000-2000 whales, owing to runaway extrapolations during years with little or no survey data in Canadian waters. We selected the best-fitting model that used climatological covariates instead.

As in the Spring season, the use of climatological covariates meant that uncertainty estimates could only account for model uncertainty, while interannual variability remained unmodeled and unaccounted-for. Standard errors and CVs were nevertheless high, likely due to the relatively low amount of survey data available and the ineffectiveness of the covariates for precisely delineating prime feeding habitat. In the 2003-2009 period, where density was clearly overestimated (see Roberts et al. (2020) for discussion), abundance was higher but CVs were lower, totaling 635 whales (CV=0.61) in August and 796 (CV=0.42) in September, than for the 2010-2018 period, which totaled 48 (CV=0.91) and 75 (CV=0.98). Maps of CVs showed similar high uncertainty, with CV > 1 nearly everywhere except in 2003-2009 in some patches of the study area, where it dipped into the 0.7-0.8 range.

These results reiterate the difficulty of modeling this region during these months. Although the situation might be improved with better covariates, e.g., of zooplankton density, the main problem is that the survey effort and sightings were too sparse to decisively address all of the complications, particularly the apparent rarity of right whales in this region in recent years relative to the 2003-2009 period. The data available to us that could be utilized for density surface modeling methodology included surveys from NEFSC—NARWSS, AMAPPS, and some of the pre-AMAPPS surveys targeting harbor porpoises—as well as a few aerial surveys conducted by New England Aquarium as part of their NLPSC program. We are aware of some additional data from U.S.-based organizations that might be utilizable with sufficient preparatory work, such as the NEFSC shipboard "large whale surveys", which were focused on small spatial subregions but conducted with protocols compatible with density surface modeling. However, the best progress would be achieved by collaborating with Canadian researchers who specialize in these regions and have conducted broader-scale surveys of Canadian waters similar to the NARWSS and AMAPPS programs. With support from NOAA, we aim to convene discussions with relevant researchers from Canada and the U.S. within the next year.

5.3.1.3.3 Winter (October-February)

Up until 2012, large numbers of right whales were often sighted in the western and central Gulf of Maine, usually in the months of fall and winter, an area hypothesized to be a right whale mating ground (Cole et al. 2013). Starting in 2012, sightings greatly decreased in this area (Figure 37). Although survey effort was patchy, there was better temporal replication than in summer, albeit not enough to try to utilize Year as a categorical covariate, as was done in the "South of Cape Hatteras" region, but we did include Era as a binary factor.

Models with contemporaneous covariates strongly outperformed those with climatological covariates, explaining 5% more deviance. We selected the best-scoring contemporaneous model, which retained seven terms but dropped the Era covariate. Remarkably, this model reproduced the apparent drop-off in density that began in the 2012/13 season (Figure 38), despite having neither of the explicitly temporal covariates (Year or Era). We believe this result was driven by the inverse relationship fitted to SST coupled with a rising trend in SST over the 2003-2018 era, which has been implicated in other changes to right whale distributions (Record et al. 2019; Meyer-Gutbrod et al. 2021).

[2] Cetacean ecologists and Canadian species managers consider Grand Manan Basin to be the deep basin east of Grand Manan island at the mouth of the Bay of Fundy (Davies et al. 2015), but geologists may refer to this area as Fundy Basin, and consider Grand Manan Basin to be the smaller, shallower basin to the west of Grand Manan island (McHone 2011). Whenever we mention "Grand Manan Basin", we mean the place referenced by cetacean ecologists.

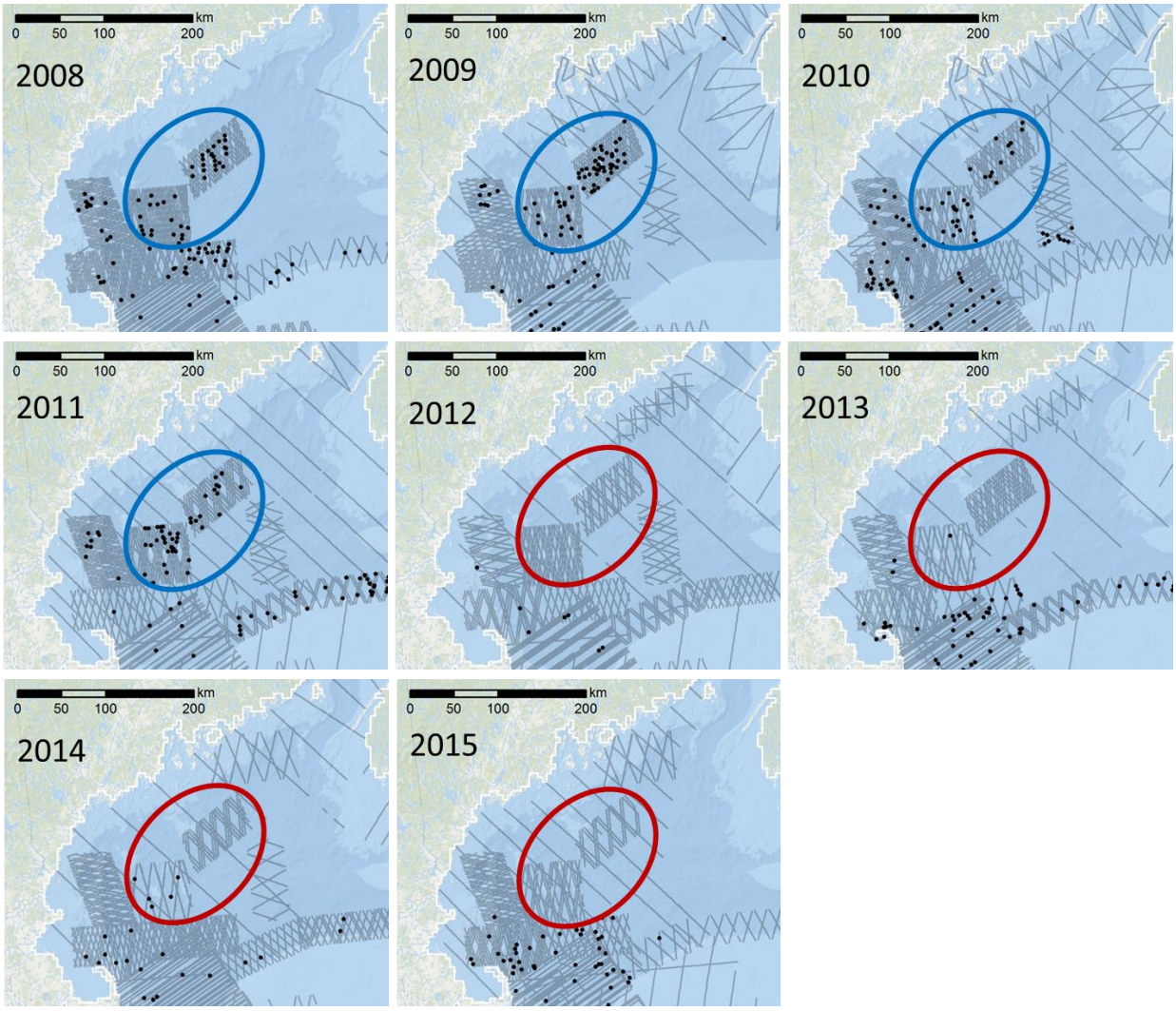


Figure 37. Surveys conducted in the Gulf of Maine between October of the year listed and September of the following year, with right whale sightings overlaid (black dots). Through 2011, right whales were sighted in large numbers in the central and western Gulf of Maine (blue ellipses), mainly in fall and winter. Starting in 2012, sightings in this area decreased substantially (red ellipses), but so did survey effort.

When total abundance was summarized by season (Figure 38) from 2003-2018, 95% confidence limits were wide because within-season variability was high, particularly in 2003/04 through 2011/12, when right whales aggregated in large numbers in the central and western Gulf of Maine early in the season, then migrated out of the modeled region (to the calving grounds or Cape Cod Bay) late in the season. Abundance was lower starting in 2012/13, when, apparently, fewer whales inhabited these areas and instead went to Cape Cod Bay (Ganley et al. 2019) or southern New England (Quintana-Rizzo et al. 2021).

Maps of CV, which accounted for both model and environmental uncertainty, ranged from about 0.4-1.0 in the central and western Gulf of Maine for the 2003-2009 era where sightings were common prior to 2012/13, and CVs were higher in the 2010-2018 era, reflecting the apparent reduction in use that occurred partway into the era. Outside of these regions of the Gulf of Maine, CVs generally exceeded 1 but predicted density and standard errors was very low, indicating no cause for concern, with one important exception: the area from Stellwagen Bank, southeast along Cape Cod, to Great South Channel. Moderate density was predicted here in every month of the season (October-February). A few sightings were reported but effort was sparse, so the outcome of moderate density with high uncertainty is appropriate. Uncertainty could be reduced by conducting additional surveying.

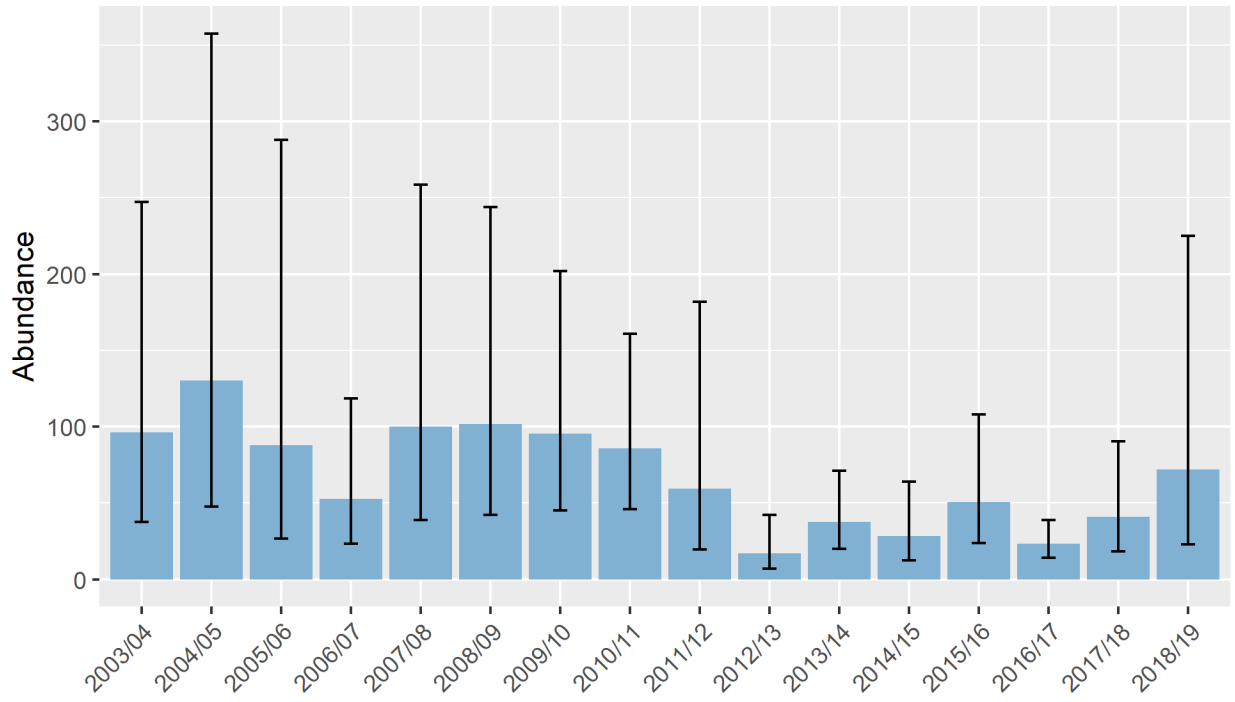


Figure 38. Mean total abundance during "Winter" season (October-February), with approximate asymptotic 95% confidence intervals, which here account for both error in model parameter estimates and within-season (month-to-month) variability.

Although the abundance predictions (Figure 38) and the surveys used to fit the model (Figure 37) suggest that right whale density was lower in 2012/13 and most subsequent winters than in years prior, these results should not be interpreted as indicating that right whales now occur only rarely in this area, or will occur only rarely in the future. CVs indicate high uncertainty, reflective of high interannual variability. For example, predicted abundance for the last modeled winter, 2018/19, when no detections were reported but surveying was very sparse (data not shown here), was actually higher than for 2011/12 (Figure 38), the last year in which large number of sightings were reported but also a year with robust surveying (Figure 37).

Acoustic monitoring of the region for the most recent years available, 2018-2021, found that whales were present in the central Gulf of Maine during this season (Figure 39), suggesting that the prediction of high density here relative to other parts of the Gulf of Maine is reliable. By contrast, no acoustic detections were reported near the patch of high density predicted in the western Gulf of Maine, near 43 °N, 70 °W, although acoustic coverage was sparser here. Finally, a number of acoustic detections occurred in the southwest corner of the study area, in the vicinity of the Stellwagen Bank National Marine Sanctuary, where acoustic coverage was denser but negligible density was predicted. These latter results provide further evidence that interannual variability in right whale distributions has been high in this region for approximately the last decade. This variability could be better characterized, and uncertainty in density predictions reduced, by increased surveying within this region, both visually and acoustically, particularly in areas that have received little coverage over the past decade.

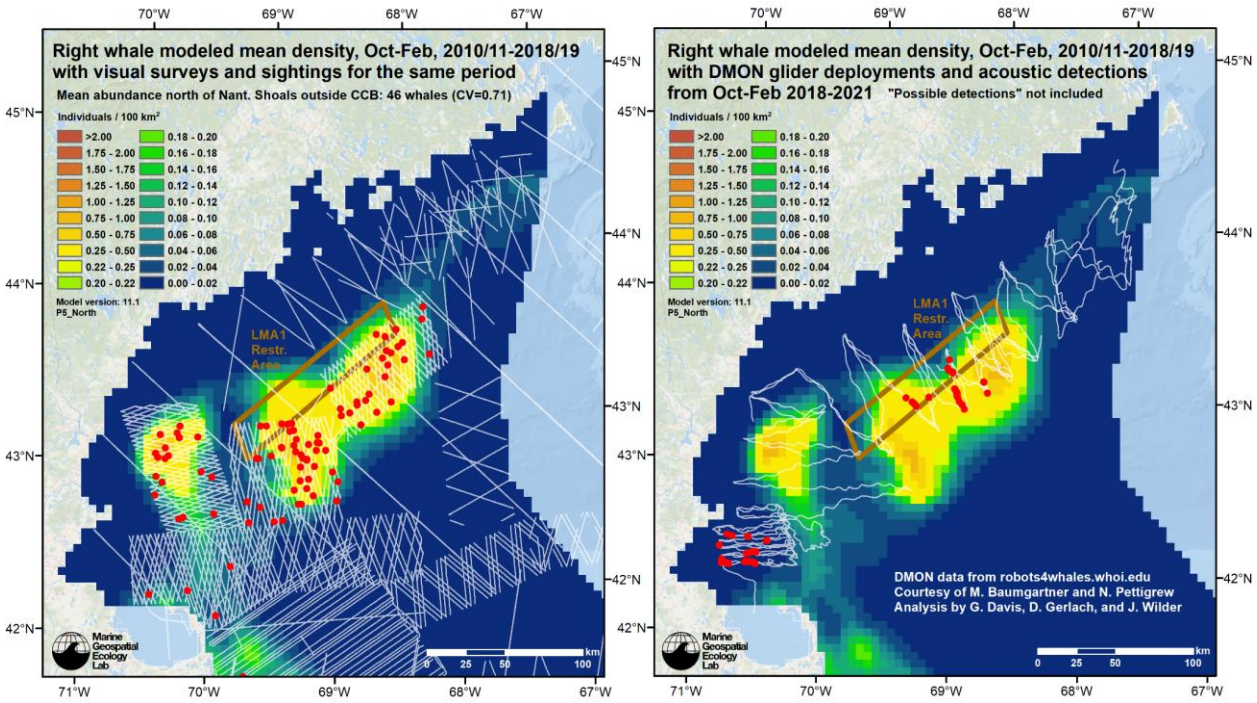


Figure 39. Left: "Winter" season (October-February) mean density prediction for the 2010-2018 era overlaid with visual survey transects and sightings from the same era that were included in the model. Right: The same prediction overlaid with tracklines of DMON-fitted Slocum gliders (Baumgartner et al. 2020), deployed during the same months but during the most recent years available, 2018-2021, and the resulting acoustic detections of right whales (acoustic activity classified as "possible detections" of right whales are not included). The acoustic data cannot be directly incorporated into density models (see Marques et al. (2013) for discussion) but can indicate where whales are present. For the benefit of readers interested in recent changes by NOAA to trap and pot fishery regulations in the region, we included the "Lobster Management Area 1 (LMA1) Restricted Area" on the maps.

5.3.2. Improving the simulation-based uncertainty estimates

In this report, we presented the simulation-based results for a univariate toy model in the northeast (Section 5.2.2.1) and for the full model for the "South of Hatteras Island" region in "Winter" (Section 5.2.2.2). In both of these models, Metropolis-Hastings sampling alone was not sufficient to yield a collection of simulations free of implausibly extreme results. In saying "implausibly extreme" we are not speaking of merely "very high" estimates of density, such as those that would be beyond the 95% or even the 99.9% quantile. We are referring to estimates so extreme that if a single such simulation were included in the collection, the mean of the entire collection would increase by one or more orders of magnitude. We addressed this problem by implementing an arbitrary but systematic trimming procedure to remove simulations that yielded extreme total abundances. This worked ok for the toy model but was not entirely satisfactory for the "South of Hatteras Island" model, as it still yielded patches of moderate to high standard error in offshore regions where point estimates were very, very low (in accordance with the view that right whales do not inhabit these offshore regions). The simulation results would thus be problematic for model users who did not restrict their area of interest to the on-shelf regions, where the results were more reasonable.

When we reached that outcome, we had no time left to continue the investigation under the current project cycle and made the decision to not release the uncertainty simulations. The questions thus remain: why does this problem happen and how can it be fixed? Additional investigation is needed, but our best current guess is that this problem occurs when models include smoothed terms for which functional plots show wide confidence intervals at one or both ends of sampled ranges and a large range on the y-axis, especially when the upper interval reaches above 0 on the y-axis. When simulated, these terms can yield functional relationships that reach high values on the y-axis, similar to what occurs in a runaway extrapolation. Because these models use a log link function, fitted relationships must be

exponentiated to transform them back to the response scale; as a consequence, relationships that reach even moderately above zero can yield high abundances when covariate values are in the right ranges.

If this diagnosis is correct, the only surefire approach we currently know of to avoid the problem is to examine fitted functional relationships as part of the model selection procedure and avoid or alter models that exhibit relationships showing this problem. Terms with such relationships could be dropped from the model, or, alternatively, the relationship could be "Winsorized" (Dixon 1960), a procedure also known as "clamping" or "clipping", in which extreme values are rounded closer to the mean. Under this scheme, a maximum (or minimum) limit is placed on the value a covariate is allowed to take, and everything beyond it is rounded down (or up) to the limit. Then, when the model is fitted, the resulting relationship is prevented from extending far enough for the upper confidence interval to reach extreme values (Figure 40). Finally, when the model is simulated, the functional relationships in the simulations are also prevented from reaching such extremes.

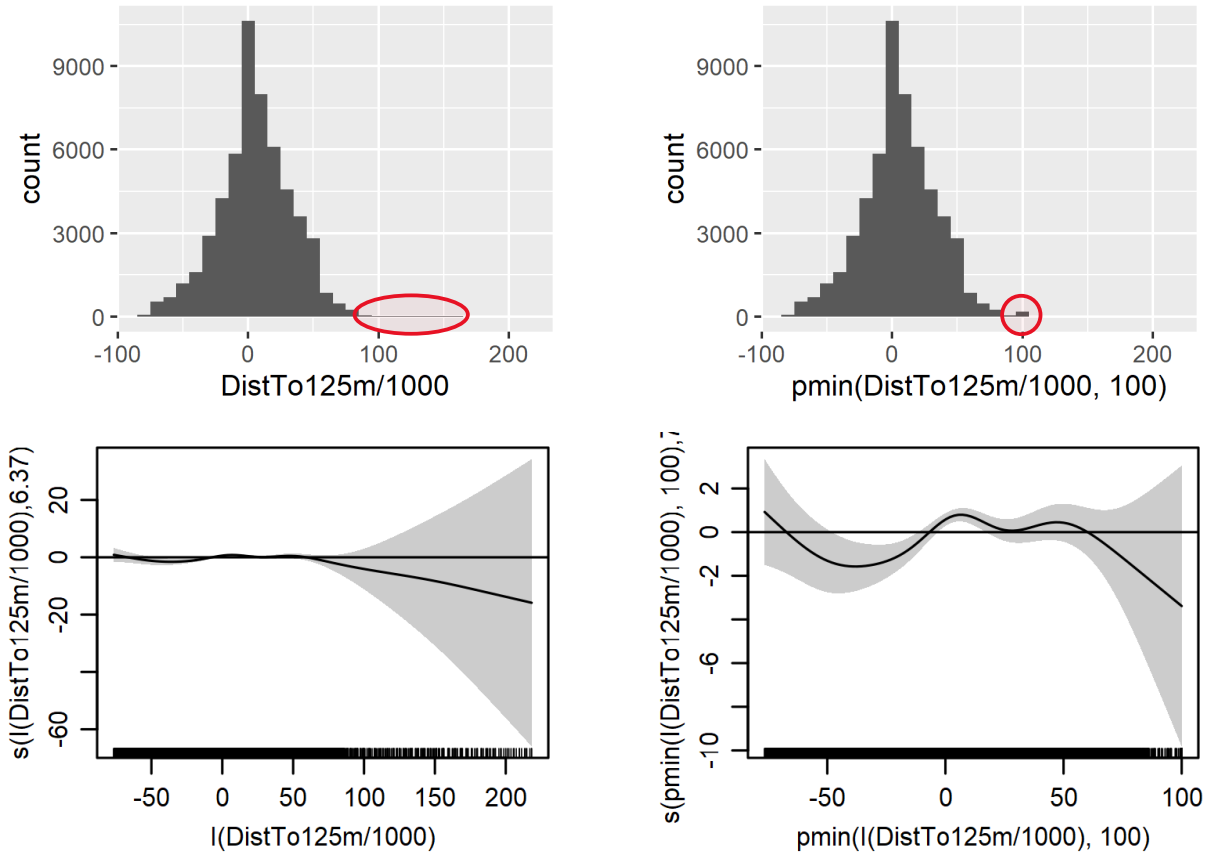


Figure 40. Example of "Winsorizing" (a.k.a. "clamping" or "clipping"), using a subset of the right whale data. The x-axis unit for all plots is kilometers. In this example, the original DistTo125m covariate (upper left) yielded a fitted relationship (lower left) that exhibited widely-expanding confidence intervals starting at about 70 km, owing to the very sparse data available beyond this distance (red ellipse on left). When the data were Winsorized by rounding all values greater than 100 km down to 100 km (upper right), the fitted relationship was essentially the same, and showed a decreasing response beyond about 50 km, but the upper confidence interval was curtailed to a value not likely to yield extreme densities. If this model were predicted across the study area, all grid cells with DistTo125m > 100 km would be rounded down to 100 km and receive the same additive contribution to density from this term. (To implement this technique, use the R `pmin` or `pmax` function to round down or up, respectively, directly in the GAM model formula. The rounding will then be automatically applied during model fitting and prediction.)

As a point of clarification, we note that Winsorization is different than classifying extreme values as out of range and excluding them from the model during fitting and prediction. That approach is more conservative, and is best used when the modeler wants to *prevent* predictions at extreme covariate values, under the precept that the functional relationship between the covariate and response cannot be determined reliably in poorly sampled ranges of the covariate. Essentially, the modeler treats poorly sampled ranges as dangerous extrapolations. By contrast, in the Winsorization approach, the modeler asserts confidence, often based on supplemental evidence or expert opinion that the relationship at the end of well-sampled range of the covariate will hold in the poorly sampled tail of the distribution.

We have long used this technique to control runaway extrapolations; for example, we applied it to fix the Massachusetts Bay extrapolation problem that prompted the production of the version 10 model (see Section 3.2 above). But we have not deeply explored its use as a means to mitigate unreasonable uncertainty simulations. An important objective of the uncertainty analysis done here is that we not refit any of the models. We wanted the version 9 density predictions to remain unchanged and assess their uncertainty as-is. (We made exceptions for the corrections and adjustments requested by users that resulted in versions 10 and 11.) Version 12 will be a full refit of the model with additional data. We will continue to investigate this problem in version 12 and potentially apply Winsorization as a means to address it, if no better solution is found.

5.3.3. Sources of uncertainty not accounted for

In this project, we estimated the combined uncertainty resulting from the statistical error in model parameter estimates (*model uncertainty*) and the variance in predictions induced by the dynamic covariates (*environmental uncertainty*). These are two of the most important sources of uncertainty but density surface modelers acknowledge others and are actively progressing on methods to characterize and account for them.

One source of uncertainty that is usually accounted for in simpler density surface models than those presented here is the variance in the probability of detection estimated by the detection function. When a single detection function is used as the first stage of a density surface model, its variance may be combined with that of the GAM that is the second stage using the "delta method" (Ver Hoef 2012; Miller et al. 2013). However, when multiple detection functions are used, each applying to a different subset of segments, as occurs with the density surface models that make up our right whale model, the delta method does not work. Recently, collaborators in the DenMod working group developed a solution to this problem (Bravington et al. 2021) that allows the variance from the detection functions to be propagated into the GAM as an additional term that affects the model's variance but not the mean estimate. We were aware of this research as it was being developed but lacked the time to implement it once it was ready, having prioritized environmental uncertainty higher in our list of objectives. We hope to address it in our next model update (version 12).

Other sources of uncertainty that have been discussed within the DenMod working group include the uncertainty in availability bias corrections (the g_0 parameter), species identifications, group size estimates, and covariate estimates. As methods are developed to address these, we may be able to incorporate them via the same variance propagation technique developed by Bravington et al. (2021) for detection functions. This is a longer term research objective that is not likely to be addressed during the Navy's AFTT Phase IV modeling cycle.

Finally, we note that the uncertainties quantified here do not account for *model selection uncertainty*. That is, the uncertainty estimates presume that the models we selected from the candidates we considered for each region would remain the "best" models if the surveys had been repeated and new data collected. We first explored model selection uncertainty in AFTT-wide models for the Navy's Phase III modeling cycle, in which we assessed abundance predictions for the five top-scoring models as a criterion for evaluating the top model used for final predictions (see Mannocci et al. (2017) species specific reports). We also addressed it in our 2018 update to the "east coast" Clymene dolphin model, in which we used AIC-weighted model averaging (Wagenmakers & Farrell 2004) to combine predictions for several models. At present, we do not plan to implement these or other approaches for the next update to the right whale model, but the topic of model selection uncertainty remains of interest to us and we may further address it in the future.

6. 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. A special thanks goes to David L. Miller, who led the development of the statistical methods we used to characterize uncertainty. We thank Brigid McKenna, Laura Ganley, and Stormy Mayo of CCS for collaborating with us to produce new December CCB density estimates on short notice at NOAA's request. We are also very grateful to the many other scientists who collaborated with us in developing the other parts of the right whale model. 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., 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, Mark Baumgartner, Neal Pettigrew, and their collaborators for allowing us to compare our density model predictions to their acoustic detections. Thanks to 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. Fleet Forces Command and NOAA Fisheries for funding the development of this model, and Navy Facilities Engineering Command Atlantic for managing this Cooperative Agreement.

7. References

Baumgartner MF, Bonnell J, Corkeron PJ, Van Parijs SM, Hotchkin C, Hodges BA, Bort Thornton J, Mensi BL, Bruner SM. 2020. Slocum Gliders Provide Accurate Near Real-Time Estimates of Baleen Whale Presence From Human-Reviewed Passive Acoustic Detection Information. Frontiers in Marine Science **7**:100.

Becker EA, Forney KA, Miller DL, Fiedler PC, Barlow J, Moore JE. 2020. Habitat-based density estimates for cetaceans in the California Current Ecosystem based on 1991-2018 survey data. NMFS-SWFSC-638, NOAA Technical Memorandum. U.S. Department of Commerce, National Oceanic and Atmospheric Administration, National Marine Fisheries Service, Southwest Fisheries Science Center, La Jolla, CA. Available from https://swfsc-publications.fisheries.noaa.gov/publications/CR/2020/2020Becker1.pdf.

Bravington MV, Miller DL, Hedley SL. 2021. Variance Propagation for Density Surface Models. Journal of Agricultural, Biological and Environmental StatisticsDOI: 10.1007/s13253-021-00438-2. Available from http://link.springer.com/10.1007/s13253-021-00438-2 (accessed March 9, 2021).

Cañadas A, Roberts JJ, Schick RS, Halpin PN. 2020. Updated Marine Species Density Models for the Arctic Study Area. Duke University Marine Geospatial Ecology Lab, Durham, NC.

Cole T, Hamilton P, Henry A, Duley P, Pace R, White B, Frasier T. 2013. Evidence of a North Atlantic right whale Eubalaena glacialis mating ground. Endangered Species Research **21**:55–64.

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.

Davies KTA, Vanderlaan ASM, Smedbol RK, Taggart CT. 2015. Oceanographic connectivity between right whale critical habitats in Canada and its influence on whale abundance indices during 1987–2009. Journal of Marine Systems **150**:80–90.

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).

Dixon WJ. 1960. Simplified Estimation from Censored Normal Samples. The Annals of Mathematical Statistics **31**:385–391.

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.

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.

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

Keller C, Garrison L, Baumstark R, Ward-Geiger L, Hines E. 2012. Application of a habitat model to define calving habitat of the North Atlantic right whale in the southeastern United States. Endangered Species Research **18**:73–87.

Keller CA, Ward-Geiger LI, Brooks WB, Slay CK, Taylor CR, Zoodsma BJ. 2006. North Atlantic Right Whale Distribution in Relation to Sea-Surface Temperature in the Southeastern United States Calving Grounds. Marine Mammal Science **22**:426–445.

Mannocci L, Roberts JJ, Miller DL, Halpin PN. 2017. 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.

McHone JG. 2011. Triassic Basin Stratigraphy at Grand Manan, New Brunswick, Canada. Atlantic Geology **47**:125–137.

Meyer-Gutbrod E, Greene C, Davies K, Johns D. 2021. Ocean Regime Shift is Driving Collapse of the North Atlantic Right Whale Population. Oceanography **34**:22–31.

Miller DL, Becker EA, Forney KA, Roberts JJ, Cañadas A, Schick RS. in prep. Estimating uncertainty in density surface models.

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.

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 EvolutionDOI: 10.1002/ece3.3406. Available from http://doi.wiley.com/10.1002/ece3.3406.

Pedersen EJ, Miller DL, Simpson GL, Ross N. 2019. Hierarchical generalized additive models in ecology: an introduction with mgcv. PeerJ **7**:e6876.

Quintana-Rizzo E et al. 2021. Residency, demographics, and movement patterns of North Atlantic right whales Eubalaena glacialis in an offshore wind energy development area in southern New England, USA. Endangered Species Research **45**:251–268.

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 May 3, 2020).

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. 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, Schick RS, Halpin PN. 2020. Final Project Report: Marine Species Density Data Gap Assessments and Update for the AFTT Study Area, 2018-2020 (Opt. Year 3), Document Version 1.4. Page 142. Duke University Marine Geospatial Ecology Lab, Durham, NC.

Ver Hoef JM. 2012. Who Invented the Delta Method? The American Statistician **66**:124–127.

Wagenmakers E-J, Farrell S. 2004. AIC model selection using Akaike weights. Psychonomic Bulletin & Review **11**:192–196.

Wood S. 2017. Generalized Additive Models: An Introduction with R, 2nd edition. CRC press, Boca Raton.

Appendix:
Right Whale Abundance Estimates for Cape Cod Bay in December

Jason Roberts (Duke MGEL) and Brigid McKenna, Laura Ganley, and Stormy Mayo (CCS)

Document Version 3 (16 February 2021)

Abstract

Since 2019, the North Atlantic right whale density model (Roberts et al. 2016, 2020) developed by a collaboration led by the Marine Geospatial Ecology Laboratory (MGEL) at Duke University has based monthly densities for Cape Cod Bay (CCB) on abundance estimates produced by the Center for Coastal Studies (CCS). CCS derived those estimates from aerial surveys flown there every winter and spring for more than two decades. That analysis, by Ganley et al. (2019), provided abundance estimates for the months of January-May but not for December, a month for which CCS was unable to conduct surveys every year, due to limitations to funding and other resources. But because right whales were occasionally observed in CCB in December, both opportunistically and by CCS on a few surveys conducted that month, the density model initially used the January abundances as proxy estimates for December. This was done for model revisions released in 2019-2020, comprising versions 8, 9, and 10 of the model.

In early 2021, the Massachusetts Division of Marine Fisheries requested that we reexamine the problem of how best to estimate abundance in December for the density model. Here, we document our new approach. First, for the years since 2003 that CCS conducted December surveys, we estimated December abundance from the observed sightings per unit effort (SPUE) by applying Ganley et al.'s (2019) abundance-SPUE relationship for January, the closest month for which this relationship was analyzed. Next, we linearly regressed the abundance of those eight Decembers on the Januaries that immediately followed, but parameter estimates were statistically insignificant, indicating that January's abundance was not a reliable predictor of December's. Finally, we explored three approaches for estimating and applying the mean abundance from the surveyed Decembers to the unsurveyed Decembers.

We conclude that at this time the best approach for estimating abundance for unsurveyed Decembers is to split the time period of interest (December 2003-2020) into two eras at the 2010/11 survey season, and then within each era use the mean abundance of surveyed Decembers for the unsurveyed Decembers. For the December 2003-2010 era, this mean is zero, based on the single survey of December 2004 which did not sight any whales. For the December 2011-2020 era, this mean is 13.85, based on seven years that had December surveys. These results should be interpreted cautiously, given the limited data available and the assumptions required, which we discuss at the end of this paper. Better results could be obtained, and uncertainty reduced, in future years by regularly conducting surveys in December, similar to what has historically been done for January-May.

Analysis

December abundance for years with December surveys

B. McKenna (CCS) summarized the surveys of CCB conducted in the month of December since 2003, the starting year of MGEL's density model. We start by reading these from McKenna's spreadsheet. Note that data for December 2020 have not been finalized by CCS yet and should be considered preliminary.

```
library(openxlsx)
df <- read.xlsx('Dec SPUE 2003-2020.xlsx', 'CCB December Flights', cols=seq(9,15),
```

```
                na.strings='n/a')
df
```

```
##    Year #.surveys sightings  nm       km SPUE.(per.100.nm) SPUE.(per.100.m)
## 1  2003         0         0   0    0.000                NA               NA
## 2  2004         2         0 590 1092.680         0.0000000     0.0000000000
## 3  2005         0         0   0    0.000                NA               NA
## 4  2006         0         0   0    0.000                NA               NA
## 5  2007         0         0   0    0.000                NA               NA
## 6  2008         0         0   0    0.000                NA               NA
## 7  2009         0         0   0    0.000                NA               NA
## 8  2010         0         0   0    0.000                NA               NA
## 9  2011         0         0   0    0.000                NA               NA
## 10 2012         2        15 612 1133.424         2.4509804     0.0013234235
## 11 2013         0         0   0    0.000                NA               NA
## 12 2014         3         7 918 1700.136         0.7625272     0.0004117318
## 13 2015         2         0 566 1048.232         0.0000000     0.0000000000
## 14 2016         2         0 612 1133.424         0.0000000     0.0000000000
## 15 2017         0         0   0    0.000                NA               NA
## 16 2018         3        18 918 1700.136         1.9607843     0.0010587388
## 17 2019         3         0 721 1335.292         0.0000000     0.0000000000
## 18 2020         1         1 306  566.712         0.3267974     0.0001764565
```

Now, for the years that surveys were flown in December, estimate abundance by applying the abundance-SPUE relationship from Ganley et al. (2019), Fig. 7A, for January (green line). Section 3.4 (p. 107) of the paper gives the slope of this line as 32646. Ganley et al. lacked the data needed to estimate this relationship for December; we are assuming January's relationship is a reasonable approximation of December's, on the basis that they are adjacent months. We discuss this further below.

```
janAbundSPUESlope <- 32646    # Abundance unit is whales; SPUE unit is whales/100 m
df['DecAbund'] <- df['SPUE.(per.100.m)'] * janAbundSPUESlope
```

January abundances for all years

We'll look at the December abundances momentarily, but first we want to add January's for comparison. Add a column with abundances from the Januaries that followed each December. For the Decembers of 2003-2016, use the Januaries from 2004-2017 estimated by Ganley et al. (2019), shown in Fig. 6 of that paper and previously provided to J. Roberts by L. Ganley. For the Decembers of 2017-2020, use the Januaries from 2018-2021, which were not included in the 2019 publication, but were subsequently estimated by B. McKenna and provided to J. Roberts. McKenna estimated these by computing the mean counts of whales per survey from those Januaries and then converting to abundance using the abundance-count relationship from Ganley et al. (2019), Fig. 7B, from January (green line). Note that data for January 2021 have not been finalized by CCS yet and should be considered preliminary.

```
df['JanAbund'] <- c(  0,           # Jan 2004 - Ganley et al. (2019)
                      7.868351809, # Jan 2005 - Ganley et al. (2019)
                      0,           # Jan 2006 - Ganley et al. (2019)
                      0,           # Jan 2007 - Ganley et al. (2019)
                      2.275324042, # Jan 2008 - Ganley et al. (2019)
                      2.11504729,  # Jan 2009 - Ganley et al. (2019)
                      2.1912406,   # Jan 2010 - Ganley et al. (2019)
                     66.47725361,  # Jan 2011 - Ganley et al. (2019)
                    121.7637878,   # Jan 2012 - Ganley et al. (2019)
                     87.1106891,   # Jan 2013 - Ganley et al. (2019)
                     48.74889565,  # Jan 2014 - Ganley et al. (2019)
```

```
                    20.90787627,    # Jan 2015 - Ganley et al. (2019)
                     6.60958422,    # Jan 2016 - Ganley et al. (2019)
                    37.03236822,    # Jan 2017 - Ganley et al. (2019)
                    69.70,          # Jan 2018 - B. McKenna (pers. comm.)
                    15.15,          # Jan 2019 - B. McKenna (pers. comm.)
                    66.66,          # Jan 2020 - B. McKenna (pers. comm.)
                   187.86)          # Jan 2021 - B. McKenna (pers. comm.)
```

Show a table of abundances for December and the January that followed. The Decembers with "NA" are those for which no surveying was done.

```
df[,c('Year', 'DecAbund', 'JanAbund')]
```

```
##    Year  DecAbund   JanAbund
## 1  2003        NA   0.000000
## 2  2004  0.000000   7.868352
## 3  2005        NA   0.000000
## 4  2006        NA   0.000000
## 5  2007        NA   2.275324
## 6  2008        NA   2.115047
## 7  2009        NA   2.191241
## 8  2010        NA  66.477254
## 9  2011        NA 121.763788
## 10 2012 43.204485  87.110689
## 11 2013        NA  48.748896
## 12 2014 13.441395  20.907876
## 13 2015  0.000000   6.609584
## 14 2016  0.000000  37.032368
## 15 2017        NA  69.700000
## 16 2018 34.563588  15.150000
## 17 2019  0.000000  66.660000
## 18 2020  5.760598 187.860000
```

Linear model of December abundance from January abundance

As shown above, surveying was done consistently in January but only occasionally in December. Let's explore how well January's abundance predicts the abundance of the December just before it. Although we don't have much data, if there is a consistent pattern of "a high January was always preceded by a high December" and *vice versa*, then we might consider using this relationship to estimate December's abundance from January's when no surveying was done in December.

First, fit a linear model relating December abundance to January abundance for the years for which December surveying was done.

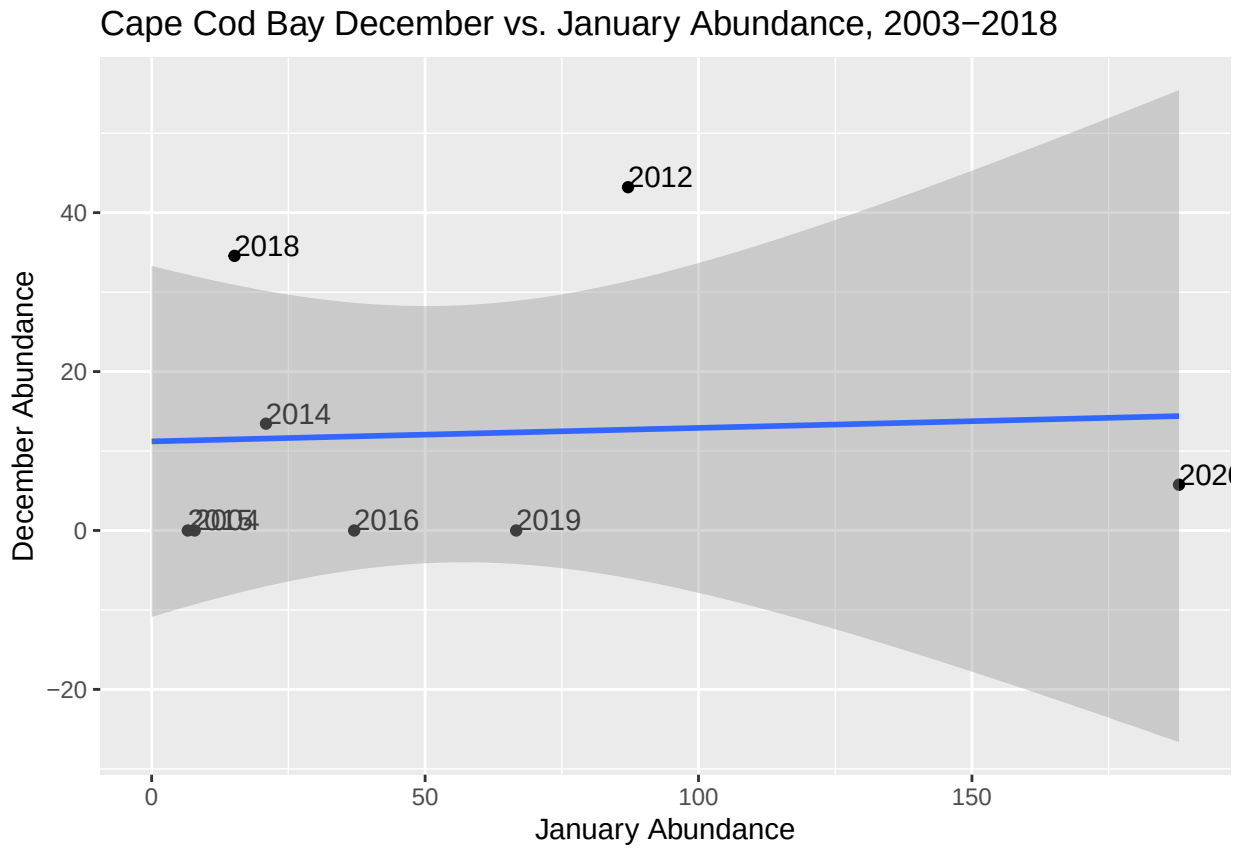
```
model <- lm(DecAbund~JanAbund, data=df, na.action=na.omit)
summary(model)
```

```
##
## Call:
## lm(formula = DecAbund ~ JanAbund, data = df, na.action = na.omit)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -12.342 -11.469  -9.979   7.180  30.516
##
## Coefficients:
```

```
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 11.21223    9.03020   1.242    0.261
## JanAbund     0.01694    0.11484   0.148    0.888
##
## Residual standard error: 18.67 on 6 degrees of freedom
##   (10 observations deleted due to missingness)
## Multiple R-squared:  0.003615,   Adjusted R-squared:  -0.1624
## F-statistic: 0.02177 on 1 and 6 DF,  p-value: 0.8875
```

Make a scatterplot and overlay the fitted model, with confidence intervals:

```
library(ggplot2)
ggplot(df, aes(JanAbund, DecAbund)) + geom_point(na.rm=TRUE) +
  geom_text(aes(label=Year), na.rm=TRUE, hjust=0, vjust=0) +
  geom_smooth(method='lm', formula=y~x, na.rm=TRUE, fullrange=TRUE) +
  labs(x='January Abundance', y='December Abundance',
       title="Cape Cod Bay December vs. January Abundance, 2003-2018")
```



The model suggests there is essentially no linear relationship between January and December abundance with the data provided. Neither of the model's coefficients is considered statistically significant ($p>0.05$) and R^2 is about 0, resulting in an almost perfectly flat relationship. As a consequence, if we predict December from January using this model, we get a narrow range of results across the time series:

```
df$DecPred1 <- predict(model, newdata=df)
summary(df$DecPred1)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   11.21   11.25   11.52   11.91   12.34   14.40
```

Alternate linear model with intercept forced to zero

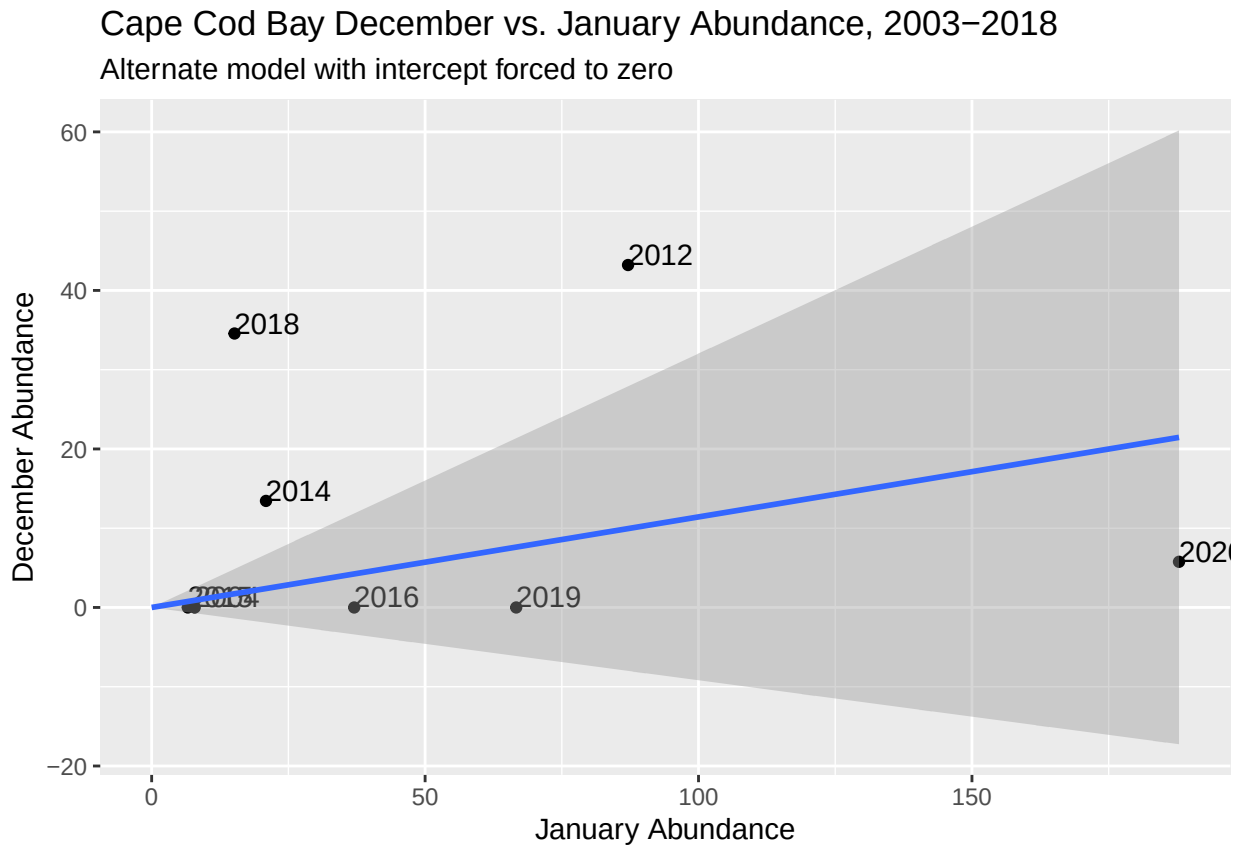
In the linear model above, we estimated both the slope and the intercept, as is usually done. The intercept is the value the model predicts when the independent variable, January abundance, is zero. In this case, the intercept was not zero; it was 11.2. So when January abundance is zero, the model still predicts 11.2 whales in December.

Although we lack comprehensive evidence, the extant data suggest that abundance in December is usually lower than January. In the data we have, this occurred 7 out of 8 years, with 2018 being the exception. If we are willing to assume that December is always lower for the purpose of this model, then we could force the model's intercept to zero and only estimate the slope. Let's see if that model performs better:

```
model2 <- lm(DecAbund~JanAbund + 0, data=df, na.action=na.omit)
summary(model2)
```

```
##
## Call:
## lm(formula = DecAbund ~ JanAbund + 0, data = df, na.action = na.omit)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -15.698  -5.076  -0.827  16.498  33.254
##
## Coefficients:
##          Estimate Std. Error t value Pr(>|t|)
## JanAbund  0.11423    0.08715   1.311    0.231
##
## Residual standard error: 19.38 on 7 degrees of freedom
##   (10 observations deleted due to missingness)
## Multiple R-squared:  0.1971, Adjusted R-squared:  0.08237
## F-statistic: 1.718 on 1 and 7 DF,  p-value: 0.2313
```

```
ggplot(df, aes(JanAbund, DecAbund)) + geom_point(na.rm=TRUE) +
  geom_text(aes(label=Year), na.rm=TRUE, hjust=0, vjust=0) +
  geom_smooth(method='lm', formula=y~x + 0, na.rm=TRUE, fullrange=TRUE) +
  labs(x='January Abundance', y='December Abundance',
       title="Cape Cod Bay December vs. January Abundance, 2003-2018",
       subtitle="Alternate model with intercept forced to zero")
```



In this model, the estimate for the slope is still not statistically significant ($p>0.05$) but the estimate is greater than zero, suggesting December's abundance increased with January's abundance, with an R^2 of 0.197 and a greater range of predictions across the time series than occurred with the first model:

```
df$DecPred2 <- predict(model2, newdata=df)
summary(df$DecPred2)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0000  0.2527  2.0594  4.7117  7.6091 21.4586
```

Third approach: just use the mean

Neither of these models yielded a statistically significant relationship, which is not surprising given the small amount of data we have and the lack of a distinct and reliable pattern in December vs. January abundance. The first model, which yielded the very flat relationship, essentially just returned the mean December abundance averaged across all of the years that December surveying was done. That's the approach we would likely take if we want to use the extant December data to produce an estimate without considering other months. So let's include that approach in our set of candidate approaches:

```
df$DecPred3 <- mean(df$DecAbund, na.rm=TRUE)
summary(df$DecPred3)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   12.12   12.12   12.12   12.12   12.12   12.12
```

However, under that approach, we treat all years across the 2003-2020 time series the same. This may not be realistic, given the changes to whale distribution patterns that occurred around 2010 in many regions. In Cape Cod Bay, a major change to seasonal abundance began in the January-May 2008 survey season, as

shown in Ganley et al. (2019) Fig. 6. So as an alternative to taking a single mean of all surveyed Decembers and applying it to all unsurveyed Decembers, we could split the time series into two eras: the Decembers prior to 2007 and the Decembers of 2007 and later. We use December 2007 as the start of the second era because it immediately preceded the January-May 2008 season.

```
mean1 <- mean(df$DecAbund[df$Year < 2007], na.rm=TRUE)
mean2 <- mean(df$DecAbund[df$Year >= 2007], na.rm=TRUE)
df$DecPred4 <- c(rep(mean1, sum(df$Year < 2007)), rep(mean2, sum(df$Year >= 2007)))
df$DecPred4
```

```
##  [1]  0.00000  0.00000  0.00000  0.00000 13.85287 13.85287 13.85287 13.85287
##  [9] 13.85287 13.85287 13.85287 13.85287 13.85287 13.85287 13.85287 13.85287
## [17] 13.85287 13.85287
```

The result of that was the first four Decembers (2003-2006) were based on the one survey of December 2004, which sighted zero whales, while the remaining Decembers were based on the surveys that occurred December 2007 and later. As a final alternative, we might assume that right whales would not have started regularly occupying Cape Cod Bay in December until 2010, when they first started occupying it in high numbers in the subsequent January (of 2011). This is also shown in Ganley et al. (2019) Fig 6.

```
mean1 <- mean(df$DecAbund[df$Year < 2010], na.rm=TRUE)
mean2 <- mean(df$DecAbund[df$Year >= 2010], na.rm=TRUE)
df$DecPred5 <- c(rep(mean1, sum(df$Year < 2010)), rep(mean2, sum(df$Year >= 2010)))
df$DecPred5
```

```
##  [1]  0.00000  0.00000  0.00000  0.00000  0.00000  0.00000  0.00000 13.85287
##  [9] 13.85287 13.85287 13.85287 13.85287 13.85287 13.85287 13.85287 13.85287
## [17] 13.85287 13.85287
```

Candidate estimates for December abundance

We developed five approaches for estimating abundance for unsurveyed Decembers. Let's display them side by side for consideration. For surveyed Decembers, we use the abundances obtained from surveying. These years' Decembers have the same value across all five approaches. Finally, create a field indicating whether we used surveyed or modeled abundance for that year.

```
df['DecMethod'] <- unname(ifelse(!is.na(df['DecAbund']), 'Surveyed', 'Modeled'))

df[!is.na(df['DecAbund']), c('DecPred1','DecPred2','DecPred3','DecPred4','DecPred5')] <-
  df[!is.na(df['DecAbund']), 'DecAbund']

# We request that R display the results with two digits but it mostly ignores our request

print(df[,c('Year', 'DecMethod', 'DecPred1','DecPred2',
            'DecPred3','DecPred4','DecPred5','JanAbund')], digits=2)
```

```
##    Year DecMethod DecPred1 DecPred2 DecPred3 DecPred4 DecPred5 JanAbund
## 1  2003   Modeled     11.2     0.00     12.1      0.0      0.0      0.0
## 2  2004  Surveyed      0.0     0.00      0.0      0.0      0.0      7.9
## 3  2005   Modeled     11.2     0.00     12.1      0.0      0.0      0.0
## 4  2006   Modeled     11.2     0.00     12.1      0.0      0.0      0.0
## 5  2007   Modeled     11.3     0.26     12.1     13.9      0.0      2.3
## 6  2008   Modeled     11.2     0.24     12.1     13.9      0.0      2.1
## 7  2009   Modeled     11.2     0.25     12.1     13.9      0.0      2.2
## 8  2010   Modeled     12.3     7.59     12.1     13.9     13.9     66.5
## 9  2011   Modeled     13.3    13.91     12.1     13.9     13.9    121.8
## 10 2012  Surveyed     43.2    43.20     43.2     43.2     43.2     87.1
```

```
## 11 2013   Modeled     12.0     5.57     12.1     13.9     13.9     48.7
## 12 2014  Surveyed     13.4    13.44     13.4     13.4     13.4     20.9
## 13 2015  Surveyed      0.0     0.00      0.0      0.0      0.0      6.6
## 14 2016  Surveyed      0.0     0.00      0.0      0.0      0.0     37.0
## 15 2017   Modeled     12.4     7.96     12.1     13.9     13.9     69.7
## 16 2018  Surveyed     34.6    34.56     34.6     34.6     34.6     15.2
## 17 2019  Surveyed      0.0     0.00      0.0      0.0      0.0     66.7
## 18 2020  Surveyed      5.8     5.76      5.8      5.8      5.8    187.9
```

Discussion

Although Cape Cod Bay has been regularly surveyed in the months of January-May since the 1990s, surveying in December has been sparse, with only eight survey seasons in the period 2003-2020 having any surveys in December. In this analysis, we estimated December abundance for all 20 seasons in this period. For seasons with December surveys, we estimated abundance directly from those data. For the remaining seasons, we estimated abundance with five candidate approaches: two linear models and three ways of estimating and assigning simple means to ranges of years

The linear models related December's abundances to the following Januaries', using the eight survey seasons in which both months were surveyed. The first model was allowed to estimate both the slope and the intercept of the relationship. Neither parameter's estimate was statistically significant, and the slope was very close to zero, indicating there was no correlation between the two. For the second model, we fixed the intercept to zero and only estimated the slope, effectively assuming that if January's abundance was zero than December's must also be zero. In this model, the slope was slightly positive but the estimate was not statistically significant.

Together, the linear models suggested there was not a strong correlation between the abundance of January and the preceding December, at least with 8 years of data used to fit the model. Given that, we believe it would be better at this time to base the abundance of unsurveyed Decembers on a simple mean of the surveyed Decembers. We formulated three approaches for doing that: one that treated all seasons in the 2003-2020 period as a single era for the purpose of estimating and applying the mean, and two that split the period into two eras, with a mean for each. One approach placed the transition to the second era at the 2007/08 survey season, on the basis that it was the first season in which Cape Cod Bay exhibited very high monthly abundances (e.g. >100 in some months). The other approach placed the transition at the 2010/11 season, on the basis that January 2011 was the first January with fairly high abundance (>50).

We conclude that at this time the best approach for estimating abundance of unsurveyed Decembers in 2003-2020 the last one; that is, to split the period into two eras at the 2010/11 season and for each era use the mean of the surveyed Decembers within the era. This accords with other evidence of changes to right whale distributions around this time, and with Ganley et al.'s (2019) result showing that 2011 was the first January with relatively high abundance. We believe this decision should be reassessed as additional data become available.

```
df$FinalDecAbund <- df$DecPred5

print(df[,c('Year', 'DecMethod', 'FinalDecAbund','JanAbund')], digits=2)
```

```
##    Year DecMethod FinalDecAbund JanAbund
## 1  2003   Modeled           0.0      0.0
## 2  2004  Surveyed           0.0      7.9
## 3  2005   Modeled           0.0      0.0
## 4  2006   Modeled           0.0      0.0
## 5  2007   Modeled           0.0      2.3
## 6  2008   Modeled           0.0      2.1
## 7  2009   Modeled           0.0      2.2
## 8  2010   Modeled          13.9     66.5
```

```
## 9  2011   Modeled          13.9    121.8
## 10 2012  Surveyed          43.2     87.1
## 11 2013   Modeled          13.9     48.7
## 12 2014  Surveyed          13.4     20.9
## 13 2015  Surveyed           0.0      6.6
## 14 2016  Surveyed           0.0     37.0
## 15 2017   Modeled          13.9     69.7
## 16 2018  Surveyed          34.6     15.2
## 17 2019  Surveyed           0.0     66.7
## 18 2020  Surveyed           5.8    187.9
```

All of the results presented here rest on several important assumptions that cannot be validated without collecting additional data. They must therefore be interpreted and applied in species management decisions cautiously. Our first assumption was that during the years December surveys, the effort available was sufficient to make reliable estimates of abundance. However, over the eight years since 2003 with Decembers surveys, a mean of only 2.25 surveys were conducted. This is less than half of the mean of 5.37 surveys conducted during the months of January-May. Further, the December estimates for all years (regardless of whether December was surveyed) rely on the assumption that the abundance-SPUE relationship for December was the same as that estimated for January by Ganley et al. (2019). Ganley et al. estimated high variability in whale dive times between months, resulting in large monthly differences in availability correction factors used in abundance estimates. If whale dive behavior in December differed substantially from that in January, the abundance-SPUE relationship would be different, and our assumption would be wrong. This assumption could eventually be validated and adjusted in the future if sufficient whale dive data were collected in Decembers of future years and Ganley et al.'s availability analysis was repeated.

Both linear models we explored as possible approaches for predicting unsurveyed Decembers assumed that December's abundance could reasonably be modeled linearly from January's. Both models assumed that the eight years used to fit them constituted a representative sample for the 12 years they were used to predict. Finally, the second model assumed that if January's abundance was zero then December's was also zero. Under these assumptions, that model, as fitted, predicted that January's abundance was always greater than or equal to December's. In the eight years used to fit the model, this was true for seven years, but in 2018, January's abundance was only half of December's.

While it is not possible to collect data for years already passed, these assumptions could be validated and adjusted in the future by conducting surveys in Decembers for future years. If December is a month for which accurate abundance and density estimates for Cape Cod Bay are critical for managing the species, we urge that surveying be performed every December with a level of effort consistent with January-May, such that Ganley et al.'s analysis be expanded to encompass December.

We plan to incorporate the new estimates into Duke's right whale density model. This will result in the release of a new version of that model, which will labeled version 11. We tentatively estimate that mean density for Cape Cod Bay for December for the 2003-2009 era will be 0.000 whales / 100 km^2, vs. 0.050 in the prior models (v8-v10), and for the 2010-2018 era will be 0.397 whales / 100 km^2, vs. 1.282 in the prior models. For the 2010-2018 era, the new mean density estimate for December would therefore be approximately 31% of the prior one. These estimates are tentative; please consult the right whale density model 11 documentation for the final values.

References

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.

Roberts JJ, Best BD, Mannocci L, Fujioka E, Halpin PN, Palka DL, Garrison LP, Mullin KD, Cole TVN, Khan CB, McLellan WM, Pabst DA, Lockhart GG (2016) Habitat-based cetacean density models for the U.S. Atlantic and Gulf of Mexico. Scientific Reports 6: 22615.

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 (Opt. Year 3), Document Version 1.4. Page 142. Duke University Marine Geospatial Ecology Lab, Durham, NC.

Appendix: Density estimates for Duke right whale density model

```
ccsCCBArea <- 4105    # Effective area covered by CCS surveys; provided by L. Ganley

df$FinalDecDensity <- df$FinalDecAbund / ccsCCBArea * 100   # whales / 100 km2
df$JanDensity <- df$JanAbund / ccsCCBArea * 100             # whales / 100 km2

print(df[,c('Year', 'DecMethod', 'FinalDecAbund','JanAbund', 'FinalDecDensity', 'JanDensity')])
```

```
##    Year DecMethod FinalDecAbund   JanAbund FinalDecDensity JanDensity
## 1  2003   Modeled      0.000000   0.000000       0.0000000 0.00000000
## 2  2004  Surveyed      0.000000   7.868352       0.0000000 0.19167727
## 3  2005   Modeled      0.000000   0.000000       0.0000000 0.00000000
## 4  2006   Modeled      0.000000   0.000000       0.0000000 0.00000000
## 5  2007   Modeled      0.000000   2.275324       0.0000000 0.05542811
## 6  2008   Modeled      0.000000   2.115047       0.0000000 0.05152369
## 7  2009   Modeled      0.000000   2.191241       0.0000000 0.05337980
## 8  2010   Modeled     13.852867  66.477254       0.3374633 1.61942153
## 9  2011   Modeled     13.852867 121.763788       0.3374633 2.96623113
## 10 2012  Surveyed     43.204485  87.110689       1.0524844 2.12206307
## 11 2013   Modeled     13.852867  48.748896       0.3374633 1.18754922
## 12 2014  Surveyed     13.441395  20.907876       0.3274396 0.50932707
## 13 2015  Surveyed      0.000000   6.609584       0.0000000 0.16101301
## 14 2016  Surveyed      0.000000  37.032368       0.0000000 0.90212834
## 15 2017   Modeled     13.852867  69.700000       0.3374633 1.69792935
## 16 2018  Surveyed     34.563588  15.150000       0.8419875 0.36906212
## 17 2019  Surveyed      0.000000  66.660000       0.0000000 1.62387333
## 18 2020  Surveyed      5.760598 187.860000       0.1403313 4.57637028
```